

UNIVERSIDAD DE SALAMANCA

ESCUELA DE DOCTORADO

Programa de Doctorado Salud y Desarrollo en los Trópicos



TESIS DOCTORAL

Identificación de genes de baja penetrancia  
implicados en la susceptibilidad a la infección  
por *Schistosoma mansoni*

Juan Hernández Goenaga

2018

UNIVERSIDAD DE SALAMANCA

ESCUELA DE DOCTORADO

Programa de Doctorado Salud y Desarrollo en los Trópicos



VNiVERSiDAD  
D SALAMANCA

CAMPUS DE EXCELENCIA INTERNACIONAL

TESIS DOCTORAL

Identificación de genes de baja penetrancia  
implicados en la susceptibilidad a la infección  
por *Schistosoma mansoni*

Tesis que presenta  
Juan Hernández Goenaga  
para optar al Grado de Doctor

2018



Prof. Dr. Antonio Muro Álvarez, Catedrático de Parasitología de la Universidad de Salamanca, Director de la Unidad de Investigación Consolidada de Castilla y León UIC.079 y Prof. Dr. Julio López Abán, Profesor Titular de Parasitología de la Universidad de Salamanca.

CERTIFICAN:

Que la Tesis Doctoral titulada “Identificación de genes de baja penetrancia implicados en la susceptibilidad a la infección por *Schistosoma mansoni*”, que se presenta para optar al grado de Doctor por la Universidad de Salamanca, ha sido realizada por **Juan Hernández Goenaga**, con D.N.I. 07973519-V, Licenciado en Ciencias Biológicas por la Universidad de Salamanca y Máster en Enfermedades Tropicales, bajo nuestra dirección en el Departamento de Biología Animal, Parasitología, Ecología, Edafología y Química Agrícola de la Universidad de Salamanca y el Centro de Investigación de Enfermedades Tropicales de la Universidad de Salamanca (CIETUS). Consideramos que reúne, a nuestro juicio, los requisitos necesarios por lo que autorizamos su presentación para ser evaluada.

Y para que así conste, a efectos legales, expiden el presente certificado en

Salamanca, a 09 de Julio de 2018

Fdo. Dr. Antonio Muro Álvarez

Fdo. Dr. Julio López Abán

*“La inteligencia es la capacidad  
de adaptarse al cambio”  
Stephen Hawking*

*A mis padres.  
A mi abuela.*

# ¡Gracias!

En primer lugar quiero hacer constar mi más sincero agradecimiento a mis directores, **Antonio Muro Álvarez** y **Julio López Abán**, por vuestro apoyo, las enseñanzas, la dedicación, la confianza y el tiempo invertidos para que esta Tesis Doctoral llegara a buen puerto.

Mi agradecimiento más profundo se lo dedico a mis padres, **José Manuel** y **María Elena**. En estos tiempos que vivimos, sin vuestro apoyo personal y económico y sin vuestro esfuerzo, esta Tesis no se habría podido realizar. **GRACIAS**.

A todas aquellas instituciones y personas que de alguna forma han colaborado en este trabajo.

A **Belén**, mi “madre” en el laboratorio, por tu apoyo, tus enseñanzas, tu confianza y sobre todo por tu compañía y amistad. A los integrantes del **Grupo de Genómica y Proteómica** del CIETUS, gracias por vuestras enseñanzas. A **Fernando Simón**, gracias a tí soy un apasionado de la Parasitología, gracias por tu apoyo y amistad. A **Javier Burguillo**, por tu inestimable ayuda, tu apoyo y tus enseñanzas, sin tí habría sido imposible. A **Pedro Fernández-Soto**, por tus enseñanzas y tu apoyo. A **Rodrigo Morchón**, por tu apoyo, tu ayuda y tu amistad.

A **Mari Ángeles**, por estar siempre que se la necesita, escuchando, aconsejando y apoyando. Simplemente, **gracias**.

Quiero acordarme especialmente de **Luis Pérez del Villar**, que inició el trabajo precioso en que se apoya esta Tesis Doctoral y que tristemente no ha podido ver los resultados, pues la enfermedad contra la que luchaste durante cinco años te venció, a pesar de tu extraordinario espíritu de lucha y calidad humana.

Gracias a **Edward** por tus enseñanzas, ideas, predisposición y tus consejos. A **Carmen**, por tus enseñanzas, consejos y amistad. A los integrantes del **Laboratorio 7** del Centro de Investigación del Cáncer, por vuestra inestimable ayuda en el proyecto del genotipado.

A mis amigos y mi familia, por vuestro apoyo incondicional. Finalmente, al resto de compañeros con los que a lo largo de estos años de desarrollo de este trabajo, compartí horas de laboratorio, alegrías por los resultados y amistad: **José Vicente, Ana Lucya, Alicia, Javier, Alba** y tantos y tantos otros que me resultaría imposible nombrarlos a todos. A lo largo de estos años me habéis ayudado y facilitado la realización de este trabajo, **GRACIAS A TODOS:**

## Índice de contenidos

1	Introducción.....	10
1.1	Biología de los esquistosomas	12
1.2	Datos sobre el genoma, transcriptoma y proteoma de los esquistosomas	15
1.3	Características epidemiológicas de la esquistosomosis	17
1.3.1	Esquistosomosis en zonas endémicas	18
1.3.2	Esquistosomosis importada	19
1.3.3	Esquistosomosis emergente en el sur de Europa	20
1.4	Mecanismos patogénicos	20
1.4.1	Mecanismos de agresión	20
1.4.2	Mecanismos de defensa	22
1.4.3	Mecanismos de evasión	24
1.5	Manifestaciones clínicas	25
1.5.1	Fase de inicio	25
1.5.2	Fase aguda	26
1.5.3	Fase crónica	26
1.6	Métodos de diagnóstico	29
1.6.1	Métodos parasitológicos	29
1.6.2	Métodos inmunológicos	31
1.6.3	Métodos moleculares	31
1.7	Tratamiento y prevención	32
1.8	Susceptibilidad/resistencia a la infección por esquistosoma	34
1.9	Análisis de fenotipos intermedios	35
2	Hipótesis y objetivos .....	37
3	Materiales y Métodos .....	40
3.1	Animales	41
3.2	Mantenimiento experimental del ciclo biológico de <i>S. mansoni</i>	41
3.3	Generación de una cohorte con heterogeneidad genética y fenotípica mediante retrocruce o <i>backcross</i>	43
3.3.1	Generación del <i>backcross</i>	43
3.4	Evaluación de la infección por <i>Schistosoma mansoni</i> en la cohorte F2BX	44
3.5	Recogida de magnitudes parasitológicas	45
3.5.1	Recuperación de vermes mediante perfusión	45
3.5.2	Determinación de la carga de huevos en hígado e intestino y cálculo de la fecundidad de las hembras	46
3.6	Magnitudes patológicas: recuento de granulomas y estimación de la superficie hepática afectada	47
3.7	Estudio de poblaciones celulares por citometría de flujo	47

3.7.1	Preparación de los leucocitos de sangre periférica para citometría de flujo	49
3.7.2	Preparación de esplenocitos para determinar poblaciones con citometría de flujo	49
3.8	Respuesta inmune humoral por ELISA	50
3.8.1	Obtención de antígeno somático de vermes adultos de <i>Schistosoma mansoni</i> (SoSmAg)	50
3.8.2	Detección de anticuerpos anti-Sm en ratones infectados con el parásito	50
3.9	Genotipado	51
3.9.1	Obtención y cuantificación de DNA de <i>S. mansoni</i>	51
3.9.2	Genotipado de los ratones <i>backcross</i> F2BX infectados con <i>S. mansoni</i>	51
3.9.3	Análisis de ligamiento e identificación de regiones de rasgo cuantitativo (QTL).	52
3.10	Análisis de datos	54
3.10.1	Exploración de datos	54
-	Estadística descriptiva	54
-	Correlaciones entre variables	54
3.10.2	Análisis multivariante	55
-	Análisis de conglomerados (Cluster analysis (CA))	55
-	Análisis de componentes principales (Principal Component Analysis (PCA))	56
-	Conglomerados por k-medias y PCA (K-means clusters and PCA (KM-PCA))	57
-	Regresión múltiple lineal por mínimos cuadrados	58
-	Regresión logística binaria	58
4	Resultados .....	60
4.1	Susceptibilidad a la infección por <i>Schistosoma mansoni</i> en la cohorte F2BX	61
4.1.1	Análisis exploratorio de los datos recopilados de la cohorte F2BX por sexos	62
4.1.2	Análisis descriptivo de las variables analizadas en la cohorte F2BX en conjunto	64
4.1.3	Análisis de correlación de variables en la cohorte F2BX	70
4.2	Análisis multivariante de los datos obtenidos con la cohorte F2BX	73
4.2.1	Análisis de conglomerados o clusters ( <i>Cluster Analysis (CA)</i> )	74
4.2.2	Análisis de componentes principales ( <i>Principal Component Analysis (PCA)</i> )	76
4.2.3	Estudio de conglomerados por k-medias y PCA ( <i>k-means clusters and PCA</i> )	77
4.3	Variación de las variables fenotípicas entre los cuatro grupos de infección	80
4.3.1	Variación entre grupos para el número de vermes totales, de vermes macho y de vermes hembra	81
4.3.2	Variación entre grupos para el número de huevos por gramo de hígado e intestino, el número de granulomas y la superficie afectada del hígado	83
4.3.3	Variación entre grupos para poblaciones celulares en sangre periférica.	84
4.3.4	Variación entre grupos para poblaciones de esplenocitos	85
4.3.5	Variación entre grupos para inmunoglobulinas en las semanas cero y nueve	86
4.3.6	Variación entre grupos para inmunoglobulinas en la semana nueve.	87
4.3.7	Análisis de correlación de las variables en los cuatro grupos de la cohorte F2BX.	89
4.4	Identificación de regiones cromosómicas asociadas a susceptibilidad y resistencia	95
4.4.1	Genotipado de la cohorte F2BX y control de calidad de los datos	95
4.4.2	Identificación de regiones genómicas asociadas a los patofenotipos.	100
4.4.3	Estudio de la ligación del daño hepático y QTL en los cromosomas 15 y 18	108

- Localización del SNP rs6169611 del cromosoma 15 de ratón	108
- Localización del SNP rs13483259 del cromosoma 18 de ratón	109
4.4.4 Identificación de regiones genómicas ligadas a la susceptibilidad a la enfermedad.	111
4.5 Análisis multivariante por regresión logística binaria como método de validación de la formación de los grupos extremos de enfermedad	113
4.6 Estudios de regresión múltiple lineal (RML) de variables de daño hepático en función de los SNPs y de otras variables fenotípicas para el total de ratones	116
4.6.1 Número de granulomas en función de las inmunoglobulinas, de las poblaciones celulares y el SNP rs6169611 del cromosoma 15	116
4.6.2 Número de granulomas en función de las inmunoglobulinas, de las poblaciones celulares y el SNP rs13483259 del cromosoma 18	118
4.7 Estudios de regresión múltiple lineal de variables de daño hepático en función de los SNPs y otras variables fenotípicas para los grupos extremos de enfermedad	120
4.7.1 Número de granulomas en función de las inmunoglobulinas, de las poblaciones celulares y el SNP rs6169611 del cromosoma 15	120
- Resultados para el Grupo 1 (menos grave)	120
- Resultados para el Grupo 4 (más grave)	122
4.7.2 Número de granulomas en función de las inmunoglobulinas, de las poblaciones celulares y el SNP rs13483259 del cromosoma 18	123
- Resultados para el Grupo 1 (menos grave)	123
- Resultados del Grupo 4 (más grave)	124
5 Discusión.....	125
6 Conclusiones .....	133
7 Bibliografía.....	135

# 1 Introducción

La esquistosomosis es una enfermedad parasitaria, englobada dentro de las enfermedades olvidadas (NTD: *Neglected Tropical Diseases*) por la OMS (Organización Mundial de la Salud). Es endémica de las zonas tropicales y es la tercera parasitosis más importante a nivel mundial en términos de morbilidad y mortalidad global, detrás de la malaria y las helmintosis intestinales (Lewis et al., 2014; King, 2015).

Los parásitos son organismos eucariotas que viven en el interior o sobre la piel de otro organismo denominado hospedador. El término parásito proviene etimológicamente del griego (παρά, junto a y σίτος, trigo, comida). La generalización biológica de su nombre tiene su origen en los médicos veterinarios romanos, que denominaban así a los animales que vivían y se alimentaban a expensas de otros. Se empleaba este término para diferenciarlos de los animales depredadores, pues aunque hay asociación con el hospedador para obtener alimento y conservar su vida, el hospedador no obtiene ningún beneficio, más bien al contrario. Estos conceptos se han ido completando mediante el estudio de la naturaleza de dicha asociación, el carácter fisiológico de la misma, los mecanismos mediante los cuales los parásitos son capaces de manipular el genoma de su hospedador, llegando a provocar cambios en los núcleos de sus células, alterando su expresión génica a favor del parásito (Cordero y Rojo, 1999). Actualmente, los estudios realizados sobre el DNA (ácido desoxirribonucleico), secuencias de genes, SNPs (*Single Nucleotide Polymorphisms*) y microsatélites, sirven para estudiar la influencia de la genética en la biología del parásito (Knight et al., 2014).

La gran mayoría de los parásitos pertenecen a los *phylum*: *Apicomplexa*, *Microspora*, *Ciliophora*, *Platyhelminthes*, *Nematoda* y *Arthropoda*. Los helmintos se encuentran en dos de ellos *Platyhelminthes* y *Nematoda*. Según las estimaciones, entre 75000 y 300000 especies de helmintos infectan a los vertebrados terrestres y acuáticos; 287 son conocidos como parásitos de seres humanos, el 95% de los cuales son especies zoonóticas, transmitidas de los animales al hombre o viceversa y aproximadamente el 25% de la población está infectada con estos organismos (Cordero y Rojo, 1999). La gran mayoría de los platelmintos parásitos pertenecen a las clases *Monogenea*, *Digenea* y *Cestoda*, las dos

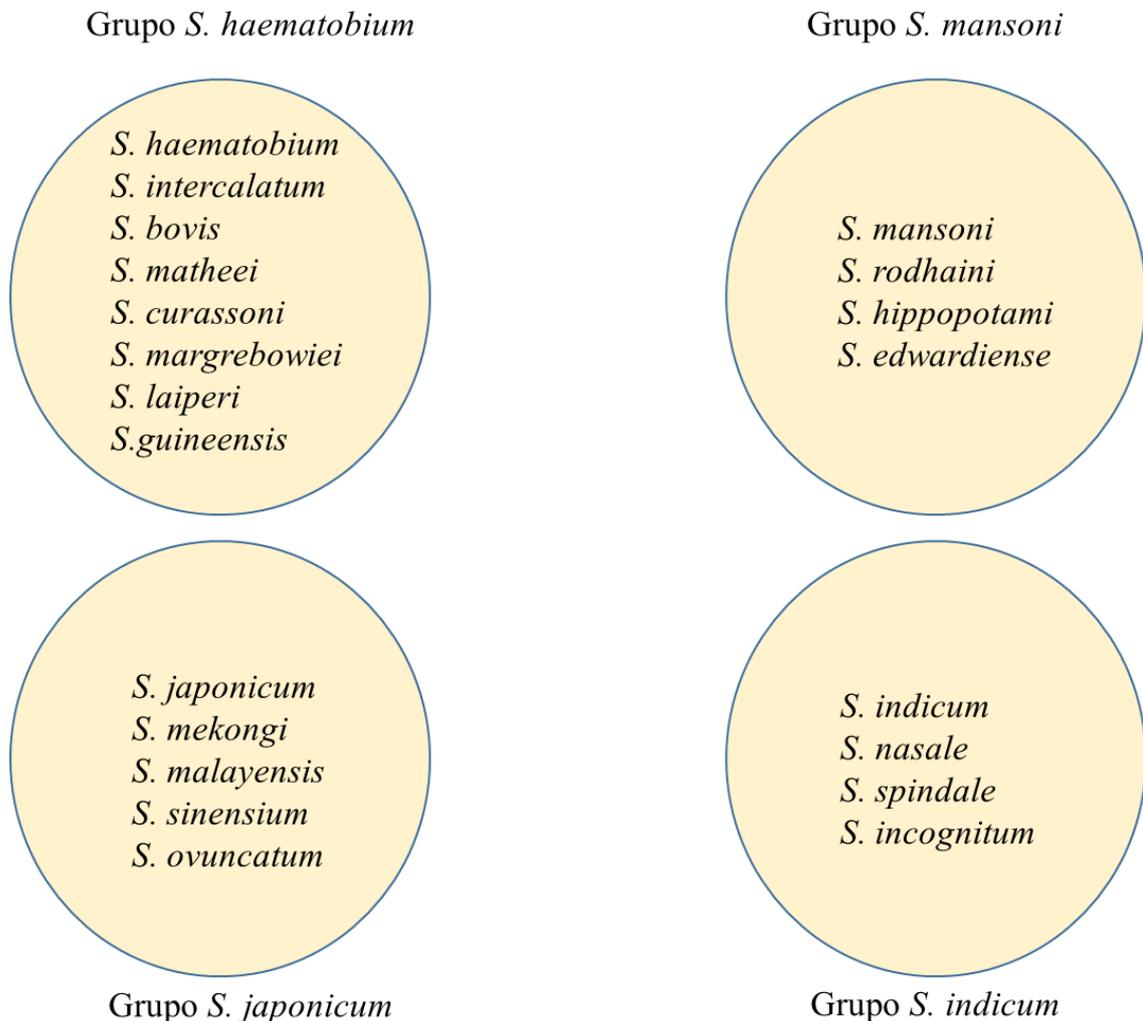
primeras son conocidas como “trematodos”. Todos los trematodos parásitos de seres humanos pertenecen a la clase *Digenea*, tienen complejos ciclos de vida, que incluyen uno o varios hospedadores intermediarios. Las especies del género *Schistosoma* incluidas dentro del *Phylum Platyhelminthes*, clase *Trematoda*, subclase *Digenea*, familia *Schistosomatidae*. Hasta el momento hay descritas 21 especies, clasificadas en cuatro grupos (Figura 1.1), según su distribución geográfica, la morfología del huevo y el hospedador intermediario (Rollinson y Southgate, 1987). La taxonomía y filogenia de *Schistosoma* está en continua revisión debido al gran número de estudios moleculares que se están llevando a cabo (Crellen et al., 2016). De esas 21 especies, hay cinco principales parásitos de seres humanos: *S. mansoni*, *S. haematobium*, *S. intercalatum*, *S. japonicum* y *S. mekongi*, aunque están descritas infecciones por *S. malayensis*, *S. mattheei* y *S. guineensis* (Webster et al., 2006). Uno de estos parásitos, *Schistosoma mansoni*, es objeto de esta tesis doctoral.

Este trabajo está dirigido a estudiar la susceptibilidad/resistencia a la infección por *Schistosoma mansoni* mediante el análisis de fenotipos intermedios. Por ello, en primer lugar se realizará una revisión sobre la esquistosomosis, donde se abordarán los estudios que se han llevado a cabo sobre susceptibilidad y resistencia a esta infección. Posteriormente, se analizará el valor del estudio de los fenotipos intermedios para entender la patogenia de diferentes enfermedades.

## 1.1 Biología de los esquistosomas

Los esquistosomas son trematodos digeneidos y difieren del resto por presentar dimorfismo sexual. El significado de su nombre es “cuerpo dividido”, que hace referencia al canal ginecóforo del macho, que plegando sus bordes laterales envuelve a la hembra, de morfología más estrecha y alargada (Figura 1.2.F) (Lewis et al., 2014). Necesitan dos hospedadores para completar su ciclo: un hospedador intermediario, caracol de agua dulce, en el cual se produce la reproducción asexual y un hospedador definitivo, donde se produce la reproducción sexual (Gryseels et al., 2006; Gryseels, 2012). Cada esquistosoma tiene una especie distinta de caracol como hospedador intermediario. *S. haematobium*: *Bulinus*, *Physopsis*, *Planorbis*. *S. mansoni*: *Biomphalaria*. *S. japonicum*: *Oncomelania*. *S. mekongi*: *Tricula*. *S. intercalatum*: *Bulinus*. Las personas adquieren la esquistosomosis al entrar en contacto directo con agua contaminada con las larvas natatorias del parásito -cercarias-.

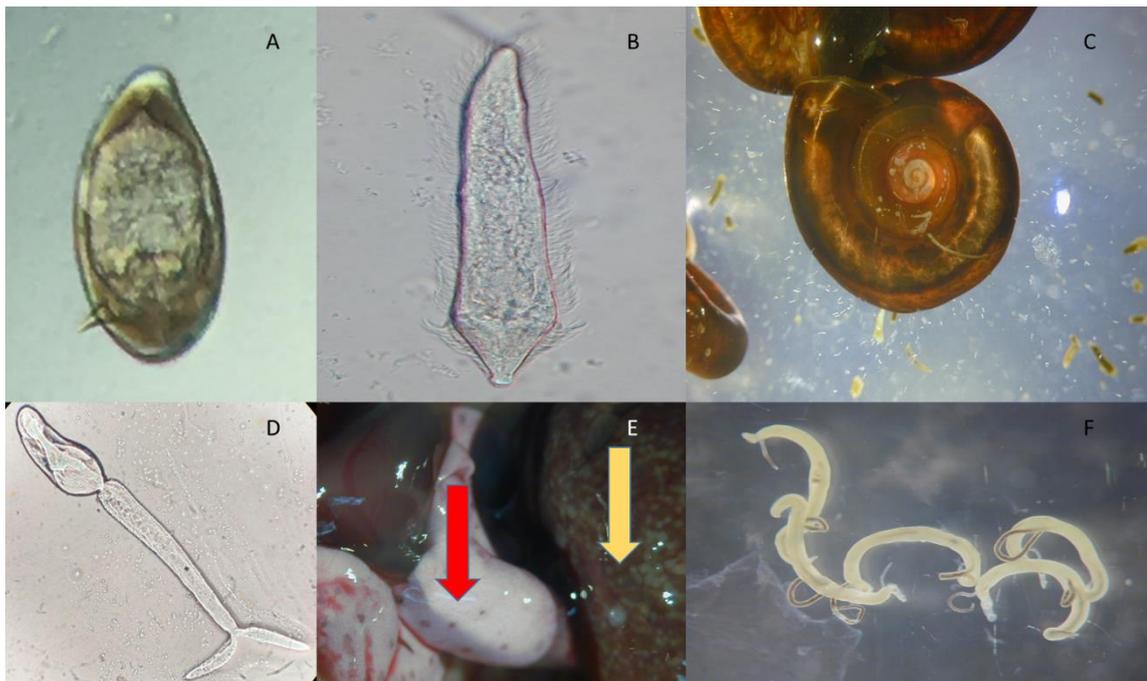
Estas cercarias son emitidas por distintas especies de caracoles acuáticos y penetran activamente por la piel del individuo.



**Figura 1.1.** Grupos y especies de *Schistosoma* (Webster et al., 2006).

Tras la penetración de la cercaria a través de la piel, ésta pierde la cola transformándose en una nueva fase denominada esquistosómula, que realiza una migración necesaria para su pre-maduración. En ese recorrido llegarán a los pulmones, para volver a la circulación general y finalmente dirigirse a los vasos de diferentes localizaciones, donde terminarán su maduración. Los vermes adultos de *S. mansoni*, *S. japonicum*, *S. mekongi* y *S. intercalatum* se asientan en las venas mesentéricas y la vena porta, mientras que los de *S. haematobium* se dirigen a los vasos que drenan la vejiga urinaria (plexo vesical). Estos

parásitos pueden permanecer en estas localizaciones durante 3-10 años, aunque se han descrito casos en los que han sobrevivido hasta 40 años (Chabasse et al., 1985). Una vez que se aparean macho y hembra, tiene lugar la reproducción y consiguiente puesta de huevos, los cuales son capaces de atravesar la pared vascular, entrar en el lumen vesical o intestinal y son expulsados con las heces (*S. mansoni*, *S. japonicum*, *S. mekongi* y *S. intercalatum*) o con la orina (*S. haematobium*). Desde la penetración de las cercarias hasta la liberación de huevos transcurren unas seis semanas (Gryseels et al., 2006). Los huevos de los esquistosomas carecen de opérculo y poseen una “espina” que varía de tamaño y posición, dependiendo de la especie de *Schistosoma* a la que pertenezca. La morfología y la localización de la “espina” son rasgos determinantes para identificar la especie. Una representación de las fases de la biología del ciclo de *S. mansoni* se muestra en la Figura 1.2.



**Figura 1.2.** Biología de *Schistosoma mansoni*. A) Huevo. B) Miracidio. C) Caracoles *Biomphalaria glabrata* emitiendo cercarias. D) Cercaria. E) Necropsia de ratón infectado, flecha roja: marca del paso de las esquistosómulas por los pulmones; flecha amarilla: granulomas en el hígado F) Parejas de adultos, machos más claros y hembras más oscuras, en el canal ginecóforo de los machos.

No todos los huevos producidos por las hembras, tras ser fecundadas por los machos, se liberan en la excreción de la orina o las heces, un porcentaje queda atrapado en diferentes tejidos formando granulomas, que son los causantes del daño originado en la fase crónica de la enfermedad. Al ser liberados al medio acuático, los huevos ya embrionados,

eclosionan y liberan la fase larvaria, denominada miracidio, que tiene capacidad natatoria y busca al hospedador intermediario, el caracol acuático. Dentro del caracol se desarrolla una generación de esporocisto I y otra de esporocisto II, que son reproducciones asexuales y que concluyen con el desarrollo y la emisión de las cercarias al medio acuático. Los hospedadores intermediarios actúan como amplificadores de la enfermedad, debido a que liberan cientos de cercarias de forma continuada (Stensgaard et al., 2013).

## 1.2 Datos sobre el genoma, transcriptoma y proteoma de los esquistosomas

En 1994, la Organización Mundial de la Salud (OMS) y *Schistosome Genome Network* dieron inicio al proyecto para secuenciar el genoma de *S. mansoni*, con el objetivo de conocer la expresión cuantitativa de los genes potenciales como dianas para diseñar vacunas y fármacos. Finalmente, en 2009 se describe la secuencia completa y se realiza el mapeo del genoma de *S. mansoni* (Berriman et al., 2009), también se anota por primera vez el genoma de *S. japonicum* (Zhou et al., 2009). En 2012 se describe el genoma de *S. haematobium* (Young et al., 2012).

El tamaño del genoma de *Schistosoma* es mayor al de la mayoría de patógenos que han sido secuenciados hasta el momento (Tabla 1.1) (Young et al., 2012). El 43% del genoma ha sido mapeado empleando la técnica de hibridación in situ (FISH) en 8 pares de cromosomas, de los cuales siete son autosomas y el octavo es un cromosoma sexual. La naturaleza repetitiva del genoma ha hecho que sea poco práctica la hibridación FISH, por ello actualmente se emplea una segunda estrategia, el mapeo HAPPY, que proporciona un equivalente in vitro de un mapa genético que puede generar un mapa físico midiendo la frecuencia de los cortes inducidos entre los marcadores (LoVerde et al., 2004). La mayoría de los genes descritos presentan una estructura con grandes intrones y con exones mucho más cortos. Es importante resaltar el gran contenido de secuencias repetitivas: 45% en *S. mansoni*, 47.2% en *S. haematobium* y 40.1% en *S. japonicum* (Tabla 1.1), como hecho importante desde el punto de vista de la relación evolutiva y filogenética con el hospedador (Crellen et al., 2016; Young et al., 2012).

El transcriptoma de *S. mansoni* se describió siguiendo la metodología de estudio multicéntrico en Brasil (Verjovski-Almeida et al., 2003), empleando lo que se conoce como EST (*Expressed Sequence Tags*), partiendo de bibliotecas de cDNA obtenidas mediante RT-PCR (*real time polymerase chain reaction*) y amplificando la región codificante de los

RNA mensajeros. Gracias al desarrollo de nuevas técnicas de secuenciación, como el RNA-seq (Mortazavi et al., 2008), se están estudiando el transcriptoma de las distintas fases del ciclo biológico del parásito (Protasio et al., 2012), como la de los vermes adultos macho (Almeida et al., 2012) y hembra (Anderson et al., 2015). Se ha estudiado el genoma de *S. japonicum*, comparándolo con el genoma del ser humano y se ha encontrado similitud en 10 secuencias entre ambos genomas, de entre el 25 hasta el 97% (Yu et al., 2008). Con esto y los resultados de los estudios de RNA-seq, parece evidente que el parásito establece una relación con su hospedador, empleando para su beneficio las señales inmunes de éste.

**Tabla 1.1.** Características principales de los genomas de *S. mansoni*, *S. haematobium* y *S. japonicum*. Adaptada de Young et al, 2012.

Características del genoma	<i>S. mansoni</i>	<i>S. haematobium</i>	<i>S. japonicum</i>
Tamaño aproximado del genoma (Mb)	381	385	403
Número de cromosomas (2n)	8	8	8
Proporción codificante del genoma (%)	4.72	4.43	4.32
Número atribuido de genes	13184	13073	13469
Tamaño de los genes (media de pb $\pm$ s.d.; rango)	13397 $\pm$ 18029; 84-240193	11952 $\pm$ 16273; 30-200000	10003 $\pm$ 12980; 150-173394
Longitud media de dominios codificantes (media de pb $\pm$ s.d.; rango)	1344 $\pm$ 1447; 60-22983	1319 $\pm$ 1502; 30-28212	1179 $\pm$ 1201; 147-24180
Número de exones por gen (media de pb $\pm$ s.d.; rango)	6.2 $\pm$ 6.24 ; 1-94	5.4 $\pm$ 5.8; 1-136	5.3 $\pm$ 4.7; 1-65
Longitud de los exones (media de pb $\pm$ s.d.; rango)	218 $\pm$ 236; 1-9291	246 $\pm$ 287; 1-9737	223 $\pm$ 256; 6-6326
Longitud de los intrones (media de pb $\pm$ s.d.; rango)	2331 $\pm$ 3200; 1-67221	2442 $\pm$ 2958; 1-68754	2058 $\pm$ 2679; 15-59770
Contenido en GC (%)	34.7	34.3	33.5
Grado de repetición (%)	45	47.2	40.1

El término proteoma aparece en la década de los 90, se emplea para describir el total de proteínas expresadas por un genoma (Wilkins et al., 1996). Para caracterizar el proteoma de parásitos se emplean la electroforesis de dos dimensiones (2D) y la espectrometría de masas. Se han llevado a cabo múltiples estudios del proteoma de las distintas fases del ciclo del parásito (Sotillo et al., 2017; Dríguez et al., 2016). Las proteínas de mayor interés son las que interactúan con las del sistema inmune del hospedador (Ludolff et al., 2014), como futuras dianas para crear nuevos fármacos o diseñar vacunas contra la enfermedad. Se diferencian tres grandes grupos de proteínas: (i) *proteínas de superficie propias del parásito*: estructurales, transportadoras de nutrientes, con actividad enzimática y receptores de ligandos con el hospedador; (ii) *proteínas de superficie propias del hospedador*, el parásito se rodea de moléculas del hospedador para evadir la respuesta inmune, entre otras hay inmunoglobulinas y proteínas del complemento como C3 (Braschi y Wilson, 2006); (iii) *proteínas de excreción-secreción*, como proteasas y proteínas de fuerte carácter inmunomodulador, que se correlacionan con funciones propias del parásito como la penetración a través de la piel (Portet et al., 2017; Da'dara et al., 2017) o con la propia patogénesis de la enfermedad, como es la infiltración granulomatosa alrededor de los huevos en los diferentes tejidos (Cass et al., 2007).

Finalmente, señalar que se han realizado estudios concretos analizando el glicoma (van Diepen et al., 2012; Smit et al., 2015), inmunoma (Mc William et al., 2012; Gaze et al., 2014) y filoma (Silva et al., 2012) de estos parásitos. Se ha avanzado en el conocimiento de la relación parásito-hospedador, ya que el estudio de los diferentes perfiles de expresión génica ha ayudado a identificar posibles marcadores celulares que interactúan con *S. mansoni* durante la infección (Adema et al., 2017).

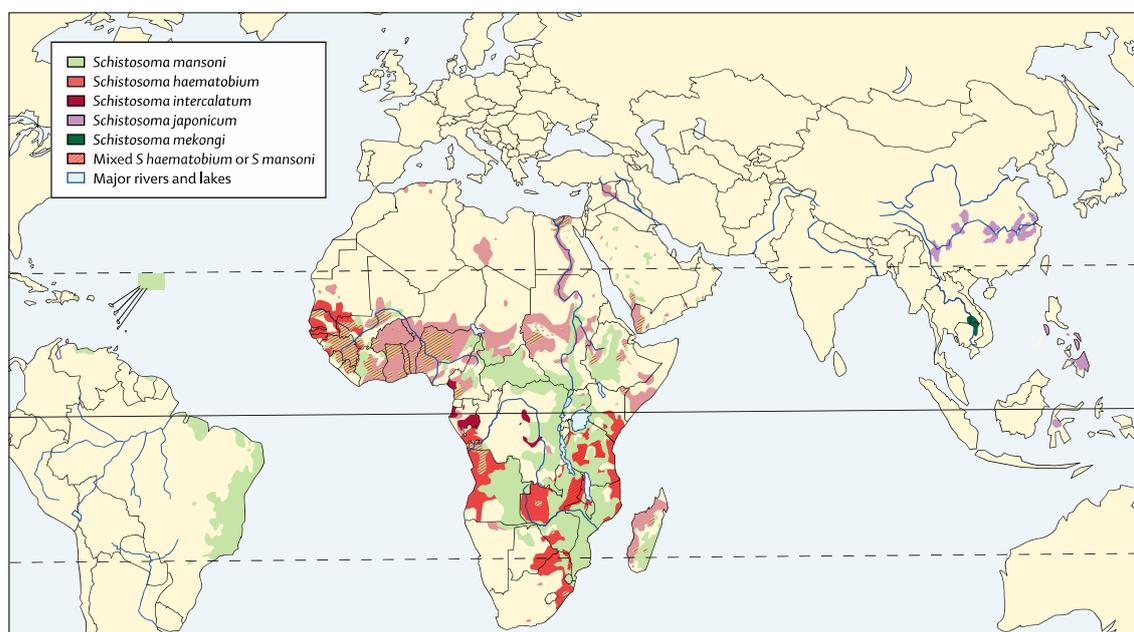
### **1.3 Características epidemiológicas de la esquistosomosis**

La esquistosomosis es una de las enfermedades más prevalentes en las zonas tropicales y subtropicales del planeta, afectando seriamente a la salud global y causando un grave impacto económico y social en los países en desarrollo (Wright et al., 2018). Hay una estimación de alrededor de 190 millones de personas infectadas en 78 países diferentes (la mayoría en el África subsahariana) (GBD Disease and Injury Incidence, 2016) y la mortalidad anual supera las 10000 personas (GBD Causes of Death, 2016). En las áreas endémicas, la mayor prevalencia se muestra en adolescentes (60-80% de infección activa en niños en edad escolar), disminuyendo generalmente en la edad adulta (20-40%) (Colley

y Secor, 2014). Más que por su mortalidad, hay autores que señalan el gran impacto de la morbilidad causada por la esquistosomosis, que se refleja en una pérdida de 1.9 millones de DALYs (incapacidad ponderada por año de vida) (GBD DALYs, 2016).

### 1.3.1 Esquistosomosis en zonas endémicas

Respecto a su distribución geográfica, en el continente americano sólo existe *S. mansoni*; en África conviven *S. haematobium*, *S. mansoni*, *S. intercalatum* y *S. guineensis* y en Asia habitan *S. japonicum* y *S. mekongi* (Figura 1.3).



**Figura 1.3.** Distribución mundial de los países con esquistosomosis humana. (Colley y Secor, Lancet 2014).

De manera global, las especies más importantes son *S. haematobium* y *S. mansoni*, ya que acumulan más del 90% de los casos de la enfermedad en el mundo (WHO, 2017). *S. haematobium* acumula más de la mitad de los casos reportados de la enfermedad y es el único agente causal de la esquistosomosis urogenital. Por otro lado, *S. mansoni* es la especie más importante en términos de morbilidad, así como la más ampliamente distribuida por el planeta (King, 2010).

Existen una serie de condicionantes para que se produzca la transmisión de la esquistosomosis (Grimes et al., 2015). El primero y más importante es que se encuentren

poblaciones de caracoles susceptibles del mantenimiento del ciclo biológico y transmisión de la enfermedad. Estos hospedadores intermediarios tienen estrategias para evadirse de la sequía, la contaminación y los cambios climáticos. Las poblaciones de caracoles tienden a ser estables, por lo que la enfermedad presenta un carácter endémico y local, pero si una especie de caracol presenta resistencia natural a la infección, la tasa de transmisión bajará (Carvalho et al., 2008). El cambio en las condiciones hidrográficas, como son las inundaciones o la construcción de grandes presas, pueden afectar a la expansión de la infección. Esto es debido a que los caracoles pueden proliferar en nuevas áreas no endémicas de esquistosomosis (Yang et al., 2018; Diakit  et al., 2017). El segundo es su asociación a la falta de desarrollo, es decir a la ausencia de ba os y letrinas, ya que para que se mantenga el ciclo de vida del par sito se necesitan aguas contaminadas con huevos expulsados por el par sito con las heces o la orina. La exposici n al agua contaminada con cercarias se asocia a actividades diarias como el lavado de utensilios, de la ropa o el ba o (por ello los ni os que viven en zona end mica son m s propensos a infectarse y reinfectarse), as  como las actividades dom sticas y las relacionadas con la agricultura, la pesca o la ganader a (King, 2010). El cambio demogr fico del campo a la ciudad no ha eliminado la enfermedad, puesto que en las periferias de las grandes ciudades de Suram rica y  frica contin a la transmisi n de la esquistosomosis (Barbosa et al., 2010; Ugbomoiko et al., 2010).

### 1.3.2 Esquistosomosis importada

Seg n los datos de las bases epidemiol gicas de *International Society of Travel Medicine* y de *Centers for Disease Control and Prevention (GeoSentinel)*, as  como del *European Network on Imported Infectious Disease Surveillance (TropNetEurop)* de 2008, es la segunda enfermedad importada m s prevalente, despu s de la malaria. La mayor a de los casos provienen de  frica occidental, Mali, Costa de Marfil, Rep blica Democr tica del Congo y Uganda. Las regiones tur sticas tienen un alto riesgo de infecci n en viajeros, como el pa s Dogon en Mali, Banfora en Burkina Faso, la orilla sur del Lago Malawi y el Parque Nacional de Omo en Etiop a (Coltart et al., 2015).

Actualmente y a causa de la movilidad internacional se ha producido un incremento en el n mero de casos importados de la enfermedad. En pa ses donde la enfermedad no es end mica, aparecen con mayor frecuencia pacientes a los que se diagnostica esta enfermedad, son viajeros a zonas rurales y perirurales de  reas end micas, sobre todo los

que realizan turismo *off-track* o de aventura (Lingscheid et al., 2017), expatriados por motivos laborales o guerras (Abou-Zeid et al., 2012) y por la emigración sur-norte. En dos estudios sobre las causas de eosinofilia realizados a inmigrantes subsaharianos se demostró que el 17% de los pacientes con eosinofilia absoluta tenían esquistosomosis (Pardo et al., 2006) y que esta infección era la causa principal de enfermedad entre los que mostraban eosinofilia relativa (Carranza-Rodríguez et al., 2008).

### **1.3.3 Esquistosomosis emergente en el sur de Europa**

Estos fenómenos son importantes a nivel epidemiológico, puesto que en Europa estaba considerada como una enfermedad importada, pero en la actualidad se puede catalogar de emergente, ya que se han dado casos de infección introducida. En 2014 se denuncia el primer caso de transmisión autóctona de esquistosomosis: una niña de 4 años con hematuria, fue llevada a consulta médica en Toulouse, Francia. No había viajado a zona endémica, pero después de varios exámenes aparecieron huevos de esquistosoma alojados en la pared de la vejiga (Berry et al., 2014). Aparecen más de 120 casos en Francia, Alemania e Italia entre residentes y turistas, que parecen haber contraído la enfermedad en Córcega. En la isla hay caracoles del género *Bulinus*, que pueden actuar como intermediarios de las especies del grupo *S. haematobium*. Se realizaron estudios moleculares, con los que se vió que las especies causantes eran *S. haematobium*, híbridos *S. haematobium-S. bovis* y *S. bovis*, todos del grupo *S. haematobium*. Un análisis más exhaustivo de las secuencias de DNA llevó a mostrar la estrecha relación de los parásitos encontrados en la isla y los procedentes de África occidental. Así, la aparición de la enfermedad en Europa se asocia con la llegada de inmigrantes de Senegal, que estaban infectados de esquistosomosis y activaron el ciclo biológico del parásito al orinar en el río (Boissier et al., 2016).

## **1.4 Mecanismos patogénicos**

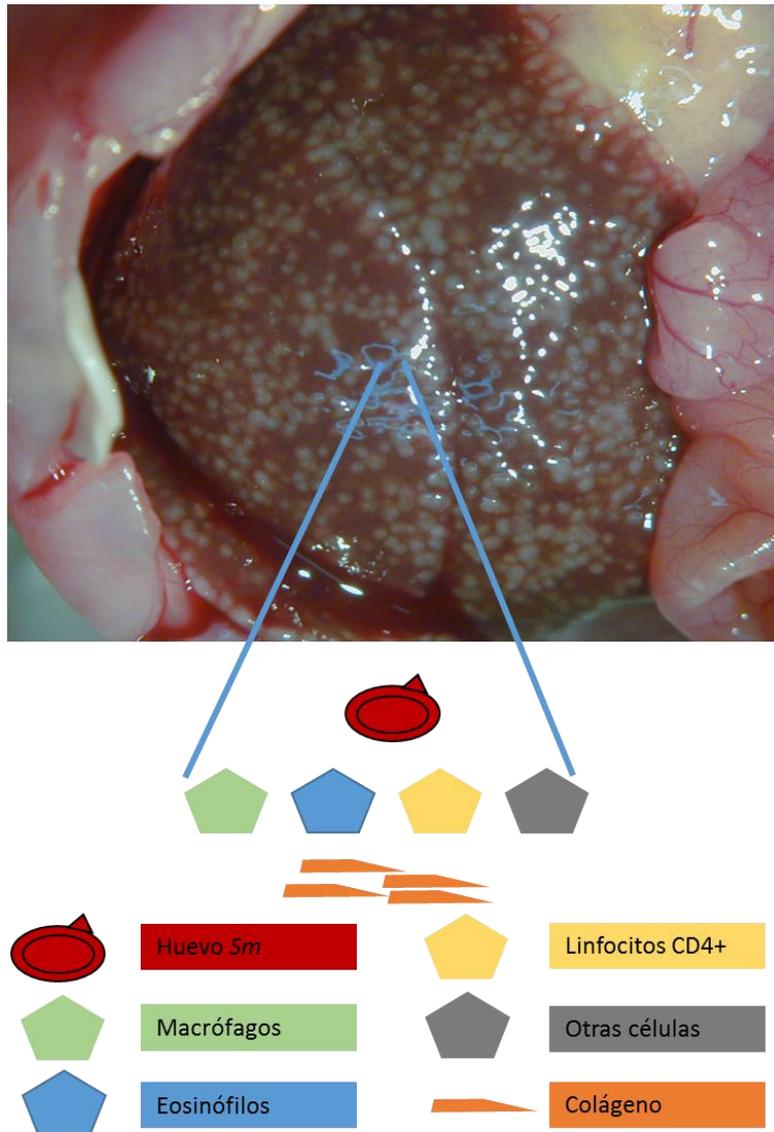
### **1.4.1 Mecanismos de agresión**

Las actividades del parásito, por las que ejerce su acción patógena se pueden dividir en varias fases: la fase de penetración cercariana, la fase de migración de la esquistosómula y la fase de formación del granuloma alrededor del huevo.

La fase de penetración cercariana es un proceso complejo y está mediado por estímulos químicos por parte del hospedador a la vez que se produce la liberación de proteasas secretadas por la cercaria. Dichas proteasas están compuestas en su mayoría por una serin-proteasa de 30 kDa, con actividad elastasa, que facilita la penetración de la cercaria al degradar la elastina de la piel. Diversos estudios realizados sobre la piel demuestran que los productos de excreción-secreción cercarianos inducen la formación de edema e infiltración de neutrófilos, además de ser capaces de degradar otras macromoléculas presentes en la piel del hospedador como la queratina, la fibronectina, la laminina o el colágeno. Sin embargo, esta reacción inflamatoria no induce la protección del individuo, sino que favorece la supervivencia del parásito (Mountford y Trottein, 2004).

En la fase de migración de la esquistosómula, ésta migra a través de la circulación sanguínea durante las 4-6 semanas siguientes a la infección, hasta que llega a su lugar de maduración a adulto. Se producen altos niveles de citocinas pro-inflamatorias (principalmente  $\text{TNF}\alpha$ , IL-1, IL-6 e  $\text{IFN}\gamma$ ) cuya máxima expresión se produce a la decimosexta semana post-infección. Esta situación se asocia a una respuesta Th1 predominante, responsable de la denominada fiebre de Katayama (Caldas et al., 2008).

La fase de formación del granuloma alrededor del huevo comienza con la puesta de huevos por parte de los adultos, a partir de la séptima semana post-infección. Esto es debido a que una gran cantidad de huevos no son liberados al exterior, quedando atrapados en el hígado, intestino o vejiga urinaria e induciendo una respuesta inmunológica Th2 y posteriormente Th17, que es la responsable de la patogenia de la fase crónica de la enfermedad. Alrededor de esta fase se produce un infiltrado celular, compuesto por macrófagos, eosinófilos, linfocitos CD4+ y colágeno. Este infiltrado da lugar a la formación de los granulomas (Figura 1.4), causantes de la patogenia de la enfermedad (Pearce y McDonald, 2002).



**Figura 1.4.** Hígado de ratón con granulomas de *S. mansoni*. Esquema de la composición del infiltrado celular alrededor del huevo del parásito (granuloma).

### 1.4.2 Mecanismos de defensa

La mayor parte de los estudios sobre la respuesta inmunológica se han llevado a cabo en personas infectadas de área endémica o empleando modelos animales.

En cuanto a la respuesta inmunitaria en seres humanos de área endémica de la enfermedad, han manifestado diferentes patrones inmunológicos frente a antígenos derivados del adulto o del huevo (Colley y Secor, 2014). En la mayoría de los estudios frente al antígeno de huevo, se ha observado elevada respuesta en las fases tempranas y disminución en fase crónica. En contraste con las respuestas originadas frente al antígeno de adulto, que se mantienen a pesar de que la infección se haga crónica. Hay estudios que señalan que la exposición continuada a antígenos de huevo a causa de reinfecciones, induce

mecanismos de regulación inmunológica que provocan que muchos de estos pacientes no desarrollen manifestaciones clínicas graves (Maizels y Yazdanbakhsh, 2003). La interpretación de los datos se adecúa a las características de las personas en estudio, dependiendo del momento de la adquisición de la infección y al hecho de haber sido tratado o no con praziquantel (fármaco de elección de la esquistosomosis). Así, la forma grave de la enfermedad se asocia con un incremento de la respuesta Th-17, con la participación de citocinas como IL-23 e IL-17 (Rutitzky et al., 2008). Se ha observado la asociación entre producción de anticuerpos, eosinófilos y resistencia a la infección en diferentes estudios epidemiológicos. No obstante, la susceptibilidad a la reinfección parece estar relacionada con el aumento de anticuerpos IgG4 (Oliveira et al., 2012). Finalmente, se ha observado asociación entre IL-10 y producción de IgG4. Estudios realizados en ratones han demostrado que bloqueando los receptores de IL-10 se induce la protección contra la reinfección (Wilson et al., 2011). Cabe destacar que la respuesta inmunitaria frente a la esquistosomosis puede complicarse, debido a que no se desarrolla de forma aislada, sino que una gran proporción de individuos afectados por la enfermedad, presentan a la vez otras infecciones como malaria, hepatitis B o C o VIH, las cuales pueden modificar la respuesta inmune de cada individuo.

En el caso de *S. mansoni*, la infección experimental de ratones comparte muchas de las características que se muestran en la infección en humanos, por lo que han sido de gran ayuda para entender la respuesta inmunológica en la esquistosomosis. Sin embargo, no ocurre lo mismo en las infecciones producidas por *S. haematobium* y *S. japonicum*. En los estudios experimentales de la esquistosomosis urinaria, los adultos no migran hasta el plexo venoso, con lo que no se produce el consiguiente depósito de huevos en la vejiga del ratón; es por ello que hay autores con modelos de infección para evitar ese problema (Fu et al., 2012). En el caso de *S. japonicum* las parejas de adultos sí llegan a su localización definitiva, aunque no hay muchos laboratorios que trabajen con *Oncomelania*, el caracol hospedador intermediario de esta especie de esquistosoma.

En cuanto a la respuesta inmunitaria contra las fases inmaduras del parásito que están migrando a través del organismo, se ha estudiado en modelo experimental murino, resultando que la respuesta más efectiva se daba en el pulmón (Wilson, 2009). Cuando los vermes adultos alcanzan su localización definitiva están más protegidos del ataque del sistema inmunológico; en esta fase desarrollan múltiples mecanismos de evasión parasitaria

que se verán detallados en el apartado siguiente. Los estudios empleando estos modelos experimentales nos han llevado al hecho de que la respuesta inmune protectora está combinada entre la generación de anticuerpos y la participación de células T (Jankovic et al., 1999). En las fases iniciales de la infección se desarrolla una respuesta predominante de tipo Th1; sin embargo, en el momento en que los adultos se han desarrollado, se han apareado los machos con las hembras y comienza la deposición de huevos, alrededor de la sexta a octava semana de infección, se modifica la respuesta bruscamente hacia el tipo Th2. La interacción del antígeno de los huevos liberados por el parásito con las células dendríticas del hospedador, fundamentalmente causado por la acción de determinados epítomos glucídicos, es lo que provoca esa modificación en el tipo de respuesta (Everts et al., 2012) (Figura 1.4). Las alteraciones hepatoesplénicas causadas por la esquistosomosis están originadas por la fibrosis hepática, esto viene de una serie de mecanismos interrelacionados que comienza con la falta de regulación de citocinas como la IL-13, la cual induce la estimulación de macrófagos por vía de la arginasa, produciendo L-ornitina y prolina, y éstas a su vez, estimulando la síntesis de colágeno, que desemboca en la fibrosis hepática. Por otro lado, cuando disminuye la respuesta de citocinas tipo Th2 como la IL-4, se manifiesta el daño tisular y la muerte del hospedador, debido al aumento de la respuesta de tipo Th1 proinflamatoria. Por ello, al aumentar las respuestas Th2, se muestran como “protectoras” para el hospedador, evitando la lesión asociada a respuestas Th1 elevadas. La IL-10 actúa en ese proceso de cambio de respuesta Th1 a Th2, que regula el proceso y limita la inflamación granulomatosa inicial. Si la infección avanza, estos mecanismos inmunomoduladores regulan la formación del granuloma, adquiriendo un menor tamaño a partir de la 12ª semana post-infección, lo que desemboca en una fibrosis más limitada que la que se da durante la fase aguda de la enfermedad. Esto se ha visto en estudios con ratones *knockout* deficientes en IL-4, demostrando que esta citocina tiene relación directa con el tamaño del granuloma (Brunet et al., 1997; Fallon et al., 2000). Además, se ha visto estimulación de la expresión de las citocinas IL-5 e IL-13 (Cheever et al., 1991), las cuales están implicadas en la fibrosis hepática grave (de Jesús et al., 2004).

### 1.4.3 Mecanismos de evasión

Los mecanismos de evasión desarrollados por los esquistosomas son variados (Pérez-Arellano et al., 2001; Jenkins et al., 2005).

*Evitar la actuación del sistema del complemento o impedir la destrucción por los macrófagos.* Se conoce la capacidad de síntesis o adquisición de moléculas reguladoras que son capaces de inactivar la convertasa de la vía alterna del complemento mediante moléculas DAF (*Decay Accelerating Factor*) del hospedador. También se conoce una proteasa con núcleo activo de serina en *S. mansoni* con similitud funcional al factor I del complemento. Impiden la destrucción por los macrófagos mediante el uso de enzimas antioxidantes como la superóxido dismutasa y la glutatión peroxidasa, cuya principal función es la protección de las superficies externas frente a la peroxidación.

*Adquirir moléculas del hospedador,* con esto consigue evitar que el sistema inmune las reconozca. El mecanismo es incorporar a la cubierta exterior del parásito lípidos del hospedador (ej: LDL, *low density lipoproteins*), los cuales dificultan la unión a los anticuerpos, a la vez que se adquieren antígenos de histocompatibilidad de clase II, antígenos de grupos sanguíneos o proteínas reguladoras del complemento.

*Generar respuestas ineficaces para el control de la infección.* Se ha demostrado la generación de anticuerpos de los isotipos IgG2, IgG4 o IgM, los cuales actúan bloqueando la citotoxicidad antiparasitaria mediada por IgE, IgG1 o IgG3.

*Eliminar antígenos durante las diferentes fases de su ciclo biológico.* Se conoce una glicoproteína de 38 kDa que se expresa en la superficie de las esquistosómulas y en el momento que éstas pasan a la fase de adulto, es liberada al medio.

*Interferir en el procesamiento y presentación antigénica.* El parásito produce prostaglandina D2, que posee la capacidad de inhibición de la migración de las células dendríticas, de forma que impide que tenga lugar este proceso (Hervé et al., 2003).

*Alterar los mecanismos efectoros.* El parásito libera una serie de proteasas que poseen capacidad de destrucción de las inmunoglobulinas por un mecanismo directo.

## **1.5 Manifestaciones clínicas**

### **1.5.1 Fase de inicio**

La fase inicial también se conoce como dermatitis cercariana o prurito del bañista. En las primeras 24 horas posteriores a la penetración de las cercarias a través de la piel es muy característica la aparición de prurito. Es más fácil su observación en viajeros, mientras que pasa desapercibida en personas residentes de áreas endémicas (Freedman et al., 2006).

El prurito puede aparecer debido a la penetración de cercarias aviares, pero en este caso el cuadro es más llamativo (Ross et al., 2002).

### 1.5.2 Fase aguda

También llamada Síndrome de Katayama, su aparición es debida a la reacción inmunológica que se desencadena frente a la fase de migración de la esquistosómula por el organismo del hospedador y tiene lugar entre las segunda y octava semanas posteriores a la penetración de las cercarias. Cursa con fiebre, lesiones cutáneas (exantema, urticaria), afectación pulmonar (tos, disnea) y eosinofilia (Ross et al., 2007). En general es autolimitada, pero hay ocasiones en que pueden persistir más de 10 semanas, presentando diarrea, pérdida de peso, dolor abdominal, hepatoesplenomegalia, etc. Es normal que aparezcan infiltrados pulmonares y engrosamiento bronquial en radiografía torácica y nódulos hepáticos hipocogénicos o hipodensos en ecografía y tomografía axial computerizada (TAC). Al igual que en la fase de inicio, la aparición de síntomas del Síndrome de Katayama es más frecuente en viajeros procedentes de áreas endémicas y está descrito clásicamente en infecciones por *S. mansoni* y *S. japonicum* con una elevada carga parasitaria, aunque actualmente se sabe que no es especie-específica, ni guarda relación con la intensidad de la infección (Ross et al., 2013).

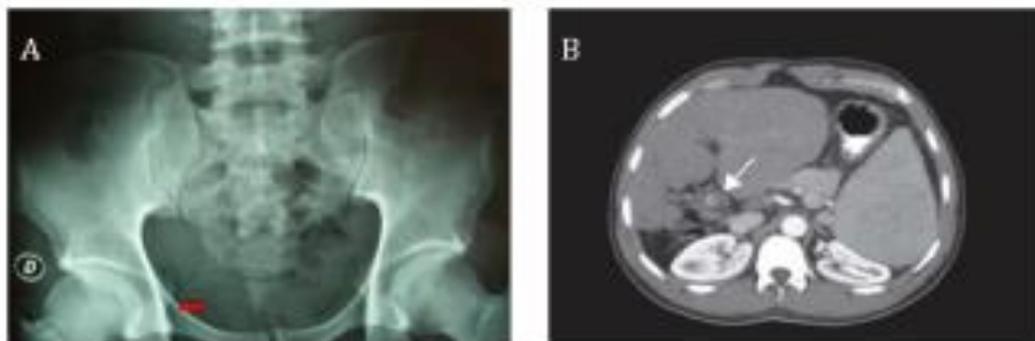
### 1.5.3 Fase crónica

Al contrario que en las fases anteriores, la fase crónica aparece con mayor frecuencia en personas residentes de áreas endémicas de esquistosomosis. De la misma forma, a diferencia de las fases anteriores, puede aparecer meses o años después de la infección y está originada por la reacción granulomatosa formada en torno a los huevos atrapados en el hígado, bazo, intestino, vejiga urinaria y otras localizaciones más lejanas como los pulmones y el sistema nervioso central. A pesar de ser más frecuente en personas residentes de área endémica, las personas sin una exposición previa pueden desarrollar esquistosomosis crónica tras una corta exposición. Por esta razón ha sido observada hasta en un 20% de los viajeros. Sin embargo, las complicaciones graves de ésta se producen tras infecciones repetidas y con alta carga parasitaria. La eosinofilia es un signo frecuente (Meltzer et al., 2006). Las principales manifestaciones clínicas se pueden agrupar en urinarias, hepatoesplénicas e intestinales. Además, pueden aparecer manifestaciones ectópicas (Gryseels et al., 2006).

Las manifestaciones urinarias se producen fundamentalmente por la localización de los huevos de *S. haematobium* en los plexos venosos del tracto urinario (Vester et al., 1997), estos huevos provocan inflamación granulomatosa, ulceración y poliposis en la mucosa uretral y vesical. Cursan clínicamente con un síndrome miccional irritativo con disuria, polaquiuria, proteinuria y especialmente hematuria terminal. Presentan síntomas más claros en niños y adultos jóvenes. El cuadro puede complicarse con litiasis y sobreinfección bacteriana. Cuando la enfermedad progresa, puede acabar provocando fibrosis y calcificación de uréteres y vejiga, produciendo hidrouréter e hidronefrosis (Khalaf et al., 2012) (Figura 1.5.A). Puede aparecer fracaso renal si el parénquima resulta afectado. Es bien conocida la asociación entre la esquistosomosis urinaria y la aparición de carcinomas de células escamosas vesicales (Parkin, 2006).

Las manifestaciones hepatoesplénicas se producen cuando los huevos de esquistosoma, principalmente *S. mansoni*, *S. japonicum* y *S. mekongi*, que se encuentran en las venas mesentéricas se incorporan a la circulación venosa portal, esto desencadena una oclusión gradual de las venas intrahepáticas que puede desembocar en una fibrosis de Symmers. Aparece hipertensión portal con ascitis, esplenomegalia, desarrollo de varices esofágicas con alto sangrado digestivo y encefalopatía (Figura 1.5.B). Sin embargo, no aparecen datos de insuficiencia hepatocelular a pesar de la localización donde se muestran las lesiones. Es por ello que no se muestran signos o estigmas de hepatopatía como ictericia, arañas vasculares, eritema palmar, atrofia testicular o ginecomastia. Además, si no hay coinfección con el virus de la hepatitis B o C, las transaminasas y la bilirrubina se mantienen en valores normales (Bica et al., 2000).

Las manifestaciones intestinales se producen en el momento en que los huevos atraviesan la pared intestinal causando hiperplasia, ulceración, formación de microabscesos y poliposis. Las lesiones se localizan frecuentemente en el intestino delgado y en el recto. En caso de que la infección esté causada por *S. intercalatum*, se asocia a lesiones intestinales moderadas. En cuanto a la clínica, causan dolor abdominal y diarrea de tipo inflamatorio con o sin sangre (Gryseels et al., 2006). Si la clínica llega a complicarse puede producirse enteropatía, desarrollar un cuadro oclusivo o suboclusivo, aparecer prolapso anal e incluso fístulas anorrectales.



**Figura 1.5.** (A) Calcificación de la vejiga (flecha roja), radiografía de abdomen en la esquistosomosis urinaria; tomada de Muro et al., 2010. (B) TAC de un paciente con hepatoesplenomegalia debida a esquistosomosis intestinal crónica y tumbosis portal secundaria (flecha blanca); tomada de Pérez del Villar Moro et al., 2012.

Las formas ectópicas más frecuentes son las cardiopulmonares, renales, genitales y las de afectación del sistema nervioso central. Además, todas las especies de esquistosomas causan afectaciones sistémicas debilitantes e inespecíficas como consecuencia de la inflamación continuada (Bustinduy et al., 2011), por lo que pueden desembocar en malnutrición, anemia, retraso en el crecimiento y trastornos en el desarrollo intelectual (King y Dangerfield-Cha, 2008).

Las manifestaciones cardiopulmonares se producen a partir de la deposición de huevos de los adultos de *Schistosoma*, con origen en las venas vesicales (en el caso de *S. haematobium*) o desde la circulación portal a través del *shunt* porto-cava (en el caso de *S. mansoni*). Esos huevos que no salen con la orina o las heces, acaban localizándose en la circulación pulmonar y pueden causar fibrosis de las arterias pulmonares, desembocando en hipertensión pulmonar con insuficiencia cardíaca derecha (Bethlem et al., 1997).

Ocasionalmente, la esquistosomosis puede provocar lesiones en el parénquima renal, en ese caso puede afectar específicamente el glomérulo y por ello causar glomerulonefritis (Barsoum, 2004). Por norma general, la glomerulonefritis está asociada a la esquistosomosis hepatoesplénica con *shunt* porto-sistémico. La más frecuente es la glomerulonefritis mesangial, que causa con una discreta proteinuria y microhematuria.

Las lesiones genitales son de tipo inflamatorio cuando están asociadas a los huevos viables y, de tipo fibroso e hipertrófico cuando los huevos no son viables o están calcificados. Si la infección cursa en mujeres, ésta afecta a vulva, vagina y cérvix. No causan dolor, sin embargo, en el caso de producirse úlceras, fístulas o si se sobreinfectan

pueden causar dispareunia y leucorrea (Poggensee et al., 2000), las cuales pueden facilitar la transmisión de otras enfermedades infecciosas, como el VIH. Si cursa en hombres, las afectaciones más frecuentes son la prostatitis crónica y la infección de las vesículas seminales. Por ello, la clínica más habitual cursa en la alteración de la eyaculación y la hematoespermia (Corachán et al., 1994).

En cuanto a las esquistosomosis que afectan al sistema nervioso central o neuroesquistosomosis, las dos manifestaciones más típicas son el síndrome cerebral y el medular. La localización de los huevos de esquistosoma en la médula espinal es más frecuente que la encefálica y puede producir varias formas de mielitis, siendo las más características la mielitis granulomatosa, la mieloradiculitis, la mielitis isquémica o vascular, siendo la más frecuente la mielitis transversa (Nobre et al., 2001). Los pacientes presentan habitualmente una vejiga neurógena, y está asociada en ocasiones a un nivel sensitivo-motor que cursa con debilidad en piernas, incapacidad para andar, dolor lumbar y parestesias (Ross et al., 2012). Cuando los huevos de *Schistosoma* spp se localizan en territorio cerebral, los síntomas comienzan como una crisis comicial y evolucionan hasta una encefalitis focal cerebral (Ferrari, 2004). Se ha descrito hasta en un 6% de los casos y se relaciona principalmente con *S. mansoni* y *S. japonicum* y, en menor medida, con *S. haematobium* (Ross et al., 2002).

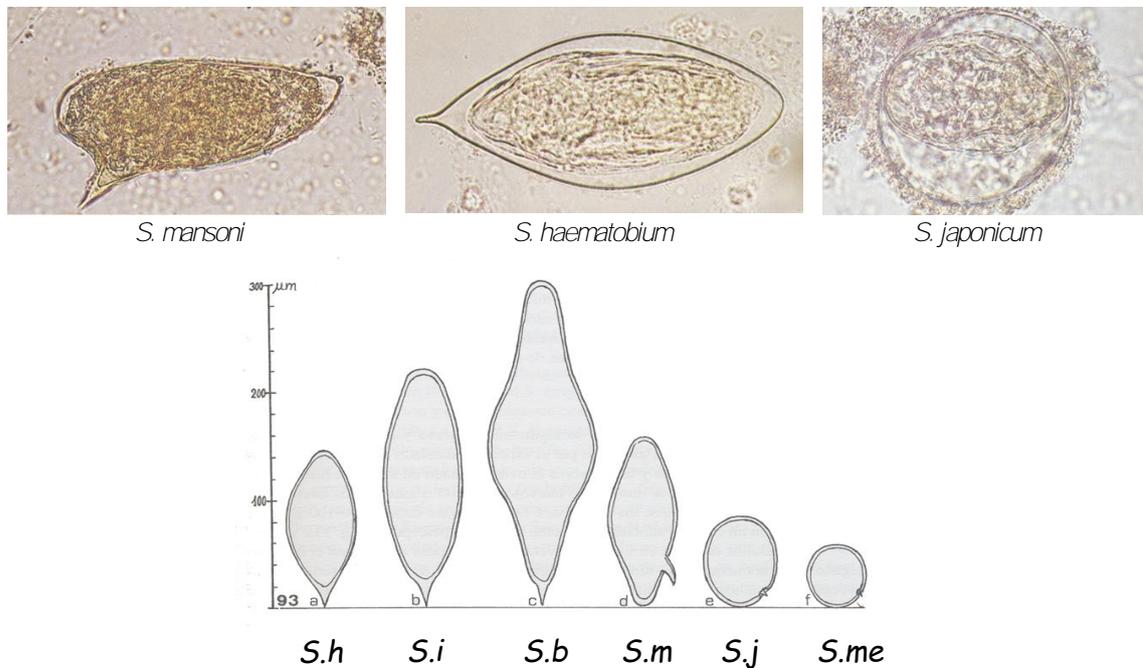
## 1.6 Métodos de diagnóstico

Para llegar a diagnosticar la enfermedad en humanos, hay que basarse en datos clínicos y epidemiológicos, aunque en la mayoría de los casos, son insuficientes. Se recurre a métodos parasitológicos, inmunológicos y moleculares.

### 1.6.1 Métodos parasitológicos

Estos métodos fueron los primeros empleados en el diagnóstico y están basados en la detección de huevos en heces en la esquistosomosis intestinal y hepática, o en orina en la esquistosomosis urogenital. Son métodos apropiados para el diagnóstico en la fase de patencia, pero no en la fase temprana, ya que la deposición de huevos no comienza hasta la sexta semana post-infección (Wang et al., 2011). Los huevos de *S. mansoni*, *S. haematobium* y *S. intercalatum* son ovoides (100-200  $\mu\text{m}$ ), mientras que los de *S. japonicum* y *S. mekongi* son redondeados (50-100  $\mu\text{m}$ ). Por otro lado, *S. haematobium* y

*S. intercalatum* presentan una espina terminal, mientras que en *S. mansoni* es lateral y en *S. japonicum* y *S. mekongi* es vestigial (Figura 1.6).



**Figura 1.6.** Características morfológicas de huevos de *Schistosoma* spp. *S. haematobium* (S.h), *S. intercalatum* (S.i), *S. bovis* (S.b), *S. mansoni* (S.m), *S. japonicum* (S.j) y *S. mekongi* (S.me).

La técnica más empleada es el examen de heces mediante la técnica de Kato-Katz y actualmente está recomendada como *gold standard* (WHO, 2013). La sencillez y bajo coste, junto con su especificidad y la posibilidad de cuantificar el número de huevos, hacen a esta técnica ser aplicable a estudios de campo en áreas endémicas. En individuos con infecciones con baja carga parasitaria, la sensibilidad disminuye (Sayasone et al., 2015). Se han utilizado otros métodos como la concentración con formol-éter (Polderman et al., 1994), la flotación en una solución saturada de cloruro sódico (Glinz et al., 2010) o la técnica del FLOTAC (Coulibaly et al., 2016). Existe una técnica mediante la cual se eclosionan los huevos, liberando los miracidios, que son atraídos hacia una fuente de luz para ser observados (Zhu et al., 2014a).

Para *S. haematobium* se emplea la observación del filtrado o sedimento de la orina en el microscopio, con alta especificidad y baja sensibilidad (Stete et al., 2012). También la detección de proteinuria y hematuria, se ha utilizado en tiras reactivas, como indicadores para el diagnóstico de la esquistosomosis urogenital (Morenikeji et al., 2014).

### 1.6.2 Métodos inmunológicos

El diagnóstico inmunológico es el método de elección en pacientes con sospecha clínica de esquistosomosis, no inmunizados previamente, para detectar anticuerpos en suero o plasma. Esta técnica es especialmente útil en situaciones de carga parasitaria muy baja (Alarcón de Noya et al., 2007). Se han desarrollado técnicas de inmunofluorescencia indirecta (IFI) (Burlandy-Soares et al., 2003), hemaglutinación indirecta (IHA) con eritrocitos recubiertos de antígenos de esquistosomas (Sorgho et al., 2005). La técnica ELISA es la más empleada, con extractos totales de vermes adultos y huevos (Sarhan et al., 2014), fracciones purificadas o proteínas (Carvalho et al., 2014b; Grenfell et al., 2013; Tanigawa et al., 2015). Para reducir la reactividad cruzada se ha intentado eliminar los carbohidratos responsables (Alarcón de Noya et al., 2007) y por otro lado aislar y purificar moléculas más específicas.

La detección de antígenos circulantes como el antígeno circulante catódico (CCA) y el antígeno circulante anódico (CAA) en sangre, orina y leche materna, han sido utilizados profusamente (Gonçalves et al., 2006). Son antígenos específicos de género, presentan pocas reacciones cruzadas con otros parásitos y permiten el diagnóstico en periodo prepatente. Se han desarrollado tiras de diagnóstico rápido sin conseguir detectar *S. haematobium* (Colley et al., 2013; Ochodo et al., 2015).

### 1.6.3 Métodos moleculares

El diagnóstico molecular está basado en la detección de ácidos nucleicos (DNA o RNA) mediante técnicas de amplificación en muestras de heces, orina o en biopsias de tejido. La técnica de referencia es la PCR convencional, diseñada sobre segmentos de genes nucleares, mitocondriales, DNA libre circulante o CFPD (*Cell-Free Parasite DNA*) tanto en heces para *S. mansoni* (Pontes et al., 2003), como en orina para *S. haematobium* (Ibironke et al., 2011). Se consiguen sensibilidades de hasta el 94% y especificidades del 100%, con utilidad desde la segunda semana post-infección (Sandoval et al., 2006). Mediante PCR también se puede detectar DNA libre circulante o CFPD (*Cell-Free Parasite DNA*) en suero, plasma, saliva, líquido cefalorraquídeo y orina, que proviene de las esquistosómulas muertas, de la renovación del tegumento de los adultos o de la desintegración de huevos (Xu et al., 2013, Kato-Hayasi et al., 2013). Se han desarrollado variantes de PCR como *Touchdown-PCR* (Helmy, 2007) o ELISA-PCR (Gomes et al.,

2010) y PCR a tiempo real (RT-PCR) (Lier et al., 2008), con capacidad para diferenciar las distintas especies de esquistosomas (Sady et al., 2015; Vinkeles et al., 2014).

También se han identificado varios micro-RNA (mi-RNAs), algunos específicos en suero de conejos y ratones infectados experimentalmente, útiles para el diagnóstico de la esquistosomosis japónica (Cheng et al., 2013) y mansónica (Hoy et al., 2014). Otros autores han mostrado técnicas como: «*RAKE assays*», «*rollingcircle amplification*» o «*DNA concatamers-based amplification*» (Zhu et al., 2014b).

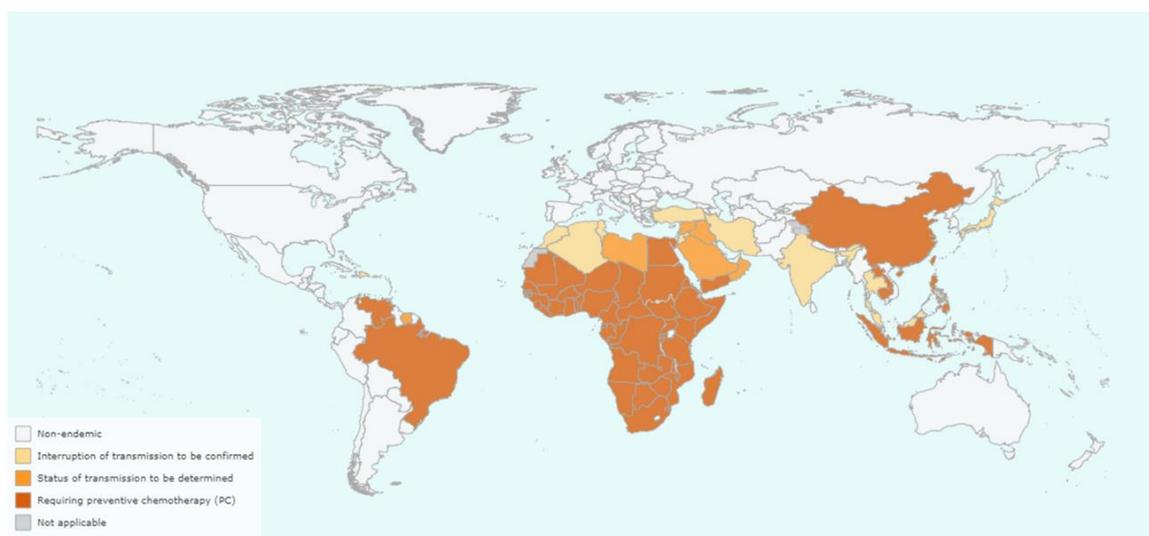
La técnica molecular de amplificación isotérmica de ácidos nucleicos tipo LAMP (*Loop-mediated isothermal amplification*) se ha empleado para la detección de DNA de *S. mansoni* (Fernández-Soto et al., 2014), *S. haematobium* “*Rapid-Heat LAMP Method*” (Gandasegui et al., 2015) y *S. japonicum* (Xu et al., 2015). Con esta técnica se han realizado estudios de campo en Brasil, para la detección de DNA de *S. mansoni* en un área de baja transmisión, así como en hospedadores intermediarios (Gandasegui et al., 2018a). También se han hecho estudios en Angola en población escolar expuesta a la infección por *S. haematobium* (Gandasegui et al., 2018b).

## 1.7 Tratamiento y prevención

El praziquantel (PZQ) es el fármaco de elección para tratar y controlar la esquistosomosis (Cioli et al., 2003). Es un medicamento de administración por vía oral, de toxicidad baja, de bajo coste y actúa rompiendo el tegumento de los vermes adultos, además de disminuyendo la producción de huevos. No obstante, no evita la reinfección, puesto que no actúa contra las formas inmaduras, las esquistosómulas y por ello no se emplea como profilaxis. Aunque el PZQ es efectivo contra todas las especies del parásito, es importante conocer la fase de la enfermedad y la especie infectante (Pérez-Arellano et al., 2007). La dosificación en fase aguda, en la que es muy eficaz frente a vermes adultos, es de 20 mg/kg cada 12 horas junto con dexametasona 20 mg/día, durante 3 días y la dosis de PZQ se repite a las 3-4 semanas. En fase crónica, la especie infectante sí influye en la dosificación del PZQ. En caso de infección por *S. mansoni* y *S. haematobium*, se recomiendan 20 mg/kg de PZQ cada 12 horas durante un día. Si la infección es por *S. japonicum*, 20 mg/kg cada 8 horas durante un día. La oxamniquina es un tratamiento alternativo, con una dosis de 10 mg/kg cada 12 horas durante 2-3 días (Muro et al., 2010).

Con la necesidad de encontrar alternativas al praziquantel, por parte de la OMS, tan sólo los derivados de la artemisina han aparecido como complemento a la terapia y profilaxis de la enfermedad (Pérez del villar et al., 2012). Otros fármacos utilizados de forma experimental son la mefloquina (Keiser et al., 2009) y la edelfosina, sola o combinada con praziquantel (Yepes et al., 2015).

En el año 1985, la OMS definió un plan para el control de la enfermedad, el tratamiento masivo o MDA “*Massive Drug Administration*” de la población en riesgo de infección, se conoce como quimioterapia preventiva o PCT “*Preventive ChemoTherapy*” (WHO, 1985). En el año 2012, a los países que llevaban programas de control, se les instó a dar un paso más para la eliminación de la enfermedad (WHO, 2013). Los objetivos para lograr esa erradicación incluyen la educación sanitaria y evitar el contacto con aguas que tengan las formas infectantes del parásito. Por ello, son necesarias mejoras en el tratamiento de aguas de consumo y residual en las zonas endémicas de la enfermedad, así como en los sistemas de irrigación, como la intubación de acequias. Aunque la aplicación de PCT ha logrado disminuir la prevalencia e impedir la transmisión en algunas zonas, no se ha logrado erradicar la enfermedad (Figura 1.8). Preocupa la disminución de la eficacia del PZQ y que se generen resistencias (Pinto-Almeida et al., 2015).



**Figura 1.8.** Mapa interactivo de la situación de la esquistosomosis de la OMS (WHO, 2017)

Por todo esto, la mejor opción para el control y prevención de la enfermedad sería encontrar una vacuna, ya que actualmente no existe ninguna vacuna comercial. Muchos estudios se han llevado a cabo por diversos grupos de investigación y a pesar de ello, sólo

hay pocos candidatos que se están evaluando en ensayos clínicos en fase I-III. I (Siddiqui y Siddiqui, 2017).

## 1.8 Susceptibilidad/resistencia a la infección por esquistosoma

La susceptibilidad a las enfermedades infecciosas está influenciada por factores dependientes del hospedador, del parásito y del ambiente en el que viven: intensidad y duración de la infección, el estado nutricional y otras enfermedades asociadas (Abath et al., 2006). En las zonas endémicas de esquistosomosis en el mundo, se observa gran heterogeneidad en la respuesta de cada individuo a la enfermedad y sólo una minoría de los pacientes, sufren las formas más graves de la enfermedad. En estas zonas, se pueden ver grupos con altos niveles de infección e individuos que desarrollan fibrosis hepática grave, lo que hace que se estime un fuerte componente genético (Burke et al., 2009). Los caracteres genéticos que afectan a la respuesta a la enfermedad son complejos y los fenotipos derivados varían ampliamente en una población. La identificación y análisis de los loci es importante para tratar de dilucidar la base molecular de estos caracteres complejos (Campino et al., 2006; Gatlin et al., 2009). Se han llevado a cabo estudios de ligamiento de regiones cromosómicas y SNPs para conocer cuanto de la base genética del individuo influye en la enfermedad. Se han podido acotar regiones genéticas implicadas tanto en la intensidad de la infección, como en el grado de fibrosis hepática inducido en un individuo (Dessein et al., 1999; Blanton et al., 2005; Burke et al., 2009; Mangano y Modiano, 2014).

Los estudios sobre la influencia genética en la esquistosomosis no son muy numerosos y solo han permitido identificar regiones cromosómicas con ligación a enfermedad grave. Una de las regiones cromosómicas identificadas en zonas endémicas de Brasil y Senegal ha sido el locus SM1 (*S. mansoni* 1) como genotipo ligado a la alta prevalencia y alto grado de intensidad a la infección en determinados grupos familiares (Marquet et al., 1996; Rodrigues et al., 1999). Los estudios disponibles han localizado SM1 en el brazo largo del cromosoma 5 humano (5q31-33). Es una región en la que se codifican importantes citocinas que tienen que ver con la respuesta Th2. En estudios de variantes genéticas se ha encontrado polimorfismos de citocinas con distinta susceptibilidad a la reinfección y alto nivel de infección en niños de distintas áreas endémicas (Kouriba et al., 2005; Gatlin et al., 2009). Otro de los locus que se ha asociado a la variabilidad en el desarrollo de esquistosomosis grave ha sido el denominado como SM2 (*S. mansoni* 2). En este

caso, este locus se encuentra relacionado con el desarrollo de fibrosis hepática grave y se localiza en el brazo largo del cromosoma 6, en la región 6q22-23. Se ha observado que se comporta como codominante con una penetrancia casi completa en jóvenes de ambos sexos con edades de 12 a 17 años de edad. En estudios de polimorfismos de los genes identificados en esa zona, se ha observado que los polimorfismos de los genes que se encuentran en esta zona, como el receptor 1 de  $IFN\gamma$  y el gen que codifica CTGF (*connective tissue growth factor*), se encuentran asociados con fibrosis hepática grave en enfermos de China, Sudán y Brasil (Chevallard et al., 2003; Dessein et al., 2009). En otros estudios se ha encontrado ligamiento de regiones cromosómicas en el cromosoma 1 (1p21-q23) y 6 (6p21-q21), con distinta intensidad de la infección por *S. mansoni*, pero no se han encontrado genes específicos (Mangano y Modiano, 2014).

Por otro lado, diversos estudios han intentado encontrar si otras regiones cromosómicas están relacionadas con la respuesta inmunológica específica contra agentes infecciosos, como es la región que codifica para el complejo mayor de histocompatibilidad (MHC: *Major Histocompatibility Complex*), pero no se han podido establecer asociaciones que permitan considerarlo como un locus mayor de susceptibilidad a la enfermedad hepática grave (Mangano y Modiano, 2014). Solamente en algunos estudios de polimorfismos de HLA (*human leucocyte antigen*) clase I y clase II, se ha podido establecer ligación de determinados alelos con enfermedad hepática grave, como en afectados por esplenomegalia grave en Egipto, Brasil y China (Huy et al., 2011).

## 1.9 Análisis de fenotipos intermedios

En las enfermedades complejas, la variabilidad de fenotipos complejos tiene herencia cuantitativa, esto es, la suma de efectos de alelos que interactúan con otros y con el medio ambiente. La herencia cuantitativa contribuye a la heterogeneidad de los rasgos complejos y a la de la herencia mendeliana. De este modo, la función del gen principal o responsable del fenotipo, es modificada por genes de baja penetrancia o modificadores, los cuales ayudan a la expresión de cualquier fenotipo, incluidos los adquiridos, como en la susceptibilidad de las enfermedades infecciosas (Blanco-Gómez et al., 2016).

La diversidad de fenotipos complejos de una población resulta de la variación fenotípica, influenciada por su medio ambiente y los componentes genéticos asociados. Los fenotipos complejos son consecuencia de diversos fenotipos intermediarios que actúan en

diferentes niveles. Hay muchas enfermedades principales como cáncer, envejecimiento, enfermedades cardíacas o enfermedades autoinmunes, que se consideran de fenotipo complejo, es decir, son consecuencia de otra serie de enfermedades de segundo orden o fenotipo intermedio, que son la causa de la enfermedad e influyen en su patología (ej: enfermedad isquémica cardíaca, influenciada por fenotipos intermedios como la presión sanguínea, la hipercolesteronemia o la susceptibilidad a agentes del tabaco, entre otras) (Libby et al., 2013).

La esquistosomosis es una enfermedad poligénica, es decir, causada por la suma de efectos de varios genes con un efecto débil, que interaccionan entre sí y con el medio, haciendo que bajo las mismas o muy similares condiciones ambientales, unos individuos desarrollen la enfermedad, a la vez que otros sean resistentes a la misma. En humanos, la identificación de estos genes modificadores de efecto débil es una tarea enormemente difícil, a causa de la heterogeneidad genética de la población y a la compleja interacción con el entorno. Serían necesarios estudios poblacionales con muestras de enorme tamaño, gran gasto de tiempo y dinero sin garantía de éxito. Con este tipo de estudios se han identificado numerosos alelos comunes. No obstante, respecto a los genes de baja penetrancia y su efecto individual tan pequeño, hay evidencias de que estos estudios de asociación de genoma completo no podrán encontrar, en un futuro próximo al menos, a estos genes responsables de la mayor parte del componente hereditario de las enfermedades poligénicas. El ratón, en cambio, ofrece un modelo más simple de variabilidad genética controlada si se emplean cepas de genética homogénea (cepas singénicas), de susceptibilidad y progresión divergentes (una susceptible y otra resistente) y sus interacciones con el medio son de más fácil control (Smith et al., 2009).

El modelo de estudio de variabilidad genética controlada se va a generar mediante el cruce retrógrado o "*backcross*" entre una cepa susceptible y otra resistente a la enfermedad. En nuestro laboratorio se ha realizado un estudio preliminar con cinco cepas de ratones singénicos (C57BL/6J, FVB/NH, DBA/2J, BALB/c y CBA/2J) y hemos identificado las dos cepas más divergentes en cuanto a susceptibilidad a la infección, que son C57BL/6J (más resistente) y CBA/2J (más susceptible) (Pérez del Villar et al., 2011). También se ha analizado la susceptibilidad a la infección en el híbrido F1B6CBA entre dichas cepas (Pérez del Villar et al., 2014).

## 2 Hipótesis y objetivos

Mutaciones en un gen de alta penetrancia, con una selección positiva a lo largo de su evolución, pueden determinar la resistencia a un agente infeccioso. Generalmente, la susceptibilidad/resistencia a un agente infeccioso viene dada por la acción sumatoria de diversos genes que interaccionan entre sí, denominados genes de baja penetrancia o modificadores.

Dada la gran variación en la prevalencia y en los grados de infección y enfermedad de la esquistosomosis en zonas endémicas, es evidente que esta parasitosis tiene un importante componente genético. Por otro lado, es difícil identificar el componente genético en el ser humano debido a la complejidad de la interacción con el ambiente y a la gran variabilidad genética entre los individuos. El modelo de ratón ofrece una gran homogeneidad genética entre los individuos de una misma cepa singénica y una gran variabilidad entre diferentes cepas, lo que permite la identificación de regiones génicas (QTLs) y su expresión (eQTLs) implicadas en la resistencia o susceptibilidad de un individuo a esta enfermedad.

Una de las metodologías más comunes para cruzar cepas consanguíneas (homocigóticas) y obtener un modelo de variabilidad genética controlada, es el cruce retrógrado o *backcross*. Se realiza un primer cruce entre cepas fenotípicamente distintas (Parental 1 AA x Parental 2 aa) y se obtiene la primera generación filial o F1, con una composición genética fija Aa (heterocigotos para todos los loci) y serán idénticos entre ellos. El *backcross* se produce cruzando un individuo de la F1 con otro procedente de una cepa parental y pueden darse dos *backcross*, dependiendo de la cepa parental que se elija. Sobre el modelo experimental murino, se aplicará metodología desarrollada recientemente con el fin de identificar los genes modificadores. Esto ayudará a identificar esos mismos genes en la población humana y a mejorar nuestro conocimiento sobre el riesgo individual, tratamiento y pronóstico de la enfermedad.

El objetivo general de este trabajo es la identificación de regiones cromosómicas implicadas en la susceptibilidad a la infección por *S. mansoni*, principalmente los genes de baja penetrancia o modificadores, mediante modelo genético murino (cruce retrógrado o *backcross*). Para lo que se definen los siguientes objetivos específicos:

- Caracterizar la susceptibilidad a la infección por *Schistosoma mansoni* mediante cruce retrógrado o *backcross* definiendo diferentes patrones de enfermedad
- Estudiar la asociación entre los patofenotipos y las magnitudes parasitológicas, patológicas e inmunológicas asociadas.
- Identificar las regiones genómicas (QTL) ligadas a los fenotipos intermedios definidos tras la infección con *Schistosoma mansoni*.

# 3 Materiales y Métodos

### 3.1 Animales

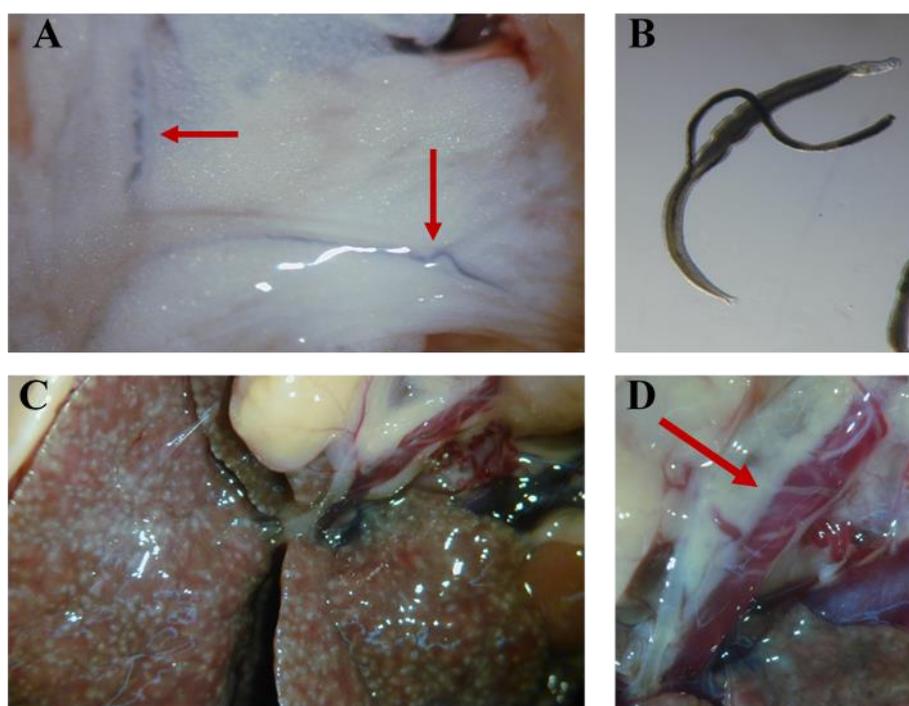
Se utilizó la cepa singénica o *inbred* de ratones CBA/2J (Charles River, Lyon) y el cruce de primera generación F1B6CBA (Charles River, Lyon) obtenido a partir de la estirpe C57BL/6J, más resistente a la infección y CBA/2J, más susceptible a la infección por *Schistosoma mansoni* (Pérez del Villar et al., 2014). Los animales fueron colocados en jaulas estándar de policarbonato y alambre, con alimento y agua *ad libitum* en un ambiente bajo 12/12-h de ciclo de luz-oscuridad a 22-25°C en las instalaciones del Servicio de Experimentación Animal (SEA) de la Universidad de Salamanca (Reg. # PAE / SA / 001), en condiciones libres de patógenos (SPF). Los procedimientos con animales cumplieron con las regulaciones españolas (L 32/2007, L 6/2013 y RD 53/2013) y la Unión Europea (Di 2010/63/CE) sobre experimentación animal. El Comité de Ética de la Universidad de Salamanca revisó y aprobó los protocolos (Ref. 15/0018). El estado de salud de los animales se controló durante los experimentos de acuerdo con las directrices de la Federación de Asociaciones Europeas de Ciencias Animales de Laboratorio (FELASA). Se hicieron todos los esfuerzos para minimizar el sufrimiento de los animales.

### 3.2 Mantenimiento experimental del ciclo biológico de *S. mansoni*

En el Centro de Investigación de Enfermedades Tropicales de la Universidad de Salamanca (CIETUS) se mantiene de forma experimental la cepa LE de *S. mansoni* en caracoles *Biomphalaria glabrata* como hospedador intermediario y ratones CD1 como hospedador definitivo. En el laboratorio de Hospedadores Intermediarios, se mantienen caracoles *Biomphalaria glabrata* que son infectados con miracidios de *S. mansoni*. Transcurridos 30 días desde la infección, en condiciones adecuadas de luz y temperatura, se produce la emisión de las cercarias, con las que se realizará la infección de ratones CD1.

Los ratones CD1 infectados con 150 cercarias de *S. mansoni* son sacrificados a los 60-90 días post-infección. Se les aplicó una sobredosis de PBS con pentobarbital sódico (P3761, Sigma) 60 mg/kg y heparina 2000 IU/L a nivel peritoneal. La extracción de vermes adultos se realizó inyectando solución salina tamponada con fosfato (PBS) con heparina

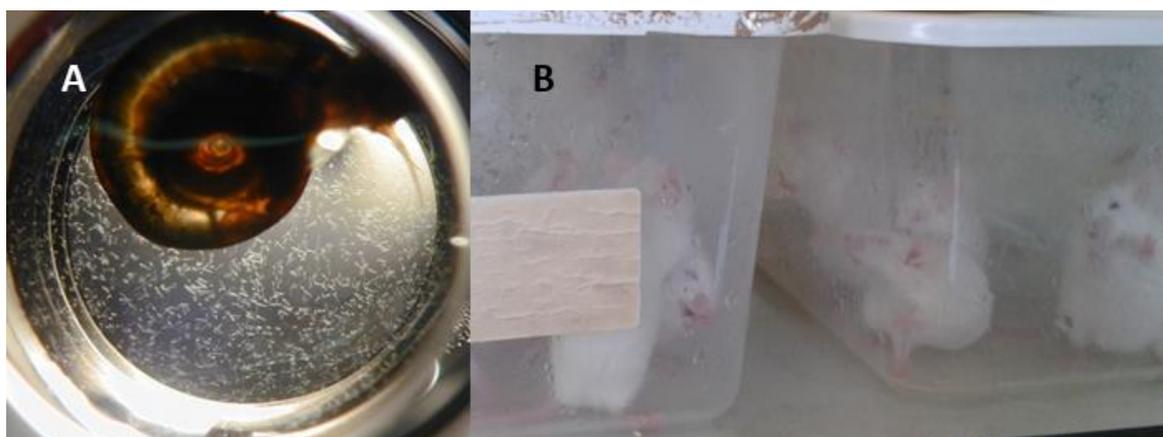
1000 IU/L, en el ventrículo izquierdo del corazón, con una jeringuilla de 10 mL y aguja de 0.50 X 16 mm (25 G). El paso del medio líquido arrastra a los parásitos, que salen al exterior por la vena porta (Figura 3.1 D), recuperándose con ayuda de unas pinzas tanto parejas de vermes adultos, como vermes libres (Mann et al., 2010) (Figura 3.1 B), también se recuperan vermes de las venas mesentéricas (Figura 3.1 A). Los hígados de los ratones infectados presentan gran número de granulomas, producidos por la reacción en el tejido en torno a los huevos eliminados por los adultos (Figura 3.1 C).



**Figura 3.1.** A) Parejas de vermes de *S. mansoni* en las venas mesentéricas. B) Pareja de vermes (macho y hembra) de *S. mansoni* obtenida de la vena porta C) Hígado con granulomas. D) Vermes adultos de *S. mansoni* en la vena porta.

Para obtener los miracidios con los que se infectan los caracoles, los hígados con los granulomas se homogeneizan en 100 mL de solución salina NaCl 0.85% y se depositan en una copa de sedimentación, completando el volumen hasta 1000 mL. Seguidamente, la mezcla se deja sedimentar durante 20 minutos, tras lo cual se retira el sobrenadante y se añade nuevamente solución salina, se deja sedimentar otros 20 minutos. Se retira de nuevo el sobrenadante hasta dejar unos 50 mL en el fondo de la copa. El sedimento obtenido se transfiere a una placa de Petri a la que se añade agua mineral y se pone bajo una fuente de luz y calor. A los 10-20 minutos los huevos comienzan a eclosionar liberando a los miracidios de su interior.

Para infectar los ejemplares de *B. glabrata* se utilizan placas de 24 pocillos (Costar® 3724, Corning Inc.). En cada pocillo de la placa se aloja un caracol de 5-6 mm de diámetro, que será infectado con 7-9 miracidios con ayuda de un estereomicroscopio Olympus SZX9 (Japón) y una pipeta Pasteur. Transcurridas 4 horas a temperatura ambiente, se transfieren los caracoles a una cubeta con agua en una sala a 27°C y se esperará entre 28 y 35 días para obtener las primeras emisiones de cercarias. Para obtener las cercarias, los caracoles infectados se colocan en un recipiente con 25 mL de agua, con luz directa y temperatura de unos 26°C. Tras unos minutos comenzará la emisión de cercarias y aproximadamente a las 2 horas de exposición se recogen y se cuentan por triplicado (Figura 3.2A). Se preparan las dosis a razón de 150 cercarias de *S. mansoni* en 25 mL de agua por ratón y se colocan en un recipiente en que la altura del agua sea entre 0.5-1 cm (Figura 3.2B). Los ratones se mantienen en el recipiente aproximadamente una hora, para que las cercarias puedan completar la infección.



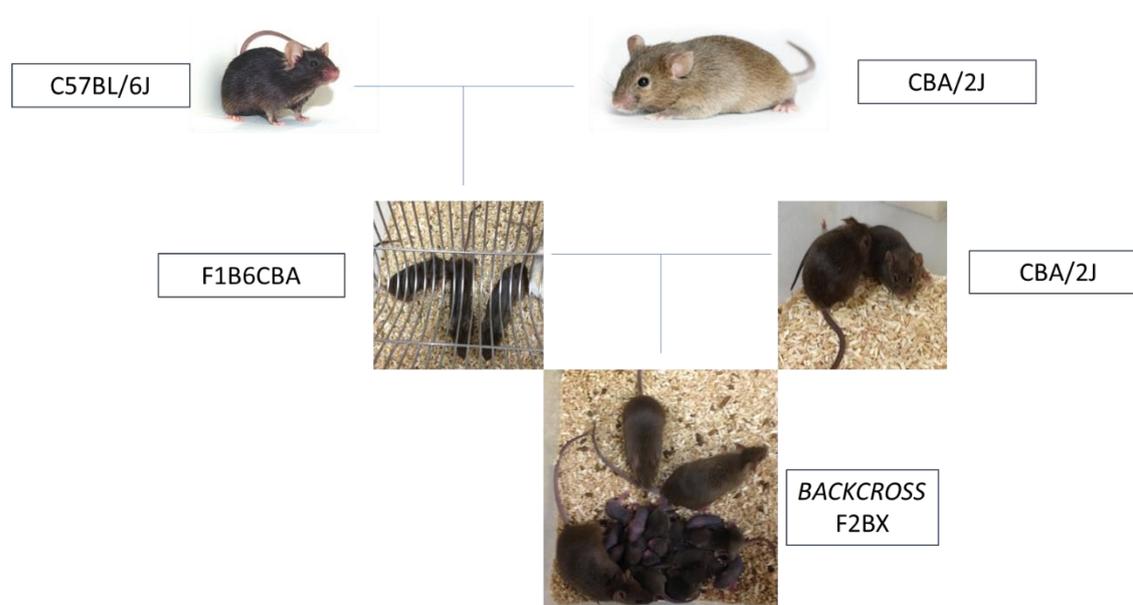
**Figura 3.2.** Emisión de cercarias de *Schistosoma mansoni* e infección de ratones. A) Ejemplar de *Biomphalaria glabrata* emitiendo cercarias. B) Ratones en contacto con agua y cercarias de *S. mansoni*.

### 3.3 Generación de una cohorte con heterogeneidad genética y fenotípica mediante retrocruce o *backcross*

#### 3.3.1 Generación del *backcross*

La cohorte en la que estudiamos la susceptibilidad a la infección por *S. mansoni*, se generó siguiendo la estrategia de retrocruce o *backcross* con la cepa parental consanguínea o singénica CBA/2J y la F1 generada del cruce de las cepas parentales CBA/2J y C57BL/6J, denominada F1B6CBA. Se generaron ratones F2BX, genéticamente únicos, pero con variabilidad genética controlada, pues son mezcla de C57BL/6J y CBA/2J en grado variable (Figura 3.3). Cada ratón pudo ser, en cada *locus*, homocigoto para el alelo CBA/2J o

heterocigoto, con un alelo de CBA/2J y otro de C57BL/6J. En dicha combinación, el componente genético procedente de la cepa CBA/2J fue mayoritario, ya que fue con la que se realizó el retrocruce, mientras que el C57BL/6J quedó reducido. Esto se hizo así para enriquecer la cohorte *backcross* en alelos de susceptibilidad a la infección por *S. mansoni*. Con aproximadamente 7 semanas de edad y peso entre 20-25 gramos se procedió a infectarlos.



**Figura 3.3.** Esquema de la estrategia seguida para la generación del *backcross*.

### 3.4 Evaluación de la infección por *Schistosoma mansoni* en la cohorte F2BX

Los ratones fueron infectados percutáneamente con  $150 \pm 8$  cercarias de *S. mansoni* por ratón con el método del anillo (Tallima et al., 2017) y fueron sacrificados a la novena semana post-infección según el diseño del experimento (Figura 3.4). A los animales se les afeitó el abdomen y se inmovilizaron con una mezcla de ketamina 50 mg/kg de peso vivo (*Imalgene1000*<sup>®</sup>, Merial), diazepam 5 mg/kg (*Valium10*<sup>®</sup>, Sant Cugat del Vallés) y atropina (B. Braun, Madrid). Una vez inmovilizados, se les colocó en posición de decúbito supino y se les humedeció el abdomen con agua estéril y luego se expusieron por el método del anillo a las cercarias (Figura 3.5). Una vez expuestos a las cercarias, los ratones se dejaron en oscuridad durante 45 minutos. El número de cercarias y su viabilidad se determinaron utilizando un microscopio estereoscópico Olympus SZX9 (Japón). Las diferentes magnitudes obtenidas de cada ratón se fueron recogiendo individualmente en una matriz de datos.



**Figura 3.4.** Diseño del experimento de infección, recogida de muestras y sacrificio de los animales a la 9ª semana post-infección.



**Figura 3.5.** Detalle de la infección percutánea de ratones con cercarias de *S. mansoni* por el método del anillo.

### 3.5 Recogida de magnitudes parasitológicas

#### 3.5.1 Recuperación de vermes mediante perfusión

Se recuperaron los vermes adultos del interior de la vena porta y las venas mesentéricas mediante la técnica de perfusión, descrita en el apartado de mantenimiento experimental del ciclo biológico de *S. mansoni*. La perfusión se realizó con ayuda de un estereomicroscopio Olympus SZX9 (Japón) y fue videomonitorizada con una cámara Olympus C-2000Z (Japón) y un televisor Samsung (Corea). También se obtuvieron vermes adultos abriendo las venas mesentéricas y recuperándolos con ayuda de unas pinzas. Se registraron el número de parejas y de vermes sueltos, tanto machos como hembras, con ayuda del estereomicroscopio y las micrografías de la carga parasitaria (Figura 3.6).



**Figura 3.6.** Fotos de los vermes recogidos de ratones infectados con *S. mansoni*.

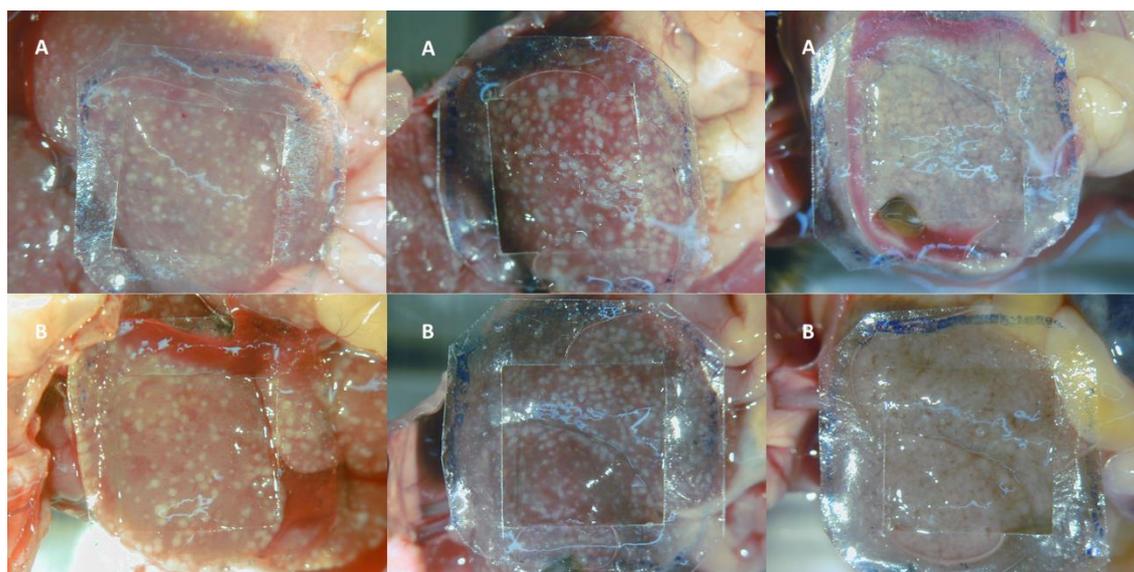
### 3.5.2 Determinación de la carga de huevos en hígado e intestino y cálculo de la fecundidad de las hembras

Se extrajeron el hígado e intestino de cada ratón, se anotó su peso y se procedió a su digestión para determinar la carga de huevos. Una cantidad pesada de hígado o intestino se introdujo en tubos de 50 mL, con 25 mL de solución salina NaCl 0.85% y con KOH al 5% y se incubaron a 37 °C en agitación hasta la mañana siguiente. A las 24 horas, se centrifugaron las muestras digeridas a 200 g en una centrífuga Mixtasel-BLT (P-Selecta, Barcelona) durante 5 minutos y se retiró el sobrenadante, dejando 5 mL de solución concentrada de huevos. Se empleó una cámara McMaster para el recuento de huevos por triplicado y se obtuvo el número de huevos por gramo de hígado e intestino.

Una vez obtenido el número de huevos por gramo de hígado e intestino, se calculó la fecundidad de las hembras dividiendo el número de huevos en hígado o intestino por el número de hembras.

### 3.6 Magnitudes patológicas: recuento de granulomas y estimación de la superficie hepática afectada

Se recogieron tres fotografías de distintas partes de cada uno de los hígados de los ratones con una cámara Olympus C-2000 Z (Japón) antes de la perfusión, para estimar el grado de lesión macroscópica de los hígados. El número de granulomas por centímetro cuadrado, fue contado por dos investigadores cualificados por separado en cada una de las tres micrografías y lo expresaron como media. Se utilizó el programa *Adobe Photoshop* para mejorar el contraste de las imágenes y *Power Point* para establecer los sectores que facilitaban el recuento. En caso de controversia se solicitaba la intervención de un tercer investigador. La superficie hepática afectada se estimó a partir del número de granulomas y la superficie media de 10 granulomas de cada una de las tres micrografías utilizando el programa *ImageJ* (Schneider et al., 2012).



**Figura 3.7.** Fotos de hígados afectados por granulomas. De izquierda a derecha de menor a mayor afectación (A: ratones machos y B: ratones hembras).

### 3.7 Estudio de poblaciones celulares por citometría de flujo

Se estudiaron las poblaciones celulares CD4, CD8, B220 y CD45 mediante citometría de flujo tanto en sangre periférica como en esplenocitos. Se emplearon en ambos casos anticuerpos monoclonales (mAb) de ratón de la serie linfoide, conjugados con isotiocianato de fluoresceína (FITC) contra CD4 (553047), con ficoeritrina (PE) contra CD8 (550798), con aloficocianina (APC) contra B220 (553092) y con proteínas de peridina-clorofila

(PerCP-Cy<sup>TM</sup>5.5) contra CD45 (550994) de BD Pharmingen<sup>TM</sup>. Los anticuerpos monoclonales marcados se aplicaron en un factor de dilución 1/50 en PBS suplementado con 2% de suero fetal de ternero (FCS), 25  $\mu$ L por pocillo y se incubaron durante 30 minutos a temperatura ambiente en oscuridad, en una placa de 96 pocillos y fondo redondeado (Corning Inc. 3789, Costar). Se añadió a cada pocillo 100  $\mu$ L de solución de lavado, se resuspendió y se centrifugó 5 minutos a 237 *g* en una centrífuga de placas *Eppendorf* 5804 (Hamburgo, Alemania). Se retiró el sobrenadante y el *pellet* se resuspendió en 200  $\mu$ L de solución de lavado. Se pasó el contenido a los tubos específicos para citometría y se mantuvo a 4°C en oscuridad. Se empleó un citómetro de flujo *FACSCalibur* (*BD Biosciences*) de cuatro colores del Servicio Central de Citometría de Flujo de la Universidad de Salamanca. Se recogió un total de 30.000 eventos de cada muestra, controlados por dispersión directa y lateral y los datos se analizaron mediante el software *GateLogic Flow Citometry Analysis* (*Inivai Technologies Pty Ltd*) (Rojas-Caraballo et al., 2014).

El principio en que se basa la citometría es simple, consiste en hacer pasar células u otras partículas en suspensión alineadas y de una en una por delante de un haz luminoso. La interacción de las células o partículas con el rayo luminoso genera señales que llegan a los detectores adecuados. La información producida puede ser de dos tipos: la generada por la dispersión de la luz y la relacionada con la emisión de luz por los fluorocromos presentes en la célula o partícula al ser excitados por el rayo luminoso. Cuando la luz incide perpendicularmente sobre cada célula, esta se dispersa de frente y hacia los lados, por lo que se detecta en dos posiciones diferentes. Por una parte se detecta la luz dispersada frontalmente (FSC, del inglés *Forward Scatter*) y por otra, la luz dispersada lateralmente con un ángulo de 90° (SSC, del inglés *Side Scatter*). La luz dispersada frontalmente se relaciona con el tamaño celular mientras que la luz dispersada lateralmente da una idea de la estructura interna de la célula, por lo que proporciona información sobre la granularidad y morfología celular. Para obtener más información sobre las células, los marcadores moleculares que expresan en superficie son marcados con anticuerpos monoclonales específicos que presentan fluorocromos para su detección. Cuando el fluorocromo capta la energía luminosa, emite una señal fluorescente de menor energía que el rayo incidente, que es detectada por un detector a una longitud de onda específica.

Después de introducir cada muestra en el citómetro de flujo, los detectores envían la información a un ordenador. Toda esta información es recogida en forma de tablas y a partir de un programa informático se puede realizar el análisis de los datos mediante la utilización de diagramas de puntos, donde cada punto se corresponde con una célula. Como el citómetro de flujo además tiene detectores de fluorescencia, también se pueden analizar los datos enfrentando en el diagrama dos fluorocromos. El software empleado dispone de herramientas, haciendo *gates* (rodeando las nubes de puntos), con las que se calculan los porcentajes de las poblaciones celulares.

Con los marcadores empleados podemos identificar las siguientes poblaciones celulares: leucocitos (CD45+), linfocitos (B220+), linfocitos T helper (CD4+) y linfocitos T citotóxicos (CD8+).

### **3.7.1 Preparación de los leucocitos de sangre periférica para citometría de flujo**

Se tomó una muestra de sangre de ratón que se mezclaba con 20  $\mu\text{L}$  de PBS con 10 IU de heparina. Después se incorporaron 35  $\mu\text{L}$  de sangre completa por pocillo a una placa de 96 pocillos y fondo redondeado (Corning Inc. 3789, Costar). Se añadieron 150  $\mu\text{L}$  de solución de lisis, formada por cloruro de amonio 154 mM (A9438, Sigma), bicarbonato potásico 10 mM (P9144, Sigma) y 0.082mM de EDTA (E6758, Sigma), se homogeneizó y se dejó 30 minutos a temperatura ambiente. Después se centrifugó a 237  $g$  en una centrífuga de placas *Eppendorf* 5804 (Hamburgo, Alemania), se decantaba y se repetía el proceso. Se incorporaron 150  $\mu\text{L}$  de solución de lavado, PBS con suero fetal bovino al 2% (PBS-FCS) y se homogeneizó de nuevo. Por último, se centrifugó a 237  $g$  y se eliminó el sobrenadante, quedando las células preparadas para ser marcadas.

### **3.7.2 Preparación de esplenocitos para determinar poblaciones con citometría de flujo**

Para obtener los esplenocitos se extrajeron los bazo de los ratones de manera aséptica en la necropsia. Se depositaron en una placa Petri con 10mL de PBS estéril a 37°C en cámara de flujo laminar FASTER BHA 48 (Ferrara, Italia). Se cortaron los extremos del bazo y con la ayuda de una aguja doblada en ángulo de 90°, se perfundió el PBS para extraer los esplenocitos y se recogió la suspensión de esplenocitos en un tubo estéril de 15 mL (Labbox, Barcelona). Se centrifugaron los tubos durante 4 minutos a 200  $g$  en una centrífuga Mixtasel-BLT (P-Selecta, Barcelona). Se retiró el sobrenadante y se resuspendió

el *pellet* en 500  $\mu\text{L}$  de solución de lisis, pasando el contenido a un vial de 1.5 mL (*Eppendorf*). Se mantuvieron 5 minutos a temperatura ambiente y se centrifugaron 5 minutos a 100 g en una centrífuga *Eppendorf* 5415R (Hamburgo, Alemania). Se retiró el sobrenadante y se resuspendió en 500  $\mu\text{L}$  de PBS con suero fetal bovino al 2%. Se tomaron 10  $\mu\text{L}$  de cada tubo de esplenocitos y se depositaron en una placa de 96 pocillos y fondo redondeado (*Costar*<sup>®</sup> 3789, *Corning Inc.*), quedando preparados para incorporar los anticuerpos monoclonales con fluorocromo.

### **3.8 Respuesta inmune humoral por ELISA**

#### **3.8.1 Obtención de antígeno somático de vermes adultos de *Schistosoma mansoni* (SoSmAg)**

Se realizó siguiendo la metodología descrita por (Abán et al., 1999) con algunas modificaciones. Se tomaron vermes adultos conservados a  $-20^{\circ}\text{C}$  y se suspendieron en PBS, a razón de 20 vermes adultos por mL de PBS. Para evitar la degradación del antígeno, se añadió una solución inhibidora de proteasas: Complete Mini EDTA-Free, Roche 04 693 159 001. A continuación, se homogeneizó en un homogeneizador de cristal y se sometió a tres ciclos sucesivos de congelación y descongelación a temperatura ambiente. Posteriormente, se llevaron a cabo cinco ciclos de 30 segundos en el sonicador (*Virsonic300*, *Virtis*) a 70 kHz. Durante el proceso de sonicación se mantuvo la solución con el antígeno en un baño de hielo. Se centrifugó con una centrífuga *Eppendorf* 5415R (Hamburgo, Alemania) a 16.000 g durante 30 minutos, a  $4^{\circ}\text{C}$  y se recogió el sobrenadante para posteriormente determinar la concentración de proteínas por el método BCA (*Bicinchoninic Acid Protein Assay, Micro BCA*<sup>™</sup> *Protein Assay Kit, Thermo Scientific, Rockford; Il, U.S.A.*), siguiendo las instrucciones del fabricante.

#### **3.8.2 Detección de anticuerpos anti-Sm en ratones infectados con el parásito**

Se tomaron muestras de plasma de los ratones antes de la infección y antes de la necropsia y se analizaron mediante ELISA indirecto para detectar anticuerpos específicos de los isotipos IgG e IgM y de los subtipos IgG1 e IgG2a. Se tapizaron microplacas *Corning Costar* de 96 pocillos de fondo plano (*Costar*<sup>®</sup> 3596, *Corning Inc*) con 5  $\mu\text{g}/\text{mL}$  de SoSmA en tampón carbonato pH 9,6 (100  $\mu\text{L}$  por pocillo). Se incubaron a  $4^{\circ}\text{C}$  durante 16 horas. Las placas se lavaron 3 veces (200  $\mu\text{L}$  por pocillo) con PBS con Tween 20 al 0.05% (PBST). Las placas se bloquearon con albúmina de suero bovino (B4287, *Sigma*) al 2% en

PBST durante 50 minutos a 37°C. Se lavaron de nuevo tres veces con PBST. Se añadieron muestras de suero diluidas a 1:100 en PBST en pocillos duplicados (100 µL por pocillo) y se incubaron durante 1h a 37°C. Tras un nuevo ciclo de 3 lavados con PBST, se usaron conjugados IgG-HRP, IgG1-HRP, IgG2a-HRP e IgM-HRP anti-ratón (Sigma) a 1:1000 en PBST y se incubaron 1 hora a 37°C (100 µL por pocillo). Las placas se lavaron de nuevo con PBST y se revelaron añadiendo H<sub>2</sub>O<sub>2</sub> (0.012%) y sustrato de ortofenilendiamina (0.04%) en tampón citrato 0.1 M, pH 5.0 (100 µL por pocillo). La reacción se detuvo con H<sub>2</sub>SO<sub>4</sub> 3N (50 µL por pocillo) y se leyó a una densidad óptica (D.O.) de 492 nm en un lector de placas *MultiSkan GO ELISA* (*Thermo Fisher Scientific*, Vantaa, Finlandia) (Rojas-Caraballo et al., 2014).

### 3.9 Genotipado

#### 3.9.1 Obtención y cuantificación de DNA de *S. mansoni*

La extracción de DNA total a partir de las colas de los ratones se realizó con el kit comercial *DNeasy<sup>®</sup> Blood & Tissue Kit* (Qiagen<sup>®</sup>, Hilden, Germany), siguiendo el protocolo para la extracción de DNA de tejidos animales recomendado por el fabricante.

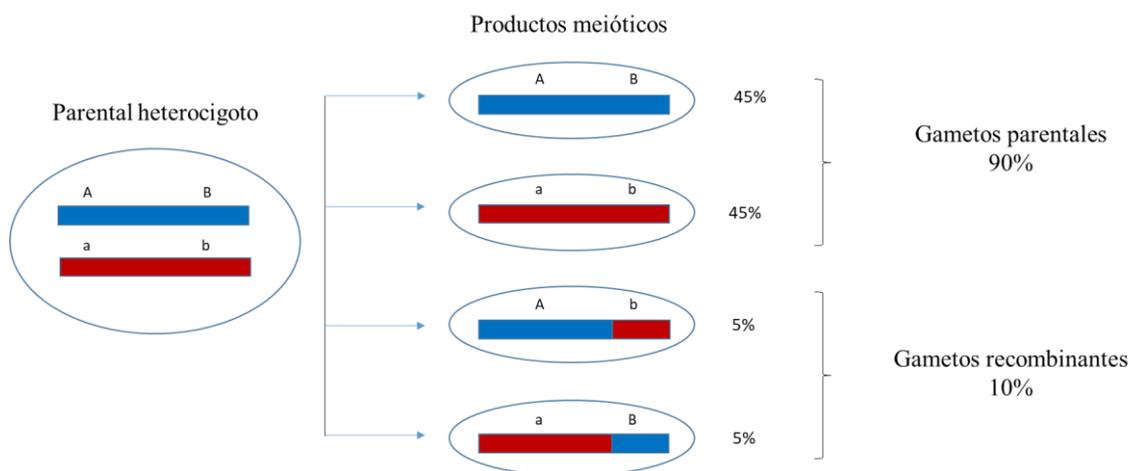
La cuantificación de la concentración obtenida de DNA en el correspondiente volumen de elución (200 µL) se llevó a cabo con 2 µL del eluido en un espectrofotómetro *NanoDrop<sup>®</sup> ND-1000* (*Termo-Fisher Scientific Inc.*).

#### 3.9.2 Genotipado de los ratones *backcross* F2BX infectados con *S. mansoni*

Para realizar el análisis de ligamiento genético e identificar las regiones del genoma asociadas a los fenotipos estudiados, se hizo un genotipado de la cohorte de ratones generada mediante cruce retrógrado o *backcross*. El genotipado se llevó a cabo mediante plataforma *Illumina* en el Centro Nacional de Genotipado (CeGen), adscrito a las instalaciones del Centro Nacional de Investigaciones Oncológicas (CNIO). Se utilizó el DNA extraído de la cola para determinar el genotipo de cada ratón en posiciones donde se dan polimorfismos de un solo nucleótido (SNPs) distribuidos a lo largo de todo el genoma. En el genotipado de la cohorte F2BX se utilizó una plataforma con 961 SNPs, todos fueron diferentes entre C57BL/6J y CBA/2J y, por tanto, informativos, porque se diseñaron a la carta.

### 3.9.3 Análisis de ligamiento e identificación de regiones de rasgo cuantitativo (QTL).

La información generada por los marcadores de genotipado de la cohorte *backcross* se empleó para la realización del *análisis de ligamiento genético* e identificar regiones genómicas de rasgo cuantitativo (QTL, *Quantitative Trait Loci*) asociadas a la variabilidad en el fenotipo de interés y a que a su vez podrían influir en su determinación. El ligamiento genético hace referencia a la cosegregación, en el momento en que se generan los gametos, de fragmentos genómicos localizados en el mismo cromosoma. Dos marcadores genéticos cualesquiera, SNPs en este caso, localizados en diferentes cromosomas segregan con independencia el uno del otro en el momento de la meiosis. Por ello, cuando se produce la gametogénesis en un individuo heterocigoto en dos marcadores *Aa* y *Bb*, si ambos marcadores están localizados en diferentes cromosomas, las frecuencias esperadas de cada uno de los genotipos en los gametos que se generen serían 25% *AB*, 25% *ab*, 25% *Ab* y 25% *aB*. Pero si la situación fuera que ambos marcadores están en el mismo cromosoma y en uno de los cromosomas homólogos estuviera el genotipo *AB* y en el otro el *ab*, los marcadores producirían cosegregación, con lo que el genotipo de los gametos sería *AB* o *ab*, salvo que se hubiera producido la recombinación entre ambos marcadores en el momento de la meiosis (Figura 3.8).



**Figura 3.8.** Segregación de marcadores genéticos ligados en el momento de la meiosis. Las frecuencias de cada producto de la meiosis representados corresponderían a dos marcadores separados por 10 *centimorgans* (cM), la frecuencia de recombinación entre ambos es del 10%.

Así, los gametos con el genotipo *Ab* o *aB* pasan a denominarse gametos recombinantes. Estos gametos recombinantes tendrían una frecuencia menor del 50% esperado en el caso de que los marcadores estuvieran en diferente cromosoma, por lo que

los marcadores están en desequilibrio de ligamiento. Debido a que la probabilidad de que ocurra un evento de recombinación entre dos marcadores en el proceso meiótico depende, principalmente de la distancia que los separa, la frecuencia de recombinación entre marcadores (se representa como  $\theta$ ) es una medida indirecta de la distancia entre ambos. La *frecuencia de recombinación* es la proporción de gametos recombinantes en función del total de gametos transmitidos. La frecuencia de recombinación máxima es  $\theta=0.5$ , que es la observada entre marcadores localizados en cromosomas diferentes y que no están ligados. Pero, si dos marcadores están ligados, entonces  $\theta < 0.5$  y cuanto menor sea la frecuencia de recombinación entre ambos, el valor de  $\theta$  será menor.

La distancia genética entre marcadores ( $d$ ) depende del número de eventos de recombinación entre éstos, la medida se expresa en *centimorgans* (cM). En nuestro estudio, para el control de calidad de los datos genéticos, se compararon las distancias genéticas estimadas en base a las frecuencias de recombinación entre marcadores observadas en la cohorte *backcross* respecto a la distancia genética correspondiente a las posiciones anotadas para esos marcadores en las bases de datos de referencia *Mouse Genome Informatics* (Cox et al., 2009; Mouse Map Converter, 2017).

El análisis de ligamiento es un método de análisis estadístico que se emplea para determinar si, en relación a los datos genéticos observados, dos marcadores están ligados o segregan de forma independiente. Este análisis es aplicable para determinar la relación entre el genotipo de marcadores genéticos y la distribución de un fenotipo de interés en una población genéticamente heterogénea. Así conseguimos identificar *regiones de rasgo cuantitativo (QTL)*, en éstas, las diferencias en el genotipo se asocian con diferencias en el fenotipo estudiado. Para el análisis de ligamiento empleamos el paquete *R/qtl* en la versión de R.3.3.3 y la interfaz RStudio. Empleamos el método de mapeo por intervalos de máxima probabilidad con el algoritmo esperanza-maximización o algoritmo EM. Para el cálculo del genotipo condicional se empleó la *función de Haldane*, con un tamaño de paso de 2.5 cM y un error de genotipado de 0.001 (Broman et al., 2003).

Empleamos el *LOD-score* para puntuar la significación estadística del ligamiento de los *QTL* encontrados. El *LOD-score* se define como el logaritmo en base 10 de la *Odd Ratio* (o razón de riesgo); ésta es el cociente entre las probabilidades de que los datos observados se hubieran generado bajo la hipótesis de que un locus esté ligado a un rasgo fenotípico, y bajo la hipótesis de que no lo esté. Un *LOD-score* igual a 3 quiere decir que el dato

observado es 1000 veces más probable bajo la hipótesis de que haya ligamiento que bajo la de independencia. Empleamos los criterios establecidos por Lander y Kruglyak (1995) para la consideración del nivel de significación de la asociación observada entre un *QTL* y el fenotipo de interés. Por ello, valores de *LOD-score* superiores a 1.4 se consideró que indicaban la presencia sugestiva de ligamiento, mientras que valores superiores a 3.3 indicaban un ligamiento estadísticamente significativo.

### 3.10 Análisis de datos

Los resultados se han analizado con el paquete estadístico *SIMFIT* en su versión *windows academic 7.3.7* (Bardsley, 2017). Este software incluye todas las opciones estadísticas y gráficas necesarias para el tratamiento de los datos experimentales obtenidos en el presente trabajo. A continuación, se describen brevemente las técnicas estadísticas utilizadas.

#### 3.10.1 Exploración de datos

##### - *Estadística descriptiva*

Con el fin de investigar las características de las distintas variables medidas en las muestras, se realizó un análisis exploratorio de dichas variables. Esta técnica exploratoria es muy útil para observar los valores de tendencia central y precisión de las variables. Así, para cada variable estudiada, se calcularon los parámetros estadísticos habituales como son la media, la mediana, desviación estándar, coeficiente de variación en porcentaje y los gráficos de caja y de barras. También se comprobó si las distintas variables seguían o no una distribución normal, para lo cual se utilizó el test de Kolmogorov-Smirnov y la superposición gráfica de la curva normal teórica al histograma de los datos. En todas las pruebas se consideró que existía una diferencia significativa cuando el p-valor para la aceptación de la hipótesis nula era inferior a 0.05 ( $p < 0.05$ ) (Bardsley, 2017).

##### - *Correlaciones entre variables*

El análisis de correlaciones se usa para estudiar la posible dependencia de dos variables entre sí, con el fin de comprobar si dos variables son completamente independientes o si por el contrario están asociadas de forma linealmente dependiente. Para ello se calcula, a partir de los datos de las dos variables (x,y), el llamado coeficiente *r* de correlación de Pearson, cuyo valor varía entre **-1** y **+1**. Un valor de *r* cercano a **-1** indica

una asociación inversa (cuando una variable aumenta la otra disminuye), siendo los valores cercanos a **0** los que corresponden a una independencia entre las variables (las dos variables están dispersas al azar en una representación x-y), y si los valores de *r* tienden a **+1** existe entonces una dependencia directa entre las variables (cuando una aumenta la otra también aumenta). Para probar si una correlación es estadísticamente significativa se utiliza un test estadístico que sigue una distribución *t de Student*, obteniéndose un p-valor, que si es  $< 0.05$  se concluye que la correlación es estadísticamente significativa (Bardsley, 2017).

### 3.10.2 Análisis multivariante

#### - Análisis de conglomerados (*Cluster analysis (CA)*)

Suele ser muy útil en investigación separar una serie de casos en grupos o categorías. Esta clasificación se basa en algún tipo de similaridad entre los casos según los valores que presenten sus variables. Para ello tiene que establecerse alguna forma de medir la distancia entre los casos, por ejemplo, para la situación de varios casos y dos variables (x,y), podría ser la distancia euclidiana (que para el caso 1 sería la hipotenusa de un triángulo recto con coordenadas  $x_1$  e  $y_1$ ). En general las llamadas técnicas de *agrupamiento o clustering* extienden este concepto de distancia entre casos a situaciones que involucran más de dos variables. Entre estas técnicas, una de las más usadas es el *Análisis de Clusters (CA)*, que hace uso de las representaciones llamadas dendrogramas (en forma de árbol), técnica que describimos brevemente a continuación.

Para clasificar los casos en *clusters* hace falta definir dos conceptos: Una métrica para calcular las distancias entre casos y un procedimiento para fusionar o unir los grupos una vez que el algoritmo los va formando. Existen varias alternativas para definir la distancia entre casos: Euclidiana, Euclidiana al cuadrado, Absoluta, etc. También se han propuesto diferentes opciones para las fusiones de los *clusters* por el algoritmo: promedio de grupo, vecino más cercano, centroide, mediana, etc. En el presente trabajo se han usado las opciones más habituales que son la distancia euclidiana y el promedio de grupo, así como el algoritmo llamado de *clustering* jerárquico aglomerativo, en el cual el primer *cluster* consiste en “n” *clusters* de un caso cada uno y el algoritmo va uniendo los casos en base a las menores distancias entre ellos, hasta llegar a un último *cluster* que engloba los “n” casos.

Cuando el algoritmo ha finalizado se hace una representación en forma de árbol llamada dendrograma, en el que se muestra como se han ido uniendo los casos y *clusters* a

semejanza de las ramas de un árbol. Este dendrograma se puede utilizar para repartir las muestras en subgrupos, para ello se suele definir un umbral arbitrariamente por el investigador que deje unos subgrupos con significado físico. Este umbral se suele representar por una línea de trazo discontinuo sobre el dendrograma (Bardsley, 2017).

- **Análisis de componentes principales (Principal Component Analysis (PCA))**

Se parte de una matriz de  $n$  casos por  $m$  variables  $x_1, x_2, x_3\dots, x_m$ , normalmente correlacionadas, para posteriormente obtener, a partir de ellas, un número  $k$  ( $k < m$ ) de variables incorreladas  $PC_1, PC_2\dots, PC_k$ , denominadas componentes principales (*Principal Components (PC)*). Estas componentes se calculan a partir de una combinación lineal de las variables originales y han de resumir dichas variables con la mínima pérdida de información, es decir, han de capturar un porcentaje de variabilidad importante. Con el fin de poder hacer una representación en el plano, es habitual utilizar solamente las dos primeras componentes principales ( $PC_1$  y  $PC_2$ ) haciendo uso de las puntuaciones (*scores*) de los casos en dichas dos componentes. Estas puntuaciones expresan los datos originales transformados en los valores de las componentes principales según las siguientes combinaciones lineales:

$$PC_1 = a_{11} x_1 + a_{12} x_2 + \dots + a_{1m} x_m$$

$$PC_2 = a_{21} x_1 + a_{22} x_2 + \dots + a_{2m} x_m$$

.....

$$PC_m = a_{m1} x_1 + a_{m2} x_2 + \dots + a_{mm} x_m$$

Es decir, las variables originales ( $x_i$ ) se multiplican por unos factores que expresan la contribución de cada variable a cada componente principal, contribuciones que se denominan cargas ( $a_i$ ). Matemáticamente, las cargas se calculan maximizando productos de matrices cuya explicación excede el presente trabajo, de modo que cada componente principal captura un porcentaje de la variabilidad total de todas las variables de partida, siendo aconsejable que las dos primeras componentes principales ( $PC_1$  y  $PC_2$ ) capturen aproximadamente un 70% de la variabilidad total. Inicialmente, se calculan tantas componentes como variables de partida, pero sólo se retienen las  $k$  componentes principales que expliquen un porcentaje alto de las variables en su conjunto. En concreto  $PC_1$  va a ser

la que más información capture,  $PC_2$  será la siguiente, luego  $PC_3... PC_m$  (Bardsley, 2017). En este trabajo se utilizaron siempre las componentes  $PC_1$  y  $PC_2$  y se analizaron las respectivas puntuaciones de los casos y las cargas de las variables.

- ***Conglomerados por k-medias y PCA (K-means clusters and PCA (KM-PCA))***

Esta técnica es una variante *del análisis de clusters* visto anteriormente, pero con ciertas diferencias que se exponen a continuación. El método utilizado en este trabajo se basa en el algoritmo de *SIMFIT* llamado *K-Means Cluster Analysis*. Este método comienza con una matriz con  $n$  casos y  $m$  variables e intenta asignar estos datos a  $K$  *clusters*. Los *clusters* se crean asignando los casos a aquellos *clusters* que minimizan, dentro de cada *cluster*, la suma de las distancias al cuadrado de los casos a los centroides de los *clusters*. El usuario debe decidir el número  $K$  de grupos que espera, así como los centroides iniciales de cada grupo para que el algoritmo comience las iteraciones de optimización. Primero el usuario carga en el programa una matriz  $n$  por  $m$  con  $n$  casos y  $m$  variables y una matriz adicional con los valores iniciales de cada variable para los centroides. En segundo lugar, el algoritmo realiza varias iteraciones para encontrar la mejor asignación de casos a los *clusters* y proporciona los valores de todas las variables en los centroides finales. En tercer lugar, el usuario puede representar los casos, los *clusters* y los centroides. Desafortunadamente, cuando hay más de dos variables, las gráficas bidimensionales de los *clusters* dependen de las variables particulares elegidas para la representación, y si se cambian las variables a representar puede ocurrir que la gráfica muestre un aspecto diferente, lo que puede dar una impresión engañosa de que algunos casos se podían haber asignado erróneamente. Por lo tanto, cuando hay más de dos variables, es mejor representar, en lugar de las variables originales, las puntuaciones de unas nuevas variables obtenidas por un análisis de componentes principales (*PCA*). Para ello *SIMFIT* realiza con los datos originales un análisis *PCA* y calcula las puntuaciones (*scores*) y las cargas (*loadings*) de los componentes principales. Este análisis de *PCA* se realiza utilizando la técnica de matriz de correlaciones y con puntuaciones de varianza unidad. Las puntuaciones así calculadas para casos y centroides se pueden representar utilizando  $PC_1$  frente a  $PC_2$  como coordenadas y dibujando radios desde los centroides a los casos, lo que permite una mejor visión de los *clusters* generados. La esperanza es que la mayor parte de la variabilidad se pueda explicar por los dos primeros componentes principales ( $PC_1$  y  $PC_2$ ). Los porcentajes de la variabilidad capturada por  $PC_1$  y  $PC_2$  se suele anotar en las leyendas de los ejes de

coordenadas y una captura acumulada en torno al 70% podría considerarse aceptable (Bardsley, 2017).

- **Regresión múltiple lineal por mínimos cuadrados**

Se asume que se ha medido una variable dependiente y que estaría explicada por varias covariables independientes  $x_i$ , es decir se considera que hay que ajustar, por el método de mínimos cuadrados, la siguiente ecuación general:

$$y = B_0 + B_1x_1 + B_2x_2 + \dots B_mx_m$$

donde  $B_0$  es un término constante que se introduce a modo de línea base y  $B_i$  son los coeficientes de cada variable. Estrictamente, las variables  $x_i$  debieran ser cuantitativas, pero a veces se permite el incluir alguna variable dicotómica.

Existen varios procedimientos para encontrar las variables que resultan estadísticamente significativas para el ajuste, así como para estimar la bondad del ajuste. En este trabajo se ha seguido el método sencillo de introducir al principio todas las variables  $x_i$  que se consideraban interesantes para explicar la variable  $y$ . A continuación, se procedía al ajuste y se analizaba la bondad del mismo en base a los p-valores de los coeficientes, al estadístico Cp de Mallows y al p-valor del ANOVA del ajuste. Paulativamente se iban suprimiendo aquellas variables iniciales que no cumplían los criterios de bondad anteriores, hasta alcanzar un número más reducido de variables que proporcionaban un modelo significativo.

Los coeficientes  $B_i$  se consideran que son estadísticamente significativos (distintos de cero) cuando su p-valor, a partir de un test *t-student*, es  $< 0.05$ , o más flexiblemente cuando es  $< 0.1$ . El estadístico Cp de Mallows dividido por el número de parámetros ( $k$ ) del modelo (incluido el término constante) mide la bondad del ajuste, de manera que cocientes  $Cp/k < 1$  indican un buen ajuste y cocientes superiores a 1 indican ajustes deficientes. En cuanto al ANOVA global del ajuste, su interpretación es la clásica de un ANOVA, en el sentido de que cuando el p-valor del test F es  $< 0.05$  se interpreta que al menos uno o más de los parámetros del modelo es significativo y que existe una reducción en la suma de los residuales al cuadrado (Bardsley, 2017).

- **Regresión logística binaria**

Esta estrategia estadística se utiliza cuando la variable dependiente  $y$  es de tipo dicotómico (éxito-fracaso, asignando con **1** a **éxito** y **0** a **fracaso**) y se asume que dicha

variable estaría relacionada con otras covariables  $x_i$  de tipo continuo o categórico. Como la variable  $y$  es dicotómica, sigue ahora una distribución binomial en lugar de normal, por lo que no se puede aplicar un ajuste por mínimos cuadrados. En su lugar, se hace un ajuste por el método de “máxima verosimilitud” a la ecuación:

$$\text{Ln} \frac{p(1)}{(1-p(1))} = B_0 + B_1x_1 + B_2x_2 + \dots B_mx_m$$

donde  $p(1)$  sería la probabilidad de éxito, frente a  $(1-p(1))$  que sería la probabilidad de fracaso. El parámetro  $B_0$  es un término constante a modo de línea base, y los parámetros  $B_i$  son los coeficientes ajustados para cada covariable “ $i$ ”.

De manera análoga a lo visto en el epígrafe anterior, el procedimiento de ajuste consiste en introducir inicialmente al algoritmo todas las variables a probar y a continuación ir suprimiendo aquellas variables que no sean significativas en base a su  $p$ -valor, que debe ser  $< 0.05$  o de una manera más flexible  $< 0.1$  (Bardsley, 2017).

Esta técnica estadística tiene un gran valor predictivo para nuevos casos, ya que midiendo las covariables  $x_i$  de un nuevo caso se puede predecir la probabilidad binomial de éxito ( $p(1)$ ) que tendría dicho caso utilizando la ecuación:

$$p(1) = \frac{1}{1 + e^{(B_0 + B_1x_1 + B_2x_2 + \dots B_mx_m)}}$$

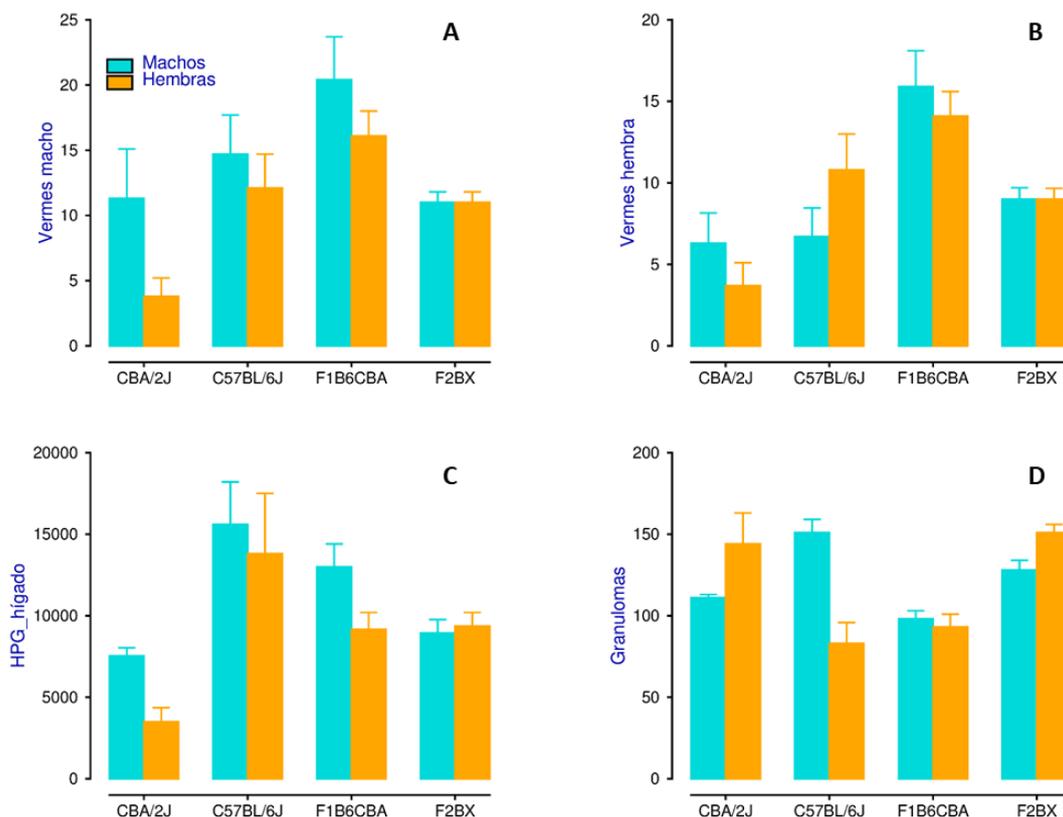
## 4 Resultados

Partimos de la generación de una cohorte de ratones con heterogeneidad genética, esto sirve como modelo de estudio capaz de reproducir, de manera simplificada, la diversidad genética que muestran las poblaciones humanas y permite caracterizar la susceptibilidad o resistencia a la infección por *S. mansoni*. Generamos la cohorte de ratones mediante cruce retrógrado o *backcross* de una cepa singénica CBA/2J, susceptible a la enfermedad, y la F1 (F1B6CBA) procedente del cruce de dos cepas singénicas (CBA/2J x C57BL/6J), con susceptibilidad intermedia. Mediante el cruce retrógrado obtenemos una cohorte F2BX cuya heterogeneidad genética provoca variabilidad fenotípica en la respuesta a la infección por *S. mansoni*. Para caracterizar la susceptibilidad, evaluamos la enfermedad en libre evolución, sin tratamientos y analizando diferentes patofenotipos de la enfermedad. Estos patofenotipos son rasgos complejos y se manifiestan por la interacción de diversos fenotipos intermedios a distintos niveles biológicos. Por ello, medimos una serie de magnitudes de la esquistosomosis que pueden describir los fenotipos intermedios, para intentar encontrar las diferencias entre los patofenotipos individuales y ver como se asocia a esa variabilidad. Para finalizar, como la heterogeneidad genética de la cohorte F2BX influye en la variabilidad fenotípica, exploraremos en busca de regiones QTL donde las diferencias genotípicas se ligan a la variabilidad, tanto en los fenotipos intermedios como en los patofenotipos analizados.

#### **4.1 Susceptibilidad a la infección por *Schistosoma mansoni* en la cohorte F2BX**

Al comparar las características fenotípicas de las cepas parentales, de la generación F1B6CBA y del *backcross* F2BX, tanto en ratones machos como en hembras (Figura 4.1), observamos que la cohorte F2BX presenta, en cuanto al número de vermes, tanto machos como hembras, números intermedios respecto a las cepas parentales y menores que los de la cohorte F1B6CBA, pero las medias son iguales entre los grupos F2BX macho y F2BX hembra, mientras que en las cepas parentales y la cohorte F1B6CBA hay diferencias entre ratones macho y hembra (Figuras 4.1 A y B). En cuanto al número de huevos por gramo de hígado, la media es menor que en la cohorte F1B6CBA, que presenta datos intermedios a

las cepas parentales (Figura 4.1 C). El número de granulomas en la cohorte F2BX, es superior al de la cohorte F1B6CBA y más parecido al de la cepa parental más susceptible CBA/2J, con mayor número de granulomas en las hembras, mientras que la cepa parental C57BL/6J, presenta mayor número de granulomas en los ratones macho (Figura 4.1 D).

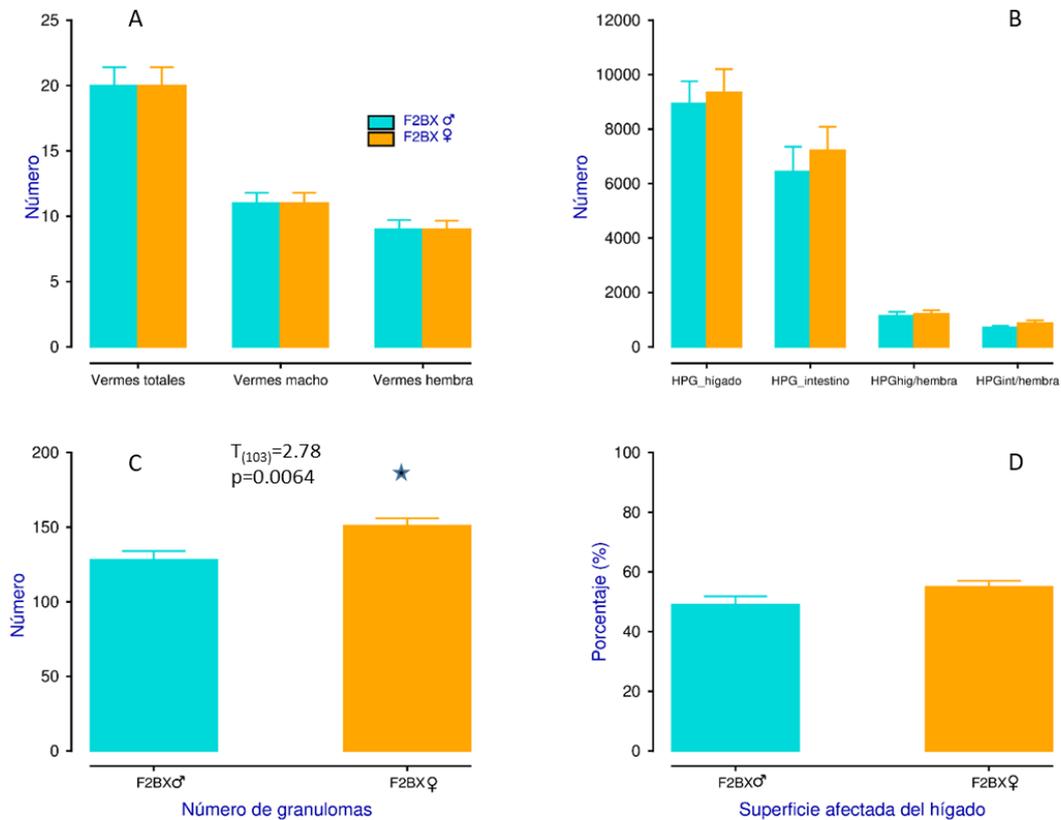


**Figura 4.1.** Comparación entre los estudios previos de las cepas parentales, el cruce en la cohorte F1B6CBA, y con los datos de nuestro cruce retrógrado o *backcross* (F2BX), en cuanto al recuento de (A) vermes macho, (B) vermes hembra, (C) huevos por gramo de hígado y (D) granulomas. Se han representado las medias y el límite de confianza superior correspondiente al error estándar de la media (EEM).

#### 4.1.1 Análisis exploratorio de los datos recopilados de la cohorte F2BX por sexos

Para estudiar las características de las variables analizadas en los ratones procedentes de la cohorte F2BX (N=105), se midieron las siguientes variables: número de vermes macho y hembra, huevos por gramo de hígado e intestino, fecundidad de los vermes hembra en hígado e intestino (calculada como el cociente entre el número de huevos por gramo de hígado e intestino y el número de vermes hembras), recuento de granulomas y estimación

de la superficie hepática afectada entre individuos de la cohorte F2BX machos y F2BX hembras, para observar si había diferencias entre ellos (Figura 4.2). Tras la realización del test t de Student de datos desapareados, sólo se detectó diferencia estadísticamente significativa entre las medias del número de granulomas entre F2BX machos y F2BX hembras ( $T_{(103)}=2.78$ ,  $p=0.0064$ ). Esto nos indica que, considerados globalmente todos los ratones de la cohorte F2BX en cuanto al sexo, sólo existen diferencias fenotípicas en el número de granulomas que presentan.



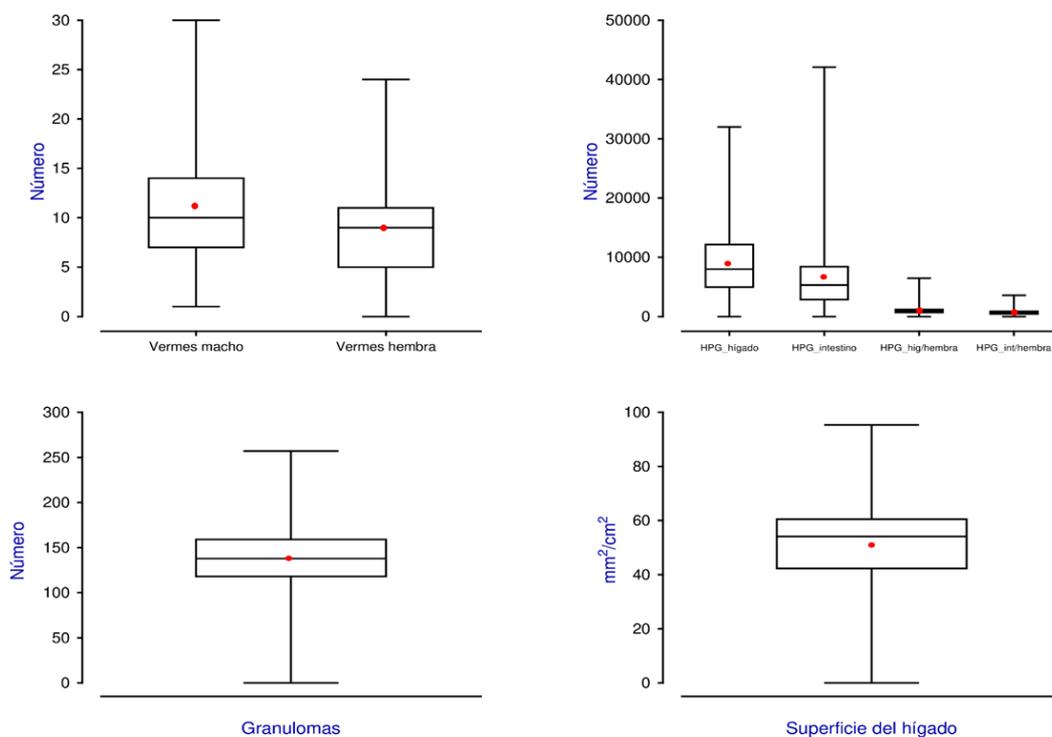
**Figura 4.2.** Representación de las variables parasitológicas (A y B) y patológicas (C y D) entre ratones machos y hembras. Se han representado las medias y el límite de confianza superior correspondiente al error estándar de la media (EEM). El estadístico T corresponde a un test t de Student entre las dos muestras. \*  $p<0.05$ .

**Tabla 4.1.** Datos de la media y el error estándar de la media (EEM) en las variables parasitológicas y patológicas entre ratones F2BX machos y F2BX hembras. Superficie afectada del hígado (Superficie), huevos por gramo de hígado (HPG\_hig), huevos por gramo de intestino (HPG\_int), fecundidad de las hembras en el hígado (HPG\_hig/H), fecundidad de las hembras en el intestino (HPG\_int/H) y error estándar de la media (EEM). El asterisco corresponde al estadístico T del test t de Student \*  $p < 0.05$

	F2BX machos		F2BX hembras	
	Media	EEM	Media	EEM
Granulomas	128	6.41	151*	5.13
Superficie	49%	2.75	55%	2.03
Vermes macho	11	0.75	11	0.82
Vermes hembra	9	0.70	9	0.66
HPG_hígado	8930	834	9348	833
HPG_intestino	6441	910	7218	870
HPG_hig/H	1140	149	1207	138
HPG_int/H	710	72	875	104

#### 4.1.2 Análisis descriptivo de las variables analizadas en la cohorte F2BX en conjunto

Con el fin de explorar en profundidad las variables parasitológicas y patológicas, se hizo un estudio de estadística descriptiva de las mismas, con la intención de caracterizar su variabilidad y sus valores de tendencia central. Los resultados se han representado en la Figura 4.3 en forma de diagramas de cajas para las diferentes variables estudiadas. Cada diagrama incluye los datos de la media y la mediana de cada variable, así como sus respectivos valores extremos. Como puede observarse en la figura, existe una gran variabilidad entre los datos recogidos para los distintos ratones.



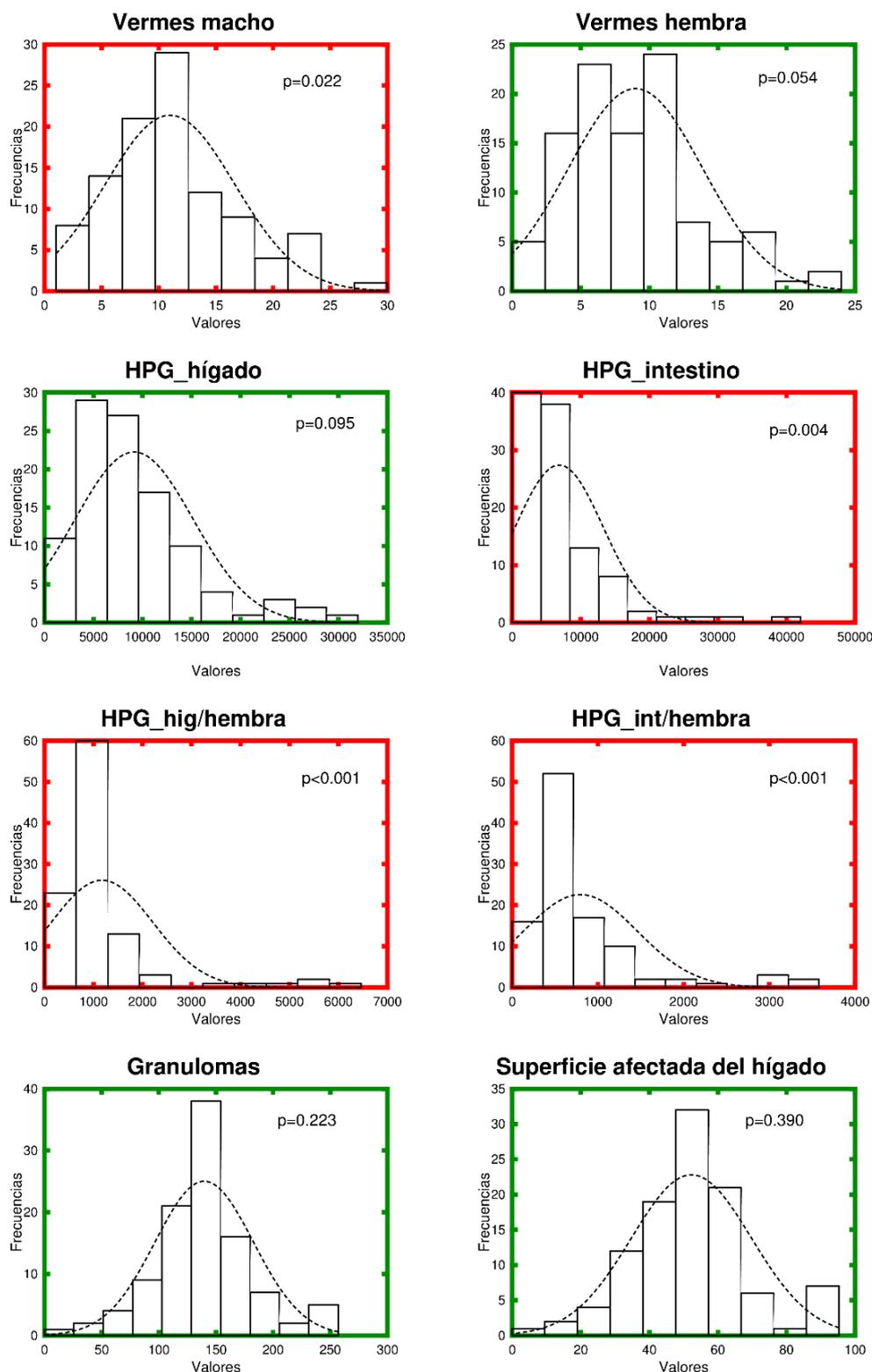
**Figura 4.3.** Estudio exploratorio de las variables parasitológicas y patológicas para la cohorte F2BX. Se han representado la mediana como una línea en el interior de la caja y la media como un círculo. Los segmentos superior e inferior representan los valores extremos superior e inferior del rango de la variable.

Se analizó si las variables seguían una distribución normal, para ello se investigó el tipo de distribución que seguían las variables parasitológicas y patológicas (Figura 4.4), así como de las variables que medían las poblaciones celulares (Figura 4.5) y de las variables que medían la respuesta inmunológica mediada por anticuerpos (Figura 4.6). Se utilizó el test de Kolmogorov-Smirnov y la superposición gráfica de la curva normal teórica al histograma de los datos. En todos los casos se consideró una diferencia significativa cuando el p-valor era inferior a 0.05 ( $p < 0.05$ ), es decir, se rechazaba la hipótesis nula de que la muestra siguiera una distribución normal.

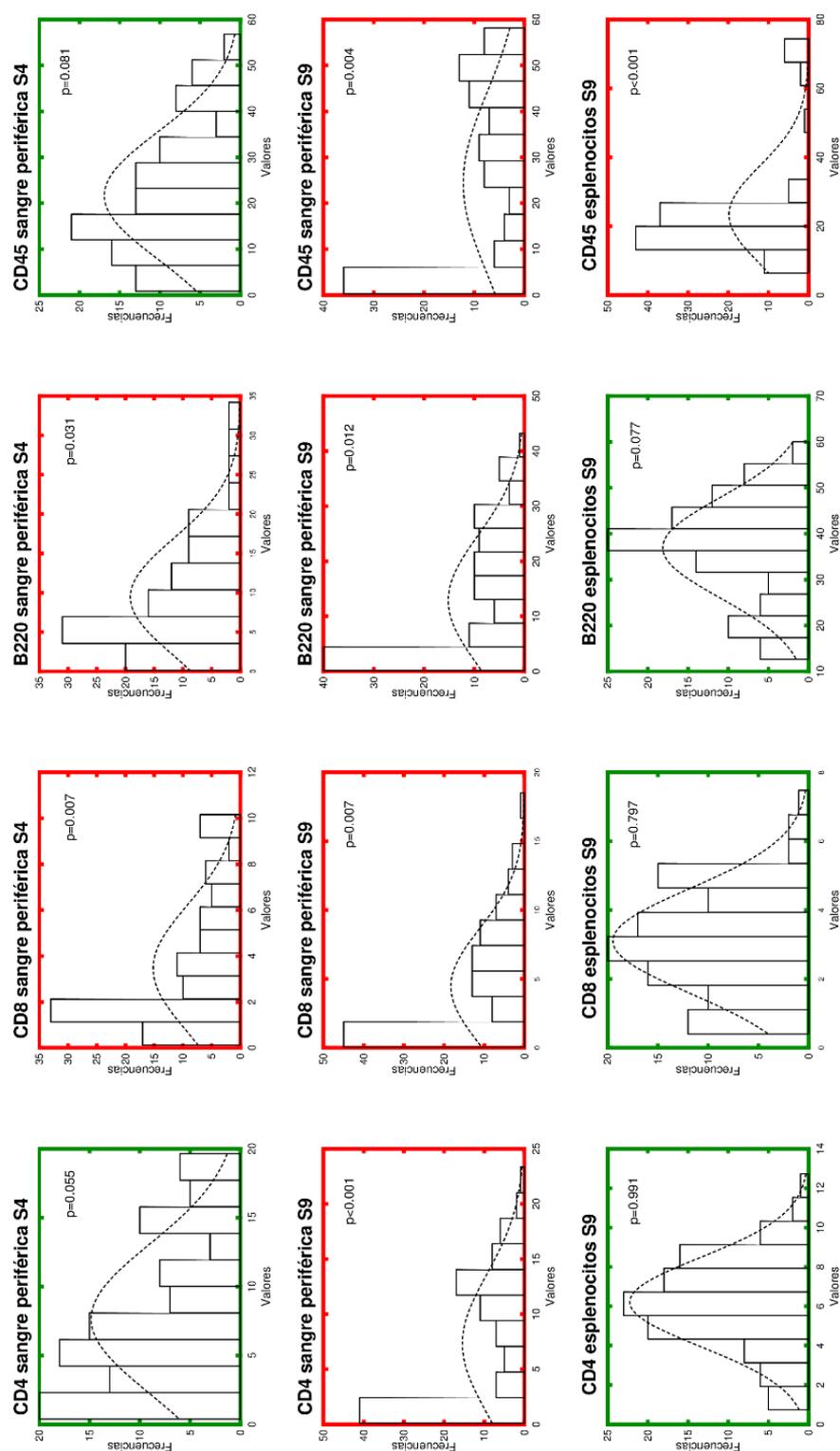
Como puede observarse en la Figura 4.4, dentro de las variables parásito-patológicas, el número de vermes hembra, el número de huevos por gramo de hígado, el número de granulomas y la superficie afectada del hígado sí que siguen una distribución normal, mientras que las variables que miden los huevos por gramo de intestino (HPG-int) y la fecundidad de las hembras en hígado e intestino (HPG\_hig/hembra y HPG\_int/hembra) son las que más se alejan de la distribución normal, estando sesgadas a la izquierda con un predominio de los valores más bajos. Este análisis de normalidad de variables ha revelado cuáles eran las variables que seguían la distribución normal esperada habitualmente para las variables biológicas y cuales otras presentaban sesgos significativos.

Por su parte, en la Figura 4.5 se han representado las pruebas de normalidad para las variables que miden las poblaciones celulares (CD4, CD8, B220 y CD45) en sangre periférica y en esplenocitos. En dicha figura se aprecia como en la semana 4 en sangre periférica siguen una distribución normal las poblaciones celulares CD4 y CD45, no así las distribuciones de CD8 ni de B220, que están sesgadas a la izquierda. En la semana 9 en sangre periférica, en ningún caso se observan distribuciones normales de las variables. En las poblaciones celulares en esplenocitos en la semana 9, siguen distribución normal todas las variables excepto la de CD45, que está sesgada hacia los valores inferiores.

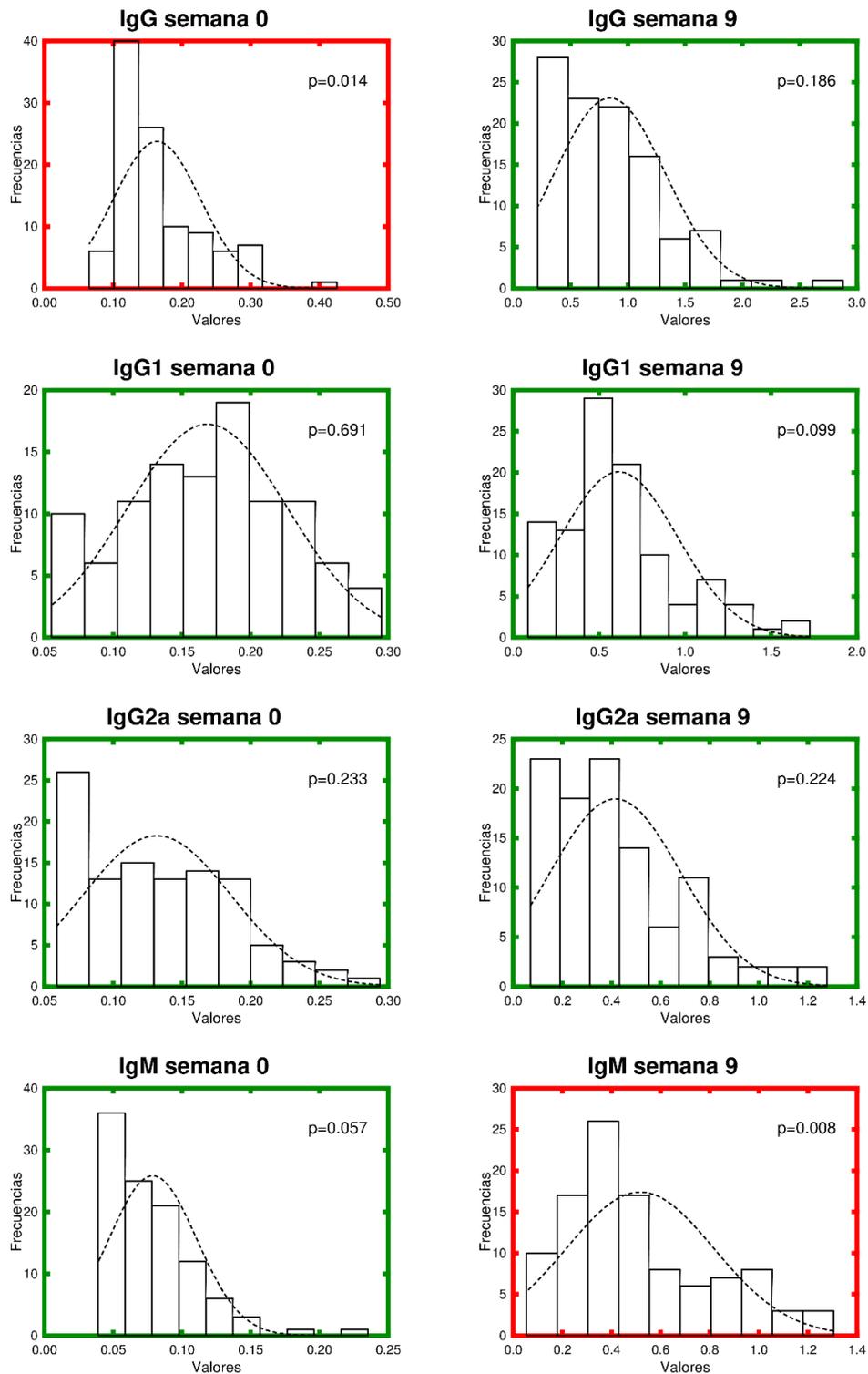
Respecto a las variables que miden la respuesta inmune humoral (IgG, IgG1, IgG2a e IgM), los resultados de las pruebas de normalidad se han representado en la Figura 4.6. En la semana 0 siguen distribución normal todas las variables excepto la IgG, cuya distribución está sesgada hacia la izquierda. En cuanto a la semana 9 todas las variables siguen distribución normal, excepto la IgM que presenta sesgo hacia los valores más bajos.



**Figura 4.4.** Pruebas de normalidad de las variables parasitológicas y patológicas. Se ha empleado el test de Kolmogorov-Smirnov y se han representado con recuadro verde las distribuciones que siguen una distribución normal y con recuadro rojo las que se alejan de dicha distribución. En todos los paneles los datos figuran en forma de histograma de frecuencias, al cual se ha superpuesto la curva teórica normal que le correspondería. Se ha incluido también el p-valor del test aplicado.



**Figura 4.5.** Pruebas de normalidad de las variables que miden las poblaciones celulares en sangre periférica y en esplenocitos. Se ha usado el test de Kolmogorov-Smirnov y se han representado con recuadro verde las distribuciones que siguen una distribución normal y con recuadro rojo las que se alejan de la distribución normal. En todos los paneles los datos figuran en forma de histograma de frecuencias, al cual se ha superpuesto la curva teórica normal que le correspondería. Se ha incluido también el p-valor del test aplicado.



**Figura 4.6.** Pruebas de normalidad de las variables inmunológicas empleando el test de Kolmogorov-Smirnov. Se han representado con recuadro verde las distribuciones que siguen una distribución normal y con recuadro rojo las que se alejan de la distribución normal. En todos los paneles los datos figuran en forma de histograma de frecuencias, al cual se ha superpuesto la curva teórica normal que le correspondiera. Se ha incluido también el p-valor del test aplicado.

### 4.1.3 Análisis de correlación de variables en la cohorte F2BX

Se realizó un estudio de correlaciones entre variables, con el fin de detectar la posible asociación de 2 variables entre sí. Se analizaron 21 variables seleccionadas del total de las 29 medidas inicialmente, debido a que las variables de poblaciones celulares en la semana 4 y las inmunoglobulinas en la semana 0 no resultaron muy indicativas. A partir de los datos de cada 2 variables (x,y) se calculó el coeficiente de correlación de Pearson (r), cuyo valor oscila entre -1 y 1. Un valor cercano a 1 indica una dependencia directa entre las variables y un valor cercano a -1 indica una asociación inversa, siendo los valores cercanos a 0 los que mostrarían independencia entre las variables. A partir del valor de “r” y mediante un test t de Student, se obtiene un p-valor, que indica si la correlación es significativa ( $p < 0.05$ ).

En la Figura 4.7.A se ha representado un mapa que muestra los resultados de las correlaciones. Las variables con una correlación estadísticamente significativa ( $p < 0.05$ ) y con asociación directa entre ellas ( $r > 0$ ) figuran en color rojo, pero si es inversa ( $r < 0$ ) figuran en color azul. El color blanco representa las correlaciones que no son significativas ( $p > 0.05$ ). A continuación, se analizan las correlaciones significativas y no significativas más importantes.

Como observamos en el panel A de la Figura 4.7.B, el número de granulomas presenta una correlación directa con la superficie afectada del hígado ( $\text{mm}^2/\text{cm}^2$ ), así como con el número de vermes total y de vermes macho; los vermes hembra se quedan con un p-valor muy cercano a la correlación significativa ( $p = 0.064$ ). Existe una correlación directa del número de granulomas con respecto al número de huevos por gramo de hígado (HPG\_hig) y la fecundidad de las hembras en el hígado (HPG\_hig/H). La superficie afectada del hígado ( $\text{mm}^2/\text{cm}^2$ ) presenta correlación en sentido directo con el resto de variables parásito-patológicas, aunque la correlación más fuerte de la superficie afectada del hígado es con el número de granulomas. El número de vermes total, de vermes macho y de vermes hembra están muy relacionados entre ellos, así como con el número de huevos en el hígado (HPG\_hig) e intestino (HPG\_int), pero presentan una correlación inversa con la fecundidad de las hembras en el hígado (HPG\_hig/H), lo que implica que, a mayor número de vermes, menor fecundidad. El número de huevos en hígado e intestino están correlacionados directamente con la fecundidad en hígado (HPG\_hig/H) e intestino

(HPG\_int/H), excepto el número de huevos por gramo de hígado y la fecundidad en el intestino que no presentan una correlación significativa.

En el panel B (Figura 4.7.B) observamos las correlaciones entre las poblaciones celulares. Las poblaciones celulares CD4, CD8, B220 y CD45, presentan correlación directa entre sí tanto en sangre periférica como en esplenocitos; si bien, en esplenocitos no existe correlación entre las poblaciones de B220 y CD45. Respecto a las poblaciones celulares en sangre periférica y en esplenocitos, se detecta correlación directa entre todas ellas, excepto en los CD45 de esplenocitos que no presentan correlación con ninguna población en la sangre periférica; a su vez las B220 de esplenocitos tampoco presentan correlación con los CD8 de sangre periférica.

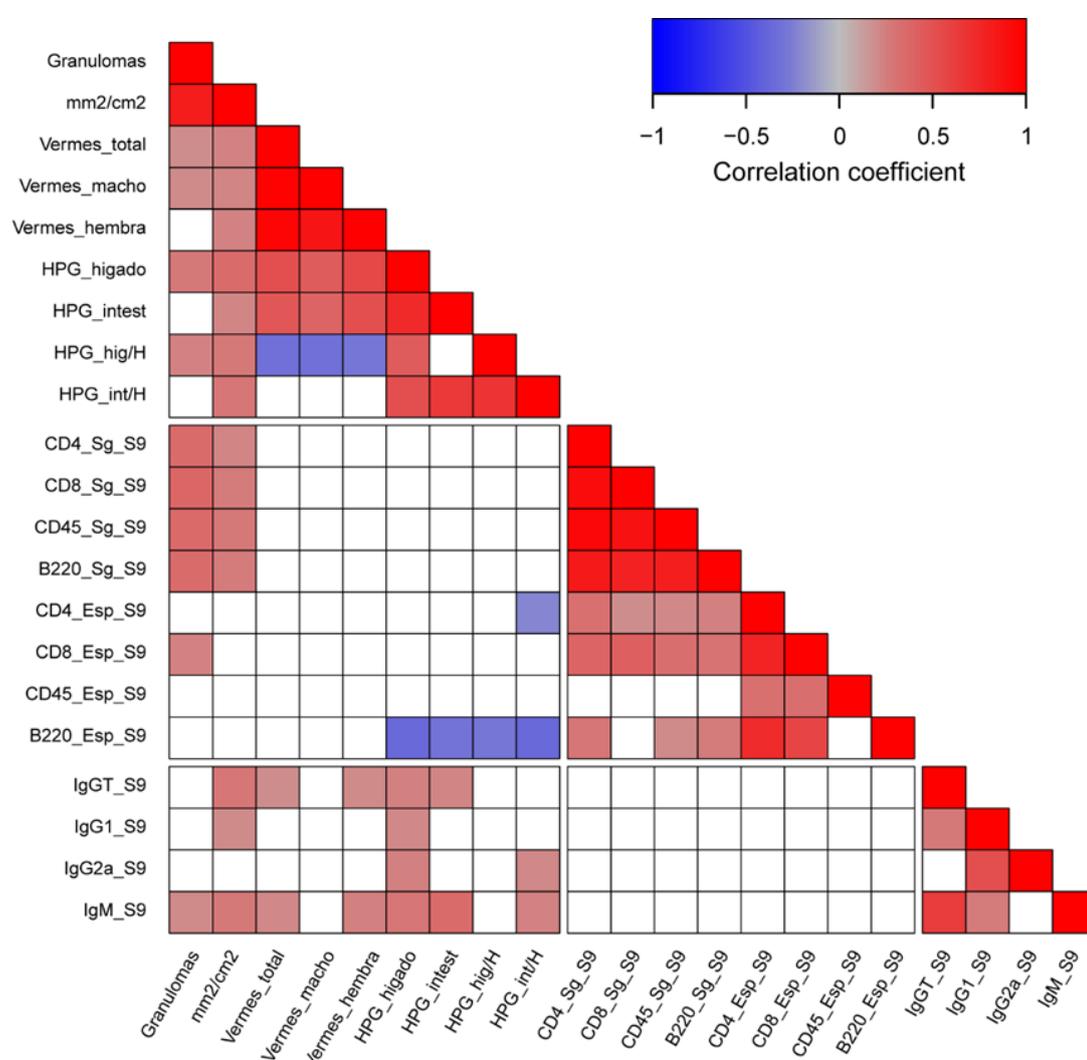
En el panel C (Figura 4.7.B) se muestran las correlaciones entre las diferentes inmunoglobulinas. Las inmunoglobulinas (IgG, IgG1, IgG2a e IgM) presentan correlación directa entre ellas, como era de esperar, excepto la IgG2a con la IgG e IgM que no presentan correlación alguna ( $p > 0.05$ ). Esto podría estar relacionado con los diferentes tipos de respuestas inmunológicas que se dan en la esquistosomosis.

En el panel D (Figura 4.7.B) aparecen las correlaciones entre las variables parásito-patológicas y las poblaciones celulares. Se observa una correlación directa entre el número de granulomas y todas las poblaciones celulares (CD4, CD8, B220 y CD45) en sangre periférica y con la población de CD8 en esplenocitos. En cuanto a la superficie afectada del hígado ( $\text{mm}^2/\text{cm}^2$ ) vemos que presenta correlación directa con las poblaciones celulares en la semana 9 en sangre periférica. El número de vermes totales, de vermes macho y de vermes hembra, no presentan correlaciones significativas con las poblaciones celulares (CD4, CD8...). Por otra parte, no se observa asociación entre el número de huevos y la fecundidad y las poblaciones celulares en sangre, pero sí existe correlación inversa con la población B220 en los esplenocitos, lo que significa que un mayor número de células B220 está relacionado con un menor número de huevos. También presenta asociación inversa entre CD4 en esplenocitos y la fecundidad en el intestino (HPG\_int/H).

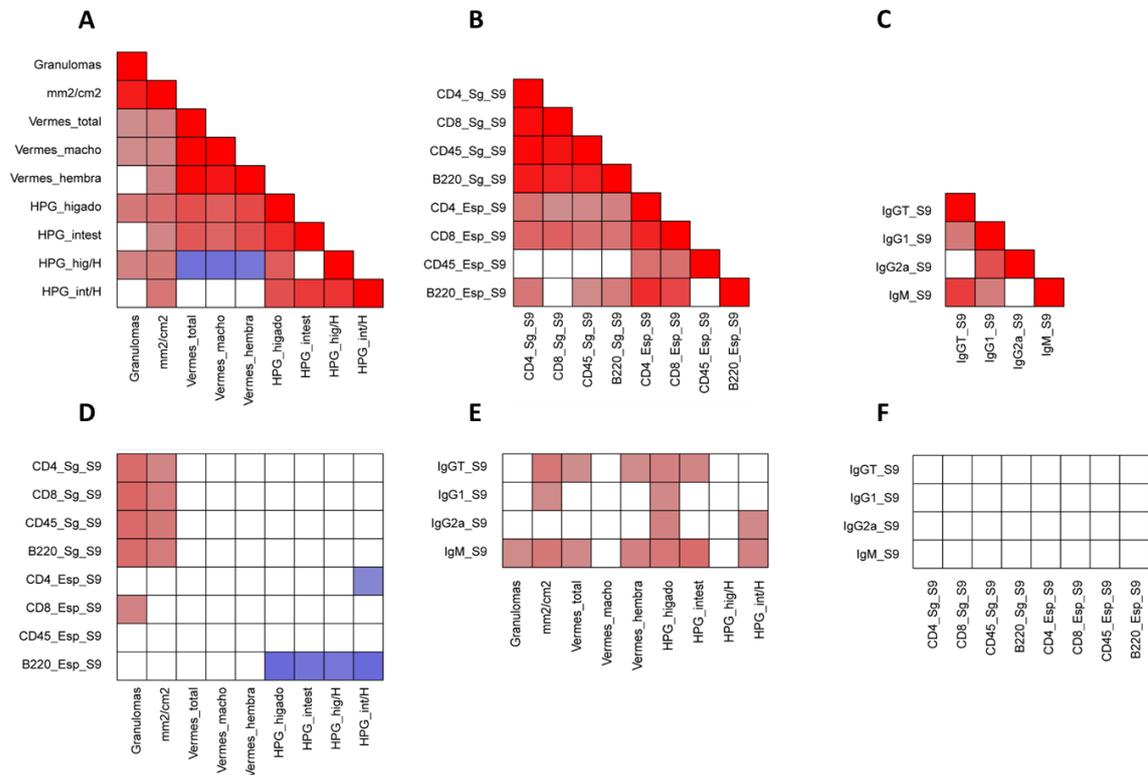
En el panel E (Figura 4.7.B) se observan las asociaciones entre las variables parásito-patológicas y las inmunoglobulinas. Se observa una correlación directa entre el número de granulomas y la IgM. La superficie afectada del hígado ( $\text{mm}^2/\text{cm}^2$ ), presenta correlación directa con la IgG, IgG1 e IgM. El número de vermes total y de vermes hembra están directamente correlacionados con la IgG e IgM. El número de huevos por gramo de hígado

presenta correlación con las cuatro inmunoglobulinas, lo cual resulta coherente con el aumento de los niveles de inmunoglobulinas como resultado de la respuesta inmune a la formación de granulomas. Por su parte, el número de huevos por gramo de intestino presenta correlación directa con la IgG e IgM. También se observa que la fecundidad de las hembras en el intestino presenta correlación directa con la IgG2a e IgM.

Finalmente, en el panel F (Figura 4.7.B) se observa la ausencia de asociación entre todas las poblaciones celulares y las inmunoglobulinas.



**Figura 4.7.A.** Análisis de correlaciones entre variables. Dentro de las correlaciones que muestran un valor estadísticamente significativo ( $p < 0.05$ ), la correlación puede ser directa ( $r > 0$ ) y se ha representado en color rojo o inversa ( $r < 0$ ) representada en color azul. En blanco se muestran las correlaciones entre variables que no son estadísticamente significativas ( $p > 0.05$ ).



**Figura 4.7.B.** Análisis de correlación entre variables. A) variables parasito-patológicas con ellas mismas, B) variables de poblaciones celulares con ellas mismas, C) variables de inmunoglobulinas con ellas mismas, D) variables parasito-patológicas con poblaciones celulares, E) variables parasito-patológicas con inmunoglobulinas y F) variables de poblaciones celulares con inmunoglobulinas. Dentro de las correlaciones que muestran un valor estadísticamente significativo ( $p < 0.05$ ), la correlación puede ser directa ( $r > 0$ ) y se ha representado en color rojo o inversa ( $r < 0$ ) representada en color azul. En blanco se muestran las correlaciones entre variables que no son estadísticamente significativas ( $p > 0.05$ ).

## 4.2 Análisis multivariante de los datos obtenidos con la cohorte F2BX

Se emplearon técnicas de análisis multivariante para agrupar los ratones en grupos o *clusters*, así como para analizar la influencia de las variables en la clasificación de los posibles grupos.

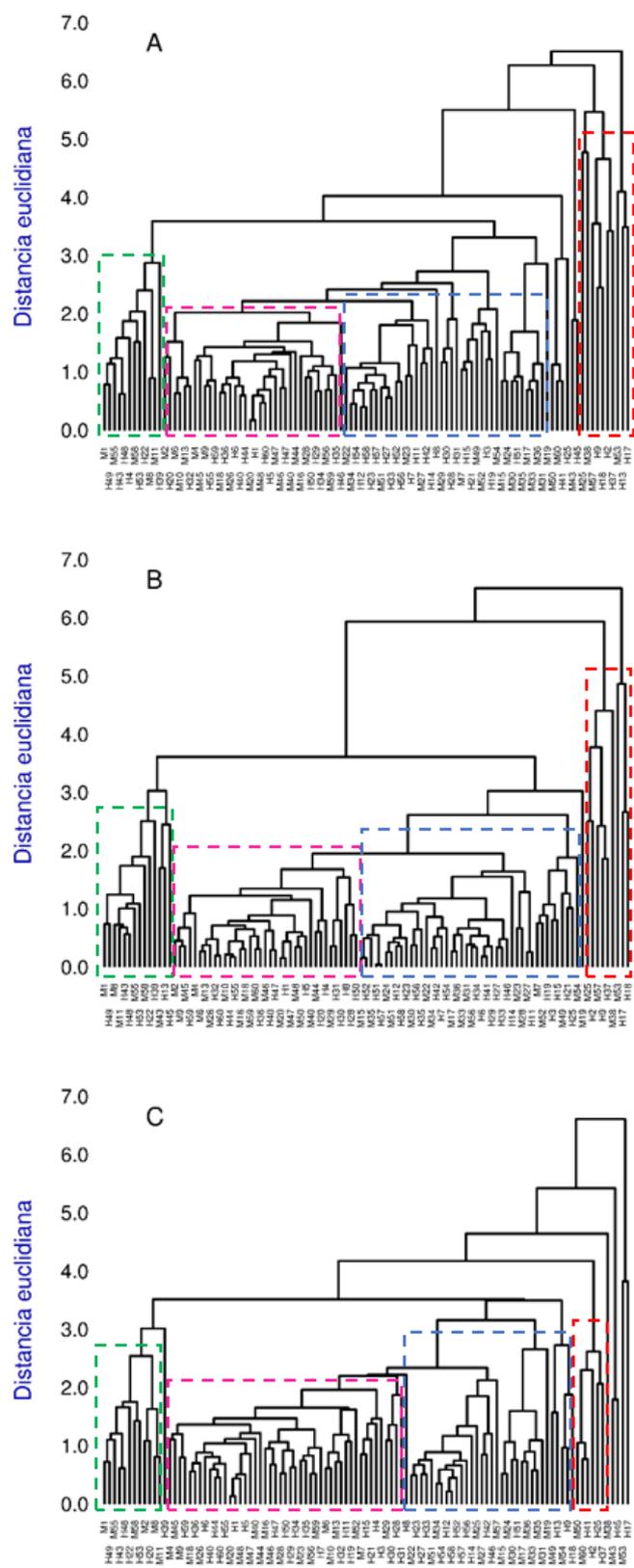
Por una parte, se estudió la similaridad entre los ratones según los valores que presentan en sus diferentes variables, tanto parasitológicas como patológicas. Como variables parasitológicas se emplearon el número de vermes macho y hembra, el número de huevos por gramo de hígado e intestino y la fecundidad de las hembras en hígado e intestino. Como variables patológicas se analizaron el número de granulomas y la superficie afectada del hígado ( $\text{mm}^2/\text{cm}^2$ ). Como técnicas de agrupamiento o *clustering* se han

empleado el “Análisis de Conglomerados” y el algoritmo de “Conglomerados por k-medias”. Por otra parte, se hizo un análisis de las variables estudiadas mediante el procedimiento de “Análisis por Componentes Principales”. Los resultados obtenidos con todas estas técnicas se describen a continuación.

#### **4.2.1 Análisis de conglomerados o clusters (*Cluster Analysis (CA)*)**

En esta técnica se calcula una matriz con las distancias entre los casos y a continuación se van agrupando por similitud entre ellos, de forma que los casos semejantes en cuanto al valor de sus variables forman agrupaciones de “ramas” que en conjunto forman un “árbol” (dendrograma), en el que las ramas cercanas se interpretan como diferentes subgrupos dentro de las muestras analizadas. Se llevó a cabo el estudio de las variables de tres formas: con todas las variables juntas (parasitológicas y patológicas), sólo con las variables parasitológicas y, por último, con las variables parasitológicas y patológicas (variables parásito-patológicas) pero sin incluir la fecundidad de las hembras en el hígado e intestino. Debido a la diferencia de magnitud de las variables, en todos los cálculos se hizo una transformación previa, estandarizando todas las variables a media cero y desviación estándar uno. Los resultados de estos tres estudios aparecen recogidos en la Figura 4.8.

La forma habitual de interpretar estos dendrogramas es levantar una línea horizontal que corte el “árbol” a una distancia determinada según el criterio del investigador. De dicha línea hacia abajo se observan los grupos de ramas y se propone el número de subgrupos que se han separado. En el caso que nos ocupa, al tratarse de ratones que presentan una gran variabilidad, es de esperar que tal procedimiento no sea fácilmente aplicable y que aparezcan algunos ratones “atípicos”. Teniendo estas dificultades, se ha recurrido a observar los subgrupos de “ramas” que se encuentran más cercanos entre sí en el dendrograma y que engloban un elevado número de ratones. Según este criterio parecen existir cuatro subgrupos de ratones que se han señalado con trazo discontinuo en la Figura 4.8 en sus tres paneles A, B y C.



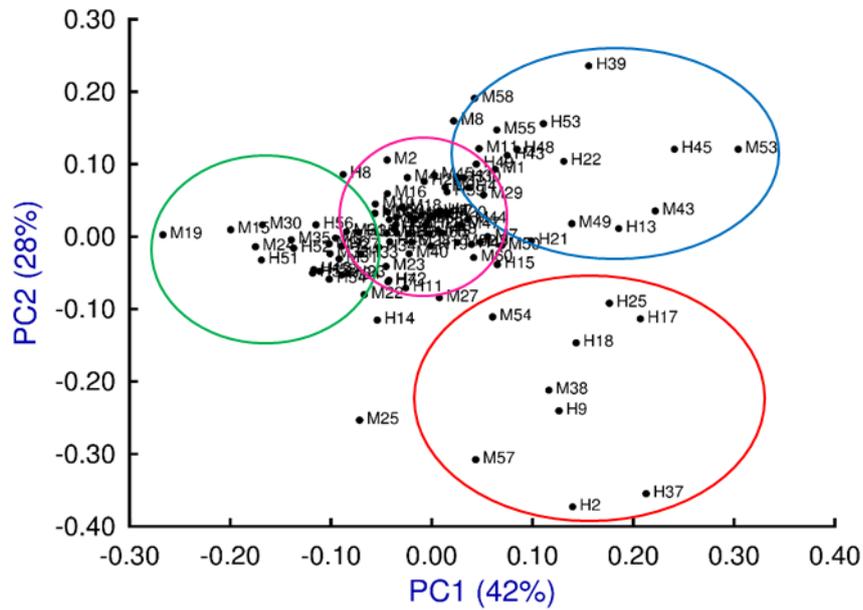
**Figura 4.8.** A) Dendrograma con todas las variables parasitológicas y patológicas. B) Dendrograma sólo con las variables parasitológicas. C) Dendrograma con las variables parasitológicas y patológicas, pero sin la fecundidad de las hembras.

#### 4.2.2 Análisis de componentes principales (*Principal Component Analysis (PCA)*)

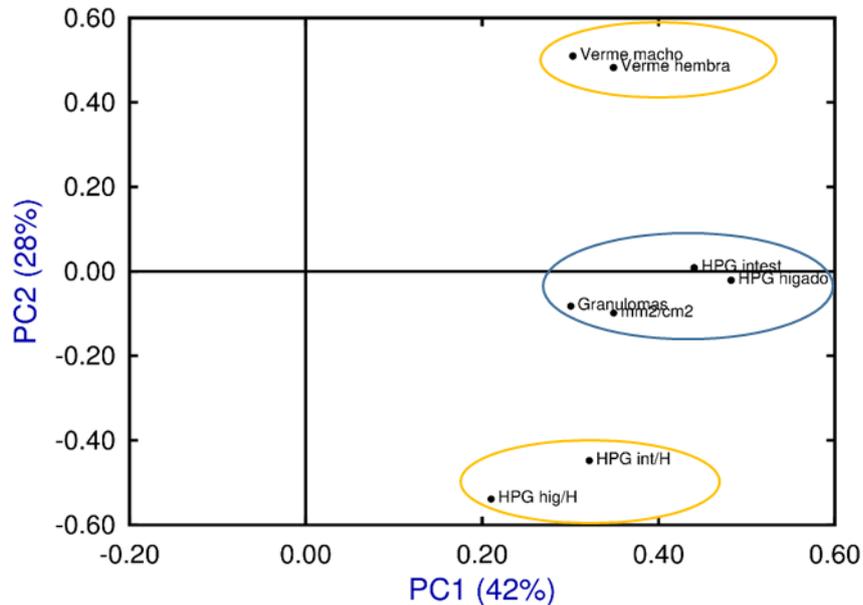
Como se ha expuesto en el apartado de Materiales y Métodos, esta técnica estadística trata de “extraer” la información contenida en muchas variables correlacionadas, para condensarlas en unas pocas variables ficticias, llamadas componentes principales (PC), las cuales pueden servir para mostrarnos cómo se comportan globalmente los casos y las variables. Normalmente, se trabaja con las dos primeras componentes, que son las que más información capturan y que se obtienen como una combinación lineal de las variables originales. En base a estas dos componentes principales, se pueden representar las puntuaciones (*scores*) de los casos en dos dimensiones (PC1 frente a PC2), y a partir de esta representación se puede deducir si los casos se agrupan de alguna forma, separándose en grupos o *clusters*. Así mismo, la técnica PCA proporciona el peso estadístico o carga (*loading*) con el que cada variable original entra a formar parte de cada componente principal, lo cual permite caracterizar la mayor importancia de unas variables frente a otras.

Se hizo un estudio por componentes principales empleando las variables parasitológicas y patológicas conjuntamente, ya que muestran un elevado grado de correlación entre ellas, tanto en asociación directa como inversa. Para evitar las diferencias en magnitud de las distintas variables, se estandarizaron todas ellas a media cero y desviación estándar uno. Los resultados se muestran en la Figura 4.9, representándose por un lado las puntuaciones de los casos (panel A) y por otro las cargas de las variables (panel B). En la representación de los casos, parece observarse la separación de los ratones en cuatro posibles grupos, en la línea de los cuatro grupos propuestos en el apartado anterior con el *análisis de clusters*. En la representación de las cargas, vemos que las dos variables patológicas, es decir, el número de granulomas y la superficie afectada del hígado ( $\text{mm}^2/\text{cm}^2$ ), así como las variables parasitológicas número de huevos por gramo de hígado (HPG hígado) e intestino (HPG intest), tienen un peso importante en la PC1. Por su parte, las otras cuatro variables parasitológicas (número de vermes macho y hembra, y la fecundidad de las hembras en el hígado (HPG hig/H) e intestino (HPG int/H)), tienen un peso importante en la PC2 y están inversamente relacionadas entre ellas dos a dos, es decir, a mayor número de vermes, menor fecundidad en hígado e intestino.

### A Análisis de componentes principales PCA



### B Cargas en PCA



**Figura 4.9.** Análisis de componentes principales (PCA). A) Representación de los casos con los 4 probables grupos en que se integren los ratones de la cohorte F2BX. B) Representación de las cargas de las variables. La variabilidad capturada fue del 70%.

#### 4.2.3 Estudio de conglomerados por k-medias y PCA (*k-means clusters and PCA*)

Como se ha comentado en los dos apartados anteriores, el análisis exploratorio con el análisis de conglomerados y de componentes principales, apunta hacia la existencia de

cuatro subgrupos de ratones en la cohorte F2BX. Para intentar caracterizar estos cuatro grupos, se empleó el algoritmo de conglomerados por k-medias. Con este método, el investigador comienza asignando unos valores medios iniciales a todas las variables para cada uno de los cuatro grupos por separado, tanto para las variables parasitológicas como patológicas (8 variables en total). Con estas estimas iniciales, el algoritmo comienza a iterar y a mover los casos de unos grupos a otros según proceda, generando al final los cuatro grupos o *clusters* a los que mejor queda asignado cada uno de los casos. Las medias iniciales proporcionadas al algoritmo se calcularon a partir de cuatro ratones tipo que presumiblemente pertenecían a los distintos grupos. Estas estimas iniciales para cada una de las variables en los cuatro grupos se muestran en la Tabla 4.2.

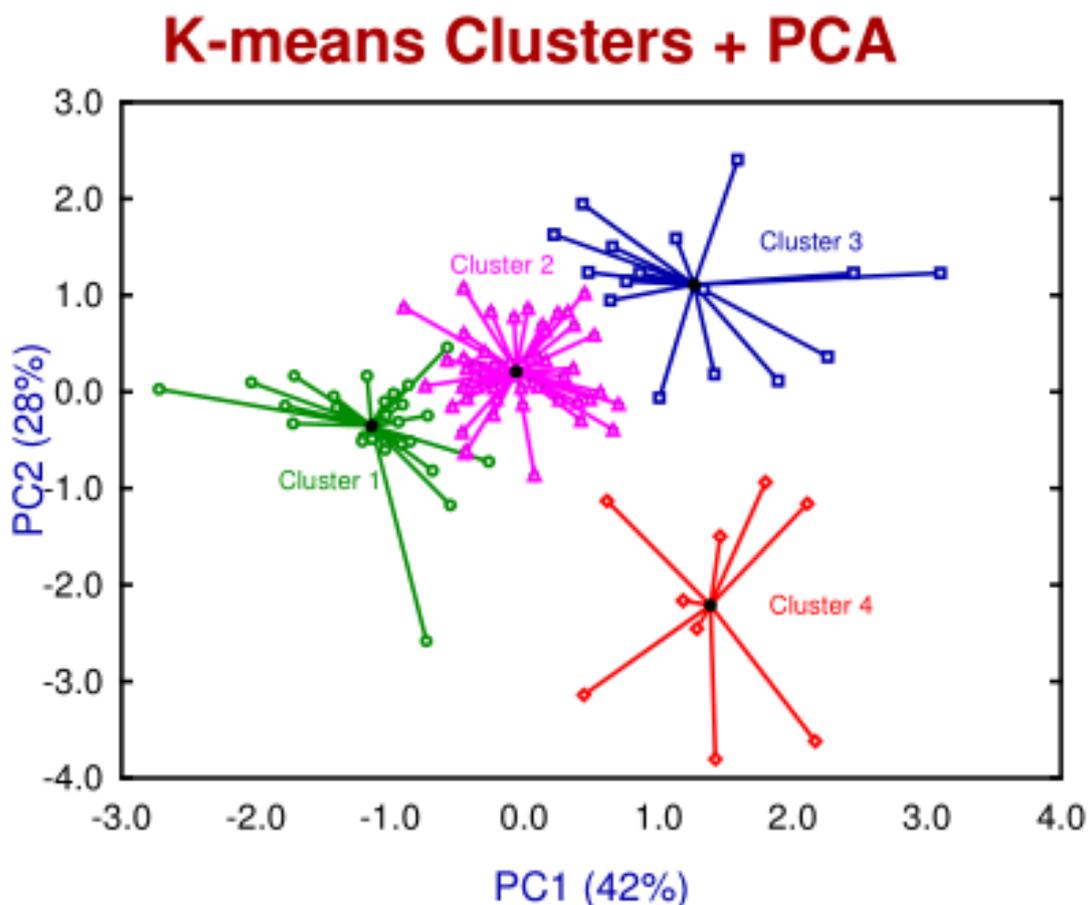
**Tabla 4.2.** Valores iniciales de las medias esperadas para las variables en los cuatro grupos de enfermedad, los cuales se asignan como estimas a los centroides de dichos grupos, a fin de que el algoritmo comience las iteraciones y forme los 4 *clusters* finales. Granulomas (Granul), superficie (Superf), vermes macho (V mac), vermes hembra (V hem), huevos por gramo de hígado (HPG\_h), huevos por gramo de intestino (HPG\_i), fecundidad de las hembras en el hígado (HPG\_h/H), fecundidad de las hembras en el intestino (HPG\_i/H).

Grupo	Granul	Superf	V mac	V hem	HPG_h	HPG_i	HPG_h/H	HPG_i/H
1	41	12.4	4	4	1585	633	396	158
2	115	46.1	9	9	8724	5150	969	572
3	183	54.8	19	16	11601	8719	725	544
4	245	92.6	17	16	23067	14706	1441	919

Debido a la diferencia de magnitud entre las variables, se hizo una estandarización previa de todas las variables y de las estimas iniciales de los centroides a media cero y desviación estándar uno.

Como se expuso en Materiales y Métodos, no es fácil representar los casos junto a los *clusters* y los centroides finales, ya que dichas representaciones en 2D dependerían de las dos variables elegidas de entre las muchas que entraron en el algoritmo, lo cual podría resultar ambiguo. En su lugar es mejor representar, en vez de las variables originales, las dos primeras componentes principales obtenidas mediante un análisis de los datos en componentes principales. Es decir, internamente el algoritmo calcula las puntuaciones de los casos y centroides en PC1 y PC2, para representar finalmente PC1 frente a PC2 y una

los casos a los centroides de cada grupo mediante radios, lo que facilita su interpretación. Esta gráfica se exhibe en la Figura 4.10, en la cual se observa la asignación de los ratones a los 4 *clusters* considerados. Los ratones parecen agruparse en cada *cluster* en función de la gravedad de la infección, del menos grave (*cluster* 1) al más grave (*cluster* 4).



**Figura 4.10.** Análisis de conglomerados por k-medias y PCA. Se observan los cuatro grupos calculados por el algoritmo tras varias iteraciones a partir de los centroides iniciales. Todos los valores se refieren a las dos primeras componentes principales calculadas internamente por el algoritmo mediante un análisis PCA. La variabilidad capturada por el procedimiento PCA fue en total del 70%.

Los valores finales de todas las variables, encontrados por el algoritmo para los cuatro centroides de los grupos, aparecen recogidos en la Tabla 4.3. Como es lógico, estos valores finales son diferentes de las estimas iniciales que aparecen en la Tabla 4.2, ya que el algoritmo ha optimizado, dentro de cada grupo, la suma de las distancias al cuadrado de cada caso a las medias de los centroides de los grupos.

**Tabla 4.3.** Valores de las medias de las variables en los centroides finales de los cuatro grupos, una vez formados dichos grupos por el algoritmo de k-medias, así como el número de ratones asignado a cada grupo (N). Granulomas (Granul), superficie (Superf), vermes macho (V mac), vermes hembra (V hem), huevos por gramo de hígado (HPG\_h), huevos por gramo de intestino (HPG\_i), fecundidad de las hembras en el hígado (HPG\_h/H), fecundidad de las hembras en el intestino (HPG\_i/H), Número de ratones asignados a cada grupo (N).

Grupo	Granul	Superf	V mac	V hem	HPG_h	HPG_i	HPG_h/H	HPG_i/H	N
1	108	34.7	5	4	3901	2383	1021	542	26
2	147	57.3	12	9	8088	5530	911	627	54
3	153	56.9	20	17	16472	13907	974	818	16
4	163	64.9	6	6	17652	15105	3565	2569	9

Al observar en la Tabla 4.3 los valores de las variables en los centroides, se aprecia que el Grupo 4, a priori el que debiera presentar los valores más altos para todas las variables relativas a la infección, presenta un número medio de vermes, tanto machos como hembras, menor que las medias de los Grupos 2 y 3, aunque sí presenta un mayor número de granulomas, así como de superficie afectada del hígado, número de huevos por gramo de hígado e intestino y fecundidad de las hembras en el hígado e intestino.

### 4.3 Variación de las variables fenotípicas entre los cuatro grupos de infección

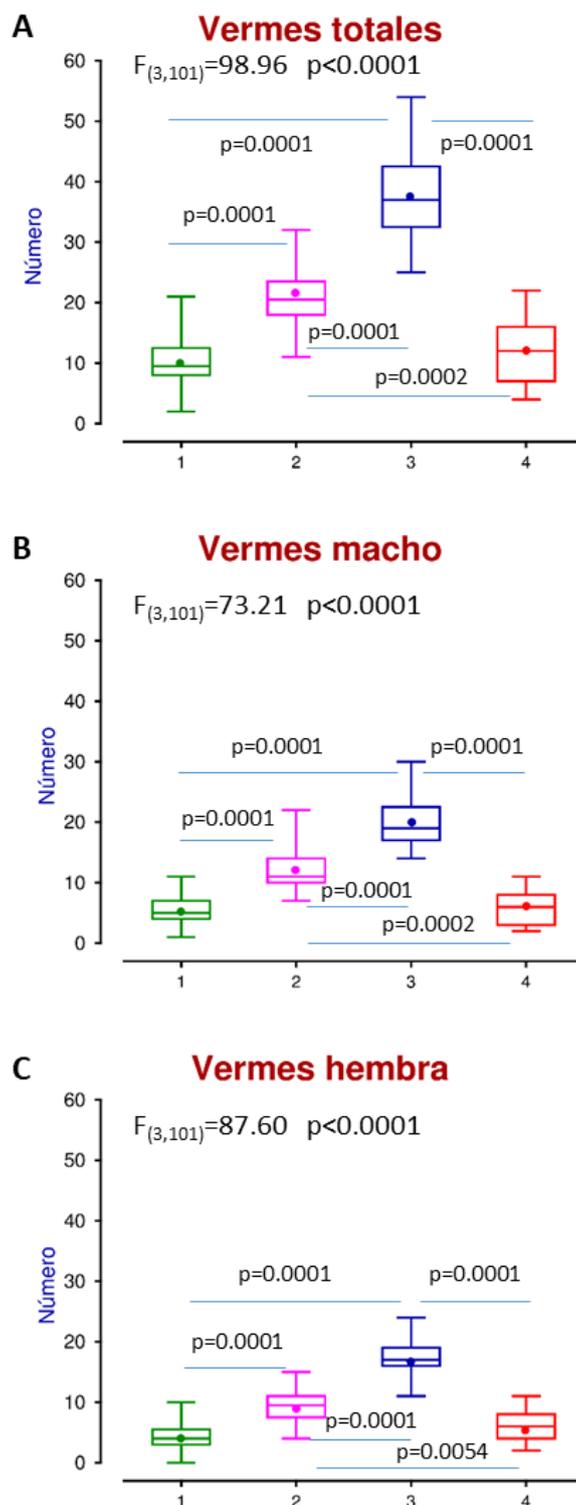
Como se ha descrito en los apartados anteriores, los estudios multivariantes apuntan en la dirección de que los ratones infectados se pueden separar en cuatro grupos, que corresponderían a una gravedad creciente de la enfermedad. A la vista de esto se procedió a comparar los valores que toman las variables en los distintos grupos. Para ello se analizarán las siguientes variables fenotípicas que caracterizan la patología de la enfermedad y que son, por un lado, las variables parasitológicas: número de vermes totales, número de vermes macho, número de vermes hembra y número de huevos por gramo de hígado e intestino y por otro lado las variables patológicas: número de granulomas y la superficie afectada del hígado. Así mismo, se analizarán las variables que midieron las poblaciones celulares en sangre periférica y en esplenocitos. Para finalizar, se llevó a cabo una comparación entre las variables que midieron la respuesta inmunológica.

El estudio de la variación de cada variable entre los cuatro grupos se realizó mediante el procedimiento estadístico ANOVA, con el fin de analizar si existen diferencias estadísticamente significativas entre los valores medios de cada variable en los cuatro grupos. Los resultados de los distintos ANOVA se representarán mediante diagramas de cajas, incluyendo en la gráfica el valor del estadístico F y el p-valor correspondiente, así como los p-valores de las comparaciones dos a dos a posteriori según el test de Tukey. En todos los casos se considera que existen diferencias estadísticamente significativas cuando el p-valor correspondiente sea menor de 0.05 ( $p < 0.05$ ).

#### **4.3.1 Variación entre grupos para el número de vermes totales, de vermes macho y de vermes hembra**

Los resultados de la comparación entre grupos para estas variables aparecen representados en la Figura 4.11. Como puede apreciarse en el panel A de dicha figura, hay diferencias significativas para el número de vermes totales entre el Grupo 1 y el 2, entre el Grupo 1 y el 3, entre el Grupo 2 y el 3, entre el Grupo 2 y el 4 y entre el 3 y el 4. La única comparación que no resultó significativa fue la que se refiere a los Grupos 1 y 4, que son el menos grave y el más grave respectivamente, pues presentan unos valores similares.

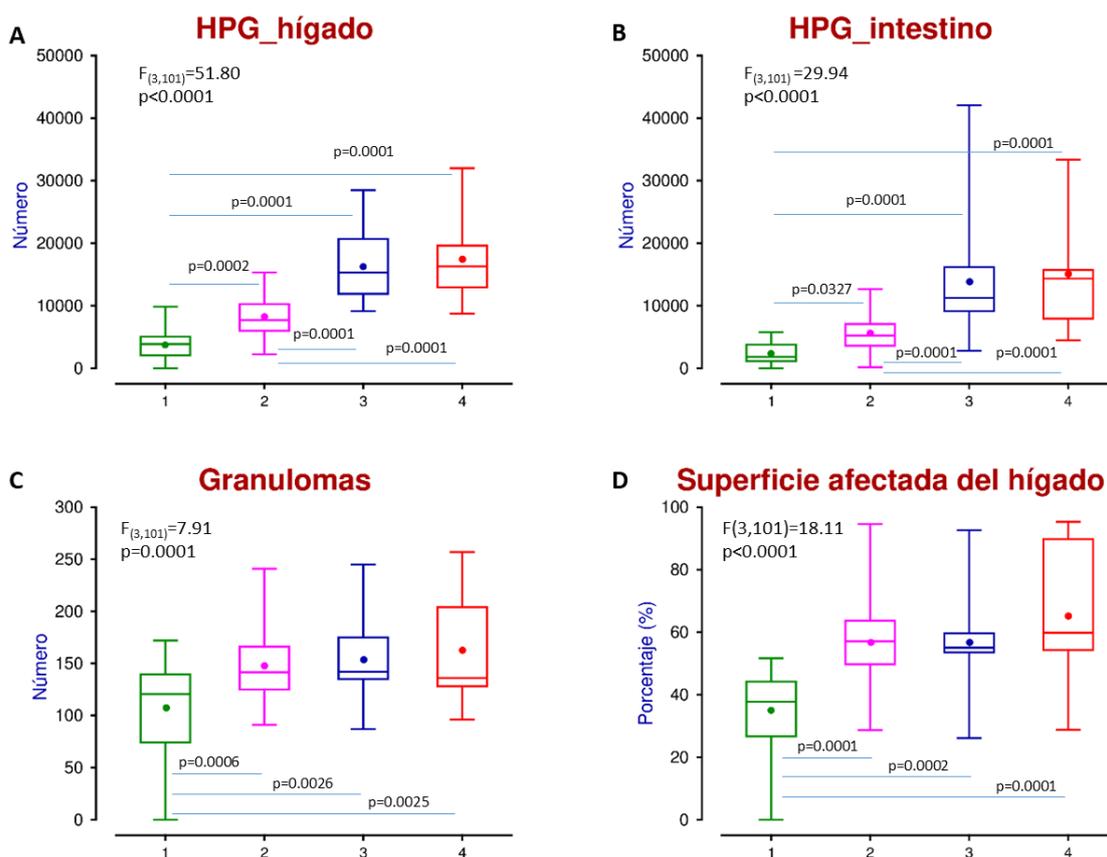
En las Figuras 4.11 B, relativa al número de vermes macho y la 4.11 C, correspondiente al número de vermes hembra, se observa un comportamiento totalmente análogo al descrito anteriormente para el número de vermes total. Así mismo, los test de Tukey entre pares de grupos, mostraron el mismo patrón de significancia que el ya expuesto para los vermes totales.



**Figura 4.11.** Comparación entre grupos de los siguientes fenotipos: (A) número de vermes totales, (B) vermes macho y (C) vermes hembra. Se representa la mediana de cada grupo como una línea en el interior de la caja y la media como un círculo, siendo los segmentos de la caja los respectivos valores extremos superior e inferior. La comparación conjunta entre las medias se realizó mediante el procedimiento ANOVA. Se muestra el valor del estadístico  $F$  del ANOVA y su  $p$ -valor, así como los  $p$ -valores de las comparaciones dos a dos a posteriori mediante el test de Tukey.

### 4.3.2 Variación entre grupos para el número de huevos por gramo de hígado e intestino, el número de granulomas y la superficie afectada del hígado

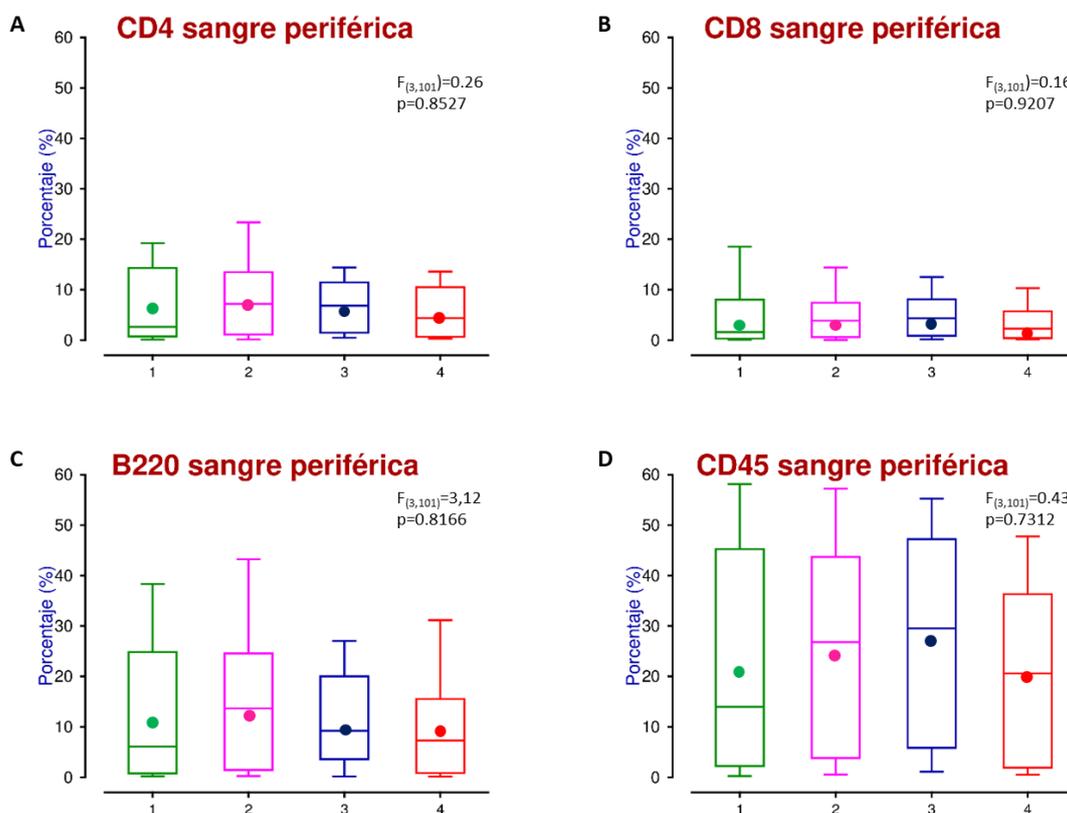
Los resultados obtenidos se muestran en la Figura 4.12. En lo que respecta al número de huevos por gramo de hígado e intestino (paneles 4.12 A y B), se detectaron diferencias significativas entre todos los grupos dos a dos, excepto entre el Grupo 3 y el Grupo 4, los dos más graves en cuanto a la infección, puesto que presentan unos valores de las medias bastante similares para estas variables entre ellos. En cuanto al número de granulomas y la superficie afectada del hígado (paneles 4.12 C y D) se observaron diferencias significativas entre el Grupo 1 y el 2, entre el 1 y el 3 y entre el 1 y el 4, pero no hay diferencia entre el Grupo 2 y el 3, entre el 2 y el 4, ni entre el 3 y el 4. Esto es, presenta diferencias entre el grupo menos grave y los otros tres, pero no hay diferencias entre los grupos en los que se presenta la enfermedad más grave/avanzada (2, 3 y 4).



**Figura 4.12.** Comparación entre grupos de los siguientes fenotipos: (A) número de huevos por gramo de hígado (HPG\_hígado), (B) número de huevos por gramo de intestino (HP\_ intestino), (C) número de granulomas y (D) superficie afectada del hígado. Se representa la mediana de cada grupo como una línea en el interior de la caja y la media como un círculo, siendo los segmentos de la caja los respectivos valores extremos superior e inferior. La comparación conjunta entre medias se realizó mediante el procedimiento ANOVA. Se muestra el valor del estadístico F del ANOVA y su p-valor, así como los p-valor de las comparaciones dos a dos a posteriori mediante el test de Tukey.

### 4.3.3 Variación entre grupos para poblaciones celulares en sangre periférica.

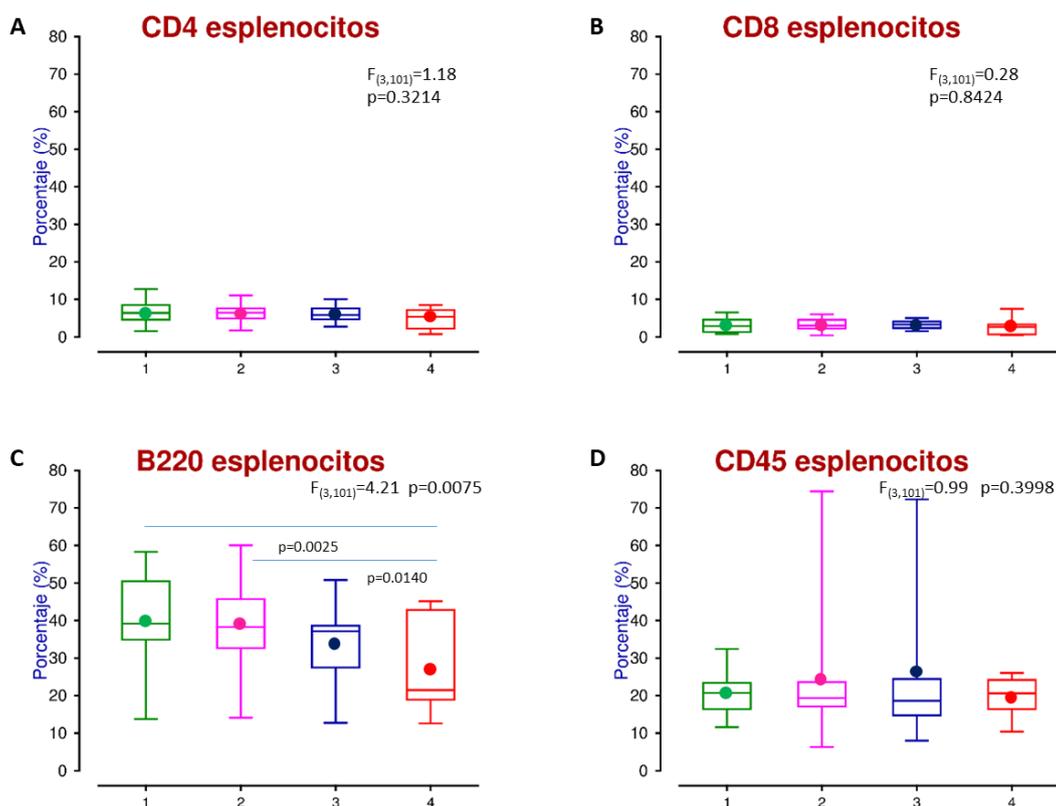
Para las variables de poblaciones celulares (CD4, CD8, B220 y CD45) en sangre periférica realizamos también el procedimiento ANOVA, calculando el estadístico F y sus p-valores. Así mismo, se muestran los p-valores de las comparaciones dos a dos a posteriori mediante el test de Tukey. En la Figura 4.13 se observan los resultados del método ANOVA para las poblaciones celulares en sangre periférica. En los cuatro paneles se observa que el estadístico F presenta unos valores muy bajos y en todos los casos el p-valor es superior a 0.05 ( $p > 0.05$ ), con lo que no se observan diferencias significativas entre grupos para ninguna de las poblaciones celulares.



**Figura 4.13.** Variación entre grupos de las poblaciones celulares en sangre periférica: (A) CD4, (B) CD8, (C) B220 y (D) CD45. Se representa la mediana de cada grupo como una línea en el interior de la caja y la media como un círculo, siendo los segmentos de la caja los respectivos valores extremos superior e inferior. La comparación conjunta entre medias se realizó mediante el procedimiento ANOVA. Se muestra el valor del estadístico F del ANOVA y su p-valor.

#### 4.3.4 Variación entre grupos para poblaciones de esplenocitos

Se midieron las mismas poblaciones celulares en esplenocitos (CD4, CD8, B220 y CD45) y se aplicó también el procedimiento ANOVA, calculándose el estadístico F y sus p-valores. Así mismo, se calcularon los p-valores de las comparaciones dos a dos a posteriori mediante el test de Tukey. En la Figura 4.14 se representan los resultados para las poblaciones de esplenocitos. No se observan diferencias significativas entre los cuatro grupos estudiados para las poblaciones celulares CD4, CD8 y CD45 en esplenocitos. Sin embargo, sí existen diferencias estadísticamente significativas ( $p < 0.05$ ) entre las medias de los grupos para las poblaciones de células B220. Así mismo, para estas células B220, se observaron diferencias significativas al calcular las comparaciones dos a dos a posteriori mediante el test de Tukey, específicamente se observan diferencias entre el Grupo 1 y el 4 y entre el Grupo 2 y el 4.

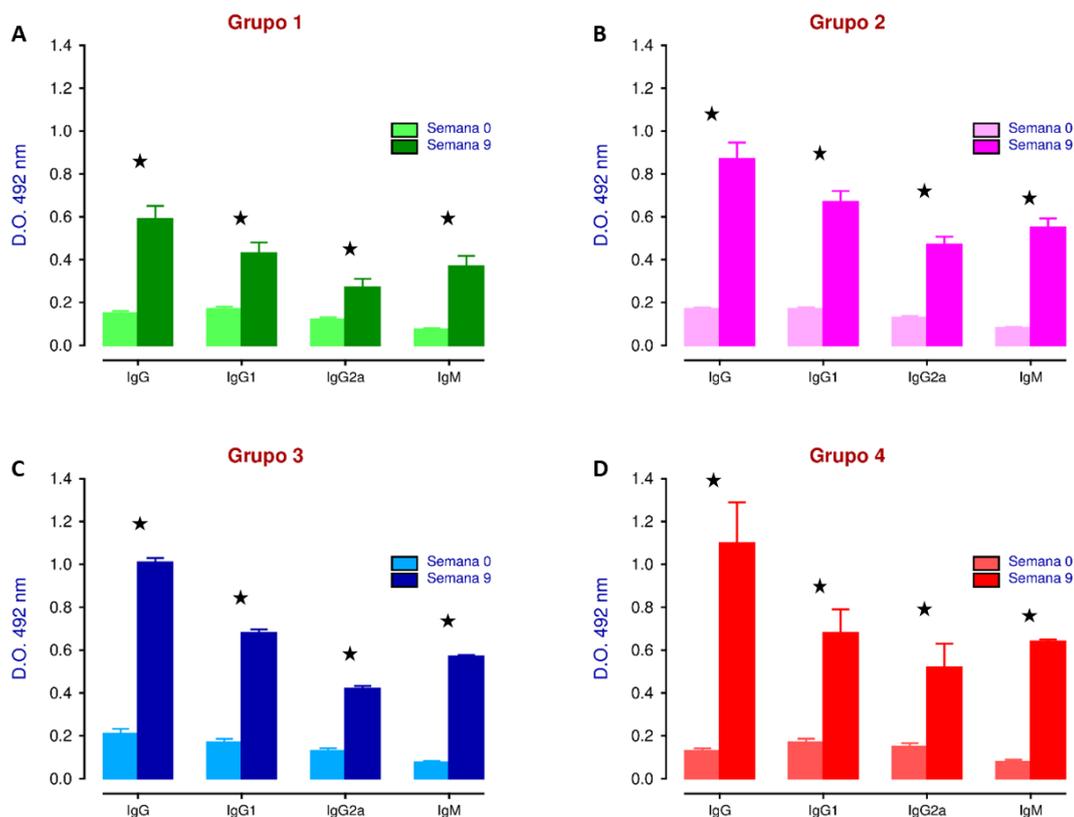


**Figura 4.14.** Variación entre grupos para las poblaciones celulares en esplenocitos. (A) CD4, (B) CD8, (C) B220 y (D) CD45. Se representa la mediana de cada grupo como una línea en el interior de la caja y la media como un círculo, siendo los segmentos de la caja los respectivos valores extremos superior e inferior. La comparación conjunta entre medias se realizó mediante el procedimiento ANOVA. Se muestra el valor del estadístico F del ANOVA y su p-valor, así como los p-valores de las comparaciones dos a dos a posteriori mediante el test de Tukey cuando son significativos.

#### 4.3.5 Variación entre grupos para inmunoglobulinas en las semanas cero y nueve

Se realizó un estudio midiendo las inmunoglobulinas específicas anti-*Schistosoma mansoni* (IgG, IgG1, IgG2a e IgM) medidas en cada uno de los grupos mediante una comparación de la densidad óptica observada con el método ELISA entre las semanas cero y nueve. Estos resultados se muestran en la Figura 4.15. Para determinar si existen diferencias entre las medias de las inmunoglobulinas se utilizó el test t de Student para datos apareados y se aprecia que, para cada uno de los cuatro grupos de infección, hay diferencias altamente significativas ( $p < 0.01$ ) entre todos los grupos para las inmunoglobulinas medidas en la semana cero y la semana nueve.

Observamos que en la semana cero para todas las inmunoglobulinas y para todos los grupos, no hay diferencias entre las medias de las densidades ópticas para cada inmunoglobulina. Sin embargo, la respuesta de IgG en la semana nueve para el Grupo 1, es menor que para los otros tres grupos. Así, se aprecia que la densidad óptica va aumentando en el Grupo 2 respecto al 1, que en el 3 es mayor que en el Grupo 2 y que en el Grupo 4 a su vez es superior a los otros tres grupos. En cuanto a la media de la densidad óptica medida para IgG1, IgG2a e IgM en la semana nueve, los valores más bajos los presenta el Grupo 1 (menos grave), pero la densidad óptica en los Grupos 2, 3 y 4 es muy similar entre ellos.



**Figura 4.15.** Diferencias entre los niveles de inmunoglobulinas anti-*S. mansoni* medidas en cada uno de los grupos entre la semana cero (previa a la infección) y la semana nueve post-infección. Los datos se presentan como la media de grupos  $\pm$  el error estándar de la media. El test empleado ha sido el test t de Student para datos apareados. En todos los casos las diferencias resultaron ser altamente significativas  $p < 0.01$ .

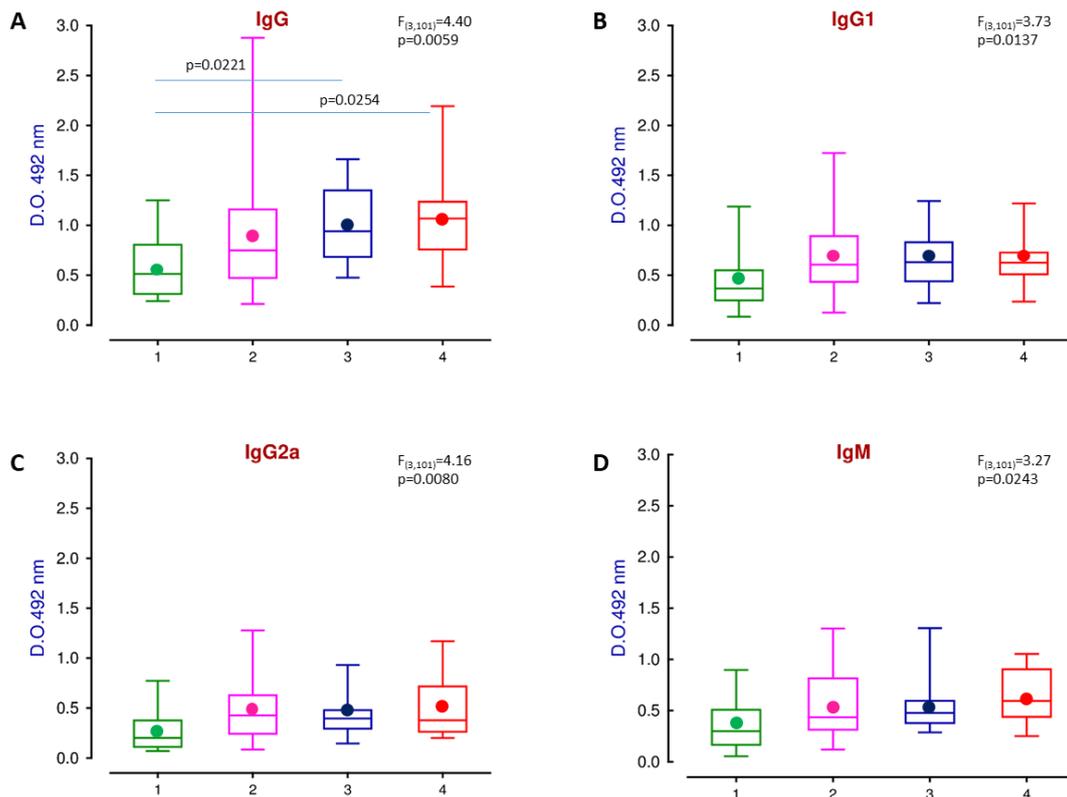
#### 4.3.6 Variación entre grupos para inmunoglobulinas en la semana nueve.

Finalmente, en cuanto a las inmunoglobulinas medidas en la semana nueve, empleamos de nuevo el procedimiento ANOVA, calculando el estadístico F y sus p-valores y el test de Tukey con sus p-valores.

En la Figura 4.16 se representan en los distintos paneles los valores de cada una de las inmunoglobulinas anti-*S. mansoni* en los cuatro grupos. Así, en el panel A, donde está representada la densidad óptica del isotipo IgG en la semana nueve, observamos que muestra diferencias estadísticamente significativas entre los grupos según ANOVA, además en las comparaciones dos a dos a posteriori se aprecian diferencias entre el Grupo 1 y el 3 y entre el Grupo 1 y el 4. En el panel B está representada la IgG1, en el panel C la IgG2a y en el panel D la IgM. En todos ellos hay diferencias estadísticamente significativas

en el procedimiento ANOVA, pero no presenta diferencias entre grupos en el test de Tukey en las comparaciones dos a dos a posteriori.

Normalmente cuando un ANOVA resulta significativo, ocurre que al menos dos de las medias incluidas en el análisis suelen ser también diferentes de forma significativa. En este caso particular, el que no se detecten por Tukey diferencias entre pares para un ANOVA significativo, se debe a que los grupos presentan tamaños de muestra muy diferentes. Además, la sensibilidad del ANOVA es mayor que la sensibilidad del test de Tukey, es decir un ANOVA detecta una menor variabilidad entre las medias que el test de Tukey, dando lugar a excepciones como la observada en la variación entre grupos para inmunoglobulinas en la semana nueve.



**Figura 4.16.** Variación entre grupos para las inmunoglobulinas en la semana nueve post-infección: (A) IgG, (B) IgG1, (C) IgG2a y (D) IgM. Se representa la mediana de cada grupo como una línea en el interior de la caja y la media como un círculo, siendo los segmentos de la caja los respectivos valores extremos superior e inferior. La comparación conjunta entre medias se realizó mediante el procedimiento ANOVA. Se muestra el valor del estadístico F del ANOVA y su p-valor, así como los p-valores de las comparaciones dos a dos a posteriori mediante el test de Tukey cuando son significativos.

#### 4.3.7 Análisis de correlación de las variables en los cuatro grupos de la cohorte F2BX.

Se realizó un estudio de correlaciones entre las distintas variables dentro de cada uno de los grupos para la infección. El objetivo era detectar la posible asociación de dos variables entre sí. Los resultados obtenidos se pueden ver en la Figura 4.17. Dentro de cada grupo se calculó el coeficiente de correlación de Pearson ( $r$ ) entre todas las variables, del mismo modo como se explicó en el Apartado 4.1.4. En dicha figura, se observa la representación de cuatro mapas que muestran los resultados de las correlaciones en cada grupo. Las variables con una correlación estadísticamente significativa ( $p < 0.05$ ) y con asociación directa entre ellas ( $r > 0$ ) se han representado en color rojo, mientras que las que presentan correlación con asociación inversa ( $r < 0$ ) figuran en azul. En color blanco figuran las correlaciones que no son significativas ( $p > 0.05$ ). A continuación, se analizan las correlaciones más importantes que se observan en los diferentes grupos.

En el Grupo 1, con la afección más leve de la cohorte, el número de granulomas está relacionado directamente con la superficie afectada del hígado ( $\text{mm}^2/\text{cm}^2$ ) y con todas las poblaciones celulares estudiadas, excepto con los CD45 en esplenocitos (CD45\_Esp\_S9). La superficie afectada del hígado ( $\text{mm}^2/\text{cm}^2$ ) está relacionada directamente con el número de granulomas, el número de huevos por gramo de hígado (HPG\_hig) y la fecundidad en el intestino (HPG\_int/H). La superficie afectada del hígado también presenta asociación directa con las poblaciones celulares, excepto con los CD45 en esplenocitos (CD45\_Esp\_S9). Por su parte, el número de vermes total, de vermes macho y de vermes hembra, están relacionados entre ellos y con los huevos por gramo de hígado (HPG\_hig) e intestino (HPG\_int). Los vermes hembra a su vez, están directamente relacionados con la IgM y están inversamente relacionados con los CD8 en esplenocitos. El número de huevos por gramo de hígado (HPG\_hig) e intestino (HPG\_int) y la fecundidad en hígado (HPG\_hig/H) e intestino (HPG\_int/H) están relacionados entre ellos, excepto el número de huevos por gramo de intestino (HPG\_int) y la fecundidad en el hígado (HPG\_hig/H), los cuales no presentan una correlación significativa. El número de huevos por gramo de intestino está relacionado directamente con la IgM e indirectamente con los B220 en esplenocitos (B220\_Esp\_S9). La fecundidad en el hígado (HPG\_hig/H) e intestino (HPG\_int/H) están relacionadas directamente con los CD45 en esplenocitos (CD45\_Esp\_S9).

En el Grupo 2, el número de granulomas está relacionado positivamente con la superficie afectada del hígado ( $\text{mm}^2/\text{cm}^2$ ), así como con las poblaciones celulares en sangre periférica. El número de vermes total, de vermes macho y de vermes hembra están relacionados entre ellos; además, los vermes totales y los vermes hembra están relacionados directamente con el número de huevos por gramo de hígado (HPG\_hig) e intestino (HPG\_int) e inversamente relacionados con la fecundidad en el hígado (HPG\_hig/H). Por su parte, los huevos por gramo de hígado e intestino presentan relación directa con la fecundidad en el hígado (HPG\_hig/H) e intestino (HPG\_int/H), salvo que no existe significación estadística entre el número de huevos por gramo de intestino y la fecundidad en el hígado (HPG\_hig/H). El número de huevos por gramo de hígado presentan a su vez una correlación inversa con las poblaciones celulares en sangre periférica y en esplenocitos, excepto con los B220 en sangre periférica (B220\_Sg\_S9), en cuyo caso no existe una correlación significativa. Por su parte, el número de huevos por gramo de intestino (HPG\_int) presentan correlación directa con las inmunoglobulinas IgG1 e IgG2a. La fecundidad en el hígado (HPG\_hig/H) presenta relación inversa con las poblaciones celulares, pero en este caso no hay asociación significativa con los B220 en sangre periférica, ni con los CD4 ni CD45 en esplenocitos. En cuanto a la fecundidad en el intestino (HPG\_int/H) sólo presenta una relación inversa con B220 en sangre periférica.

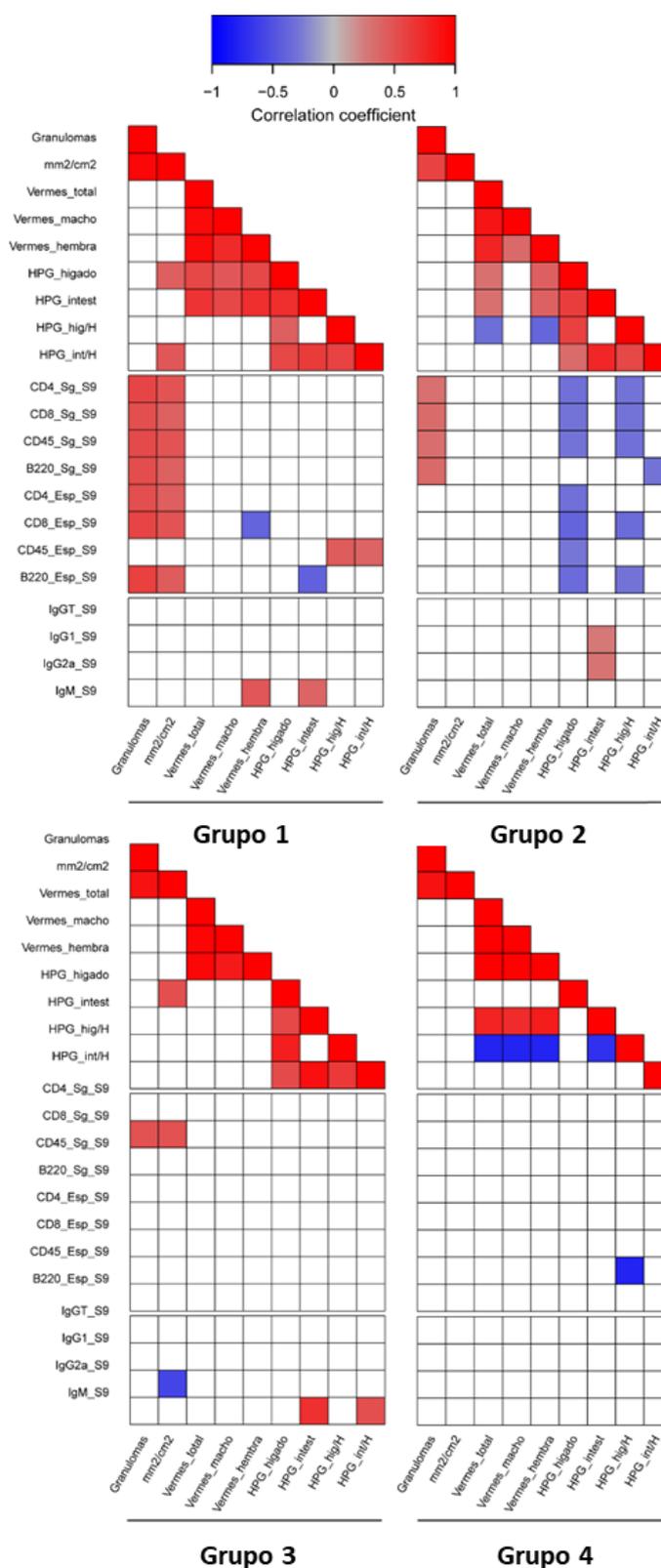
El Grupo 3 presenta una correlación directa entre número de granulomas y superficie afectada del hígado ( $\text{mm}^2/\text{cm}^2$ ). El número de granulomas también está relacionado directamente con los CD8 en sangre periférica (CD8\_Sg\_S9). Respecto a la superficie afectada del hígado ( $\text{mm}^2/\text{cm}^2$ ), ésta presenta una relación directa con el número de huevos por gramo de hígado (HPG\_hig); a la vez que presenta una asociación directa con los CD8 en sangre periférica (CD8\_Sg\_S9), pero inversa con la IgG2a. El número de vermes total, de vermes macho y de vermes hembra están relacionados directamente entre ellos. Por otra parte, el número de huevos por gramo de hígado (HPG\_hig) e intestino (HPG\_int) están directamente relacionados con la fecundidad en el hígado (HPG\_hig/H) e intestino (HPG\_int/H), a excepción del número de huevos por gramo de intestino con la fecundidad en el hígado, que no presentan correlación. Finalmente, la IgM mostró una relación directa significativa con el número de huevos por gramo de intestino y la fecundidad en el intestino.

En el Grupo 4, se observó una relación directa entre el número de granulomas y la superficie afectada del hígado ( $\text{mm}^2/\text{cm}^2$ ). Por su parte, el número de vermes total, de

vermes macho y de vermes hembra presentan relación directa entre ellos y a su vez con el número de huevos por gramo de intestino (HPG\_int), pero la asociación es inversa con la fecundidad en el hígado (HPG\_hig/H). Así mismo, se detectó una relación inversa entre el número de huevos por gramo de intestino (HPG\_int) y la fecundidad en el hígado (HPG\_hig/H). Finalmente, se observó que la fecundidad en el hígado (HPG\_hig/H) muestra asociación inversa con la población celular de los CD45 en esplenocitos (CD45\_Esp\_S9). En el Grupo 4, salvo esta única asociación inversa entre CD45 y la fecundidad en el hígado, no hay correlaciones significativas con las variables de poblaciones celulares ni con las de inmunoglobulinas.

En conjunto, se puede concluir que, en todos los grupos hay asociación directa entre el número de granulomas y la superficie afectada del hígado ( $\text{mm}^2/\text{cm}^2$ ). También se observó una relación directa entre el número de vermes total, de vermes macho y de vermes hembra en todos los grupos. La asociación entre el número de huevos tanto en hígado como en intestino y la fecundidad de las hembras, varía entre los cuatro grupos, pudiendo ser directa, caso del Grupo 1 y 3 o inversa en algunos casos, como en los Grupos 2 y 4. Para las poblaciones celulares, pocas variables presentan correlación estadística significativa, ya sea directa o inversa y van disminuyendo en número desde el Grupo 1 al Grupo 4. Finalmente, en cuanto a las inmunoglobulinas, presentan pocas correlaciones en todos los grupos o incluso ninguna, como el caso del Grupo 4, esto podría deberse a que el tamaño de esta muestra es pequeño ( $n=9$ ).

Con el fin de analizar aún más las características de los cuatro grupos de infección, se estudió como variaba cada variable a lo largo de los cuatro grupos. Para ello se tomó la mediana máxima de cada variable dentro de los cuatro grupos como referencia. A continuación, se calculó la proporción en porcentaje que la mediana de cada grupo tenía respecto de la mediana máxima, lo cual es indicativo de la variación que ha experimentado la variable a lo largo de los grupos. Es decir, el valor máximo de la mediana de cada variable dentro de cada grupo representaría el 100% y los distintos porcentajes de la variable en cada grupo nos indicarán la distancia respecto al máximo. Cuanto menor sea la diferencia entre las medianas, el color representado en la gráfica será más rojo. Por el contrario, cuanto mayor sea la diferencia entre el valor máximo y los porcentajes de cada uno de los otros grupos, el color representado será más azul.

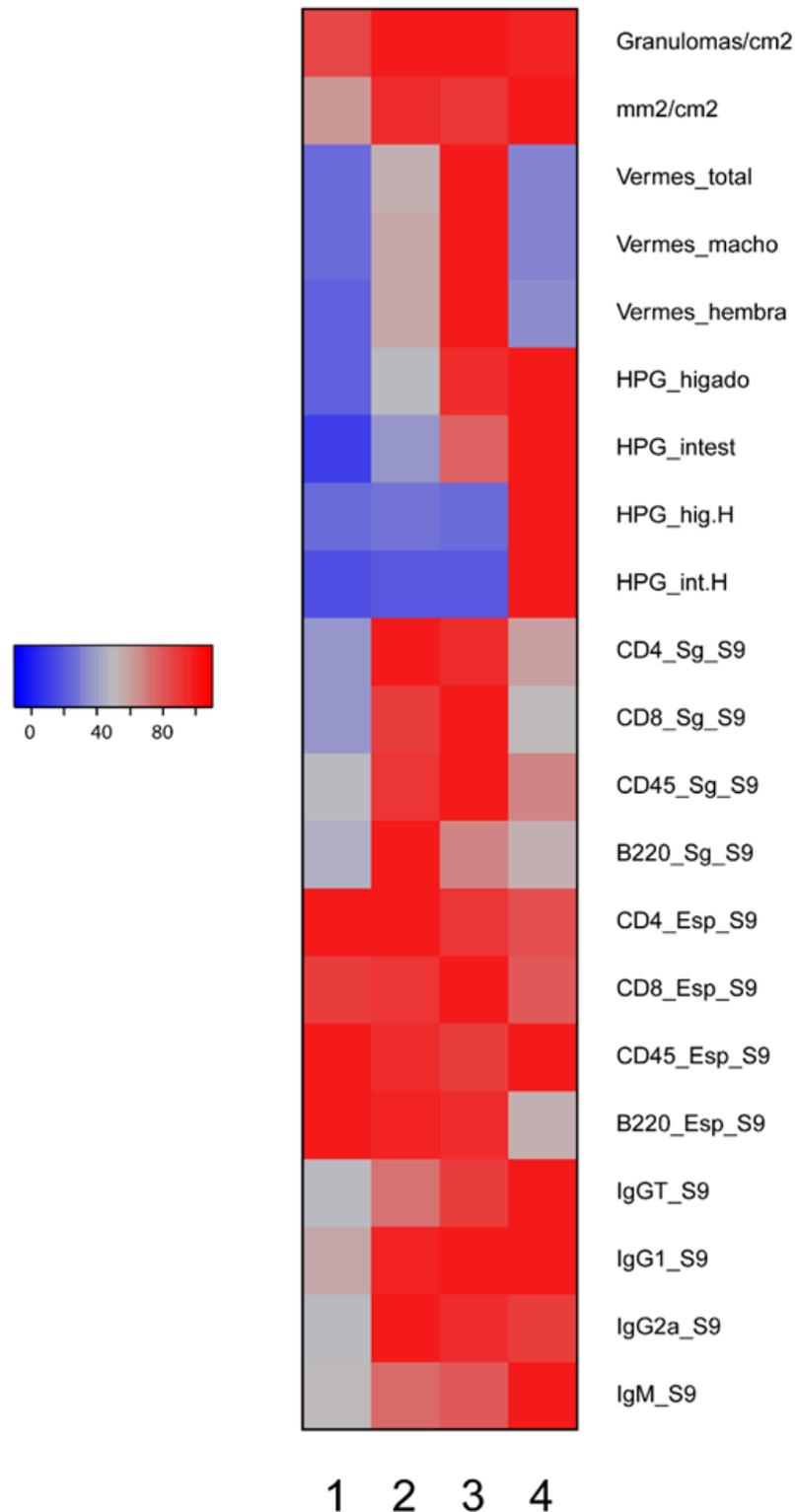


**Figura 4.17.** Análisis de correlaciones entre variables para los 4 grupos de ratones. Se presentan las variables que tienen una correlación estadísticamente significativa ( $p < 0.05$ ) en función de si esa correlación es directa ( $r > 0$ ), en cuyo caso aparecen representadas en color rojo o inversa ( $r < 0$ ) en cuyo caso figuran en color azul. Las variables que no presentan correlación estadísticamente significativa ( $p > 0.05$ ) se han representado en color blanco.

Estos resultados se muestran en forma de mapa en la Figura 4.18. En cuanto a las variables parásito-patológicas, el Grupo 4 tiene los valores de referencia (medianas máximas) en la mayoría de las variables, excepto en las de número de vermes, tanto totales, como machos y hembras, donde el Grupo 3 es el que presenta los valores máximos de mediana en estas tres variables; el Grupo 2 tiene valores intermedios (50-60%) y el Grupo 1 y 4 presentan valores entre el 20 y el 35% del valor máximo. Esto concuerda con lo visto anteriormente para la formación de los grupos por conglomerados de k-medias + PCA. Para la variable número de granulomas, los 4 grupos están entre el 85 y el 100% del valor máximo. Para la superficie afectada del hígado ( $\text{mm}^2/\text{cm}^2$ ), no hay mucha diferencia entre las medianas de los Grupos 2, 3 y 4, las cuales están entre el 92 y el 100%, a diferencia del Grupo 1, que presenta valores más bajos (63%), como era de esperar por ser el grupo de menor gravedad. Sin embargo, en las variables número de huevos por gramo de hígado (HPG\_hig) e intestino (HPG\_int), sí hay diferencia para las medianas de los Grupos 1 y 2, cuyos porcentajes están entre el 10 y el 50% de su valor máximo, mientras que las medianas de los Grupos 3 y 4 presentan valores entre el 80 y el 100%. Para la fecundidad de las hembras en el hígado (HPG\_hig/H) e intestino (HPG\_int/H) hay muchas diferencias entre las medianas del Grupo 4 y las de los Grupos 1, 2 y 3, que presentan valores entre el 15 y el 30% de la mediana del Grupo 4.

Para las variables que miden las poblaciones celulares en sangre periférica, los valores máximos de las medianas están en los Grupos 2 y 3, los cuales muestran porcentajes del 70 al 100%, el Grupo 4 presenta porcentajes de entre el 50 y el 70% del máximo y el Grupo 1 muestra porcentajes de entre el 35 y el 50%. Para las poblaciones celulares en esplenocitos, no hay ninguna mediana máxima en el Grupo 4, pero los valores de las medianas se encuentran para todos los grupos en valores muy semejantes entre el 80 y el 100%, excepto para las células B220 del Grupo 4 (55%).

Para las variables que miden la respuesta inmunológica, las medianas máximas las presenta el Grupo 4 para la IgG y la IgM; mientras que el Grupo 3 presenta el máximo para la IgG1. Por su parte, el Grupo 2 tiene la mediana máxima para la IgG2a. Hay diferencias del 50% entre los porcentajes de las medianas del Grupo 1 y las del Grupo 4, mientras que las medianas de los Grupos 2, 3 y 4 presentan menos diferencias entre ellas, pues sus porcentajes están entre el 70 y el 100%.



**Figura 4.18.** Análisis de correlaciones de las medianas de las variables entre los cuatro grupos de ratones. Se toma como valor de referencia la mediana máxima de cada variable dentro de los cuatro grupos. Se representa la proporción, en porcentaje, que la mediana de cada grupo tiene respecto a la máxima, lo cual indica la variación experimentada a lo largo de los grupos. El color representado cuando la diferencia entre medianas es baja, es más rojo. Por el contrario, cuando la diferencia entre medianas es mayor, el color representado es más azul.

## 4.4 Identificación de regiones cromosómicas asociadas a susceptibilidad y resistencia

### 4.4.1 Genotipado de la cohorte F2BX y control de calidad de los datos

Se llevó a cabo el genotipado de la cohorte de ratones generada por *backcross* F2BX, para realizar el análisis de ligamiento genético e identificar las regiones del genoma asociadas a los fenotipos estudiados. Se enviaron 92 muestras de DNA extraído de la cola de los ratones al Centro Nacional de Genotipado (CeGen) y se empleó una plataforma con 961 SNPs (polimorfismos de un sólo nucleótido).

Previamente al análisis de ligamiento, realizamos un control de calidad de los datos remitidos por el CeGen, para lo que empleamos el paquete *R/qtl* en la versión de R3.3.3 y la interfaz RStudio. Para ello, preparamos una matriz con los datos fenotípicos y los datos del genotipado de los ratones. Los datos del fenotipo son los de nuestras variables y no hay que prepararlos de ninguna manera (son cuantitativos). Los datos del genotipo hay que convertirlos a un código binario. En nuestro caso, los homocigotos se convierten a 0 y los heterocigotos se convierten a 1. En el cromosoma X, para los F2BX hembras seguimos el mismo código de conversión; pero para los F2BX machos, si es homocigoto para la cepa hacia la que retrocruza lo convertimos en 0 y si es heterocigoto para la otra cepa lo convertimos en 2. Cuando no hay resultados, se tipifica como NA. Una vez que tenemos hecho esto, cruzamos los datos del fenotipo con el genotipo y procedemos a realizar el control de calidad.

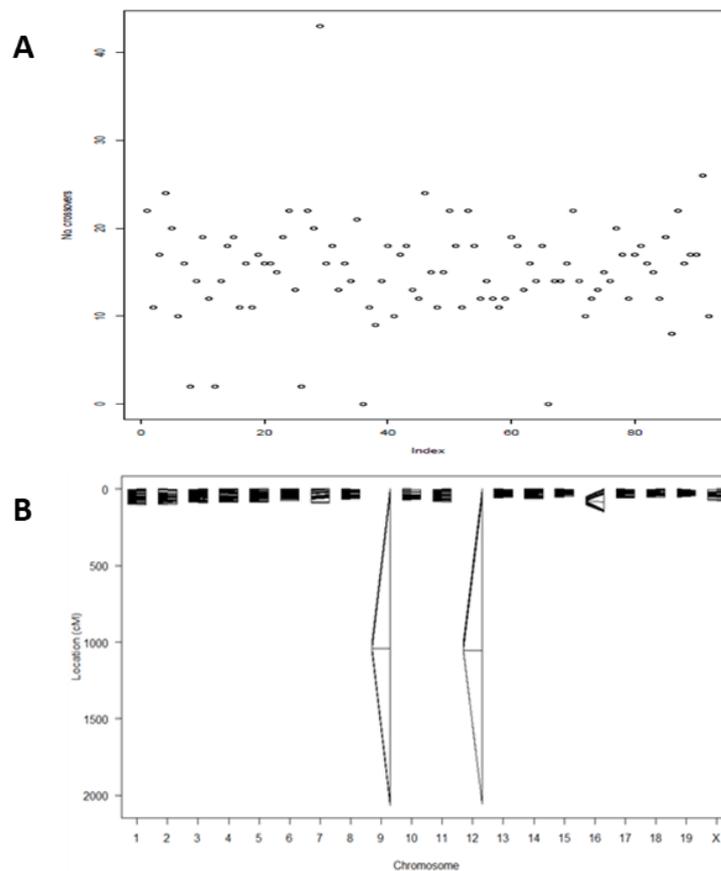
Para el control de calidad, llevamos nuestra matriz de datos a RStudio y cargamos el paquete “*qtl*” y nos identifica 92 individuos, 961 SNPs, 30 fenotipos [29 variables estudiadas, más una cualitativa para el sexo: macho (1) y hembra (0) y nos reconoce el cruce como *backcross* (BX)]. Se van realizando diferentes pasos, en primer lugar se inspeccionan las frecuencias de cada marcador (mostrar los SNPs que muestren una desviación sobre la frecuencia que cabría esperar en una población *backcross*, que sería 1:1). A continuación, se considera la probabilidad de que haya un error de genotipado en cada marcador, frente a que sea correcto, empleando la información de los marcadores que lo rodean (nos devuelve aquellos para los que la posibilidad de error está por encima del umbral).

Otra forma de identificar errores de genotipado es cuantificar los sobrecruzamientos [individuos con un número de sobrecruzamientos anormalmente bajo (menor de 5) o

elevado (mayor de 30), se consideran sospechosos]. También podemos comparar los genotipos de cada ratón dos a dos para identificar errores.

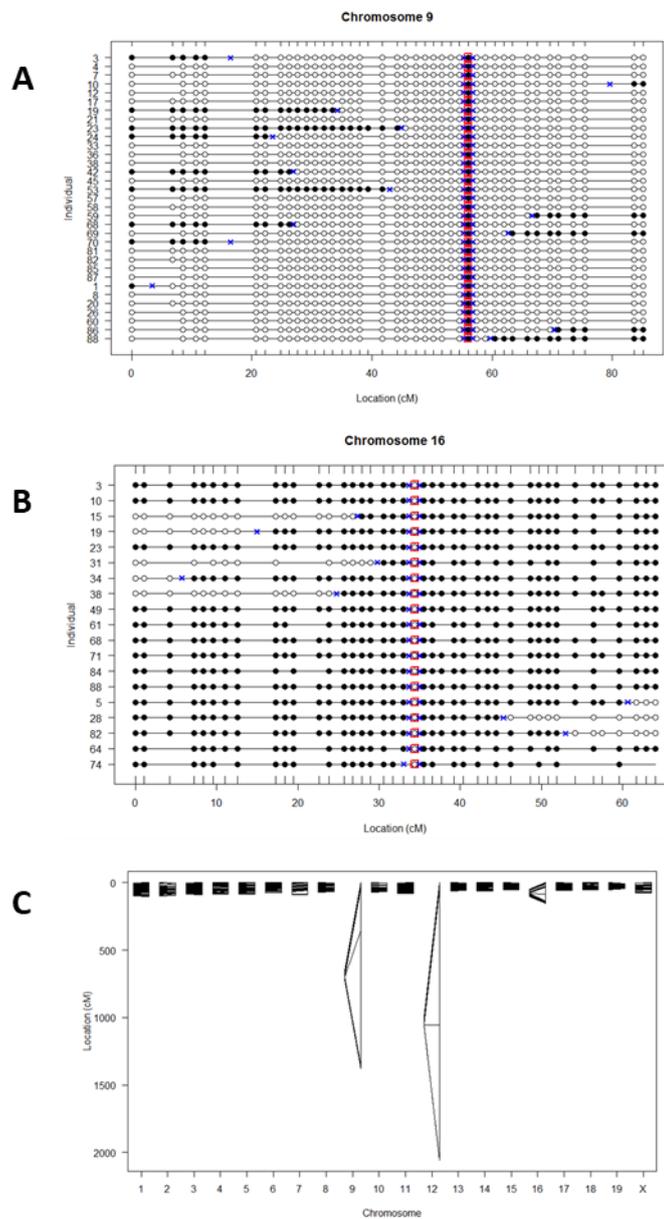
El siguiente paso es realizar un mapa genético con la información suministrada en las bases de datos para estimar las fracciones de recombinación entre los distintos SNPs, lo cual nos permite identificar problemas. Podemos generar el mapa genético esperado para nuestra población *backcross* y después generamos el mapa genético estimado empleando la información observada (fracción de recombinación). Finalmente, comparamos el mapa genético esperado (real) y el estimado, lo que permite identificar problemas con nuestra matriz de datos.

Realizamos cuatro controles de calidad. En el primer control de calidad tenemos 92 individuos, 961 SNPs y 30 fenotipos. Encontramos varios ratones con pocos sobrecruzamientos ( $Sc < 5$ ) y uno con un número de sobrecruzamientos muy elevado ( $Sc = 43$ ), como se aprecia en la Figura 4.19 A. La diferencia entre el mapa genético esperado y el estimado es grande, como podemos ver en la Figura 4.19 B.



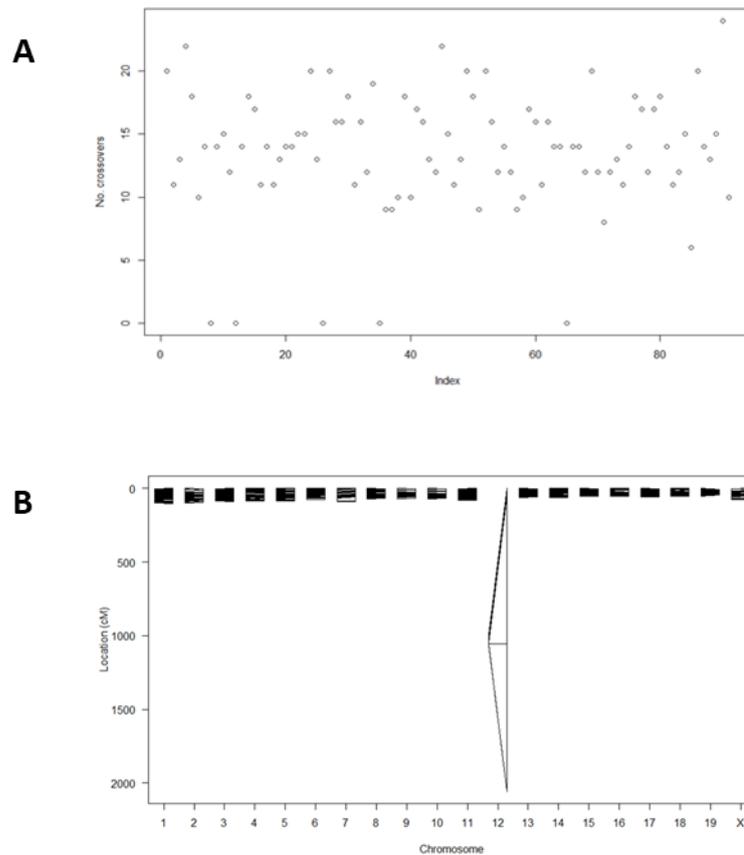
**Figura 4.19.** Resultados del primer control de calidad. A) Representación de los sobrecruzamientos. B) Comparación entre el mapa genético estimado y el esperado para el primer control de calidad.

En el segundo control de calidad, eliminamos de la matriz de datos el ratón M40 (por excesivos sobrecruzamientos) Así, tenemos 91 individuos, 961 SNPs y 30 fenotipos. Encontramos 5 ratones con pocos sobrecruzamientos y además, dos SNPs presentan bastantes errores, como podemos observar en la Figura 4.20 A y 4.20 B. Por último, obtenemos una comparativa entre mapa genético esperado y estimado diferente a la observada en la Figura 4.19, como podemos apreciar en la Figura 4.20 C. Por presentar bastantes errores, eliminamos los SNPs rs13480331 del cromosoma 9 y el rs4187220 del cromosoma 16, pero mantenemos los ratones con pocos sobrecruzamientos.



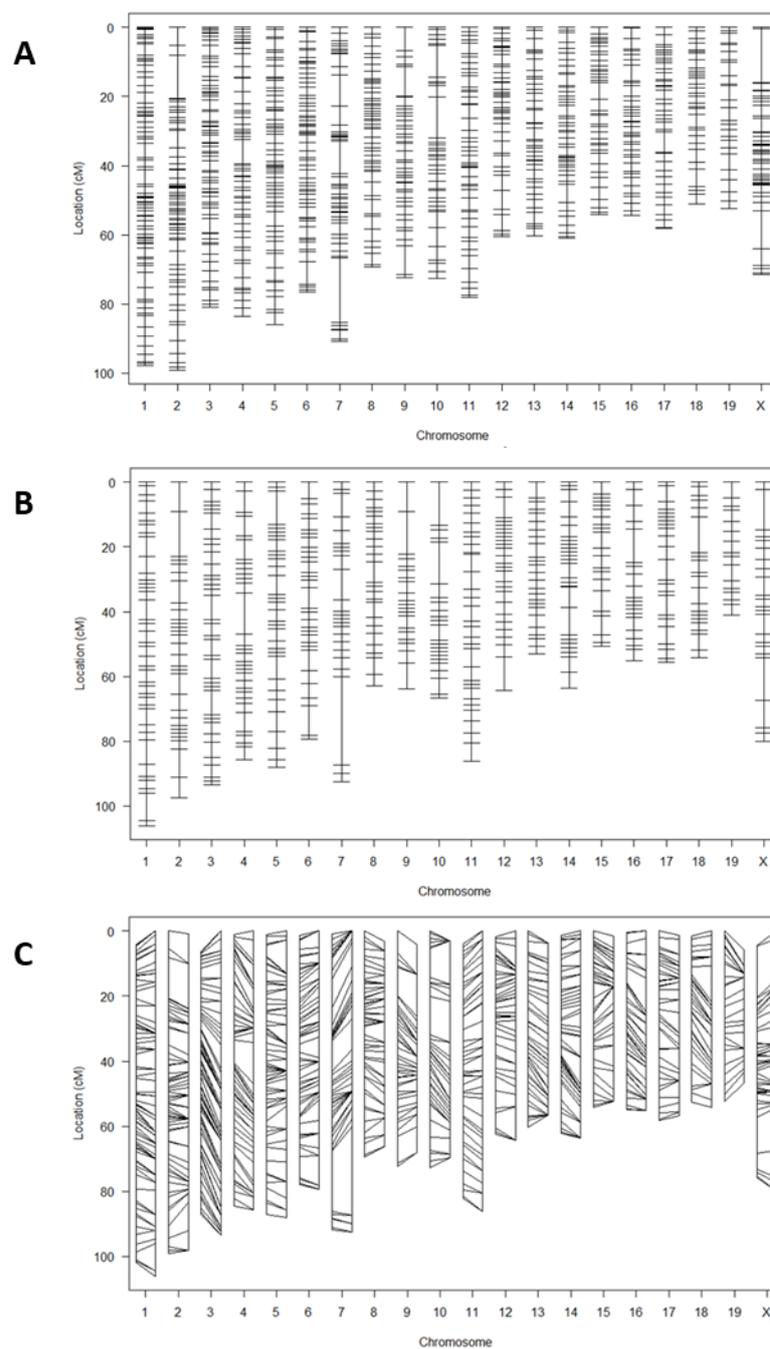
**Figura 4.20.** Resultados del segundo control de calidad. A) SNP problemático en el cromosoma 9. B) SNP problemático en el cromosoma 16. C) Comparación entre el mapa genético estimado y el esperado para el segundo control de calidad.

En el tercer control de calidad, después de los SNPs eliminados, encontramos 91 individuos, 959 SNPs y 30 fenotipos. En este caso, siguen apareciendo una serie de ratones que presentan 0 sobrecruzamientos (pueden ser animales parentales que se han colado en el estudio), como se muestra en la Figura 4.21 A, así como un SNP problemático. Obtenemos la comparativa de mapa genético estimado y esperado que se observa en la Figura 4.21 B.



**Figura 4.21.** Resultados del tercer control de calidad. A) Representación de los sobrecruzamientos. B) Comparación entre el mapa genético estimado y el esperado para el tercer control de calidad.

En el cuarto control de calidad, eliminados los ratones M11, M17, M35, M50 y H29 y el SNP rs 13481642 del cromosoma 12. De esta manera, incluimos 86 individuos, 958 SNPs y 30 fenotipos al introducir la matriz de datos en RStudio. Una vez realizadas las comprobaciones, no muestra ratones con exceso o defecto de sobrecruzamientos, ni SNPs problemáticos. En la Figura 4.22 vemos el mapa genético estimado en el panel A, el mapa genético esperado en el panel B y la comparación entre ambos mapas en el panel C.



**Figura 4.22.** Resultados del cuarto control de calidad. A) Mapa genético estimado. B) Mapa genético esperado. C) Comparación entre mapa genético esperado y estimado.

Después de los controles de calidad obtenemos una matriz de datos con las magnitudes fenotípicas de 86 ratones y 958 SNPs, con la que procedemos a realizar el análisis de ligamiento genético de los patofenotipos.

#### 4.4.2 Identificación de regiones genómicas asociadas a los patofenotipos.

Realizamos un análisis de ligamiento genético sobre los patofenotipos descritos en la cohorte F2BX; esto es, las variables parasitológicas y patológicas, las de poblaciones celulares y las de inmunoglobulinas. El análisis de ligamiento se emplea para determinar si dos marcadores están ligados o segregan de forma independiente. Esto es aplicable en la determinación de la relación entre el genotipo de marcadores genéticos y la distribución de un fenotipo de interés en una población heterogénea. Con ello podemos identificar regiones de rasgo cuantitativo (QTL- *quantitative trait loci*), en las que las diferencias en el genotipo se asocian con diferencias en el fenotipo estudiado.

Conseguimos identificar así una serie de regiones genómicas, asociadas a la variabilidad en el comportamiento frente a la enfermedad, a las que denominamos eQTL (expresión-QTL). Los resultados se recogen en la Tabla 4.4, para las variables parasitopatológicas, en la Tabla 4.5 para las variables de poblaciones celulares en sangre periférica y esplenocitos a la semana nueve post-infección y en la Tabla 4.6 para las variables de inmunoglobulinas a la semana nueve post-infección.

Encontramos al menos un eQTL asociado a todos los patofenotipos estudiados, excepto para la fecundidad de las hembras en el intestino (HPG\_int/H). Los patofenotipos para los que encontramos más eQTL fueron: la fecundidad de las hembras en el hígado (HPG\_hig/H), la superficie afectada del hígado (mm<sup>2</sup>/cm<sup>2</sup>) y la IgG2a. La presencia de un alelo procedente de C57BL/6J (AB) o la doble presencia del alelo CBA/2J (AA) en estas regiones genómicas, se asoció en cada eQTL con un mejor o peor comportamiento frente a la enfermedad, según se especifica en las Tablas 4.4, 4.5 y 4.6. En dichas tablas, se observa la siguiente información: el patofenotipo estudiado, la región genómica (eQTL) identificada, el cromosoma (Cr) en que aparece, el SNP en que se muestra (marcador en pico), los SNPs que entran en el intervalo de confianza 95% , la posición que ocupa en el cromosoma (cM), el *LOD-score* en el intervalo de confianza (siempre superior a 1.4 en el SNP en el que salta), la media y el error estándar de la media (EEM) de los ratones para el dato de cada patofenotipo y finalmente el genotipo que presenta mayor valor para cada patofenotipo (homocigoto o heterocigoto).

Como podemos apreciar en las Tablas 4.4, 4.5 y 4.6, se identifican 18 regiones genómicas eQTL, para las que en algún lugar de los 20 cromosomas (19 autosomas y el

sexual) de los ratones ha resultado un *LOD-score* superior a 1.4, en el que se muestra el pico y el SNP correspondiente, así como su intervalo de confianza al 95%.

En la Tabla 4.4A se muestran las regiones genómicas (eQTL) asociadas a la variabilidad en los patofenotipos de las variables patológicas. Se observa que para los granulomas (Granulomas.cm2) presenta un pico en el cromosoma 15 para el SNP rs6169611 con *LOD-score* de 1.51 y el genotipo con mayor valor es el AA. También presenta un pico en el cromosoma 18 para el SNP rs13483259 con *LOD-score* de 1.48 y el genotipo con mayor valor es AB.

La superficie afectada del hígado (mm2/cm2) presenta cuatro SNPs con *LOD-score* superior a 1.4. En el cromosoma 2, el SNP rs13476772 presenta un *LOD-score* de 1.47 y el genotipo con mayor valor es AA. En el cromosoma 15, para el SNP rs13482664 muestra un *LOD-score* de 2.03 y el genotipo de mayor valor es AA. Por otro lado, en el cromosoma 17, con un *LOD-score* de 1.54 para el SNP 13482845, el genotipo de mayor valor es el AB. Finalmente, en el cromosoma 18, el SNP rs13483259 presenta un *LOD-score* de 1.54 y el genotipo de mayor valor es AB.

**Tabla 4.4.A** eQTL asociados con la variabilidad en los patofenotipos de las variables parásito-patológicas de la cohorte F2BX. Cr: cromosoma; cM: Centimorgan; EEM: error estándar de la media; AA: genotipo homocigoto con dos alelos de la cepa CBA/2J; AB: genotipo heterocigoto con un alelo de CBA/2J y otro de C57BL/6J.

Fenotipo	eQTL	Cr	Marcador en pico	Intervalo de confianza (0.05)	Posición (cM)	LOD Score	Media ± EEM		Genotipo con mayor valor
							AA	AB	
Granulomas.cm2	eQTL1	15	rs6169611	rs13482425	4.09	0.46	152.61±6.49	127.64±6.64	AA
					32.17	1.51			
				c15.loc40	44.09	0.45			
	eQTL2	18	rs13483259	rs13483197	4.80	1.08	128.71±6.31	154.42±6.90	AB
					14.18	1.48			
				rs13483455	53.04	0.11			
mm2/cm2	eQTL3	2	rs13476772	c2.loc20	23.68	0.09	56.63±2.73	46.70±2.52	AA
					63.25	1.47			
	eQTL1	15	rs13482664	rs6257221	102.73	0.34	56.84±2.56	45.31±2.62	AA
					7.08	0.91			
				c15.loc37.5	41.59	0.95			
	eQTL4	17	rs13482845	rs13482845	2.52	1.54	45.55±2.80	55.65±2.49	AB
					2.52	1.54			
				rs13483171	60.65	0.50			
	eQTL2	18	rs13483259	rs13483197	4.80	0.80	46.59±2.55	56.78±2.80	AB
					14.18	1.54			
			rs13483466	55.84	0.01				

En la Tabla 4.4B se muestran las regiones genómicas (eQTL) asociadas a la variabilidad en los patofenotipos de las variables parasitológicas. Así, se observa que los vermes totales muestran en el cromosoma 6 el SNP rs13478921 con un *LOD-score* de 1.98 y el genotipo de mayor valor AA y para el cromosoma 7, el SNP rs13479382 con *LOD-score* de 1.57 y genotipo con mayor valor AB. Por otro lado, los vermes macho presentan en el cromosoma 6 un *LOD-score* de 2.01 para el SNP rs13478799 y genotipo AA y en el cromosoma 7, el SNP rs13479382 y *LOD-score* de 1.64, con genotipo de mayor valor AB. Finalmente, los vermes hembra presentan en el cromosoma 6, para el SNP c6.loc57.5 un *LOD-score* de 2.11 y genotipo de mayor valor AA.

Los huevos retenidos en el hígado (HPG\_higado), presentan en el cromosoma 8, un *LOD-score* de 2.82 para el SNP rs13479680 y genotipo de mayor valor AB y en el cromosoma 17, para el SNP rs13474581 un *LOD-score* de 1.41 y genotipo AB. Mientras, los huevos atrapados en el intestino (HPG\_intest), muestran tres SNPs con *LOD-score* superior a 1.4. En el cromosoma 3, el SNP rs13477093 presenta un *LOD-score* de 1.67 y genotipo AB; el cromosoma 6 muestra un *LOD-score* de 1.50 para el SNP rs3023064 y genotipo AA; mientras que en el cromosoma 8, para el SNP rs13479680 el *LOD-score* es de 1.61 y genotipo con mayor valor AB.

Finalmente, la fecundidad de las hembras en relación a los huevos retenidos en el hígado (HPG\_hig\_hembra) presenta cuatro SNPs con *LOD-score* superior a 1.4. En el cromosoma 1, el SNP rs13475979, muestra un *LOD-score* de 1.49 y genotipo AA; en el cromosoma 2, el SNP rs6257221 presenta un *LOD-score* de 1.56 y genotipo AA; el cromosoma 5, muestra un *LOD-score* de 2.37 para el SNP rs13478485 y genotipo de mayor valor AB y finalmente el cromosoma 11, en el SNP rs13481156 presenta un *LOD-score* de 1.41 y genotipo AB. En cambio, no hay ninguna región cromosómica asociada a la fecundidad basada en huevos en la pared del intestino (HPG\_int\_hembra).

El número de vermes total, el número de vermes macho y el número de vermes hembra, presentan en el cromosoma X, para el mismo SNP rs13484110 y con un *LOD-score* respectivo de 2.00 para los vermes totales, 2.28 para los vermes macho y 1.45 para los vermes hembra un genotipo de mayor valor para los homocigotos, tanto en F2BX machos como F2BX hembras

**Tabla 4.4.B** eQTL asociados con la variabilidad en los patofenotipos de las variables parásito-patológicas de la cohorte F2BX. Cr: cromosoma; cM: Centimorgan; EEM: error estándar de la media; AA/AY: genotipo homocigoto con dos alelos CBA/2J; AB/BY: genotipo heterocigoto un alelo CBA/2J y otro C57BL/6J.

Fenotipo	eQTL	Cr	Marcador en pico	Intervalo de confianza (0.05)	Posición (cM)	LOD Score	Media ± EEM				Genotipo con mayor valor
							AA	AB	AY	BY	
Verme_totales	eQTL5	6	rs13478921	c6.loc2.5	4.38	0.78	13.09±0.95	9.47±0.79			AA
				c6.loc70	45.60	1.98					
				rs13479105	71.88	0.76					
	eQTL6	7	rs13479382	rs13479105	2.66	0.38	9.31±0.84	12.73±0.88			AB
			rs13479572	51.74	1.57						
	eQTL7	X	rs13483769	rs13483713	93.54	0.05	12.81±1.20	10.46±1.08	13.04±1.28	7.72±1.22	AY/AA/AB/BY
			rs13484110	4.30	0.30						
Verme_macho	eQTL5	6	rs13478799	c6.loc2.5	4.38	1.10	13.11±0.92	9.31±0.80			AA
				c6.loc60	30.04	2.01					
				rs13479105	61.88	1.07					
	eQTL6	7	rs13479382	rs13479105	2.66	0.31	9.31±0.84	12.73±0.88			AB
			c7.loc77.5	51.74	1.64						
	eQTL7	X	rs13483769	rs13483736	80.16	0.02	12.81±1.20	10.46±1.08	13.01±1.27	7.66±1.23	AY/AA/AB/BY
			rs13484110	24.21	0.43						
Verme_hembra	eQTL5	6	c6.loc57.5	c6.loc27.5	29.38	1.14	10.48±0.73	7.29±0.70			AA
				rs13479093	59.38	2.11					
				rs13483713	78.35	1.06					
	eQTL7	X	rs13483769	rs13483713	4.30	0.29	9.67±1.03	8.15±0.93	10.73±1.10	7.07±1.05	AY/AA/AB/BY
			rs13484110	30.62	1.45						
HPG_higado	eQTL8	8	rs13479680	rs13479622	8.41	2.09	6971.09±883.81	11563.40±843.15			AB
				rs13479788	19.01	2.82					
				rs13482845	31.94	2.01					
	eQTL4	17	rs13474581	c17.loc50	2.52	0.85	7481.65±966.22	10806.63±839.62			AB
				4.55	1.41						
HPG_intest	eQTL9	3	rs13477093	rs13476954	1.54	0.12	5161.36±810.57	8679.62±891.95			AB
				rs13477527	20.78	1.67					
				rs3023064	rs6172481	82.40	0.13				
	eQTL5	6	rs3023064	rs13479016	1.88	0.87	8442.27±874.16	5138.64±832.61			AA
			rs13479622	7.99	1.50						
	eQTL8	8	rs13479680	rs13480014	59.60	0.62	4953.33±870.26	8304.55±829.49			AB
				8.41	1.20						
HPG_hig_hembra	eQTL10	1	rs13475979	rs13475821	18.39	0.06	1685.07±199.20	1031.32±142.09			AA
				rs13476316	47.01	1.49					
				c2.loc15	99.38	0.08					
	eQTL3	2	rs6257221	rs6257221	18.68	0.04	1552.30±159.78	922.42±167.33			AA
			rs13478428	102.73	1.56						
	eQTL11	5	rs13478485	rs13480863	102.73	1.56	954.78±142.56	1752.94±185.19			AB
			c5.loc77.5	49.64	2.33						
	eQTL12	11	rs13481156	rs6225297	60.10	2.37	967.33±159.50	1570.61±169.12			AB
				rs13480863	80.21	0.83					
				rs6225297	55.29	1.41					
HPG_intest_hembra					83.98	0.31					

En la Tabla 4.5 se muestran los resultados de las regiones genómicas (eQTL) asociadas a la variabilidad en los patofenotipos de las variables de poblaciones celulares tanto en sangre periférica como en esplenocitos en la semana 9 post-infección.

En CD4 en sangre (CD4\_Sg\_S9) muestra un *LOD-score* de 1.46 para el SNP rs13481826 en el cromosoma 13 y un genotipo de mayor valor AB. En CD8 en sangre (CD8\_Sg\_S9) presenta dos SNP, uno en el cromosoma 14, rs8251327 y *LOD-score* de 1.69, además de genotipo AA y otro en el cromosoma 17, con *LOD-score* de 1.58 para el SNP rs13482998 y genotipo AA. Para CD45 en sangre (CD45\_Sg\_S9) muestra en el cromosoma 14 un *LOD-score* de 1.52 para el SNP rs13481826 y un genotipo AA. Finalmente, para B220 en sangre (B220\_Sg\_S9), presenta dos SNP con *LOD-score* superior a 1.4, uno en el cromosoma 13 y *LOD-score* de 1.63 para el SNP 13481826 y genotipo AB y otro para el cromosoma 19, en el SNP rs13483665 y *LOD-score* de 1.87, con genotipo de mayor valor AA.

Para CD4 en esplenocitos (CD4\_Esp\_S9) muestra dos SNPs, uno en el cromosoma 2, con *LOD-score* de 1.45 para el SNP rs13476358 y genotipo AB y otro en el cromosoma 5, para el SNP rs13478264 y *LOD-score* de 2.58 para el genotipo AB. Para las células CD8 en esplenocitos (CD8\_Esp\_S9) presenta tres SNP con valor superior a 1.4. En el cromosoma 5 para el SNP rs13478264, con *LOD-score* de 2.67 y genotipo de mayor valor AB; en el cromosoma 11, el SNP rs13481122, con *LOD-score* de 1.66 y genotipo AB y en el cromosoma 13, para el SNP rs4229866, con *LOD-score* de 1.68 y genotipo de mayor valor AB. Para CD45 en esplenocitos (CD45\_Esp\_S9) muestra en el cromosoma 11 el SNP rs13481072 con *LOD-score* de 1.46 y genotipo AB. Finalmente, en B220 en esplenocitos (B220\_Esp\_S9), presenta en el cromosoma 1 el SNP rs13475821, con *LOD-score* de 1.95 y genotipo con mayor valor AB.

**Tabla 4.5.** eQTL asociados con la variabilidad en los patofenotipos de las variables poblaciones celulares en sangre periférica y esplenocitos en la semana nueve post-infección de la cohorte F2BX. Cr: cromosoma; cM: Centimorgan; EEM: error estándar de la media; AA: genotipo homocigoto con dos alelos de la cepa CBA/2J; AB: genotipo heterocigoto con un alelo de CBA/2J y otro de C57BL/6J.

Fenotipo	eQTL	Cr	Marcador en pico	Intervalo de confianza (0.05)	Posición (cM)	LOD Score	Media ± EEM		Genotipo con mayor valor
							AA	AB	
CD4_Sg_S9	eQTL13	13	rs13481826	rs13481713	6.72	0.99	0.058±0.009	0.093±0.01	AB
				c13.loc50	30.69	1.46			
CD8_Sg_S9	eQTL14	14	rs8251327	rs13482045	5.53	0.54	0.058±0.006	0.032±0.007	AA
				rs13482415	28.08	1.69			
	eQTL4	17	rs13482998	rs13482845	66.55	0.09	0.060±0.007	0.035±0.006	AA
				rs13483126	2.52	0.45			
CD45_Sg_S9	eQTL14	14	rs8251327	rs13482045	5.53	0.16	0.30±0.028	0.19±0.03	AA
				rs13482415	28.08	1.52			
B220_Sg_S9	eQTL13	13	rs13481826	rs13481713	6.72	1.00	0.10±0.17	0.17±0.019	AB
				c13.loc47.5	30.69	1.63			
	eQTL15	19	rs13483665	rs13483595	54.22	0.03	0.18±0.018	0.10±0.017	AA
				rs4232268	27.23	0.85			
CD4_Esp_S9	eQTL3	2	rs13476358	rs13476332	45.23	1.87	0.056±0.003	0.069±0.004	AB
				rs13476919	56.42	1.15			
	eQTL11	5	rs13478264	c5.loc2.5	30.75	2.58	0.055±0.003	0.071±0.003	AB
				rs13478444	51.83	1.34			
CD8_Esp_S9	eQTL11	5	rs13478264	rs13478130	9.25	1.01	0.026±0.002	0.038±0.002	AB
				rs4225380	30.75	2.67			
	eQTL12	11	rs13481122	rs13480872	47.74	0.81	0.026±0.002	0.036±0.002	AB
				c11.loc67.5	7.03	0.04			
	eQTL13	13	rs4229866	rs13481772	48.96	1.66	0.027±0.002	0.036±0.002	AB
				rs13482027	73.37	0.07			
CD45_Esp_S9	eQTL12	11	rs13481072	rs13480863	19.13	0.16	0.19±0.02	0.27±0.02	AB
				rs6225297	35.76	1.68			
B220_Esp_S9	eQTL10	1	rs13475821	rs13475697	64.90	0.11	0.33±0.022	0.41±0.016	AB
				rs13476132	1.61	1.45			

En la Tabla 4.6 se exponen las regiones genómicas (eQTL) asociadas con la variabilidad en los patofenotipos de producción de inmunoglobulinas frente a la infección en la semana nueve post-infección.

Así observamos que para la IgG (IgGT\_S9), presenta en el cromosoma 11 el SNP rs13481062 con *LOD-score* de 2.03 y genotipo de mayor valor AB y además, en el cromosoma 17, para el SNP rs13482845 un *LOD-score* de 3.42 y genotipo AB. Para la IgG1 (IgG1\_S9), muestra dos SNPs, uno en el cromosoma 16, rs4204623, con *LOD-score*

de 2.85 y genotipo AB y otro en el cromosoma 19, rs13483612, con *LOD-score* de 1.74 y genotipo de mayor valor AB.

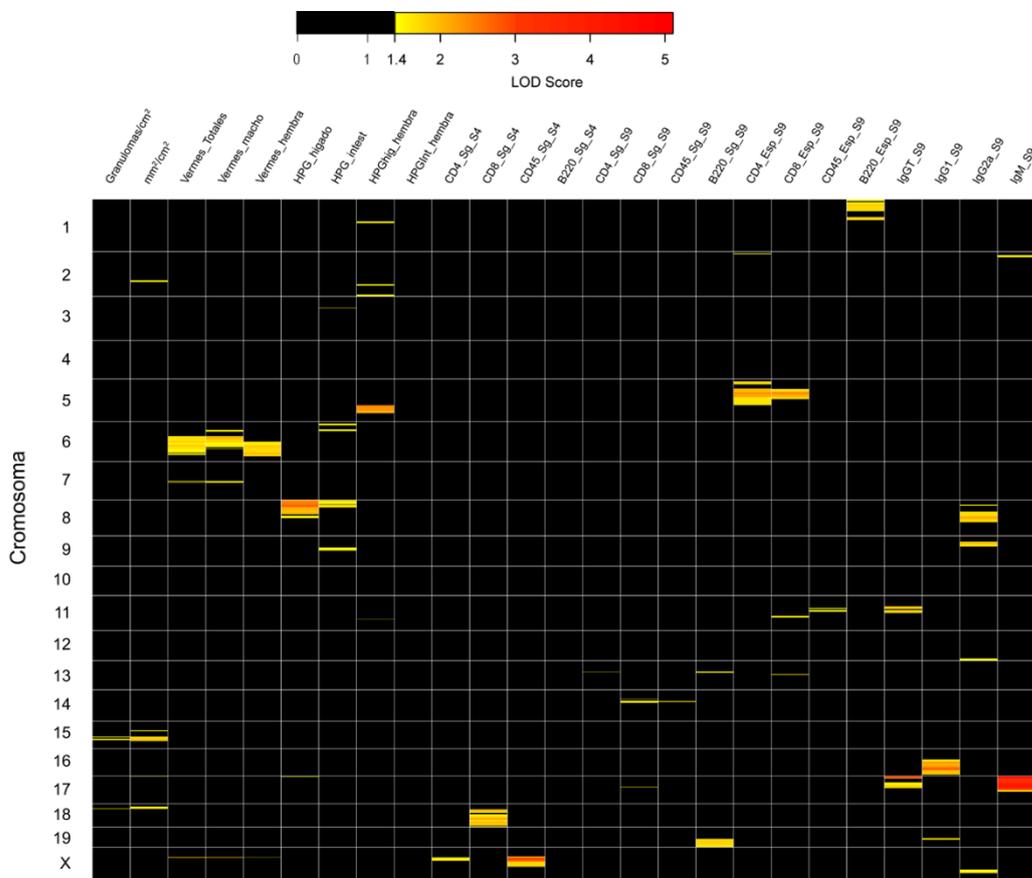
Por otro lado, para la IgG2a, presenta tres SNPs con *LOD-score* superior a 1.4. En el cromosoma 8, en el SNP rs13479846, muestra un *LOD-score* de 2.13 y genotipo de mayor valor AB. En el cromosoma 9, el SNP rs4227609 presenta *LOD-score* de 2.31 y genotipo AB. En el cromosoma 12, el SNP c12.loc57.5, muestra un *LOD-score* de 1.54 y genotipo de mayor valor AA. Finalmente, señalar que para la IgG2a presenta en el cromosoma X el SNP rs13483959 un *LOD-score* de 1.44 y el genotipo de mayor valor es para los heterocigotos, tanto en F2BX machos como en F2BX hembras.

En cuanto a la IgM (IgM\_S9), se observan dos SNPs, uno en el cromosoma 2, con *LOD-score* de 1.57 para el SNP c2.loc20 y genotipo AB y otro en el cromosoma 17, para el SNP rs13482845, con el mayor *LOD-score* de todos los SNPs, 4.96 y un genotipo de mayor valor AB.

**Tabla 4.6.** eQTL asociados con la variabilidad en los patofenotipos de las variables inmunoglobulinas semana nueve post-infección de la cohorte F2BX. Cr: cromosoma; cM: Centimorgan; EEM: error estándar de la media; AA/AY: genotipo homocigoto con dos alelos CBA/2J; AB/BY: genotipo heterocigoto con un alelo CBA/2J y otro C57BL/6J.

Fenotipo	eQTL	Cr	Marcador en pico	Intervalo de confianza (0.05)	Posición (cM)	LOD Score	Media ± EEM				Genotipo con mayor valor
							AA	AB	AY	BY	
IgGT_S9	eQTL12	11	rs13481062	rs13480863	5.87	0.59	0.7±0.07	1.02±0.07			AB
				rs13481146	38.91	2.03					
	eQTL4	17	rs13482845	rs13482845	52.76	0.46	0.15±0.01	0.17±0.009			AB
				rs13482998	2.52	3.42					
IgG1_S9	eQTL16	16	rs4204623	rs4182798	2.52	3.42	0.48±0.04	0.70±0.04			AB
				rs3708391	20.63	1.86					
	eQTL15	19	rs13483612	rs13483503	30.52	1.87	0.49±0.04	0.67±0.04			AB
				rs4232247	42.53	2.85					
IgG2a_S9	eQTL8	8	rs13479846	rs13479630	54.32	1.90	0.31±0.036	0.47±0.037			AB
				rs6356426	4.08	0.70					
	eQTL17	9	rs4227609	rs13480103	32.11	1.74	0.31±0.035	0.48±0.036			AB
				rs6292067	40.74	0.36					
	eQTL18	12	c12.loc57.5	rs13481278	10.31	0.66	0.46±0.036	0.32±0.037			AA
				rs3024062	37.83	2.13					
	eQTL7	X	rs13483959	cX.loc12.5	48.79	0.69	0.33±0.047	0.52±0.052	0.34±0.055	0.37±0.053	AB/BY/AY/AA
				cX.loc70	9.44	0.42					
IgM_S9	eQTL3	2	c2.loc20	rs13476332	24.58	2.31	0.067±0.0049	0.087±0.0045			AB
				c2.loc87.5	48.51	0.69					
	eQTL4	17	rs13482845	rs13482845	59.92	1.54	0.069±0.0044	0.089±0.0051			AB
				c17.loc20	63.03	1.44					

A modo de resumen de todos los datos presentados en las tablas anteriores, se muestra la Figura 4.23, en la que se representa un *heatmap* de los cromosomas, a la izquierda y todos los patofenotipos en la parte superior y los puntos en que aparece un SNP en el que el valor es superior a 1.4, en una escala de intensidad de amarillo a rojo. Cuanto más alto es el valor del *LOD-score*, el color representado es más rojo y cuanto más cercano es a 1.4 es más amarillo. En negro están los marcadores para los que el *LOD-score* es inferior a 1.4.



**Figura 4.23.** *Heatmap* con los *LOD-scores* resultado del análisis de ligamiento entre los marcadores genéticos y los fenotipos. Cada punto del *heatmap* representa el *LOD-score* del ligamiento con el fenotipo para cada marcador genético estudiado. Estos marcadores se distribuyen ordenados a lo largo de los 20 cromosomas del ratón (izquierda), ordenados según su posición. Los *LOD-scores* < 1.4 se representan en color negro. Sólo se incluyen los fenotipos de respuesta al tratamiento para los que se encontró algún QTL con  $LOD-score \geq 1.4$ . Se representan en amarillo los que están más cerca de 1.4 y va cambiando a naranja o rojo según aumenta su valor.

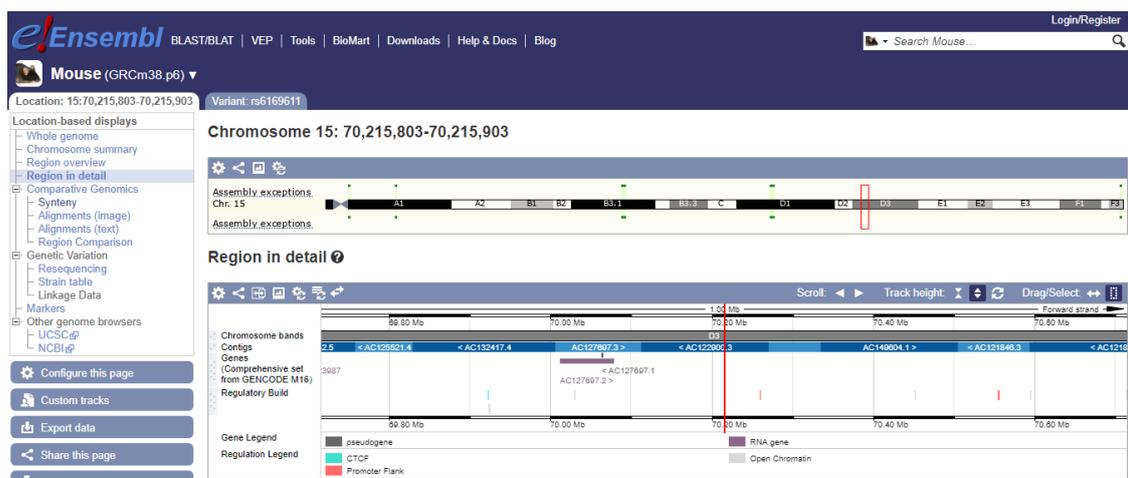
#### 4.4.3 Estudio de la ligación del daño hepático y QTL en los cromosomas 15 y 18

Como conclusión de estos resultados, podemos indicar que tanto el SNP rs6169611 del cromosoma 15, como el SNP rs13483259 del cromosoma 18, están íntimamente asociados al número de granulomas, variable más representativa del daño hepático.

Empleando la herramienta bioinformática *Ensembl* (<https://www.ensembl.org/index.html>), se buscaron los cromosomas y las regiones donde se presentan los SNPs seleccionados, analizando la sintenia (conservación de conexiones genéticas entre animales poco relacionados) entre ratón y humano.

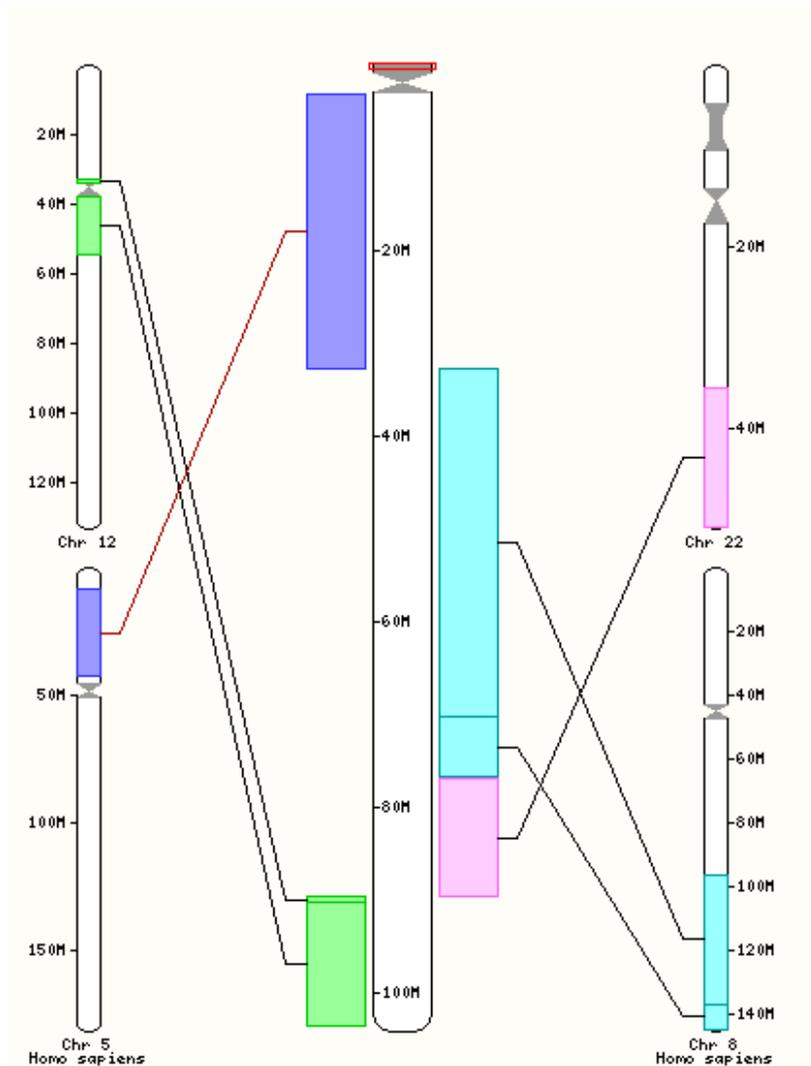
##### - Localización del SNP rs6169611 del cromosoma 15 de ratón

Para el SNP rs 6169611 del cromosoma 15, los resultados obtenidos aparecen en la Figura 4.24, donde se muestra el cromosoma 15 y la región donde se encuentra el SNP encuadrada en color rojo.



**Figura 4.24.** Representación de la localización del SNP rs6169611 del cromosoma 15 del ratón. Región del SNP encuadrada en rojo.

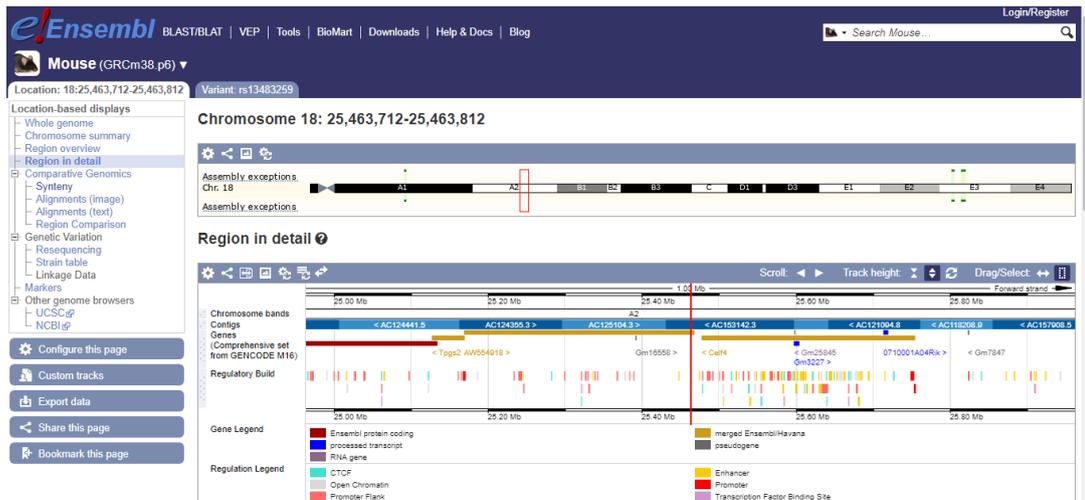
En la Figura 4.25 se muestra la sintenia entre el cromosoma 15 del ratón, con el SNP localizado en la parte superior, recuadro rojo y las regiones de los cromosomas humanos en los que estarían conectados. Como se observa en la figura, aparece sintenia en los cromosomas 5, 8, 12 y 22 de humanos, pero no muestra que el SNP rs6169611 aparezca conservado en los cromosomas humanos.



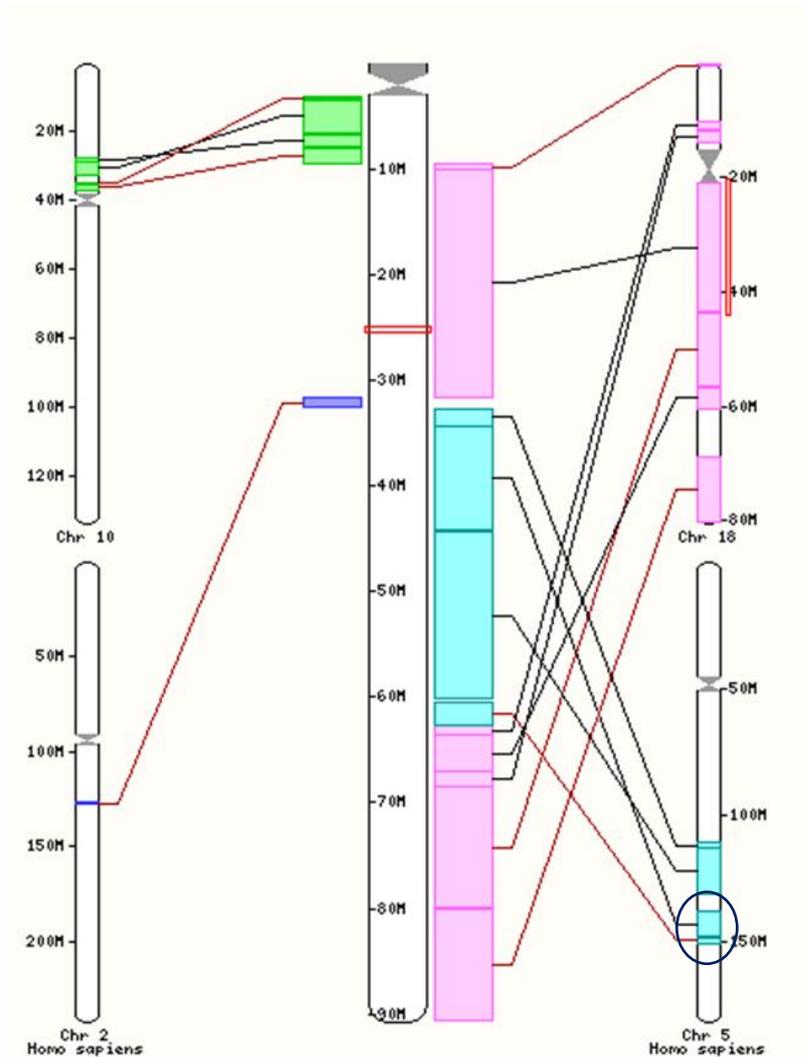
**Figura 4.25.** Representación de la sintenia entre el cromosoma 15 de ratón y los cromosomas humanos 5, 8, 12 y 22. En recuadro rojo, región del cromosoma 15 del ratón donde se encuentra el SNP.

**- Localización del SNP rs13483259 del cromosoma 18 de ratón**

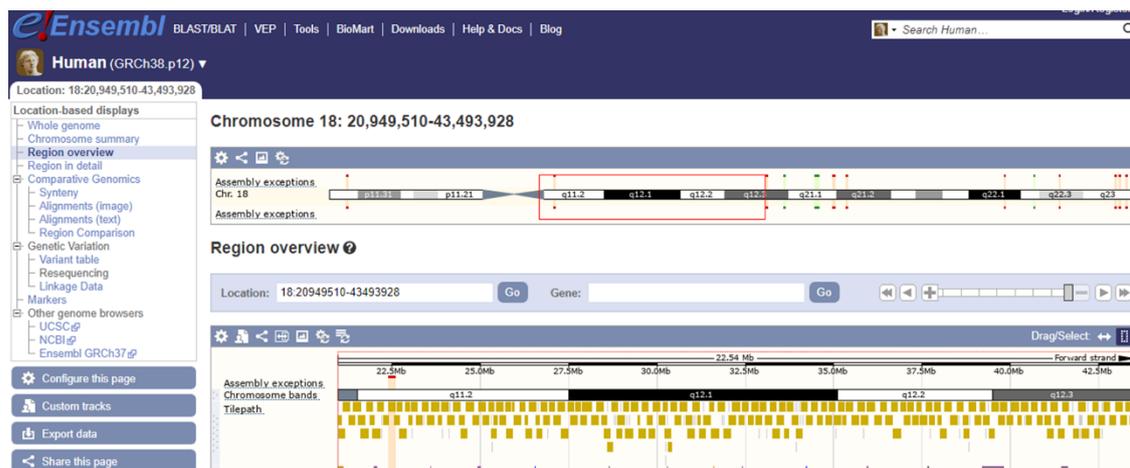
En cuanto al SNP rs 13483259 del cromosoma 18, en la Figura 4.26 se representa su localización en dicho cromosoma del ratón (en recuadro rojo). En la Figura 4.27 se muestra la sintenia entre el cromosoma 18 del ratón, con el SNP marcado con recuadro rojo y las regiones de los cromosomas humanos en los que estarían conectados. Como se observa en la figura, aparece sintenia en los cromosomas 2, 5, 10 y 18 de humanos. En el cromosoma 18 de humanos se muestra la región conservada donde está el SNP del ratón, que corresponde a la localización 18q11.2-12.3, como se observa en la Figura 4.28. En el cromosoma 5 humano, hay una región asociada a la intensidad de la infección, locus SM1 (5q31-33) y que está marcada con un círculo azul en el cromosoma 5 humano, pero como se observa en la Figura 4.27, no está relacionada con el SNP rs13483259.



**Figura 4.26.** Representación de la localización del SNP rs13483259 del cromosoma 18 del ratón. Región del SNP encuadrada en rojo.



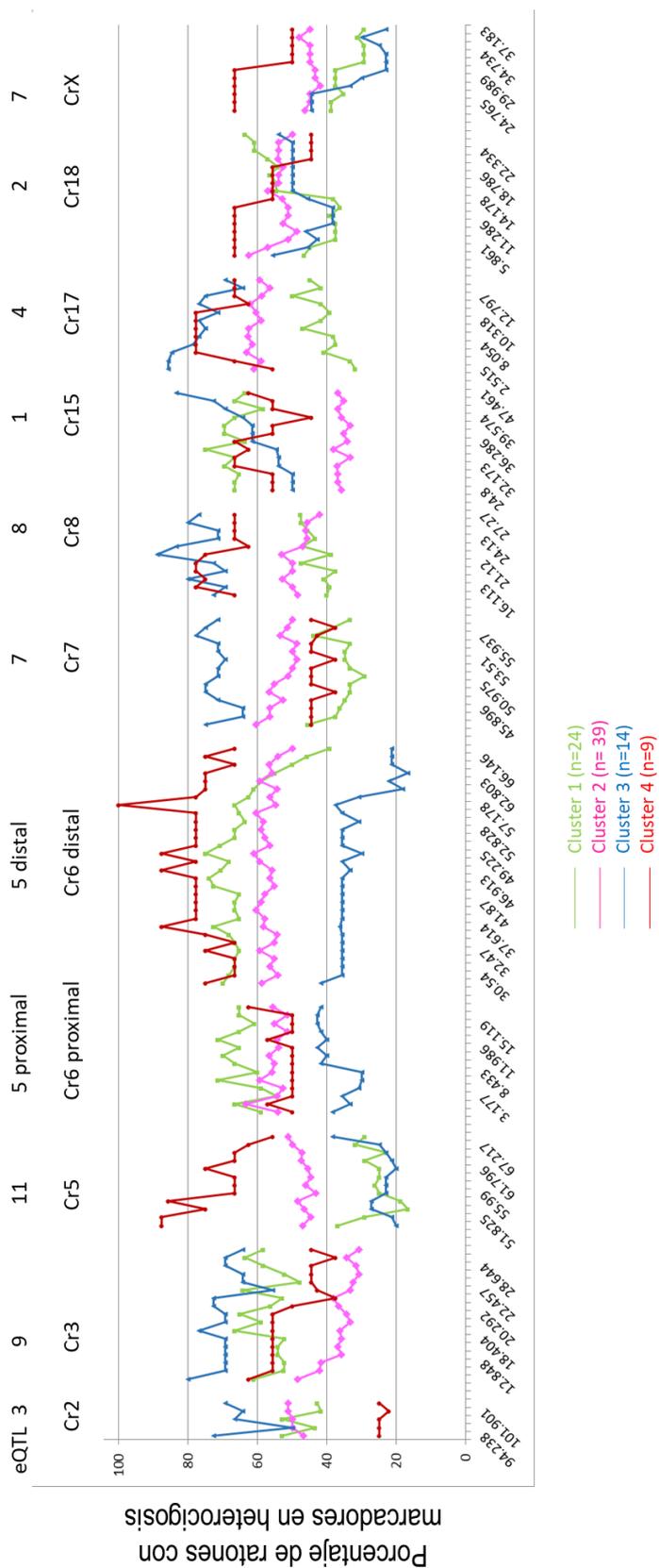
**Figura 4.27.** Representación de la sintenia entre el cromosoma 18 de ratón y los cromosomas 2, 5, 10 y 18 de humano. En recuadro rojo el SNP rs13483259, que aparece en el cromosoma 18 de humano. Con círculo azul, locus SM1 (5q31-33) asociado a la intensidad de la infección, en el cromosoma 5 humano.



**Figura 4.28.** Localización del SNP rs 13483259 del cromosoma 18 de ratón en el cromosoma 18 de humano, 18q11.2-12.3.

#### 4.4.4 Identificación de regiones genómicas ligadas a la susceptibilidad a la enfermedad.

Como el componente genético influye en la progresión de la enfermedad, estudiamos los QTLs (*quantitative trait loci*) asociados a la esquistosomosis por análisis de ligamiento en la cohorte F2BX. Para ello, identificamos los marcadores genéticos asociados a los grupos definidos de ratones según la gravedad frente a la enfermedad (Grupos 1 a 4). Se muestran 11 loci, de los 18 descritos en el Apartado 4.4.2, asociados a la enfermedad, denominados eQTL (expresiónQTL). Significativamente, como se muestra en la Figura 4.29, el Grupo 4 incluyó mayor porcentaje de ratones heterocigotos para eQTL11 (Cr 5), eQTL5 distal (Cr 6 distal) y para el eQTL7 (Cr X). Además, en el eQTL3 (Cr2), el menor porcentaje de heterocigotos lo presenta este Grupo 4, a la vez que en dicho eQTL el mayor porcentaje de heterocigosis lo presenta el Grupo 3 (el anterior al más grave). También observamos que para los eQTL8 (Cr 8) y el eQTL4 (Cr17), los grupos con mayor porcentaje de ratones heterocigotos eran el 3 y el 4, los más graves frente a la enfermedad. Del mismo modo, en el eQTL4 (Cr17), se observa que el Grupo 1, el menos grave, es el que menor porcentaje de heterocigotos presenta. Por otro lado, en el eQTL5 distal (Cr6 distal), el Grupo 3 es el que menor número de ratones heterocigotos presenta, pues en este eQTL el mayor porcentaje de ratones heterocigotos lo presenta el Grupo 4 (más grave).



**Figura 4.29.** Distinción de los grupos definidos a nivel genético. La figura muestra el porcentaje de ratones con un marcador genético heterocigótico particular, diferencialmente presente en los cuatro grupos de ratones, según la gravedad de la enfermedad.

#### 4.5 Análisis multivariante por regresión logística binaria como método de validación de la formación de los grupos extremos de enfermedad

Con el fin de confirmar la clasificación de los ratones en un grupo más grave (denominado en su momento como Grupo 4), frente al grupo menos grave (Grupo 1), propuestos en el apartado “Análisis de conglomerados por k-medias”, se ha empleado la técnica multivariante de regresión logística binaria (RLB), que resulta adecuada para este objetivo. Así mismo, esta técnica permite estudiar qué covariables influyen en la clasificación de los dos grupos y sirve para crear un modelo explicativo y también predictivo a partir del modelo final encontrado por la regresión.

Como se ha expuesto en “Materiales y Métodos”, el procedimiento RLB se basa en realizar un ajuste de “Probabilidad de Máxima Verosimilitud” a la ecuación:

$$\ln \frac{p(1)}{(1-p(1))} = B_0 + B_1x_1 + B_2x_2 + \dots B_mx_m$$

donde  $p(1)$  sería, en nuestro caso, la probabilidad de pertenencia de un ratón al Grupo 4 (tipificado con el valor “1”), frente a  $(1-p(1))$ , que sería la probabilidad de que dicho ratón pertenezca al Grupo 1 (tipificado con valor “0”). El parámetro  $B_0$  es un término constante a modo de línea base y los parámetros  $B_i$  son los coeficientes estimados para el ajuste de cada covariable “i”. Como covariables  $x_i$  iniciales se eligieron las siguientes:  $x_1$  = número de granulomas (granulomas),  $x_2$  = superficie afectada del hígado (superficie) y  $x_3$  = número de huevos por gramo de hígado (HPG\_hígado), ya que serían las variables que podrían haber determinado la segregación de los casos en los dos grupos (mayor y menor enfermedad).

Debido a la diferencia de magnitud entre las variables, la variable número de huevos por gramo de hígado ha sido necesario normalizarla dividiendo sus valores por 100, con el fin de que todas las variables fuesen homogéneas, ya que sino unas variables predominarían sobre otras y sesgarían el ajuste de regresión. El procedimiento de ajuste consiste en introducir al algoritmo todas las variables a probar y a continuación ir suprimiendo las variables que no sean significativas en base a su p-valor, que debe ser  $p < 0.05$  ó de una manera más flexible  $p < 0.1$  ó  $p < 0.2$ .

Tras intentar obtener el ajuste con las tres variables arriba mencionada, hubo que suprimir la variable número de huevos por gramo de hígado (HPG\_hígado) debido a que

mostraba un p-valor elevado. Los resultados finales obtenidos con las variables que permanecieron como significativas en el modelo de RLB, se han recogido en la Tabla 4.7.

**Tabla 4.7.** Parámetros del ajuste a un modelo de RLB que analiza la pertenencia de los ratones a los grupos de enfermedad 4 (más grave) y 1 (menos grave). Los Límites de confianza (LC) se refieren al 95%.

Parámetro	Valor	LC al 95% inferior	LC al 95% superior	Error estándar	p-valor
Constante	-4.58	-9.52	0.36	2.42	0.0678
Granulomas	-0.05	-0.12	0.02	0.03	0.1649
Superficie	0.22	0.04	0.40	0.09	0.0201

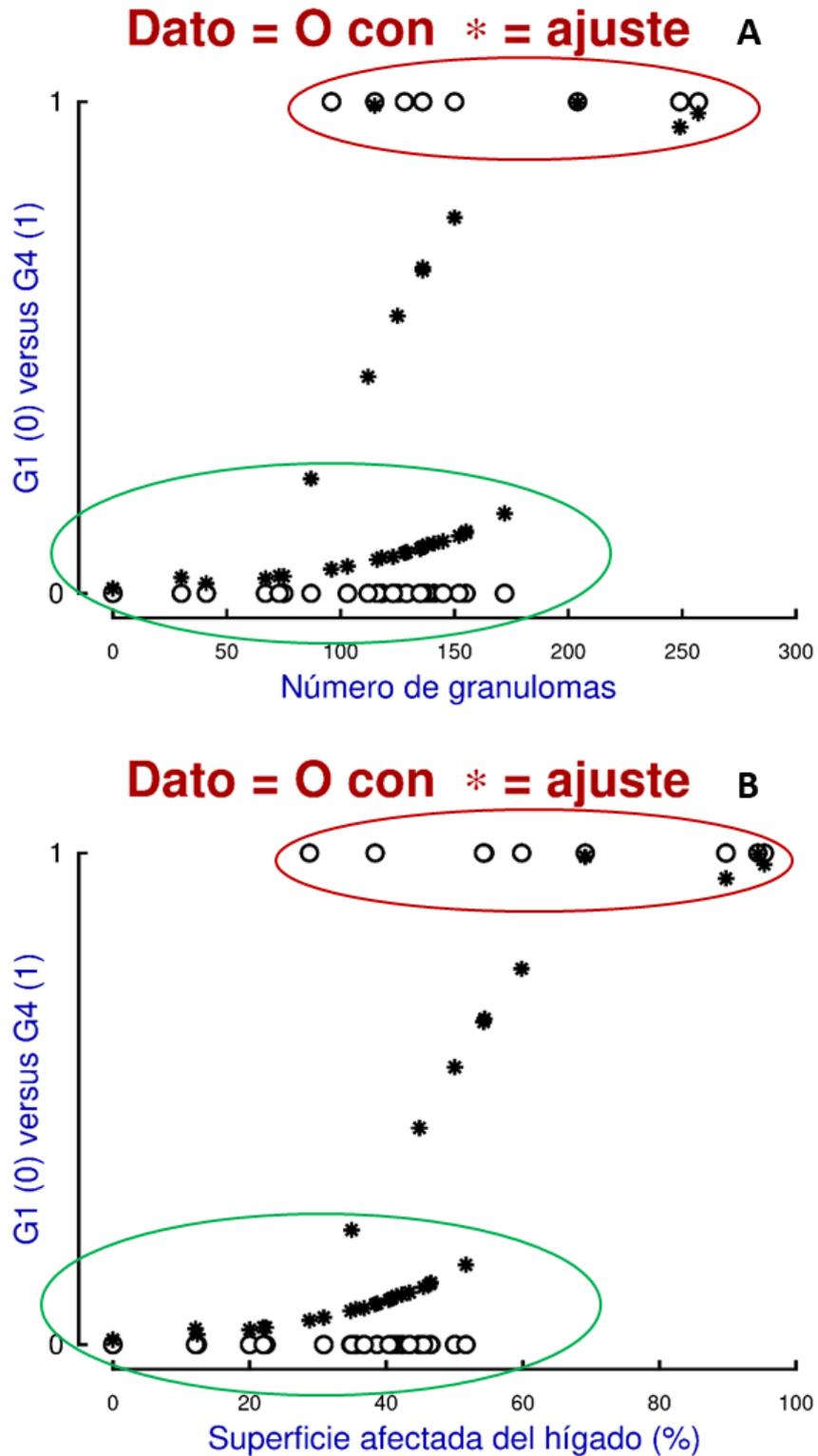
El ajuste se puede aceptar como adecuado, ya que los p-valores de los parámetros podrían considerarse tolerables con un criterio un poco más flexible que p-valores < 0.05, a la vez que los residuales resultaron ser moderados.

Las representaciones gráficas para los ajustes se muestran en la Figura 4.30, donde se pueden observar los datos reales, mostrados como círculos, junto con el valor ajustado por el modelo, mostrado como asterisco. En el panel A se muestra el ajuste del número de granulomas con los Grupos 1 (0) y 4 (1) y en el panel B observamos el ajuste de la superficie afectada del hígado para dichos grupos.

En resumen, la regresión RLB confirmaría la existencia de los dos grupos extremos de enfermedad (1 y 4). Así mismo, la regresión RLB ha generado un modelo que tiene como covariables predictoras significativas el número de granulomas y la superficie afectada del hígado. Este modelo podría tener a su vez fines predictivos, ya que para un nuevo ratón del que se conocieran el número de granulomas y la superficie afectada del hígado, se podría determinar la probabilidad binomial que tendría de pertenecer al grupo más grave ( $p(1)$ ), en base a la ecuación:

$$p(1) = \frac{1}{1 + e^{(-4.58 - 0.05Granulomas + 0.22Superficie)}}$$

Pero la utilidad de estas propiedades predictivas requeriría de ulteriores estudios, con series de entrenamiento y series de prueba, con un tamaño de muestra mayor al actual.



**Figura 4.30.** Regresión logística binaria para validación de la formación de los grupos de ratones, Grupo 1 (0), menos grave vs Grupo 4 (1), más grave. El dato real se muestra con un círculo y el ajustado por el modelo con un asterisco. La elipse de color rojo superior indica los ratones que se ajustan al Grupo 4, y la inferior en verde, enmarca los ratones que se ajustan al Grupo 1. En el centro se encuentran los casos atípicos, que no se ajustan bien a los grupos.

#### 4.6 Estudios de regresión múltiple lineal (RML) de variables de daño hepático en función de los SNPs y de otras variables fenotípicas para el total de ratones

Se consideró que un objetivo de interés, sería el confirmar que los SNPs, detectados en los diferentes cromosomas, estaban asociados a la variable fenotípica propuesta por el genotipado. A su vez, resultaría interesante analizar si, para esa variable fenotípica asociada con el SNP, existían otras covariables adicionales que pudieran también estar asociadas junto con el SNP a la variable fenotípica. Con este propósito, se utilizó la técnica de “Regresión Múltiple Lineal (RML)” que, como es sabido, es una extensión de la regresión lineal simple al caso de más de una covariable. Este método realiza un ajuste de mínimos cuadrados a la ecuación genérica:

$$y = B_0 + B_1x_1 + B_2x_2 + \dots B_mx_m$$

donde  $y$  es la variable respuesta,  $x_i$  son las covariables predictoras y  $B_0$  es un término constante.

##### 4.6.1 Número de granulomas en función de las inmunoglobulinas, de las poblaciones celulares y el SNP rs6169611 del cromosoma 15

Se preparó una matriz para el conjunto de ratones, que incluía los datos del número de granulomas como variable  $y$ , y como variables  $x_i$  los valores de IgM, IgG1, IgG2a, CD4, CD8 y B220 y los datos del genotipado para el SNP rs6169611 del cromosoma 15. Se depuró la matriz anterior eliminando los ratones para los que no existía resultado en el genotipado (NA), obteniéndose una matriz final con 80 ratones.

Debido a que la diferencia de magnitud de las variables  $x_i$  no era excesiva, se realizó la regresión con los datos originales sin transformar, incluidos los datos del SNP, que son datos binarios (0 para los homocigotos y 1 para los heterocigotos).

De las variables iniciales que entran en el modelo, se van “eliminando” paulatinamente las que presentan p-valores elevados, que indican variables no significativas para el modelo. Las variables finales que entraron como significativas en el presente caso fueron IgM, CD8 y el SNP. Los valores de los parámetros obtenidos se muestran en la Tabla 4.8.

**Tabla 4.8.** Modelo RML para el número de granulomas en función del SNP rs6169611 del cromosoma 15 y otras variables significativas.

Parámetro	Valor	LC al 95% inferior	LC al 95% superior	Error estándar	p-valor
Constante	114.17	90.16	138.18	12.05	0.0000
IgM	29.83	-0.46	60.12	15.21	0.0535
CD8	3.52	1.46	5.59	1.04	0.0011
rs6169611	-15.59	-33.55	2.37	9.02	0.0878

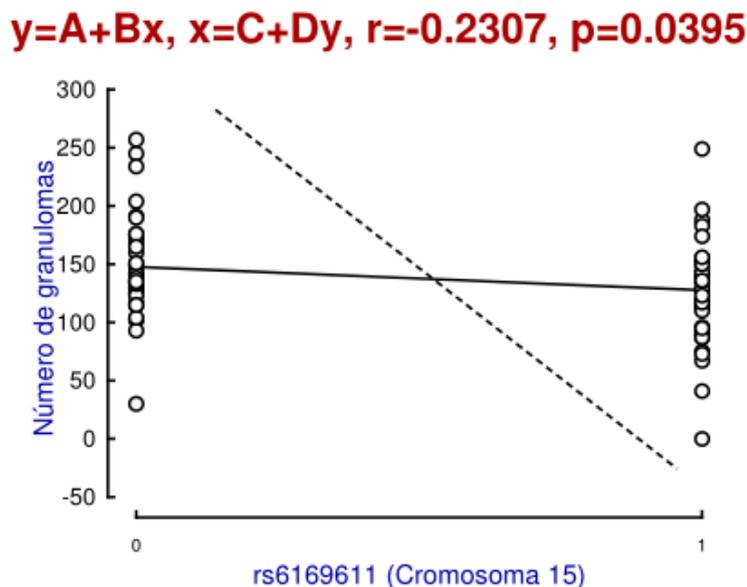
ANOVA					
Tipo	Grados libertad	Suma de cuadrados	Cuadrado medio	F	p-valor
Total	79	151769			
Regresión	3	30973.1	1032.4	6.50	0.0006
Residual	76	120796	1589.4		

Los parámetros para las variables resultan aproximadamente significativos ( $p < 0.1$ ), por lo que se puede admitir este modelo como aceptable. La ecuación final sería:

$$\text{Número de granulomas} = 114 + 29.8 \text{ IgM} + 3.52 \text{ CD8} - 15.6 \text{ rs6169611}$$

A parte del interés explicativo de las variables incluidas en el modelo, esta ecuación también serviría, como en el caso de la RLB, para predecir el número de granulomas a partir de datos obtenidos de IgM, CD8 y rs6169611 en nuevos ratones. Para validar este modelo con fines de predicción, habría que hacer estudios más adelante con un mayor número de ratones.

Dado el interés del SNP, en la Figura 4.31 se muestra la representación de la correlación entre el número de granulomas y el SNP rs6169611 del cromosoma 15. Como era de esperar, por la información aportada por el genotipado, existe una correlación significativa entre el número de granulomas y el SNP rs6169611 del cromosoma 15. Se han omitido las correlaciones con las otras variables en aras de la brevedad.



**Figura 4.31.** Representación de la correlación entre el SNP rs6169611 del cromosoma 15 y el número de granulomas. Se muestra el p-valor ( $p < 0.05$ ) y el coeficiente de correlación de Pearson ( $r$ ). En este caso la correlación es inversa ( $r < 0$ ). Las dos rectas representan la regresión lineal simple de  $y$  vs.  $x$  (línea continua) y de  $x$  vs.  $y$  (línea discontinua).

#### 4.6.2 Número de granulomas en función de las inmunoglobulinas, de las poblaciones celulares y el SNP rs13483259 del cromosoma 18

De nuevo, se preparó una matriz para el conjunto de ratones, que incluía los datos del número de granulomas como variable  $y$ , y como variables  $x_i$  los valores con los datos fenotípicos de IgM, IgG1, IgG2a, CD4, CD8 y B220 y los datos obtenidos del genotipado para el SNP rs13483259 del cromosoma 18. Se depuró la matriz eliminando los ratones para los que no existía resultado en el genotipado y se obtuvo una matriz con 75 ratones.

Como en el caso anterior, dado que la diferencia de magnitud de las variables  $x_i$  no era excesiva, la regresión se llevó a cabo con los datos originales sin transformar, incluidos los datos del SNP, datos binarios (0 para homocigotos y 1 para heterocigotos).

De las variables iniciales que entran en el modelo, se fueron “eliminando” las que presentaban p-valores altos. Las variables finales que entraron como significativas en el presente caso fueron, como en el caso del cromosoma 15, IgM, CD8 y el SNP. Los valores obtenidos se muestran en la Tabla 4.9.

**Tabla 4.9.** Modelo RML para el número de granulomas en función del SNP rs13483259 del cromosoma 18 y otras variables.

Parámetro	Valor	LC al 95% inferior	LC al 95% superior	Error estándar	p-valor
Constante	97.66	75.47	119.84	11.13	0.0000
IgM	27.11	-2.92	57.15	15.06	0.0761
CD8	3.33	1.29	5.38	1.03	0.0018
rs13483259	24.41	6.64	42.17	8.91	0.0078

ANOVA					
Tipo	Grados libertad	Suma de cuadrados	Cuadrado medio	F	p-valor
Total	74	136360			
Regresión	3	32384.4	10794.8	7.37	0.0002
Residual	71	103976	1464.4		

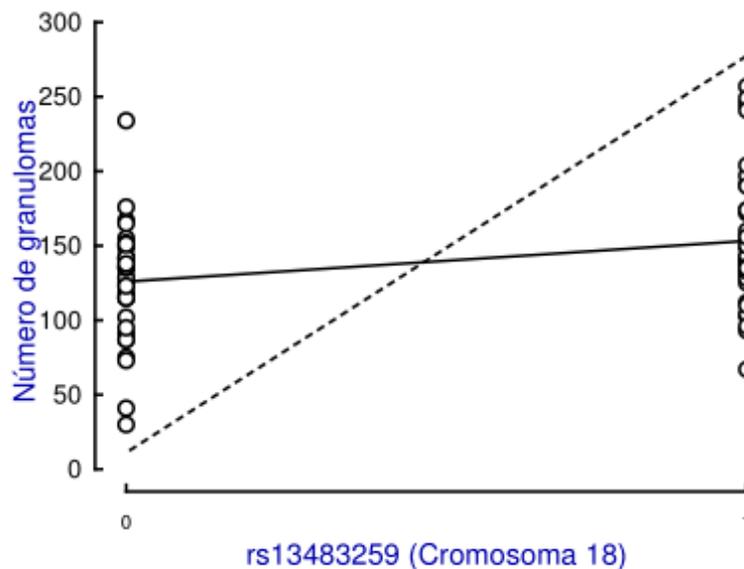
Salvo para IgM, que presenta un p-valor aproximadamente significativo ( $p < 0.1$ ), el resto de p-valores son claramente significativos ( $p < 0.05$ ), por lo que se puede admitir este modelo como aceptable. La ecuación final sería:

$$\text{Número de granulomas} = 97.7 + 27.1 \text{ IgM} + 3.33 \text{ CD8} + 24.4 \text{ rs13483259}$$

Esta ecuación nos ayudaría en estudios posteriores, midiendo los valores de IgM, CD8 y rs13483259 en nuevos ratones, para poder determinar el número de granulomas que presentarían.

De nuevo, debido al interés que presenta el dato del SNP, en la Figura 4.32 se muestra la representación de la correlación entre el número de granulomas y el SNP rs13483259 del cromosoma 18. Como era de esperar, por la relación ya aportada por el genotipado, se observa la correlación significativa entre el número de granulomas y el SNP rs 13483259 del cromosoma 18.

$$y=A+Bx, x=C+Dy, r= 0.3205, p=0.0051$$



**Figura 4.32.** Representación de la correlación entre el SNP rs13483259 del cromosoma 18 y el número de granulomas. Se muestra el p-valor ( $p<0.05$ ) y el coeficiente de correlación de Pearson ( $r$ ). En este caso la correlación es directa ( $r>0$ ). Las dos rectas representan la regresión lineal simple de  $y$  vs.  $x$  (línea continua) y de  $x$  vs.  $y$  (línea discontinua).

#### 4.7 Estudios de regresión múltiple lineal de variables de daño hepático en función de los SNPs y otras variables fenotípicas para los grupos extremos de enfermedad

Un último aspecto a considerar respecto a los SNPs y su relación con otras variables fenotípicas, sería el analizar estas relaciones para los denominados grupos extremos de la enfermedad, el menos grave (Grupo 1) frente al grupo más grave (Grupo 4). Con este objetivo, se exponen a continuación los resultados obtenidos para los dos grupos con las dos variables de daño hepático más importantes, como son el número de granulomas y la superficie afectada del hígado.

##### 4.7.1 Número de granulomas en función de las inmunoglobulinas, de las poblaciones celulares y el SNP rs6169611 del cromosoma 15

###### - Resultados para el Grupo 1 (menos grave)

Se preparó una matriz para los ratones del Grupo 1, que incluía los datos del número de granulomas como variable  $y$ , y como variables  $x_i$  los valores de IgM, IgG1, IgG2a, CD4, CD8 y B220 y los datos del genotipado para el SNP rs6169611 del cromosoma 15. Se

depuró la matriz anterior eliminando los ratones para los que no existía resultado en el genotipado (NA) y se obtuvo una matriz final con 23 ratones. La regresión se llevó a cabo con los datos sin transformar, ya que la diferencia de magnitud entre las variables  $x_i$  no era excesiva.

De manera análoga a lo ya descrito en apartados anteriores, de las variables iniciales, se van eliminando sucesivamente las que presentan p-valores elevados. Las variables finales que se han considerado en el presente caso fueron IgM, CD8 y el SNP. Los valores de los parámetros obtenidos se muestran en la tabla 4.8.

De las variables iniciales que entran en el modelo, se van “eliminando” paulatinamente las que presentan p-valores elevados. Las variables finales que entran en el modelo son CD4 y el SNP y los valores obtenidos se muestran en la Tabla 4.11. Por el interés que tiene el SNP, se ha dejado intencionadamente en el modelo de regresión, pero en este caso el SNP presenta un p-valor muy elevado y no es significativamente distinto de cero, es decir, no interviene en el modelo.

**Tabla 4.11.** Modelo RML para el número de granulomas en función del SNP rs6169611 del cromosoma 15 y otras variables significativas para el Grupo 1.

Parámetro	Valor	LC al 95% inferior	LC al 95% superior	Error estándar	p-valor
Constante	83.15	43.51	122.79	19.00	0.0003
CD4	3.44	0.88	6.00	1.23	0.0110
rs6169611	1.40	-36.38	39.19	18.11	0.9390

ANOVA					
Tipo	Grados libertad	Suma de cuadrados	Cuadrado medio	F	p-valor
Total	22	42515.7			
Regresión	2	12390.0	6195.0	4.11	0.0319
Residual	20	30125.7	1506.3		

En este caso y debido a que el p-valor para el SNP es muy elevado, la ecuación de regresión sólo debe considerar la variable CD4 y sería:

$$\text{Número de granulomas} = 83.2 + 3.44 \text{ CD4}$$

- **Resultados para el Grupo 4 (más grave)**

Se preparó una matriz para los ratones del Grupo 4, que incluía los datos del número de granulomas como variable  $y$ , y como variables  $x_i$  los valores de IgM, IgG1, IgG2a, CD4, CD8 y B220 y los datos del genotipado para el SNP rs6169611 del cromosoma 15. No hubo que depurar la matriz original, puesto que no había ratones para los que no existiera resultado en el genotipado (NA) y se obtuvo una matriz final con 9 ratones. La regresión se llevó a cabo como en estudios anteriores con los datos sin transformar. De nuevo, se fueron eliminando aquellas variables iniciales que tenían p-valores elevados y no eran significativas. Las variables finales que entran en el modelo son IgM, IgG1, IgG2a y el SNP y los valores obtenidos se muestran en la Tabla 4.12.

**Tabla 4.12.** Modelo RML para el número de granulomas en función del SNP rs6169611 del cromosoma 15 y otras variables significativas para el Grupo 4.

Parámetro	Valor	LC al 95% inferior	LC al 95% superior	Error estándar	p-valor
Constante	350.89	299.48	402.30	18.52	0.0000
IgM	110.01	82.73	137.28	9.82	0.0004
IgG1	-226.42	-274.57	-178.27	17.34	0.0002
IgG2a	23.37	-3.38	50.12	9.64	0.0723
rs6169611	-172.97	-201.67	-144.28	10.34	0.0001

ANOVA					
Tipo	Grados libertad	Suma de cuadrados	Cuadrado medio	F	p-valor
Total	8	27556.2			
Regresión	4	27338.3	6834.58	125.46	0.0002
Residual	4	217.91	54.48		

Para este Grupo 4, que es el más grave en cuanto a enfermedad, han resultado significativas un mayor número de variables que con el Grupo 1 anterior (menor gravedad), lo cual era de esperar, ya que a un mayor grado de infección, un mayor número de variables debiera verse afectado. La ecuación de regresión ahora sería:

$$\text{Número de granulomas} = 351 + 110 \text{ IgM} - 226 \text{ IgG1} + 23.4 \text{ IgG2a} - 173 \text{ rs6169611}$$

#### 4.7.2 Número de granulomas en función de las inmunoglobulinas, de las poblaciones celulares y el SNP rs13483259 del cromosoma 18

##### - Resultados para el Grupo 1 (menos grave)

Siguiendo el procedimiento aplicado en los casos anteriores, se preparó la matriz para el Grupo 1, con los datos del número de granulomas como variable  $y$ , y como variables  $x_i$  los valores de IgM, IgG1, IgG2a, CD4, CD8 y B220 y del SNP rs13483259 del cromosoma 18. Se depuraron los datos, obteniéndose una matriz con 21 ratones. El estudio se llevó a cabo sin transformar los datos. Después de ir eliminando sucesivamente las variables que presentaban p-valores elevados, las variables que entran al modelo son CD4 y el SNP y los valores obtenidos se muestran en la Tabla 4.13. De nuevo, como en el Grupo 1 del estudio anterior, el SNP se mantiene en el modelo de regresión pese a mostrar un p-valor elevado, por lo que tampoco interviene en el modelo.

**Tabla 4.13.** Modelo RML para el número de granulomas en función del SNP rs13483259 del cromosoma 18 y otras variables significativas para el Grupo 1.

Parámetro	Valor	LC al 95% inferior	LC al 95% superior	Error estándar	p-valor
Constante	86.25	60.27	112.22	12.36	0.0000
CD4	2.79	0.47	5.12	1.11	0.0213
rs13483259	14.77	-17.52	47.07	15.37	0.3493

ANOVA					
Tipo	Grados libertad	Suma de cuadrados	Cuadrado medio	F	p-valor
Total	20	29962.7			
Regresión	2	9352.5	4676.3	4.08	0.0345
Residual	18	20610.2	1145.0		

La ecuación de regresión sólo considera la variable CD4, puesto que el p-valor del SNP es elevado.

$$\text{Número de granulomas} = 86.3 + 2.79 \text{ CD4}$$

**- Resultados del Grupo 4 (más grave)**

Se preparó la matriz para los ratones del Grupo 4, que incluía los datos del número de granulomas como variable  $y$ , y como variables  $x_i$  los valores de IgM, IgG1, IgG2a, CD4, CD8 y B220 y los datos del genotipado para el SNP rs13483259 del cromosoma 18. No hubo que depurar la matriz original, puesto que no había ratones para los que no existiera resultado en el genotipado (NA) y se obtuvo una matriz final con 9 ratones. La regresión se llevó a cabo como en estudios anteriores con los datos sin transformar. De nuevo, se fueron eliminando aquellas variables iniciales que tenían p-valores elevados y no eran significativas. Las variables finales que entran en el modelo son IgM, IgG1, CD4, CD8 y el SNP y los valores obtenidos se muestran en la Tabla 4.14. En este caso, pese a que la constante y la IgG1 presentan valores ligeramente superiores a 0.05, se mantienen en el modelo.

**Tabla 4.14.** Modelo RML para el número de granulomas en función del SNP rs13483259 del cromosoma 18 y otras variables significativas para el Grupo 4.

Parámetro	Valor	LC al 95% inferior	LC al 95% superior	Error estándar	p-valor
Constante	-461.67	-937.62	14.27	149.55	0.0538
IgM	590.87	140.59	1041.14	14.49	0.0250
IgG1	136.59	-13.42	286.60	47.14	0.0626
CD4	78.83	8.89	148.77	21.98	0.0371
CD8	-120.93	-230.84	-11.01	34.54	0.0394
rs13483259	239.66	52.17	427.15	58.91	0.0268

ANOVA					
Tipo	Grados libertad	Suma de cuadrados	Cuadrado medio	F	p-valor
Total	8	27556.2			
Regresión	5	25229.6	5045.9	6.51	0.0771
Residual	3	2326.6	775.5		

En este caso sólo eliminamos las variables IgG2a y B220, con lo que la ecuación sería:

$$\text{Número de granulomas} = -462 + 591 \text{ IgM} + 137 \text{ IgG1} + 78.8 \text{ CD4} - 121 \text{ CD8} + 240 \text{ rs13483259}$$

## 5 Discusión

Los individuos de todas las especies presentan diferencias que afectan a nivel anatómico, fisiológico, del desarrollo, de la conducta y de la susceptibilidad o resistencia a diferentes enfermedades. La mayoría de estas diferencias no responden a una herencia mendeliana dominante o recesiva de un sólo gen, sino que están bajo el control de múltiples genes, definidos como rasgos complejos (Abiola et al., 2003). En principio, la susceptibilidad a las enfermedades infecciosas responde a un modelo de rasgos complejos (Lander y Kruglyak, 1995), pero la problemática de este modelo reside en que un mismo genotipo puede determinar diferentes fenotipos, ya que los rasgos complejos presentan una incompleta penetrancia y están sujetos a efectos del azar y del medio ambiente.

Existen publicaciones donde se estudia la infección experimental de *Schistosoma mansoni* en distintas cepas de ratón, en su mayoría realizados en los años 80 (Cheever et al., 1983; Dean et al., 1981; Cheever et al., 1987). Estos trabajos se centran en el análisis de magnitudes histopatológicas, principalmente granulomas y fibrosis hepática, y no están orientados a la identificación de marcadores genéticos implicados en la susceptibilidad o resistencia a la infección. Por ello, en trabajos previos en nuestro grupo de investigación (Pérez del Villar et al., 2011), se caracterizó la cepa más susceptible y más resistente a la infección por *S. mansoni* entre cinco cepas parentales singénicas estudiadas (C57BL/6J, FVB/NH, DBA/2J, BALB/c y CBA/2J). En el análisis fenotípico de las magnitudes parasitológicas y patológicas, asociadas a la resistencia, la cepa C57BL/6J, presentó mayor número de vermes y mayor número de huevos por gramo de hígado. Sin embargo, la cepa CBA/2J presentó características asociadas a la susceptibilidad, con menor número de vermes y de huevos por gramo de hígado. En un trabajo posterior, Pérez del Villar *et al* (2014) estudiaron la susceptibilidad/resistencia en el cruce F1B6CBA, entre la cepa más susceptible (CBA/2J) y la más resistente (C57BL/6J) del estudio anterior (Pérez del Villar et al., 2014). La cohorte F1B6CBA, presentó mayor número de vermes que las cepas parentales y datos intermedios en cuanto al número de huevos por gramo de hígado. Sin embargo, se observaba un menor número de granulomas, asemejándose a las hembras de la cepa parental más resistente.

Siguiendo con la secuencia experimental y como primer objetivo de esta Tesis Doctoral, se procedió a realizar un *backcross* entre la cepa híbrida F1B6CBA y la cepa parental susceptible a la infección por *S. mansoni*, CBA/2J, con la finalidad de generar diversidad genética y con el último fin de identificar regiones en el genoma susceptibles a la infección por *S. mansoni*. Se analizaron magnitudes parasitológicas como el número de vermes totales y separados por sexos, el número de huevos por gramo de hígado e intestino y la fecundidad, definida como el cociente entre el número de huevos en tejidos y el número de vermes hembra recogidos. Estas magnitudes indican, en principio, el grado de infección. Además, se estudian magnitudes denominadas “patológicas”, como el número de granulomas en superficie hepática, así como la superficie afectada del hígado, indicadoras de lesión hepática y por tanto del nivel de gravedad.

La respuesta inmunológica contra la infección por *S. mansoni* es muy compleja (Colley y Secor, 2014), ya que participan diferentes tipos de células del sistema inmune, orquestadas por moléculas conectoras o efectoras (inmunoglobulinas), que varían respecto a la etapa de la enfermedad (aguda/crónica) o la fase del ciclo biológico (cercaria, esquistosómula, adultos). En este trabajo se han estudiado poblaciones celulares en sangre periférica y esplenocitos mediante citometría de flujo (CD45, CD4, CD8, B220), así como diferentes isotipos y subtipos de anticuerpos específicos contra *S. mansoni* (IgM, IgG, IgG1 e IgG2a).

Los datos obtenidos en ratones F2BX muestran menor número de vermes, tanto machos como hembras, y menor número de huevos por gramo de hígado que los observados en ratones F1B6CBA. Sin embargo, presentan mayor número de granulomas, tendiendo más hacia el comportamiento observado con la cepa parental más susceptible a la infección, datos esperados “a priori” cuando se genera el *backcross* con el cruce de la cepa parental más susceptible a la infección por el parásito. Si comparamos entre F2BX machos y F2BX hembras, observamos que no existen diferencias entre las medias de ninguna de las variables parasitológicas o patológicas estudiadas, con la excepción del número de granulomas. Esto podría indicar una asociación entre la patología ocasionada por *S. mansoni* y el sexo del individuo. En su conjunto, los datos obtenidos muestran una gran variabilidad, con valores extremos en los límites superior e inferior, que es lo que se pretendía conseguir con la generación de la cohorte cruzada F2BX.

En segundo lugar, se realizó un estudio de correlaciones entre 21 magnitudes fenotípicas, descartándose ocho de las 29 analizadas en el estudio, por resultar poco indicativas. Se excluyeron las variables inmunológicas como la detección de anticuerpos específicos medidas en la semana previa a la infección y las poblaciones celulares analizadas en la cuarta semana post-infección. Por tanto, todas las magnitudes de respuesta inmune se estudiaron en la novena semana post-infección. Las magnitudes parasitológicas presentan una correlación directa entre ellas, con la excepción del número de vermes y la fecundidad, cuya relación es inversa. Esta situación, a mayor número de vermes hembras existe menor producción de huevos, se como *crowding effect*, fenómeno que ha sido observado en diferentes helmintosis (Fong et al., 2018). Es evidente que las diferentes poblaciones celulares estudiadas guardan una estrecha correlación en sangre periférica, como se podía esperar. Sin embargo, en esplenocitos no existe correlación entre B220 y CD45, marcadores de linfocitos B y T respectivamente. También hay que señalar la ausencia de correlación entre IgG2a con IgG e IgM. Este dato podría explicarse por la generación de diferentes respuestas Th1/Th2 que se producen en esta infección (Anthony et al., 2012).

Es interesante estudiar las correlaciones existentes entre las magnitudes que traducen infección/enfermedad y la respuesta inmunológica generada. En este sentido existen correlaciones positivas entre poblaciones celulares y marcadores de daño hepático, sin que exista esta relación cuando se compara con el número de vermes. Además, cabe destacar que a mayor número de células B220 menor número de huevos tanto en hígado como en intestino, y también menor fecundidad de las hembras. Este dato no se corresponde cuando se analizan los datos de los diferentes isotipos y subtipos de inmunoglobulinas con las magnitudes parásito-patológicas estudiadas.

Uno de los objetivos prioritarios de esta Tesis Doctoral era definir grupos o *clusters* de animales con diferentes niveles de enfermedad atendiendo a las variables fenotípicas analizadas, específicamente las denominadas parásito-patológicas. Empleamos tres técnicas de análisis multivariante. Mediante el “Análisis de conglomerados” analizamos la distancia entre casos, con representación en forma de árbol, dendrograma (Pérez-López, 2005). Cuando empleamos la técnica de “Análisis de componentes principales”, se obtiene una representación en el plano, con una combinación lineal de las variables originales (Armitage et al., 2002). Con el método de “Análisis de conglomerados por k-medias” se

asignan los datos a  $K$  clusters, basado en el algoritmo de SIMFIT (Bardsley, 2017). Los resultados obtenidos tras el análisis con estas técnicas presumen la formación de cuatro subgrupos diferentes. El Grupo 1 sería aquel en el que los ratones presentan la enfermedad con menor gravedad, mientras que el Grupo 4 correspondería a los ratones que muestran la enfermedad con mayor gravedad. Finalmente, los Grupos 2 y 3 reunirían a los ratones en los que la enfermedad mostró un comportamiento intermedio. Hay que considerar que los individuos se han agrupado teniendo en cuenta todos los patofenotipos en su conjunto, lo cual no significa que todos los patofenotipos hayan quedado perfectamente separados, dada la gran variabilidad existente entre los ratones y su respuesta a la infección.

Analizando las diferencias encontradas entre los cuatro grupos definidos, se observa que existe una tendencia ascendente respecto a la afectación hepática y al número de huevos en tejidos. El Grupo 1 presenta los menores índices de lesión tanto en hígado como en intestino mientras que en el Grupo 4 se observan los mayores grados de afectación, indicando altos niveles de gravedad. Llama la atención que entre el grupo más leve (Grupo 1) y el más grave (Grupo 4) no existen diferencias significativas respecto al número de vermes, tanto totales como separados por sexo, aunque como hemos indicado anteriormente el nivel de lesión hepática es totalmente diferente. Este hecho indicaría que el número de vermes no está relacionado directamente con las lesiones histopatológicas derivadas de la infección por *S. mansoni*. Una vez más, observamos que el número de vermes no guarda relación con los efectos patológicos desencadenados por la puesta de huevos, fruto de la fecundidad de las hembras, y como consecuencia la producción de una reacción inflamatoria alrededor de ellos, que origina la formación de granulomas. Por todo ello, no puede considerarse el número de vermes totales una magnitud implicada en la gravedad de la enfermedad producida por este parásito.

Múltiples estudios llevados a cabo en ratón muestran que los linfocitos CD4 son esenciales para la formación del granuloma y para orquestar la respuesta inmune frente a *S. mansoni* (Pearce y MacDonald, 2002) (Wilson et al., 2007). Además, se postula que los linfocitos CD8 están implicados en la modulación de la respuesta inmune Th2, característica de la fase crónica de la enfermedad, minimizando el daño inmunológico asociado a la formación del granuloma (Pedras-Vasconcelos y Pearce, 1996). En nuestro estudio observamos que las poblaciones celulares CD4 y CD8 son predominantes en los grupos de gravedad intermedia (Grupos 2 y 3) tanto en sangre periférica como en

esplenocitos, disminuyendo drásticamente en el grupo más grave (Grupo 4). Estos resultados coinciden con los datos obtenidos previamente en nuestro laboratorio tanto con cepas parentales como con ratones F1B6CBA (Pérez del Villar et al., 2014). Similar patrón es observado en los diferentes grupos respecto al estudio de las poblaciones celulares B220. Este resultado está en consonancia con lo observado en estudios anteriores, donde se muestra un alto porcentaje de B220 en la cepa parental C57BL/6J, resistente a la infección por *S. mansoni*, y un bajo porcentaje (como en el caso que nos ocupa) de células B220 en la cepa parental CBA/2J (Pérez del Villar et al., 2011; Pérez del Villar et al., 2014). Estos resultados, en su conjunto, indican la importancia de estas poblaciones celulares en la progresión de la enfermedad.

En cuanto a la detección específica de anticuerpos anti-*S. mansoni*, observamos que existen bajos niveles de detección de los isotipos IgG e IgM en el grupo de menor gravedad (Grupo 1). Estos niveles van aumentando progresivamente, hasta alcanzar los más elevados en el grupo de afectación más grave (Grupo 4). Se conoce desde hace años que diferentes subtipos de inmunoglobulinas son indicadores indirectos de respuestas inmunológicas. Así, niveles altos de IgG1 están asociados a respuestas Th2 (Stevens et al., 1988), mientras que elevados niveles de IgG2a se relacionan con respuestas Th1 (Lawrence et al., 1994). Nuestros resultados indican que no existen diferencias entre los dos subtipos estudiados en los grupos intermedios (Grupos 2 y 3), mientras que en el grupo más grave (grupo 4), se detectan menores niveles de IgG2a. Estos resultados indican que los altos niveles de anticuerpos relacionados con respuestas Th2, no están relacionados con la gravedad de la infección.

Una vez obtenidas las distancias de recombinación en la cohorte F2BX, se identificaron las regiones eQTL ligadas a los patofenotipos. Sin embargo, los patofenotipos son complejos y manifiestan el efecto de distintos fenotipos intermedios, que actúan a diferentes niveles de organización biológica. Por este motivo, los resultados de este análisis no recogieron todos los determinantes genéticos que aportaron a la variabilidad en los patofenotipos. El análisis de ligamiento sobre cada fenotipo sólo fue capaz de revelar aquellas regiones que contribuyeron al valor de éste con la fuerza suficiente. Los determinantes genéticos de la variabilidad en los fenotipos intermedios, que no mostraron asociación con la fuerza suficiente con los patofenotipos, no podrían ser encontrados por

este análisis, y por este motivo serían una de las causas de la heredabilidad perdida (Blanco-Gómez et al., 2016).

Se procedió a realizar diferentes controles de calidad basados en el estudio de la frecuencia de cada marcador, en la detección de errores de genotipado, en la cuantificación de sobrecruzamientos y en la generación de un mapa genético adecuado. Iniciamos el estudio con 92 ratones, 961 SNPs y 30 fenotipos y tras los cuatro controles de calidad establecidos finalizamos con 86 animales, 958 SNPs y 30 fenotipos. Hay que considerar que en este trabajo se estudian un número importante de SNPs que triplica lo analizado en otras publicaciones (Castellanos-Martín et al., 2015).

Se identificaron 18 regiones cromosómicas asociadas a las diferentes magnitudes fenotípicas estudiadas con *LOD-score* por encima de 1.4. Respecto a las variables relacionadas con el daño hepático (granulomas y superficie afectada) encontramos asociación con cuatro eQTL en cuatro cromosomas (2, 15, 17 y 18). Llama la atención que la región eQTL2 en el mismo marcador, está asociada tanto con la presencia de granulomas como con la afectación de la superficie hepática. Además, la región eQTL1 también está asociada con las dos variables fenotípicas, pero en distinta localización cromosómica.

Los vermes totales, así como los vermes macho y hembra se encuentran relacionados con tres eQTL, localizados en los cromosomas 6, 7 y X. Respecto al cromosoma X, los homocigotos machos y hembras presentan mayor número de vermes que los heterocigotos de ambos sexos. Sin embargo, no existen diferencias en el número de vermes entre ratones F2BX machos y hembras. Estos datos no son muy concluyentes y es difícil extraer conclusiones veraces. Finalmente, en cuanto a las variables parasitológicas estudiadas, observamos la asociación de eQTL8 del cromosoma 8 con los huevos en hígado e intestino en el mismo marcador.

Cuando se analizan las eQTL relacionadas con las poblaciones celulares estudiadas, se observa que no existe ninguna coincidencia entre las obtenidas en sangre periférica y en esplenocitos. Por el contrario, existe coincidencia en la eQTL4, localizada en el cromosoma 17 y en el mismo marcador, relacionada tanto con la detección de IgG e IgM como con la afectación de la superficie hepática.

A continuación, analizamos la sintenia *in silico* mediante *Ensembl* para estudiar la homología entre ratón y ser humano. Elegimos para este estudio las eQTL relacionadas con el daño hepático que corresponden a eQTL1 localizada en el cromosoma 15 con SNP

rs6169611 y eQTL2 localizada en el cromosoma 18 con SNP rs13483259. Respecto a eQTL1, a pesar de que existe sintenia con el genoma humano, no hay representación de regiones homólogas. Por tanto, no podemos extrapolar regiones similares entre las dos especies. En cuanto a eQTL2, existen regiones homólogas en el genoma humano, localizadas también en el cromosoma 18 humano en la región 18q11.2-12.3. Estudios previos en humanos han demostrado asociación de ciertas regiones cromosómicas localizadas en los cromosomas 1, 5 y 6 (Mangano y Modiano, 2014). Hay que destacar que no todas las regiones estudiadas estaban asociadas con la susceptibilidad a la infección por *S. mansoni*, a excepción del locus SM2 (6q22-23) relacionado con la fibrosis hepática. Por todo ello, se deduce que la región identificada en este estudio en modelo experimental *backcross* y con homología en el genoma humano no había sido descrita en estudios previos.

Por último, analizamos la asociación eQTL con los cuatro patofenotipos de gravedad. Encontramos que el Grupo 4 (mayor gravedad) muestra un elevado porcentaje de ratones heterocigóticos para la mayor parte de los SNPs que mostraban ligación. Además, utilizando técnicas de análisis multivariante por regresión logística binaria y múltiple lineal fuimos capaces de construir modelos predictivos de susceptibilidad a la infección por *S. mansoni*. De esta manera, se han obtenido ecuaciones que predicen la susceptibilidad mediante la combinación de variables fenotípicas y genotípicas.

## 6 Conclusiones

1.- Se ha generado una cohorte F2BX mediante *backcross*, con una gran heterogeneidad, útil para estudiar la influencia genética en la susceptibilidad a la infección producida por *Schistosoma mansoni*.

2.- Se han definido cuatro patofenotipos de enfermedad en función del daño hepático producido e independientemente del número de vermes totales recuperados.

3.- Se ha comprobado la importancia de las poblaciones celulares en la progresión de la enfermedad y se ha observado que elevados niveles de anticuerpos relacionados con respuestas Th2, no están asociados con la gravedad de la infección.

4.- De las 18 eQTL encontradas inicialmente, se ha identificado una región localizada en el cromosoma 18 de ratón relacionada directamente con el daño hepático, que presenta sintenia con una región homóloga en el cromosoma 18 humano, siendo la primera descripción realizada en este tipo de estudios.

## 7 Bibliografía

- Abán JL, Ramajo V, Arellano JL, Oleaga A, Hillyer GV, Muro A. A fatty acid binding protein from *Fasciola hepatica* induced protection in C57/BL mice from challenge infection with *Schistosoma bovis*. *Vet Parasitol.* 1999 Jun 15;83(2):107-21. PMID: 10392967
- Abath FG, Morais CN, Montenegro CE, Wynn TA, Montenegro SM *Trends Parasitol.* 2006 Feb;22(2):85-91. PMID: 16380294.
- Abiola O, Angel JM, Avner P, Bachmanov AA, Belknap JK, Bennett B, Blankenhorn EP et al. The nature and identification of quantitative trait loci: a community's view. *Nat Rev Genet.* 2003 Nov;4(11):911-6. PMID: 14634638
- Abou-Zeid AH, Abkar TA, Mohamed RO. Schistosomiasis and soil-transmitted helminths among an adult population in a war affected area, Southern Kordofan state, Sudan. *Parasit Vectors.* 2012 Jul 3;5:133. doi: 10.1186/1756-3305-5-133. PMID: 22759923
- Adema CM, Hillier LW, Jones CS et al. Whole genome analysis of a schistosomiasis-transmitting freshwater snail. *Nat Commun.* 2017 May 16;8:15451. doi: 10.1038/ncomms15451. Aug 23;8:16153. PMID: 28508897
- Alarcón de Noya B, Ruiz R, Losada S, Colmenares C, Contreras R, Cesari IM, Noya O. Detection of schistosomiasis cases in low-transmission areas based on coprologic and serologic criteria The Venezuelan experience. *Acta Trop.* 2007 Jul;103(1):41-9. Epub 2007 May 18. PMID: 17606217
- Almeida GT, Amaral MS, Beckedorff FC, Kitajima JP, DeMarco R, Verjovski-Almeida S. Exploring the *Schistosoma mansoni* adult male transcriptome using RNA-seq. *Exp Parasitol.* 2012 Sep;132(1):22-31. doi: 10.1016/j.exppara.2011.06.010. PMID: 21745473
- Anderson L, Amaral MS, Beckedorff F, Silva LF, Dazzani B, Oliveira KC, Almeida GT, Gomes MR, Pires DS, Setubal JC, DeMarco R, Verjovski-Almeida S. *Schistosoma mansoni* Egg, Adult Male and Female Comparative Gene Expression Analysis and Identification of Novel Genes by RNA-Seq. *PLoS Negl Trop Dis.* 2015 Dec 31;9(12):e0004334. doi: 10.1371/journal.pntd.0004334. PMID: 26719891
- Anthony BJ, Ramm GA, McManus DP. Role of resident liver cells in the pathogenesis of schistosomiasis. *Trends Parasitol.* 2012 Dec;28(12):572-9. doi: 10.1016/j.pt.2012.09.005. PMID: 23099112
- Armitage P., Berry G. y Matthews J. N. S. *Statistical Methods in Medical Research.* 4<sup>a</sup> Ed. Blackwell Science. Oxford, U.K. 2002.
- Barbosa CS, Araújo KC, Sevilla MA, Melo F, Gomes EC, Souza-Santos R. Current epidemiological status of schistosomiasis in the state of Pernambuco, Brazil. *Mem Inst Oswaldo Cruz.* 2010 Jul;105(4):549-54. PMID: 20721507
- Bardsley, W.G. (2017), Manchester Univeristy. Simfit statistical package. Free download in: <https://simfit.org.uk/>
- Barsoum R. The changing face of schistosomal glomerulopathy. *Kidney Int.* 2004 Dec;66(6):2472-84. PMID: 15569345

- Berriman M, Haas BJ, LoVerde PT, Wilson RA, Dillon GP, Cerqueira GC, et al. The genome of the blood fluke *Schistosoma mansoni*. *Nature*. 2009 Jul 16;460(7253):352-8. doi: 10.1038/nature08160. PMID: 19606141
- Berry A, Moné H, Iriart X, Mouahid G, Aboo O, Boissier J, Fillaux J, et al. Schistosomiasis haematobium, Corsica, France. *Emerg Infect Dis*. 2014 Sep;20(9):1595-7. doi: 10.3201/eid2009.140928. PMID: 25153697
- Bethlem EP, Schettino Gde P, Carvalho CR. Pulmonary schistosomiasis. *Curr Opin Pulm Med*. 1997 Sep;3(5):361-5. PMID: 9331538
- Bica I, Hamer DH, Stadecker MJ. Hepatic schistosomiasis. *Infect Dis Clin North Am*. 2000 Sep;14(3):583-604, viii. PMID: 10987111
- Blanco-Gómez A, Castillo-Lluva S, Del Mar Sáez-Freire M, Hontecillas-Prieto L, Mao JH, Castellanos-Martín A, Pérez-Losada J. Missing heritability of complex diseases: Enlightenment by genetic variants from intermediate phenotypes. *Bioessays*. 2016 Jul;38(7):664-73. doi: 10.1002/bies.201600084. PMID: 27241833
- Blanton RE, Salam EA, Ehsan A, King CH, Goddard KA. Schistosomal hepatic fibrosis and the interferon gamma receptor: a linkage analysis using single-nucleotide polymorphic markers. *Eur J Hum Genet*. 2005 May;13(5):660-8. PMID: 15756299
- Boissier J, Grech-Angelini S, Webster BL, Allienne JF, Huyse T, Mas-Coma S, et al. Outbreak of urogenital schistosomiasis in Corsica (France): an epidemiological case study. *Lancet Infect Dis*. 2016 Aug;16(8):971-9. doi: 10.1016/S1473-3099(16)00175-4. PMID: 27197551
- Braschi S, Wilson RA. Proteins exposed at the adult schistosome surface revealed by biotinylation. *Mol Cell Proteomics*. 2006 Feb;5(2):347-56. PMID: 16269422
- Broman KW, Wu H, Sen S, Churchill GA. R/qtl: QTL mapping in experimental crosses. *Bioinformatics*. 2003 May 1;19(7):889-90. PMID: 12724300
- Brunet LR, Finkelman FD, Cheever AW, Kopf MA, Pearce EJ. IL-4 protects against TNF-alpha-mediated cachexia and death during acute schistosomiasis. *J Immunol*. 1997 Jul 15;159(2):777-85. PMID: 9218595
- Burke ML, Jones MK, Gobert GN, Li YS, Ellis MK, McManus DP. Immunopathogenesis of human schistosomiasis. *Parasite Immunol*. 2009 Apr;31(4):163-76. doi: 10.1111/j.1365-3024.2009.01098.x. PMID: 19292768,
- Burlandy-Soares LC, de Souza Dias LC, Kanamura HY, de Oliveira EJ, Ciaravolo RM. Schistosomiasis mansoni: follow-up of control program based on parasitologic and serologic methods in a Brazilian community of low endemicity. *Mem Inst Oswaldo Cruz*. 2003 Sep;98(6):853-9. PMID: 14595468
- Bustinduy AL, Thomas CL, Fiutem JJ, Parraga IM, Mungai PL, Muchiri EM, Mutuku F, Kitron U, King CH. Measuring fitness of Kenyan children with polyparasitic infections using the 20-meter shuttle run test as a morbidity metric. *PLoS Negl Trop Dis*. 2011 Jul;5(7):e1213. doi: 10.1371/journal.pntd.0001213. PMID: 21750742
- Caldas IR, Campi-Azevedo AC, Oliveira LF, Silveira AM, Oliveira RC, Gazzinelli G. Human schistosomiasis mansoni: immune responses during acute and chronic phases of the infection. *Acta Trop*. 2008 Nov-Dec;108(2-3):109-17. doi: 10.1016/j.actatropica.2008.05.027. PMID: 18577364
- Campino S, Kwiatkowski D, Dessein A. Mendelian and complex genetics of susceptibility and resistance to parasitic infections. *Semin Immunol*. 2006 Dec;18(6):411-22. PMID: 17023176

- Carranza-Rodriguez C, Pardo-Lledias J, Muro-Alvarez A, Pérez-Arellano JL. Cryptic parasite infection in recent West African immigrants with relative eosinophilia. *Clin Infect Dis*. 2008 Mar 15;46(6):e48-50. doi: 10.1086/528865. PMID: 18260784
- Carvalho GB, Pacífico LG, Pimenta DL, Siqueira LM, Teixeira-Carvalho A, Coelho PM, Pinheiro Cda S, Fujiwara RT, Oliveira SC, Fonseca CT. Evaluation of the use of C-terminal part of the *Schistosoma mansoni* 200kDa tegumental protein in schistosomiasis diagnosis and vaccine formulation. *Exp Parasitol*. 2014b Apr;139:24-32. doi: 10.1016/j.exppara.2014.02.003. PMID: 24560833
- Carvalho OS, Jannotti-Passos LK, Caldeira RL. Importância epidemiológica e biologia molecular aplicada ao estudo dos moluscos do gênero *Biomphalaria*. In OS Carvalho, PMZ Coelho, HL Lenzi. *Schistosoma mansoni* e esquistossomose: uma visão multidisciplinar. Fiocruz. Rio de Janeiro. 2008.
- Cass CL, Johnson JR, Califf LL, Xu T, Hernandez HJ, Stadecker MJ, Yates JR 3rd, Williams DL. Proteomic analysis of *Schistosoma mansoni* egg secretions. *Mol Biochem Parasitol*. 2007 Oct;155(2):84-93. PMID: 17644200
- Castellanos-Martín A, Castillo-Lluva S, Sáez-Freire Mdel M, Blanco-Gómez A, Hontecillas-Prieto L et al. Unraveling heterogeneous susceptibility and the evolution of breast cancer using a systems biology approach. *Genome Biol*. 2015 Feb 21;16:40. doi: 10.1186/s13059-015-0599-z. PMID: 2585329
- Chabasse D, Bertrand G, Leroux JP, Gauthey N, Hocquet P. Developmental bilharziasis caused by *Schistosoma mansoni* discovered 37 years after infestation. *Bull Soc Pathol Exot Filiales*. 1985;78(5):643-7. PMID: 393663
- Cheever AW, Dunn MA, Dean DA, Duvall RH. Differences in hepatic fibrosis in ICR, C3H, and C57BL/6 mice infected with *Schistosoma mansoni*. *Am J Trop Med Hyg*. 1983 Nov;32(6):1364-9. PMID: 6650737
- Cheever AW, Duvall RH, Hallack TA Jr, Minker RG, Malley JD, Malley KG. Variation of hepatic fibrosis and granuloma size among mouse strains infected with *Schistosoma mansoni*. *Am J Trop Med Hyg*. 1987 Jul;37(1):85-97. PMID: 3111283
- Cheever AW, Xu YH, Sher A, Macedonia JG. Analysis of egg granuloma formation in *Schistosoma japonicum*-infected mice treated with antibodies to interleukin-5 and gamma interferon. *Infect Immun*. 1991 Nov;59(11):4071-4. PMID: 1834572
- Cheng G, Luo R, Hu C, Cao J, Jin Y. Deep sequencing-based identification of pathogen-specific microRNAs in the plasma of rabbits infected with *Schistosoma japonicum*. *Parasitology*. 2013 Dec;140(14):1751-61. doi: 10.1017/S0031182013000917. PMID: 23942009
- Chevillard C, Moukoko CE, Elwali NE, Bream JH, Kouriba B, Argiro L, Rahoud S, Mergani A, Henri S, Gaudart J, Mohamed-Ali Q, Young HA, Dessein AJ. IFN-gamma polymorphisms (IFN-gamma +2109 and IFN-gamma +3810) are associated with severe hepatic fibrosis in human hepatic schistosomiasis (*Schistosoma mansoni*). *J Immunol*. 2003 Nov 15;171(10):5596-601. PMID: 14607968
- Cioli D, Pica-Mattoccia L. Praziquantel. *Parasitol Res*. 2003 Jun;90 Supp 1:S3-9. PMID: 12811543
- Colley DG, Binder S, Campbell C, King CH, Tchuem Tchuenté LA, N'Goran EK, Erko B, Karanja DM, Kabatereine NB, van Lieshout L, Rathbun S. A five-country evaluation of a point-of-care circulating cathodic antigen urine assay for the prevalence of *Schistosoma mansoni*. *Am J Trop Med Hyg*. 2013 Mar;88(3):426-32. doi: 10.4269/ajtmh.12-0639. PMID: 23339198

- Colley DG, Secor WE. Immunology of human schistosomiasis. *Parasite Immunol.* 2014 Aug;36(8):347-57. doi: 10.1111/pim.12087. PMID: 25142505
- Coltart CE, Chew A, Storrar N, Armstrong M, Suff N, Morris L, Chiodini PL, Whitty CJ. Schistosomiasis presenting in travellers: a 15 year observational study at the Hospital for Tropical Diseases, London. *Trans R Soc Trop Med Hyg.* 2015 Mar;109(3):214-20. doi: 10.1093/trstmh/tru195. PMID: 25575554
- Corachan M, Valls ME, Gascon J, Almeda J, Vilana R. Hematospermia: a new etiology of clinical interest. *Am J Trop Med Hyg.* 1994 May;50(5):580-4. PMID: 8203707
- Cordero M, Rojo F A. *Parasitología General.* Madrid. Mc-Graw-Hill-Interamericana.1999
- Coulibaly JT, Ouattara M, Becker SL, Lo NC, Keiser J, N'Goran EK, Ianniello D, Rinaldi L, Cringoli G, Utzinger J. Comparison of sensitivity and faecal egg counts of Mini-FLOTAC using fixed stool samples and Kato-Katz technique for the diagnosis of *Schistosoma mansoni* and soil-transmitted helminths. *Acta Trop.* 2016 Dec;164:107-116. doi: 10.1016/j.actatropica.2016.08.024. Epub 2016 Aug 31. PMID: 27591137
- Cox A, Ackert-Bicknell CL, Dumont BL, Ding Y, Bell JT, Brockmann GA, Wergedal JE, Bult C, Paigen B, Flint J, Tsaih SW, Churchill GA, Broman KW. A new standard genetic map for the laboratory mouse. *Genetics.* 2009 Aug;182(4):1335-44. doi: 10.1534/genetics.109.105486. PMID: 19535546
- Crellen T, Allan F, David S, et al. Whole genome resequencing of the human parasite *Schistosoma mansoni* reveals population history and effects of selection. *Sci Rep.* 2016 Feb 16;6:20954. doi: 10.1038/srep20954. PMID: 26879532
- Da'dara AA, Siddons G, Icaza M, Wang Q, Skelly PJ. How schistosomes alter the human serum proteome. *Mol Biochem Parasitol.* 2017 Jul;215:40-46. doi: 10.1016/j.molbiopara.2016.12.007. PMID: 28011341
- de Jesus AR, Magalhães A, Miranda DG, Miranda RG, Araújo MI, de Jesus AA, Silva A, Santana LB, Pearce E, Carvalho EM. Association of type 2 cytokines with hepatic fibrosis in human *Schistosoma mansoni* infection. *Infect Immun.* 2004 Jun;72(6):3391-7. PMID: 15155645
- Dean DA, Bukowski MA, Cheever AW. Relationship between acquired resistance, portal hypertension, and lung granulomas in ten strains of mice infected with *Schistosoma mansoni*. *Am J Trop Med Hyg.* 1981 Jul;30(4):806-14. PMID: 7258492
- Dessein A, Chevillard C, Arnaud V, Hou X, Hamdoun AA, Dessein H, He H, Abdelmaboud SA, Luo X, Li J, Varoquaux A, Mergani A, Abdelwahed M, Zhou J, Monis A, Pitta MG, Gasmelseed N, Cabantous S, Zhao Y, Prata A, Brandt C, Elwali NE, Argiro L, Li Y. Variants of CTGF are associated with hepatic fibrosis in Chinese, Sudanese, and Brazilians infected with schistosomes. *J Exp Med.* 2009 Oct 26;206(11):2321-8. doi: 10.1084/jem.20090383. PMID: 19822645
- Dessein AJ, Hillaire D, Elwali NE, Marquet S, Mohamed-Ali Q, Mirghani A, Henri S, Abdelhameed AA, Saeed OK, Magzoub MM, Abel L. Severe hepatic fibrosis in *Schistosoma mansoni* infection is controlled by a major locus that is closely linked to the interferon-gamma receptor gene. *Am J Hum Genet.* 1999 Sep;65(3):709-21. PMID: 10441577
- Diakitè NR, Winkler MS, Coulibaly JT, Guindo-Coulibaly N, Utzinger J, N'Goran EK. Dynamics of freshwater snails and *Schistosoma* infection prevalence in schoolchildren during the construction and operation of a multipurpose dam in central Côte d'Ivoire. *Infect Dis Poverty.* 2017 May 4;6(1):93. doi: 10.1186/s40249-017-0305-3. PMID: 28468667

- Driguez P, McManus DP, Gobert GN. Clinical implications of recent findings in schistosome proteomics. *Expert Rev Proteomics*. 2016;13(1):19-33. doi: 10.1586/14789450.2016.1116390. PMID: 26558506
- Everts B, Hussaarts L, Driessen NN, Meevissen MH, Schramm G, van der Ham AJ, van der Hoeven B, Scholzen T, Burgdorf S, Mohrs M, Pearce EJ, Hokke CH, Haas H, Smits HH, Yazdanbakhsh M. Schistosome-derived omega-1 drives Th2 polarization by suppressing protein synthesis following internalization by the mannose receptor. *J Exp Med*. 2012 Sep 24;209(10):1753-67, S1. PMID: 22966004
- Fallon PG, Richardson EJ, McKenzie GJ, McKenzie AN. Schistosome infection of transgenic mice defines distinct and contrasting pathogenic roles for IL-4 and IL-13: IL-13 is a profibrotic agent. *J Immunol*. 2000 Mar 1;164(5):2585-91. PMID: 10679097
- Fernández-Soto P, Gandasegui Arahuetes J, Sánchez Hernández A, López Abán J, Vicente Santiago B, Muro A. A loop-mediated isothermal amplification (LAMP) assay for early detection of *Schistosoma mansoni* in stool samples: a diagnostic approach in a murine model. *PLoS Negl Trop Dis*. 2014 Sep 4;8(9):e3126. doi: 10.1371/journal.pntd.0003126. PMID: 25187956
- Ferrari TC. Involvement of central nervous system in the schistosomiasis. *Mem Inst Oswaldo Cruz*. 2004;99(5 Suppl 1):59-62. PMID: 15486636
- Fong CR, Moron NA, Kuris AM. Two's a crowd? Crowding effect in a parasitic castrator drives differences in reproductive resource allocation in single vs double infections. *Parasitology*. 2017 Apr;144(5):662-668. doi: 10.1017/S003118201600233X. PMID: 27928977
- Freedman DO, Weld LH, Kozarsky PE, Fisk T, Robins R, von Sonnenburg F, Keystone JS, Pandey P, Cetron MS; GeoSentinel Surveillance Network. Spectrum of disease and relation to place of exposure among ill returned travelers. *N Engl J Med*. 2006 Jan 12;354(2):119-30. Erratum in: *N Engl J Med*. 2006 Aug 31;355(9):967. PMID: 16407507
- Fu CL, Odegaard JI, Herbert DR, Hsieh MH. A novel mouse model of *Schistosoma haematobium* egg-induced immunopathology. *PLoS Pathog*. 2012;8(3):e1002605. doi: 10.1371/journal.ppat.1002605. PMID: 22479181
- Gandasegui J, Fernández-Soto P, Carranza-Rodríguez C, Pérez-Arellano JL, Vicente B, López-Abán J, Muro A. The Rapid-Heat LAMP Pellet Method: A Potential Diagnostic Method for Human Urogenital Schistosomiasis. *PLoS Negl Trop Dis*. 2015 Jul 31;9(7):e0003963. doi: 10.1371/journal.pntd.0003963. PMID: 26230990
- Gandasegui J, Fernández-Soto P, Dacal E, Rodríguez E, Saugar JM, Yepes E, Aznar-Ruiz-de-Alegría ML, Espasa M, Ninda A, Bocanegra C, Salvador F, Sulleiro E, Moreno M, Vicente B, López-Abán J, Muro A. Field and laboratory comparative evaluation of a LAMP assay for the diagnosis of urogenital schistosomiasis in Cubal, Central Angola. *Trop Med Int Health*. 2018b Jun 19. doi: 10.1111/tmi.13117. PMID: 29920859
- Gandasegui J, Fernández-Soto P, Muro A, Simões Barbosa C, Lopes de Melo F, Loyo R, de Souza Gomes EC. A field survey using LAMP assay for detection of *Schistosoma mansoni* in a low-transmission area of schistosomiasis in Umbuzeiro, Brazil: Assessment in human and snail samples. *PLoS Negl Trop Dis*. 2018a Mar 13;12(3):e0006314. doi: 10.1371/journal.pntd.0006314. PMID: 29534072
- Gatlin MR, Black CL, Mwinzi PN, Secor WE, Karanja DM, Colley DG. Association of the gene polymorphisms IFN-gamma +874, IL-13 -1055 and IL-4 -590 with patterns of

- reinfection with *Schistosoma mansoni*. *PLoS Negl Trop Dis*. 2009;3(2):e375. doi: 10.1371/journal.pntd.0000375. PMID: 19190772
- Gaze S, Driguez P, Pearson MS, Mendes T, Doolan DL, Trieu A, McManus DP, Gobert GN, Periago MV, Correa Oliveira R, Cardoso FC, Oliveira G, Nakajima R, Jasinskas A, Hung C, Liang L, Pablo J, Bethony JM, Felgner PL, Loukas A. An immunomics approach to schistosome antigen discovery: antibody signatures of naturally resistant and chronically infected individuals from endemic areas. *PLoS Pathog*. 2014 Mar 27;10(3):e1004033. doi: 10.1371/journal.ppat.1004033. PMID: 24675823
- GBD 2016 Causes of Death Collaborators. Global, regional, and national age-sex specific mortality for 264 causes of death, 1980-2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet*. 2017 Sep 16;390(10100):1151-1210. doi: 10.1016/S0140-6736(17)32152-9. PMID: 28919116
- GBD 2016 DALYs and HALE Collaborators. Global, regional, and national disability-adjusted life-years (DALYs) for 333 diseases and injuries and healthy life expectancy (HALE) for 195 countries and territories, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet*. 2017 Sep 16;390(10100):1260-1344. doi: 10.1016/S0140-6736(17)32130-X. PMID: 28919118
- GBD 2016 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet*. 2017 Sep 16;390(10100):1211-1259. doi: 10.1016/S0140-6736(17)32154-2. PMID: 28919117
- Glinz D, Silué KD, Knopp S, Lohourignon LK, Yao KP, Steinmann P, Rinaldi L, Cringoli G, N'Goran EK, Utzinger J. Comparing diagnostic accuracy of Kato-Katz, Koga agar plate, ether-concentration, and FLOTAC for *Schistosoma mansoni* and soil-transmitted helminths. *PLoS Negl Trop Dis*. 2010 Jul 20;4(7):e754. doi: 10.1371/journal.pntd.0000754. PMID: 20651931
- Gomes LI, Dos Santos Marques LH, Enk MJ, de Oliveira MC, Coelho PM, Rabello A. Development and evaluation of a sensitive PCR-ELISA system for detection of schistosoma infection in feces. *PLoS Negl Trop Dis*. 2010 Apr 20;4(4):e664. doi: 10.1371/journal.pntd.0000664. PMID: 20421918
- Gonçalves MM, Barreto MG, Peralta RH, Gargioni C, Gonçalves T, Igreja RP, Soares MS, Peralta JM. Immunoassays as an auxiliary tool for the serodiagnosis of *Schistosoma mansoni* infection in individuals with low intensity of egg elimination. *Acta Trop*. 2006 Nov;100(1-2):24-30. PMID: 17069742
- Grenfell RF, Martins W, Drummond SC, Antunes CM, Voietta I, Otoni A, Oliveira AA, Silva-Moraes V, Oliveira ER, Oliveira E, Lambertucci JR, Fonseca CT, Coelho PM. Acute schistosomiasis diagnosis: a new tool for the diagnosis of schistosomiasis in a group of travelers recently infected in a new focus of *Schistosoma mansoni*. *Rev Soc Bras Med Trop*. 2013 Mar-Apr;46(2):208-13. doi: 10.1590/0037-8682-0064-2012. PMID: 23740077
- Grimes JE, Croll D, Harrison WE, Utzinger J, Freeman MC, Templeton MR. The roles of water, sanitation and hygiene in reducing schistosomiasis: a review. *Parasit Vectors*. 2015 Mar 13;8:156. doi: 10.1186/s13071-015-0766-9. PMID: 25884172
- Gryseels B, Polman K, Clerinx J, Kestens L. Human schistosomiasis. *Lancet*. 2006 Sep 23;368(9541):1106-18. PMID: 16997665

- Gryseels B. Schistosomiasis. *Infect Dis Clin North Am*. 2012 Jun;26(2):383-97. doi: 10.1016/j.idc.2012.03.004. PMID: 22632645
- He H, Isnard A, Kouriba B, Cabantous S, Dessein A, Doumbo O, Chevillard C. A STAT6 gene polymorphism is associated with high infection levels in urinary schistosomiasis. *Genes Immun*. 2008 Apr;9(3):195-206. doi: 10.1038/gene.2008.2. PMID: 18273035
- Helmy MM. Touchdown PCR, ELISA and stool examination for early diagnosing of *Schistosoma mansoni* in mice. *J Egypt Soc Parasitol*. 2007 Dec;37(3):903-13. PMID: 18383792
- Hervé M, Angeli V, Pinzar E, Wintjens R, Faveeuw C, Narumiya S, Capron A, Urade Y, Capron M, Riveau G, Trottein F. Pivotal roles of the parasite PGD2 synthase and of the host D prostanoid receptor 1 in schistosome immune evasion. *Eur J Immunol*. 2003 Oct;33(10):2764-72. PMID: 14515260
- Hoy AM, Lundie RJ, Ivens A, Quintana JF, Nausch N, Forster T, Jones F, Kabatereine NB, Dunne DW, Mutapi F, Macdonald AS, Buck AH. Parasite-derived microRNAs in host serum as novel biomarkers of helminth infection. *PLoS Negl Trop Dis*. 2014 Feb 20;8(2):e2701. doi: 10.1371/journal.pntd.0002701. PMID: 24587461
- Huy NT, Hamada M, Kikuchi M, Lan NT, Yasunami M, Zamora J, Hirayama K. Association of HLA and post-schistosomal hepatic disorder: a systematic review and meta-analysis. *Parasitol Int*. 2011 Dec;60(4):347-56. doi: 10.1016/j.parint.2011.05.008. PMID: 21664486
- Ibironke OA, Phillips AE, Garba A, Lamine SM, Shiff C. Diagnosis of *Schistosoma haematobium* by detection of specific DNA fragments from filtered urine samples. *Am J Trop Med Hyg*. 2011 Jun;84(6):998-1001. doi: 10.4269/ajtmh.2011.10-0691. PMID: 21633040
- Jankovic D, Wynn TA, Kullberg MC, Hieny S, Caspar P, James S, Cheever AW, Sher A. Optimal vaccination against *Schistosoma mansoni* requires the induction of both B cell- and IFN-gamma-dependent effector mechanisms. *J Immunol*. 1999 Jan 1;162(1):345-51. PMID: 9886405
- Jenkins SJ, Hewitson JP, Jenkins GR, Mountford AP. Modulation of the host's immune response by schistosome larvae. *Parasite Immunol*. 2005 Oct-Nov;27(10-11):385-93. PMID: 16179032
- Kato-Hayashi N, Yasuda M, Yuasa J, Isaka S, Haruki K, Ohmae H, Osada Y, Kanazawa T, Chigusa Y. Use of cell-free circulating schistosome DNA in serum, urine, semen, and saliva to monitor a case of refractory imported schistosomiasis hematobia. *J Clin Microbiol*. 2013 Oct;51(10):3435-8. doi: 10.1128/JCM.01219-13. PMID: 23884992
- Keiser J, Chollet J, Xiao SH, Mei JY, Jiao PY, Utzinger J, Tanner M. Mefloquine--an aminoalcohol with promising antischistosomal properties in mice. *PLoS Negl Trop Dis*. 2009;3(1):e350. doi: 10.1371/journal.pntd.0000350. PMID: 19125172
- Khalaf I, Shokeir A, Shalaby M. Urologic complications of genitourinary schistosomiasis. *World J Urol*. 2012 Feb;30(1):31-8. doi: 10.1007/s00345-011-0751-7. PMID: 21909645
- King CH, Dangerfield-Cha M. The unacknowledged impact of chronic schistosomiasis. *Chronic Illn*. 2008 Mar;4(1):65-79. doi: 10.1177/1742395307084407. PMID: 18322031
- King CH. Health metrics for helminth infections. *Acta Tropica*. 2015;141, Part B:150–160.

- King CH. Parasites and poverty: the case of schistosomiasis. *Acta Trop.* 2010 Feb;113(2):95-104. doi: 10.1016/j.actatropica.2009.11.012. PMID: 19962954
- Knight M, Arican-Goktas HD, Ittiprasert W, Odoemelum EC, Miller AN, Bridger JM. Schistosomes and snails: a molecular encounter. *Front Genet.* 2014 Jul 21;5:230. doi: 10.3389/fgene.2014.00230. PMID: 25101114
- Kouriba B, Chevillard C, Bream JH, Argiro L, Dessein H, Arnaud V, Sangare L, Dabo A, Beavogui AH, Arama C, Traoré HA, Doumbo O, Dessein A. Analysis of the 5q31-q33 locus shows an association between IL13-1055C/T IL-13-591A/G polymorphisms and *Schistosoma haematobium* infections. *J Immunol.* 2005 May 15;174(10):6274-81. PMID: 15879126
- Lander E, Kruglyak L. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat Genet.* 1995 Nov;11(3):241-7. PMID: 7581446
- Lawrence RA, Allen JE, Osborne J, Maizels RM. Adult and microfilarial stages of the filarial parasite *Brugia malayi* stimulate contrasting cytokine and Ig isotype responses in BALB/c mice. *J Immunol.* 1994 Aug 1;153(3):1216-24. PMID: 7913112
- Lewis F, Tucker M. In: Toledo R, Fried B, Ed. *Schistosomiasis*. vol. 766. Springer, New Yor, 2014.
- Libby P. Mechanisms of acute coronary syndromes and their implications for therapy. *N Engl J Med.* 2013 May 23;368(21):2004-13. doi: 10.1056/NEJMra1216063. PMID: 23697515
- Lier T, Johansen MV, Hjelmevoll SO, Vennervald BJ, Simonsen GS. Real-time PCR for detection of low intensity *Schistosoma japonicum* infections in a pig model. *Acta Trop.* 2008 Jan;105(1):74-80. PMID: 18036505
- Lingscheid T, Kurth F, Clerinx J et al. Schistosomiasis in European Travelers and Migrants: Analysis of 14 Years TropNet Surveillance Data. *Am J Trop Med Hyg.* 2017 Aug;97(2):567-574. doi: 10.4269/ajtmh.17-0034. PMID: 2872263
- LoVerde PT, Hirai H, Merrick JM, Lee NH, El-Sayed N. *Schistosoma mansoni* genome project: an update. *Parasitol Int.* 2004 Jun;53(2):183-92. PMID: 15081950
- Ludolf F, Patrocínio PR, Corrêa-Oliveira R, Gazzinelli A, Falcone FH, Teixeira-Ferreira A, Perales J, Oliveira GC, Silva-Pereira RA. Serological screening of the *Schistosoma mansoni* adult worm proteome. *PLoS Negl Trop Dis.* 2014 Mar 20;8(3):e2745. doi: 10.1371/journal.pntd.0002745. PMID: 24651847
- Maizels RM, Yazdanbakhsh M. Immune regulation by helminth parasites: cellular and molecular mechanisms. *Nat Rev Immunol.* 2003 Sep;3(9):733-44. PMID: 12949497
- Mangano VD, Modiano D. Host genetics and parasitic infections. *Clin Microbiol Infect.* 2014 Dec;20(12):1265-75. doi: 10.1111/1469-0691.12793. PMID: 25273270 .
- Mann VH, Morales ME, Rinaldi G, Brindley PJ. Culture for genetic manipulation of developmental stages of *Schistosoma mansoni*. *Parasitology.* 2010 Mar;137(3):451-62. doi: 10.1017/S0031182009991211. Epub 2009 Sep 21. Review. PMID: 19765348
- Marquet S, Abel L, Hillaire D, Dessein H, Kalil J, Feingold J, Weissenbach J, Dessein AJ. Genetic localization of a locus controlling the intensity of infection by *Schistosoma mansoni* on chromosome 5q31-q33. *Nat Genet.* 1996 Oct;14(2):181-4. PMID: 8841190
- McWilliam HE, Driguez P, Piedrafita D, McManus DP, Meeusen EN. Novel immunomic technologies for schistosome vaccine development. *Parasite Immunol.* 2012 May;34(5):276-84. doi: 10.1111/j.1365-3024.2011.01330.x. PMID: 22486551

- Meltzer EI, Percik R, Shatzkes J, Sidi Y, Schwartz E. Eosinophilia among returning travelers: a practical approach. *Am J Trop Med Hyg.* 2008 May;78(5):702-9. PMID: 18458300
- Morenikeji O, Quazim J, Omoregie C, Hassan A, Nwuba R, Anumudu C, Adejuwon S, Salawu O, Jegede A, Odaibo A. A cross-sectional study on urogenital schistosomiasis in children; haematuria and proteinuria as diagnostic indicators in an endemic rural area of Nigeria. *Afr Health Sci.* 2014 Jun;14(2):390-6. doi: 10.4314/ahs.v14i2.15. PMID: 25320589
- Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods.* 2008 Jul;5(7):621-8. doi: 10.1038/nmeth.1226. PMID: 18516045
- Mountford AP, Trottein F. Schistosomes in the skin: a balance between immune priming and regulation. *Trends Parasitol.* 2004 May;20(5):221-6. PMID: 15105022
- Mouse Map Converter 2017. [Internet]. U.S.A. The Jackson Laboratory. [Acceso 15/06/2017]. URL disponible en: <http://cgd.jax.org/mousemapconverter/>
- Muro A, Pérez del Villar L, Velasco V, Pérez-Arellano JL. Infecciones por trematodos. *Medicine* 2010; 10: 3717-28.
- Nobre V, Silva LC, Ribas JG, Rayes A, Serufo JC, Lana-Peixoto MA, Marinho RF, Lambertucci JR. Schistosomal myeloradiculopathy due to *Schistosoma mansoni*: report on 23 cases. *Mem Inst Oswaldo Cruz.* 2001;96 Suppl:137-41. PMID: 11586439
- Ochodo EA, Gopalakrishna G, Spek B, Reitsma JB, van Lieshout L, Polman K, Lamberton P, Bossuyt PM, Leeflang MM. Circulating antigen tests and urine reagent strips for diagnosis of active schistosomiasis in endemic areas. *Cochrane Database Syst Rev.* 2015 Mar 11;(3):CD009579. doi: 10.1002/14651858.CD009579.pub2. PMID: 25758180
- Oliveira RR, Figueiredo JP, Cardoso LS, Jabar RL, Souza RP, Wells MT, Carvalho EM, Fitzgerald DW, Barnes KC, Araújo MI, Glesby MJ. Factors associated with resistance to *Schistosoma mansoni* infection in an endemic area of Bahia, Brazil. *Am J Trop Med Hyg.* 2012 Feb;86(2):296-305. doi: 10.4269/ajtmh.2012.11-0204. PMID: 22302866
- Pardo J, Carranza C, Muro A, Angel-Moreno A, Martín AM, Martín T, Hernández-Cabrera M, Pérez-Arellano JL. Helminth-related Eosinophilia in African immigrants, Gran Canaria. *Emerg Infect Dis.* 2006 Oct;12(10):1587-9. PMID: 17176579
- Parkin DM. The global health burden of infection-associated cancers in the year 2002. *Int J Cancer.* 2006 Jun 15;118(12):3030-44. PMID: 16404738
- Pearce EJ, MacDonald AS. The immunobiology of schistosomiasis. *Nat Rev Immunol.* 2002 Jul;2(7):499-511. PMID: 12094224
- Pedras-Vasconcelos JA, Pearce EJ. Type 1 CD8+ T cell responses during infection with the helminth *Schistosoma mansoni*. *J Immunol.* 1996 Oct 1;157(7):3046-53. PMID: 8816414
- Pérez- Arellano JL, Espinoza EY, Sánchez MM, Muro A. Evasion mechanisms of parasites. *Res Rev Parasitol* 2001; 61:4-16.
- Pérez del Villar L, Burguillo FJ, López-Abán J, Muro A. Systematic review and meta-analysis of artemisinin based therapies for the treatment and prevention of schistosomiasis. *PLoS One.* 2012;7(9):e45867. doi: 10.1371/journal.pone.0045867. Epub 2012 Sep 21. Review. PMID: 23029285

- Pérez del Villar L, Gutiérrez M, Vicente B, Castellanos A, Pérez-Losada J, Muro A. Identifying phenotypes involved in resistance to *Schistosoma mansoni* infection in inbred mice. *Trop Med Int Health* 2011, 16: 214---5.
- Pérez del Villar L, Vicente B, Blanco-Gómez A, Castellanos A, Pérez-Losada J, Muro A. Identifying phenotypes involved in susceptibility to *Schistosoma mansoni* infection in F1B6CBA mice. *Acta Parasitol.* 2014 Sep;59(3):529-39. doi: 10.2478/s11686-014-0277-4. PMID: 25119369
- Pérez del Villar L, Virginia V, Pérez-Arellano JL, Muro A. Infecciones difícilmente transmisibles IV: Esquistosomosis. *Manual de Enfermedades Importadas.* 2012; 1:483-491.
- Pérez López C., Métodos estadísticos avanzados con SPSS. Ediciones Paraninfo S. A. Madrid. 2005.
- Pérez-Arellano JL, Hernández-Cabrera M, Pisos-Álamo E, Carranza-Rodríguez C, Castillo-de-Vera M, Aparicio-Azcárraga P. Tratamiento de las enfermedades parasitarias (II): Helmintosis y ectoparasitosis. *Información Terapéutica del Sistema Nacional de Salud* 2007; 31, 55.
- Pinto-Almeida A, Mendes T, Armada A, Belo S, Carrilho E, Viveiros M, Afonso A. The Role of Efflux Pumps in *Schistosoma mansoni* Praziquantel Resistant Phenotype. *PLoS One.* 2015 Oct 7;10(10):e0140147. doi: 10.1371/journal.pone.0140147. PMID: 26445012
- Poggensee G, Krantz I, Kiwelu I, Feldmeier H. Screening of Tanzanian women of childbearing age for urinary schistosomiasis: validity of urine reagent strip readings and self-reported symptoms. *Bull World Health Organ.* 2000;78(4):542-8. PMID: 10885183
- Polderman AM, Panday UG, Ramkisoen S, van Lieshout L, Oostburg BF. A sedimentation-selective filtration method for the diagnosis of light infections with *Schistosoma mansoni*. *Acta Trop.* 1994 Dec;58(3-4):221-7. PMID: 7709861
- Pontes LA, Oliveira MC, Katz N, Dias-Neto E, Rabello A. Comparison of a polymerase chain reaction and the Kato-Katz technique for diagnosing infection with *Schistosoma mansoni*. *Am J Trop Med Hyg.* 2003 Jun;68(6):652-6. PMID: 12887022
- Portet A, Pinaud S, Tetreau G, Galinier R, Cosseau C, Duval D, Grunau C, Mitta G, Gourbal B. Integrated multi-omic analyses in *Biomphalaria-Schistosoma* dialogue reveal the immunobiological significance of FREP-SmPoMuc interaction. *Dev Comp Immunol.* 2017 Oct;75:16-27. doi: 10.1016/j.dci.2017.02.025. PMID: 28257854
- Protasio AV, Tsai IJ, Babbage A, Nichol S, Hunt M, Aslett MA, De Silva N, et al. A systematically improved high quality genome and transcriptome of the human blood fluke *Schistosoma mansoni*. *PLoS Negl Trop Dis.* 2012 Jan;6(1):e1455. doi: 10.1371/journal.pntd.0001455. PMID: 22253936
- Rodrigues V Jr, Piper K, Couissinier-Paris P, Bacelar O, Dessein H, Dessein AJ. Genetic control of schistosome infections by the SM1 locus of the 5q31-q33 region is linked to differentiation of type 2 helper T lymphocytes. *Infect Immun.* 1999 Sep;67(9):4689-92. PMID: 10456917
- Rojas-Caraballo J, López-Abán J, Pérez del Villar L, Vizcaíno C, Vicente B, Fernández-Soto P, del Olmo E, Patarroyo MA, Muro A. In vitro and in vivo studies for assessing the immune response and protection-inducing ability conferred by *Fasciola hepatica*-derived synthetic peptides containing B- and T-cell epitopes. *PLoS One.* 2014 Aug 14;9(8):e105323. doi: 10.1371/journal.pone.0105323. PMID: 25122166

- Rollinson D, Southgate VR. The genus *Schistosoma*: a taxonomic appraisal. In Rollinson D, Simpson AJG Ed. *The Biology of Schistosomes. From genes to latrines* Academic Press Ltd, London, 1987.
- Ross AG, Bartley PB, Sleight AC, Olds GR, Li Y, Williams GM, McManus DP. Schistosomiasis. *N Engl J Med*. 2002 Apr 18;346(16):1212-20. PMID: 11961151
- Ross AG, McManus DP, Farrar J, Hunstman RJ, Gray DJ, Li YS. Neuroschistosomiasis. *J Neurol*. 2012 Jan;259(1):22-32. doi: 10.1007/s00415-011-6133-7. PMID: 21674195
- Ross AG, Olds GR, Cripps AW, Farrar JJ, McManus DP. Enteropathogens and chronic illness in returning travelers. *N Engl J Med*. 2013 May 9;368(19):1817-25. doi: 10.1056/NEJMra1207777. PMID: 23656647
- Ross AG, Vickers D, Olds GR, Shah SM, McManus DP. Katayama syndrome. *Lancet Infect Dis*. 2007 Mar;7(3):218-24. PMID: 17317603
- Rutitzky LI, Bazzone L, Shainheit MG, Joyce-Shaikh B, Cua DJ, Stadecker MJ. IL-23 is required for the development of severe egg-induced immunopathology in schistosomiasis and for lesional expression of IL-17. *J Immunol*. 2008 Feb 15;180(4):2486-95. PMID: 18250458
- Sady H, Al-Mekhlafi HM, Ngui R, Atroosh WM, Al-Delaimy AK, Nasr NA, Dawaki S, Abdulsalam AM, Ithoi I, Lim YA, Chua KH, Surin J. Detection of *Schistosoma mansoni* and *Schistosoma haematobium* by Real-Time PCR with High Resolution Melting Analysis. *Int J Mol Sci*. 2015 Jul 16;16(7):16085-103. doi: 10.3390/ijms160716085. PMID: 26193254
- Sandoval N, Siles-Lucas M, Pérez-Arellano JL, Carranza C, Puente S, López-Abán J, Muro A. A new PCR-based approach for the specific amplification of DNA from different *Schistosoma* species applicable to human urine samples. *Parasitology*. 2006a Nov;133(Pt 5):581-7. PMID: 16834820
- Sarhan RM, Aminou HA, Saad GA, Ahmed OA. Comparative analysis of the diagnostic performance of adult, cercarial and egg antigens assessed by ELISA, in the diagnosis of chronic human *Schistosoma mansoni* infection. *Parasitol Res*. 2014 Sep;113(9):3467-76. doi: 10.1007/s00436-014-4017-3. PMID: 25028207
- Sayasone S, Utzinger J, Akkhavong K, Odermatt P. Repeated stool sampling and use of multiple techniques enhance the sensitivity of helminth diagnosis: a cross-sectional survey in southern Lao People's Democratic Republic. *Acta Trop*. 2015 Jan;141(Pt B):315-21. doi: 10.1016/j.actatropica.2014.09.004. PMID: 25225157
- Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods*. 2012 Jul;9(7):671-5. PMID: 22930834
- Siddiqui AA, Siddiqui SZ. Sm-p80-Based Schistosomiasis Vaccine: Preparation for Human Clinical Trials. *Trends Parasitol*. 2017 Mar;33(3):194-201. doi: 10.1016/j.pt.2016.10.010. PMID: 27865740
- Silva LL, Marcet-Houben M, Nahum LA, Zerlotini A, Gabaldón T, Oliveira G. The *Schistosoma mansoni* phylome: using evolutionary genomics to gain insight into a parasite's biology. *BMC Genomics*. 2012 Nov 13;13:617. doi: 10.1186/1471-2164-13-617. PMID: 23148687
- Smit CH, van Diepen A, Nguyen DL, Wuhrer M, Hoffmann KF, Deelder AM, Hokke CH. Glycomic Analysis of Life Stages of the Human Parasite *Schistosoma mansoni* Reveals Developmental Expression Profiles of Functional and Antigenic Glycan Motifs. *Mol*

- Cell Proteomics. 2015 Jul;14(7):1750-69. doi: 10.1074/mcp.M115.048280. PMID: 25883177
- Smith PM, Shainheit MG, Bazzone LE, Rutitzky LI, Poltorak A, Stadecker MJ. Genetic control of severe egg-induced immunopathology and IL-17 production in murine schistosomiasis. *J Immunol.* 2009 Sep 1;183(5):3317-23. doi: 10.4049/jimmunol.0901504. PMID: 19675160
- Sorgho H, Bahgat M, Poda JN, Song W, Kirsten C, Doenhoff MJ, Zongo I, Ouédraogo JB, Ruppel A. Serodiagnosis of *Schistosoma mansoni* infections in an endemic area of Burkina Faso: performance of several immunological tests with different parasite antigens. *Acta Trop.* 2005 Feb;93(2):169-80. PMID: 15652331
- Sotillo J, Doolan D, Loukas A. Recent advances in proteomic applications for schistosomiasis research: potential clinical impact. *Expert Rev Proteomics.* 2017 Feb;14(2):171-183. doi: 10.1080/14789450.2017.1271327. PMID: 27937063
- Stensgaard AS, Utzinger J, Vounatsou P, Hürlimann E, Schur N, Saarnak CF, et al. Large-scale determinants of intestinal schistosomiasis and intermediate host snail distribution across Africa: does climate matter? *Acta Trop.* 2013 Nov;128(2):378-90. doi: 10.1016/j.actatropica.2011.11.010. PMID: 22142789
- Stete K, Krauth SJ, Coulibaly JT, Knopp S, Hattendorf J, Müller I, Lohourignon LK, Kern WV, N'goran EK, Utzinger J. Dynamics of *Schistosoma haematobium* egg output and associated infection parameters following treatment with praziquantel in school-aged children. *Parasit Vectors.* 2012 Dec 21;5:298. doi: 10.1186/1756-3305-5-298. PMID: 23259435
- Stevens TL, Bossie A, Sanders VM, Fernandez-Botran R, Coffman RL, Mosmann TR, Vitetta ES. Regulation of antibody isotype secretion by subsets of antigen-specific helper T cells. *Nature.* 1988 Jul 21;334(6179):255-8. PMID: 2456466
- Tallima H, Dvořák J, Kareem S, Abou El Dahab M, Abdel Aziz N, Dalton JP, El Ridi R. Protective immune responses against *Schistosoma mansoni* infection by immunization with functionally active gut-derived cysteine peptidases alone and in combination with glyceraldehyde 3-phosphate dehydrogenase. *PLoS Negl Trop Dis.* 2017 Mar 27;11(3):e0005443. doi: 10.1371/journal.pntd.0005443. eCollection 2017 Mar. PMID: 28346516
- Tanigawa C, Fujii Y, Miura M, Nzou SM, Mwangi AW, Nagi S, Hamano S, Njenga SM, Mbanefo EC, Hirayama K, Mwau M, Kaneko S. Species-Specific Serological Detection for Schistosomiasis by Serine Protease Inhibitor (SERPIN) in Multiplex Assay. *PLoS Negl Trop Dis.* 2015 Aug 20;9(8):e0004021. doi: 10.1371/journal.pntd.0004021. PMID: 26291988
- Ugbomoiko US, Ofoezie IE, Okoye IC, Heukelbach J. Factors associated with urinary schistosomiasis in two peri-urban communities in south-western Nigeria. *Ann Trop Med Parasitol.* 2010 Jul;104(5):409-19. doi: 10.1179/136485910X12743554760469. PMID: 20819309
- van Diepen A, Van der Velden NS, Smit CH, Meevissen MH, Hokke CH. Parasite glycans and antibody-mediated immune responses in *Schistosoma* infection. *Parasitology.* 2012 Aug;139(9):1219-30. doi: 10.1017/S0031182012000273. PMID: 22423613
- Verjovski-Almeida S, DeMarco R, Martins EA, Guimarães PE, Ojopi EP, et al. Transcriptome analysis of the acelomate human parasite *Schistosoma mansoni*. *Nat Genet.* 2003 Oct;35(2):148-57. PMID: 12973350

- Vester U1, Kardorff R, Traoré M, Traoré HA, Fongoro S, Juchem C, Franke D, Korte R, Gryseels B, Ehrich JH, Doebling E. Urinary tract morbidity due to *Schistosoma haematobium* infection in Mali. *Kidney Int.* 1997 Aug;52(2):478-81. PMID: 9264005
- Vinkeles Melchers NV, van Dam GJ, Shaproski D, Kahama AI, Brienen EA, Vennervald BJ, van Lieshout L. Diagnostic performance of *Schistosoma* real-time PCR in urine samples from Kenyan children infected with *Schistosoma haematobium*: day-to-day variation and follow-up after praziquantel treatment. *PLoS Negl Trop Dis.* 2014 Apr 17;8(4):e2807. doi: 10.1371/journal.pntd.0002807. PMID: 24743389
- Wang C, Chen L, Yin X, Hua W, Hou M, Ji M, Yu C, Wu G. Application of DNA-based diagnostics in detection of schistosomal DNA in early infection and after drug treatment. *Parasit Vectors.* 2011 Aug 24;4:164. doi: 10.1186/1756-3305-4-164. PMID: 21864384
- Webster BL, Southgate Vr, Littlewood DT. A revision of the interrelationships of *Schistosoma* including the recently described *Schistosoma guineensis*. *Int J Parasitol.* 2006 Jul;36(8):947-55. PMID: 16730013
- WHO. Schistosomiasis. Progress. Report. 2001-2011 and Strategic Plan 2012-2020. Geneva: World Health Organization, 2013.
- WHO. Status of schistosomiasis endemic countries 2016. [Internet]. Ginebra, Suiza. World Health Organization. 01/01/2017 [Acceso 14/05/2018] URL disponible en: [http://apps.who.int/neglected\\_diseases/ntddata/sch/sch.html](http://apps.who.int/neglected_diseases/ntddata/sch/sch.html).
- WHO. The control of schistosomiasis: report of a WHO Expert Committee. Geneva: World Health Organization. 1985.
- Wilkins MR, Sanchez JC, Gooley AA, Appel RD, Humphery-Smith I, Hochstrasser DF, Williams KL. Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. *Biotechnol Genet Eng Rev.* 1996;13:19-50. PMID: 8948108
- Wilson MS, Cheever AW, White SD, Thompson RW, Wynn TA. IL-10 blocks the development of resistance to re-infection with *Schistosoma mansoni*. *PLoS Pathog.* 2011 Aug;7(8):e1002171. doi: 10.1371/journal.ppat.1002171. PMID: 21829367
- Wilson MS, Mentink-Kane MM, Pesce JT, Ramalingam TR, Thompson R, Wynn TA. Immunopathology of schistosomiasis. *Immunol Cell Biol.* 2007 Feb-Mar;85(2):148-54. PMID: 17160074
- Wilson RA. The saga of schistosome migration and attrition. *Parasitology.* 2009 Oct;136(12):1581-92. doi: 10.1017/S0031182009005708. PMID: 19265564
- Wright JE, Werkman M, Dunn JC, Anderson RM. Current epidemiological evidence for predisposition to high or low intensity human helminth infection: a systematic review. *Parasit Vectors.* 2018 Jan 31;11(1):65. doi: 10.1186/s13071-018-2656-4. PMID: 29382360
- Xu J, Guan ZX, Zhao B, Wang YY, Cao Y, Zhang HQ, Zhu XQ, He YK, Xia CM. DNA detection of *Schistosoma japonicum*: diagnostic validity of a LAMP assay for low-intensity infection and effects of chemotherapy in humans. *PLoS Negl Trop Dis.* 2015 Apr 13;9(4):e0003668. doi: 10.1371/journal.pntd.0003668. PMID: 25874964
- Xu J, Liu AP, Guo JJ, Wang B, Qiu SJ, Sun H, Guan W, Zhu XQ, Xia CM, Wu ZD. The sources and metabolic dynamics of *Schistosoma japonicum* DNA in serum of the host. *Parasitol Res.* 2013 Jan;112(1):129-33. doi: 10.1007/s00436-012-3115-3. PMID: 22983220

- Yang Y, Gao J, Cheng W, Pan X, Yang Y, Chen Y, Dai Q, Zhu L, Zhou Y, Jiang Q. Three Gorges Dam: polynomial regression modeling of water level and the density of schistosome-transmitting snails *Oncomelania hupensis*. *Parasit Vectors*. 2018 Mar 14;11(1):183. doi: 10.1186/s13071-018-2687-x. PMID: 29540206
- Yepes E, Varela-M RE, López-Abán J, Rojas-Caraballo J, Muro A, Mollinedo F. Inhibition of Granulomatous Inflammation and Prophylactic Treatment of Schistosomiasis with a Combination of Edelfosine and Praziquantel. *PLoS Negl Trop Dis*. 2015 Jul 20;9(7):e0003893. doi: 10.1371/journal.pntd.0003893. PMID: 26191954
- Young ND, Jex AR, Li B, Liu S, Yang L, Xiong Z, Li Y, Cantacessi C, Hall RS, et al. Whole-genome sequence of *Schistosoma haematobium*. *Nat Genet*. 2012 Jan 15;44(2):221-5. doi: 10.1038/ng.1065. PMID: 22246508
- Yu F, Li Y, Liu L, Li Y. Comparative genomics of human-like *Schistosoma japonicum* genes indicates a putative mechanism for host-parasite relationship. *Genomics*. 2008 Feb;91(2):152-7. PMID: 18083328
- Zhou et al. *Schistosoma japonicum* Genome Sequencing and Functional Analysis Consortium. The *Schistosoma japonicum* genome reveals features of host-parasite interplay. *Nature*. 2009 Jul 16;460(7253):345-51. doi: 10.1038/nature08140. PMID: 19606140
- Zhu HQ, Xu J, Zhu R, Cao CL, Bao ZP, Yu Q, Zhang LJ, Xu XL, Feng Z, Guo JG. Comparison of the miracidium hatching test and modified Kato-Katz method for detecting *Schistosoma japonicum* in low prevalence areas of China. *Southeast Asian J Trop Med Public Health*. 2014 Jan;45(1):20-5. PMID: 24964649
- Zhu L, Liu J, Cheng G. Role of microRNAs in schistosomes and schistosomiasis. *Front Cell Infect Microbiol*. 2014 Nov 11;4:165. doi: 10.3389/fcimb.2014.00165. PMID: 25426450

