



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

MÁSTER EN MEJORA GENÉTICA ANIMAL Y BIOTECNOLOGÍA DE LA
REPRODUCCIÓN

**ESTUDIO DE REGIONES GENÓMICAS
ASOCIADAS A LA VARIANZA AMBIENTAL
DEL TAMAÑO DE CAMADA EN CONEJOS**

Proyecto Final de Máster
Cristina Casto Rebollo

Universitat Politècnica de València
Julio 2018

Directora: Dra. Noelia Ibáñez Escriche

ÍNDICE GENERAL

ABSTRACT	7
RESUMEN.....	8
RESUM.....	10
1. REVISIÓN BIBLIOGRÁFICA	12
2. OBJETIVOS.....	23
ARTÍCULO 1: ESTUDIO DE REGIONES GENÓMICAS ASOCIADAS A LA VARIANZA AMBIENTAL EN EL TAMAÑO DE CAMADA EN CONEJOS.....	24
1. INTRODUCCIÓN	24
2. MATERIALES Y MÉTODOS	25
2.1. Material fenotípico y genotípico.	25
2.2. Procesado de datos.	26
2.2.1. Control de calidad y filtrado de muestras.....	26
2.2.2. Control de calidad y filtrado de SNPs.....	27
2.2.3. Imputación de genotipos faltantes.....	27
2.3. Análisis de asociación del genoma completo (GWAS).	28
2.3.1. Método de regresión marcador por marcador (SMR).	28
2.3.2. Método de regresión de múltiples marcadores (MMR).	29
2.4. Búsqueda de genes candidatos.	30
2.5. Estudio de desequilibrio de ligamiento.	30
2.6. Validaciones.	30
2.6.1. Test de permutación.	30
2.6.2. GWAS dentro de línea.	31
2.6.3. GWAS en otra población.	32
3. RESULTADOS	33
3.1. Análisis de asociación del genoma completo (GWAS).	33
3.2. Búsqueda de genes candidatos.	35

3.3. Estudio del desequilibrio de ligamiento.....	38
3.4. Validaciones.....	39
4. DISCUSIÓN.....	41
5. CONCLUSIÓN.....	51
6. REFERENCIAS.....	52
7. GLOSARIO.....	61
8. DATOS SUPLEMENTARIOS.....	64
Figura S1: Gráfico de componentes principales que muestra la estratificación de la población a través del PC1 y el PC2.....	64
Figura S2: Análisis del genoma completo para el tamaño de camada por regresión marcador a marcador y regresión con múltiples marcadores.....	65
Figura S3: Análisis del genoma completo para el número de embriones implantados por regresión marcador a marcador y regresión con múltiples marcadores.....	66
Figura S4: Análisis del genoma completo para la supervivencia embrionaria por regresión marcador a marcador y regresión con múltiples marcadores.....	67
Figura S5: Análisis del genoma completo para la supervivencia fetal por regresión marcador a marcador y regresión con múltiples marcadores.....	68
Figura S6: Análisis del genoma completo para la tasa de ovulación por regresión marcador a marcador y regresión con múltiples marcadores.....	69
Figura S7: Representación gráfica del posicionamiento aproximado del SNP Affx-151984290 dentro del gen <i>DOCK2</i>	69
Figura S8: Análisis del genoma completo para la tasa de ovulación por regresión marcador a marcador y regresión con múltiples marcadores.....	70
Tabla S1: Genes localizados en las regiones genómicas asociadas a V_E del tamaño de camada.....	71
Tabla S2: Funciones de los genes candidatos identificados para V_E del tamaño de camada, según las bases de datos Ensembl, GeneCards y KEGG.....	71
Tabla S3: Genes localizados en las regiones genómicas asociadas al tamaño de camada (TC), el número de embriones implantados (EI) y la supervivencia embrionaria (ES) y fetal (FS).....	73

Tabla S4: SNPs con diferente efecto de sustitución alélica entre la línea para baja (L) y alta (H) V_E del tamaño de camada. 74

ÍNDICE DE TABLAS

Tabla 1. Análisis descriptivos de los caracteres varianza ambiental (V_E), tamaño de camada (TC), tasa de ovulación (TO), embriones implantados (EI) y supervivencia embrionaria (ES) y fetal (FS) para la línea de alta y baja V_E del tamaño de camada.....26

Tabla 2. Análisis descriptivos para el carácter V_E del TC de las líneas para alta capacidad uterina (HUC), baja capacidad uterina (LUC) y una línea control (V).....32

Tabla 3. Regiones genómicas candidatas para los caracteres de tamaño de camada (TC), supervivencia fetal (FS), supervivencia embrionaria (ES) y embriones implantados (EI) según su posición en el cromosoma (Chr) en pares de bases.....35

Tabla 4. Genes localizados en las regiones genómicas asumidas como relevantes según su localización en el cromosoma (Chr) por pares de bases (PB) para V_E del TC.....36

Tabla 5. Genes candidatos identificados para los caracteres de tamaño de camada (TC), supervivencia prenatal (PS), supervivencia embrionaria (ES) y embriones implantados (EI)..37

ÍNDICE DE FIGURAS

Figura 1. Regiones genómicas asociadas a V_E del tamaño de camada.....33

Figura 2. Desequilibrio de ligamiento (r^2) para las regiones genómicas candidatas de V_E del tamaño de camada (TC).....39

Figura 3. Variación del porcentaje de varianza explicada por las regiones genómicas candidatas en función del valor del número de SNPs con efecto incluido en el modelo.....42

Figura 4. Respuesta inflamatoria inducida por daño en el tejido y la entrada de patógenos.....45

Figura 5. Rutas de señalización activadas por el desarrollo de una respuesta inflamatoria local.....49

“In biology, nothing is clear, everything is too complicated, everything is a mess, and just when you think you understand something, you peel off a layer and find deeper complications beneath. Nature is anything but simple.”

Richard Preston-The Hot Zone

ABSTRACT

In the last few years the genetic determination of environmental variance (V_E) has become more relevant for animal breeding. The direct relation of V_E with the environmental sensitivity and the opportunity to decrease V_E by selection can allow animals to improve their capacity to cope with changes of the environmental conditions. Results of a twelve-generation divergent selection experiment for V_E of litter size (LS) in rabbits show that does of the low V_E line tolerated external stressors more effectively than those of the high V_E line. However, it is not known what genes and metabolic pathways are regulating V_E of LS in rabbits. To identify genomic regions associated to V_E of litter size, 288 does of the high (149) and low (139) lines at generations 11th and 12th were genotyped using the Affymetrix Axiom OrCunSNP. V_E was calculated as the within-doe variance of litter size, after litter size pre-correction by year-season and three levels of parity-lactation effects. Association analyses were performed using two different methods: (a) single marked regression corrected by genotype relatedness matrix (GRM) excluding the chromosome of the SNP tested and (b) multiple marked regression with line as fixed effect and given a π value=0.9997. In both cases, three regions were identified in chromosomes 1, 3 and 14. In these regions, seven genes were highlighted as candidate genes (*CAT*, *DOCK2*, *LCP2*, *SLIT3*, *ST3GAL6*, *LMO2*, *GPR15* and *CD59*). Functions of these genes are mainly involved on immunological processes related to inflammatory response such as formation of selectin ligands, activation complement system, leukocytes migration and reduction of reactive species of oxygen. Associated regions were also identified in traits with a likely correlated response to selection for V_E of litter size as implanted embryo (EI), LS and embryonic (ES) and foetal survival (FS). Associated regions were identified in all traits except for TO. *ZP1* and *VEGFC* gene was highlighted among all genes. *ZP1* is in the chromosome 1 in a genomic region associated to EI and FS and its main functions are involved to ensure the structural integrity of the zona pellucida in the oocyte. *VEGFC* is associated to LS and its role is involved in angiogenesis during embryo development. Results of this research showed that selection by V_E of LS could modify allelic frequencies of genes involved on immunological response. Its support the assumption that animal's sensitivity to environment by selection of V_E of LS can be modified. Further these results also show that selection by V_E of TC might influence genes that control the performance of reproductive traits as LS, EI, ES and FS.

Keywords: environmental variance, environmental sensitivity, immunological process, genomic, rabbit.

RESUMEN

En los últimos años ha habido un creciente interés por estudiar el control genético de la varianza ambiental (V_E) en el campo de la mejora genética animal. La posibilidad de reducir la V_E por selección unido a su relación con la sensibilidad ambiental permite aumentar la capacidad de los animales para hacer frente a cambios en las condiciones ambientales. Un experimento de selección divergente en conejos para V_E del tamaño de camada (TC) durante 12 generaciones fue realizado por la Universidad Miguel Hernández de Elche. En estas líneas se observó que las hembras pertenecientes a la línea para baja V_E del TC toleraban de una forma más eficiente los factores de estrés externos, comparado con la línea para alta V_E del TC. Sin embargo, no se conocen los genes y rutas que podrían estar regulando la V_E del TC. El objetivo de este estudio fue identificar regiones genómicas asociadas a V_E del TC. Para ello se procedió a genotipar 288 hembras pertenecientes a las líneas de alta (149) y baja (139) V_E del TC de las generaciones 11 y 12 usando el chip de Affymetrix Axiom OrCunSNP. La V_E se calculó, tras la pre-corrección del tamaño de camada por los efectos año estación y tres niveles de parto-lactación, como la variación del tamaño de camada dentro de hembra. Los análisis de asociación se realizaron utilizando dos metodologías diferentes: (a) asociación marcador por marcador incluyendo en el modelo la matriz de relaciones de parentesco genómicas (GRM) construida excluyendo el cromosoma donde se encuentra el SNP a testar y (b) método de regresión con múltiples marcadores considerando la línea como efecto fijo y asumiendo un valor de $\pi=0.9997$. Por ambos métodos se identificaron tres regiones genómicas localizadas en los cromosomas 1, 3 y 14. En estas regiones, siete genes fueron seleccionados como genes candidatos (*CAT*, *DOCK2*, *LCP2*, *SLIT3*, *ST3GAL6*, *LMO2*, *GPR15* y *CD59*). La mayoría de las funciones de estos genes están relacionadas con procesos inmunológicos involucrados en la respuesta inflamatoria como la formación de ligandos de selectinas, la activación del sistema del complemento, la inducción de la migración leucocitaria y la reducción de especies reactivas del oxígeno. De igual forma, también se realizaron estudios de asociación sobre caracteres con respuesta correlacionada a la selección por V_E del TC como el tamaño de camada, el número de embriones implantados (EI), la tasa de ovulación (TO) y la supervivencia embrionaria (ES) y fetal (FS). En estos estudios se identificaron regiones genómicas asociadas en todos los caracteres exceptuando la TO. De todos los genes identificados destacan el gen *ZPI* y el gen *VEGFC*, asociado al carácter EI y ES y al TC, respectivamente. *ZPI* está relacionado con el mantenimiento de la integridad de la zona pelúcida del ovocito y *VEGFC* con la angiogénesis

durante el desarrollo embrionario. Los resultados obtenidos en este estudio sugieren que la selección por V_E modificaría las frecuencias alélicas de genes involucrados en la respuesta inmunológica. Estos resultados apoyarían la hipótesis de que se puede modificar la susceptibilidad de los animales al ambiente a través de la selección de V_E del TC. Además, los resultados también indicarían que la selección por V_E del TC puede influenciar genes que afecten el rendimiento en caracteres reproductivos como TC, EI, ES y FS.

Palabras claves: varianza ambiental, sensibilidad ambiental, procesos inmunológicos, genómica, conejo.

RESUM

En els últims anys hi ha hagut un creixent interès per estudiar el control genètic de la variància ambiental (V_E) en el camp de la millora genètica animal. La possibilitat de reduir la V_E per selecció unit a la seva relació amb la sensibilitat ambiental permet augmentar la capacitat dels animals per fer front a canvis en les condicions ambientals. Un experiment de selecció divergent en conills per V_E de Tamany de Camada (TC) durant 12 generacions va ser realitzat per la Universitat Miguel Hernández d'Elx. En aquestes línies es va observar que les femelles pertanyents a la línia per a baixa V_E del TC toleraven d'una manera més eficient els factors d'estrès externs, comparat amb la línia per a alta V_E del TC. No obstant això, no es coneixen els gens i rutes que podrien estar regulant la V_E del TC. L'objectiu d'aquest estudi va ser identificar regions genòmiques associades a v_e del TC. Per a això es va procedir a genotipar 288 femelles pertanyents a les línies d'alta (149) i baixa (139) V_E del TC de les generacions 11 i 12 usant el xip de Affymetrix Axiom OrCunSNP. La V_E es va calcular, després de la pre-correcció del TC pels efectes any estació i tres nivells de part-lactació, com la variació de TC dins de femella. Les anàlisis d'associació es van realitzar utilitzant dues metodologies diferents: (a) associació marcador per marcador incloent en el model la matriu de relacions de parentiu genòmiques (GRM) construïda excloent el cromosoma on es troba el SNP a testar i (b) mètode de regressió amb múltiples marcadors considerant la línia com a efecte fix i assumint un valor de $\pi = 0,9997$. Per tots dos mètodes es van identificar tres regions genòmiques localitzades en els cromosomes 1, 3 i 14. En aquestes regions, set gens van ser seleccionats com a gens candidats (CAT, DOCK2, LCP2, SLIT3, ST3GAL6, LMO2, GPR15 i CD59). La majoria de les funcions d'aquests gens estan relacionades amb processos immunològics involucrats en la resposta inflamatòria com la formació de lligands de selectinas, l'activació del sistema del complement, la inducció de la migració leucocitària i la reducció d'espècies reactives de l'oxigen. De la mateixa manera, també es van realitzar estudis d'associació sobre caràcters amb resposta correlacionada a la selecció per V_E del TC com el TC, el nombre d'embrions implantats (EI), la taxa d'ovulació (TO) i la supervivència embrionària (ES) i fetal (FS). En aquests estudis es van identificar regions genòmiques associades en tots els caràcters exceptuant la TO. De tots els gens identificats destaquen el gen ZP1 i el gen VEGFC, associat al caràcter EI i ES i al TC, respectivament. ZP1 està relacionat amb el manteniment de la integritat de la zona pel·lúcida de l'oòcit i VEGFC amb l'angiogènesi durant el desenvolupament embrionari. Els resultats obtinguts en aquest estudi suggereixen que la selecció per v_e modificaria les freqüències al·lèliques de gens involucrats en la resposta immunològica. Aquests resultats donarien suport

a la hipòtesi que es pot modificar la susceptibilitat dels animals a l'ambient a través de la selecció de V_E del TC. A més, els resultats també indicarien que la selecció per V_E del TC pot influenciar gens que afectin el rendiment en caràcters reproductius com TC, EI, ES i FS.

Paraules Claus: variància ambiental, sensibilitat ambiental, processos immunològics, genòmica, conills.

1. REVISIÓN BIBLIOGRÁFICA

El modelo clásico para el estudio de caracteres cuantitativos asume que:

$$V_P = V_G + V_E, \quad (1)$$

donde V_P es la variabilidad fenotípica de los individuos, V_G es la variabilidad genética y V_E es la variabilidad ambiental, la cual recoge todas las variaciones no genéticas que afectan al fenotipo. Este modelo considera la ausencia de posibles interacciones y correlaciones entre el genotipo y el ambiente y asume que aquellos individuos que comparten ambiente poseen una varianza ambiental idéntica. Sin embargo, hay diversas situaciones donde estos supuestos no se cumplen. Por ejemplo, una correlación genotipo ambiente puede aparecer cuando no hay aleatoriedad en los ambientes (Falconer and Mackay, 1996). Un ejemplo clásico, en vacuno lechero, es el tratamiento preferencial de los ganaderos a determinadas vacas. Por otra parte, la interacción genotipo ambiente (GxE) aparece cuando ambientes distintos no afectan de la misma forma el rendimiento de diferentes genotipos (Falconer and Mackay, 1996). Igualmente, trabajos como el de Hill (1984) han mostrado que animales separados en ambientes diferentes pueden variar tanto en media como en V_E del carácter, lo que se conoce como heterogeneidad ambiental. Hay estudios que muestran que parte de esta heterogeneidad del ambiente puede deberse a un control genético de la V_E (Ibáñez-Escriche et al., 2008; Ros et al., 2004; Sorensen and Waagepetersen, 2003). Además, existen evidencias que sugieren que las diferencias encontradas tanto en V_E como en la interacción GxE están relacionadas directamente con la capacidad de los individuos de responder a los estímulos ambientales, es decir, con su sensibilidad ambiental (SanCristobal-Gaudy et al., 1998; Waddington, 1942; Wu, 1997).

La sensibilidad ambiental puede definirse de dos formas diferentes, como cambios en diferentes ambientes de la media fenotípica de un genotipo dado (sensibilidad macro-ambiental), o como diferencias en la varianza residual de diferentes genotipos en el mismo ambiente (sensibilidad micro-ambiental) (Ros et al., 2004). Cada una de estas definiciones implica la asunción de distintos factores ambientales afectando el fenotipo del individuo, así como dos modelos estadísticos diferentes. La primera definición implica una modificación de la respuesta del fenotipo producida por una interacción entre el genotipo del animal y factores macro-ambientales (GxE) (Wakchaure et al., 2016). Estos factores macro-ambientales suelen ser conocidos y predecibles y pueden definirse en categorías discretas o cuantificarse (Temperatura, sistema de producción, etc.) (Mulder et al., 2013b). Generalmente, este tipo de

sensibilidad ambiental se suele estudiar mediante modelos de norma de reacción o modelos multicarácter donde cada ambiente se analiza como un carácter diferente (Falconer, 1990; Mulder et al., 2013b; Pigliucci, 1992).

La segunda definición de sensibilidad ambiental implica que los factores ambientales no se pueden determinar, son desconocidos o no categorizados y se definen como efectos micro-ambientales (Mulder et al., 2013b). Estos factores se representan como una suma de efectos o combinación de todos los factores ambientales a los que el individuo está expuesto a lo largo de su vida (Ros et al., 2004). La sensibilidad micro-ambiental estudia las diferencias presentadas en V_E dados factores ambientales desconocidos e implica la modelización de la varianza residual asumiendo un control genético de la V_E (Hill, 1984). En estudios de mejora genética animal se ha observado que una baja V_E está relacionada con individuos menos susceptibles a los factores del microambiente (Blasco et al., 2017; Strandberg et al., 2013; Vandenplas et al., 2013). Ejemplos de sensibilidad al micro-ambiente son comunes en mejora genética de plantas, donde clones de una misma planta presentan diferencias en V_E bajo las mismas condiciones ambientales (Wu, 1997). Estas diferencias de V_E entre individuos también se han observado en animales para caracteres como el peso al nacimiento en ratones (Formoso-Rafferty et al., 2016) o el tamaño de camada en conejo (Blasco et al., 2017), entre otros (Hill and Mulder, 2010).

En los últimos años el control genético de la sensibilidad micro-ambiental ha despertado un gran interés, tanto en el campo de la mejora genética animal como en el campo de la genética evolutiva. Desde una perspectiva evolutiva se considera que una variabilidad de la respuesta entre genotipos es ventajosa ya que permite mantener una variabilidad genética en las poblaciones (Wu, 1997). Esto se debe a que parte de la sensibilidad ambiental se considera heredable y permite a las especies evolucionar gracias al mantenimiento de sus capacidades de adaptación a los cambios y fluctuaciones en el ambiente (Wolf et al., 1998). En términos de competencia entre individuos, se podría considerar que mantener unos rangos de variabilidad fenotípica reducida (homogeneidad) podría suponer un beneficio para el animal, ya que en diversas situaciones se favorece a los individuos con los valores fenotípicos más cercanos al óptimo (Zhang and Hill, 2005). Sin embargo, mantener fenotipos con rangos pequeños de variabilidad supone un elevado coste para la fitness del animal, debido a la necesidad de afrontar los cambios en el microambiente (Zhang and Hill, 2005). En principio se asume que existe un equilibrio en los valores de la V_E debido a un balance de las fuerzas que aumentan (mutaciones)

y disminuyen la variabilidad (deriva génica y selección) (Zhang and Hill, 2005). Sin embargo, un conocimiento más profundo de las fuerzas que actúan sobre la V_E es necesario para afianzar los conocimientos sobre los mecanismos de adaptación de las especies.

Desde el punto de vista de la mejora genética animal, el interés del estudio y caracterización de la V_E se centró durante muchos años en la disminución de la variabilidad de los caracteres para la homogeneización de la producción (Ibáñez-Escriche et al., 2008; Marjanovic et al., 2016; SanCristobal-Gaudy et al., 1998). Homogeneizar los productos ganaderos permitiría optimizar el manejo y aumentar la eficiencia económica del mercado ya que se reducirían costes y se aumentarían los beneficios para algunos caracteres (Ej. peso al sacrificio en cerdos) (Ibáñez-Escriche et al., 2008a; Mulder et al., 2008). Además, la selección genética por baja V_E también podría aumentar la heredabilidad del carácter (Hill and Mulder, 2010; Yi et al., 2014). No obstante, el interés por la V_E se ha visto incrementado, en los últimos años, debido a la relación de la V_E con la resiliencia de los animales (Argente et al., 2014; Blasco et al., 2017; García et al., 2012). La resiliencia es la capacidad que tienen los individuos de afrontar las perturbaciones en el ambiente con un mantenimiento de su rendimiento productivo (Mulder, 2016). Recientes estudios han demostrado que líneas seleccionadas para baja V_E presentan mejores respuestas ante las perturbaciones ambientales que las seleccionadas para alta V_E (Argente et al., 2014; García et al., 2012). Por tanto, la V_E puede ser una medida de la resiliencia de los individuos y podría usarse como objetivo de selección incrementando también el bienestar de los animales (Broom, 2008). Asimismo, una mejor comprensión de la variabilidad genética de la V_E nos permitiría optimizar e implementar programas de selección con un aumento de la respuesta genética (Blasco et al., 2017; Formoso-Rafferty et al., 2016; Garreau et al., 2008).

El beneficio aparente de considerar la V_E como un componente relevante en los programas de mejora genética animal ha originado que numerosos estudios hayan estimado el valor aditivo de la V_E para diversos caracteres: Sorensen y Waagepetersen (2003) para el tamaño de camada (TC) en cerdo, Ros et al. (2005) para peso adulto en *Helix Aspersa*, Ibáñez-Escriche et al. (2008) para peso al sacrificio en cerdos a los 175 días y Vandenplas et al. (2013) para diferentes parámetros relacionados con el rendimiento lechero en ganado Holstein, entre otros (Argente et al., 2013; Ibáñez-Escriche et al., 2008a; Marjanovic et al., 2016). En general, los resultados de estos estudios muestran una heredabilidad baja para la V_E , entorno al 0.10, pero un coeficiente de variación de 0.2 o más (Hill and Mulder, 2010). Por tanto, hay evidencias

estadísticas que apoyan el determinismo genético de la V_E , lo que nos permitirá seleccionar los individuos en función de la variabilidad de un fenotipo al ambiente (Blasco et al., 2017; Formoso-Rafferty et al., 2016). En todos estos estudios, el modelo utilizado para la determinación genética de V_E fue el modelo jerárquico con estructura genética en la varianza residual descrito por SanCristobal-Gaudy et al. (1998):

$$P = \mu + A_m + \chi \exp \left[\frac{1}{2} (\log(\sigma_e^2) + A_{v,exp}) \right] \quad (2)$$

donde A_m es el efecto aditivo sobre la media poblacional (μ) y $A_{v,exp}$ es el efecto aditivo sobre el logaritmo natural de la escala de la varianza ambiental ($\log(\sigma_e^2)$). χ es un parámetro de escala asumido con una desviación normal $N \sim (0,1)$. El efecto aditivo se considera una normal bivalente para $\sigma_{A_m}^2$ y $\sigma_{A_{v,exp}}^2$ con $COV(A_m, A_{v,exp}) = r_A \sigma_{A_v} \sigma_{A_m,exp}$ definida a través de la matriz A de relaciones aditivas. Las principales razones por las que se aplica este modelo son porque permite, gracias al $\log(\sigma_e^2)$, eliminar la dependencia del factor de escala y además de evitar valores negativos en la varianza (SanCristobal-Gaudy et al., 1998). No obstante, este modelo es de gran complejidad y su resolución no es directa. De hecho, el desarrollo de métodos que permiten la estimación de la varianza genética de V_E es reciente. Desde un punto de vista frecuentista se ha desarrollado un método de estimación aproximado usando modelos dobles jerárquicos (Marjanovic et al., 2016; Mulder et al., 2013b; Ronnegard et al., 2010) y resuelto por REML (Lee and Nelder, 2006). La resolución de este modelo aplicando estadística bayesiana es directa, pero requiere la implementación de algoritmos de muestreo considerablemente complejos (Waagepetersen et al., 2008). En la actualidad, hay dos software que permiten estudiar la estructura genética de la varianza ambiental ASReml v4.1 (Gilmour et al., 2015) y GSEVM v.2 (Ibáñez-Escriche et al., 2010). Sin embargo, los modelos usados para el estudio del control genético de V_E están altamente parametrizados, por lo que las estimas pueden ser muy sensibles a los efectos asumidos en el modelo (Ibáñez-Escriche et al., 2008b; Sorensen and Waagepetersen, 2003). Además, Yang et al. (2011) muestran que estos modelos pueden no ser robustos y que las estimas de los parámetros genéticos como la correlación genética entre media y varianza son dependientes de la escala utilizada en el fenotipo. Así mismo, la alta complejidad e hiperparametrización de los modelos hace prácticamente inviable la estima de las correlaciones genéticas entre la V_E de un carácter y la de otros caracteres de interés. No obstante, antes de implementar la selección por V_E es fundamental conocer la posible respuesta correlacionada con otros caracteres. Por tanto, estudios experimentales centrados directamente en el análisis de la V_E eran necesarios para evaluar de una forma más precisa el

control genético de la varianza ambiental y apoyar así las evidencias estadísticas hasta ahora aportadas.

La primera evidencia experimental a favor de la estructura genética de V_E fue expuesta por Mackay y Lyman (2005) mediante un estudio llevado a cabo en *Drosophila*. En este estudio se comparó la V_E del número de pelos abdominales en líneas homocigotas y heterocigotas para el carácter. Los resultados mostraron una sensibilidad ambiental de V_E (inducida por temperatura) más elevada en la línea homocigota que en la heterocigota. Así mismo, Garreau et al. (2008) estudió la V_E mediante la creación de líneas divergentes para alta y baja V_E del peso al nacimiento en conejos, usando como criterio de selección el logaritmo de los residuos al cuadrado y observando una respuesta del 16% de la media tras 4 generaciones de selección. Además, estos autores estudiaron la respuesta correlacionada para caracteres como: el peso al destete, el número de nacidos vivos, el número de nacidos vivos destetados por camadas, la mortalidad al nacimiento y la mortalidad entre parto y destete. Los resultados mostraron que una disminución en la variabilidad del peso al nacimiento en conejos disminuía la mortalidad de los gazapos tanto al nacimiento como antes del destete. Del mismo modo, Formoso-Rafferty et al. (2016) en un experimento realizado con líneas seleccionadas divergentemente para V_E del peso al nacimiento en ratón (considerado como fenotipo de la madre) observó que las líneas de menor V_E ambiental también mostraban una menor mortalidad antes del destete, además de un mejor tamaño de camada (1.65 gazapos más a la octava generación). Esta selección fue realizada usando como criterio de selección el valor genético de la V_E .

En otro experimento, llevado a cabo por Blasco et al. (2017), también se generaron líneas divergentes para alta y baja V_E , en este caso para el TC en conejo. Este estudio usó directamente como criterio de selección la varianza fenotípica del tamaño de camada dentro de hembra y registró una respuesta a la selección con respecto a la media de 4.5% por generación. La selección por V_E del TC mostró, también, diferencias entre las líneas divergentes para caracteres reproductivos como el TC, el número de embriones implantados (EI) o la supervivencia embrionaria (ES) y prenatal (PS) (Argente et al., 2017). No se observó respuesta para la tasa de ovulación (TO) y supervivencia fetal (FS) (Argente et al., 2017). Estudios utilizando estas mismas líneas identificaron una relación entre la V_E y la resiliencia del animal a través de análisis cuantitativos de biomarcadores asociados con la “*fitness*” y el estado fisiológico del animal (Cortisol en plasma, Proteína C reactiva (CRP), Proteína amiloide sérica A (SAA) y

recuento de linfocitos y neutrófilos) (Argente et al., 2014; García et al., 2012). Todos estos biomarcadores están involucrados en la respuesta inflamatoria inducida por daños en el tejido o la presencia de patógenos (Goldsby et al., 2003). El inicio de esta respuesta se caracteriza por la liberación de citocinas proinflamatorias (IL-1, IL-6, TNF- α) que inducen una respuesta a nivel del hipotálamo, el hígado y la médula ósea (Goldsby et al., 2003). La respuesta en el hipotálamo se caracteriza por la síntesis de hormonas como la prostaglandina (para provocar la fiebre) y la hormona adrenocorticotrópica (ACTH), que induce la producción de cortisol en la corteza suprarrenal. Asimismo, la respuesta provocada en la médula ósea induce la producción de leucocitos que serán requeridos en el tejido inflamado (Goldsby et al., 2003). Los procesos desencadenados en el hígado se promueven tanto por las citocinas liberadas como por los corticoesteroides (como el cortisol) generados por la vía hipotalámica ante la respuesta a la inflamación (Goldsby et al., 2003). En el hígado se promueve la síntesis de proteínas de fase aguda como la CRP y la SAA, entre otras. Estas proteínas permitirán la activación del sistema del complemento que facilitará el reconocimiento de los patógenos por las células fagocíticas, para proceder a su eliminación (Goldsby et al., 2003). En dicho estudio se observó que la línea para baja V_E del TC presentaba niveles basales inferiores a la línea para alta V_E del TC en biomarcadores relacionados con la respuesta inflamatoria (García et al., 2012). Además, Argente et al. (2014) también determinaron que la línea para baja V_E presentaba una mejor respuesta frente a la vacunación con virus atenuados de enfermedades como la Mixomatosis y la enfermedad hemorrágica, ambas endémicas de conejo. Los estudios experimentales de V_E corroboran las evidencias estadísticas de su control genético. Además, sus resultados refuerzan la relación entre la V_E y la resiliencia. No obstante, apenas se conocen que genes y rutas biológicas podrían estar interviniendo en la V_E y como pueden afectar a otros caracteres relacionados con éste.

La publicación de un borrador del genoma completo de los seres humanos en el 2000 marcó el inicio de una era liderada por el análisis de datos genómicos a gran escala. El desarrollo computacional y de herramientas estadísticas potentes permitió a través de la bioinformática hacer frente a la cantidad masiva de datos generados por las técnicas de secuenciación de alto rendimiento (secuenciación masiva y genotipado). Análisis estadísticos centrados en el estudio de la asociación entre el código genético de todo el genoma de los individuos y su fenotipo (GWAS) permitió aumentar el conocimiento sobre las bases genéticas que controlaban los mecanismos biológicos implicados en la diversidad de las especies y las enfermedades. En el campo de la mejora genética animal, estos avances favorecieron la aplicación de evaluaciones

genéticas por medio del estudio de la información aportada por la identificación de polimorfismos de un único nucleótido (SNPs), llamada selección genómica. Esto ha hecho posible generar nuevas estrategias de selección e incrementar la respuesta genética en las poblaciones. Además, el uso de estudios GWAS ha facilitado la identificación de regiones genómicas asociadas a caracteres de interés (QTL) permitiendo aumentar el conocimiento sobre los procesos que determinan las diferencias en los caracteres de interés en la producción animal. Un ejemplo de ello es el interés por determinar las regiones genómicas asociadas a la V_E (vQTL) y su relación con la capacidad de adaptación de los animales al ambiente.

En la actualidad existen evidencias de vQTL identificados en diferentes especies animales (Morgante et al., 2015; Mulder et al., 2013a; Sell-Kubiak et al., 2015; Vandenplas et al., 2013; Wijga et al., 2012). En *Drosophila Melanogaster* se estudió la posible segregación de vQTLs para la resistencia a la inanición, al tiempo de recuperación de un coma inducido por frío y para la respuesta al sobresalto (Morgante et al., 2015). Se encontraron 36 genes en las regiones asociadas a la sensibilidad micro-ambiental. La mayoría de estos genes están relacionados con el desarrollo y funcionamiento del sistema nervioso y con el comportamiento del animal. De todos ellos destacó el gen candidato *Hsp90*, responsable de la respuesta a estrés térmico (Morgante et al., 2015). En cerdos, se han localizado vQTLs para el número de nacidos vivos totales (Sell-Kubiak et al., 2015). En particular, este estudio identificó un gen candidato relacionado con estrés térmico, el *HSPCB*, también en *Drosophila* (Morgante et al., 2015), y el factor de crecimiento endotelial vascular (*VEGFA*) (Sell-Kubiak et al., 2015), ambos correlacionados positivamente. En vacuno lechero se ha estudiado los vQTLs asociados a la incidencia de mastitis. S.Wijga et al. (2012) identificó regiones genómicas asociadas a la desviación estándar del recuento del número de células somáticas. Los resultados de este estudio mostraron dos regiones genómicas de 0.2 Mb alrededor de los SNPs significativos localizados en el cromosoma 6 y 18 de vacuno. Los genes localizados en esta región se han relacionado con procesos inmunológicos involucrados en la respuesta del sistema innato, como la migración y adhesión de los leucocitos. Asimismo, un estudio realizado por Mulder et al. (2013) analizó la asociación genómica con la variabilidad en el recuento del número de células somáticas e identificó una región genómica asociada en el cromosoma 9 de vacuno, con un SNP que explicaba un 3.5% de la varianza genética cuando se corrige por la matriz de relaciones de parentesco. No obstante, este estudio no realizó búsqueda de genes candidatos.

Además de en animales, en humanos también se han llevado a cabo estudios de asociación para identificar vQTLs relacionados con los caracteres altura e índice de masa corporal (Yang et al., 2012). En este caso se utilizaron datos provenientes de un meta-análisis de 38 estudios, con un total de 133 154 individuos de una población Europea. Los resultados de este estudio revelaron la importancia del gen candidato *FTO* en la variabilidad del índice de masa corporal entre personas, además de la relación ya conocida con la obesidad (Kilpeläinen et al., 2011). Estos resultados concordarían con las interacciones entre el ambiente (actividad física y la dieta) y el gen *FTO*, previamente descritas y relacionadas con la metilación del DNA (Bell et al., 2010). En este trabajo se indica que el estudio del control genético y caracterización de la V_E podría mejorar la respuesta de los tratamientos aplicados en los pacientes por identificación de los factores de riesgo asociados a V_E (Yang et al., 2012).

Las identificaciones de regiones genómicas asociadas a los caracteres de variabilidad, mencionas anteriormente, se realizaron a través de análisis de asociación del genoma completo (GWAS). Numerosas metodologías pueden usarse para la identificación de QTLs que estén segregando en las poblaciones (Gondro et al., 2013). No obstante, la mayoría explotan el desequilibrio de ligamiento (LD) existente entre polimorfismos de un solo nucleótido (SNPs) y los supuestos QTLs asociados al carácter (Bush and Moore, 2012; Vukcevic, 2009). A través de la estadística frecuentista y bayesiana se pueden realizar estudios de regresión de un único marcador o estudios de regresión de múltiples marcadores que permiten implementar estos análisis (Vukcevic, 2009).

Los análisis de regresión sobre un único marcador (SMR) (Gondro et al., 2013) asumen que los QTLs son unidades independientes y con un gran efecto sobre el carácter. Los efectos de los marcadores se consideran como efecto fijo del modelo y sólo se incluye el efecto de un único marcador en cada análisis (Burton et al., 2007). Generalmente, el modelo usado (Gondro et al., 2013) para los análisis es:

$$\mathbf{y} = \mathbf{W}\mathbf{b} + \mathbf{X}\mathbf{g} + \mathbf{e} \quad (3)$$

donde \mathbf{W} es la matriz de incidencia del vector de efectos sistemáticos \mathbf{b} , \mathbf{X} es la matriz de diseño de los efectos de los marcadores codificada según el número de copias del alelo minoritario (0, 1, 2), \mathbf{g} es el vector con el efecto del marcador y \mathbf{e} es el vector de los residuos, considerados independientes y distribuidos de forma $N \sim (\mathbf{0}, \mathbf{I}\sigma_e^2)$. La asociación estadística se calcula comparando el modelo causal (efecto del SNP testado) con el modelo nulo (no efecto del SNP

testado) a través de mínimos cuadrados generalizados (MLS) por la prueba F de Fisher o por la prueba chi-cuadrado y se obtiene el valor de significación (p-valor) del SNP testado (Yang et al., 2014). El valor de significancia (α) es fijado por el investigador y nos permite determinar el umbral a partir del cual los valores del p-valor apoyan el modelo causal (p-valor $< \alpha$), indicando una asociación estadística relevante entre el marcador y el carácter. Clásicamente, el valor del p-valor se fija a 0.05, pero para test múltiples se realizan correcciones del α mediante ajustes por FDR o por Bonferroni (Gondro et al., 2013). En el caso del ajuste por Bonferroni el umbral se determina en función de:

$$\alpha_B = \frac{\alpha}{M} \quad (4)$$

donde α_B es el nuevo umbral de significación, $\alpha = 0.05$ y M es el número total de SNPs considerado en el modelo. De esta manera se considera que el número total de SNPs testados por el modelo son independientes entre sí, lo que no es cierto ya que existen dependencias lineales entre ellos (Wray et al., 2013). Una solución a este problema es la estimación del número de bloques de SNPs en equilibrio de ligamiento. De forma general en mejora genética animal, el umbral de significación considerado para los análisis de asociación es $\alpha_B=1e-4$.

Uno de los aspectos que hay que tener en cuenta cuando se realizan GWAS con SMR es la asociación errónea de SNPs (error Tipo I) debido a factores de confusión como por ejemplo la estratificación de la población (J. Yang et al., 2011b). Una manera de solucionar este problema es la inclusión en el modelo de la matriz de relaciones de parentesco genómica (GRM) en los modelos SMR (J. Yang et al., 2011a), aunque puede disminuir la potencia de asociación (McCarthy et al., 2008). No obstante, la exclusión de la GRM del cromosoma del SNP a testar en el análisis permite aumentar la potencia de los análisis (Yang et al., 2014). Otro de los elementos a considerar es que las correlaciones entre SNPs conlleva a una sobrestimación de los efectos de los marcadores en SMR. Esto es debido a que un marcador puede recoger también los efectos de los SNPs que están en desequilibrio de ligamento con él (Bush and Moore, 2012). Una posibilidad de estimar el valor del efecto de los SNPs teniendo en cuenta las correlaciones entre marcadores es aplicar los análisis de regresión con múltiples marcadores (MMR) (Meuwissen et al., 2001). Estos métodos se resuelven a través de algoritmos bayesianos y asumen distribuciones para los efectos de los SNPs, considerados como variables aleatorias (Gondro et al., 2013; Meuwissen et al., 2001). En general, los métodos MMR se han utilizado principalmente en predicción. No obstante, hay estudios que muestran que los MMR presentan ventajas en la identificación de SNPs asociados a caracteres, respecto a los métodos de

regresión marcador por marcador (López de Maturana et al., 2014). En los últimos años, la aplicación de MMR para GWAS en mejora genética animal se ha incrementado (De los Campos et al., 2013), siendo el Bayes B el modelo más aplicado. En este modelo se asume que un porcentaje de los SNPs tienen efecto nulo, con una probabilidad π , mientras que el resto de SNPs con efecto y varianza se asumen con probabilidad $1-\pi$ (Gondro et al., 2013). Aquellos SNPs que presenten efectos elevados sobre el carácter tienen una alta probabilidad de ser incluidos de nuevo en el modelo, a medida que se realizan nuevas iteraciones. Además, aunque la correlación entre SNPs no es asumida a priori, sí que se tiene en cuenta cuando se estima sus distribuciones posteriores. El valor de π sirve para calcular el número de SNPs incluidos en el modelo por iteración:

$$SNP_{modelo} = SNP_{totales} \times (1 - \pi) \quad (5)$$

Este varía entre estudios y es dependiente del número total de SNPs y el número de datos disponibles para realizar el análisis (López de Maturana et al., 2014). Por otra parte, las distribuciones marginales posteriores obtenidas permiten, mediante el cálculo del Bayes Factor, realizar un test de hipótesis obteniendo la relevancia estadística de cada SNP, pudiéndose comparar así con las relevancias estadísticas obtenidas por SMR (Wakefield, 2009). Un valor de BF >10 indica una asociación fuerte, mientras que un BF > 100 indica una asociación decisiva de los marcadores con el QTL (Kass and Raftery, 1995).

Hay que tener en cuenta que los estudios GWAS, independientemente del método aplicado (SMR o MMR), son análisis exploratorios que nos permiten visualizar si existe un posible QTL segregando en la población. Posteriores análisis siempre son necesarios después de realizar un GWAS para validar los resultados obtenidos (Gondro et al., 2013). Esto es debido a que la mayoría de los caracteres cuantitativos analizados en mejora genética animal son explicados por múltiples QTLs de pequeño efecto cada uno. Además, para tener una precisión elevada de los efectos de los SNPs es necesario que el número de datos a analizar sea elevado, factor limitante en la mayoría de los estudios. Por otra parte, la mayoría de los SNPs que se determinan como significativos no son las variantes causales, si no variantes en LD con la variante causal (Bush and Moore, 2012; Zare et al., 2014). Generalmente esto se produce debido a que los chips de genotipado se construyen con polimorfismos representativos en la mayor parte de la población.

Hasta hace muy poco, los estudios GWAS estaban limitados en conejo por la ausencia de chips de SNPs así como de la caracterización del genoma completo del conejo. No obstante, la reciente publicación del genoma del conejo en 2014 y posteriormente la de un chip comercial de 200K brindan la oportunidad de realizar estudios para la identificación de QTLs asociados a caracteres de interés en esta especie (Sosa Madrid and Santacreu, 2016). El conejo doméstico (*Oryctolagus cuniculus*) es una especie de interés tanto desde un punto de vista productivo (producto de consumo humano) como en investigación (organismo modelo de otras especies). En mejora genética animal, el conejo es apreciado como modelo animal debido a que presenta unas características idóneas para una fácil manipulación y manejo: un tamaño mediano e intervalo generacional corto (9 meses). Su tamaño mediano permite también minimizar errores asociados al ambiente debido a que pueden mantenerse un número elevado de animales bajo un mismo ambiente (Misma nave). Por otra parte, su corto intervalo generacional unido a su elevado tamaño de camada (media= 8.22 y CV=0.22) (Sosa Madrid and Santacreu, 2016) permite obtener una respuesta a la selección rápida y explotar de forma intensiva dicha especie. Además, la creación de líneas divergentes, extraordinariamente útiles en experimentos de mejora genética, son muy factibles en esta especie (Blasco et al., 2017; García et al., 2016; Garreau et al., 2008). Una de las grandes ventajas del uso de líneas divergentes (en el mismo ambiente) es que las diferencias observadas entre ellas para un carácter seleccionado se deben a los cambios en las frecuencias alélicas de los genes que regulan dicho carácter. En la Universidad Miguel Hernández de Elche se han desarrollado dos líneas de conejos seleccionadas divergentemente durante 12 generaciones para la V_E del TC (Blasco et al., 2017). Los resultados de este experimento, mencionado previamente, han mostrado una elevada respuesta por generación, así como una relación directa entre la V_E y la resiliencia. El objetivo de este estudio es la identificación de regiones genómicas asociadas a la V_E del TC en conejo aprovechando este material genético excepcional junto con las nuevas herramientas biotecnológicas disponibles. Asimismo, se pretende determinar si dichas regiones están relacionadas con genes que pudieran estar implicados en el control de la “*fitness*” y la resiliencia de los animales, fenotipos de interés en la producción animal actual.

2. OBJETIVOS

1. Identificar regiones genómicas asociadas a la V_E del TC en líneas seleccionadas divergentemente para alta y baja V_E del TC.
2. Determinar si las regiones genómicas asociadas a V_E del TC podrían influir en la capacidad de los individuos para adaptarse a nuevas condiciones ambientales.
3. Identificar regiones genómicas asociadas a caracteres reproductivos en líneas seleccionadas divergentemente para alta y baja V_E del TC: tamaño de camada (TC), el número de embriones implantados (EI), la tasa de ovulación (TO) y la supervivencia embrionaria (ES) y fetal (FS).

ARTÍCULO 1: ESTUDIO DE REGIONES GENÓMICAS ASOCIADAS A LA VARIANZA AMBIENTAL EN EL TAMAÑO DE CAMADA EN CONEJOS

1. INTRODUCCIÓN

En los últimos años ha habido un creciente interés en el campo de la mejora genética animal por estudiar el control genético de la varianza ambiental (V_E). Numerosos estudios han mostrado evidencias estadísticas de la existencia de un efecto aditivo genético influenciando a la V_E (Ibáñez-Escriche et al., 2008; Lallias et al., 2017; Ros et al., 2004; Sorensen et al., 2015), observándose además una relación con la resiliencia y la “*fitness*” del animal, es decir, con su capacidad de adaptación a nuevas condiciones ambientales (Argente et al., 2014; García et al., 2012; Mulder et al., 2013a). Una mayor resiliencia permitiría al animal mantener su rendimiento a pesar de los factores de estrés externos, suponiendo un beneficio a nivel productivo (Blasco et al., 2017; Mulder and Rashidi, 2017).

La Universidad Miguel Hernández de Elche llevó a cabo un experimento de selección divergente en conejos para V_E en el tamaño de camada (TC) durante doce generaciones. Los resultados de este experimento mostraron una respuesta a la selección del 4.5% de la media del carácter por generación (Blasco et al., 2017). Estos resultados evidencian claramente la existencia de factores genéticos regulando la V_E del TC. Además, un estudio sobre esta misma población de Argente et al. (2017) determinó que la selección por variabilidad en el TC mostraba una respuesta correlacionada negativa para caracteres reproductivos como el TC, el número de embriones implantados (EI) y la supervivencia embrionaria (ES), fetal (FS) y prenatal (PS). De la misma manera, también se observó una correlación de V_E con la respuesta inmunológica en esta población. La línea de baja V_E del TC tuvo una mejor respuesta inmunológica que la línea de alta V_E (Argente et al., 2014; Blasco et al., 2017). Este resultado indicaría que una selección dirigida hacia la reducción de la V_E podría aumentar las capacidades de los individuos para adaptarse a nuevas condiciones ambientales (Blasco et al., 2017). No obstante, todavía no se conocen los genes ni las rutas y mecanismos implicados en la regulación de la V_E del TC. Por tanto, el objetivo de este estudio fue identificar regiones genómicas asociadas a V_E en estas líneas y determinar su posible relación con la resiliencia y la “*fitness*” del animal.

2. MATERIALES Y MÉTODOS

2.1. Material fenotípico y genotípico.

Los datos usados en este estudio pertenecen a un experimento de la Universidad Miguel Hernández de Elche de selección divergente para alta (H) y baja (L) V_E del TC durante 12 generaciones (Blasco et al., 2017). El material genético utilizado proviene de 288 hembras de las generaciones 11 (281) y 12 (7) que fueron genotipadas con el chip de Affymetrix Axiom OrcunSNP para un total de 199 692 SNPs. De estas hembras, 149 eran de la línea para H V_E y 139 para L V_E y tenían registrados 1254 partos de un total de 13788 pertenecientes a 3070 hembras de todas las generaciones.

El carácter estudiado V_E se calculó como la varianza del TC (número de nacidos totales) dentro de hembra. Este cálculo se realizó bajo la asunción de que el valor genético de la hembra y el efecto permanente es el mismo en todos los partos de la hembra. Previamente se realizó una pre-corrección del TC (Blasco et al., 2017) para los efectos de año estación (AE; 9 niveles) y de parto lactación (PL; 3 niveles) usando los registros de todas las generaciones. De esta manera se eliminaron posibles efectos sistemáticos que pudieran estar afectando a la varianza. El modelo lineal aplicado fue:

$$TC = AE + PL + e, \quad (6)$$

donde a partir de la media de la estima de los residuos (e) obtenidos con el software R (R Core Team, 2017), se calculó V_E por hembra usando el estimador de mínimo riesgo cuadrático (Blasco et al., 2017):

$$V_E = \frac{1}{n+1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (7)$$

Siendo x_i el TC pre-correcto de una hembra en el parto i y n el número total de partos de la hembra, cuyo valor puede oscilar entre 2 y 12 partos.

Además de los estudios de asociación realizados sobre el carácter V_E , también se llevaron a cabo estudios de asociación para varios caracteres reproductivos con una posible respuesta correlacionada con V_E : TC, tasa de ovulación (TO), número de embriones implantados (EI),

supervivencia embrionaria (ES) y supervivencia fetal (FS). Los datos utilizados para los caracteres TO, EI, ES y FS se tomaron en el segundo parto de cada hembra. En la Tabla 1 se presentan los estadísticos descriptivos de los datos analizados por línea.

En el caso del carácter TC se realizó una corrección con toda la base de datos, previa al análisis de asociación, de los datos por los factores ambientales y los efectos permanentes. Para ello se ajustó un modelo animal con efectos permanentes utilizando el software TM (Legarra et al., 2008). En los análisis de asociación se utilizó la media posterior de los valores aditivos más los residuales como fenotipo corregido.

Tabla 1. Análisis descriptivos de los caracteres varianza ambiental (V_E), tamaño de camada (TC), tasa de ovulación (TO), embriones implantados (EI) y supervivencia embrionaria (ES) y fetal (FS) para la línea de alta y baja V_E del tamaño de camada

Carácter	Alta V_E						Baja V_E					
	N ^(a)	μ ^(b)	M_e ^(c)	CV ^(d)	Min ^(e)	Max ^(f)	N	μ	M_e	CV	Min	Max
V_E	139	3.92	2.91	0.94	1.33e-6	19.39	130	2.35	2.06	0.83	3.82e-5	8.73
TC	643	6.51	6	0.47	1	15	593	7.71	8	0.31	1	13
TO	124	12.65	13	0.17	6	20	124	12.68	12.50	0.19	7	19
EI	124	9.05	10	0.34	2	15	124	10.77	11	0.24	4	16
ES	124	71.94	75.96	0.32	14.29	100.00	124	85.24	89.44	0.17	29.41	100.00
PS	124	66.09	71.43	0.38	10.00	100.00	124	70.79	73.03	0.30	18.18	100.00

^(a) Número total de datos, ^(b)media, ^(c)mediana, ^(d)coeficiente de variación y ^(e)máximo y ^(f)mínimo del rango de los datos

2.2. Procesado de datos.

2.2.1. Control de calidad y filtrado de muestras.

Los controles de calidad sobre las muestras genotipadas se realizaron usando la plataforma Axiom Analysis Suite 3.1 de ThermoFisher Scientific y el software PLINK v1.9 (Chang et al., 2015). Los criterios usados en la plataforma Axiom Analysis Suite 3.1 para excluir las muestras con baja calidad de genotipado fueron: señal de contraste en los locus no polimórficos (DQC)<0.89, eficiencia del genotipado (genotipos faltantes por individuo o QC) <96%, promedio de la eficiencia del genotipado por placa <98.5% y porcentajes de muestras que superan DQC y QC <95%. En PLINK v1.9 (Chang et al., 2015) se fijó que el número de

genotipos faltantes por individuos debía ser menor del 97%. Posteriormente se estudió la estructura de la población mediante un análisis de componentes principales para determinar posibles individuos “*outliers*” usando el paquete “SNPRelate” del software R (R Core Team, 2017). Se generó un gráfico con los componentes PC1 y PC2 y se determinó los “*clusters*” correspondientes al efecto de selección para la línea H y L (Figura S1). Aquellos individuos asociados a una línea diferente a la asignada en la base de datos fueron excluidos para los posteriores análisis. Finalmente, un total de 269 hembras pasaron los controles de calidad y fueron incluidas en los análisis (130 de la línea L y 139 de la línea H).

2.2.2. Control de calidad y filtrado de SNPs.

La plataforma Axiom Analysis Suite 3.1 se utilizó para filtrar aquellos SNPs que no se localizaran en una zona del mapa físico o genómico conocida en el genoma de referencia del conejo (OryCun 2.0). El resto de los controles de calidad sobre los SNPs se realizaron con el software PLINK v1.9 (Chang et al., 2015) excluyéndose aquellos que no cumplieran los siguientes criterios: genotipos faltantes por SNP menor del 5% y frecuencia del alelo minoritario (MAF) mayor del 5%. Finalmente, un total de 94 282 SNPs se tuvieron en cuenta para los análisis.

2.2.3. Imputación de genotipos faltantes.

La imputación de los genotipos faltantes se realizó con el software Beagle v.4.1 (Browning and Browning, 2016). En total se imputaron 10 437 genotipos faltantes de SNP. Ninguno de los SNP imputados fue descartado en posteriores análisis debido a que la calidad de la imputación (R^2) en todos los casos fue superior a 0.9.

2.3. Análisis de asociación del genoma completo (GWAS).

El análisis de asociación del genoma completo se realizó aplicando dos métodos diferentes: método de regresión marcador por marcador y método de regresión de múltiples marcadores.

2.3.1. Método de regresión marcador por marcador (SMR).

El análisis de regresión marcador a marcador se realizó con el software GCTA v1.91.4beta (Yang et al., 2011a) a partir del siguiente modelo lineal mixto (MLM):

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{z}_j\alpha_j + \mathbf{Z}\mathbf{g} + \mathbf{e}. \quad (8)$$

Donde \mathbf{X} es la matriz de incidencia del vector de efectos fijos \mathbf{b} que incluye la media, \mathbf{z}_j es el vector de genotipos del SNP j codificado según el número de copias del alelo de referencia (0,1,2), α_j es el efecto de sustitución alélica del SNP j y \mathbf{Z} es la matriz de incidencia de los efectos genético \mathbf{g} que siguen una distribución $N(\mathbf{0}, \mathbf{GRM}\sigma_g^2)$ donde la \mathbf{GRM} es la matriz de relaciones de parentesco genómica excluyendo el cromosoma del SNP j y σ_g^2 es reestimado en cada cálculo de la \mathbf{GRM} . La matriz \mathbf{GRM} fue calculada usando el método desarrollado por Yang et al. (2010). La matriz \mathbf{GRM} se incluyó en el modelo porque permite corregir por posibles efectos de estratificación de la población (Yang et al., 2014). Finalmente, a partir de este modelo el programa calculó la estima del efecto de los SNPs y su p-valor asociado.

El umbral de significación (α) utilizado para determinar si un SNP estaba asociado a V_E se ajustó con una corrección por Bonferroni teniendo en cuenta el número de SNPs independientes en la población. Para ello se calculó el valor del r^2 para el desequilibrio de ligamiento (LD) entre pares de SNPs con el software PLINK v1.9 (Chang et al., 2015), para una distancia no mayor a 1Mb. El número de SNPs independientes se determinó analizando el LD de los SNPs con el software R (R Core Team, 2017) y asumiendo que para un $r^2 < 0.5$ no existe LD entre SNPs (Carneiro et al., 2011). Finalmente, se estableció un umbral de significación con un p-valor $< 1.025e-5$ asumiendo un α inicial de 0.05 y un número de SNPs independientes de 4876.

2.3.2. Método de regresión de múltiples marcadores (MMR).

El análisis de regresión de múltiples marcadores se realizó con el software GenSel (Garrick and Fernando, 2013) según el siguiente modelo:

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \sum_{j=1}^k \mathbf{z}_j \alpha_j \delta_j + \mathbf{e} \quad (9)$$

Donde \mathbf{X} es la matriz de incidencia del vector de efectos fijos \mathbf{b} que incluye la media y el efecto línea, \mathbf{z}_j es la matriz de incidencia del SNP j codificado según el alelo de referencia, α_j es el efecto del SNP j , δ_j es una variable de modelización que indica los SNP con efecto y varianza que se incluyen en el modelo ($\delta_j = 1$) y los que no ($\delta_j = 0$) y \mathbf{e} es el vector de residuos. En el modelo bayesiano aplicado (Bayes B) se asume que los efectos de los SNPs son independientes con una distribución $N \sim (0, \sigma_{\alpha_j}^2)$ y que, con probabilidad π , hay un porcentaje de marcadores que no tienen efecto ni varianza, mientras que el resto de los marcadores, con probabilidad $1 - \pi$, sí que tienen efecto. La $\sigma_{\alpha_j}^2$ para estos marcadores sigue una distribución $IG(\alpha, \beta)$ (Blasco, 2017).

En este análisis el valor de π se fijó en 0.9997 (Sell-Kubiak et al., 2015). En media, un total de 28 SNPs con efecto se incluyeron en el modelo cada iteración. El número total de iteraciones se fijó en 550 000 teniendo en cuenta un “lag” de 100 y un “burn-in” en 150 000. La varianza genotípica y ambiental a priori se fijó en 1.69 y 7.73, respectivamente. Estos parámetros corresponden a los estimados con un modelo animal clásico.

La relevancia estadística de la asociación entre los marcadores individuales y el carácter se evaluó calculando el Bayes Factor (BF) según (Kass and Raftery, 1995):

$$BF = \frac{\hat{p}_j \times \pi}{(1 - \hat{p}_j) \times (1 - \pi)} \quad (10)$$

Donde \hat{p}_j es la probabilidad posterior de un SNP en el locus j dada la H_0 mientras que $1 - \hat{p}_j$ es la probabilidad dada H_1 . En el análisis se consideró que un $BF > 10$ determina una evidencia fuerte a favor de la hipótesis de asociación siendo un $BF > 100$ una evidencia decisiva (Kass and Raftery, 1995).

2.4. Búsqueda de genes candidatos.

La búsqueda de genes candidatos en las regiones genómicas asociadas se realizó a través de la base de datos de Ensembl (Zerbino et al., 2018). Las regiones genómicas estudiadas fueron las correspondientes a las ventanas genómicas que explicaban un mayor porcentaje de varianza genética a través del método de RMM y que además se solapaban en ambos estudios de asociación. El genoma de referencia utilizado fue el de la especie en estudio, el conejo doméstico (*Oryctolagus cuniculus*). Se revisó la ontología genética de cada gen localizado en las regiones (Función biológica y molecular) y se validaron con la base de datos GeneCards (Rappaport et al., 2017) y KEGG (Kanehisa et al., 2017)

Posteriormente, aquellos SNPs que en ambos métodos mostraron una elevada relevancia estadística fueron localizados en el genoma del conejo. El objetivo fue ver si se encontraban dentro de la secuencia genética de alguno de los genes considerados como candidatos. En los casos afirmativos se procedió al análisis del transcripto de dichos genes para ver si la localización del SNP correspondía a zonas intrónicas o exónicas del gen. Para ello se usó también la base de datos Ensembl (Zerbino et al., 2018).

2.5. Estudio de desequilibrio de ligamiento.

El estudio del desequilibrio de ligamiento se realizó a partir de la función LDheatmap() (Shin et al., 2006) de R (R Core Team, 2017) con ficheros generados con PLINK v1.9 (Chang et al., 2015). Las regiones usadas para el estudio de LD fueron las ventanas genómicas en las que se había realizado una búsqueda de genes candidatos. Se ignoró comparaciones entre pares de SNPs para distancias mayores a 1Mb.

2.6. Validaciones.

2.6.1. Test de permutación.

El test de permutación se realizó mediante el software PLINK v1.9 (Chang et al., 2015). El objetivo de este estudio fue eliminar la estructura de genealogía de los datos para observar si los resultados obtenidos en los GWAS no eran debidos a asociaciones aleatorias en la

población. Para ello, el programa generó 100 000 set de datos fenotípicos a partir del fichero de fenotipos reales para V_E , mediante la permutación de los fenotipos entre los individuos. De esta manera, se eliminó la relación entre el genotipo y el fenotipo de los individuos generándose un set de datos correspondientes a la H_0 . Posteriormente, se realizó un análisis de asociación marcador por marcador con los resultados obtenidos según el modelo:

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{z}_j\alpha_j + \mathbf{e} \quad (11)$$

Donde \mathbf{y} es el vector de datos fenotípicos permutado (modificado en cada iteración), \mathbf{X} es la matriz de incidencia de los efectos sistemáticos \mathbf{b} que incluyen la media y el efecto línea, \mathbf{z}_j es el vector de incidencia del SNP j codificado según el número de copias del alelo de referencia (0,1,2), α_j es el efecto del SNP j y \mathbf{e} es el vector de los efectos de los residuos. Test de hipótesis son realizados para calcular la probabilidad de permutación (Knijnenburg et al., 2009) de los SNPs según:

$$P_{jperm} = \frac{\sum_{n=1}^{N_T} (x_{jn}^* > x_0)}{N_T} \quad (12)$$

Donde P_{perm} es la probabilidad de permutación de cada SNP j , N_T es el número total de iteraciones de permutaciones realizadas, x_n^* es el p-valor para cada SNP j para la permutación n (asumiendo que H_0 es cierta) y x_0 es el p-valor de la asociación del SNP j de los fenotipos sin permutar. P_{perm} es por tanto la probabilidad de que un SNP posea un valor de x_n^* superior al establecido por el análisis de asociación para los datos reales (x_0). Si P_{perm} es elevada indicaría que la probabilidad de que x_n^* sea mayor que x_0 es elevada por lo que las relevancias estadísticas observadas para los fenotipos sin permutar son generadas al azar (Knijnenburg et al., 2009). Para considerar que x_0 ha sido producto de la selección y no del azar, el valor de P_{jperm} deber ser muy pequeño e inferior a 0.02.

2.6.2. GWAS dentro de línea.

Dentro de línea (L y H) se realizó un GWAS marcador por marcador para el carácter V_E utilizando el mismo modelo usado para analizar los datos de las dos líneas conjuntamente (Ecuación 8). Los ficheros de genotipos y fenotipos para la L V_E y para H V_E se crearon por separado y se generó un fichero con los alelos de referencia usados para el análisis de las líneas conjuntamente. El análisis se realizó con el programa GCTA v1.91.4beta (Yang et al., 2011a)

teniendo en cuenta el uso de los mismos alelos de referencia para cada SNP en ambas líneas. Para cada línea se realizó una lista con todos los SNPs localizados en las regiones genómicas identificadas por el GWAS para las líneas en conjunto. Posteriormente se comprobó si existían diferencias en los efectos de sustitución alélica para esos SNPs comparando los resultados obtenidos para cada línea. Para ello se determinaron los intervalos de confianza generados para el efecto de sustitución alélica de esos SNPs en función de $\pm 2SE$.

2.6.3. GWAS en otra población.

Los mismos análisis de asociación para V_E del TC fueron realizados con genotipos pertenecientes a un experimento de selección divergente por capacidad uterina (UC) durante 12 y de una línea control seleccionada por tamaño de camada al destete durante 14 generaciones (V) (Sosa Madrid and Santacreu, 2016). La base de datos de fenotipos incluyó un total de 4654 registros de partos procedentes de 1413 hembras. Los genotipos usados para los análisis de asociación pertenecen a 184 hembras pertenecientes a la línea de baja UC (85), la línea de alta UC (66) y una población control (V) con 30 animales. Los parámetros estadísticos del carácter V_E del TC de esta población se muestran en la Tabla 2. El tamaño de camada fue pre-correctado por los efectos fijos AE y PL. El fenotipo V_E del TC fue calculado como la varianza del TC dentro de hembra, de la misma forma que indica el apartado “Material genotípico y fenotípico” de esta misma sección.

A partir de las regiones genómicas identificadas en esta población se realizó una comparación con las regiones asociadas a V_E del TC. Además, se determinó si existían diferencias con respecto al efecto de sustitución alélica del SNP más relevante identificado para V_E del TC a partir de los intervalos de confianza generados en función de $\pm 2SE$.

Tabla 2. Análisis descriptivos para el carácter V_E del TC de las líneas para alta capacidad uterina (HUC), baja capacidad uterina (LUC) y una línea control (V).

Línea	N ^(a)	μ ^(b)	M_e ^(c)	CV ^(d)	Min ^(e)	Max ^(f)
HUC	85	4.13	2.84	0.97	0.01	22.75
LUC	66	3.17	2.11	0.91	0.21	10.24
V	30	4.49	3.07	1.06	0.21	20.29

^(a) Número total de datos, ^(b)media, ^(c)mediana, ^(d)coeficiente de variación y ^(e)máximo y ^(f)mínimo del rango de los datos

3. RESULTADOS

3.1. Análisis de asociación del genoma completo (GWAS).

Los dos métodos de asociación (SMR y MMR) identificaron regiones genómicas asociadas a V_E del TC según los umbrales de relevancia estadística fijados (Figura 1a y 1b).

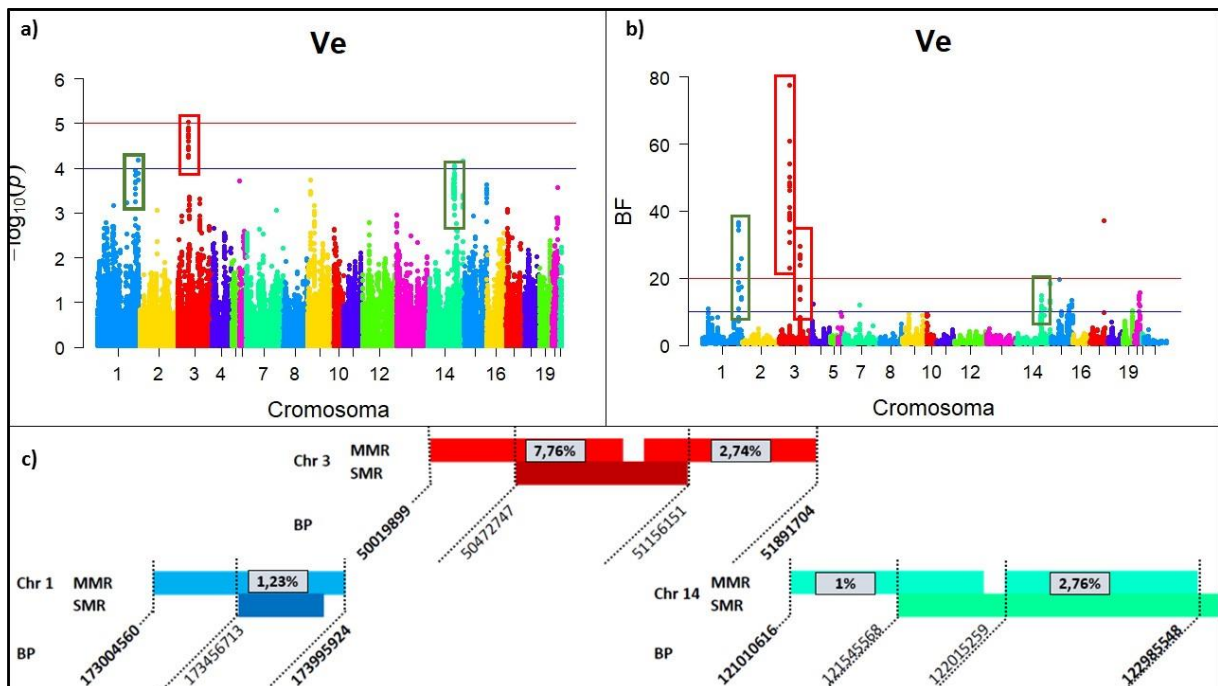


Figura 1. Regiones genómicas asociadas a V_E del tamaño de camada. a) Manhattan plot de regresión marcador por marcador (SMR) con dos umbrales de significación estadística fijados para un $-\log_{10}(p)$ -valor de 4 y 5. b) Manhattan plot por regresión de múltiples marcadores (MMR) con dos umbrales de relevancia estadística fijados en un BF de 10 y 20. En ambos casos los recuadros rojos muestran las regiones con mayor importancia en la asociación (chr 3) y los recuadros verdes las asociaciones en otras regiones importantes (chr 1 y chr 14). c) Ilustración de las regiones asociadas a V_E del tamaño de camada que coinciden en ambos métodos, localizadas en los cromosomas 1, 3 y 14. Se indica tanto su localización en pares de bases (BP) como el porcentaje de varianza genética explicada según MMR.

A través del método SMR se determinaron 3 regiones genómicas con un p-valor menor de $1E-4$ (Figura 1a), valor sugerido en la bibliografía para estudios con poblaciones seleccionadas en producción animal. Teniendo en cuenta el umbral más restrictivo ($1.025E-5$), calculado considerando los SNPs en equilibrio de ligamiento y aplicando la corrección de múltiple test de Bonferroni, se identificó una región en el cromosoma 3 como la más relevante estadísticamente para nuestro carácter (Figura 1a). Esta región, localizada entre la megabase (Mb) 50.4 y 51.2 (Figura 1c) incluye un total de 27 SNPs. Entre ellos, se encontró el SNP Affx-151984290, el más significativo de todos los SNPs incluidos en el análisis (94 282) con un p-

valor de $9.46E-6$. Las regiones genómicas localizadas en el cromosoma 1 y 14 (Figura 1) incluyen 57 y 60 SNPs respectivamente.

El análisis realizado con MMR mostró 9 regiones genómicas asociadas a V_E más que por SMR. Regiones con un $BF > 10$ se identificaron en los cromosomas 1, 3, 4, 7, 14, 15, 17, 19 y 20 (Figura 1b). Regiones con un $BF > 20$ sólo se encontraron en el cromosoma 1, 3 y 17 (Figura 1b). Además, con el análisis MMR se obtuvieron los porcentajes de varianza genética (V_G) explicada por ventanas genómicas de aproximadamente 1 Mb a lo largo de todo el genoma. De todas ellas, las regiones que más porcentaje de V_G explicaban eran ventanas genómicas que se solapaban con parte de las regiones genómicas identificadas también por SMR (Figura 1c), en los cromosomas 1, 3 y 14. Dos ventanas genómicas consecutivas se identificaron en el cromosoma 3 entre la Mb 50 y 52 explicando un 7.76% y un 2.74% de la V_G (Figura 1c) e incluyendo 62 y 47 SNPs, respectivamente. Si estas ventanas se considerasen como una única región, explicaría un 10.50% de la V_G total. Por el método MMR, el SNP con mayor relevancia estadística ($BF = 77.4$) se localizó en la ventana genómica del cromosoma 3 que explicaba un 2.74% de la V_G total (Figura 1c). Este SNP (Affx-151984290) fue también identificado por el método SMR como el SNP más relevante estadísticamente. Las ventanas genómicas identificadas en el cromosoma 1 y 14 (dos regiones consecutivas) explicaban un 1.23% y un 3.78% con un total de 57 y 101 SNPs, respectivamente (Figura 1c). Las demás regiones identificadas con un $BF > 10$ poseen un porcentaje de varianza menor del 1% por lo que se descartaron como zonas candidatas para la realización de posteriores análisis. Finalmente, aquellas regiones que fueron identificadas por ambos métodos se asumieron como regiones asociadas a V_E del TC. Por tanto, se determinaron las siguientes regiones genómicas de estudio para la búsqueda de genes candidatos: 1:173004560-173995924, 14:121010616-123019882 y 3:50019899-51891704.

En cuanto a los estudios GWAS realizados sobre los caracteres con una posible respuesta correlacionada, no se encontraron ninguna región genómica coincidente con las regiones candidatas para V_E del TC, indicadas anteriormente. No obstante, en todos los casos se identificaron regiones genómicas con un p-valor $< 5E-4$ y un $BF > 10$ por SMR y MMR respectivamente (Figura S2-S6). Por MMR también se identificaron ventanas genómicas que explicaban más de un 1% de la V_G para todos los caracteres evaluados (TC, TO, EI, ES y FS). En el caso del TC, dos ventanas genómicas consecutivas en el cromosoma 2 entre la Mb 54 y 56 explicaron un 7.46% de la V_G . Para FS, tres ventanas consecutivas en el cromosoma 13 entre

la Mb 4 y 5 explicaron un 10.97% de V_G . Del mismo modo, la ventana con mayor porcentaje de V_G explicada para ES se localizó en el cromosoma 7 entre la Mb 113 y 114 con un valor de 3.84%. Además, una ventana genómica para este carácter en el cromosoma 1, que explica un 2.86% de V_G coincidió exactamente con la ventana genómica que más porcentaje de V_G explicó para EI (6.72%). Esta ventana se encuentra localizada en el cromosoma 1 entre las pares de bases 193003955 y 193977844. Un resumen de las regiones genómicas encontradas asociadas con los caracteres TC, TO, EI, ES y FS se muestran en la Tabla 3. Para TO, no se asignó ninguna región genómica como asociada. Ninguna de las regiones identificadas para este carácter coincidió con los métodos SMR y MMR. Además, no se obtuvo ninguna ventana genómica con elevado porcentaje de V_G explicada. La máxima ventana obtenida explicó un porcentaje de 1.13% y estaba localizada en el cromosoma 19.

Tabla 3. Regiones genómicas candidatas para los caracteres de tamaño de camada (TC), supervivencia fetal (FS), supervivencia embrionaria (ES) y embriones implantados (EI) según su posición en el cromosoma (Chr) en pares de bases.

Carácter	Chr	Inicio	Fin
TC	2	46048634	46996706
	2	54020450	55998050
	12	125270544	126997876
FS	4	5016033	5993707
	13	4002200	6988431
	14	51028460	51998323
ES	1 ^(a)	193003955	193977844
	7	113062175	113987079
	14	39010763	40996168
EI	1 ^(a)	193003955	193977844
	3	99047991	99515286
	15	72013791	72986567

^a Regiones solapantes para los caracteres ES y EI.

3.2. Búsqueda de genes candidatos.

Un total de 39 genes fueron identificados en las regiones candidatas para V_E del TC: 12 genes localizados en el cromosoma 1, 10 en el cromosoma 3 y 17 en el cromosoma 14 (Tabla S1). De todos ellos, aquellos que presentaron relación con procesos inmunológicos fueron valorados como posibles genes candidatos asociados a V_E del TC. Este criterio se fijó debido a que estudios previos indicaron que los procesos inmunológicos podían estar involucrados en la

resiliencia y la “*fitness*” de los animales procedentes de las líneas utilizadas en este estudio (Argente et al., 2014; Agustín Blasco et al., 2017; García ML et al., 2012). Por tanto, en un primer filtraje, 29 genes (Tabla 4) fueron seleccionados como posibles genes candidatos. No obstante, se realizó un filtraje más restrictivo y se consideró estudiar aquellos genes que tanto en la base de datos GeneCards como en KEGG fueran asignados como genes que participaran en procesos biológicos relacionados con la inmunología y la reducción del estrés oxidativo en las células, a excepción del gen *SLIT3*. Este gen se seleccionó debido a su posible relación con procesos regulados por el cortisol, biomarcador de la respuesta a estrés utilizado en estudios previos (Argente et al., 2014; García ML et al., 2012). Finalmente, 8 genes de los representados en la Tabla 4 se propusieron como genes asociados a V_E del TC debido a su relación con procesos inmunológicos que desencadenan la respuesta inflamatoria celular y a una relación directa con procesos de defensa celular frente al estrés oxidativo (Tabla S2). La respuesta inflamatoria es la primera barrera de defensa ante la entrada de patógenos en el organismo y está relacionada con algunas funciones inmunológicas desarrolladas por los genes identificados como: desarrollo de las células hematopoyéticas, señalización mediada por quimiocinas, migración de leucocitos, formación de ligandos de selectina, activación de las células dendríticas mieloides y mastocitos, respuesta al cortisol, inhibición del complejo de ataque a membrana (MAC) y regulación positiva de la fagocitosis, entre otras (Tabla S2). Por otra parte, el estrés oxidativo desencadenado por especies reactivas del oxígeno (ROS) provoca daños en los ácidos nucleicos, proteínas y lípidos. Las concentraciones de ROS suelen elevarse en los tejidos donde se desarrolla una respuesta inflamatoria ya que es el sistema de defensa que usan los fagocitos para eliminar los patógenos externos.

Tabla 4. Genes localizados en las regiones genómicas asumidas como relevantes según su localización en el cromosoma (Chr) por pares de bases (PB) para V_E del TC.

Chr	PB	Genes
1	173004560-173995924	<i>KIAA1549L</i> , <i>C11orf91</i> , <i>CD59^a</i> , <i>LMO2^a</i> , <i>CAPRINI</i> , <i>NAT10</i> , <i>ABTB2</i> , <i>CAT^(b)</i> , <i>ELF5</i> , <i>EHF</i>
3	50019899-51891704	<i>CPNE3</i> , <i>CNGB3</i> , <i>CNBD1</i> , <i>SLIT3^a</i> , <i>SPDL1</i> , <i>DOCK2^a</i> , <i>FAM196B</i> , <i>FOXH1</i> , <i>LCP2^a</i> , <i>KCNMB1</i> , <i>KCNIP1</i>
14	121545568-122985548	<i>DCBLD2</i> , <i>ST3GAL6^a</i> , <i>CPOX</i> , <i>GPR15^a</i> , <i>CLDND1</i> , <i>GABRR3</i> , <i>ARL6</i> , <i>EPHA6</i>

^a Genes candidatos seleccionados debido a su directa relación con procesos inmunológicos desencadenantes de la respuesta inflamatoria. ^(b) Gen candidato seleccionado debido a su relación directa con procesos involucrados en la reducción del estrés oxidativo en las células.

Con relación a los genes candidatos propuestos, es interesante destacar que el estudio de la localización de los SNPs más relevantes para V_E del TC reveló que el SNP Affx-151984290 se encontraba posicionado en una región intrónica determinada por los exones 23 y 24 del gen DOCK2 (Figura S7). La mayoría de las funciones de este gen están involucradas en procesos inmunológicos desencadenados por la activación de las proteínas Rac1 y Rac2 y relacionadas con la dinámica del citoesqueleto de actina, procesos necesarios para hacer frente a la respuesta inflamatoria (Chen et al., 2018).

La búsqueda de genes candidatos para los caracteres TC, EI, ES y FS se llevó a cabo en las regiones genómicas descritas en la Tabla 3, donde se identificaron un total de 25, 39, 45 y 53 genes respectivamente (Tabla S3). Aquellos genes relacionados con el desarrollo embrionario o la respuesta inmunológica se muestran en la Tabla 5. Entre algunas de sus funciones cabe destacar que *NKX3-1* y *NKX2-6* son genes homeobox relacionados con el desarrollo cardíaco del embrión y del crecimiento de las células endoteliales en los tejidos de la próstata, respectivamente. *AH11* participa en la diferenciación del ectodermo, al igual que *ASTN1*, y *FRAS1* en la organogénesis durante el desarrollo del embrión. En la región genómica coincidente del cromosoma 1 para ES y EI (Tabla 3) se localizó el gen *ZP1*. Este gen codifica para unas glicoproteínas que conforman la zona pelúcida del ovocito, y adquiere funciones importantes durante la fertilización y el desarrollo preimplantacional. Por último, el gen *VEGFC* localizado en una región asociada con el TC, pertenece a la familia VEGF y está involucrado en la angiogénesis vascular del sistema linfático y venoso durante la embriogénesis.

Tabla 5. Genes candidatos identificados para los caracteres de tamaño de camada (TC), supervivencia fetal (FS), supervivencia embrionaria (ES) y embriones implantados (EI)

Caracteres	Genes
TC	<i>NKX3-1</i> ^a , <i>NKX2-6</i> ^a , <i>DDX60</i> ^b , <i>VEGFC</i> ^m , <i>ALDH8A1</i> ^c , <i>HBSIL</i> ^b , <i>MYB</i> ^b , <i>APH11</i> ^a , <i>MAP7</i>
FS	<i>NCOA6</i> ^c , <i>PIGU</i> , <i>ITCH</i> ^b , <i>RALY</i> ^a , <i>NECAB3</i> , <i>E2F1</i> , <i>ASTN1</i> ^a , <i>ABL2</i> ^b , <i>TORIAIP1</i> , <i>CEP350</i> , <i>TIPARP</i> , <i>PTX3</i> ^b
ES	<i>MS4A1</i> ^b , <i>PTGDR2</i> ^b , <i>ZP1</i> ^a , <i>TMEM109</i> , <i>CD3</i> ^b , <i>FOSL1</i> ^b , <i>CDCA7</i> , <i>OLA1</i> , <i>CIR1</i> , <i>PLSCR4</i> ^b
EI	<i>MS4A1</i> ^b , <i>PTGDR2</i> ^b , <i>ZP1</i> ^a , <i>TMEM109</i> , <i>CD3</i> ^b , <i>FOSL1</i> ^b , <i>SEPT11</i> ^b , <i>FRAS1</i> ^a , <i>CXCL13</i> ^b , <i>CCNG2</i> , <i>CNOT6L</i>

^a Genes directamente relacionados con procesos involucrados en el desarrollo embrionario. ^b Genes directamente relacionados con procesos inmunológicos. ^c Genes involucrados en procesos relacionados con el ácido retinoico. ^m Gen de la misma familia que el *VEGFA* determinado en otro estudio como relacionado con V_E del TC (Sell-Kubiak et al., 2015).

3.3. Estudio del desequilibrio de ligamiento.

El estudio de desequilibrio de ligamiento (LD) en las regiones genómicas más relevantes para V_E del TC mostró diferentes patrones de LD para cada región asociada. En el cromosoma 3, entre la Mb 50 y 52 se observa un cambio en el patrón de LD entre la línea de H V_E (Figura 2b) y L V_E (Figura 2a). En la línea de H V_E el bloque de LD es mayor que en la línea de L V_E donde se observa que se ha producido una disminución de las correlaciones entre SNP y diferenciado dos bloques de LD. No obstante, el bloque de LD que contiene los 2 SNPs más relevantes estadísticamente de la región (Affx-151984290 y Affx-151987366) se mantiene en ambas líneas en fuerte desequilibrio de ligamiento. En el cromosoma 14 el patrón de LD es más claro ya que se aprecian dos bloques con elevada correlación entre todos los SNPs (Figura 2d). El mayor de ellos contiene los SNPs con mayor relevancia estadística de la región; Affx-151948159 (BF=14.73), Affx-151855761 (BF=13.72) y Affx-151973522 (BF=13.38) (Figura 2d). Por último, en el gráfico de LD del cromosoma 1 entre la Mb 173 y 174 (Figura 2c) no se distingue ningún bloque claro de LD ya que es una zona predominante por la ausencia de LD entre los SNPs. No obstante, podemos apreciar que los SNPs más relevantes de dicha región, el Affx-151911039 (BF=36.72) y el Affx-151923816 (BF=35.7) se encuentran en la misma zona, a una distancia de aproximadamente 10 kb y en elevado LD.

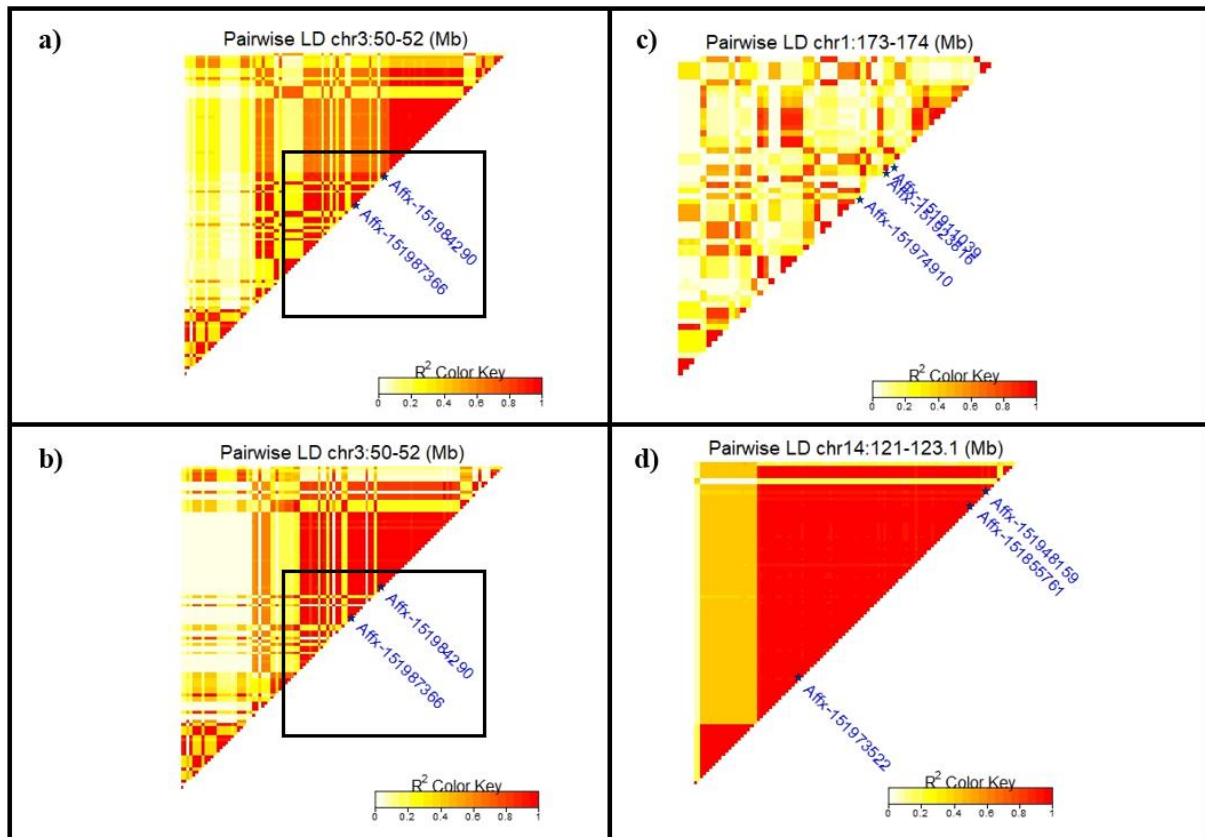


Figura 2. Desequilibrio de ligamiento (r^2) para las regiones genómicas candidatas de V_E del tamaño de camada (TC). a) Representación gráfica del disequilibrio de ligamiento (LD) de la región genómica comprendida entre las megabases (Mb) 50 y 52 del cromosoma 3 del conejo para la línea de baja V_E del TC. b) Representación gráfica del LD de la región genómica comprendida entre la Mb 50 y 52 del cromosoma 3 de la línea de baja V_E del TC. c) LD para la región genómica del cromosoma 1 entre la Mb 173-174 para la línea H V_E y L V_E del TC conjuntamente. d) LD para la región genómica del cromosoma 14 entre la Mb 122-123.1 para la línea H V_E y L V_E del TC conjuntamente. En todas las figuras se indican los SNP más significativos de la región (azul) y en las gráficas a) y b) el recuadro indica la zona en LD entre ambas líneas que no modifica su patrón de LD.

3.4. Validaciones.

El test de permutación y los estudios de asociación dentro de la línea H V_E y L V_E apoyaron las evidencias encontradas a favor de las regiones genómicas asociadas a V_E del TC, en los cromosomas 1, 3 y 14. Mediante el test de permutación se observó que las regiones genómicas con elevada relevancia estadística no se debían a procesos determinados por azar sino a eventos relacionados con la selección realizada para V_E del TC en las líneas. La mayoría de los SNPs localizados en estas regiones mostraron una probabilidad de asociación con la V_E del TC por azar menor al 2% ($P_{perm} < 0.02$). En el cromosoma 3, 72 SNPs de 109 mostraron $P_{perm} < 0.02$, al igual que 23 SNPs de los 57 identificados en el cromosoma 1 y 82 SNPs de los 101 identificados en el cromosoma 14. El valor mínimo de P_{perm} para dichas regiones fue de $1E-5$, $1.10E-4$ y $1.20E-4$, respectivamente. El SNP Affx-151984290, SNP más relevante en el estudio,

mostró un valor de P_{perm} de $1E-5$, por lo que la relevancia estadística observada para este SNP es muy improbable que se deba a eventos generados por azar. No obstante, es probable que la relación genotipo-fenotipo no se haya perdido por completo durante el test, debido al elevado grado de parentesco entre los individuos. Aun así, el test de permutación indicaría que gran parte de las asociaciones detectadas con la V_E del TC se deben a la selección y no al azar.

En los análisis de asociación realizados dentro de cada línea se observó la dirección que seguían los efectos de sustitución alélica de los SNPs en las regiones candidatas. Todos los SNPs, exceptuando algunos localizados en la región del cromosoma 14 entre las pares de bases 121545568 y 122873676 (Tabla S4), seguían la misma dirección. Los SNPs localizados en esta región tienen efectos de sustitución alélica contrarios entre líneas (positivo para H V_E y negativo para L V_E) e intervalos de confianza no solapantes. Esto podría indicar un proceso de selección diferente para la línea de H V_E y la línea de L V_E para esos SNPs (Tabla S4). Un ejemplo de ello es el SNP Affx-151794850 localizado en el cromosoma 14 en la posición 12222613 pb cuyo intervalo para el efecto de sustitución alélica de la línea H V_E ([0.61-2.39]) y L V_E ([-1.69-0.15]) no se solapa (Anexo XI). Las diferencias encontradas podrían deberse a las diferencias de MAF entre las líneas (0.36 para H V_E y 0.08 para L V_E). Además, dentro de la línea L V_E , para estos SNPs sólo hay 1 individuo homocigoto para el alelo de referencia seleccionado. La mayoría de SNPs son homocigotos para el alelo alternativo (84.62%), con alguna proporción de heterocigotos (14.62%).

El estudio de asociación de V_E para el TC en una población seleccionada por capacidad uterina (UC) no replicó los resultados obtenidos en nuestro estudio. Las regiones genómicas asociadas con el carácter en esta población no fueron coincidentes con las obtenidas en la población de estudio. Las regiones genómicas identificadas en esta población se localizaron en las regiones 2:95023603-95974017 y 13:90038780-90996436 (Figura S8). Además, también se comprobó el efecto de sustitución alélica del SNP Affx-151984290 para ver si en ambas poblaciones se había determinado el mismo sentido de selección. Sin embargo, se observó un cambio en la dirección del efecto de sustitución alélica con respecto a nuestra población. El efecto de sustitución alélica en la población de validación fue de -1.28 con un error estándar de 0.63, mientras que en la población de estudio fue de 1.44 con un error de 0.32.

4. DISCUSIÓN

La creación de líneas divergentes para V_E del TC en conejos durante 12 generaciones nos ha permitido disponer de un material genético único para estudiar el control genético de la V_E del TC. Además, con el uso de dos metodologías estadísticas diferentes (SMR y MMR), para la identificación de las regiones genómicas asociadas, hemos podido obtener resultados más fiables por comparación entre ambos métodos. Nótese que la estima de los efectos de los SNPs es dependiente del modelo utilizado (López de Maturana et al., 2014). De esta manera, las regiones genómicas con mayor evidencia de asociación a V_E del TC se identificaron en los cromosomas 1, 3 y 14 del genoma del conejo.

Los efectos de los SNPs obtenidos con el método SMR se sobreestimaron con respecto al modelo BayesB. Esto es normal porque al incluirse un único SNP en cada modelo, este LD produce una sobrestimación del efecto del SNP ya que se captura parte de los efectos de las variantes con los que está correlacionado. Por otra parte, no se observó una inflación de los SNPs asociados con la V_E . Esto indicaría que no hubo factores de confusión. El factor de confusión más habitual es la estratificación de la población. No obstante, una ventaja de esta metodología es la utilización de la GRM para la corrección del modelo por posibles efectos de estratificación de la población que aumenta el error de tipo I. Así mismo, excluir el cromosoma del SNP a testar de la GRM permitió aumentar la potencia del análisis para la detección de más variantes con efectos (Yang et al., 2014).

La aplicación del modelo BayesB (método MMR) permitió definir el grado de ajuste “*shrinkage*” del modelo dependiendo de los datos disponibles. De esta manera, conforme a lo encontrado por Lopez de Maturana et al. (2014), se aumentó la potencia de detección de los SNPs con efecto, respecto al método SMR. En el presente estudio se consideró aproximadamente 30 SNPs con efecto y varianza en el modelo, dado a priori un valor de $\pi=0.9997$. Sin embargo, aunque a través de este método se permite identificar la mayor parte de los SNPs con pequeño y gran efecto asociados al carácter, también se aumenta el número de falsos positivos. La asunción de valores de π distintos genera resultados diferentes bajo los mismos datos (Figura 3) debido a que se modifica el número de SNP considerados con efecto y varianza en el modelo. Por ejemplo, un valor de $\pi=0.999$ (inferior a 0.9997) considera un mayor número de SNPs con efecto y varianza en el modelo (aproximadamente 95 SNPs, si se considera 199 692 SNP totales). En este estudio hemos comprobado que a medida que este

valor de π disminuye, el porcentaje de V_G explicada por las ventanas genómicas se reduce (Figura 3). Esto puede ser debido a que se reparten los efectos de las variables entre un mayor número de parámetros. Por tanto, el uso de un valor de π es dependiente del número total de SNPs, del modelo y del número de datos disponibles para la estima de los efectos de los SNPs. No obstante, a pesar de que el porcentaje de V_G explicada por las ventanas genómicas del estudio se modificaban según el valor de π considerado, el ranking determinado para esas ventanas genómicas se mantenía (Figura 3). Esto podría indicar que dichas regiones son realmente importantes para la determinación de las diferencias fenotípicas observadas sobre el carácter. Cabe destacar que el porcentaje de V_G explicada por las ventanas genómicas debería usarse como valor cualitativo para la comparación entre regiones genómicas relevantes identificadas en los estudios. El uso del porcentaje de V_G explicada como una medida cuantitativa no aporta ninguna información consistente, en este caso, ya que está condicionada por el valor de π .

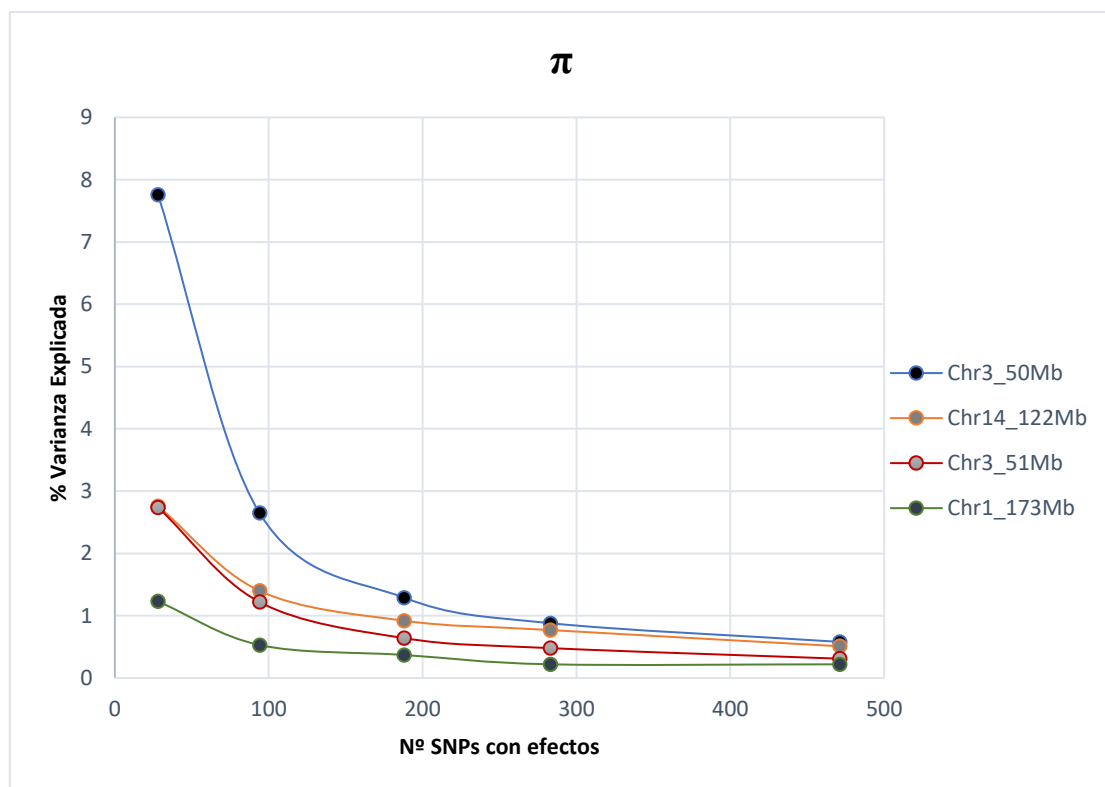


Figura 3. Variación del porcentaje de varianza explicada por las regiones genómicas candidatas en función del valor del número de SNPs con efecto incluido en el modelo. Se observa como el porcentaje de varianza explicada por las regiones genómicas candidatas se ve reducida a medida que aumenta el número de SNPs con efecto y varianza, es decir, a medida que se reduce el valor de π .

Los resultados mostraron que ambos métodos identificaron regiones genómicas para V_E del TC localizadas en; 1:173004560-173995924, 3:50019899-51891704 y 14:121010616-123019882. Analizando los gráficos de LD para estas zonas se podría asumir que las regiones localizadas en el cromosoma 3 y 14 han sido seleccionadas mayoritariamente en ambas líneas, ya que presentan grandes bloques de LD (Figura 2a y 2c). Esto nos indica, unido a la elevada V_G explicada por dichas regiones (10.5% y 3.76% respectivamente), que existe una asociación de estas regiones genómicas con la V_E del TC en esta población. No obstante, la asociación puede producirse porque en estas ventanas se encuentran localizados los vQTLs causales o porque pueden estar en LD con éstos. Para la región localizada en el cromosoma 1, la ausencia de grandes bloques, así como la elevada correlación entre los dos SNPs más significativos de la región (Affx-151923816 y Affx-151974910) (Figura 2c), podría indicarnos que el porcentaje de V_G explicada por toda esa ventana puede estar sólo determinada por estos 2 SNPs. Sin embargo, la existencia de una correlación de estos SNPs con una variante fuera de esta región es posible.

Por otra parte, los estudios realizados para validar los resultados obtenidos refuerzan la hipótesis de la existencia de una asociación entre las regiones genómicas identificadas y el carácter V_E . No obstante, no son resultados determinantes. En el caso del test de permutación, la elevada relación de parentesco entre los individuos hacía difícil la disociación completa de la relación entre el genotipo y el fenotipo del animal. Además, el número de datos es bastante reducido, por lo que, aun aumentando el número de iteraciones del modelo, las combinaciones de genotipo-fenotipos permutados son más reducidas. En el caso de los estudios de asociación realizados dentro de línea, el factor más limitante fue el número de datos disponibles, 130 y 139 para la línea de H V_E y para la línea de L V_E , respectivamente. Además, la mayoría de los SNPs con efectos de sustitución alélica diferentes entre ambas líneas (Tabla S4) poseen un MAF bastante bajo (0.07), que unido al elevado LD entre estos SNPs (Figura 2d) y al reducido número de datos, impide una correcta estima de los efectos de los SNPs a través del método SMR. Aun así, ambas líneas poseen valores positivos en los intervalos de confianza para el efecto de sustitución alélica, lo que podría indicar un mismo sentido de selección en ambas líneas, aunque no se solapen los intervalos. Además, el bajo MAF de la línea para L V_E (0.07) con respecto a la línea para H V_E (0.36) para la región genómica identificada en el cromosoma 14 podría ser un indicativo más de que es una asociación relevante para la regulación del carácter. Esta diferencia en frecuencias alélicas podría deberse a un claro efecto de la selección realizada por V_E del TC sobre esta región.

Un resultado notable para este estudio es que la búsqueda de genes candidatos nos permitió evidenciar que las regiones genómicas escogidas como asociadas a V_E del TC eran zonas involucradas en la regulación de procesos inmunológicos. En estas regiones se identificaron los genes *SLIT3*, *DOCK2*, *ST3GAL6*, *CD59*, *LCP2*, *LMO2* y *GRP15*, involucrados en procesos relacionados con la respuesta inflamatoria (Figura 4). Este tipo de respuesta es desencadenada por el sistema inmune innato y determina una de las primeras líneas de defensa ante patógenos externos (Goldsby et al., 2003). En una de estas regiones (cromosoma 1) también se identificó el gen *CAT* (codificante para la enzima catalasa). Este gen ejecuta una acción antioxidante en el organismo mediante la reducción del daño oxidativo en las células. También actúa a nivel de la respuesta inflamatoria reduciendo los niveles de peróxido de hidrógeno (H_2O_2) generados por las células fagocíticas para hacer frente a la infección (Wang et al., 2013).

Otros resultados interesantes se generaron a partir de la localización de los SNPs relevantes en el cromosoma. El SNP Afx-151984290 (con mayor relevancia estadística obtenida tanto por SMR y MMR) se encuentra dentro de la secuencia genética del gen *DOCK2* en la posición 51156151 pb. Este gen participa en procesos determinantes en la migración de los leucocitos hacia el lugar donde se ha iniciado la respuesta inflamatoria (Gadea and Blangy, 2014). Sin embargo, no podemos afirmar que sea el SNP causal debido a que se encuentra localizado en una secuencia intrónica y en una región genómica donde el LD entre los SNPs es muy elevada. La relevancia estadística observada puede ser debida a su elevado grado de asociación con el carácter o al LD con los SNPs de su región, de los que captura parte de su efecto. El SNP Afx-151923816, identificado en el cromosoma 1 en la posición 173762846 pb, se encuentra localizado dentro de la zona 3' UTR del gen *CAT*, gen con función antioxidante a nivel celular. La localización de este SNP en la región 3'UTR podría indicar que modula los niveles de expresión del ARNm que codifica la enzima Catalasa, puesto que esta zona determina la estabilidad de las moléculas de ARNm (Geissler and Grimson, 2016). Además, se ha visto que este gen está involucrado en la disminución de la inflamación a través de la eliminación de agentes oxidantes usados por las células inmunitarias para combatir las infecciones (Wang et al., 2013).

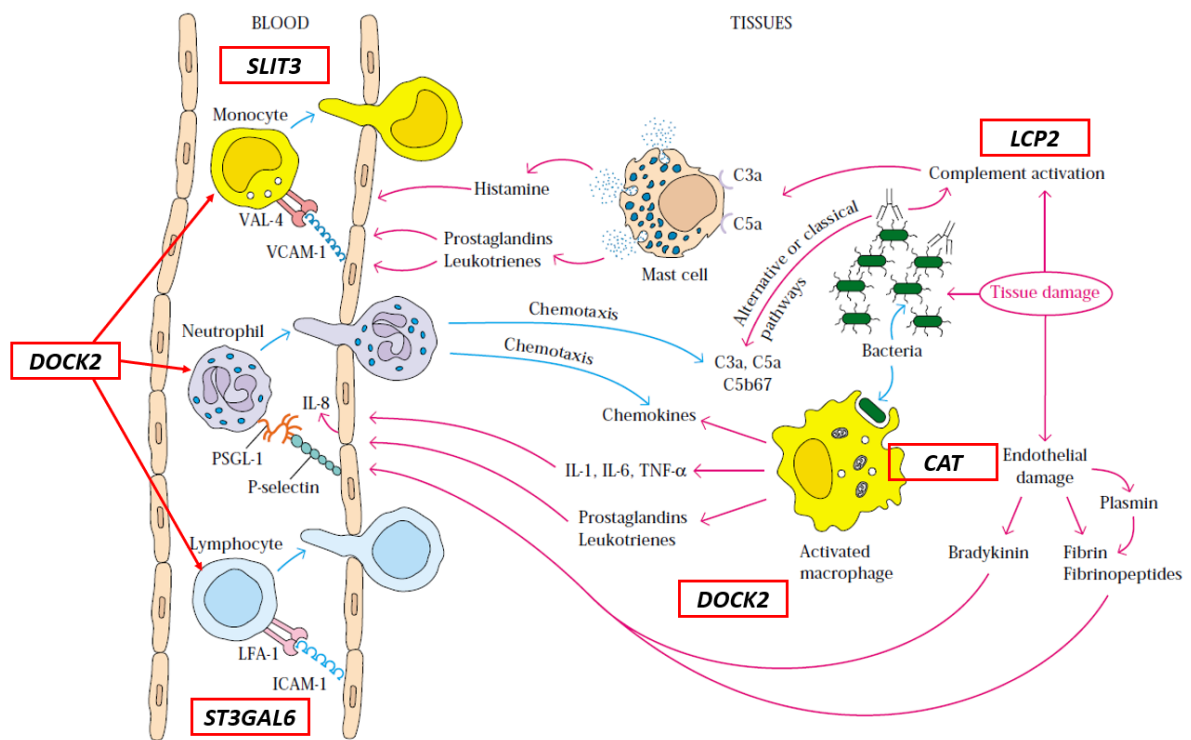


Figura 4. Respuesta inflamatoria inducida por daño en el tejido y la entrada de patógenos. En esta imagen se observan los mecanismos principales desencadenantes de una respuesta inflamatoria. Ante un daño en el tejido el organismo activa 3 rutas principales: la activación del sistema del complemento, la activación de las células fagocíticas y la generación de proteínas de daño endotelial que actuarán potenciando la respuesta inflamatoria. En todos los casos, el objetivo es inducir la migración de los leucocitos hacia el tejido y mitigar la infección (Goldsby et al., 2003). Los genes candidatos involucrados en esta respuesta se encuentran marcados con recuadros rojos y están posicionado donde desarrollan su función.

La relación entre los genes candidatos identificados y la respuesta inflamatoria es de gran relevancia para esta población donde se ha observado que la línea para L V_E del TC responde mejor ante patógenos externos que la línea para H V_E del TC (Argente et al., 2014). La respuesta inflamatoria se caracteriza por la estimulación de una cascada de acontecimientos desencadenados por la presencia de patógenos extraños o daño en el tejido. Estos acontecimientos pueden ser inducidos a través de la activación del complemento, la activación de los macrófagos o la producción de proteínas de fase aguda (como la CRP y SAA) que potencian los dos mecanismos anteriores (Figura 4) (Goldsby et al., 2003). Por tanto, el objetivo principal de esa respuesta es el reclutamiento de leucocitos hacia el tejido donde se ha desarrollado la respuesta inflamatoria para poder hacer frente a la infección (proceso denominado extravasación leucocitaria). Diversos estudios han determinado que los genes *DOCK2*, *ST3GAL6*, *LCP2* y *SLIT3* juegan un papel importante en el desarrollo de esta respuesta primaria (Figura 4) (Chen et al., 2018; Geutskens et al., 2010; Jiang et al., 2018; Munro et al., 1992). En concreto, los genes *DOCK2*, *SLIT3* y *ST3GAL6* se han caracterizado como genes que

actúan a diferentes niveles promoviendo la extravasación leucocitaria gracias a la coordinación de las interacciones necesarias entre estas células y el endotelio de los vasos sanguíneos (Figura 4).

La función principal del gen *DOCK2* (Chen et al., 2018; Gadea and Blangy, 2014) es inducir una polarización de la membrana de los leucocitos, así como una remodelación del citoesqueleto de actina de estos. Así se promueve la extravasación de los monocitos y macrófagos. Además, actúa a nivel de los macrófagos ya que permite su activación a través de mecanismos mediados por el receptor FcγR. Todo ello se lleva a cabo mediante la activación de proteínas GTPasas pequeñas como Rac1 y Rac2 que convierten el GDP en GTP (Gadea and Blangy, 2014). Así mismo, *ST3GAL6* también promueve la extravasación leucocitaria, aunque este gen actúa modificando los ligandos de selectina Sialil-Lewis (Munro et al., 1992) a través de la adición de una molécula de ácido salicílico (Natoni et al., 2016). Este proceso se induce en respuesta a citocinas proinflamatorias como la IL-6 y el TNF-α y permite la interacción de los leucocitos con las células endoteliales de los vasos sanguíneos, permitiendo su avance hasta el lugar donde se desarrolla la infección (Goldsby et al., 2003). El gen *SLIT3* también actúa a este nivel, aunque codificando una proteína de matriz extracelular en las células endoteliales que permitirá, a través de la interacción con proteínas de tipo Robo1, la extravasación leucocitaria a partir de la activación de proteínas RhoA (Geutskens et al., 2010).

Por otro lado, el gen *LCP2* participa en procesos relacionados con la activación del sistema del complemento (Jiang et al., 2018). Este sistema actúa a varios niveles, aunque en términos generales su función es potenciar la respuesta inflamatoria a través de la activación de las células mastocíticas (que producen histamina para permitir la vasodilatación) y de la formación del complejo de ataque a membrana (MAC) (Goldsby et al., 2003). La formación del MAC induce la lisis de las células dianas mediante la formación de poros en su membrana. De esta manera se posibilita la eliminación de complejos inmunes y restos de células muertas fagocitadas (Goldsby et al., 2003). En este sentido es necesario tener un regulador que impida la lisis de células que no deberían ser eliminadas. El gen *CD59* codifica unas glicoproteínas que se localizan en la superficie de la membrana de las células que inhiben el MAC y evitan que se lisen las células normales (Michielsen et al., 2018). Además, también es necesario eliminar los agentes oxidantes libres en las células generadas por las células fagocíticas en respuesta a una activación inducida por la presencia de patógenos. Las células fagocíticas al activarse generan una especie de vesículas denominadas lisosomas que contienen especies reactivas del oxígeno

(ROS) como el peróxido de hidrógeno y el anión superóxido, además de otras especies reactivas intermediarias del nitrógeno. Estos lisosomas se unen con las vesículas que contienen los patógenos fagocitados formando el fagolisosoma donde, a través de las ROS, se procederá a digerir los agentes infecciosos (Goldsby et al., 2003). El contenido digerido por las células fagocíticas es liberado al medio, por lo que es necesario que se desarrolle una función antioxidante para que estas especies no generen daño celular. En este caso, el gen *CAT* (codificante de la enzima catalasa) actúa como agente antioxidante para reducir los elevados niveles de ROS durante la respuesta inflamatoria y evitar así el daño celular (Wang et al., 2013). Por último, tanto el gen *CD59* como el *LMO* actúan regulando la hematopoyesis, es decir, la proliferación y diferenciación de las células que componen el sistema inmunitario. Para ello, *LMO2* actúa a nivel de la fase de replicación del ADN (Fase S) promoviendo la progresión de la división celular (Sincennes et al., 2016). Además, se ha observado que una sobreexpresión de este gen en las células T genera un fenotipo de preleucemia en pacientes (Sincennes et al., 2016).

Los resultados encontrados en este estudio apoyarían la hipótesis de que la V_E del TC puede ser una medida de resistencia frente a las posibles infecciones del día a día generadas por cambios en el microambiente (Blasco et al., 2017). Esta hipótesis concuerda con conclusiones generadas en estudios previos que cuantificaron las diferencias entre las líneas seleccionadas para $H V_E$ y $L V_E$ del TC para los niveles de monocitos, linfocitos y neutrófilos, así como una serie de biomarcadores (CRP, Cortisol y SAA) indicativos del estado fisiológico del animal (García et al., 2012) y que son determinantes en el desarrollo de la respuesta inmunológica (Figura 5) (Goldsby et al., 2003). En estos estudios se observó que la línea de $L V_E$ presentaba unos niveles basales de leucocitos inferiores a la línea de $H V_E$ (García et al., 2012), así como una mejor respuesta a la inoculación de virus atenuados endémicos de la especie (Argente et al., 2014). Todo esto podría indicar que la selección aplicada en conejos para V_E del TC afecta a mecanismos que permiten a los animales adaptarse a las perturbaciones de las condiciones ambientales, lo que es sinónimo de resiliencia (Mulder, 2016). No obstante, estudios funcionales son necesarios para poder validar si realmente la asociación encontrada entre la V_E del TC y la resiliencia de los animales es cierta. Así mismo, otros estudios basados en análisis de secuenciación masiva y perfiles epigenéticos y de expresión también ayudarían a comprender mejor los mecanismos detrás de estas asociaciones.

Los genes candidatos de este estudio no coinciden con genes identificados en regiones genómicas asociadas a la V_E en otras especies (Morgante et al., 2015; Mulder et al., 2013a; Sell-Kubiak et al., 2015; Vandenplas et al., 2013). No obstante, en estos estudios también se han encontrado genes candidatos implicados en el sistema inmunológico. Tanto H.A. Mulder et al. (2013) como Morgante et al. (2005) propusieron como gen candidato los genes HSPCB y Hsp90 respectivamente. Estos genes pertenecen a la familia HSP que responde ante estímulos inducidos por choques térmicos. Esta familia juega un papel importante en el desarrollo del sistema innato en respuesta a estrés celular ya que participa regulando la activación de los leucocitos y protegiendo a las células frente a ROS (Zhang and Mosser, 2009). En otro estudio, también se identificaron regiones candidatas donde se encontraron genes que también regulaban procesos del sistema inmunitario, como la fagocitosis (Vandenplas et al., 2013). Los resultados de este estudio y de los estudios mencionados indicarían que la selección por V_E favorece predominantemente la selección de genes involucrados en el control de la respuesta innata del sistema inmunitario, permitiendo a su vez aumentar las capacidades de los individuos para afrontar los cambios en el ambiente. Las discrepancias entre las regiones genómicas identificadas en cada caso pueden deberse a la especie utilizada o al carácter de estudio. No obstante, también puede deberse a que el número de individuos utilizados en este estudio (269) no nos permita tener la potencia suficiente para que las regiones genómicas donde se localizan genes de la familia HSP se identifiquen como relevantes.

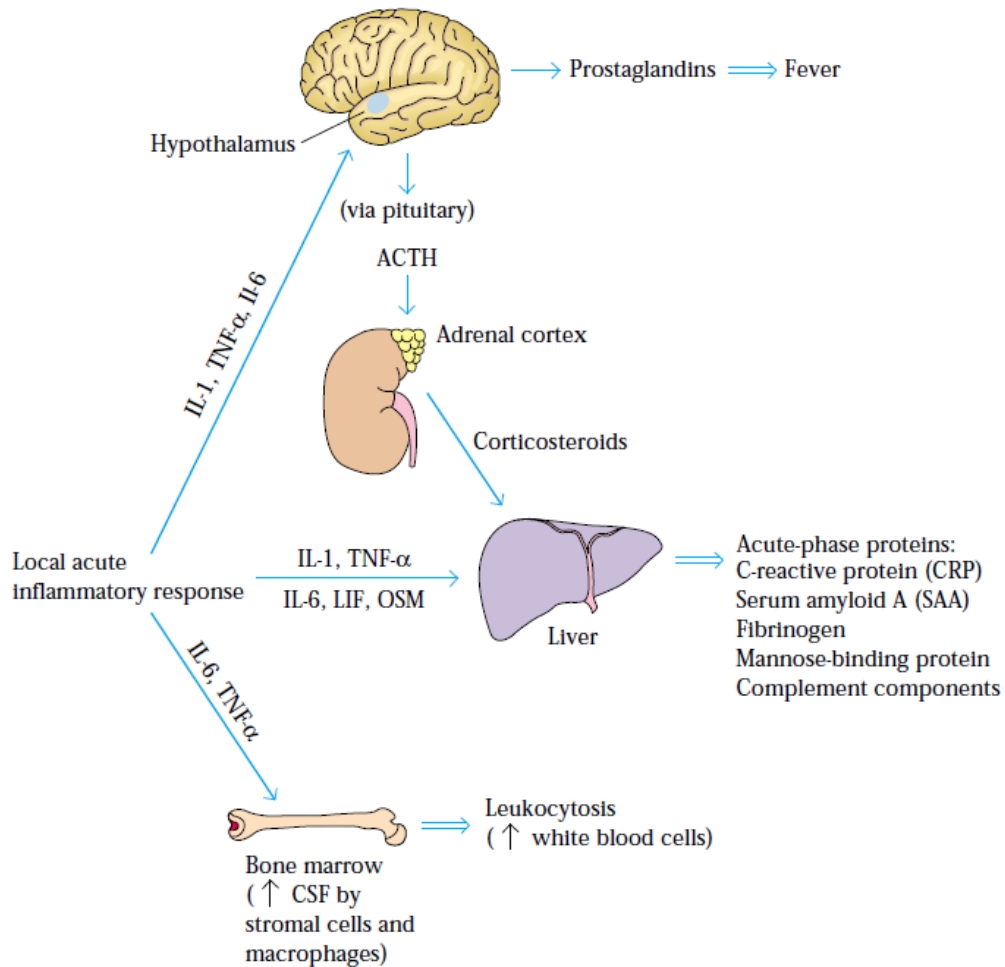


Figura 5. Rutas de señalización activadas por el desarrollo de una respuesta inflamatoria local. La respuesta inflamatoria activa tres rutas a partir de la liberación de citocinas que permitirá amplificar la respuesta y combatir de una forma más efectiva la inflamación. La amplificación de la respuesta inflamatoria se induce a nivel de la médula ósea, por la proliferación de leucocitos, en el hipotálamo, mediante la generación de hormonas, y a través del hígado y la generación de proteínas de fase aguda (Goldsby et al., 2003).

Finalmente, los resultados de los estudios de asociación realizados para caracteres con una posible respuesta correlacionada permitieron identificar regiones genómicas relevantes asociadas a TC, FS, ES y EI (Tabla 3), aunque no eran coincidentes con las identificadas para V_E del TC. En estas regiones se localizaron tanto genes involucrados en el desarrollo embrionario como genes relacionados con algunos de los procesos inmunológicos determinados por el sistema inmune innato (Tabla 5). Entre todos los resultados obtenidos, cabe destacar que para EI y ES se observó una región solapante localizada en el cromosoma 1 en la región comprendida entre las pares de base 193003955 y 193977844. Además, en ambos casos el SNP con mayor relevancia estadística es el mismo (Affx-151898544) y se encuentra localizado dentro de la región genómica coincidente (Tabla 3) en la posición 193660523 pb. Este SNP presentó una relevancia estadística de un $BF=197.68$ y $BF=331.99$ para ES y EI

respectivamente. Estos resultados concuerdan con las respuestas observadas por Argente et al. (2017), ya que los caracteres con mayor respuesta correlacionada con la V_E del TC (EI, ES y TC) son los que poseen asociaciones estadísticas más relevantes (Argente et al., 2017). Esto se debe a que una mayor divergencia entre las líneas permite una mayor potencia estadística para detectar asociaciones.

La búsqueda de genes candidatos para los caracteres correlacionados identificaron genes relacionados tanto con la respuesta inmunológica como con el desarrollo embrionario. No obstante, caben destacar el gen *ZP1* y el gen *VEGFC*. La función desarrollada por el gen *ZP1* (en la región 1:193003955-193977944) se ha relacionado con el mantenimiento de la integridad de la zona pelúcida del ovocito (Epifano et al., 1995). Mutaciones en un gen de la misma familia (*ZP3*) se han relacionado con defectos en la maduración de los ovocitos e infertilidad en las hembras (Chen et al., 2017). El gen *VEGFC* (en la región 2:54020450-55998050 pb) pertenece a la familia de los VEGF, familia implicada en el desarrollo del endotelio vascular y que se ha visto que cumple funciones importantes en la vasculogénesis y angiogénesis del útero materno, así como en el desarrollo embrionario. Concretamente, un estudio realizado en ratones observó que el gen *VEGFC* jugaba un papel importante en la etapa preimplantacional del desarrollo embrionario (D4-D7) regulando el éxito de la implantación de los blastocistos en el útero materno (Saikia et al., 2016). Tanto *ZP1* como *VEGFC* podrían ser genes responsables de la supervivencia, crecimiento y desarrollo del embrión, estableciendo el éxito del embarazo. En un estudio realizado para la identificación de vQTL para la varianza ambiental en el tamaño de camada en cerdos se observó una asociación del gen *VEGFA*, perteneciente a la misma familia que el gen *VEGFC* (VEGF), el cual está implicado en el desarrollo de las conexiones vasculares de la placenta (Sell-Kubiak et al., 2015). Resultados expuestos por Argente et al. (2017) muestran que las diferencias en el TC entre las líneas de H y L V_E del TC son establecidas en la etapa preimplantacional del desarrollo embrionario, lo que concuerda con las funciones que desarrollan los genes asociados a EI, ES y TC. Además, un estudio de Burkus et al. (2015) observó que la exposición de hembras a condiciones de estrés en la etapa preimplantacional podría ocasionar una inhibición de la adhesión del blastocisto al endometrio debido a un retraso del proceso de “*hatching*”. Este proceso se caracteriza por la rotura de la zona pelúcida y la salida del embrión, por lo que el gen *ZP1* tiene un papel relevante en este proceso y su actuación podría verse modificada debido a cambios en las condiciones ambientales de la hembra, durante el periodo preimplantacional.

5. CONCLUSIÓN

El estudio de asociación realizado ha revelado regiones genómicas asociadas a la varianza ambiental del tamaño de camada. Genes como *DOCK2*, *CAT*, *SLIT3*, *LCP2*, *LMO2* y *CD59* han sido localizados en las regiones asociadas a V_E del TC. Todos ellos presentan funciones importantes durante el desarrollo de una respuesta inflamatoria en el organismo. Diferencias en la actividad de estos genes podrían indicar diferencias en las capacidades de los animales para hacer frente a las nuevas condiciones del ambiente. Por tanto, las diferencias de resiliencia de las líneas utilizadas en este estudio podrían deberse a una regulación diferente del sistema inmunitario a nivel de la respuesta inflamatoria, primera línea de defensa frente a agentes infecciosos. Por otro lado, se identificaron regiones genómicas asociadas a caracteres reproductivos como el tamaño de camada, el número de embriones implantados y la supervivencia embrionaria y fetal. Las regiones que mostraron mayor relevancia estadística fueron aquellos caracteres que presentaron mayor respuesta correlacionada con la V_E del TC. Genes como el *ZPI* y el *VEGFC* fueron destacados debido a su relación con procesos involucrados en la fase de preimplantación del embrión, ya que podrían explicar las diferencias en estos caracteres entre las líneas para alta y baja V_E del TC. En esta etapa se ha visto que el animal es muy susceptible a su estado fisiológico y a las condiciones de estrés al que este expuesto. Los resultados mostrados en este estudio podrían indicar que la selección por V_E del TC modificaría tanto la resiliencia como caracteres reproductivos de los animales a través de un control del sistema inmunológico. No obstante experimentos funcionales son necesarios para corroborarlo.

6. REFERENCIAS

- Argente, M.J., García, M., Muelas, R., Blasco, A., 2013. Divergent selection for residual variance of litter size in rabbits. *Proc. 10th World Rabbit Congr. Genet. Appl. to Livest. Prod.* 56.
- Argente, M.J., García, M.L., Zbynovska, K., Petruska, P., Capcarova, M., Blasco, A., 2014. Effect of Selection for Residual Variance of Litter Size on Hematology Parameters as Immunology indicators in Rabbits. *Proceedings, 10th World Congr. Genet. Appl. to Livest. Prod.*
- Argente, M.J., Calle, E.W., García, M.L., Blasco, A., 2017. Correlated response in litter size components in rabbits selected for litter size variability. *J. Anim. Breed. Genet.* 134, 505–511. <https://doi.org/10.1111/jbg.12283>
- Bell, C.G., Finer, S., Lindgren, C.M., Wilson, G.A., Rakyán, V.K., Teschendorff, A.E., Akan, P., Stupka, E., Down, T.A., Prokopenko, I., Morison, I.M., Mill, J., Pidsley, R., Deloukas, P., Frayling, T.M., Hattersley, A.T., McCarthy, M.I., Beck, S., Hitman, G.A., 2010. Integrated genetic and epigenetic analysis identifies haplotype-specific methylation in the FTO type 2 diabetes and obesity susceptibility locus. *PLoS One* 5, e14040. <https://doi.org/10.1371/journal.pone.0014040>
- Blasco, A., 2017. *Bayesian Data Analysis for Animal Scientists. The Basics.* Springer Nature. <https://doi.org/10.1007/978-3-319-54274-4>
- Blasco, A., Martínez-Álvarez, M., García, M.L., Capcarova, M., Zbynovska, K., Petruska, P., Ibáñez-Escriche, N., Argente, M.J., 2017. Selection for Genetic Environmental Sensitivity of Litter Size Changes Resilience in Rabbits. *Proc. World Congr. Genet. Appl. to Livest. Prod.* 11.89.
- Blasco, A., Martínez-Álvarez, M., García, M.L., Ibáñez-Escriche, N., Argente, M.J., 2017. Selection for environmental variance of litter size in rabbit. *Genet. Sel. Evol.* 49, 48. <https://doi.org/10.1186/s12711-017-0323-4>
- Broom, D.M., 2008. Welfare assessment and relevant ethical decisions: Key concepts. *Annu. Rev. Biomed. Sci.* 10, 79–90. <https://doi.org/10.5016/1806-8774.2008.v10pT79>
- Browning, B.L., Browning, S.R., 2016. Genotype Imputation with Millions of Reference Samples. *Am. J. Hum. Genet.* 98, 116–126. <https://doi.org/10.1016/j.ajhg.2015.11.020>
- Burton, P.R., Clayton, D.G., Cardon, L.R., Craddock, N., Deloukas, P., Duncanson, A., Kwiatkowski, D.P., McCarthy, M.I., Ouwehand, W.H., Samani, N.J., Todd, J.A., Donnelly, P., Barrett, J.C., Davison, D., Easton, D., Evans, D., 2007. Genome-wide

- association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447, 661–678. <https://doi.org/10.1038/nature05911>
- Bush, W.S., Moore, J.H., 2012. Chapter 11: Genome-Wide Association Studies. *PLoS Comput. Biol.* 8, e1002822. <https://doi.org/10.1371/journal.pcbi.1002822>
- Carneiro, M., Afonso, S., Geraldes, A., Garreau, H., Bolet, G., Boucher, S., Tircazes, A., Queney, G., Nachman, M.W., Ferrand, N., 2011. The genetic structure of domestic rabbits. *Mol. Biol. Evol.* 28, 1801–1816. <https://doi.org/10.1093/molbev/msr003>
- Chang, C.C., Chow, C.C., Tellier, L.C.A.M., Vattikuti, S., Purcell, S.M., Lee, J.J., 2015. Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience* 4, 1–16. <https://doi.org/10.1186/s13742-015-0047-8>
- Chen, T., Bian, Y., Liu, X., Zhao, S., Wu, K., Yan, L., Li, M., Yang, Z., Liu, H., Zhao, H., Chen, Z.J., 2017. A Recurrent Missense Mutation in ZP3 Causes Empty Follicle Syndrome and Female Infertility. *Am. J. Hum. Genet.* 101, 459–465. <https://doi.org/10.1016/j.ajhg.2017.08.001>
- Chen, Y., Meng, F., Wang, B., He, L., Liu, Y., Liu, Z., 2018. Dock2 in the development of inflammation and cancer. *Eur. J. Immunol.* 00, 1–8. <https://doi.org/10.1002/eji.201747157>
- De los Campos, G., Hickey, J.M., Pong-Wong, R., Daetwyler, H.D., Calus, M.P.L., 2013. Whole-genome regression and prediction methods applied to plant and animal breeding. *Genetics* 193, 327–345. <https://doi.org/10.1534/genetics.112.143313>
- Epifano, O., Liang, L.F., Dean, J., 1995. Mouse Zp1 encodes a zona pellucida protein homologous to egg envelope proteins in mammals and fish. *J. Biol. Chem.* 270, 27254–27258. <https://doi.org/10.1074/jbc.270.45.27254>
- Falconer, D., Mackay, T.F.C., 1996. *Introduction to Quantitative Genetics*, 4th ed. Edinburgh Gate, Harlow, U.K. Addison Wesley Longman Limited.
- Falconer, D.S., 1990. Selection in different environments: Effects on environmental sensitivity (reaction norm) and on mean performance. *Genet. Res.* 56, 57–70. <https://doi.org/10.1017/S0016672300028883>
- Formoso-Rafferty, N., Cervantes, I., Ibáñez-Escriche, N., Gutiérrez, J.P., 2016. Genetic control of the environmental variance for birth weight in seven generations of a divergent selection experiment in mice. *J. Anim. Breed. Genet.* 133, 227–237. <https://doi.org/10.1111/jbgs.12174>
- Gadea, G., Blangy, A., 2014. Dock-family exchange factors in cell migration and disease. *Eur. J. Cell Biol.* 93, 466–477. <https://doi.org/10.1016/j.ejcb.2014.06.003>

- García, M.L., Blasco, A., Argente, M.J., 2016. Embryologic changes in rabbit lines selected for litter size variability. *Theriogenology* 86, 1247–1250.
<https://doi.org/10.1016/j.theriogenology.2016.04.065>
- García ML, Argente MJ, Muelas R, Birlanga V, Blasco A, 2012. Effect of Divergent Selection for Residual Variance of Litter Size on Health Status and Welfare. *Proc. 10th World Rabbit Congr.* 103–106.
- Garreau, H., Bolet, G., Larzul, C., Robert-Granié, C., Saleil, G., SanCristobal, M., Bodin, L., 2008. Results of four generations of a canalising selection for rabbit birth weight. *Livest. Sci.* 119, 55–62. <https://doi.org/10.1016/j.livsci.2008.02.009>
- Garrick, D.J., Fernando, R.L., 2013. Genome-Wide Association Studies and Genomic Prediction. *Methods Mol. Biol.* 1019, 275–298. <https://doi.org/10.1007/978-1-62703-447-0>
- Geissler, R., Grimson, A., 2016. A position-specific 3' UTR sequences that accelerates mRNA decay 13, 1075–1077. <https://doi.org/10.1080/15476286.2016.1225645>
- Geutskens, S.B., Hordijk, P.L., van Hennik, P.B., 2010. The chemorepellent Slit3 promotes monocyte migration. *J. Immunol.* 185, 7691–7698.
<https://doi.org/10.4049/jimmunol.0903898>
- Gilmour, A.R., Gogel, B.J., Cullis, B.R., Welham, S.J., Thompson, R., 2015. ASReml User Guide Release 4.1 Structural Specification. VSN Int. Ltd, Hemel Hempstead, HP1 1ES, UK.
- Goldsby, R.A., Kindt, T.J., Osborne, B.A., Kuby, J., 2003. *Immunology*, 5th ed. W.H Freeman and Company.
- Gondro, C., Werf, J. Van Der, Hayes, B., 2013. Genome-Wide Association Studies and Genomic Prediction, in: *Methods in Molecular Biology*. Media Springer Science+ Business LLC. <https://doi.org/10.1007/978-1-62703-447-0>
- Hill, W.G., 1984. On selection among groups with heterogeneous variance. *Anim. Prod.* 39, 473–477.
- Hill, W.G., Mulder, H.A., 2010. Genetic analysis of environmental variation. *Genet. Res. (Camb)*. 92, 381–395. <https://doi.org/10.1017/S0016672310000546>
- Ibáñez-Escriche, N., García, M., Sorensen, D., 2010. GSEVM v.2: MCMC software to analyze genetically structured environmental variance models. *J. Anim. Breed. Genet.* 127, 249–251. <https://doi.org/10.1111/j.1439-0388.2009.00846.x>
- Ibáñez-Escriche, N., Moreno, A., Nieto, B., Piqueras, P., Salgado, C., Gutiérrez, J.P., 2008a. Genetic parameters related to environmental variability of weight traits in a selection

- experiment for weight gain in mice; signs of correlated canalised response. *Genet. Sel. Evol.* 40, 279–293. <https://doi.org/10.1186/1297-9686-40-3-279>
- Ibáñez-Escriche, N., Sorensen, D., Waagepetersen, R., Blasco, A., 2008b. Selection for environmental variation: A statistical analysis and power calculations to detect response. *Genetics* 180, 2209–2226. <https://doi.org/10.1534/genetics.108.091678>
- Ibáñez-Escriche, N., Varona, L., Sorensen, D., Noguera, J.L., 2008. A study of heterogeneity of environmental variance for slaughter weight in pigs. *Animal* 2, 19–26. <https://doi.org/10.1017/S1751731107001000>
- Jiang, H., Dong, L., Gong, F., Gu, Y., Zhang, H., Fan, D., Sun, Z., 2018. Inflammatory genes are novel prognostic biomarkers for colorectal cancer. *Int. J. Mol. Med.* 42, 1–13. <https://doi.org/10.3892/ijmm.2018.3631>
- Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., Morishima, K., 2017. KEGG: New perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* 45, D353–D361. <https://doi.org/10.1093/nar/gkw1092>
- Kass, R.E., Raftery, A.E., 1995. Bayes Factors. *J. Am. Stat. Assoc.* 90, 773–795.
- Kilpeläinen, T.O., Qi, L., Brage, S., Sharp, S.J., Sonestedt, E., Demerath, E., Ahmad, T., Mora, S., Kaakinen, M., Sandholt, C.H., Holzapfel, C., Autenrieth, C.S., Hyppönen, E., Cauchi, S., He, M., Kutalik, Z., Kumari, M., Stančáková, A., 2011. Physical activity attenuates the influence of FTO variants on obesity risk: A meta-analysis of 218,166 adults and 19,268 children. *PLoS Med.* 8, e1001116. <https://doi.org/10.1371/journal.pmed.1001116>
- Knijnenburg, T.A., Wessels, L.F.A., Reinders, M.J.T., Shmulevich, I., 2009. Fewer permutations, more accurate P-values. *Bioinformatics* 25, 161–168. <https://doi.org/10.1093/bioinformatics/btp211>
- Lallias, D., Quillet, E., Bégout, M.-L., Aupérin, B., Khaw, H.L., Millot, S., Valotaire, C., Kernéis, T., Labbé, L., Prunet, P., Dupont-Nivet, M., 2017. Genetic variability of environmental sensitivity revealed by phenotypic variation in body weight and (its) correlations to physiological and behavioral traits. *PLoS One* 12, e0189943. <https://doi.org/10.1371/journal.pone.0189943>
- Lee, Y., Nelder, J.A., 2006. Double hierarchical generalized linear models. *Appl. Stat.* 55, 139–185.
- Legarra, A., Varona, L., López de Maturana, E., 2008. TM Threshold model. *Geno Toul Bioinforma.*
- López de Maturana, E., Ibáñez-Escriche, N., González-Recio, Ó., Marenne, G., Mehrban, H.,

- Chanock, S.J., Goddard, M.E., Malats, N., 2014. Next generation modeling in GWAS: comparing different genetic architectures. *Hum. Genet.* 133, 1235–1253.
<https://doi.org/10.1007/s00439-014-1461-1>
- Marjanovic, J., Mulder, H.A., Khaw, H.L., Bijma, P., 2016. Genetic parameters for uniformity of harvest weight and body size traits in the GIFT strain of Nile tilapia. *Genet. Sel. Evol.* 48:41, 1–10. <https://doi.org/10.1186/s12711-016-0218-9>
- McCarthy, M.I., Abecasis, G.R., Cardon, L.R., Goldstein, D.B., Little, J., Ioannidis, J.P.A., Hirschhorn, J.N., 2008. Genome-wide association studies for complex traits: Consensus, uncertainty and challenges. *Nat. Rev. Genet.* 9, 356–369.
<https://doi.org/10.1038/nrg2344>
- Meuwissen, T.H.E., Hayes, B.J., Goddard, M.E., 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157, 1819–1829. <https://doi.org/11290733>
- Michielsen, L.A., Budding, K., Drop, D., van de Graaf, E.A., Kardol-Hoefnagel, T., Verhaar, M.C., van Zuilen, A.D., Otten, H.G., 2018. Reduced expression of membrane complement regulatory protein CD59 on leukocytes following lung transplantation. *Front. Immunol.* 8, 1–9. <https://doi.org/10.3389/fimmu.2017.02008>
- Morgante, F., Sørensen, P., Sorensen, D.A., Maltecca, C., Mackay, T.F.C., 2015. Genetic architecture of micro-environmental plasticity in *Drosophila melanogaster*. *Sci. Rep.* 5, 09785. <https://doi.org/10.1038/srep09785>
- Mulder, H.A., 2016. Genomic selection improves response to selection in resilience by exploiting genotype by environment interactions. *Front. Genet.* 7, 1–11.
<https://doi.org/10.3389/fgene.2016.00178>
- Mulder, H.A., Bijma, P., Hill, W.G., 2008. Selection for uniformity in livestock by exploiting genetic heterogeneity of residual variance. *Genet. Sel. Evol.* 40, 37–59.
<https://doi.org/10.1051/gse:2007034>
- Mulder, H.A., Crump, R.E., Calus, M.P.L., Veerkamp, R.F., 2013a. Unraveling the genetic architecture of environmental variance of somatic cell score using high-density single nucleotide polymorphism and cow data from experimental farms. *J. Dairy Sci.* 96, 7306–7317. <https://doi.org/10.3168/jds.2013-6818>
- Mulder, H.A., Rashidi, H., 2017. Selection on resilience improves disease resistance and tolerance to infections. *J. Anim. Sci.* 95, 3346–3358.
<https://doi.org/10.2527/jas.2017.1479>
- Mulder, H.A., Rønnegård, L., Fikse, W.F., Veerkamp, R.F., Strandberg, E., 2013b. Estimation of genetic variance for macro- and micro-environmental sensitivity using double

- hierarchical generalized linear models. *Genet. Sel. Evol.* 45, 23.
<https://doi.org/10.1186/1297-9686-45-23>
- Munro, J.M., Lo, S.K., Corless, C., Robertson, M.J., Lee, N.C., Barnhill, R.L., Weinberg, D.S., Bevilacqua, M.P., 1992. Expression of sialyl-Lewis X, an E-selectin ligand, in inflammation, immune processes, and lymphoid tissues. *Am. J. Pathol.* 141, 1397–1408.
- Natoni, A., Macauley, M.S., O'Dwyer, M.E., 2016. Targeting Selectins and Their Ligands in Cancer. *Front. Oncol.* 6, 93. <https://doi.org/10.3389/fonc.2016.00093>
- Pigliucci, M., 1992. Modelling phenotypic plasticity. I. Linear and higher-order effects of dominance, drift, environmental frequency and selection on a one-locus, two-allele model. *J. Genet.* 71, 135–150. <https://doi.org/10.1007/BF02927893>
- R Core Team, 2017. R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing. <https://doi.org/10.1007/978-3-540-74686-7>
- Rappaport, N., Fishilevich, S., Nudel, R., Twik, M., Belinky, F., Plaschkes, I., Stein, T.I., Cohen, D., Oz-Levi, D., Safran, M., Lancet, D., 2017. Rational confederation of genes and diseases: NGS interpretation via GeneCards, MalaCards and VarElect. *Biomed. Eng. Online* 16, 72. <https://doi.org/10.1186/s12938-017-0359-2>
- Ronnegard, L., Felleki, M., Fikse, F., Mulder, H., Strandberg, E., 2010. Genetic heterogeneity of residual variance - estimation of variance components using double hierarchical generalized linear models. *Genet. Sel. Evol.* 42, 8.
- Ros, M., Sorensen, D., Waagepetersen, R., Dupont-nivet, M., Sancristobal, M., Bonnet, J., Mallard, J., 2004. Evidence for Genetic Control of Adult Weight Plasticity in the Snail *Helix aspersa*. *Genetics* 168, 2089–2097. <https://doi.org/10.1534/genetics.104.032672>
- Saikia, P.J., Das, D., Mize, D., Das, M., Sarma, H.N., 2016. Spatiotemporal expression of Vascular Endothelial Growth Factor-C in mice fetal-maternal tissues during periimplantation (D4–D7). *Middle East Fertil. Soc. J.* 22, 115–124.
<https://doi.org/10.1016/j.mefs.2016.10.001>
- SanCristobal-Gaudy, M., Elsen, J.M., Bodin, L., Chevalet, C., 1998. Prediction of the response to a selection for canalisation of a continuous trait in animal breeding. *Genet. Sel. Evol.* 30, 423–451. <https://doi.org/10.1051/gse:19980502>
- Sell-Kubiak, E., Duijvesteijn, N., Lopes, M.S., Janss, L.L.G., Knol, E.F., Bijma, P., Mulder, H.A., 2015. Genome-wide association study reveals novel loci for litter size and its variability in a Large White pig population. *BMC Genomics* 16, 1049.
<https://doi.org/10.1186/s12864-015-2273-y>
- Shin, J.-H., Blay, S., Graham, J., McNeney, B., 2006. LDheatmap : An R Function for

- Graphical Display of Pairwise Linkage Disequilibria Between Single Nucleotide Polymorphisms. *J. Stat. Softw.* 16, Code Snippet 3. <https://doi.org/10.18637/jss.v016.c03>
- Sincennes, M.-C., Humbert, M., Grondin, B., Lisi, V., Veiga, D.F.T., Haman, A., Cazaux, C., Mashtalir, N., Affar, E.B., Verreault, A., Hoang, T., 2016. The LMO2 oncogene regulates DNA replication in hematopoietic cells. *Proc. Natl. Acad. Sci.* 113, 1393–1398. <https://doi.org/10.1073/pnas.1515071113>
- Sorensen, D., Waagepetersen, R., 2003. Normal linear models with genetically structured residual variance heterogeneity : a case study. *Genet. Res. Cambridge* 82, 207–222. <https://doi.org/10.1017/S0016672303006426>
- Sorensen, P., de los Campos, G., Morgante, F., Mackay, T.F.C., Sorensen, D., 2015. Genetic Control of Environmental Variation of Two Quantitative Traits of *Drosophila melanogaster* Revealed by Whole-Genome Sequencing. *Genetics* 201, 487–497. <https://doi.org/10.1534/genetics.115.180273>
- Sosa Madrid, B.S., Santacreu, M.A., 2016. Análisis Genómico de Líneas Divergentes Seleccionadas para Capacidad Uterina en Conejo, in: Universitat Politècnica de València (Ed.), Máster Internacional En Mejora Genética Animal y Biotecnología de La Reproducción.
- Strandberg, E., Felleki, M., Fikse, W.F., Franzén, J., Mulder, H.A., Rönnegård, L., Urioste, J.I., Windig, J.J., 2013. Statistical tools to select for robustness and milk quality. *Adv. Anim. Biosci.* 4:3, 606–611. <https://doi.org/10.1017/S2040470013000162>
- Vandenplas, J., Bastin, C., Gengler, N., Mulder, H.A., 2013. Genetic variance in micro-environmental sensitivity for milk and milk quality in Walloon Holstein cattle. *J. Dairy Sci.* 96, 5977–5990. <https://doi.org/10.3168/jds.2012-6521>
- Vukcevic, D., 2009. Bayesian and Frequentist Methods and Analyses of Genome-Wide Association Studies. St Hugh's Coll. Dep. Stat. Univ. Oxford D.Phil. Th, 1–240.
- Waagepetersen, R., Ibáñez-Escriche, N., Sorenson, D., 2008. A comparison of strategies for Markov chain Monte Carlo computation in quantitative genetics. *Genet. Sel. Evol.* 40, 161–176. <https://doi.org/10.1051/gse:2007042>
- Waddington, C.H., 1942. Canalization of development and the inheritance of acquired characters. *Nature* 150, 563–565. <https://doi.org/10.1038/150563a0>
- Wakchaure, R., Ganguly, S., Praveen, P., 2016. Genotype X Environment Interaction in Animal Breeding: A Review. *Biodivers. Conserv. Chang. Clim.* Chapter 3, 60–73.
- Wakefield, J., 2009. Bayes factors for Genome-wide association studies: Comparison with P-values. *Genet. Epidemiol.* 33, 79–86. <https://doi.org/10.1002/gepi.20359>

- Wang, C., Yue, X., Lu, X., Liu, B., 2013. The role of catalase in the immune response to oxidative stress and pathogen challenge in the clam. *Fish & Shellfish Immunol.* 34, 91–99. <https://doi.org/10.1016/j.fsi.2012.10.013>
- Wijga, S., Bastiaansen, J.W.M., Wall, E., Strandberg, E., de Haas, Y., Giblin, L., Bovenhuis, H., 2012. Genomic associations with somatic cell score in first-lactation Holstein cows. *J. Dairy Sci.* 95, 899–908. <https://doi.org/10.3168/jds.2011-4717>
- Wolf, J.B., Brodie, E.D., Cheverud, J.M., Moore, A.J., Wade, M.J., 1998. Evolutionary consequences of indirect genetic effects. *Trends Ecol. Evol.* 13, 64–69. [https://doi.org/10.1016/S0169-5347\(97\)01233-0](https://doi.org/10.1016/S0169-5347(97)01233-0)
- Wray, N.R., Yang, J., Hayes, B.J., Price, A.L., Goddard, M.E., Visscher, P.M., 2013. Pitfalls of predicting complex traits from SNPs. *Nat. Rev. Genet.* 14, 507–515. <https://doi.org/10.1038/nrg3457>
- Wu, R.L., 1997. Genetic control of macro-and micro-environmental sensitivities in *Populus*. *Theor. Appl. Genet.* 94, 104–114. <https://doi.org/10.1007/s001220050388>
- Yang, J., Lee, S.H., Goddard, M.E., Visscher, P.M., 2011a. GCTA: A tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* 88, 76–82. <https://doi.org/10.1016/j.ajhg.2010.11.011>
- Yang, J., Loos, R.J.F., Powell, J.E., Medland, S.E., Elizabeth, K., Chasman, D.I., Rose, L.M., Thorleifsson, G., Mägi, R., Waite, L., Smith, A.V., Laura, M., 2012. FTO genotype is associated with phenotypic variability of body mass index. *Nature* 490, 267–272. <https://doi.org/10.1038/nature11401>.
- Yang, J., Weedon, M.N., Purcell, S., Lettre, G., Estrada, K., Willer, C.J., Smith, A. V., Ingelsson, E., O’Connell, J.R., Mangino, M., Mägi, R., Madden, P.A., Heath, A.C., Nyholt, D.R., Martin, N.G., Montgomery, G.W., Frayling, T.M., Hirschhorn, J.N., McCarthy, M.I., Goddard, M.E., Visscher, P.M., 2011b. Genomic inflation factors under polygenic inheritance. *Eur. J. Hum. Genet.* 19, 807–812. <https://doi.org/10.1038/ejhg.2011.39>
- Yang, J., Zaitlen, N.A., Goddard, M.E., Visscher, P.M., Price, A.L., 2014. Advantages and pitfalls in the application of mixed-model association methods. *Nat. Genet.* 46, 100–106. <https://doi.org/10.1038/ng.2876>
- Yi, G., Liu, W., Li, J., Zheng, J., Qu, L., Xu, G., Yang, N., 2014. Genetic analysis for dynamic changes of egg weight in 2 chicken lines. *Poult. Sci.* 93, 2963–2969. <https://doi.org/10.3382/ps.2014-04178>
- Zare, Y., Shook, G.E., Collins, M.T., Kirkpatrick, B.W., 2014. Genome-wide association

analysis and genomic prediction of *Mycobacterium avium* subspecies *paratuberculosis* infection in US Jersey cattle. *PLoS One* 9, e88380.

<https://doi.org/10.1371/journal.pone.0088380>

- Zerbino, D.R., Achuthan, P., Akanni, W., Amode, M.R., Barrell, D., Bhai, J., Billis, K., Cummins, C., Gall, A., Girón, C.G., Gil, L., Gordon, L., Haggerty, L., Haskell, E., Hourlier, T., Izuogu, O.G., Janacek, S.H., Juettemann, T., 2018. Ensembl 2018. *Nucleic Acids Res.* 46, Database issue, D754-D761. <https://doi.org/10.1093/nar/gkx1098>
- Zhang, X., Mosser, D., 2009. Macrophage activation by endogenous danger signals. *J. Pathol* 214, 161–178. <https://doi.org/10.1002/path.2284>
- Zhang, X.S., Hill, W.G., 2005. Evolution of the environmental component of the phenotypic variance: Stabilizing selection in changing environments and the cost of homogeneity. *Evolution* (N. Y). 59, 1237–1244. <https://doi.org/10.1111/j.0014-3820.2005.tb01774.x>

7. GLOSARIO

CD: moléculas de la superficie de leucocitos y plaquetas que son distinguibles con anticuerpos monoclonales y puede ser usado para diferenciar las poblaciones celulares.

CÉLULAS DENDRÍTICAS: células presentadoras de antígeno a los linfocitos T.

CITOCINAS: proteínas producidas por las células en respuesta a una gran variedad de estímulos y que son capaces de alterar de alguna manera el comportamiento de otras células. La naturaleza de las células sobre las que ejercen su efecto viene determinada por la presencia de receptores específicos. Estos pueden localizarse en la superficie de las células que las producen (efecto autocrino), de células vecinas (efecto paracrino) o en otros órganos y tejidos (efecto semejante a las hormonas).

COMPLEJO DE ATAQUE A LA MEMBRANA: complejo formado por los factores del complemento C5b-C9 que se insertan en las membranas celulares permitiendo la creación de poros e induciendo la lisis celular.

COMPLEMENTO: grupo de proteínas séricas involucradas en el control de la inflamación, activación de fagocitos y del complejo de ataque a membrana.

CORTISOL: hormona corticoide que segregan las glándulas suprarrenales de muchos mamíferos en respuesta a estrés.

ESPECIES REACTIVAS DEL OXÍGENO: moléculas pequeñas derivadas del oxígeno altamente reactivas debido a la presencia de electrones desapareados (no compensados) que producen daño celular.

ESTRÉS OXIDATIVO: estado de desequilibrio celular debido a la producción de especies reactivas del oxígeno y la capacidad del organismo de hacer frente a ellos.

EXTRAVASACIÓN LEUCOCITARIA: migración de los leucocitos al lugar donde se ha desarrollado una respuesta inflamatoria para mitigar la infección.

FAGOCITOSIS: proceso por el cual ciertas células y organismos unicelulares capturan y digieren partículas nocivas o alimento.

GTP: molécula energética similar al ATP necesarias para el funcionamiento de ciertas enzimas y que tras su utilización se convierte en GDP.

HISTAMINA: amina idazólica liberada por mastocitos y basófilos e involucrada en las respuestas locales del sistema inmunitario. También regula funciones normales en el estómago y actúa como neurotransmisor en el sistema nervioso central.

INMUNIDAD: Estado de capacidad de defensa de un individuo en principio sensible frente a sustancias antigénicas, adquirido de forma activa o pasiva.

LEUCOCITOS: comúnmente conocidos como glóbulos blancos. Son células producidas por la médula ósea, que son responsables de las defensas del sistema inmunológico contra las agresiones externas, como las bacterias o virus infecciosos.

LINFOCITOS: Célula móvil perteneciente a la línea linfoide con gran núcleo y escaso citoplasma. Realiza múltiples funciones en los mecanismos de defensa inmunológica. Se distinguen dos tipos: los T o timodependientes y los B dependientes de la "Bursa" o de la médula ósea.

NEUTRÓFILOS: célula fagocítica circulante perteneciente a la línea mieloide. Es la primera célula del sistema inmunitario en llegar al sitio de inflamación

MACRÓFAGO: Leucocito mononuclear que interviene en la captación, transformación y presentación del antígeno a los linfocitos inmunocompetentes y que posee capacidad fagocítica.

MASTOCÍSTOS: Célula presente sobre todo en el tejido conectivo que posee en su citoplasma histamina, serotonina y heparina. Tras la fijación de anticuerpos tipo IgE a la membrana y subsiguiente reacción con el antígeno específico, liberan estas sustancias para favorecer la migración de leucocitos.

MIGRACIÓN LEUCOCITARIA: distribución de los leucocitos del linaje mieloide (principalmente monocitos y neutrófilos) desde su sitio de maduración en la médula ósea, hacia los lugares donde hay lesión o infección. Los linfocitos migran desde sus sitios de maduración hacia los órganos linfoides secundarios donde se diferenciarán a linfocitos efectores tras la presentación de antígenos. Una vez activados, migrarán al sitio de inflamación para llevar a cabo su función protectora

MONOCITOS: Fagocito mononuclear circulante. Tras el inicio de una respuesta inflamatoria migra al tejido afectado y se diferencia a macrófago.

OPSONIZACIÓN: Acción facilitadora de la fagocitosis por la que macrófagos y polimorfonucleares neutrófilos presentan en su membrana receptores (CR1, CR3 y probablemente CR4) capaces de unir la molécula C3b y sus derivados de manera que si el C3b está fijado sobre la superficie de un germen, los fagocitos pueden conectar con éste mediante sus receptores para C3b facilitándose la fagocitosis.

PROTEÍNA AMILOIDE SÉRICA A (SAA): es una proteína de fase aguda y bajo peso molecular, unida a lipoproteínas de alta densidad (HDL) en el plasma. Es capaz de aglutinar y de opsonizar bacterias, así como activar el complemento por lo que se incluye dentro de los mecanismos de defensa inespecíficos.

PROTEÍNA C REACTIVA (CRP): b-Globulina análoga a los anticuerpos que se encuentra en el suero de pacientes con inflamaciones agudas. Es una proteína de fase aguda. Es capaz de aglutinar y de opsonizar bacterias, así como activar el complemento por lo que se incluye dentro de los mecanismos de defensa inespecíficos.

PROTEÍNAS DE FASE AGUDA: proteínas que incrementan de forma importante su concentración durante una infección o proceso inflamatorio.

SELECTINA: glicoproteínas que se encuentran en la superficie de la mayoría de las células y median la adhesión célula a célula o la adhesión de la célula con la matriz extracelular. Actúan como receptores y fluctúan entre estados de alta y baja afinidad con sus respectivos ligandos, determinándose la especificidad para cada molécula de adhesión.

8. DATOS SUPLEMENTARIOS

Figura S1: Gráfico de componentes principales que muestra la estratificación de la población a través del PC1 y el PC2.

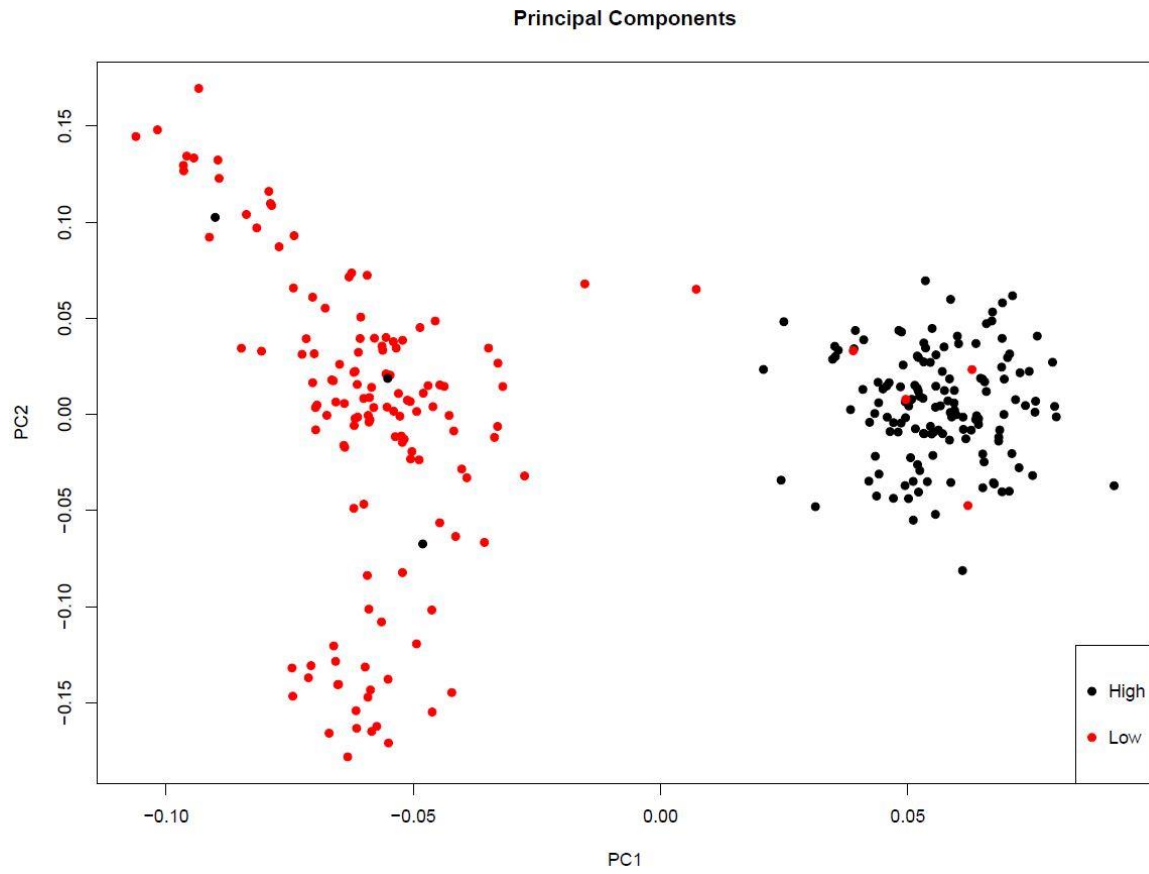


Figura S2: Análisis del genoma completo para el tamaño de camada por regresión marcador a marcador y regresión con múltiples marcadores.

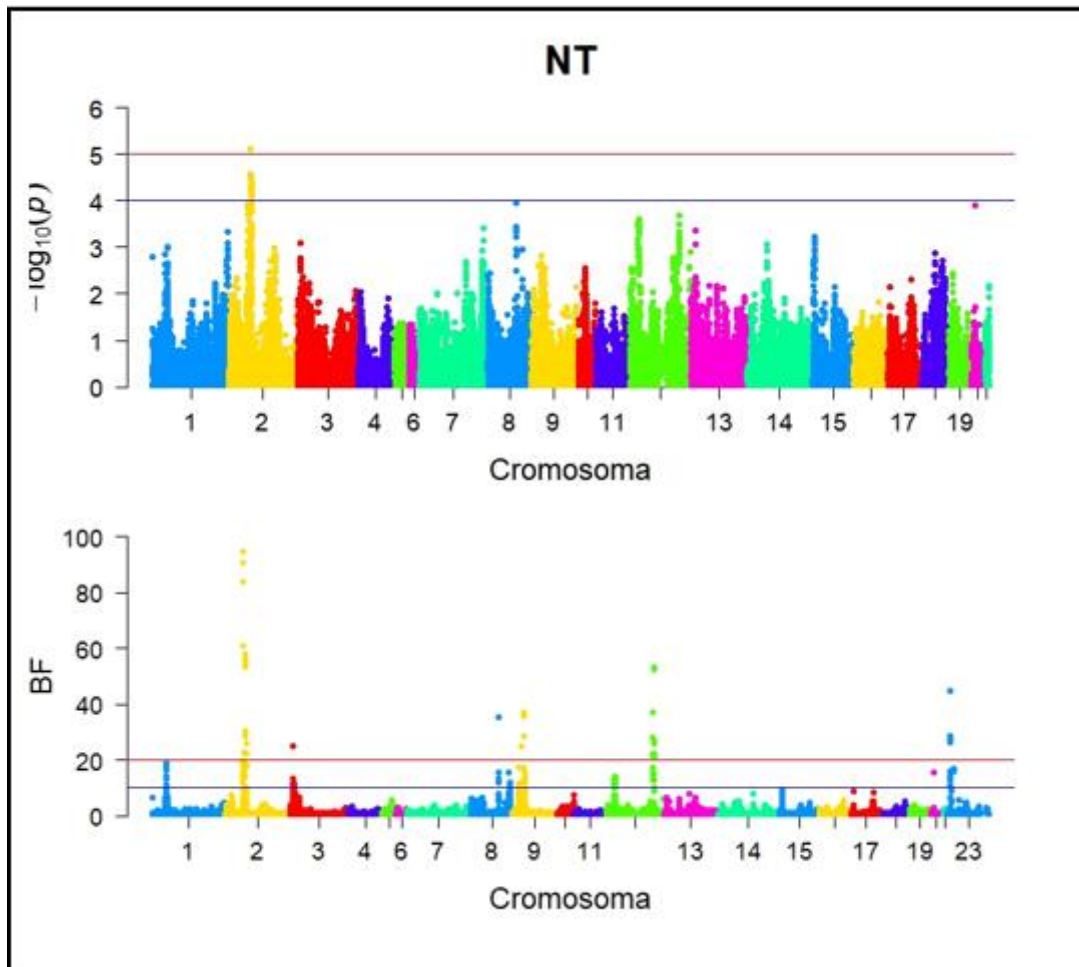


Figura S3: Análisis del genoma completo para el número de embriones implantados por regresión marcador a marcador y regresión con múltiples marcadores.

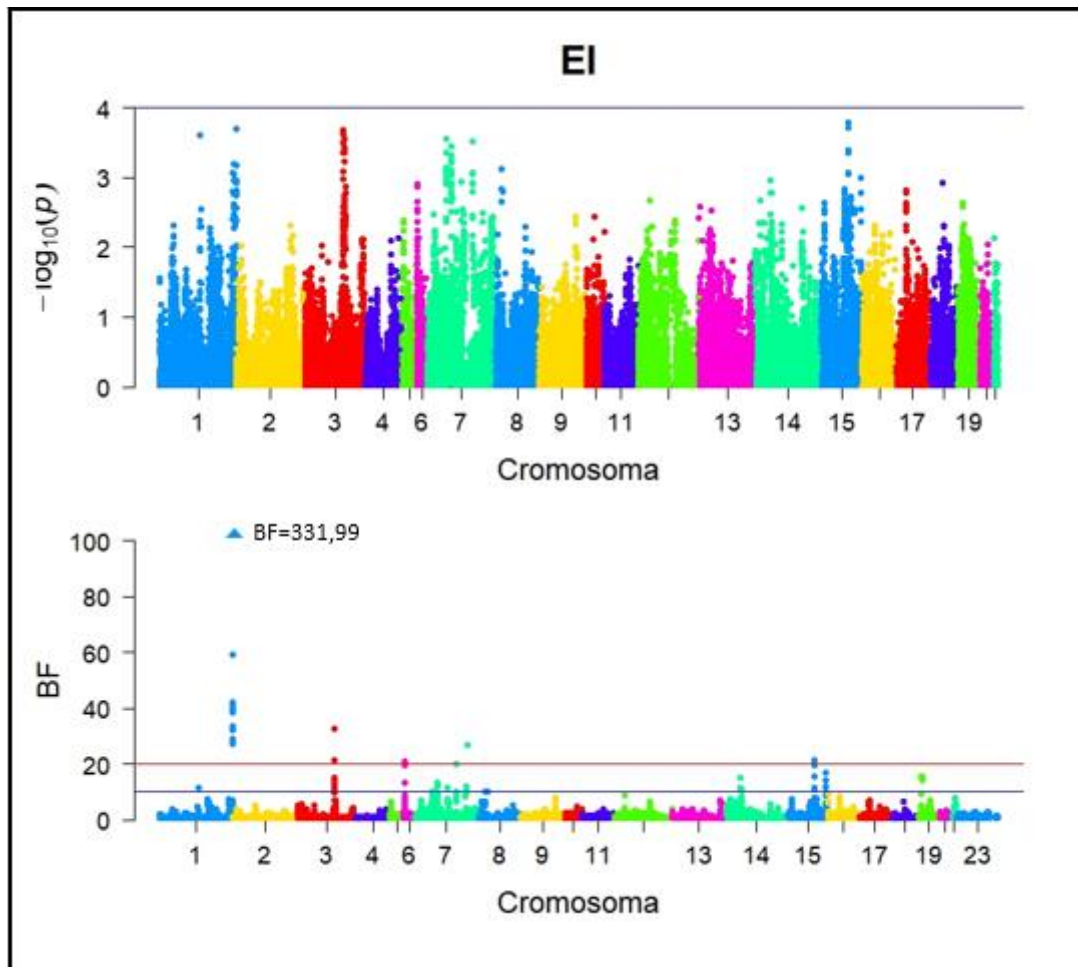


Figura S4: Análisis del genoma completo para la supervivencia embrionaria por regresión marcador a marcador y regresión con múltiples marcadores.

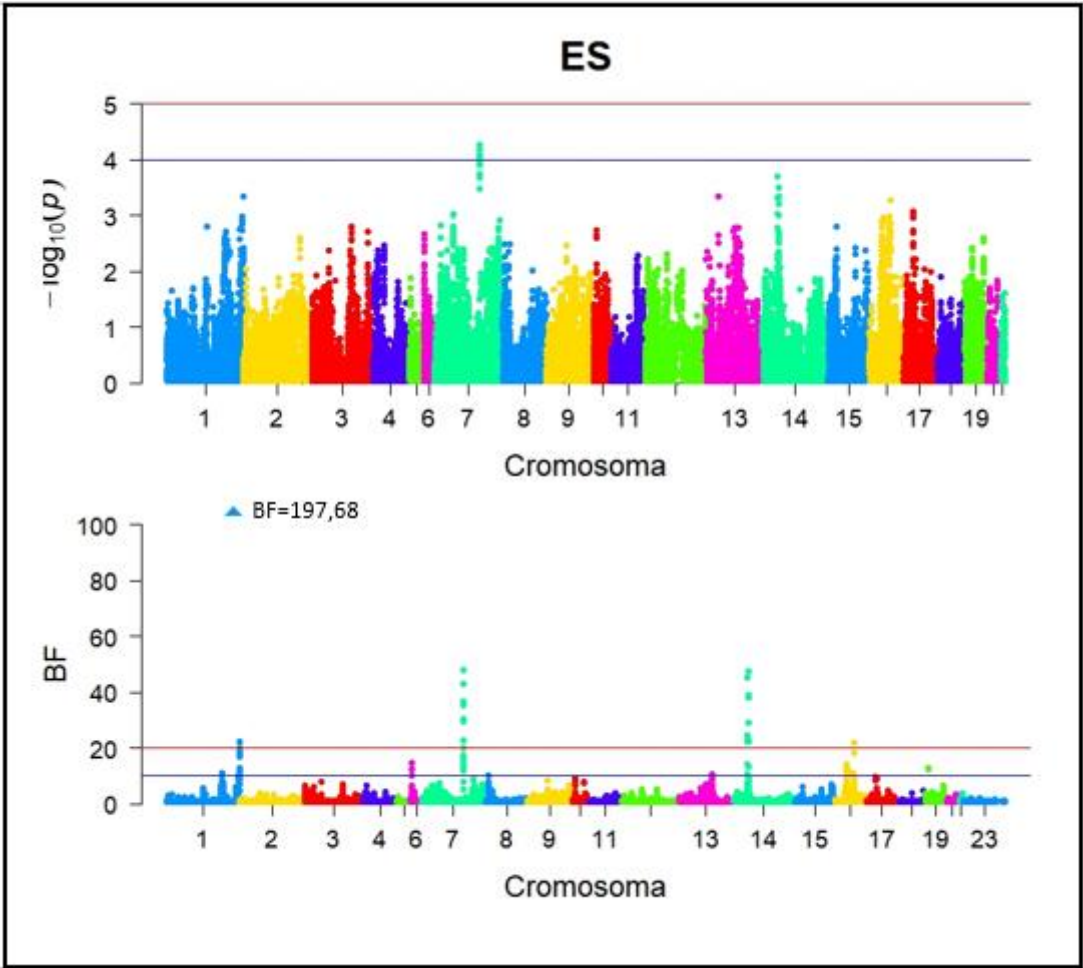


Figura S5: Análisis del genoma completo para la supervivencia fetal por regresión marcador a marcador y regresión con múltiples marcadores.

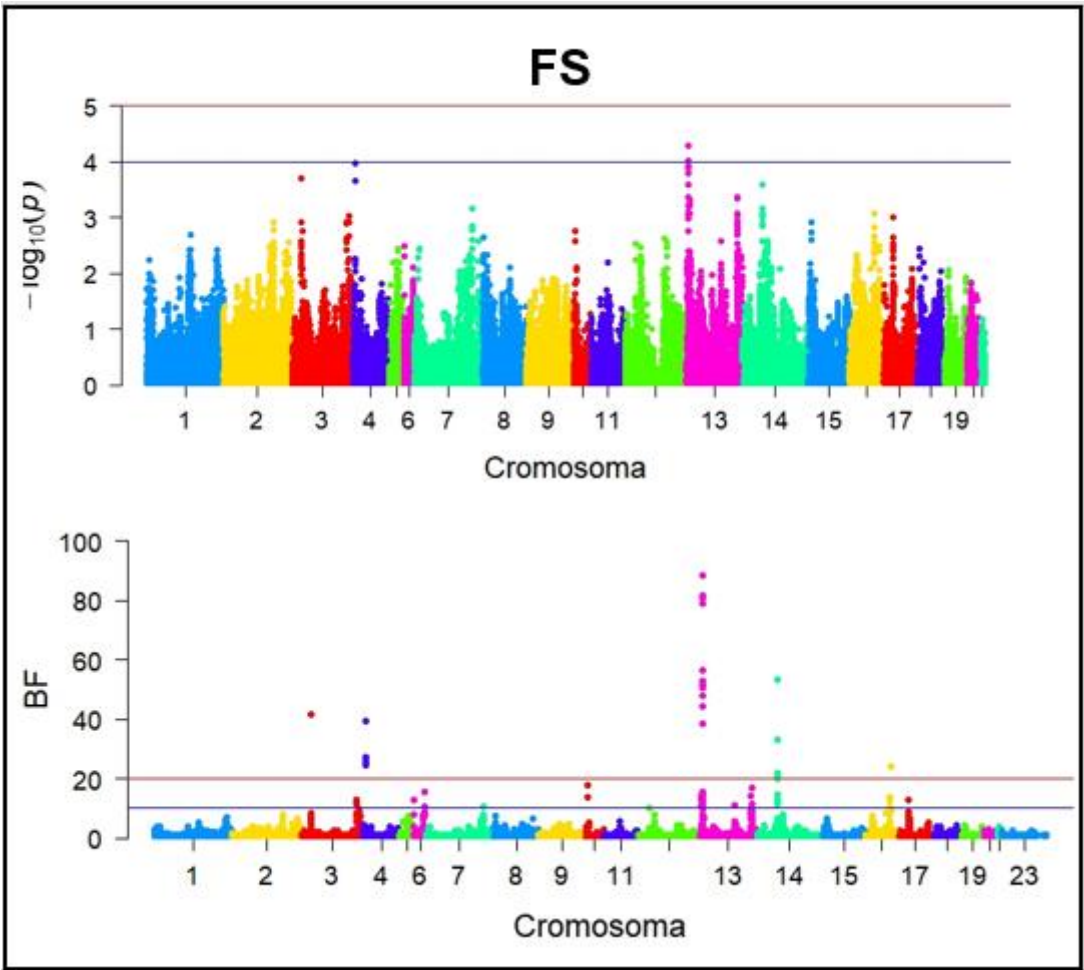


Figura S6: Análisis del genoma completo para la tasa de ovulación por regresión marcador a marcador y regresión con múltiples marcadores.

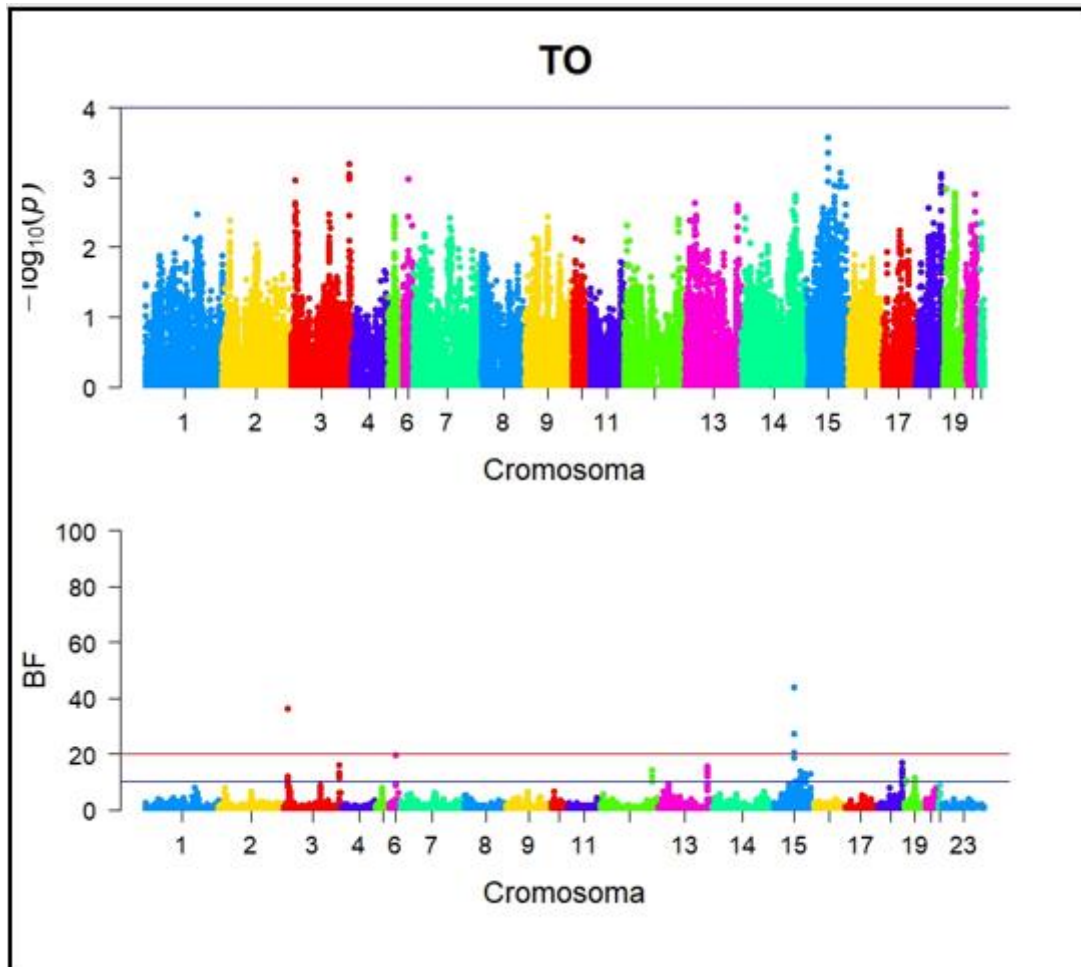


Figura S7: Representación gráfica del posicionamiento aproximado del SNP Afx-151984290 dentro del gen *DOCK2*.

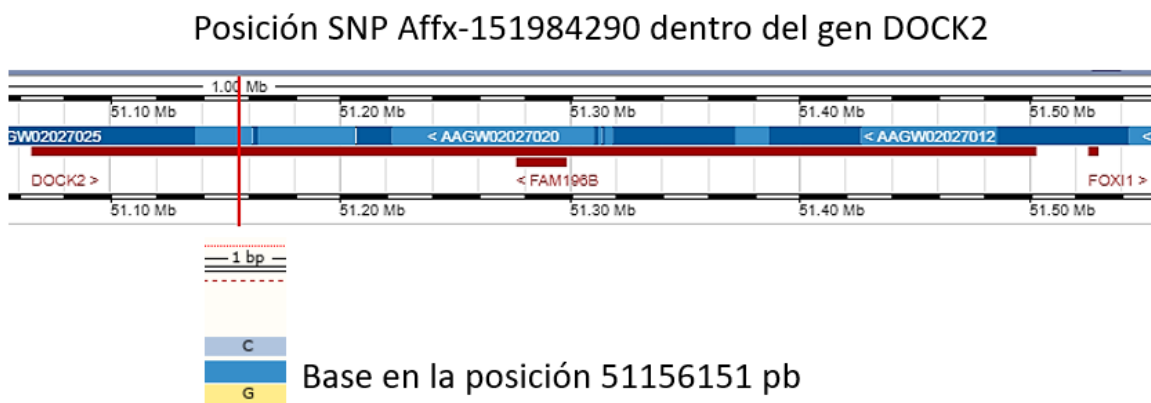


Figura S8: Análisis del genoma completo para la tasa de ovulación por regresión marcador a marcador y regresión con múltiples marcadores.

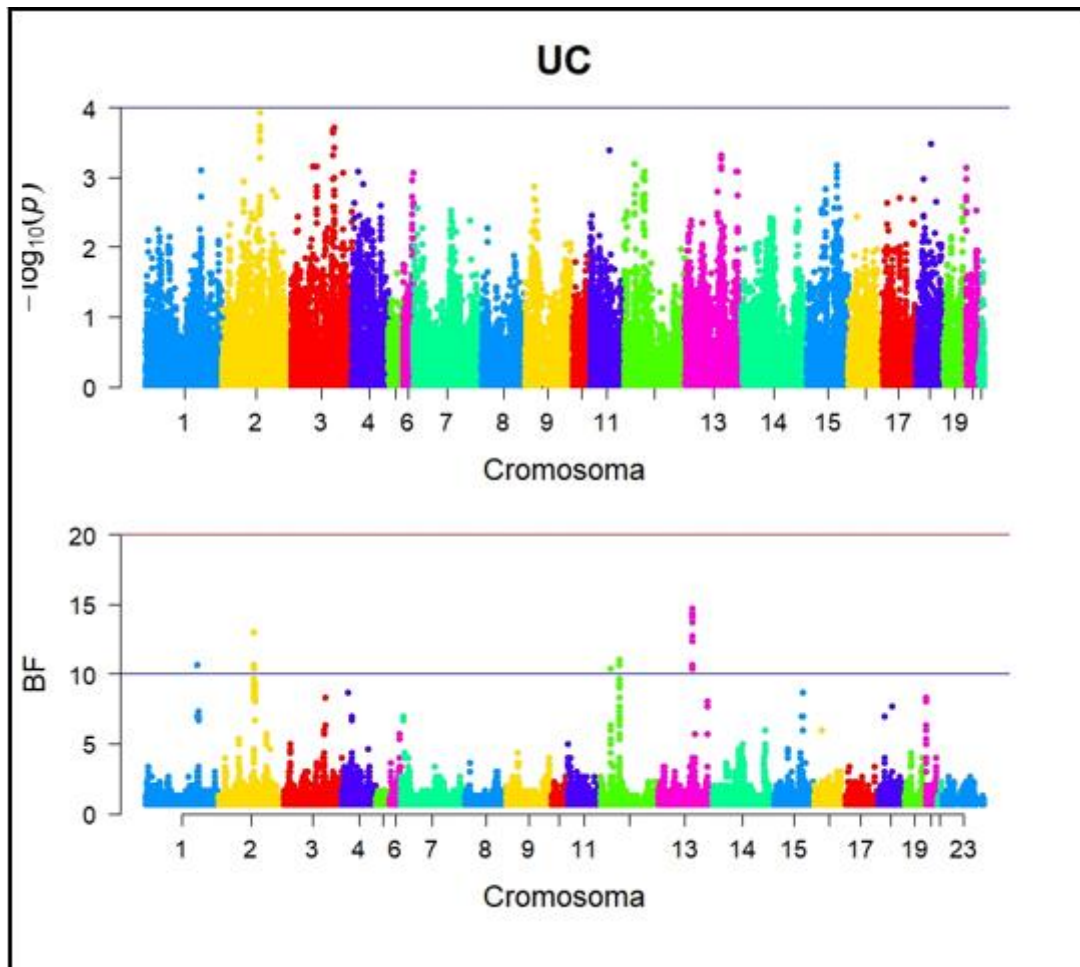


Tabla S1: Genes localizados en las regiones genómicas asociadas a V_E del tamaño de camada.

Regiones	Genes
1:173004560-173995924	<i>KIAA1549L</i> , <i>C11orf91</i> , <i>CD59</i> , <i>ENSOCUG00000014606</i> , <i>ENSOCUG00000026650</i> , <i>LMO2</i> , <i>CAPRINI</i> , <i>NAT10</i> , <i>ABTB2</i> , <i>CAT</i> , <i>ELF5</i> , <i>EHF</i>
3:50019899-51891704	<i>SLIT3</i> , <i>ENSOCUG00000018975</i> , <i>SPDL1</i> , <i>DOCK2</i> , <i>RF00100</i> , <i>FAM196B</i> , <i>FOXI1</i> , <i>LCP2</i> , <i>KCNMBI</i> , <i>KCNIP1</i>
14:121010616-1223019882	<i>COL8A1</i> , <i>ENSOCUG00000003981</i> , <i>DCBLD2</i> , <i>ST3GAL6</i> , <i>CPOX</i> , <i>CLDND1</i> , <i>OR5K4</i> , <i>GPR15</i> , <i>OR5K3</i> , <i>ENSOCUG00000017027</i> , <i>ENSOCUG00000021889</i> , <i>ENSOCUG00000001770</i> , <i>GABRB3</i> , <i>ENSOCUG00000017257</i> , <i>RIOX2</i> , <i>ENSOCUG00000012878</i> , <i>ENSOCUG00000006253</i>

Tabla S2: Funciones de los genes candidatos identificados para V_E del tamaño de camada, según las bases de datos Ensembl, GeneCards y KEGG.

Gen	Procesos Biológicos
<i>DOCK2</i>	<p>Establecimiento de la polaridad de las células T</p> <p>Formación de la sinapsis inmunológica</p> <p>Activación de las células dendríticas mieloides involucradas en la respuesta inmune</p> <p>Quimiotaxis</p> <p>Organización del citoesqueleto de actina</p> <p>Proliferación y activación de las células T</p> <p>Regulación positiva de las señales mediadas por GTPase pequeñas</p> <p>Macropinocitosis</p> <p>Selección positiva y negativa de las células T en el timo</p> <p>Fagocitosis mediada por el receptor FcyR</p> <p>Señalización por quimiocinas</p> <p>Remodelación del citoesqueleto de actina requerido para la migración de los linfocitos en respuesta a quimiocinas</p>
<i>LCP2</i>	<p>Desarrollo y activación de las células T</p> <p>Activación de las células mastocíticas mediante la señalización por FceRI</p> <p>Señalización a través del receptor de las células T</p> <p>Activación plaquetaria</p> <p>Regulación positiva de la actividad de la proteína quinasa</p> <p>Citotoxicidad mediada por las células Natural Killer</p>
<i>LMO2</i>	<p>Desarrollo de las células hematopoyéticas</p>

	<p>Respuesta celular a los estímulos</p> <p>Transcripción de mRNA promovido por la RNA polimerasa II</p> <p>Regulación positiva de la transcripción</p>
<i>SLIT3</i>	<p>Regulación negativa de la proliferación y el crecimiento celular</p> <p>Morfogénesis de órganos animales</p> <p>Regulación negativa de la expresión génica</p> <p>Respuesta celular a hormonas</p> <p>Quimiotaxis negativa</p> <p>Respuesta al cortisol</p> <p>Procesos de apoptosis involucrados en la luteolisis</p> <p>Regulación negativa de las rutas de señalización mediada por quimiocinas</p> <p>Guía y extensión axónica</p> <p>Rutas de señalización mediadas por Roundabout (Robo)</p>
<i>ST3GAL6</i>	<p>Formación de los ligandos de selectina y Sialyl Lewis X</p> <p>Respuesta celular a interleucina 6</p> <p>Glicosilación de proteínas</p> <p>Procesos metabolismo glicolipídico</p>
<i>CD59</i>	<p>Inhibidor del complemento del complejo de ataque a membrana (MAC)</p> <p>Cascada de coagulación y del complemento</p> <p>Linaje de las células hematopoyéticas</p>
<i>GPR15</i>	<p>Migración de las células T</p> <p>Transducción de señales</p> <p>Rutas de señalización acopladas al receptor de la proteína G</p> <p>Mediador de la entrada viral en las células hospedadoras</p>
<i>CAT</i>	<p>Respuesta a especies reactivas del oxígeno</p> <p>Procesos metabólicos de triglicéridos</p> <p>Respuesta a estrés oxidativo</p> <p>Procesos metabólicos del colesterol</p> <p>Respiración aeróbica</p> <p>Protección UV</p> <p>Regulación positiva de las rutas de señalización de quinasas 3-fosfatidilinositol</p> <p>Procesos metabólicos de hemoglobina</p> <p>Regulación negativa y positiva de la actividad del factor de transcripción NF-KappaB</p> <p>Procesos catabólicos del peróxido de hidrógeno</p> <p>Regulación negativa de procesos apoptóticos</p> <p>Tetramerización de proteínas</p> <p>Homotetramerización de proteínas</p> <p>Procesos de óxido-reducción</p> <p>Detoxificación de oxidantes celulares</p>

Tabla S3: Genes localizados en las regiones genómicas asociadas al tamaño de camada (TC), el número de embriones implantados (EI) y la supervivencia embrionaria (ES) y fetal (FS).

Caracteres	Genes
TC	<i>NKX2-6, NKX3-1, SLC25A37, ENTPD4, DDX60, PALLD, VEGFC, ENSOCUG00000023661, NEIL3, ENSOCUG0000002866, ENSOCUG00000023661, RF00001, ENSOCUG00000024301, ENSOCUG00000025270, ENSOCUG00000022735, ALDH8A1, HBS1L, MYB, AH11, PDE7B, ENSOCUG00000023560, ENSOCUG00000016363, ENSOCUG00000016373, MAP7, RF00001</i>
ES	<i>MS4A1, MS4A12, MS4A8, MS4A15, MS4A10, CCDC86, PTGDR2, ZP1, PRPF19, TEMEM109, SLC1543, TEMEM132A, CD6, FOSL1, C11orf68, DRAP1, EIF1AD, CST6, GAL3SR3, ENSOCUG00000012559, ENSOCUG00000015915, ENSOCUG0000002558, ENSOCUG00000029566, ENSOCUG00000029368, ENSOCUG00000026537, ENSOCUG00000022025, ENSOCUG00000025097, ENSOCUG00000015722, ENSOCUG00000016490, ENSOCUG00000022795, CDC47, SP3, OLA1, CIR1, ENSOCUG00000026850, RF00001, ENSOCUG00000022719, PLOD2, PLSCR4, PLSCR5, ENSOCUG00000026408, ENSOCUG00000023944, ENSOCUG00000029550, ENSOCUG00000012892, ENSOCUG00000024040</i>
EI	<i>MS4A1, MS4A12, MS4A8, MS4A15, MS4A10, CCDC86, PTGDR2, ZP1, PRPF19, TEMEM109, SLC1543, TEMEM132A, CD6, FOSL1, C11orf68, DRAP1, EIF1AD, CST6, GAL3SR3, ENSOCUG00000012559, ENSOCUG00000015915, ENSOCUG0000002558, ENSOCUG00000029566, ENSOCUG00000029368, ENSOCUG00000026537, ENSOCUG00000022025, ENSOCUG00000025097, ENSOCUG00000015722, ENSOCUG00000016490, ENSOCUG00000022795, SEPT11, FRAS1, CXCL13, CCNG2, CCNI, CNOT6L, MRPL1, ENSOCUG00000021783, ENSOCUG00000001139</i>
FS	<i>NCOA6, ENSOCUG00000009634, PIGU, ENSOCUG00000021027, ITCH, ENSOCUG00000008592, ASIP, EIF2S2, RALY, CHMP4B, ZNF341, NECAB3, E2F1, CBFA2T2, ACTL10, ASTN1, BRINP2, SEC16B, RASAL2, TEX35, RALGPS2, ANGPTL1, FAM20B, TOR3A, ABL2, SOAT1, AXDND1, NPHS2, TDRD5, FAM163A, TOR1AIP1, TOR1AIP2, CEP350, TIPARP, CCNL1, VEPHI, PTX3, ENSOCUG00000027732, RF00001, RF01210, RF00001, ENSOCUG00000023110, ENSOCUG00000023165</i>

Tabla S4: SNPs con diferente efecto de sustitución alélica entre la línea para baja (L) y alta (H) V_E del tamaño de camada.

CHR ^a	BP ^b	A1 ^c	A2 ^d	MAF H ^e	MAF L ^f	α H ^g	α L ^h	IC H ⁱ	IC L ^j		
14	119411302	G	T	0.35	0.08	1.45	-0.68	0.54	2.35	-1.63	0.26
14	119438702	T	C	0.35	0.08	1.45	-0.68	0.54	2.35	-1.63	0.26
14	119481491	T	C	0.35	0.08	1.40	-0.68	0.49	2.31	-1.63	0.26
14	119493313	T	C	0.35	0.07	1.42	-0.67	0.52	2.32	-1.63	0.29
14	119510003	T	A	0.35	0.08	1.45	-0.68	0.54	2.35	-1.63	0.26
14	119519332	T	C	0.35	0.08	1.45	-0.68	0.54	2.35	-1.63	0.26
14	119528607	T	G	0.35	0.08	1.45	-0.68	0.54	2.35	-1.63	0.26
14	119627380	G	A	0.35	0.08	1.45	-0.68	0.54	2.35	-1.63	0.26
14	119634851	G	A	0.35	0.08	1.45	-0.68	0.54	2.35	-1.63	0.26
14	119647325	G	A	0.36	0.08	1.43	-0.68	0.52	2.34	-1.63	0.26
14	119653763	C	T	0.35	0.08	1.41	-0.68	0.50	2.32	-1.63	0.26
14	119674677	C	T	0.35	0.08	1.45	-0.68	0.54	2.35	-1.63	0.26
14	119698422	T	C	0.35	0.08	1.45	-0.68	0.54	2.35	-1.63	0.26
14	119721590	C	T	0.35	0.08	1.45	-0.52	0.54	2.35	-1.43	0.38
14	121545568	T	C	0.36	0.09	1.41	-0.55	0.51	2.31	-1.38	0.28
14	121553471	T	C	0.35	0.08	1.45	-0.72	0.55	2.35	-1.65	0.20
14	121557805	A	G	0.36	0.08	1.46	-0.68	0.56	2.35	-1.63	0.26
14	121607394	G	A	0.36	0.08	1.46	-0.68	0.56	2.35	-1.63	0.26
14	121709194	A	G	0.36	0.08	1.46	-0.68	0.56	2.35	-1.63	0.26
14	121722271	C	T	0.36	0.08	1.46	-0.68	0.56	2.35	-1.63	0.26
14	121742713	A	G	0.35	0.08	1.48	-0.68	0.58	2.39	-1.63	0.26
14	121768614	T	C	0.36	0.08	1.46	-0.77	0.56	2.35	-1.70	0.16
14	121775490	T	C	0.36	0.08	1.46	-0.68	0.56	2.35	-1.63	0.26
14	121800537	A	C	0.36	0.08	1.46	-0.68	0.56	2.35	-1.63	0.26
14	121811327	A	C	0.36	0.08	1.47	-0.68	0.58	2.36	-1.63	0.26
14	121819229	T	C	0.36	0.08	1.46	-0.68	0.56	2.35	-1.63	0.26
14	121835127	A	G	0.36	0.08	1.46	-0.68	0.56	2.35	-1.63	0.26
14	121846466	T	C	0.36	0.08	1.46	-0.68	0.56	2.35	-1.63	0.26
14	121853055	C	T	0.37	0.08	1.54	-0.68	0.64	2.45	-1.63	0.26
14	121869256	T	C	0.36	0.08	1.46	-0.68	0.56	2.35	-1.63	0.26
14	121890492	C	T	0.36	0.08	1.46	-0.68	0.56	2.35	-1.63	0.26
14	121900079	G	A	0.36	0.08	1.46	-0.68	0.56	2.35	-1.63	0.26
14	121920845	A	C	0.36	0.08	1.46	-0.68	0.56	2.35	-1.63	0.26
14	121926782	C	A	0.36	0.08	1.46	-0.68	0.56	2.35	-1.63	0.26
14	121939533	A	G	0.36	0.08	1.46	-0.68	0.56	2.35	-1.63	0.26
14	122015259	A	G	0.35	0.08	1.45	-0.68	0.55	2.35	-1.63	0.26
14	122025602	G	A	0.36	0.08	1.41	-0.78	0.52	2.31	-1.70	0.15
14	122035750	C	A	0.36	0.08	1.51	-0.71	0.60	2.41	-1.62	0.20
14	122051784	T	C	0.36	0.08	1.46	-0.68	0.56	2.35	-1.63	0.26
14	122057144	A	C	0.35	0.08	1.46	-0.68	0.57	2.35	-1.63	0.26

14	122067396	A	G	0.36	0.08	1.46	-0.68	0.56	2.35	-1.63	0.26
14	122081554	T	C	0.36	0.08	1.46	-0.68	0.56	2.35	-1.63	0.26
14	122100330	G	A	0.36	0.08	1.46	-0.68	0.56	2.35	-1.63	0.26
14	122105989	C	A	0.36	0.08	1.46	-0.68	0.56	2.35	-1.63	0.26
14	122116380	A	G	0.36	0.08	1.46	-0.68	0.56	2.35	-1.63	0.26
14	122122006	C	A	0.36	0.08	1.46	-0.63	0.56	2.35	-1.56	0.29
14	122134774	T	C	0.36	0.08	1.46	-0.68	0.56	2.35	-1.63	0.26
14	122143041	A	G	0.36	0.08	1.46	-0.68	0.56	2.35	-1.63	0.26
14	122150465	C	T	0.36	0.08	1.46	-0.68	0.56	2.35	-1.63	0.26
14	122158178	G	A	0.36	0.08	1.46	-0.62	0.56	2.35	-1.55	0.31
14	122197634	A	G	0.36	0.08	1.46	-0.68	0.56	2.35	-1.63	0.26
14	122205185	T	C	0.36	0.08	1.46	-0.68	0.56	2.35	-1.63	0.26
14	122226131	C	A	0.36	0.08	1.50	-0.77	0.61	2.39	-1.69	0.15
14	122260566	A	G	0.36	0.08	1.46	-0.68	0.56	2.35	-1.63	0.26
14	122267612	G	A	0.35	0.09	1.44	-0.64	0.55	2.34	-1.53	0.25
14	122273203	A	G	0.36	0.08	1.46	-0.68	0.56	2.35	-1.63	0.26
14	122285369	A	G	0.36	0.08	1.42	-0.68	0.53	2.31	-1.63	0.26
14	122336825	T	C	0.36	0.08	1.46	-0.68	0.56	2.35	-1.63	0.26
14	122365220	G	A	0.36	0.08	1.46	-0.68	0.56	2.35	-1.63	0.26
14	122381971	T	G	0.36	0.08	1.46	-0.68	0.56	2.35	-1.63	0.26
14	122411378	G	A	0.37	0.08	1.26	-0.68	0.39	2.13	-1.63	0.26
14	122432920	G	A	0.36	0.08	1.45	-0.64	0.55	2.35	-1.55	0.26
14	122471382	A	G	0.36	0.08	1.46	-0.68	0.56	2.35	-1.63	0.26
14	122478735	A	G	0.36	0.08	1.46	-0.68	0.56	2.35	-1.63	0.26
14	122489762	C	T	0.35	0.10	1.49	-1.03	0.59	2.39	-1.87	-0.18
14	122534940	T	C	0.36	0.08	1.46	-0.68	0.56	2.35	-1.63	0.26
14	122544769	C	T	0.35	0.08	1.49	-0.68	0.59	2.38	-1.63	0.26
14	122562966	T	C	0.36	0.08	1.46	-0.68	0.56	2.35	-1.63	0.26
14	122575971	C	A	0.36	0.08	1.46	-0.68	0.56	2.35	-1.63	0.26
14	122609443	T	C	0.37	0.08	1.31	-0.79	0.44	2.17	-1.71	0.14
14	122616950	G	A	0.36	0.08	1.46	-0.68	0.56	2.35	-1.63	0.26
14	122633902	T	C	0.36	0.08	1.46	-0.68	0.56	2.35	-1.63	0.26
14	122665196	T	G	0.36	0.08	1.46	-0.61	0.56	2.35	-1.45	0.22
14	122673025	C	T	0.36	0.08	1.47	-0.61	0.56	2.37	-1.45	0.22
14	122686546	A	C	0.35	0.08	1.37	-0.69	0.46	2.29	-1.57	0.20
14	122693897	A	G	0.36	0.08	1.46	-0.61	0.56	2.35	-1.45	0.22
14	122709049	G	A	0.36	0.08	1.46	-0.61	0.56	2.35	-1.45	0.22
14	122720589	G	A	0.36	0.08	1.46	-0.61	0.56	2.35	-1.45	0.22
14	122727864	T	C	0.36	0.08	1.46	-0.68	0.56	2.35	-1.63	0.26
14	122744666	C	T	0.36	0.08	1.46	-0.68	0.56	2.35	-1.63	0.26
14	122751385	G	A	0.36	0.08	1.46	-0.68	0.56	2.35	-1.63	0.26
14	122772616	G	A	0.36	0.08	1.46	-0.68	0.56	2.35	-1.63	0.26
14	122795884	A	G	0.36	0.08	1.46	-0.68	0.56	2.35	-1.63	0.26
14	122809573	T	G	0.36	0.08	1.53	-0.68	0.62	2.44	-1.63	0.26
14	122819188	A	G	0.36	0.08	1.46	-0.68	0.56	2.35	-1.63	0.26

14	122851522	T	C	0.36	0.08	1.46	-0.68	0.56	2.35	-1.63	0.26
14	122873676	A	G	0.33	0.07	1.50	-0.58	0.61	2.39	-1.53	0.38
14	162957181	G	A	0.28	0.25	2.06	0.00	0.97	3.15	-0.56	0.57
14	162965885	A	G	0.28	0.25	2.06	0.00	0.97	3.15	-0.56	0.57
14	162972545	T	C	0.28	0.25	2.06	0.00	0.97	3.15	-0.56	0.57
14	162977721	A	G	0.28	0.25	2.06	0.00	0.97	3.15	-0.56	0.57
14	163015138	A	G	0.28	0.25	2.06	0.00	0.97	3.15	-0.56	0.57
14	163021732	G	A	0.28	0.25	2.06	0.00	0.97	3.15	-0.56	0.57
14	163039420	T	C	0.28	0.25	2.06	0.00	0.97	3.15	-0.56	0.57
14	163104347	A	G	0.28	0.26	2.06	0.08	0.97	3.15	-0.49	0.65
14	163137148	T	C	0.28	0.25	2.06	0.09	0.97	3.15	-0.48	0.66

^aCromosoma, ^bPosición en pares de bases, ^cAlelo de referencia, ^dAlelo alternativo, ^eFrecuencia del alelo minoritario para la línea de H V_E del TC, ^fFrecuencia del alelo minoritario para la línea de L V_E del TC, ^g Efecto de sustitución alélica de la línea H para V_E del TC. ^hEfecto de sustitución alélica d la línea L V_E del TC, ⁱIntervalo de confianza para la línea H V_E del TC, ^jIntervalo de confianza para la línea H V_E del TC.