GENOMIC ANALYSIS OF DIVERGENTLY SELECTED EXPERIMENTAL LINES IN RABBITS

Ph.D. Thesis by

Bolívar Samuel Sosa Madrid

Supervisors: **Prof. Dr. Noelia Ibáñez Escriche**
Prof. Dr. Agustín Blasco Mateu

Institute for Animal Science and Technology
Universitat Politècnica de València

València, 23th March 2020

GENOMIC ANALYSIS OF DIVERGENTLY SELECTED EXPERIMENTAL LINES IN RABBITS

This thesis has been submitted in fulfilment of the requirements for the degree of Doctor with International Mention at the Universitat Politècnica de València.

Esta tesis ha sido escrita y presentada como uno de los requisitos para optar al grado de Doctor con Mención Internacional por la Universitat Politècnica de València.

By

Bolívar Samuel Sosa Madrid

Thesis Supervisors

Prof. Dr. Noelia Ibáñez Escriche **Prof. Dr. Agustín Blasco Mateu**

València, 23th March 2020

One is all, All is one

FULLMETAL ALCHEMIST: BROTHERHOOD

ACKNNOWLEDGEMENTS

Yo agradezco a “DIOS” y a mi familia, quienes han sido un apoyo fundamental durante el doctorado. “DIOS”, tú me ayudaste a tomar la decisión de estudiar en el extranjero y tú has guiado mis pasos principalmente en los momentos difíciles. Mamá, Papá, “Salo” y “Gwendo”, gracias por recordarme siempre de dónde vengo, lo que valgo y lo que es importante en la vida. Yo estoy orgulloso de ser panameño, de ustedes, de nuestra cultura, costumbres e idioma. Panamá es un país pequeño lleno de personas que tienen un corazón “sólido” y “sabrosón”.

También, quiero agradecer al grupo de mejora genética animal de la UPV, a los investigadores principales: María Antonia, Pilar, Agustín y Noelia, por todas sus enseñanzas. Gracias a Agustín por todo durante este tiempo de mi doctorado. Un especial gracias a Noelia, por el tiempo que dedicaste a enseñarme genómica estadística; realmente aprecio todo lo que has hecho por mí a nivel personal.

Gracias a todos mis compañeros del Instituto de Ciencia y Tecnología Animal. Un millón de gracias a los compañeros de genética, Marina Martínez, Agostina, Cristina, “Mafe”, Enrico, David, Ana y los demás, quienes me acompañaron en estos últimos cuatro años. También quiero agradecer a los compañeros del grupo de acuicultura, Víctor “El Pastor”, German, Guillem, Miriam y los demás. También gracias Richard, Ximo y Andrea por las graciosas conversaciones.

Un especial gracias a Federico, Verónica, Marina Morini, Antonella, Ilaria y Rosalia por su trabajo en este experimento; especialmente a Marina que me acompañó en el arduo trabajo durante 5 meses, y estuvo presente cuando proclame mi frase célebre “Que asco de vida”. También quiero agradecer a John Hickey y su grupo, especialmente a Roger y Mara, el tiempo dedicado a mi formación en mi primera estancia en Edinburgh. Para mi es y será un recuerdo especial. A Chris Haley y su grupo, en especial a Pau, Yanni, Charlie y Carmen, un millón de gracias, ya que con ustedes aprendí mucho de GWAS y el trabajo grupal, lo cual superó mis expectativas. Gracias a Luis Varona por enseñarme sobre huellas de selección y mostrarme qué tan genial puede ser Zaragoza.

A “Froy”, Maru, Leonela y los demás amigos de Panamá, gracias por su apoyo en los últimos cuatro años. A Jeffry, Judith, Latifa y Kenza, un millón de gracias, ya que ustedes fueron un pilar esencial para continuar realizando el doctorado. A los amigos que conocí en Edinburgh, Piotr, Dariuz, Alberto, Yuyao y Liam, mil gracias, conocerlos me ayudo a ser feliz durante mi estancia en esa hermoso lugar que es Scotland. Un especial gracias a José “El Duque” Medina, tu amistad fue esencial para terminar la tesis este último año. Supongo que no molaba estar más de 10 minutos escuchándome hablar de genética, pero admítelo a veces era interesante. Gracias a Andrea y "Duque por la portada. Gracias a todos los amigos que tengo alrededor del mundo y a los que he olvidado mencionar.

En fin, esta tesis se la dedico a “DIOS”, a mi familia y a mis amigos cercanos.

In English:

I thank GOD and my family, who have been an essential support during the doctorate. GOD, you helped me take the decision of studying abroad and you have guided my steps mainly in difficult times. I would like to thank Mom, Dad, "Salo" and "Gwendo" for always reminding me where I came from, the worth of myself and what is important in life. I am so proud to be Panamanian, of my family, of our culture, customs and language. Panama is a small country full of people who have a "solid" and "sabrosón" heart.

I thank to the Group of Animal Breeding from the Polytechnic University of Valencia, thanks to the main researchers: Maria Antonia, Pilar, Agustín and Noelia for their teaching. I must express my very profound gratitude to Noelia for your patience to teach me quantitative genetics and genomics. I know it was not easy, and I really appreciate everything you have done for me.

Thanks to all my colleagues from the Institute for Animal Science and Technology. Thanks a million to fellows of the Group of Animal Breeding, Marina Martínez, Agostina, Cristina, "Mafe", Enrico, David, Ana and other fellows. I am also thankful to the colleagues of the Aquaculture Group, Victor “El Pastor”, German, Guillem, Miriam and the others. I would also like to thank Richard, Ximo and Andrea for the funny conversations.

I must express a special thanks to Federico, Veronica, Marina Morini, Antonella, Ilaria and Rosalia, for working with me in the experiments at the rabbit farm and the meat lab; especially to Marina, my principal workmate over tough five months. You were present when I said my famous phrase “How disgusting my life”. I will not forget it. I would like to thank John Hickey and your group, particularly to Roger and Mara, for your time invested in my learning during my first stay in Edinburgh. That time was and will be a special memory for me. Thanks to Chris Haley and your group, particularly to Pau, Yanni, Charlie and Carmen, thanks a million, since I learnt with you a lot of GWAS and work as a team. That stay exceeded my expectation. I would like to thank Luis Varona for teaching methods of selection signatures, and also, for showing me the phenomenal Zaragoza.

Thanks to “Froy”, Maru, Leonela and other friends in Panama for your support in the last four years. To Jeffry, Judith, Latifa and Kenza, thanks a million, since you were an essential piece to keep going with my doctorate. Thanks to my friends that I knew in Edinburgh: Piotr, Dariuz, Alberto, Yuyao y Liam, thanks a million, I was happy with you during my long stay in beautiful Scotland. I would like to express a special thanks to José “El Duque” Medina, your friendship was crucial in the last year of my thesis. I suppose that it is not nice listening to me talk about genetics over 10 minutes. Thanks to my friends around the world and those friends that I have forgotten.

Finally, this thesis is dedicated to GOD, my family and close friends.

CONTENTS

ACKNNOWLEDGEMENTS ..... **I**
CONTENTS ..... **III**
LIST OF FIGURES ..... **VII**
LIST OF TABLES ..... **IX**
LIST OF ABBREVIATIONS ..... **XI**
ABSTRACT **1**
RESUMEN **3**
RESUM **7**
RÉSUMÉ **9**
CHAPTER ONE **13**
1. GENERAL INTRODUCTION **13**
1.1. RABBIT PRODUCTION AND BREEDING 13
1.1.1. Overview *13*
1.1.2. Traits of interest in rabbit production *13*
1.1.3. Relationship between growth, meat and litter size traits *14*
1.2. RABBITS AS ANIMAL EXPERIMENTAL MODEL 16
1.3. DIVERGENT SELECTION EXPERIMENTS 16
1.4. MOLECULAR GENETICS FOR UNCOVERING THE IMPORTANCE OF GENOMICS IN RABBIT 17
1.4.1. Genome wide association study *19*
1.4.1.1. Previous to GWAS: genetic markers and genomic analyses . 19
1.4.1.2. Methods and strategies 20
Designing association studies 21
Frequentist inference 24
Bayesian inference 28
1.4.2. Genome scan study *33*
1.4.2.1. Methods of selection signatures 33
Differentiation amongst populations 34
Reduction of the local genomic variability 35
Modification of allele frequency spectrums 35
Extension of linkage disequilibrium 36
1.4.3. Post-genomic analyses, validation and application of findings *37*
1.5. REFERENCES 39
CHAPTER TWO **55**

2. SCOPE OF THE STUDY 55
2.1. SPECIFIC OBJECTIVES OF THIS THESIS 55
CHAPTER THREE 57
3. A GENOME-WIDE ASSOCIATION STUDY IN DIVERGENTLY SELECTED LINES IN RABBITS REVEALS NOVEL GENOMIC REGIONS ASSOCIATED WITH LITTER SIZE TRAITS 57
3.1. ABSTRACT 59
3.2. BACKGROUND 59
3.3. MATERIAL AND METHODS 61
3.3.1. Ethical statement 61
3.3.2. Animals and phenotypes 61
3.3.3. Genotypes and quality control 62
3.3.4. Statistical analysis 62
3.3.5. Linkage disequilibrium, pathways and functional enrichment analyses 65
3.4. RESULTS AND DISCUSSION 65
3.4.1. Descriptive statistics of phenotypic data 65
3.4.2. Description of genomic data 66
3.4.3. Prior choice 68
3.4.4. Genomic windows associated with litter size traits 68
3.4.4.1. Total number born and number born alive 68
3.4.4.2. Implanted embryos 70
3.4.4.3. Ovulation rate 71
3.4.5. Associated SNPs in genomic regions 74
3.4.6. Linkage disequilibrium analysis 78
3.4.7. Gene search and functional annotations 81
3.5. CONCLUSIONS 85
3.6. ABBREVIATIONS 85
3.7. REFERENCES 85
3.8. ADDITIONAL FILES 92
CHAPTER FOUR 95
4. GENOMIC REGIONS INFLUENCING INTRAMUSCULAR FAT IN DIVERGENTLY SELECTED RABBIT LINES 95
4.1. ABSTRACT 97
4.2. BACKGROUND 97
4.3. MATERIAL AND METHODS 99

4.3.1. Ethical statement *99*
4.3.2. Animals and phenotypes *99*
4.3.3. Genotypes and quality control *100*
4.3.4. Genome-wide association study *101*
4.3.5. Linkage disequilibrium and functional gene analysis *104*
4.4. RESULTS 104
4.4.1. Genomic data *104*
4.4.2. GWAS for intramuscular fat *105*
4.4.3. Functional annotation analysis and putative candidate genes. ... *111*
4.5. DISCUSSION 112
4.6. CONCLUSIONS AND IMPLICATIONS 117
4.7. ABBREVIATIONS 117
4.8. AVAILABILITY OF DATA AND MATERIAL 117
4.9. REFERENCES 118
4.10. ADDITIONAL FILES 125
CHAPTER FIVE 139
5. THE EFFECT OF DIVERGENT SELECTION FOR INTRAMUSCULAR FAT IN THE DOMESTIC RABBIT GENOME. 139
5.1. ABSTRACT 141
5.2. IMPLICATIONS 142
5.3. BACKGROUND 142
5.4. MATERIAL AND METHODS 143
5.4.1. Ethical statement *143*
5.4.2. Animals, genotyping data and quality control *144*
5.4.3. Divergence between lines *144*
5.4.4. Detection of signatures of selection *144*
5.4.4.1. Fixation Index (F_{st}) 145
5.4.4.2. Cross Population Composite Likelihood Ratio Test (XP-CLR) 145
5.4.4.3. Cross Population Extended Haplotype Homozygosity Test (XP - EHH) 146
5.4.5. Enrichment analysis of functional annotation, and gene ontology (GO) terms *147*
5.5. RESULTS AND DISCUSSION 147
5.5.1. Genome-wide scan for populations from divergent selection *149*

5.5.2. Underlying selected genes and gene ontology terms for divergent selection *157*
5.6. CONCLUSIONS 160
5.7. ABBREVIATIONS 160
5.8. REFERENCES 160
5.9. ADDITIONAL FILES 165
CHAPTER SIX 175
6. GENERAL DISCUSSION 175
REFERENCES 179
CHAPTER SEVEN 183
7. CONCLUSIONS AND IMPLICATIONS 183

LIST OF FIGURES

CHAPTER THREE: FIRST ARTICLE

Figure 1. Multidimensional scaling plot of the genomic data.....67

Figure 2. Manhattan plot for total number born (TNB) using the percentage of genomic variance explained by each non-overlapping one megabase window.....69

Figure 3. Manhattan plot for number born alive (NBA) using the percentage of genomic variance explained by each non-overlapping one megabase window..... 69

Figure 4. Manhattan plot for implanted embryos (IE) using the percentage of genomic variance explained by each non-overlapping one megabase window..... 70

Figure 5. Manhattan plot for ovulation rate (OR) using the percentage of genomic variance explained by each non-overlapping one megabase window.....71

Figure 6. Manhattan plot for implanted embryos (IE) using the Bayes factors by each SNP along the rabbit chromosomes.....75

Figure 7. Manhattan plot for total number born (TNB) using the Bayes factors by each SNP along the rabbit chromosomes.....75

Figure 8. Linkage disequilibrium plot of chromosome 11..... 79

Figure 9. Linkage disequilibrium plot of chromosome 17..... 80

CHAPTER FOUR: SECOND ARTICLE

Figure 1. Multidimensional scaling plot of genomic data..... 106

Figure 2. Manhattan Plot for each model..... 110

Figure 3. Linkage disequilibrium blocks from main relevant associated polymorphisms..... 111

Additional file 1: Figure S1. Linkage disequilibrium (LD) block of the associated genomic region on OCU13..... 125

Additional file 2: Figure S2. Linkage disequilibrium (LD) block of an associated genomic region on OCU8..... 126

Additional file 3: Figure S3. Assessment of genotypes for the three relevant SNPs within genomic regions associated with intramuscular fat..... 127

CHAPTER FIVE: THIRD ARTICLE

Figure 1. Single nucleotide polymorphism (SNP) density within 1-Mb window size using 89,968 SNPs after quality control for the intramuscular fat genomic data in rabbits..... 148

Figure 2. Multidimensional scaling plot of intramuscular fat genomic data in rabbits..... 149

Figure 3. Manhattan plot of 500 kb window for every method of selection signatures in rabbits..... 153

Figure 4. Venn diagram of methods of selection signatures in rabbits..... 156

LIST OF TABLES

CHAPTER ONE: GENERAL INTRODUCTION

Table 1. Selection response of divergent selection experiments in rabbits..... 18

CHAPTER THREE: FIRST ARTICLE

Table 1. Prior variances for Bayes B method.....64

Table 2. Descriptive statistics of little size traits.....66

Table 3. Genomic windows associated with total number born (TNB), number born alive (NBA), implanted embryos (IE), and ovulation rate (OR) in rabbits. 72

Table 4. Relevant polymorphisms (SNPs) for total number born (TNB) and implanted embryos (IE)..... 76

Table 5. Top five functional enrichment from the analyses performed through DAVID online web..... 83

Additional file 1: Table S1. Annotated genes in the genomic regions associated with litter size traits. 92

CHAPTER FOUR: SECOND ARTICLE

Table 1. Allocation of SNPs after quality control and average distance amongst contiguous SNPs on every chromosome.....105

Table 2. Relevant polymorphisms (SNPs) and genomic windows associated with intramuscular fat..... 108

Table 3. Summary of relevant genomic regions associated with intramuscular fat, and annotated rabbit genes.....113

Additional file 4: Table S1. Genes found in the genomic regions associated with intramuscular fat..... 128

Additional file 5: Table S2. Functions of genes identified in this study through Enrichr and DAVID..... 131

CHAPTER FIVE: THIRD ARTICLE

Table 1. Correlations between the levels of window size within methods of selection signatures in rabbits..... 151

Table 2. SNP name (SNP_ID), rabbit chromosome (OCU), cluster (genomic region), SNP physical position in megabase and values for three methods of selection signatures based on detection in at least two methods using a cut-off of 1%..... 154

Table 3. Biological processes of highlighted genes identified in at least two methods of selection signatures for cut-off of 1%..... 158

Additional file 1: Table S1. Top five GO terms of biological processes using cut-off of 1% for each method of selection signatures..... 165

Additional file 2: Table S2. Annotated genes in the selection signatures of intramuscular fat in rabbits.....167

Additional file 2: Table S2. First ten GO terms of biological processes using genes located in genomic regions that exceeded a cut-off of 1% in at least two methods of selection signatures in rabbits..... 173

LIST OF ABBREVIATIONS

200K SNP	two hundred thousand single nucleotide polymorphisms
ACER2	alkaline ceramidase 2
ADG	average daily gain
AI	artificial insemination
APOLD1	apolipoprotein L domain containing 1
BF	Bayes factor
BGLR	Bayesian Generalized Linear Regression
BMM	Bayesian multi-marker
BMMR	Bayesian multi marker regression
BMP4	bone morphogenetic protein 4
CCT5	chaperonin containing TCP1, subunit 5 (epsilon)
CDKN3	cyclin-dependent kinase inhibitor 3
CLR	composite likelihood ratio
CM	cumulative mortality
DAVID	Database for Annotation, Visualization and Integrated Discovery
DENND4C	DENN domain containing 4C
DNA	deoxyribonucleic acid
DQC	dish quality control
EBV – ES	estimated breeding value for the enteropathy score
EHH	extended haplotype homozygosity
EMMAX	efficient mixed-model association expedited
ERO1	endoplasmic reticulum oxidoreductase 1
ES	embryo survival
ESR	estrogen receptor
FASTA	family-based score test association
FCR	feed conversion rate
FS	foetal survival
FSHB	follicle stimulating hormone subunit Beta
F_{st}	Wright's fixation index
FUT1	fucosyltransferase 1 (H Blood Group)
GANC	glucosidae alpha neutral C
GCTA	genome wide complex trait analysis
GEMMA	genome wide efficient mixed model association

GO	gene ontology
GPRC5	G protein-coupled receptor class C group 5
GRAMMAR	genome wide rapid association using mixed model and regression
GRM	genomic relationships matrix
GV	genomic variance
GWAS	genome wide association study
Hp	measure of heterozygosity
HSD17B4	progesterone receptor - PGR, hydroxysteroid (17-beta) dehydrogenase 4
HWE	Hardy Weinberg equilibrium
IE	implanted embryos
iHS	integrated haplotype score
IMF	intramuscular fat
K88	K88 fimbriae gene of Escherichia coli
KEGG	Kyoto Encyclopedia of genes and Genomes
KRAS	KRAS proto-oncogene, GTPase
LD	linkage disequilibrium
LTH	*longissimus thoracis et lumborum*
MAF	minor allele frequency
Mb	Megabase
MCMC	Markov chain Monte Carlo
MDS	multidimensional scaling
MSTN	myostatin
MTMR2	myotubularin related protein 2
NBA	number born alive
NBD	number born dead
NGS	next generation sequencing
NRAMP	natural resistance-associated macrophage protein
OCU	the rabbit chromosome
OFS	open-field score
OR	ovulation rate
OVGP1	oviductal glycoprotein 1
PDE6H	phosphodiesterase 6H
PGR	progesterone receptor
PLA2G4B	phospholipase A2 group IVB
PLBD1	phospholipase B domain containing 1

PLIN2	Perilipin 2
PPA	posterior probabilities of association
PRLR	prolactin receptor
PS	prenatal survival
PTGDR	prostaglandin D2 receptor (DP)
PTGER2	prostaglandin E receptor 2 (subtype EP2), 53kDa
QC	quality control
QTL	quantitative trait locus
RBP4	Retinol Binding Protein 4
REACTA	regional heritability advanced complex trait analysis
REHH	relative extended haplotype homozygosity
RFLP	restriction fragment length polymorphism
ROH	runs of homozygosity
RORA	RAR related orphan receptor A
RPS6	ribosomal protein S6
RRGA	Ras related GTP binding A
Rsb	the standardized ratio of integrated EHH from two populations
RTF1	RNA Polymerase-Associated Protein RTF1 Homolog
SFC	the site frequency spectrum
SLA	Src Like Adaptor
SLC24A2	solute-carrier gene family 24 member 2
SMR	single marker regression
SNP	single nucleotide polymorphism
SNV	single nucleotide variant
ST8SIA6	ST8 alpha-N-acetyl-neuraminide alpha-2,8-sialytransferase 6
STYX	serine/threonine/tyrosine interacting protein
TASSEL	trait analysis by association, evolution and linkage
TIMP1	TIMP metallopeptidase inhibitor 1
TNB	total number born
UC	uterine capacity
ULO	unilaterally ovariectomized
varLD	variation of linkage disequilibrium
VIM	Vitemin
WPPA	window posterior probabilities of association
XP-CLR	cross population composite likelihood ratio test
XP-EHH	cross population extended haplotype homozygosity test

ABSTRACT

Divergent selection can alter frequencies of genetic markers in opposite directions, leading to intermediate allelic frequencies when both divergent lines are jointly considered in the genetic analyses. Therefore, divergent selection experiments increase the detection power for genome wide association study (GWAS) and for genomic scan studies through methods of selection signatures.

At the Universitat Politècnica de València, two independent divergent selection experiments were carried out in rabbits, one for uterine capacity and the other one for intramuscular fat. Both experiments attained successful selection responses, being 1.50 kits for uterine capacity at 10th generation and 3.10 standard deviations for intramuscular fat at 9th generation, respectively. Animals from these experiments were used for performing genomic analyses of litter size traits and intramuscular fat. Genotypes were obtained by means of a high-density single nucleotide polymorphism (SNP) array of 200K.

Bayesian GWASs using Bayes B model was used to analyse genomic data of litter size traits of the uterine capacity experiment with 181 does. The associations were tested by computing Bayes factors for each SNP, and by computing percentages of the genomic variance for each 1-Mb non-overlapping window. The GWASs uncovered SNPs associated with total number born and implanted embryos. Moreover, relevant genomic regions were revealed for total number born (1 region), number born alive (1 region), implanted embryos (3 regions), and ovulation rate (5 regions). The percentages of genomic variance that accounted for these litter size traits were 39.48%, 10.36%, 37.21%, and 3.95%, respectively, under a model excluding line effect; and 7.36%, 1.27%, 15.87%, and 3.95%, respectively, under a model with line effect. The genomic region located on the rabbit chromosome (OCU) 17 at 70.0 - 73.3 Mb was deemed as a novel quantitative trait locus (QTL) of reproductive traits in rabbits, since this region was found overlapped for total number born, number born alive and implanted embryos. Bone morphogenetic protein 4 gene, *BMP4*, is the main promising candidate gene within the novel QTL.

A combination of GWASs were performed for analysing the genomic data of the intramuscular fat experiment with 480 rabbits. The GWAS methods included a

Bayesian method, Bayes B model; and a frequentist method, single marker regressions with the data adjusted by genomic relatedness. This study revealed four relevant genomic regions in OCU1 (1 region), OCU8 (2 regions) and OCU13 (1 region) associated with intramuscular fat. The most important associated region was on OCU8 at 24.59 - 26.95 Mb, and accounted for 7.34% of the genomic variance. The low percentage explained by the main relevant genomic regions indicates a large polygenic component for intramuscular fat. Functional analyses retrieved genes linked to pathways and function of energy, carbohydrate and lipid metabolisms. In addition, a genome scan study was performed using rabbits from the divergent selection experiment for intramuscular fat, and using three methods of selection signatures: Wright's fixation index (F_{st}), cross population composite likelihood ratio (XP-CLR) and cross population extended haplotype homozygosity (XP-EHH). The results showed multiple selection signatures across the rabbit genome. None of these selection signatures agreed with the associated genomic regions from GWAS findings. In synthesis, the results of both experiments, GWAS and genome scan study, suggest that the genomic architecture of intramuscular fat in rabbit seems to be highly polygenic and their causative variants would be hardly detectable.

This study demonstrates that detection of causative variants and associated genetic markers depends on the hypothetical genomic architectures of traits, regardless of the successful responses attained in the two divergent selection experiments. Apart from the novel QTL for litter size, none of genomic regions explained a large part of the genomic variances of litter size traits and of intramuscular fat in rabbits. Thus, all analysed traits have a large polygenic component. Further analyses and studies will be needed to bear out the findings of the current research study. Hitherto, these findings would not have worthwhile implications for the rabbit breeding programs.

RESUMEN

La selección divergente puede cambiar las frecuencias de los marcadores genéticos en direcciones opuestas, produciéndose frecuencias alélicas intermedias en estos marcadores cuando ambas líneas divergentes son consideradas conjuntamente en los análisis genéticos. Por lo tanto, los experimentos de selección divergente aumentan el poder de detección para estudio de asociación de genoma completo (GWAS) y para estudios de escaneo genómico por medio de métodos de huellas de selección.

En la Universitat Politècnica de València, dos experimentos de selección divergente independientes entre sí fueron realizados en conejos, uno para la capacidad uterina y el otro para la grasa intramuscular. Ambos experimentos lograron exitosas respuestas de selección, siendo 1.50 gazapos para la capacidad uterina en la décima generación y 3.10 desviaciones estándar para la grasa intramuscular en la novena generación, respectivamente. Los animales que provienen de estos experimentos fueron utilizados para llevar a cabo análisis genómicos de los caracteres de tamaño de camada y de la grasa intramuscular. Los genotipos fueron obtenidos usando una plataforma de alta densidad de 200k de polimorfismos de nucleótido único (SNP).

GWASs bayesianos, utilizando el modelo Bayes B, se implementaron para analizar datos genómicos de los caracteres de tamaño de camada del experimento de capacidad uterina con 181 hembras. Las asociaciones fueron evaluadas calculando los factores de Bayes para cada SNP, y calculando los porcentajes de la varianza genómica para cada ventana no solapada de 1-Mb. Los GWASs descubrieron SNPs asociados con el número total de gazapos al parto y los embriones implantados. Además, se revelaron regiones genómicas relevantes para el número total de gazapos al parto (1 región), el número de nacidos vivos (1 región), los embriones implantados (3 regiones) y la tasa de ovulación (5 regiones). Los porcentajes de varianza genómica que explicaban los anteriores caracteres de tamaño de camada fueron 39.48%, 10.36%, 37.21% y 3.95%, respectivamente, en un modelo que excluye el efecto línea; y 7.36%, 1.27%, 15.87% y 3.95%, respectivamente, en un modelo con el efecto línea. La región genómica localizada en el cromosoma del conejo (OCU) 17 en 70.0 - 73.3 Mb se consideró como un nuevo locus de carácter cuantitativo (QTL) asociado a

caracteres reproductivos en conejos, ya que esta región fue encontrada solapada para el número total de gazapos al parto, el número de nacidos vivos y los embriones implantados. El gen de la proteína morfogenética ósea 4, *BMP4*, es el principal gen candidato prometedor dentro del nuevo QTL.

Una combinación de GWASs fueron implementados para analizar los datos genómicos del experimento de la grasa intramuscular con 480 conejos. Los métodos de GWASs incluyeron un método bayesiano, modelo Bayes B; y un método frecuentista, regresiones de marcadores únicos con los datos ajustados por el parentesco genómico. Este estudio reveló cuatro regiones genómicas relevantes en OCU1 (1 región), OCU8 (2 regiones) y OCU13 (1 región) asociadas con la grasa intramuscular. La región asociada más importante estaba en OCU8 en 24.59 - 26.95 Mb, y explicó el 7.34% de la varianza genómica. El bajo porcentaje explicado por las principales regiones genómicas relevantes indica un gran componente poligénico para la grasa intramuscular. Los análisis funcionales recuperaron genes vinculados con las rutas y funciones de los metabolismos de energía, carbohidratos y lípidos. Además, se realizó un estudio de escaneo genómico usando conejos del experimento de selección divergente para grasa intramuscular, y usando tres métodos de huellas de selección: índice de fijación de Wright (F_{st}), coeficiente de verosimilitud compuesto entre poblaciones (XP-CLR) y extensión de homocigosidad de los haplotipos entre poblaciones (XP-EHH). Los resultados mostraron múltiples huellas de selección en todo el genoma del conejo. Ninguna de estas huellas de selección concuerda con las regiones genómicas asociadas con la grasa intramuscular, provenientes de los resultados de los GWASs. En síntesis, los resultados de ambos experimentos, GWAS y el estudio de escaneo genómico, sugieren que la arquitectura genómica de la grasa intramuscular en el conejo parece ser altamente poligénica y sus variantes causales serían apenas detectables.

Este estudio demuestra que la detección de variantes causales y marcadores genéticos asociados depende de las hipotéticas arquitecturas genómicas de los caracteres, independientemente de las exitosas respuestas logradas en los dos experimentos de selección divergente. Aparte del nuevo QTL para el tamaño de la camada, ninguna de las regiones genómicas explicaba una gran parte de las varianzas genómicas de los caracteres del tamaño de la camada y de la grasa intramuscular en conejos. Por lo tanto, todos los caracteres analizados tienen

un gran componente poligénico. Análisis y estudios adicionales serán necesarios para confirmar los hallazgos del actual estudio de investigación. Hasta la fecha, estos hallazgos no tendrían implicaciones factibles para los programas de cría de conejos.

RESUM

La selecció divergent pot alterar les freqüències dels marcadors genètics en direccions oposades, donant lloc a freqüències al·lèliques intermèdies quan les dos línies divergents es consideren conjuntament en els anàlisis genètics. Per tant, els experiments de selecció divergents augmenten el poder de detecció en estudis d'associació de genoma ampli (GWAS) i en estudis d'exploració genòmica a través de mètodes de signatures de selecció.

A la Universitat Politècnica de València, es van dur a terme dos experiments independents de selecció divergent en conills, un per a la capacitat uterina i l'altre per al greix intramuscular. Els dos experiments van aconseguir respostes de selecció reeixides, sent 1.50 llorigons per a la capacitat uterina en la desena generació i 3.10 desviacions estàndard per al greix intramuscular en la novena generació, respectivament. Els animals d'aquests experiments es van usar per a realitzar anàlisis genòmics de caràcters de grandària de ventrada i greix intramuscular. Els genotips es van obtindré per mitjà d'una matriu d'alta densitat de polimorfisme d'un sol nucleòtid (SNP) de 200k.

GWASs bayesians, utilitzant el model Bayes B, es van implementar per a analitzar dades genòmiques de caràcters de grandària de ventrada de l'experiment de capacitat uterina amb 181 conilles femelles. Les associacions es van provar calculant els factors de Bayes per a cada SNP, i calculant els percentatges de la variància genòmica per a cada finestra no superposada d'1-Mb. Els GWASs van descobrir SNPs associats amb el número total de llorigons al part i els embrions implantats. A més, es van revelar regions genòmiques rellevants per al número total de llorigons al part (1 regió), el número de nascuts vius (1 regió), els embrions implantats (3 regions) i la taxa d'ovulació (5 regions). Els percentatges de variància genòmica que explicaven els anteriors caràcters de grandària de ventrada van ser 39.48%, 10.36%, 37.21% i 3.95%, respectivament, sota un model que exclou l'efecte de línia; i 7.36%, 1.27%, 15.87% i 3.95%, respectivament, sota un model amb efecte de línia. La regió genòmica situada en el cromosoma del conill (OCU) 17 en 70.0 - 73.3 Mb es va considerar com un nou locus de caràcters quantitatius (QTL) associat a caràcters reproductius en conills, ja que aquesta regió es va superposar per al número total de llorigons al part, el número de nascuts vius i els embrions

implantats. El gen de la proteïna morfogenètica òssia 4, *BMP4*, és el principal gen candidat prometedor dins del nou QTL.

Una combinació de GWASs es van implementar per a analitzar les dades genòmiques de l'experiment del greix intramuscular amb 480 conills. Els mètodes GWASs van incloure un mètode bayesià, model Bayes B; i un mètode frecuentista, regressions de marcadors únics amb les dades ajustades pel parentiu genòmico. Aquest estudi va revelar quatre regions genòmiques rellevants en OCU1 (1 regió), OCU8 (2 regions) i OCU13 (1 regió) associades amb el greix intramuscular. La regió associada més important estava en OCU8 en 24.59 - 26.95 Mb, i va explicar el 7.34% de la variància genòmica. El baix percentatge explicat per les principals regions genòmiques rellevants indica un gran component poligènic per al greix intramuscular. Els anàlisis funcionals van recuperar gens relacionats amb les rutes i la funció d'energia, metabolismes de carbohidrats i lípids. A més, es va realitzar un estudi d'exploració del genoma usant conills de l'experiment de selecció divergent per a greix intramuscular, i usant tres mètodes de signatures de selecció: índex de fixació de Wright (F_{st}), coeficient de versemblança compost entre poblacions (XP-CLR) i extensió de homocigosidad dels haplotipos entre poblacions (XP-EHH). Els resultats van mostrar múltiples petjades de selecció en tot el genoma del conill. Cap d'aquestes petjades de selecció concorda amb les regions genòmiques associades a partir dels resultats dels GWASs. En síntesi, els resultats dels dos experiments, GWASs i estudi d'exploració del genoma, suggereixen que l'arquitectura genòmica del greix intramuscular en el conill sembla ser altament poligènica i les seues variants causals serien a penes detectables.

Aquest estudi demostra que la detecció de variants causals i marcadors genètics associats depèn de les hipotètiques arquitectures genòmiques dels caràcters, independentment de les respostes reeixides en els dos experiments de selecció divergents. A part del nou QTL per a la grandària de la ventrada, cap de les regions genòmiques explicava una gran part de les variacions genòmiques dels caràcters de la grandària de la ventrada i del greix intramuscular en conills. Per tant, tots els caràcters analitzats tenen un gran component poligènic. Anàlisi i estudis addicionals seran necessaris per a confirmar les troballes de l'actual estudi d'investigació. Fins ara, aquestes troballes no tindrien implicacions valuoses per als programes de cria de conills.

RÉSUMÉ

La sélection divergente peut modifier les fréquences des marqueurs génétiques dans des directions opposées. Les études génétiques des lignées divergentes mènent à des fréquences alléliques intermédiaires. Par conséquent, les études d'association génétique pangénomiques (GWAS) et de balayage génomique, basées sur des expériences de sélection divergente, présentent un meilleur pouvoir de détection.

A l'Université Polytechnique de Valence, deux expériences indépendantes de la sélection divergente ont été menées chez le lapin, une pour la capacité utérine et une autre pour la graisse intramusculaire. Les deux expériences ont rapporté des réponses positives à la sélection, respectivement, 1.50 lapinous pour la capacité utérine à la 10-ème génération et 3.1 écart-type pour la graisse intramusculaire à la 9-ème génération. Des individus issus de ces expériences ont été utilisés pour effectuer des analyses génomiques de la taille de la portée et de la graisse intramusculaire. Les génotypes ont été obtenus en utilisant des puces à ADN de haute densité de 200K polymorphismes mono-nucléotidiques (SNP).

Des GWASs bayésiennes, du type Bayes B, ont été utilisées pour l'analyse génomique de la capacité utérine, et ont portées sur 181 individus. Les associations ont été testées en calculant le facteur de Bayes pour chaque SNP ainsi que le pourcentage de la variance génomique des fenêtres non superposés de 1-Mb. Ces analyses ont permis de détecter les SNPs associés au nombre total des naissances et des embryons implantés. De plus, des régions génomiques importantes ont été détectées pour le nombre total des naissances (une région), le nombre des nés-vivants (une région), les embryons implantés (3 régions), et le taux d'ovulation (5 régions). Les pourcentages de la variance génomique expliquées par les régions associées a ces caractères ont été respectivement de 39.48%, 10.36%, 37.21%, et 3.95% sous un modèle sans l'effet lignée, alors qu'elles ont été respectivement de 7.36%, 1.27%, 15.87% et 3.95% en incluant l'effet lignée. La région génomique 70.0-73.3 Mb au niveau du chromosome de lapin (OCU) 17 a été considérée comme un nouveau locus à caractère quantitatif (QTL) pour les caractères reproductifs du lapin. Cette région résulte de la superposition des régions associées aux nombre de naissances totales, des nés-

vivants et le taux des embryons implantés. Le gène de la protéine morphogénétique osseuse 4, *BMP4*, est le principal gène candidat prometteur au niveau de ce QTL.

Plusieurs analyses GWASs ont été performées pour l'analyse des données génomiques de la graisse intramusculaire de 480 individus. Ils ont inclus une méthode bayésienne, le Bayes B, une méthode fréquentiste et une régression à un seul marqueur avec les données corrigées pour la parenté génomique. Cette étude a révélé quatre régions génomiques pertinentes, OCU1 (une région), OCU8 (2 régions) et OCU13 (une région), associées à la graisse intramusculaire. La plus importante région a été OCU8 à 24.59-26.95 Mb en expliquant 7.34% de la variance génomique. Certes c'est un faible pourcentage, toutefois, il indique une importante composition polygénique de la graisse intramusculaire. Les analyses fonctionnelles ont identifié des gènes liés à des voies et fonctions énergétiques, ainsi qu'au métabolisme glucidique et lipidique. En plus, un balayage du génome a été performé en utilisant des lapins issus de l'expérience de la sélection divergente pour la graisse intramusculaire et trois méthodes des signatures génétiques de la sélection à savoir l'index de fixation de Wright (F_{ST}), le rapport de vraisemblance composite entre populations (XP-CLR) et l'extension de l'homozygotie de l'haplotype entre population (XP-EHH). De nombreuses signatures génétiques de la sélection ont été détectées tout au long du génome du lapin. Cependant, aucune n'a coïncidé avec les régions génomiques obtenues par l'analyse GWAS. Ainsi, à partir des résultats de l'analyse GWAS et du balayage du génome, parait-il que la graisse intramusculaire chez le lapin est très polygénique et la variante causale serait difficile à détecter.

Cette étude montre que la détection des variantes causales et des marqueurs génétiques associés dépend de l'architecture génomique hypothétique du caractère, en dépit des réponses positives obtenues par les deux expériences de la sélection divergente. A part le nouveau QTL détecté pour la taille de la portée, aucune région génomique n'a expliqué une importante part de la variance génomique de la taille de la portée et de la graisse intramusculaire. Par conséquent, les caractères étudiés présentent de larges composantes polygéniques. D'autres analyses seront nécessaires pour confirmer les résultats de la présente étude. Jusqu'à présent, ces résultats n'auraient pas

d'importantes implications dans les programmes d'amélioration génétique du lapin.

CHAPTER ONE

1. GENERAL INTRODUCTION

1.1. Rabbit production and breeding

European domestic rabbit (*Oryctolagus cuniculus*) belongs to the family *Leporidae* and the order *Lagomorpha*. At world level, the main purposes of raising rabbits are meat, wool, fur and an animal model for experimental research (McNitt *et al.*, 2013).

1.1.1. Overview

The domestication of rabbits was a dynamic process including several events; such as the transportation of rabbits to Mediterranean islands, rabbits housed in Roman leporaria, and they also reared under an organized reproduction system into hutches in southern France (Irving-Pease *et al.*, 2018). A genomic analysis from wild French and domestic rabbits suggests that rabbit domestication could be traced between 12,200 years and 17,000 years ago. Two consecutive bottlenecks occurred in the rabbit farming, impairing the genetic and phenotypic variability of fitness and adaptation traits (Carneiro et al., 2011; Irving-Pease et al., 2018). In this sense, the study of the domestication process showed phenotypic changes linked to polygenic background, essentially in behavioural traits influencing the brain and neuronal development (Carneiro et al., 2014). Besides, the domestication process would lead to changes in the litter size (Carneiro *et al.*, 2011).

1.1.2. Traits of interest in rabbit production

The rabbit, being a prolific species, presents beneficial features for breeding such as the small space of housing, a short productive cycle and handling facilities (McNitt *et al.*, 2013; Cullere and Dalle Zotte, 2018). At trade level, rabbit production involves items such as meat, fur, pet, and lab animals (McNitt *et al.*, 2013). Regarding meat production, consumer demand for rabbit meat has been decreasing in the last years, and hence, rabbit industry requires new strategies of production and marketing (Cullere and Dalle Zotte, 2018). Economic weights, derived from profit functions, have been estimated. The main traits having high economic weights are feed conversion rate (FCR) during

fattening and litter size (Prayaga and Eady, 2000; Cartuche *et al.*, 2014). Several experiments have been carried out to improve litter size; however, selection responses have been low, around 0.1 kits per generation (see in: Khalil & Al-Saef, 2008; Argente, 2016; Badawy *et al.*, 2018). As FCR is difficult and expensive to measure, FCR has been usually improved as a correlated response by selecting average daily gain (ADG). Although the genetic and phenotypic correlations between them are rather moderate, around 0.48; as also are their heritabilities, around 0.21 to 0.31 (Piles *et al.*, 2004; Blasco *et al.*, 2018). Conventional selection through ADG has attained positive genetic progress (see review by Blasco *et al.*, 2018), *e.g.* for ADG, around 3.9 g/d after 11 generations of selection in a Spanish rabbit line (Piles M. and Blasco A., 2003), and for FCR, -0.015 per generation in French rabbit lines under ad libitum feeding (Drouilhet *et al.*, 2016).

Lately, carcass and meat quality traits have been investigated for main livestock species because of the possibility of obtaining an additional economic value, supported by consumer's perceptions (Hocquette *et al.*, 2010; Pannier *et al.*, 2018). These most noticeable traits of meat quality are intramuscular fat and fatty acid profiles (Pena *et al.*, 2016; Cullere and Dalle Zotte, 2018; Blasco *et al.*, 2018). These traits are not selection criteria in rabbits since their recording is expensive and it has to be done in a large number of candidate animals to selection. Besides, rabbit meat has very low fat (Hernández, 2008; Martínez-Álvaro *et al.*, 2016a), and consequently, for example, an increase in intramuscular fat would not be perceptible by consumers (Martínez-Álvaro *et al.*, 2016a, 2016b).

1.1.3. Relationship between growth, meat and litter size traits

Animal breeding programmes for prolific livestock species generally comprise several traits, in which the traits are first chosen according to their economic values, and then according to their genetic parameters and correlations. Estimates of correlations between growth and carcass traits present a wide rank of values in pigs (Hoque *et al.*, 2007; Gilbert *et al.*, 2007; Kouba and Sellier, 2011), and rabbits (Hernández *et al.*, 2004; Larzul *et al.*, 2005; Su G. *et al.*, 2010). For instance, in rabbits, ADG has low-to-moderate genetic correlations with carcass traits (Larzul *et al.*, 2005; Nagy *et al.*, 2006); and besides, it has

null-to-low genetic correlations with meat quality traits (Hernández *et al.*, 2004; Blasco *et al.*, 2018). Thus, selection for growth rate in rabbit could have little effect in carcass and meat quality in rabbits.

Regarding the reproductive traits, they present low-to-moderate genetic correlations with growth traits in pigs (Holm *et al.*, 2004; Skovsted *et al.*, 2005; Rothschild and Ruvinsky, 2011), minks (Karimi *et al.*, 2018), mice (Malik, 1984), rats (Eisen, 1976), rabbits (Garreau *et al.*, 2004; Peiró *et al.*, 2019) and all polytocous species. In rabbits, estimations of genetic correlations amongst reproductive traits and body weight traits within lines were low (García and Baselga, 2002; Peiró *et al.*, 2019) or inconsistent (Camacho and Baselga, 1990), since studies suggested positive and negative correlations between them (Camacho and Baselga, 1990; Bünger *et al.*, 2005). For instance, genetic correlations between litter size and weaning weight, slaughter weight, and growth rate were -0.11, 0.03 and 0.11, respectively (Peiró *et al.*, 2019). García & Baselga (2002) found very low genetic correlation (0.06) between ADG and litter size. Furthermore, correlated responses have been null-to-low for growth and carcass traits in experiments of litter size selection in pigs (Estany *et al.*, 2002; Petry *et al.*, 2004) and rabbits (García and Baselga, 2002; Peiró *et al.*, 2019).

The estimated correlations amongst litter size and meat quality traits in pigs are low (Beaulieu *et al.*, 2010). Hitherto in rabbits, there is no known information about the correlation among litter size and meat quality traits. A preliminary unpublished study, at Universitat Politècnica de València, suggests low correlations between total number born with average daily gain and intramuscular fat, which have been estimated using rabbit data of a divergent selection experiment (Martínez-Álvaro *et al.*, 2019, personal communication).

Despite numerous research studies on selection for alternatives traits in order to improve rabbit meat production, the selection scheme and criteria have not changed. Breeding programs in commercial rabbit lines consist of a cross of three ways, in which the maternal line comes from a cross of two synthetic lines selected for number of kits at birth or weaning. The terminal sire line comes from a synthetic line selected for growth rate (Baselga, 2004; McNitt *et al.*, 2013).

1.2. Rabbits as animal experimental model

Rabbits are also used for human medical research as animal model (Calasans-Maia *et al.*, 2009; Mapara *et al.*, 2012). The availability of new genomic tools (*e.g.* next generation sequencing, NGS) in research has led to an increase of proteomic and genomic information in rabbits, being useful in specific fields of human research. Currently, muscular and circulatory systems have been fields of human research in which the rabbits have been useful as animal model (Miller *et al.*, 2014); for example, the implementation of new generation transgenic methods by employing rabbit models of cardiac disease (Bősze *et al.*, 2016). Thus, rabbits are very important for production and as an experimental model, which furthers investigation in the fields of genetics and genomics in this species.

1.3. Divergent selection experiments

Divergent selection experiments, as one-way selection experiments with control populations, are used to compute selection responses (Hill and Caballero, 1992). Furthermore, divergent selection experiments present advantage against one-way selection, such as: (1) to reduce the number of generations of selection, obtaining a double selection response of one-way selection whether there is symmetry in the response; (2) populations can be used as a control population of each other in the cases of symmetric selection responses, since experiments present populations with high and low performance for the selected trait; and (2) to reveal physiological information of selected traits, *e.g.* physiological boundaries of selected traits whether there are asymmetric responses (Hill, 1972; Bohren, 1975; Walsh and Lynch, 2018). The asymmetry in the response can be detected by a control population or by comparing the estimated genetic values in the last generation of selection; though this latter is strongly dependent on the mixed model techniques (Blasco *et al.*, 2018).

Various divergent selection experiments have been carried out in rabbits, having positive results of selection responses (Table 1). Particularly, the Universitat Politènica de València has performed pioneering research of divergent selection experiments in rabbits. These experiments presented symmetric selection responses (Blasco *et al.*, 1993, 2018; Argente *et al.*, 1997); for example, in an experiment of divergent selection for uterine capacity was attained a positive

response of ~1.5 kits after ten generations of selection (Blasco *et al.*, 2005). Furthermore, in a divergent-based experiment for intramuscular fat was obtained a response of 2.4 phenotypic standard deviations after eight generations of selection (Martínez-Álvaro *et al.*, 2018). Selection can modify frequencies of genetic markers in opposite directions leading to intermediate allelic frequencies when both lines are jointly regarded in genetic analyses. Under this condition, the detection power can be increased for genome wide association (GWAS) and genomic scan studies, *e.g.* methods of selection signatures (Kessner and Novembre, 2015). Thereby, exceptional biological samples can be obtained from divergently selected populations in order to study the relationship between the selected trait and the rabbit genome. This approach has been implemented in other livestock species, such as poultry (Johansson *et al.*, 2010; Grams *et al.*, 2015) and pigs (Kim *et al.*, 2015a), displaying positive results with respect to genome-wide analyses.

1.4. Molecular genetics for uncovering the importance of genomics in rabbit

Genetic improvement of livestock has been relying on selective breeding with superior phenotypes. With the advent of the genomic tools and techniques, the traditional artificial selection has been modified due to the interest in applying genomic information on animal breeding programs (Gurgul *et al.*, 2014; Singh *et al.*, 2014; Mehrban *et al.*, 2017). Animal genomics is not only genomic selection, but it encompasses the study of polymorphisms, genes, functional genomics and linkage disequilibrium by methods such as parentage determination, genome-wide association study, signatures of selection, genetic expression, marker-assisted selection, and so on (Gurgul *et al.*, 2014). Genomic information is mainly worthy for traits having specific features related to the data collection and related to some specific populations or breeds (Ibáñez-Escriche and Gonzalez-Recio, 2011; Schmid and Bennewitz, 2017). This information can be relevant for traits of low-to-moderate heritability, traits of sex limited expression, and when a trait is collected late in an individual's life (Singh *et al.*, 2014). Most quantitative or complex traits in livestock breeding are controlled by many genes; therefore, some genomic analyses, *e.g.* GWAS and genome scan studies, can shed light on how the livestock genome is linked to

the variability of economically important traits, since these studies can be performed using the whole genome of particular livestock.

Table 1. Selection response of divergent selection experiments in rabbits.

Trait (selection criterion)	Selection response	Number of generations	Reference
112-days body weight	126.60 grams	4	(Mgheni and Christensen, 1985)
Number of foetus (dead embryos between implantation and birth)	- 0.50 foetus	4	(Bolet *et al.*, 1994) (Santacreu *et al.*, 1994)
Average daily gain	8.40 g/d	3	(Moura *et al.*, 1997)
Feed conversion rate	0.20	3	(Moura *et al.*, 1997)
Uterine capacity	1.50 kits	10	(Blasco *et al.*, 2005)
63-days body weight	450 g	6	(Larzul *et al.*, 2005)
Locomotor activity score in the open field	~ 50 OFS	8	(Daniewski and Jezierski, 2003)
Total fleece weight	80.95 grams	Overlapped generation* High line = 3.90 Low line = 3.64	(Rafat *et al.*, 2009) (Rafat *et al.*, 2007)
Computer tomography (thigh muscle weight)	~ 27 cm^3	2	(Szendro *et al.*, 2012)
Intramuscular fat	0.39 grams	6	(Martínez-Álvaro *et al.*, 2016a)
Residual variance	1.67 $kits^2$	10	(Blasco *et al.*, 2017)
Variability of birth weight	3.9 grams	10	(Bodin *et al.*, 2010)
Longevity index (number of AI)	+0.75 AI (32 days)	2 (does was kept until 8th litter)	(Larzul *et al.*, 2014)
Digestive disorders (EBV – ES)	~ 1.7% CM	1	(Garreau *et al.*, 2012)

OFS = open-field score; AI = artificial insemination; EBV – ES = estimated breeding value for the enteropathy score; CM = cumulative mortality; Overlapped generation* = equivalent number of discrete generations in the selection for total fleece weight of high and low lines.

1.4.1. Genome wide association study

Nowadays, the GWAS is the leading strategy used to find genetic markers throughout the whole genome associated, with important economic production traits. GWAS exploits the linkage disequilibrium (LD) between the genetic causative variants of the alleles of a trait and genetic markers. GWAS tries to find effective signals of association of these genetic markers with a given trait when it is controlled by particular causative variants large enough to be detected (Spencer *et al.*, 2009; Ball, 2013).

1.4.1.1. Previous to GWAS: genetic markers and genomic analyses

Molecular markers, also termed genetic markers, are polymorphic fragments of DNA with a certain location in the genome (Vignal *et al.*, 2002; Stram, 2014). Some genetic markers are: amplified fragment length polymorphism, restriction fragment length polymorphism (RFLP), copy number variants, single nucleotide polymorphisms (SNPs), insertion/deletion, minisatellites and microsatellites, and sequences (Teneva and Petrovic, 2010; Singh *et al.*, 2014). Several techniques are used to obtain polymorphisms which are still used according to their price and automatisation (Vignal *et al.*, 2002; Stram, 2014).

In the past, linkage analyses were the main association studies for genetic maps. They were based on the recombination rate of genetic markers such as RFLP or microsatellites (Hearne *et al.*, 1992; Witte, 2010; Singh *et al.*, 2014). Linkage analyses presented problems in complex traits for discovering the causative variants, since the associated regions were large, limited by the annotation on the genetic map, and had a low resolution because of the limited number of meiosis within families (Tabor *et al.*, 2002; Witte, 2010).

Another method of association analysis is the candidate gene study. This method evaluates the association between a trait of interest with some genetic markers placed inside a known gene (Tabor *et al.*, 2002; Patnala *et al.*, 2013). The principal problem of this method has been its dependence on the prior biological information of genes for using the candidate genes. This method ignores most of the genome and therefore the analyses are probably missing many causative regions, leading to many false-positive associations (Zhu and

Zhao, 2007; Witte, 2010). Despite these drawbacks, some genetic markers within genes associated with a trait were identified by this method. For instance, there were genetic markers located in the genes or gene regions *K88*, *FUT1*, *SLA* and *NRAMP* (associated with immune response or disease resistance traits), and *ESR*, *PRLR*, *RBP4* and *FSHB* (associated with litter size) in pigs (Rothschild *et al.*, 2007). Genes with genetic markers showing associations in rabbits were: *OVGP1* (Merchán *et al.*, 2009), *PGR* (Peiró *et al.*, 2010) and *TIMP1* (Argente *et al.*, 2010). These markers were associated with litter size traits, having small effects (Merchán *et al.*, 2009; Argente *et al.*, 2010; Peiró *et al.*, 2010). In addition, a genetic marker in the MSTN gene was associated with carcass traits in rabbits (Sternstein *et al.*, 2014). To our knowledge, these associated markers within these genes have not been validated in additional rabbit populations. Besides, when these markers were found, there were limitations such as non-availability of SNP platforms, low resolution of the genomic map and little functional information in rabbits, hindering exhaustive researches throughout the rabbit genome (Ibáñez-Escriche and Gonzalez-Recio, 2011; Miller *et al.*, 2014).

With the development of NGS, the SNP became the most popular polymorphism. Currently, it is the most abundant of all marker systems, both in animals and plants (Vignal *et al.*, 2002; Gurgul *et al.*, 2014; Sharma *et al.*, 2015). SNP is defined as a single nucleotide variant (SNV) with two alleles (biallelic genetic marker), at a specific position on a given chromosome, wherein the least common allele has a frequency of about 1% or greater (Vignal *et al.*, 2002; Stram, 2014). An advantage of the SNPs in animal breeding studies was the creation of low cost SNP platforms for genotyping the main livestock species (Sharma *et al.*, 2015). SNPs have received attention due to their genetically stability and amenability to high-throughput automated analysis (McCarthy and Hirschhorn, 2008; Singh *et al.*, 2014). Besides, SNPs are widely used for implementing GWAS and genomic scan studies, albeit in rabbits the high cost of SNP platforms hampered its usefulness for research studies (Ibáñez-Escriche and Gonzalez-Recio, 2011).

1.4.1.2. Methods and strategies

GWASs depend chiefly on LD between genes and genetic markers, *i.e.* SNPs. In this way, GWASs can capture the effect of several genes when SNPs have high

LD with the causative variants of these genes (Spencer *et al.*, 2009; De Los Campos *et al.*, 2013). The GWAS analyses also rely on several factors, *e.g.* type of trait (case-control traits or quantitative traits), the genomic architecture of traits (additive, dominant, traits with imprinting, etc.), the population features (populations under natural or artificial selection), sample size, and inference methods (frequentist or Bayesian inferences). These factors are important before performing a genomic analysis, since they determine the scope, strengths, limitations and strategies of the experiments (Witte, 2010; Hayes, 2013).

Designing association studies

The factors influencing GWAS results are described in the following paragraphs:

Choosing the SNP-array density. In rabbits, a SNP platform of ~200k SNP was made available to genotype rabbit at research level (Blasco and Pena, 2018). High SNP-density platforms are useful for GWASs when a genome presents a wide variability of LD between SNPs. For example, populations with large hypothetical effective population size would need to use platforms of high SNP-density in order to analyse the whole genome. This is especially important for across-breed GWAS experiments due to their high genomic variability. The high SNP-density platforms (*e.g.* SNP density of 777K in cattle), having a homogenous distribution of SNPs across a given genome, can increase the detection power of association between SNPs with causative variants of a particular trait (Spencer *et al.*, 2009; Schmid and Bennewitz, 2017). These platforms can increase the number of SNPs close to causative variants, which would also increase the detection power (Schmid and Bennewitz, 2017). Nonetheless, this depends mainly on the analysed trait and its genomic architecture. In addition, a shorter physical genomic distance between SNPs could not be interpreted as a high LD between SNPs and causative variants improving detection power and consequently to the GWAS results. For instance, platforms higher than 650 k SNP in human can produce a negligible change in improving detection power (Spencer *et al.*, 2009).

Taking into account the genomic map. The resolution of the genomic map is important for identifying the genes around associated SNPs. There are a few known maps in rabbits: a linkage map by crossing Giant Gent and New Zealand

White (Sternstein *et al.*, 2015), a microsatellite-based integrated genetic and cytogenetic map (Chantry-Darmon *et al.*, 2006), a physical map of sequence assembly having 2X coverage (https://may2009.archive.ensembl.org/Oryctolagus_cuniculus/Info/Index), and the current physical map of 7X coverage, named OryCun2.0 (Miller *et al.*, 2014; Carneiro *et al.*, 2014). Despite the availability of the rabbit genome map, the functional annotations of the genes in this species are much less broad than in other species (humans, mice, and cattle), which hinders the identification of promising candidate genes (Craig *et al.*, 2012; Miller *et al.*, 2014).

Knowing the features of populations. Samples for GWAS experiments must represent statistically independent units drawn from populations (Sul *et al.*, 2018). GWAS can present problems of confounding factors, such as population structure and cryptic relatedness, when samples come from several populations or high inbreeding populations (Price *et al.*, 2010). Population structure occurs when systematic differences are present in allele frequencies between subpopulations in a population because of different ancestry, especially in case-control association studies. These systematic differences can cause the method applied to GWAS to assign strong association signals to genetic markers that are not actually causal for the trait. In a nutshell, population structure produces a high rate of false positive associations named as inflation of type I errors (Price *et al.*, 2010; Sul *et al.*, 2018). The cryptic relatedness appears when the relationships between individuals are unknown or ignored by researchers, albeit the individuals shared or had common ancestries. This confounding factor can also increase the rate of false positive associations. This can affect the association analysis under the presence of family-based samples, *e.g.* parent-offspring, full-siblings (Voight and Pritchard, 2005; Sul *et al.*, 2018). Nevertheless, the modelling of relatedness can be carried out using genomic data, pedigrees or both (Price *et al.*, 2010; Li and Zhu, 2013; Eu-ahsunthornwattana *et al.*, 2014). Another essential point of features of populations is to know whether the analysed populations are currently under selection. This is linked to tests and filters of quality controls, affecting the GWAS results. For instance, Hardy-Weinberg equilibrium test indicates genotyping errors in non-selected populations; on the contrary, this test can remove relevant SNPs associated with a given trait when populations are under selection (Marees *et al.*, 2018).

Choosing the sample size. Large sample sizes must ensure that a given study of association clearly distinguishes between real and spurious associations. GWASs look for strong evidence of association. Therefore, GWASs with low detection power lead to a higher presence of spurious associations. Frequently, at least 1000 genotyped and phenotyped individuals are required for statistical analysis (Ball, 2013; Schmid and Bennewitz, 2017). The samples should have animals as unrelated as possible avoiding the problem of confounding factors. The sample size for GWASs may be reduced when phenotypic records of traits are collected in experimental populations through exhaustive phenotyping at a molecular level (*e.g.*, metabolic traits or gene expression traits) (Schmid and Bennewitz, 2017). When phenotypic data come from experiments of divergent selection, the sample size may also be reduced since these experiments increase the detection power (Kessner and Novembre, 2015).

In synthesis, researchers should take into account for designing an association study: (a) the SNP-array density, (b) the genomic map, (c) the features of population, and (d) the sample size. All of them are directly or indirectly affecting the detection power which depends on LD between the markers and a given causative variant, and also, on the genomic architecture. This latter includes the effect size of causative variants and the proportion of total phenotypic variance accounted for by every causative variant (De Los Campos *et al.*, 2013; Hayes, 2013). Causative variants of large effect can be easily detectable even using small sample sizes when they are segregating in a population. Besides, causative variants with both a high minor allele frequency (MAF) and LD can be detected regardless of their effect sizes (López de Maturana *et al.*, 2014). Note that LD has been always an essential factor. In a nutshell, true marker associations cannot be detected in the absence of high LD.

There are numerous GWASs that were published after the availability of SNP platforms in livestock species with positive results (Sharma *et al.*, 2015). Nonetheless, there has been a problem concerning the lack of replication of the positive findings from GWASs. This lack of replication of genetic markers, previously associated, shows the need for larger samples, control of the population's differences and stronger statistical evidence priors. Bearing in mind this GWAS issue and a reasonable cost, the incorporation of genome

sequences with imputation strategies has been investigated as a new strategy for GWAS in order to place the causative variants. However, this method would rely strongly on the quality of sequencing and imputation (see review by Schmid & Bennewitz, 2017).

Frequentist inference

The GWAS in the past was mainly carried out by using of single marker regression (SMR) models and frequentist statistic, *i.e.* p-values. SMR consist of performing marker-by-marker analysis, testing every association between genetic marker or SNP and a given phenotypic database for a trait. P-values, used by frequentists, are defined as the probability than under conceptual future repetitions of study would generate stronger evidence for association than actually observed test statistic when the null hypothesis were true. P-values are usually easily calculated (Stephens and Balding, 2009; Ball, 2013). But, how to use and interpret the p-value? The test of association is done by testing of every SNP effect, being different from zero, which result in a p-value. These values are mistakenly construed as "the probability of being wrong". In addition, p-values are commonly misinterpreted as a measure of significance, as also the probability of the null hypothesis being true, *e.*g. no having SNPs associated with a given trait; and moreover, as evidence that supports the alternative hypothesis (Ball, 2013; Blasco, 2017; Schmid and Bennewitz, 2017). More details of common misinterpretation for p-values are in Blasco (2017). Furthermore, p-values depend on sample size, test setup and experimental design. In turn, these values also change by experiment; thus, p-values are troublesome in a sequential sampling application (Ball, 2013).

The number of phenotypic data (sample size) and the allele frequency of SNPs influence the estimation of allelic effect of the SNPs. This problem, termed "large p small n", arises when the number of genetic markers (p) vastly exceeds the number of samples with phenotypic record (n). This situation can lead to issues of high-dimensional feature space and consequently, problems of computation (De Los Campos *et al.*, 2013; Shen *et al.*, 2013; Mei and Wang, 2016). Under the frequentist inference, the models would fit the whole genome based on penalized likelihood (linear mixed models and ridge regression). Although, common GWASs normally avoid high-dimensional models and turns the issue into

multiple testing problem instead. Thus, the routines for implementing of SMR sacrifices both detection and prediction power (Shen *et al.*, 2013; Mei and Wang, 2016). This problem is utterly more important in whole-genome regression and prediction methods than in GWASs (De Los Campos *et al.*, 2013).

Under a predefined significance threshold, *e.g.* 0.05, p-values fall below this threshold are considered to be significant. Choosing a significance threshold is a problem in GWAS since the threshold should regard and account for the huge amount of multiple testing performing (Ball, 2013; Schmid and Bennewitz, 2017). The most common corrections of multiple testing are the Bonferroni correction (or an alternative option adjusted by LD blocks), permutation testing and the false discovery rate - FDR (Hayes, 2013; Schmid and Bennewitz, 2017; Marees *et al.*, 2018). The Bonferroni correction is very stringent, especially in high SNP-density platforms (*e.g.* over 700K SNPs), and considers all independent tests each other. As this assumption is wrong, alternative approaches are used to discover associations having a trade-off between the detection power and the inflation of type I errors, such as the above mentioned corrections (Li *et al.*, 2012; Schmid and Bennewitz, 2017; Marees *et al.*, 2018). In human, the current conventional threshold is 5×10^{-8} (Ball, 2013). This is based on threshold of 0.05 fitted for the equivalent number of independent comparisons conforming to a dense set of genetic markers. Moreover, short LD blocks are widely present in the human genome, explained mainly by few non-inbreeding matings. This corroborates the presence of a large number of independent testing for most of traits in humans (Duggal *et al.*, 2008; Ball, 2013). In livestock, the number of long LD blocks increases strongly under artificial selection and mating between close relatives, named as inbreeding populations. Commercial livestock populations have also small effective population size. Hence, the number of independent testing in livestock is commonly much lower than in human (Schmid and Bennewitz, 2017). Thus, this suggested the inclusion of a lower "suggestive" threshold, 1×10^{-4}, for livestock GWASs (Sahana *et al.*, 2011; Bertolini *et al.*, 2018; Do *et al.*, 2018).

Population structure and cryptic relatedness augment false positive associations in GWAS. The main methods uncovering and controlling the confounding factors are: genomic control, structured association, multidimensional scaling and principal component analyses. The correction of

the tests can be accomplished by adjusting the inflation of test by the population structure, or by including principal components as fixed effects in the model (Li and Zhu, 2013). These methods deal efficiently with the problems derived from population structure; however, decline their impact on cryptic relatedness with family-based data (Eu-ahsunthornwattana *et al.*, 2014; Sul *et al.*, 2018). Different frequentist models are used for GWAS. An alternative way to avoid cryptic relatedness and consequently the inflation of type I errors is applying linear mixed models with a random polygenic effect of the individuals (Li and Zhu, 2013; Sul *et al.*, 2018). The polygenic effects are normally distributed with mean zero and an additive genetic variance multiplied by a variance-covariance matrix. This latter may be based on either pedigree or genomic kinship - $\boldsymbol{GRM}$ (VanRaden, 2008). The estimator of $\boldsymbol{GRM}$ varies from field to field according to use of allelic frequencies and statistical modelling (Wang, 2016; Goudet *et al.*, 2018). For example, in human genetics, estimators are frequently used from equations of Yang *et al.* (2010) and Ritland (1996), and in the models for animal breeding programs are commonly used the procedures of VanRaden (2008) and Yang *et al.* (2010). The main difference between them lies in the assumption on the contribution of the genetic markers and their allelic frequencies. The $\boldsymbol{GRM}$ of VanRaden (2008) is calculated from the following equation:

$$\boldsymbol{GRM} = \frac{\boldsymbol{ZZ'}}{2\sum_{i=1}^{k} p_i q_i}$$

where $\boldsymbol{Z}$ is an incidence matrix of marker effects, $\boldsymbol{Z'}$ is an inverse of incidence matrix of marker effects, k is the number of genetic markers, p_i is the allelic frequency of one allele at marker i, and q_i is the allelic frequency of another allele at marker i (VanRaden, 2008; Clark and Van Der Werf, 2013; Goudet *et al.*, 2018). A second matrix creates $\boldsymbol{GRM}$ weighting every marker differently (VanRaden, 2008; Yang *et al.*, 2010). The $\boldsymbol{GRM}$ would be:

$$\boldsymbol{GRM} = \frac{1}{k}\sum_{i=1}^{k} \boldsymbol{GRM}_i = \frac{1}{k}\sum_{i=1}^{k} \frac{\boldsymbol{z_i z_i'}}{2\, p_i q_i}$$

where $\boldsymbol{z_i}$ is a vector with genotypes for genetic marker i, $\boldsymbol{z_i'}$ is an inverse of a vector with genotypes for genetic marker i. The rest of parameters in the equation are the same than in $\boldsymbol{GRM}$ of VanRaden (2008). The estimates of $\boldsymbol{GRM}$

elements of this second matrix are relative to the base population, in which the average of the relationship between individuals is zero. This ***GRM*** presents problems compared with the first one: it is very sensitive to small allelic frequencies because this gives high weight to very rare alleles; besides, it gives large effects to genetic markers with small allelic frequencies because the ***GRM*** assumes that the contribution of every marker to the overall ***GRM*** are identical. Hence, this ***GRM*** sets different a priori variances to the genetic markers, depending on their frequencies (Clark and Van Der Werf, 2013; Legarra, 2016; Goudet *et al.*, 2018). On the other hand, there is a concern about whether a given SNP to be tested for its association (and the other SNPs in high LD with the given SNP) should be used to establish the ***GRM*** or not. The initial hypothesis poses that if the tested SNP is included in the GRM, this could increase the rate of false negatives. Therefore, several authors have recommended the exclusion of all SNPs that are located on the same chromosome as the SNP to be tested from the GRM (Yang *et al.*, 2014; Schmid and Bennewitz, 2017). Nevertheless, other authors suggest that double-fitting of the SNP effects is a less severe problem than previously thought, having negligible impact on the results (Gianola *et al.*, 2016).

There are various models, methods and software for GWAS by SMR. Some of them involve linear regression models, such as TASSEL, TASSEL+P3D, EMMAX, FaST-LMM$_{full}$, FaST-LMMlow$_{full}$, GenABEL (FASTA and GRAMMAR), GEMMA and so on (see reviews by: Li & Zhu, 2013; Eu-ahsunthornwattana *et al.*, 2014; Yang *et al.*, 2014). Currently, the most common software is Genome wide Complex Trait Analysis - GCTA (Yang *et al.*, 2011). Another approach is instead to identify a particular SNP, to identify a genomic region. This approach calculates the heritability of each genomic region using consecutive SNPs within each particular region. It takes into account that a large proportion of genetic variation is captured by common linked SNPs clusters (Caballero *et al.*, 2015; Visscher *et al.*, 2017). Under this condition, the method recommended is the estimation of the heritability of genomic regions using regional heritability software - REACTA (Caballero *et al.*, 2015; Shirali *et al.*, 2016). However, all of the above mentioned methods still present the problems concerning the p-values and multiple testing. Nowadays, numerous researchers ask for the retirement of the p-value and statistical significance as the procedures of interpretation of results when an actual effect exists (Amrhein *et al.*, 2019;

McShane *et al.*, 2019; Wasserstein *et al.*, 2019). Besides, some researchers have suggested that p-values and statistical significance should be replaced by indicators showing stronger evidence for each research hypothesis, *e*.g. test using Bayesian statistic (Stephens and Balding, 2009; Ball, 2013; Blasco, 2017).

Bayesian inference

The Bayesian inference is increasingly used in GWAS because this statistical school employs probability distributions given the observed data (Stephens and Balding, 2009; Ball, 2013; Blasco, 2017). Bayesian inference is based on the use of probability for stating uncertainty. Calculating posterior probability distributions are carried out by specifying probability statement based in the observed data and prior distributions that are normally subjective priors. These priors summarize knowledge about unknown before the observed data are considered. Moreover, the Bayesian inference provides a description of how existing knowledge is changed by experience. In brief, statisticians use Bayes' theorem to turn the prior distribution into a posterior distribution. This latter distribution is used to describe and obtain the results of a given research study (Sorensen and Gianola, 2002; Stephens and Balding, 2009; Blasco, 2017).

Bayesian multi-marker (BMM) models fit a large number of SNPs simultaneously as random effects in the model under distinct shrinkage assumptions (Gianola, 2013; López de Maturana *et al.*, 2014; Schmid and Bennewitz, 2017). The main difference amongst BMM models is essentially the information on prior distributions of SNPs effect sizes. These models include Bayes A, B (Meuwissen *et al.*, 2001), LASSO (Park and Casella, 2008), C, Cπ, D, Dπ (Habier *et al.*, 2011), R (Erbe *et al.*, 2012), IM (Wilson-Wells and Kachman, 2016) and so on. A brief description of the main BMM is presented below:

- Bayes A: assumes a t distribution of SNPs effects, depending on the degree of freedom and the scale parameters, characterized by thick-tailed prior. All SNPs contributes to the trait variance (Meuwissen *et al.*, 2001).
- Bayes B: the same as Bayes A but assuming a small proportion of SNPs that have effects on the trait variance. The proportion is assigned by the researcher. Every SNP has a specific variance (Meuwissen *et al.*, 2001).

- LASSO: assumes a shrinkage under a double exponential distribution. The method estimates the parameters under penalized regressions (Park and Casella, 2008).
- Bayes C and Cπ: the first is as the Bayes B model but assuming a normal distribution of SNPs effects. All SNPs effects have a common variance. In Bayes Cπ, the proportion of SNPs with effect are estimated derived from data, as the proportion is treated as an unknown parameter (Habier *et al.*, 2011).

Nowadays, Bayes C, B and R are most commonly applied models for Bayesian GWASs (Sharma *et al.*, 2015; Fernando *et al.*, 2017; Schmid and Bennewitz, 2017). The Bayes B reduces the effects of sampling noise, producing associations inside or very close to major genes (De los Campos *et al.*, 2013; Ros-Freixedes *et al.*, 2016; Fernando *et al.*, 2017). This could be convenient when the sample number is small for traits presenting evidence of major genes. GWASs using Bayesian inference are more robust in detecting of genetic markers associated with a given trait than using frequentist inference. These GWASs can efficiently deal with the "large *p* small *n*" problems (De Los Campos *et al.*, 2013) and confounding factors (Toosi *et al.*, 2018). Hence, BMM methods demonstrate a larger detection power and smaller type I errors compared with SMR methods (López de Maturana *et al.*, 2014). However, BMM methods depend on the variance scaling factor, the degrees of freedom, and the proportion of the SNPs having a zero effect on the variance of a given trait (denoted as π in Bayes B, C and D). In this sense, the wrong choice of these former parameters could entail some problems on the detection power and precision of the results, over- or under-estimating the SNPs effects (Habier *et al.*, 2011; Lehermeier *et al.*, 2013; Gianola, 2013). Moreover, the length of Markov chain Monte Carlo (MCMC), being a computational method, can influence the sampling for creating the posterior distribution and, consequently, the SNP effect estimation. Standard MCMC presents high computational cost and is inefficient in the presence of high-dimensional data. Therefore, other alternatives have been proposed and developed, *e.g.* Bayesian hierarchical variable selection (Schmid and Bennewitz, 2017; Zhao *et al.*, 2019).

For Bayes B, C, and D, the inference about each SNP association with a given trait can be tested using Bayes factors. These are obtained using the marginal

posterior distribution of each SNP effect (Stephens and Balding, 2009; Habier *et al.*, 2011; Schmid and Bennewitz, 2017). Bayes factors are defined as the ratio between the posterior odds ratio and the prior odds ratio derived from:

$$prior\ odds\ ratio\ = \frac{\mathrm{P}(H_1)}{\mathrm{P}(H_0)}$$

$$BF\ = \frac{\mathrm{P}(y \mid H_1)}{\mathrm{P}(y \mid H_0)}$$

$$posterior\ odds\ ratio\ = \frac{\mathrm{P}(H_1 \mid y)}{\mathrm{P}(H_0 \mid y)} = \frac{\frac{\mathrm{P}(y \mid H_1)\ \cdot\ \mathrm{P}(H_1)}{\mathrm{P}(y)}}{\frac{\mathrm{P}(y \mid H_0)\ \cdot\ \mathrm{P}(H_0)}{\mathrm{P}(y)}}$$

$$posterior\ odds\ ratio\ = \frac{\mathrm{P}(y \mid H_1)}{\mathrm{P}(y \mid H_0)}\ \cdot\ \frac{\mathrm{P}(H_1)}{\mathrm{P}(H_0)} = BF\ \cdot\ \frac{\mathrm{P}(H_1)}{\mathrm{P}(H_0)}$$

$$posterior\ odds\ ratio\ =\ BF \cdot prior\ odds\ ratio$$

$$BF\ = \frac{posterior\ odds\ ratio}{prior\ odds\ ratio}$$

where $\Pr(H_1 \mid y)$ is the probability of an alternative hypothesis under the observed data, $\Pr(H_0 \mid y)$ is the probability of a hypothesis of a zero-effect SNP under the observed data, $\Pr(y \mid H_1)$ is the probability of the data under an alternative hypothesis, $\Pr(y \mid H_0)$ is the probability of the data under a hypothesis of a zero-effect SNP, $\Pr(H_1)$ is the prior probability of the alternative hypothesis, $\Pr(H_0)$ is the prior probability of the hypothesis of a zero-effect SNP, and BF denotes the Bayes factors (Sorensen and Gianola, 2002; Wakefield, 2012; Blasco, 2017). The use of *BF* allows to compare the posterior probabilities of two hypotheses (Stephens and Balding, 2009; Blasco, 2017). *BF* using Bayes B, C or D can be calculated considering the posterior probability of each SNP at locus j, denote as $\widehat{p_j}$, and a prior π, representing the proportion of the SNPs having a zero-effect on the variance of a given trait (Ros-Freixedes *et al.*, 2016):

$$BF = \frac{posterior\ odds\ ratio}{prior\ odds\ ratio} = \frac{\frac{\mathrm{P}(H_1 \mid y)}{\mathrm{P}(H_0 \mid y)}}{\frac{\mathrm{P}(H_1)}{\mathrm{P}(H_0)}} = \frac{\frac{\widehat{p_j}}{(1 - \widehat{p_j})}}{\frac{(1 - \pi)}{\pi}}$$

$$BF = \frac{\frac{\widehat{p_j}}{(1 - \widehat{p_j})}}{\frac{(1 - \pi)}{\pi}}$$

The evidence of association of a given SNP can be considered "substantial" when its BF is above 3.2, "strong" above 10, and "decisive" above 100 (Kass and Raftery, 1995). The chief problem of *BF* is that the probabilities of the hypotheses are sensitive to the prior information of the unknown parameters within the hypotheses. Moreover, *BF* are only ratios of posterior probabilities when prior probabilities are the same. *BF* will be difficult to compute and to interpret under other conditions, particularly in complex models (Stephens and Balding, 2009; Blasco, 2017). Posterior probabilities of association, $\mathrm{P}(H_1 \mid y)$, are represented as PPA by some researchers. PPA are better indicators for defining the association importance rather than the use of *BF*. These latter (*BF*) has to be very large to provide convincing evidence for an association as π is taken generally so small (Stephens and Balding, 2009; Purfield *et al.*, 2014). The PPA can be expressed using BF and π. Hence, PPA does not depend on detection power and sample size (Stephens and Balding, 2009). PPA are easy of calculating through the following equations:

$$PO = BF \cdot \frac{(1 - \pi)}{\pi} = \frac{\widehat{p_j}}{(1 - \widehat{p_j})}$$

$$\mathrm{PPA} = \frac{PO}{(PO + 1)} = \frac{\frac{\mathrm{P}(H_1 \mid y)}{\mathrm{P}(H_0 \mid y)}}{\frac{\mathrm{P}(H_1 \mid y)}{\mathrm{P}(H_0 \mid y)} + 1} = \frac{\frac{\widehat{p_j}}{(1 - \widehat{p_j})}}{\frac{\widehat{p_j}}{(1 - \widehat{p_j})} + 1} = \widehat{p_j}$$

Where PO is the posterior odds ratio, and the rest of the parameters means the same than in the above equations. In a nutshell, PPA is $\widehat{p_j}$, the posterior probability of each SNP at a SNP at locus j.

On the other hand, consecutive SNPs that surround a causative variant of a gene can jointly explain better the effect of this causative variant than using a single SNP. Defining genomic windows, using a physical distance or the number of consecutive SNPs, can improve the detection power. This approach also reduces the sampling error due to the presence of reasonable hitchhiking phenomena derive from the LD produced by selection (Garrick and Fernando, 2013; Beissinger *et al.*, 2015; Hoban *et al.*, 2016). There are three types of approaches for delineating boundaries of windows: sliding-windows, distinct-windows and the optimization of window size using smoothing spline techniques. Sliding-windows consist of defining each window with a constant window length and an interval for moving along the genome. This interval is lower than the window size leading to overlapping of adjacent windows. This approach allows to refine the precise locations of windows associated with a given trait (Tang *et al.*, 2009). However, as the number of tests increases, an inflation of type I errors is produced (Beissinger *et al.*, 2015; Hoban *et al.*, 2016). Distinct windows consist of defining each window with a constant window length without being overlapped along the genome (Garrick and Fernando, 2013; Beissinger *et al.*, 2015). This approach shows lower inflation of type I errors than the sliding-window approach (Beissinger *et al.*, 2015; Hoban *et al.*, 2016). The optimization of windows size using smoothing spline techniques determines the ideal window size along the genome from the genomic database. This approach can attain a greater number of detected QTL, controlling the inflation of type I errors at the same time, compared with the two former approaches (Beissinger *et al.*, 2015). The importance of the associations can be defined using the conventional threshold of 1% of genomic variance explained by a window (Garrick and Fernando, 2013) or the window posterior probability of association (WPPA). This latter criterion is appropriate to calculate evidence of association for a window, since WPPA is not influenced by the SNP-density in the model (Fernando *et al.*, 2017; Schmid and Bennewitz, 2017). The implementation of Bayesian GWAS can be performed using GenSel (Garrick and Fernando, 2013) and BGLR software (Pérez and De Los Campos, 2014).

1.4.2. Genome scan study

Most traits in livestock lack major genes ruling the trait variation (Sharma *et al.*, 2015; Goddard *et al.*, 2016; Schmid and Bennewitz, 2017). Revealing causative genes by GWAS is like finding a needle in the haystack. The underlying causative genes have only been identified in few exceptional cases (Schmid and Bennewitz, 2017). The implementation of genomic scans studies using animals from divergent selections can improve or confirm GWAS results (Qanbari and Simianer, 2014; Kim *et al.*, 2015a). In this way, both results, from GWAS and genomic scan study, would give a better strong evidence of the positions of putative causative genes.

1.4.2.1. Methods of selection signatures

In absence of phenotypic information, genomic scan analyses contribute to elucidate selective signals across the genome of livestock breeds. Selection signatures are defined as genomic regions that harbour outstanding sequence variants; therefore, they are or have been under either artificial or natural selection leaving particular patterns of DNA behind (Qanbari and Simianer, 2014). These patterns derive from "selective sweep" processes when an allele of beneficial mutation (or SNV) chains the alleles of SNPs located in the vicinity of this beneficial mutation. This phenomenon increases LD between these SNPs and consequently an allele of the beneficial mutation rise to high frequency, jointly with these nearby SNPs (see review by Biswas & Akey, 2006; Oleksyk *et al.*, 2010; Qanbari & Simianer, 2014). The procedures for detecting selection signatures are based on several null hypotheses of absence of selection (Biswas and Akey, 2006; Qanbari and Simianer, 2014). Similar than in GWAS, these require high statistical power to detect "selective sweeps" (Schwarzenbacher *et al.*, 2012; Jacobs *et al.*, 2016). Hence, some factors are crucial for choosing the procedures, *e.g.* recombination rate, mutation rate, timing of selection, number of genetic markers and the type of trait (sex limited, fitness and so on) (Oleksyk *et al.*, 2010). The methods of selection signatures can cluster in the following categories:

Differentiation amongst populations

Genome scans of genetic markers between breed groups selected for simply inherited traits may disclose the genomic regions that contributed to the observed phenotypic divergence (Beaumont and Balding, 2004; Johansson *et al.*, 2010; Kessner and Novembre, 2015). These methods detect chiefly the deviation of the loss of heterozygosity amongst populations. However, they are very sensitive to "ascertainment bias". This systematic bias is introduced due to selection criteria for SNPs of genotyping platforms. Since SNPs are chosen according to their MAF, those SNPs will present low MAFs in regions under natural or artificial selection. Hence, these SNPs will be under-represented or no included in the genotyping platforms (see review by Biswas & Akey, 2006; Qanbari & Simianer, 2014).

Wright's fixation index (F_{st}) is the most used metric of genetic differentiation (Wright, 1949). When selection favours a particular causative variant, several markers close to this causative variants show large F_{st} values (Biswas and Akey, 2006). In populations selected for and against a trait, it is more efficient searching for a number of consecutive SNP with large F_{st} rather than analysing each SNPs separately (Qanbari and Simianer, 2014). Nonetheless, we must take into account that F_{st} depends on the allele frequency of the SNP before selection; hence, the genomic information in the base population is important for detecting most causative variants with this method (Pritchard *et al.*, 2010).

Another method, called SelEstim by Vitalis *et al.* (2014), analyses the differences among the allele frequencies of several populations using a Bayesian approach. This method is based on a diffusion model approximation, *i.e.* an island model, for the distribution of allele frequency in a population subdivided into a number of groups of closely related individuals (demes), exchanging migrants. SelEstim estimates the parameters $\boldsymbol{k}_{ij}$ that indicates which allele is selected for, in the ith deme at the locus j; the parameter $\boldsymbol{\delta}_j$ which denotes the average effects of selection at locus j (over all demes); and the hyper-parameter $\boldsymbol{\lambda}$ which represents the genome-wide effect of selection over all demes and loci. The model-based method distinguish the strength and the type of selection acting on segregating polymorphisms. These parameters are estimated from the

posterior distributions using a MCMC algorithm for sampling from the joint posterior distribution of the model parameters (Vitalis *et al.*, 2014).

Reduction of the local genomic variability

These methods are used to study patterns along the genome of only one population, *i.e.* intra–population study. The main idea is to identify genomic regions with a systematically local reduced variation relative to the average across the genome. These methods present an advantage, in which a causative variant and their associated genetic markers can be detected when these are already fixed. Under this condition, these detections are impossible in GWASs (see review by Qanbari & Simianer, 2014). This category includes: deviation of molecular genomic kinship, measure of heterozygosity (H_p) and runs of homozygosity (ROH) (see review by Oleksyk *et al.*, 2010). This latter method is defined as steady homozygous segments of DNA sequence. Besides, its importance lies mainly in quantifying and understanding inbreeding in humans, livestock, and plants (Peripolli *et al.*, 2017; Ceballos *et al.*, 2018). This method is strongly linked to adaptation or fitness traits, *e.g.* litter size and immune response (Kim *et al.*, 2015b; Saura *et al.*, 2015).

Modification of allele frequency spectrums

These methods detect signals of "selective sweeps" which model the allele frequency spectrum across a given genome. The allele frequency distribution of a given set of loci is termed the "site frequency spectrum" (SFC). Under a non-recombination model, SFC can be skewed by natural or artificial selection. In this way, SFC that results from a recent selective fixation may be very different from that produced under the hypothesis of genetic neutrality. Hence, two separate sides of allele frequencies may be observed in a chromosome region, showing an excess of both low- and high-frequencies of each genetic marker. Namely, these methods are based on searching for noticeable shifts in the allele frequency spectrum or haplotype structure in a single population (see reviews by Bamshad & Wooding, 2003; Akey, 2009; Qanbari & Simianer, 2014). Some statistics based on the spectrum of allele frequencies are: Taijama's D (TD), Fay and Wu H test, Fu and Li D test and composite likelihood ratio (CLR) test (see reviews by Biswas & Akey, 2006; Qanbari & Simianer, 2014). The CLR uses coalescent simulations to derive a distribution of the test under the null

hypothesis of absence of selection. Nevertheless, this statistic assumes a uniform mutation rate and recombination rate across the genome, which would be an incorrect hypothesis (Williamson *et al.*, 2007). A drawback of these methods is their high sensitivity to "ascertainment bias", similar to the methods of differentiation amongst populations (Chen *et al.*, 2010; Qanbari and Simianer, 2014). This pitfall can be avoided using sequence data or using SNP arrays greater than 50K SNP (Qanbari and Simianer, 2014).

"The cross-population composite likelihood ratio test" (XP-CLR) is a method which allows comparing an objective population against a reference population. This likelihood method models the multilocus allele frequency differentiation between the two populations (Chen *et al.*, 2010). XP-CLR models the genetic drift under neutrality using Brownian motion which is a non-Markov random process described by stochastic integral equations (Chen *et al.*, 2010; Morozov and Skripkin, 2011). Besides, XP-CLR uses a deterministic model to approximate the effect of a "selective sweep" on genetic markers in the vicinity of the SNP allele chosen as reference point. This method looks for chromosome regions in the genome, where the shift of allele frequency at the locus happened too quickly to be due to random drift (Chen *et al.*, 2010). The XP-CLR is a robust method since can improve unfavourable effects of "ascertainment bias" by modelling the SNP ascertainment schemes (Chen *et al.*, 2010; Qanbari and Simianer, 2014).

Extension of linkage disequilibrium

The methods are based on the decay of LD and the extension of haplotypes in the genome. These methods focus on the study of the correlations amongst neighbouring genetic markers (Oleksyk *et al.*, 2010). They are efficient for detecting ongoing and nearly fixed "selective sweeps". Essentially, they detect hard "selective sweeps" where a new mutation arises on a haplotype that quickly sweeps toward fixation before a recombination process breaks up the haplotype (Sabeti *et al.*, 2007; Szpiech and Hernandez, 2014). Some statistics of this category are: extended haplotype homozygosity (EHH), integrated haplotype score (|iHS|), variation of linkage disequilibrium (varLD) and relative extended haplotype homozygosity (REHH) (see review by Biswas & Akey, 2006; Oleksyk *et al.*, 2010; Qanbari & Simianer, 2014). These methods are the least sensitive

to "ascertainment bias" (Chen *et al.*, 2010; Qanbari and Simianer, 2014). A popular method is "the cross-population extended haplotype homozygosity test" (XP-EHH) (Sabeti *et al.*, 2007). As in XP-CLR, this method compares two populations distinguishing nearby loci with "selective sweeps" in which the selected alleles have risen to high frequency or fixation in the objective population, whilst these alleles tend toward non-extreme frequencies in the reference population. This method assumes that two populations had a common ancestors population and proceeded from a genomic divergence process (Sabeti *et al.*, 2007; Pritchard *et al.*, 2010; Schwarzenbacher *et al.*, 2012).

1.4.3. Post-genomic analyses, validation and application of findings

Increasing the number of SNP (high SNP-density) can lead to many spurious associations in both GWASs and genomic scan studies (Spencer *et al.*, 2009; Sul *et al.*, 2018). After genomic analyses, it is important to bear out the findings by means of (a) implementing several genomic studies (Schwarzenbacher *et al.*, 2012), (b) validation through biological information (Ioannidis *et al.*, 2009), and (c) replication of the findings using several independent populations within a given species (Schmid and Bennewitz, 2017).

A combination of methods of selection signatures may improve the detection of genomic regions showing signals of hard "selective sweep". These methods should be chosen according to the lower sensitivity to "ascertainment bias" (Schwarzenbacher *et al.*, 2012; Qanbari and Simianer, 2014). Besides, genomic regions overlapped between selection signatures and GWAS signals may provide a way to validate promising genetic markers and QTLs (Schwarzenbacher *et al.*, 2012).

After positive findings from GWASs and genomic scan study, the adequate process involves the validation and the replication of these results (Sharma et al., 2015). In this way, the next step is the searching for genes within associated genomic regions using genomic maps (Ioannidis *et al.*, 2009; Schmid and Bennewitz, 2017). Discovering causative variants must have a biological interpretation. The search of genes with causative variants linked to associated SNPs must also point out a consistent and intelligible interpretation of the

pathways, as also cause-effect factors according to knowledge of molecular biology (Schmid and Bennewitz, 2017). That would be the most important way of validation for putative causative variants, inferred from the results of gene expression and references of previous studies in a particular species (McCarthy and Hirschhorn, 2008; Ioannidis *et al.*, 2009). In the case of absence of genomic information in the associated genomic regions, the next step would be to make refining or sequencing of these genomic regions. In particular, researchers should analyse the promoter and exonic regions of the genes within the associated genomic regions. These procedures must initially carry out using the same population in which the associated genomic regions were identified.

GWASs and genomic scan results help to outline further strategies of research and characterize the genomic architecture for a given trait of interest (Ioannidis *et al.*, 2009; Schmid and Bennewitz, 2017). The final objective of these genomic experiments is to detect associated markers with a particular trait, located these markers very close to the underlying genes and to the mutation, within a given gene, affecting the phenotypic trait (Witte, 2010; Sharma *et al.*, 2015). The association and the causation as synonymous are rare from GWASs results. In this way, the causation of a SNV (or causative variant) can be validated using several independent populations as long as a beneficial allele is segregating within those populations. For instance, if a relevant SNP with large effect was identified by genomic analyses using a Spanish rabbit population, researchers must identify the same associated SNP when they repeat the analyses using French, Polish and Hungarian rabbit populations. In this way, they can deem this SNP as a causative variant. Namely, the causative variants must show the same magnitude of effect with the same algebraic sign in all analyzed rabbit populations. In addition, positive results from GWASs and genomic scan studies must be validated by the concordance test and the complete linkage disequilibrium test (Ioannidis *et al.*, 2009; Schmid and Bennewitz, 2017)

Results of GWASs are troublesome for their application in animal breeding programs. Frequently, different regions and different genes are found associated with a given trait in different breeds, due to different genomic architectures and the polygenic nature of complex traits. Nonetheless, there is a few GWASs with outstanding results in livestock production, *e.g.* GWAS for diseases (Meyers *et al.*, 2010; Sharma *et al.*, 2015; Schmid and Bennewitz, 2017). An example was

the GWAS for osteopetrosis in red Angus cattle. The results showed a significant SNP cluster associated with the disease on bovine chromosome 4. This SNP cluster was considered as QTL. The refining and validation of this associated QTL conducted to discover a deletion mutation which causes loss of SLC4A2 function inducing premature cell death and stillbirth. This finding considerably improved the beef production in red Angus cattle, since it was possible to remove the animals showing the unfavourable allele (Meyers *et al.*, 2010).

Anyway, the usefulness of GWASs on animal breeding programs is actually negligible for most quantitative traits, according to mentioned references in this "General Introduction" (view reviews by Sharma *et al.*, 2015; Schmid & Bennewitz, 2017; Georges *et al.*, 2019). Finding all causative variants does not ensure an increase of genetic progress for a given quantitative trait. For instance, we can suppose that a GWAS discovered the 1000 causative variants controlling the variability of that trait. After that, we would only select for the beneficial additive alleles of causative variants. Although this GWAS revealed the effect of causative variants, the study thoroughly ignores the effects of interactions amongst the causative variants. Thus, this selection using the 1000 causative variants will produce a null or the same genetic progress than the traditional selection in the following generation of selection.

1.5. References

Akey, J. M., 2009. "Constructing genomic maps of positive selection in humans: Where do we go from here?" *Genome Research*, 19 (5): 711–722.

Amrhein, V., S. Greenland, and B. McShane, 2019. "Scientists rise up against statistical significance." *Nature*, 567 (7748): 305–307.

Argente, M.-J., 2016. Major Components in Limiting Litter Size, pp. 87–114 in *Insights from Animal Reproduction*, edited by R. Payan Carreira. InTech.

Argente, M. J., M. Merchán, R. Peiró, M. L. García, M. a. Santacreu *et al.*, 2010. "Candidate gene analysis for reproductive traits in two lines of rabbits divergently selected for uterine capacity." *Journal of Animal Science*, 88 (3): 828–836.

Argente, M. J., M. A. Santacreu, A. Climent, G. Bolet, and A. Blasco, 1997. "Divergent selection for uterine capacity in rabbits." *Journal of Animal Science*, 75 (9): 2350–2354.

Badawy, A. Y., R. Peiró, A. Blasco, and M. A. Santacreu, 2019. "Correlated responses on litter size traits and survival traits after two-stage selection for ovulation rate and litter size in rabbits." *Animal*, 13 (3): 453–459.

Ball, R. D., 2013. Designing a GWAS: Power, Sample Size, and Data Structure, pp. 37–98 in *Genome-Wide Association Studies and Genomic Prediction*, edited by C. Gondro, J. van der Werf, and B. Hayes. Humana Press, Totowa, NJ.

Bamshad, M., and S. P. Wooding, 2003. "Signatures of natural selection in the human genome." *Nature Reviews Genetics*, 4 (2): 99–110.

Baselga, M., 2004. Genetic Improvement of Meat Rabbits. Programmes and Diffusion, pp. 1–13 in *8th World Rabbit Congress*, Puebla, Mexico.

Beaulieu, A. D., J. L. Aalhus, N. H. Williams, and J. F. Patience, 2010. "Impact of piglet birth weight, birth order, and litter size on subsequent growth performance, carcass quality, muscle composition, and eating quality of pork." *Journal of Animal Science*, 88 (8): 2767–2778.

Beaumont, M. A., and D. J. Balding, 2004. "Identifying adaptive genetic divergence among populations from genome scans." *Molecular Ecology*, 13 (4): 969–980.

Beissinger, T. M., G. J. Rosa, S. M. Kaeppler, D. Gianola, and N. de Leon, 2015. "Defining window-boundaries for genomic analyses using smoothing spline techniques." *Genetics Selection Evolution*, 47 (1): 30.

Bertolini, F., G. Schiavo, G. Galimberti, S. Bovo, M. D'Andrea *et al.*, 2018. "Genome-wide association studies for seven production traits highlight genomic regions useful to dissect dry-cured ham quality and production traits in Duroc heavy pigs." *Animal*, 12 (9): 1777–1784.

Biswas, S., and J. M. Akey, 2006. "Genomic insights into positive selection." *Trends in Genetics*, 22 (8): 437–446.

Blasco, A., 2017." *Bayesian Data Analysis for Animal Scientists*." Springer International Publishing, Cham.

Blasco, A., M. J. Argente, C. S. Haley, and M. A. Santacreu, 1994. "Relationships between components of litter size in unilaterally ovariectomized and intact rabbit does." *Journal of Animal Science*, 72 (12): 3066–3072.

Blasco, A., J. P. Bidanel, G. Bolet, C. S. Haley, and M. A. Santacreu, 1993. "The genetics of prenatal survival of pigs and rabbits: a review." *Livestock Production Science*, 37 (1–2): 1–21.

Blasco, A., M. Martínez-Álvaro, M. L. García, N. Ibáñez-Escriche, and M. J.

Argente, 2017. "Selection for environmental variance of litter size in rabbits." *Genetics Selection Evolution*, 49 (1): 48.

Blasco, A., I. Nagy, and P. Hernández, 2018. "Genetics of growth, carcass and meat quality in rabbits." *Meat Science*, 145 178–185.

Blasco, A., J. A. Ortega, A. Climent, and M. A. Santacreu, 2005. "Divergent selection for uterine capacity in rabbits. I. Genetic parameters and response to selection." *Journal of Animal Science*, 83 (10): 2297–2302.

Blasco, A., and R. N. Pena, 2018. Current Status of Genomic Maps: Genomic Selection/GBV in Livestock, pp. 61–80 in *Animal Biotechnology 2*, edited by H. Niemann and C. Wrenzycki. Springer International Publishing, Cham.

Bodin, L., M. Garcia, G. Saleil, G. Bolet, and H. Garreau, 2010. Results of 10 generations of canalising selection for rabbit birth weight, in *Proc. 9th World Congress on Genetics Applied to Livestock Production.*, Leipzig, Germany.

Bohren, B. B., 1975. "Designing artificial selection experiments for specific objectives." *Genetics*, 80 (1): 205–220.

Bolet, G., M. A. Santacreu, M. J. Argente, A. Climent, and A. Blasco, 1994. Divergent Selection for Uterine Efficiency in Unilaterally Ovariectomized Rabbits. I. Phenotypic and Genetic Parameters., pp. 261–264 in *5th World Congress on Genetics Applied to Livestock Production.*, Guelph.

Bősze, Z., P. Major, I. Baczkó, K. E. Odening, L. Bodrogi *et al.*, 2016. "The potential impact of new generation transgenic methods on creating rabbit models of cardiac diseases." *Progress in Biophysics and Molecular Biology*, 121 (2): 123–130.

Bünger, L., R. M. Lewis, M. F. Rothschild, A. Blasco, U. Renne *et al.*, 2005. "Relationships between quantitative and reproductive fitness traits in animals." *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360 (1459): 1489–1502.

Caballero, A., A. Tenesa, and P. D. Keightley, 2015. "The nature of genetic variation for complex traits revealed by GWAS and regional heritability mapping analyses." *Genetics*, 201 (4): 1601–1613.

Calasans-Maia, M. D., M. L. Monteiro, F. O. Áscoli, and J. M. Granjeiro, 2009. "The rabbit as an animal model for experimental surgery." *Acta Cirurgica Brasileira*, 24 (4): 325–328.

Camacho, J., and M. Baselga, 1990. Genetic correlation between reproductive and growth traits in rabbits, pp. 366–369 in *Proc.: 4th World Congress on Genetics Applied to the Livestock Production*, Edinburgh, Scotland.

Carneiro, M., S. Afonso, A. Geraldes, H. Garreau, G. Bolet *et al.*, 2011. "The Genetic Structure of Domestic Rabbits." *Molecular Biology and Evolution*, 28 (6): 1801–1816.

Carneiro, M., C. J. Rubin, F. Di Palma, F. W. Albert, J. Alföldi *et al.*, 2014. "Rabbit genome analysis reveals a polygenic basis for phenotypic change during domestication." *Science*, 345 (6200): 1074–1079.

Cartuche, L., M. Pascual, E. A. Gómez, and A. Blasco, 2014. "Economic weights in rabbit meat production." *World Rabbit Science*, 22 (3): 165–177.

Ceballos, F. C., P. K. Joshi, D. W. Clark, M. Ramsay, and J. F. Wilson, 2018. "Runs of homozygosity: windows into population history and trait architecture." *Nature Reviews Genetics*, 19 (4): 220–234.

Chantry-Darmon, C., C. Urien, H. de Rochambeau, D. Allain, B. Pena *et al.*, 2006. "A first-generation microsatellite-based integrated genetic and cytogenetic map for the European rabbit (Oryctolagus cuniculus) and localization of angora and albino." *Animal Genetics*, 37 (4): 335–341.

Chen, H., N. Patterson, and D. Reich, 2010. "Population differentiation as a test for selective sweeps." *Genome Research*, 20 (3): 393–402.

Clark, S. A., and J. Van Der Werf, 2013. Genomic best linear unbiased prediction (gBLUP) for the estimation of genomic breeding values, pp. 321–330 in *Genome-Wide Association Studies and Genomic Prediction*, edited by C. Gondro, J. van der Werf, and B. Hayes. Humana Press, Totowa, NJ.

Craig, D. B., S. Kannan, and A. A. Dombkowski, 2012. "Augmented annotation and orthologue analysis for Oryctolagus cuniculus: Better Bunny." *BMC Bioinformatics*, 13 (1): 84.

Cullere, M., and A. Dalle Zotte, 2018. "Rabbit meat production and consumption: State of knowledge and future perspectives." *Meat Science*, 143 137–146.

Daniewski, W., and T. Jezierski, 2003. "Effectiveness of divergent selection for open-field activity in rabbits and correlated response for body weight and fertility." *Behavior Genetics*, 33 (3): 337–345.

Do, D. N., F. S. Schenkel, F. Miglior, X. Zhao, and E. M. Ibeagha-Awemu, 2018. "Genome wide association study identifies novel potential candidate genes for bovine milk cholesterol content." *Scientific Reports*, 8 (1): 13239.

Drouilhet, L., C. S. Achard, O. Zemb, C. Molette, T. Gidenne *et al.*, 2016. "Direct and correlated responses to selection in two lines of rabbits selected for feed efficiency under ad libitum and restricted feeding: I. Production traits and

gut microbiota characteristics." *Journal of Animal Science*, 94 (1): 38–48.

Duggal, P., E. M. Gillanders, T. N. Holmes, and J. E. Bailey-Wilson, 2008. "Establishing an adjusted p-value threshold to control the family-wide type 1 error in genome wide association studies." *BMC Genomics*, 9 (1): 516.

Eisen, E. J., 1976. "Results of growth curve analyses in mice and rats." *Journal of Animal Science*, 42 (4): 1008–1023.

Erbe, M., B. J. Hayes, L. K. Matukumalli, S. Goswami, P. J. Bowman *et al.*, 2012. "Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels." *Journal of Dairy Science*, 95 (7): 4114–4129.

Estany, J., D. Villalba, M. Tor, D. Cubiló, and J. L. Noguera, 2002. "Correlated response to selection for litter size in pigs: II. Carcass, meat, and fat quality traits1." *Journal of Animal Science*, 80 (10): 2566–2573.

Eu-ahsunthornwattana, J., E. N. Miller, M. Fakiola, S. M. B. Jeronimo, J. M. Blackwell *et al.*, 2014. "Comparison of methods to account for relatedness in genome-wide association studies with family-based data." *PLoS Genetics*, 10 (7): e1004445.

Fernando, R., A. Toosi, A. Wolc, D. Garrick, and J. Dekkers, 2017. "Application of whole-genome prediction methods for genome-wide association studies: a Bayesian approach." *Journal of Agricultural, Biological and Environmental Statistics*, 22 (2): 172–193.

García, M. L., and M. Baselga, 2002. "Estimation of correlated response on growth traits to selection in litter size of rabbits using a cryopreserved control population and genetic trends." *Livestock Production Science*, 78 (2): 91–98.

Garreau, H., S. Brard, J. Hurtaud, E. Guitton, L. Cauquil *et al.*, 2012. Divergent selection for digestive disorders in two commercial rabbit lines: Response of crossbred young rabbits to an experimental inoculation of Escherichia coli O-103, pp. 153–157 in *Proc. 10th World Rabbit Congr.*, Sharm El-Sheikh, Egypt.

Garreau, H., M. Piles, C. Larzul, M. Baselga, H. de Rochambeau *et al.*, 2004. Selection of maternal lines: last results and prospects, pp. 14–25 in *Proc. 8th World Rabbit Congress*, Puebla, Mexico.

Garrick, D. J., and R. L. Fernando, 2013. Implementing a QTL Detection Study (GWAS) Using Genomic Prediction Methodology, pp. 275–298 in *Genome-Wide Association Studies and Genomic Prediction*, edited by C. Gondro, J.

van der Werf, and B. Hayes. Humana Press, Totowa, NJ.

Georges, M., C. Charlier, and B. Hayes, 2019. "Harnessing genomic information for livestock improvement." *Nature Reviews Genetics*, 20 (3): 135–156.

Gianola, D., 2013. "Priors in Whole-Genome Regression: The Bayesian Alphabet Returns." *Genetics*, 194 (3): 573–596.

Gianola, D., M. I. Fariello, H. Naya, and C.-C. Schön, 2016. "Genome-Wide Association Studies with a Genomic Relationship Matrix: A Case Study with Wheat and Arabidopsis." *G3: Genes,Genomes,Genetics*, 6 (10): 3241–3256.

Gilbert, H., J. P. Bidanel, J. Gruand, J. C. Caritez, Y. Billon *et al.*, 2007. "Genetic parameters for residual feed intake in growing pigs, with emphasis on genetic relationships with carcass and meat quality traits." *Journal of Animal Science*, 85 (12): 3182–3188.

Goddard, M. E., K. E. Kemper, I. M. MacLeod, A. J. Chamberlain, and B. J. Hayes, 2016. "Genetics of complex traits: prediction of phenotype, identification of causal polymorphisms and genetic architecture." *Proceedings of the Royal Society B: Biological Sciences*, 283 (1835): 20160569.

Goudet, J., T. Kay, and B. S. Weir, 2018. "How to estimate kinship." *Molecular Ecology*, 27 (20): 4121–4135.

Grams, V., R. Wellmann, S. Preuß, M. A. Grashorn, J. B. Kjaer *et al.*, 2015. "Genetic parameters and signatures of selection in two divergent laying hen lines selected for feather pecking behaviour." *Genetics Selection Evolution*, 47 (1): 77.

Gurgul, A., E. Semik, K. Pawlina, T. Szmatoła, I. Jasielczuk *et al.*, 2014. "The application of genome-wide SNP genotyping methods in studies on livestock genomes." *Journal of Applied Genetics*, 55 (2): 197–208.

Habier, D., R. L. Fernando, K. Kizilkaya, and D. J. Garrick, 2011. "Extension of the bayesian alphabet for genomic selection." *BMC Bioinformatics*, 12 (1): 186.

Hayes, B., 2013. Overview of statistical methods for genome-wide association studies (GWAS), pp. 149–169 in *Genome-Wide Association Studies and Genomic Prediction*, edited by C. Gondro, J. H. van der Werf, and B. Hayes. Humana Press, Totowa, NJ.

Hearne, C., S. Ghosh, and J. Todd, 1992. "Microsatellites for linkage analysis of genetic traits." *Trends in Genetics*, 8 (8): 288–294.

Hernández, P., 2008. Enhancement of nutritional quality and safety in rabbit

meat, pp. 10–13 in *Proc.: 9th World Rabbit Congress*, Verona, Italy.

Hernández, P., S. Aliaga, M. Pla, and A. Blasco, 2004. "The effect of selection for growth rate and slaughter age on carcass composition and meat quality traits in rabbits." *Journal of Animal Science*, 82 (11): 3138–3143.

Hill, W. G., 1972. "Estimation of realised heritabilities from selection experiments. I. Divergent selection." *Biometrics*, 28 (3): 747.

Hill, W. G., and A. Caballero, 1992. "Artificial selection experiments." *Annual Review of Ecology and Systematics*, 23 (1): 287–310.

Hoban, S., J. L. Kelley, K. E. Lotterhos, M. F. Antolin, G. Bradburd *et al.*, 2016. "Finding the genomic basis of local adaptation: Pitfalls, practical solutions, and future directions." *The American Naturalist*, 188 (4): 379–397.

Hocquette, J. F., F. Gondret, E. Baéza, F. Médale, C. Jurie *et al.*, 2010. "Intramuscular fat content in meat-producing animals: development, genetic and nutritional control, and identification of putative markers." *Animal*, 4 (2): 303–319.

Holm, B., M. Bakken, G. Klemetsdal, and O. Vangen, 2004. "Genetic correlations between reproduction and production traits in swine." *Journal of Animal Science*, 82 (12): 3458–3464.

Hoque, M. A., K. Suzuki, H. Kadowaki, T. Shibata, and T. Oikawa, 2007. "Genetic parameters for feed efficiency traits and their relationships with growth and carcass traits in Duroc pigs." *Journal of Animal Breeding and Genetics*, 124 (3): 108–116.

Ibañez-Escriche, N., and O. Gonzalez-Recio, 2011. "Review. Promises, pitfalls and challenges of genomic selection in breeding programs." *Spanish Journal of Agricultural Research*, 9 (2): 404–413.

Ioannidis, J. P. A., G. Thomas, and M. J. Daly, 2009. "Validating, augmenting and refining genome-wide association signals." *Nature Reviews Genetics*, 10 (5): 318–329.

Irving-Pease, E. K., L. A. F. Frantz, N. Sykes, C. Callou, and G. Larson, 2018. "Rabbits and the specious origins of domestication." *Trends in Ecology & Evolution*, 33 (3): 149–152.

Jacobs, G. S., T. J. Sluckin, and T. Kivisild, 2016. "Refining the use of linkage disequilibrium as a robust signature of selective sweeps." *Genetics*, 203 (4): 1807–1825.

Johansson, A. M., M. E. Pettersson, P. B. Siegel, and Ö. Carlborg, 2010. "Genome-wde effects of long-term divergent selection." *PLoS Genetics*, 6

(11): e1001188.

Karimi, K., M. Sargolzaei, G. S. Plastow, Z. Wang, and Y. Miar, 2018. "Genetic and phenotypic parameters for litter size, survival rate, gestation length, and litter weight traits in American mink." *Journal of Animal Science*, 96 (7): 2596–2606.

Kass, R. E., and A. E. Raftery, 1995. "Bayes Factors." *Journal of the American Statistical Association*, 90 (430): 773–795.

Kessner, D., and J. Novembre, 2015. "Power analysis of artificial selection experiments using efficient whole genome simulation of quantitative traits." *Genetics*, 199 (4): 991–1005.

Khalil, M. H., and A. M. Al-Saef, 2008. Methods, criteria, techniques and genetic responses for rabbit selection: A review, pp. 1–22 in *Proc. 9th World Rabbit Congress*, Verona, Italy.

Kim, E.-S., R. Ros-Freixedes, R. N. Pena, T. J. Baas, J. Estany *et al.*, 2015a. "Identification of signatures of selection for intramuscular fat and backfat thickness in two Duroc populations." *Journal of Animal Science*, 93 (7): 3292–3302.

Kim, E. S., T. S. Sonstegard, C. P. Van Tassell, G. Wiggans, and M. F. Rothschild, 2015b. "The relationship between runs of homozygosity and inbreeding in Jersey cattle under selection." *PLoS ONE*, 10 (7): 1–17.

Kouba, M., and P. Sellier, 2011. "A review of the factors influencing the development of intermuscular adipose tissue in the growing pig." *Meat Science*, 88 (2): 213–220.

Larzul, C., V. Ducrocq, F. Tudela, H. Juin, and H. Garreau, 2014. "The length of productive life can be modified through selection: An experimental demonstration in the rabbit." *Journal of Animal Science*, 92 (6): 2395–2401.

Larzul, C., F. Gondret, S. Combes, and H. de Rochambeau, 2005. "Divergent selection on 63-day body weight in the rabbit: response on growth, carcass and muscle traits." *Genetics Selection Evolution*, 37 (1): 105.

Legarra, A., 2016. "Comparing estimates of genetic variance across different relationship models." *Theoretical Population Biology*, 107 26–30.

Lehermeier, C., V. Wimmer, T. Albrecht, H. J. Auinger, D. Gianola *et al.*, 2013. "Sensitivity to prior specification in Bayesian genome-based prediction models." *Statistical Applications in Genetics and Molecular Biology*, 12 (3): 375–391.

Li, M. X., J. M. Y. Yeung, S. S. Cherny, and P. C. Sham, 2012. "Evaluating the

effective numbers of independent tests and significant p-value thresholds in commercial genotyping arrays and public imputation reference datasets." *Human Genetics*, 131 (5): 747–756.

Li, G., and H. Zhu, 2013. "Genetic studies: The linear mixed models in genome-wide association studies." *The Open Bioinformatics Journal*, 7 (1): 27–33.

López de Maturana, E., N. Ibáñez-Escriche, Ó. González-Recio, G. Marenne, H. Mehrban *et al.*, 2014. "Next generation modeling in GWAS: comparing different genetic architectures." *Human Genetics*, 133 (10): 1235–1253.

de Los Campos, G., J. M. Hickey, R. Pong-Wong, H. D. Daetwyler, and M. P. L. Calus, 2013. "Whole-genome regression and prediction methods applied to plant and animal breeding." *Genetics*, 193 (2): 327–345.

Malik, R. C., 1984. "Genetic and physiological aspects of growth, body composition and feed efficiency in mice: A review." *Journal of Animal Science*, 58 (3): 577–590.

Mapara, M., B. S. Thomas, and K. M. Bhat, 2012. "Rabbit as an animal model for experimental research." *Dental Research Journal*, 9 (1): 111–118.

Marees, A. T., H. de Kluiver, S. Stringer, F. Vorspan, E. Curis *et al.*, 2018. "A tutorial on conducting genome-wide association studies: Quality control and statistical analysis." *International Journal of Methods in Psychiatric Research*, 27 (2): 1–10.

Martínez-Álvaro, M., M. F. Betancur-Zambrano, A. Blasco, and N. Ibáñez-Escriche, 2019. "Correlated responses to selection for intramuscular fat on litter size." *Personal Communication*,.

Martínez-Álvaro, M., A. Blasco, and P. Hernández, 2018. "Effect of selection for intramuscular fat on the fatty acid composition of rabbit meat." *Animal*, 12 (10): 2002–2008.

Martínez-Álvaro, M., P. Hernández, and A. Blasco, 2016a. "Divergent selection on intramuscular fat in rabbits: Responses to selection and genetic parameters." *Journal of Animal Science*, 94 (12): 4993.

Martínez-Álvaro, M., V. Penalba, A. Blasco, and P. Hernández, 2016b. "Effect of divergent selection for intramuscular fat on sensory traits and instrumental texture in rabbit meat." *Journal of Animal Science*, 94 (12): 5137–5143.

McCarthy, M. I., and J. N. Hirschhorn, 2008. "Genome-wide association studies: potential next steps on a genetic journey." *Human Molecular Genetics*, 17 (R2): R156–R165.

McNitt, J. I., S. D. Lukefahr, P. R. Cheeke, and N. M. Patton, 2013." *Rabbit*

Production." CABI Wallingford and Boston, Boston.

McShane, B. B., D. Gal, A. Gelman, C. Robert, and J. L. Tackett, 2019. "Abandon Statistical Significance." *The American Statistician*, 73 (sup1): 235–245.

Mehrban, H., D. H. Lee, M. H. Moradi, C. IlCho, M. Naserkheil *et al.*, 2017. "Predictive performance of genomic selection methods for carcass traits in Hanwoo beef cattle: impacts of the genetic architecture." *Genetics Selection Evolution*, 49 (1): 1.

Mei, B., and Z. Wang, 2016. "An efficient method to handle the 'large p, small n' problem for genomewide association studies using Haseman–Elston regression." *Journal of Genetics*, 95 (4): 847–852.

Merchán, M., R. Peiró, M. J. Argente, M. A. Santacreu, M. L. García *et al.*, 2009. "Analysis of the oviductal glycoprotein 1 polymorphisms and their effects on components of litter size in rabbits." *Animal Genetics*, 40 (5): 756–758.

Meuwissen, T. H. E., B. J. Hayes, and M. E. Goddard, 2001. "Prediction of total genetic value using genome-wide dense marker maps." *Genetics*, 157 (4): 1819–1829.

Meyers, S. N., T. G. McDaneld, S. L. Swist, B. M. Marron, D. J. Steffen *et al.*, 2010. "A deletion mutation in bovine SLC4A2 is associated with osteopetrosis in Red Angus cattle." *BMC Genomics*, 11 (1): 337.

Mgheni, M., and K. Christensen, 1985. "Selection experiment on growth and litter size in rabbits." *Acta Agriculturae Scandinavica*, 35 (3): 278–286.

Miller, I., C. Rogel-Gaillard, D. Spina, L. Fontanesi, and A. de Almeida, 2014. "The rabbit as an experimental and production animal: From genomics to proteomics." *Current Protein & Peptide Science*, 15 (2): 134–145.

Morozov, A. N., and A. V. Skripkin, 2011. "Spherical particle Brownian motion in viscous medium as non-Markovian random process." *Physics Letters A*, 375 (46): 4113–4115.

Moura, A. S., M. Kaps, D. W. Vogt, and W. R. Lamberson, 1997. "Two-way selection for daily gain and feed conversion in a composite rabbit population." *Journal of Animal Science*, 75 (9): 2344.

Nagy, I., N. Ibáñez, R. Romvári, W. Mekkawy, S. Metzger *et al.*, 2006. "Genetic parameters of growth and in vivo computerized tomography based carcass traits in Pannon White rabbits." *Livestock Science*, 104 (1–2): 46–52.

Oleksyk, T. K., M. W. Smith, and S. J. O'Brien, 2010. "Genome-wide scans for footprints of natural selection." *Philosophical Transactions of the Royal*

Society B: Biological Sciences, 365 (1537): 185–205.

Pannier, L., G. E. Gardner, R. A. O'Reilly, and D. W. Pethick, 2018. "Factors affecting lamb eating quality and the potential for their integration into an MSA sheepmeat grading model." *Meat Science*, 144 43–52.

Park, T., and G. Casella, 2008. "The Bayesian Lasso." *Journal of the American Statistical Association*, 103 (482): 681–686.

Patnala, R., J. Clements, and J. Batra, 2013. "Candidate gene association studies: a comprehensive guide to useful in silico tools." *BMC Genetics*, 14 (1): 39.

Peiró, R., A. Y. Badawy, A. Blasco, and M. A. Santacreu, 2019. "Correlated responses on growth traits after two-stage selection for ovulation rate and litter size in rabbits." *Animal*, 13 (3): 1–6.

Peiró, R., A. Herrler, M. A. Santacreu, M. Merchán, M. J. Argente *et al.*, 2010. "Expression of progesterone receptor related to the polymorphism in the PGR gene in the rabbit reproductive tract." *Journal of Animal Science*, 88 (2): 421–427.

Pena, R., R. Ros-Freixedes, M. Tor, and J. Estany, 2016. "Genetic marker discovery in complex traits: A field example on fat content and composition in pigs." *International Journal of Molecular Sciences*, 17 (12): 2100.

Pérez, P., and G. De Los Campos, 2014. "Genome-wide regression and prediction with the BGLR statistical package." *Genetics*, 198 (2): 483–495.

Peripolli, E., D. P. Munari, M. V. G. B. Silva, A. L. F. Lima, R. Irgang *et al.*, 2017. "Runs of homozygosity: current knowledge and applications in livestock." *Animal Genetics*, 48 (3): 255–271.

Petry, D. B., J. W. Holl, and R. K. Johnson, 2004. "Responses to 19 generations of litter size selection in the NE Index line. II. Growth and carcass responses estimated in pure line and crossbred litters." *Journal of Animal Science*, 82 (7): 1895–1902.

Piles, M., E. A. Gómez, O. Rafel, J. Ramon, and A. Blasco, 2004. "Elliptical selection experiment for the estimation of genetic parameters of the growth rate and feed conversion ratio in rabbits." *Journal of Animal Science*, 82 (3): 654–660.

Piles M., and Blasco A., 2003. "Response to selection for growth rate in rabbits estimated by using a control cryopreserved population." *World Rabbit Science*, 11 (2): 53–62.

Prayaga, K. C., and S. J. Eady, 2000. "Rabbit farming for meat production in

Australia: Preliminary estimates of economic values for production traits." *Asian-Australasian Journal of Animal Sciences*, 13 357–359.

Price, A., N. Zaitlen, D. Reich, and N. Patterson, 2010. "New approaches to population stratification in genome-wide association studies." *Nature Reviews Genetics*, 11 (7): 459–463.

Pritchard, J. K., J. K. Pickrell, and G. Coop, 2010. "The genetics of human adaptation: hard Sweeps, soft Sweeps, and polygenic adaptation." *Current Biology*, 20 (4): R208–R215.

Purfield, D. C., D. G. Bradley, J. F. Kearney, and D. P. Berry, 2014. "Genome-wide association study for calving traits in Holstein-Friesian dairy cattle." *Animal*, 8 (2): 224–235.

Qanbari, S., and H. Simianer, 2014. "Mapping signatures of positive selection in the genome of livestock." *Livestock Science*, 166 133–143.

Rafat, S. A., D. Allain, and H. de Rochambeau, 2009. "Genetic description of a divergent selection experiment in Angora rabbits with overlapping generations." *Journal of Animal Breeding and Genetics*, 126 (3): 189–197.

Rafat, S. A., D. Allain, R. G. Thébault, and H. de Rochambeau, 2007. "Divergent selection for fleece weight in French Angora rabbits: Non-genetic effects, genetic parameters and response to selection." *Livestock Science*, 106 (2–3): 169–175.

Ritland, K., 1996. "Estimators for pairwise relatedness and individual inbreeding coefficients." *Genetical Research*, 67 (2): 175–185.

Ros-Freixedes, R., S. Gol, R. N. Pena, M. Tor, N. Ibáñez-Escriche *et al.*, 2016. "Genome-wide association study singles out SCD and LEPR as the two main loci influencing intramuscular fat content and fatty acid composition in Duroc pigs (R. Davoli, Ed.)." *PLOS ONE*, 11 (3): e0152496.

Rothschild, M. F., Z. Hu, and Z. Jiang, 2007. "Advances in QTL mapping in pigs." *International Journal of Biological Sciences*, 3 (3): 192–197.

Rothschild, M. F., and A. Ruvinsky, 2011." *The genetics of the pig.*" CABI, Wallinford, United Kingdom.

Sabeti, P. C., P. Varilly, B. Fry, J. Lohmueller, E. Hostetter *et al.*, 2007. "Genome-wide detection and characterization of positive selection in human populations." *Nature*, 449 (7164): 913–918.

Sahana, G., B. Guldbrandtsen, and M. S. Lund, 2011. "Genome-wide association study for calving traits in Danish and Swedish Holstein cattle." *Journal of Dairy Science*, 94 (1): 479–486.

Santacreu, M. A., M. J. Argente, A. Climent, A. Blasco, and G. Bolet, 1994. Divergent selection for uterine efficiency in unilaterally ovariectomized rabbits. II. Response to selection, pp. 265–267 in *Proc. 5th World Congress on Genetics Applied to Livestock Productio*, Ontario, Canada.

Saura, M., A. Fernández, L. Varona, A. I. Fernández, M. de Cara *et al.*, 2015. "Detecting inbreeding depression for reproductive traits in Iberian pigs using genome-wide data." *Genetics Selection Evolution*, 47 (1): 1.

Schmid, M., and J. Bennewitz, 2017. "Invited review: Genome-wide association analysis for quantitative traits in livestock – a selective review of statistical models and experimental designs." *Archives Animal Breeding*, 60 (3): 335–346.

Schwarzenbacher, H., M. Dolezal, K. Flisikowski, F. Seefried, C. Wurmser *et al.*, 2012. "Combining evidence of selection with association analysis increases power to detect regions influencing complex traits in dairy cattle." *BMC Genomics*, 13 (1): 48.

Sharma, A., J. S. Lee, C. G. Dang, P. Sudrajad, H. C. Kim *et al.*, 2015. "Stories and challenges of genome wide association studies in livestock — A Review." *Asian-Australasian Journal of Animal Sciences*, 28 (10): 1371–1379.

Shen, X., M. Alam, F. Fikse, and L. Rönnegård, 2013. "A novel generalized ridge regression method for quantitative genetics." *Genetics*, 193 (4): 1255–1268.

Shirali, M., R. Pong-Wong, P. Navarro, S. Knott, C. Hayward *et al.*, 2016. "Regional heritability mapping method helps explain missing heritability of blood lipid traits in isolated populations." *Heredity*, 116 (3): 333–338.

Singh, U., R. Deb, R. R. Alyethodi, R. Alex, S. Kumar *et al.*, 2014. "Molecular markers and their applications in cattle genetic research: A review." *Biomarkers and Genomic Medicine*, 6 (2): 49–58.

Skovsted, K., M. Henryon, L. Rydhmer, J. Jensen, and F. X. Solanes, 2005. "Growth rate of growing pigs is weakly correlated genetically with litter size, while the amount of genetic variation for growth rate changes with litter size." *Acta Agriculturae Scandinavica, Section A - Animal Science*, 55 (2–3): 66–73.

Sorensen, D., and D. Gianola, 2002." *Likelihood, Bayesian, and MCMC Methods in Quantitative Genetics*." Springer New York, New York, NY.

Spencer, C. C. A., Z. Su, P. Donnelly, and J. Marchini, 2009. "Designing Genome-Wide Association Studies: Sample Size, Power, Imputation, and the Choice of Genotyping Chip." *PLoS Genetics*, 5 (5): e1000477.

Stephens, M., and D. J. Balding, 2009. "Bayesian statistical methods for genetic association studies." *Nature Reviews Genetics*, 10 (10): 681–690.

Sternstein, I., M. Reissmann, D. Maj, J. Bieniek, and G. A. Brockmann, 2015. "A comprehensive linkage map and QTL map for carcass traits in a cross between Giant Grey and New Zealand White rabbits." *BMC Genetics*, 16 (1): 16.

Sternstein, I., M. Reissmann, D. Maj, J. Bieniek, and G. A. Brockmann, 2014. "A new single nucleotide polymorphism in the rabbit (Oryctolagus cuniculus) myostatin (MSTN) gene is associated with carcass composition traits." *Animal Genetics*, 45 (4): 596–599.

Stram, D. O., 2014." *Design, Analysis, and Interpretation of Genome-Wide Association Scans*." Springer New York, New York, NY.

Su G., J. B. Kjaer, U. T. Brenøe, and P. Sørensen, 2010. "Estimates of genetic parameters in Danish white rabbits using an animal model: I. Growth and carcass traits." *World Rabbit Science*, 7 (2): 59–64.

Sul, J. H., L. S. Martin, and E. Eskin, 2018. "Population structure in genetic studies: Confounding factors and mixed models." *PLOS Genetics*, 14 (12): e1007309.

Szendro, Z. S., S. Z. Metzger, I. Nagy, A. Szabó, Z. S. Petrási *et al.*, 2012. "Effect of divergent selection for the computer tomography measured thigh muscle volume on productive and carcass traits of growing rabbits." *Livestock Science*, 149 (1–2): 167–172.

Szpiech, Z. A., and R. D. Hernandez, 2014. "selscan: An efficient multithreaded program to perform EHH-based scans for positive selection." *Molecular Biology and Evolution*, 31 (10): 2824–2827.

Tabor, H. K., N. J. Risch, and R. M. Myers, 2002. "Candidate-gene approaches for studying complex genetic traits: practical considerations." *Nature Reviews Genetics*, 3 (5): 391–397.

Tang, R., T. Feng, Q. Sha, and S. Zhang, 2009. "A variable-sized sliding-window approach for genetic association studies via principal component analysis." *Annals of Human Genetics*, 73 (6): 631–637.

Teneva, A., and M. P. Petrovic, 2010. "Application of molecular markers in livestock improvement." *Biotechnology in Animal Husbandry*, 26 (3–4): 135–154.

Toosi, A., R. L. Fernando, and J. C. M. Dekkers, 2018. "Genome-wide mapping of quantitative trait loci in admixed populations using mixed linear model

and Bayesian multiple regression analysis." *Genetics Selection Evolution*, 50 (1): 32.

VanRaden, P. M., 2008. "Efficient methods to compute genomic predictions." *Journal of Dairy Science*, 91 (11): 4414–4423.

Vignal, A., D. Milan, M. SanCristobal, and A. Eggen, 2002. "A review on SNP and other types of molecular markers and their use in animal genetics." *Genetics Selection Evolution*, 34 (3): 275–305.

Visscher, P. M., N. R. Wray, Q. Zhang, P. Sklar, M. I. McCarthy *et al.*, 2017. "10 years of GWAS discovery: Biology, function, and translation." *American Journal of Human Genetics*, 101 (1): 5–22.

Vitalis, R., M. Gautier, K. J. Dawson, and M. A. Beaumont, 2014. "Detecting and measuring selection from gene frequency data." *Genetics*, 196 (3): 799–817.

Voight, B. F., and J. K. Pritchard, 2005. "Confounding from cryptic relatedness in case-control association studies." *PLOS Genetics*, 1 (3): e32.

Wakefield, J., 2012. "Commentary: Genome-wide significance thresholds via Bayes factors." *International Journal of Epidemiology*, 41 (1): 286–291.

Walsh, B., and M. Lynch, 2018." *Evolution and selection of quantitative traits*." Oxford University Press.

Wang, J., 2016. "Pedigrees or markers: Which are better in estimating relatedness and inbreeding coefficient?" *Theoretical Population Biology*, 107 4–13.

Wasserstein, R. L., A. L. Schirm, and N. A. Lazar, 2019. "Moving to a world beyond " p < 0.05."" *The American Statistician*, 73 (sup1): 1–19.

Williamson, S. H., M. J. Hubisz, A. G. Clark, B. A. Payseur, C. D. Bustamante *et al.*, 2007. "Localizing recent adaptive evolution in the human genome." *PLoS Genetics*, 3 (6): e90.

Wilson-Wells, D. F., and S. D. Kachman, 2016. "A bayesian GWAS method utilizing haplotype clusters for a composite breed population." *Conference on Applied Statistics in Agriculture*, 0–13.

Witte, J. S., 2010. "Genome-Wide Association Studies and Beyond." *Annual Review of Public Health*, 31 9–20.

Wright, S., 1949. "The genetical structure of populations." *Annals of Eugenics*, 15 (1): 323–354.

Yang, J., B. Benyamin, B. P. McEvoy, S. Gordon, A. K. Henders *et al.*, 2010. "Common SNPs explain a large proportion of the heritability for human

height." *Nature Genetics*, 42 (7): 565–569.

Yang, J., S. H. Lee, M. E. Goddard, and P. M. Visscher, 2011. "GCTA: A tool for genome-wide complex trait analysis." *American Journal of Human Genetics*, 88 (1): 76–82.

Yang, J., N. A. Zaitlen, M. E. Goddard, P. M. Visscher, and A. L. Price, 2014. "Advantages and pitfalls in the application of mixed-model association methods." *Nature Genetics*, 46 (2): 100–106.

Zhao, Y., H. Zhu, Z. Lu, R. C. Knickmeyer, and F. Zou, 2019. "Structured genome-wide association studies with Bayesian hierarchical variable selection." *Genetics*, 212 (2): 397–415.

Zhu, M., and S. Zhao, 2007. "Candidate gene identification approach: Progress and challenges." *International Journal of Biological Sciences*, 3 (7): 420–427.

CHAPTER TWO

2. SCOPE OF THE STUDY

At the Universitat Politècnica de València, the Animal Breeding Unit carried out two independent experiments of divergent selection for uterine capacity and for intramuscular fat, respectively. The divergently selected populations are an exceptional material to identify causative variants because they increase the detection power compared with a one-way selected population. The outline of the current research study was to disclose potential genomic regions and putative causative variants controlling the traits which were selection criteria. The findings of this study could suggest new research studies focused likely on rabbit breeding programs.

2.1. Specific objectives of this thesis

- To identify genetic markers (SNPs) and genomic regions associated with uterine capacity and its correlated traits, using Bayesian GWAS and the rabbit lines of the divergent selection experiment for uterine capacity.

- To identify genetic markers (SNPs) and genomic regions associated with intramuscular fat, using distinct GWAS approaches and the rabbit lines of the divergent selection experiment for intramuscular fat.

- To detect genomic regions associated with intramuscular fat through genome scan studies, as the methods of selection signatures, taking samples from divergently selected populations.

CHAPTER THREE

3. A GENOME-WIDE ASSOCIATION STUDY IN DIVERGENTLY SELECTED LINES IN RABBITS REVEALS NOVEL GENOMIC REGIONS ASSOCIATED WITH LITTER SIZE TRAITS

Authors: Bolívar Samuel Sosa-Madrid[1], María Antonia Santacreu[1], Agustín Blasco[1], Luca Fontanesi[2], Romi Natacha Pena[3], and Noelia Ibáñez-Escriche[1].

Institutional affiliations

[1] Institute for Animal Science and Technology, Universitat Politècnica de València, 46022 Valencia, Spain.

[2] Department of Agricultural and Food Sciences, Division of Animal Sciences, University of Bologna, 40127 Bologna, Italy.

[3] Departament de Ciència Animal, Universitat de Lleida–Agrotecnio Center, E-25198 Lleida, Catalonia, Spain.

The content of this chapter has been published in *Journal of Animal Breeding and Genetics.*

DOI: https://doi.org/10.1111/jbg.12451

3.1. Abstract

Uterine capacity (UC), defined as the total number of kits from unilaterally ovariectomized does at birth, has a high genetic correlation with litter size. The aim of our research was to identify genomic regions associated with litter size traits through a genome-wide association study using rabbits from a divergent selection experiment for UC. A high-density SNP array (200K) was used to genotype 181 does from a control population, high and low UC lines. Traits included total number born (TNB), number born alive (NBA), number born dead, ovulation rate (OR), implanted embryos (IE), and embryo, foetal and prenatal survivals at second parity. We implemented the Bayes B method and the associations were tested by Bayes factors and the percentage of genomic variance (GV) explained by windows. Different genomic regions associated with TNB, NBA, IE, and OR were found. These regions explained 7.36%, 1.27%, 15.87%, and 3.95% of GV, respectively. Two consecutive windows on chromosome 17 were associated with TNB, NBA, and IE. This genomic region accounted for 6.32% of GV of TNB. In this region, we found the *BMP4, PTDGR, PTGER2, STYX* and *CDKN3* candidate genes which presented functional annotations linked to some reproductive processes. Our findings suggest that a genomic region on chromosome 17 has an important effect on litter size traits. However, further analyses are needed to validate this region in other maternal rabbit lines.

Keyword: *divergent selection, genome-wide association study, litter size, quantitative trait loci, rabbits, uterine capacity.*

3.2. Background

Litter size has high economic importance in all polytocous livestock species, including rabbits (Cartuche et al., 2014) and swine (Quinton et al., 2006). However, the selection response for this complex trait, as well for several other reproduction traits, is small. For example, in rabbit selection experiments for litter size the response can be 0.1 kits per generation (see review Khalil & Al-Saef, 2008). This situation encouraged the application of alternative selection strategies based on litter size components such as uterine capacity (UC) (Argente et al., 1997), ovulation rate (OR) (Laborda et al., 2012), or selection

using independent culling levels for OR and litter size (Ziadi et al., 2013; Badawy et al., 2018).

UC is the prenatal survival when the OR is not a limiting factor of litter size and the uterine horn is crowded with embryos (Blasco et al., 1994; Argente et al., 1997). This trait can be measured as total number of kits at birth under these conditions (Christenson et al., 1987; Mocé et al., 2004), since does have a double cervix preventing intrauterine transmigration; and thus, only one uterine horn remains functional and crowed, duplicating its OR when ovariectomies are implemented (Blasco et al., 1994; Argente et al., 1997). From 1991 to 1998, the Animal Science Department of "Universitat Politècnica de València" carried out an experiment of divergent selection for UC. After ten generations of selection, the divergence between the two divergent lines (high and low UC lines) was 1.50 kits for UC (Blasco et al., 2005), with a correlated response in litter size of 2.35 kits (Santacreu et al., 2005). Approximately one-half of the response in UC was obtained in the first two generations suggesting the presence of a major locus with large effect segregating in these populations (Argente et al., 2003; Blasco et al., 2005). Thus, a candidate gene strategy was carried out to characterize this locus by comparing polymorphisms and expression levels between the two UC lines of some promising candidates (Peiró et al., 2008; Argente et al., 2010; Ballester et al., 2013). Some of these genes (progesterone receptor - *PGR*, hydroxysteroid (17-beta) dehydrogenase 4 - *HSD17B4*, and Endoplasmic Reticulum Oxidoreductase 1 - *ERO1*) showed different expression levels in the oviduct of the two UC line, remarkably overexpressed in the low UC line, but these result could not identify any putative causal mutations (Peiró et al., 2008; Argente et al., 2010; Ballester et al., 2013).

The recent availability of an updated rabbit reference genome (Carneiro et al., 2014) and a high-density single nucleotide polymorphisms (SNP) array (Blasco & Pena, 2018) has opened new possibilities for more comprehensive genomic analyses in this species, similar to what is possible in all other major livestock species. Together with these tools, several methods for genome-wide association analyses have been also already developed and applied in many different species (Fan et al., 2010). Among them, genome-wide association studies (GWAS) using multi-marker regression approaches can attain better power detection to identify genomic regions associated with a trait than the classical approach of

single maker simple regression (López de Maturana et al., 2014; Toosi et al., 2018).

In this study, we designed a GWAS in rabbit based on the described extreme and divergent lines for UC and applied a Bayesian multi-marker regression approach to identify quantitative trait loci (QTL) affecting litter size traits in this species.

3.3. Material and Methods

3.3.1. Ethical statement

Animal manipulations and the experimental procedures were approved by the Ethical Committee of the Universitat Politècnica de València, according to Council Directives 98/58/EC (European Economic Community, 1998).

3.3.2. Animals and phenotypes

Animals came from an experiment of divergent selection for uterine capacity and a cryopreserved control population (Santacreu et al., 2005; Blasco et al., 2005). After ten generations of selection for uterine capacity, the selection was relaxed. For the current study, we collected blood samples from non-ULO female rabbits. The study involves 90 does of the high UC line, 69 does of the low UC line and 30 does of the control population. All samples of high and low UC lines came from the 11th and 12th generations (Mocé et al., 2005; Santacreu et al., 2005). The base population of divergent lines for UC came from the 12th generation of a line selected for number of kits at weaning (named V line). The control population was derived from cryopreserved embryos from the 13th and 15th generations of the V line. The embryos were transferred to receptor does to produce a control population which was contemporary to UC females from 11th generation (Santacreu et al., 2005).

The traits were recorded at the second parity: NBA, as the number of alive kits at parity; NBD, as the number of dead kits; TNB, as the sum of NBA and NBD; OR, calculated as the number of corpora lutea; IE, calculated as the number of implantation sites by laparoscopy at day 12 of the gestation; embryo survival (ES), computed as a ratio IE/OR; foetal survival (FS), as a ratio TNB/IE; and

prenatal survival (PS), as a ratio TNB/OR (Mocé et al., 2005; Santacreu et al., 2005).

3.3.3. Genotypes and quality control

Genomic DNA was isolated from blood using Favorgen Kit (FABGK 001-2; Favorgen Biotech Corp., Taiwan). We collected 189 samples with a minimum concentration of 20 ŋg/µl and minimum volume of 45 µl. The concentrations were estimated with Nanodrop ND-1000 (NanoDrop Technologies, Wilmington, DE, USA) and borne out with PicoGreen (Invitrogen Corp. Carlsbad, C.A.). The threshold values for the integrity of DNA were 1.8 OD260 /OD280 and 1.5 OD260 /OD320. The genotyping was performed in The National Genotyping Centre of "Universidad de Santiago de Compostela".

Does were genotyped using the Affymetrix Axiom OrcunSNP Array (Affymetrix, Inc. Santa Clara, CA, USA) (Blasco & Pena, 2018). The SNP array contains 199,692 molecular markers. Quality control (QC) and genotype calling from raw data in the form of CEL files were implemented with Axiom Analysis Suite v. 4.0 and reanalysed by ZANARDI (Marras et al., 2017). The SNP quality control was performed using the following criteria: call rate ≥ 0.95, P-value > 1.0E-7 for the χ^2 test for Hardy Weinberg equilibrium, MAF ≥ 0.03 and only SNPs with known chromosome position. Animal samples were excluded from the dataset for values of dish quality control (DQC) < 0.89, missing genotype frequency > 0.03, Plate QC ≤ 0.96 or for failing a Mendelian segregation test. Missing genotypes were imputed by BEAGLE v4.1. SNPs with imputation quality score R^2 > 0.75 were included (Browning & Browning, 2009). After quality control, genotyping data for association analysis consisted of 181 samples and 117,791 SNPs.

3.3.4. Statistical analysis

Preceding to GWAS, we carried out a classical multidimensional scaling plot (Borg & Groenen, 2005) to find putative outliers or the presence of population stratification. The associations between SNPs and phenotypic traits were obtained using Bayes B Method. Briefly, this method computes all SNPs effects jointly and assumed for each marker a different genomic variance (Garrick & Fernando, 2013; Lehermeier et al., 2013). The following statistical model was used for the GWAS analysis:

$$\boldsymbol{y} = \boldsymbol{\mu} + \boldsymbol{X}\,\boldsymbol{b} + \sum_{j=1}^{k} \boldsymbol{z_j}\,\alpha_j\,\delta_j + \boldsymbol{e}$$

in which $\boldsymbol{y}$ is the vector of the phenotypic values; $\boldsymbol{\mu}$ is the trait mean, $\boldsymbol{X}$ is the incidence matrix for systematic effects; $\boldsymbol{b}$ is the vector with the systematic effects of year-season (five levels), line (high UC, low UC or control) and physiological state (lactating or non-lactating does); k is the total number SNP after quality control; $\boldsymbol{z_j}$ is the vector including the genotypic covariate for each SNP or locus j (0, 1 or 2 reference alleles); α_j is the random allele substitution effect for SNPj, which conditional on σ_α^2 is assumed normally distributed N(0 , $\boldsymbol{I} \cdot \sigma_\alpha^2$); δ_j is the random 0/1 variable that represents the presence (δ_j = 1, with probability 1-π) and the absence (δ_j = 0, with probability π) of the SNP in the model for a given iteration of the Markov chain; and $\boldsymbol{e}$ is the vector of the residual values with a normal distribution N(0 , $\boldsymbol{I} \cdot \sigma_e^2$) (Onteru et al., 2012; Cesar et al., 2014). The genomic variance for every SNP was denoted as σ_α^2 and the residual variance as σ_e^2. In Bayesian approaches, variance parameters can be treated as unknown, but having assumed prior distributions (Garrick & Fernando, 2013). In our study, we assigned the prior genomic variance of the SNPs derived from the estimated total genetic variance (Lehermeier et al., 2013). The prior variances for each trait were retrieved from previous experiments (García & Baselga, 2002; Blasco et al., 2005; Ragab et al., 2014) and are displayed in Table 1. A model including line effect can cause a reduction of the statistical power of the divergent selection experiment. The line effect can mistakenly capture markers effects with opposite frequencies between lines. Hence, GWAS analyses were repeated using a model without line effect.

The π value defines the proportion of SNPs having zero effects in each iteration. We performed several analyses before defining this parameter. The π values were evaluated within range of 0.99 to 0.9995. The π values were very high according to the limited number of animals in this study (Ros-Freixedes et al.,

2016). In addition, we also performed GWAS at chromosome level with π = 0.95 in order to corroborate the results consistency.

The parameters of the model were estimated with marginal posterior distributions using Markov chain Monte Carlo (MCMC). After some exploratory analyses, a total of 825,000 iterations were performed, with a burn-in period of 225,000 iterations. Only one sample every 60 iterations was saved to avoid the high correlation between consecutive samples. The GenSel® v. 4.90 software (Garrick & Fernando, 2013) was used for the GWAS analysis.

Table 1. Prior variances for Bayes B method.

Trait	σ_a^2	σ_e^2
Ovulation rate	1.5913	3.3816
Implanted embryos	1.6638	5.8987
Embryo survival	11.56×10^{-4}	27.74×10^{-4}
Foetal survival	8.96×10^{-4}	55.24×10^{-4}
Prenatal survival	2.25×10^{-4}	22.75×10^{-4}
Total number born	0.6495	5.2554
Number born alive	0.8589	9.8198
Number born dead	0.1261	0.6652

σ_a^2: additive genetic variance; σ_e^2: residual variance.

In our study, 2,171 genomic windows were allocated to the 21 autosomes and the chromosome "X", containing around 54 SNP markers by each one. Genomic windows were defined for each chromosome according to the rabbit genetic map of OryCun2.0 assembly, and the percentage of the genomic variance explained for non-overlapping genomic windows of one megabase was calculated by marginal posterior density (Onteru et al., 2012; Garrick & Fernando, 2013; Cesar et al., 2014). The genomic windows that explained at least 0.5% of the genomic variance of each trait and with a probability being higher than zero at chromosome level of at least 0.70 were considered to be putative QTL. This threshold of 0.5% was 10 times higher than the expected percentage of genomic variance explained for one window (Onteru et al., 2013; Cesar et al., 2014). In addition, we considered relevant those SNPs markers that overcome at least a

Bayes factor of 10, a value commonly considered as evidence of association (Kass & Raftery, 1995; Stephens & Balding, 2009; Ros-Freixedes et al., 2016). The posterior probability of association (PPA) suggested was not used as criterion of association since the low number of records with a high number of SNPs leads always low PPA values, even for real associations (Stephens & Balding, 2009: see their supplementary information). Hence, additional information such as the results consistency for different models and priors was used to identified the genomic regions associated to the traits.

3.3.5. Linkage disequilibrium, pathways and functional enrichment analyses

The analysis of LD was performed in order to assess its pattern within the consecutive associated windows. The aim of this analysis was to provide support for the association evidence. Hence, those windows with a great span of LD ($r^2 > 0.5$) and with SNPs associated within this LD block were considered as a true association with the trait. We assumed that these SNPs are a tag of the same causal variant. In addition, the LD analysis was performed within line, in order to understand the selection process. The R LDheatmap package was used for this analysis (Shin et al., 2006).

The position of the candidate genes was determined for each QTL using UCSC Rabbit Genome Browser (Rosenbloom et al., 2015). The gene annotations were provided by Ensembl Genes 97 database using Biomart Software (Aken et al., 2016) and "GenerCards" (Stelzer et al., 2016). Moreover, the functional enrichment analyses were performed by Gene Ontology (GO) (Ashburner et al., 2000) and "Database for Annotation, Visualization and Integrated Discovery" (DAVID) v 6.8 (Jiao et al., 2012).

3.4. Results and Discussion

3.4.1. Descriptive statistics of phenotypic data

Descriptive statistics for litter size traits of the rabbit lines of UC divergent selection experiment are shown in Table 2. The mean and standard deviation across lines for litter size traits were similar to other rabbit lines (Piles et al., 2006; Elmaghraby & Elkholya, 2010; Ragab et al., 2014). Apart from OR, there

were phenotypic differences between lines in all the traits. The most noticeable differences were for TNB with mean (standard deviation) of 10.11 (2.71), 7.01 (3.08), and 9.57 (2.82) kits for the high UC line, the low UC line and the control population, respectively; and for IE with 13.08 (2.65), 10.96 (3.04), and 12.07 (2.88) embryos; and for PS with 0.69 (0.17), 0.51 (0.21), and 0.65 (0.21), respectively.

Table 2. Descriptive statistics of little size traits.

Trait	N[1]	Mean	SD[2]	Min[3]	Max[4]
Ovulation rate (OR)	157	14.85	2.52	9.00	22.00
Implanted embryos (IE)	158	12.15	2.98	3.00	19.00
Embryo survival (ES)	154	0.82	0.17	0.25	1.00
Foetal survival (FS)	158	0.75	0.19	0.09	1.00
Prenatal survival (PS)	157	0.62	0.21	0.06	1.00
Total number born (TNB)	183	8.87	4.18	1.00	17.00
Number born alive (NBA)	183	8.25	3.98	0.00	15.00
Number born dead (NBD)	183	0.62	0.89	0.00	11.00

N[1]: Number of records; SD[2]: Standard deviation; Min[3]: Minimum; Max[4]: Maximum.

3.4.2. Description of genomic data

A total of 181 rabbits from the two UC lines and for a control line were genotyped with the Affymetrix Axiom OrcunSNP Array, which interrogates 199,692 SNPs. The criteria to exclude SNPs for the GWAS analysis were: minor allele frequency smaller than 0.03 (16.37%), unmapped SNPs (15.82%), mono-high resolution (8.65%), and call rate smaller than 0.95 (8.05%). After filtering, only 59% of SNPs in the array remained. This number was appropriate, taking into account the small phenotypic data size and the selection process performed before the UC experiment (Blasco *et al.*, 1994). Besides, the rabbit lines from "Universitat Politècnica de València" were not considered to design the actual SNP-array. Thus, an important number of SNPs (17,282) was fixed in the experimental UC lines. The average distance between SNPs was 18.90 kb along the genome leading to a LD average around 0.79 for 100 kb, and 0.76 when all genomic data in consecutive pairs SNPs were used. This value seems to be high considering

that an average distance of 98 kb showed a LD of 0.5, calculated within rabbit strains (Carneiro *et al.*, 2011).

The multidimensional scaling analysis using genomic data found an evident population stratification (Figure 1). This analysis identified three clusters corresponding to the high UC line, the low UC line, and the control population, respectively. The first two principal components jointly explained 23.6% of the total variance. This would indicate that SNPs captured the population stratification of this experiment. Bayesian multi-marker regression models are quite robust to population stratification (Toosi *et al.*, 2018). Although the inclusion of line effect reduced the power obtained by the divergent selection, we included the line effect in order to avoid the possible drift effect and check the consistency of the results. We are aware that this type of correction is very stringent. So, we also performed the analysis without line. The variance explained for the main associated region increased considerably (Table 3). However, the conclusions our findings did not change. The regions identified as associated were identical and with the similar order of importance which showed results consistency with and without line effect.

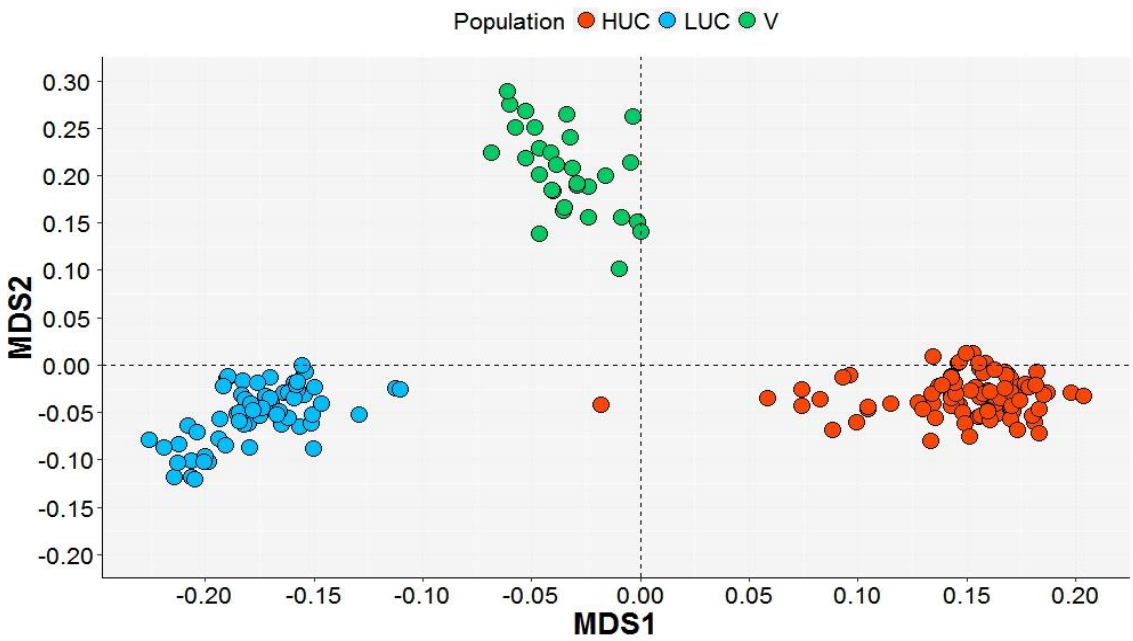


Figure 1. Multidimensional scaling plot of the genomic data. The first component (MDS1) explained 16.73% of the genomic variance and the second component (MDS2) explained 6.90% of the genomic variance. Populations: high uterine capacity line (HUC), low uterine capacity line (LUC) and control population or line selected for number of kits at weaning (V).

3.4.3. Prior choice

The exploratory analysis of the π value under the model without line effect showed similar results across π values, being not sensitive to them. By contrast, the model with line effect showed a greater increase of shrinkage led to a lower number of windows overcoming the relevant threshold. Additionally, the percentage of the genomic variance explained by these associated windows was greater when the π value was greater. For instance: using a π value of 0.9995 the analysis reported four consecutive genomic windows associated with TNB that explained 16.3% of the genomic variance, whilst using 0.9992, 0.9975, 0.995, and 0.99, these explained 7.4%, 2.8%, 1.4% and 0.6%, respectively. However, the ranking of the relevant genomic windows did not change. Therefore, the π value used in this study was 0.9992 based on the average number of SNPs in the model per iteration (119) and the total number of samples (181).

3.4.4. Genomic windows associated with litter size traits

The GWAS analyses showed associated genomic windows for TNB, NBA, IE, and OR. No associations were evidenced for NBD, ES, FS, and PS.

3.4.4.1. Total number born and number born alive

The genomic windows associated with TNB are located on chromosome 17 (windows 1903, 1904, 1905 and 1906) (Figure 2). Two of them (1905 and 1906) also showed association with NBA (Figure 3). The genomic variance explained by these two windows was 6.32% for TNB and 1.27% for NBA (Table 3). This result would be in agreement with the high genetic correlation found between NBA and TNB (0.964 +/- 0.008) (García & Baselga, 2002).

The associated genomic region (70.0 - 73.3 Mb) seems to have a major effect on TNB in the UC lines. This could make sense since half of response of selection was obtained in the first two generations of UC divergent selection (Blasco et al., 2005). This region accounted for up to 38.82% and 10.36 % of the genomic variance for TNB and NBA, respectively, under a model excluding the line effect.

In addition, the genomic variance explained by all these genomic windows had a probability of being greater than zero at chromosome level of at least 0.95, except the 1906 being greater than 0.75.

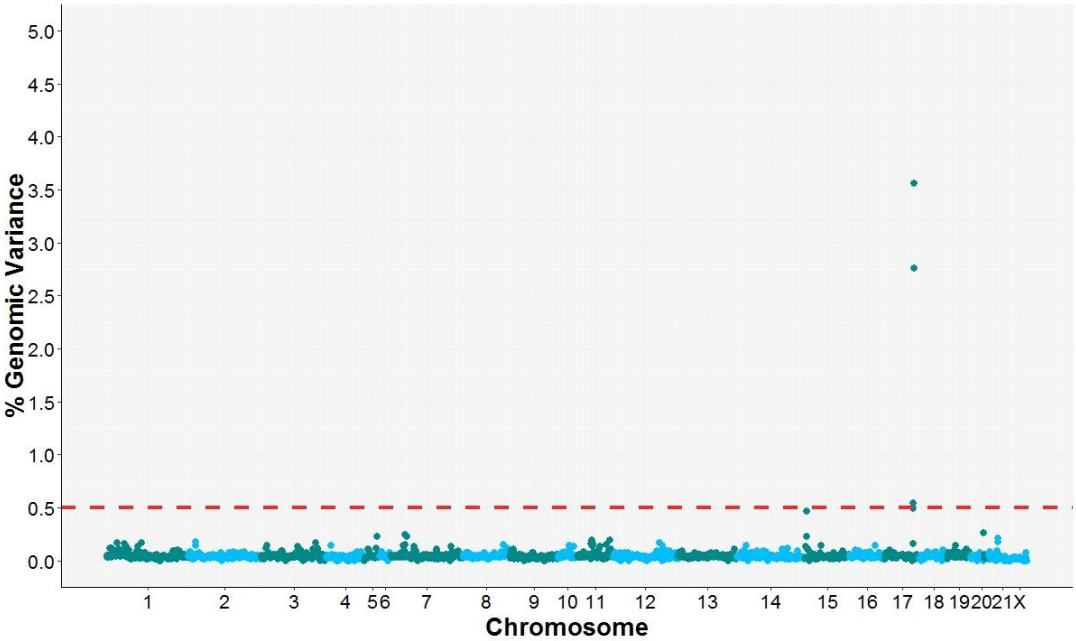


Figure 2. Manhattan plot for total number born (TNB) using the percentage of genomic variance explained by each non-overlapping one megabase window.

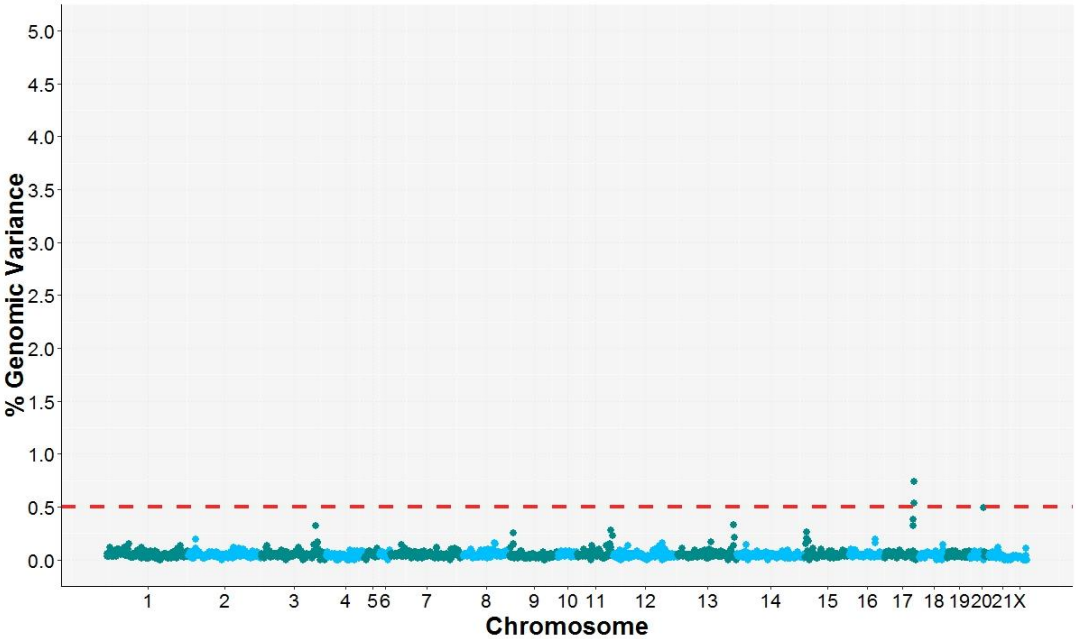


Figure 3. Manhattan plot for number born alive (NBA) using the percentage of genomic variance explained by each non-overlapping one megabase window.

Other genomic regions with a smaller effect size than the region associated on chromosome 17 could not have been identified due to the small sample size. In swine, GWAS analyses for TNB and NBA have reported QTLs in several chromosomes. However, the sample size in these studies was greater (>600), and in both studies, third terminal crossbred lines were used (Onteru et al., 2012; Schneider et al., 2012), generating a much higher LD in their population than in our lines.

3.4.4.2. Implanted embryos

A large relevant genomic region for IE was found on chromosome 11 (Figure 4). This region involved five associated genomic windows (35.2 – 39.0 Mb), from window 1143 to 1147, accounting for 10% of the genomic variance of IE (Table 3). Besides, the same genomic region on chromosome 17 associated with TNB and NBA explained 5.37% (32.23 % without line) of the genomic variance of IE. Therefore, this region could have a pleiotropic effect on these three litter size traits (TNB, NBA, and IE). These results could be related to the correlated response to selection for IE, shown in the UC divergent selection experiment (Santacreu et al., 2005; Blasco et al., 2005) which is in agreement with the moderate to high genetic correlation between IE and UC (0.66) (Blasco et al., 2005) and IE and TNB (0.46) (Laborda et al., 2012).

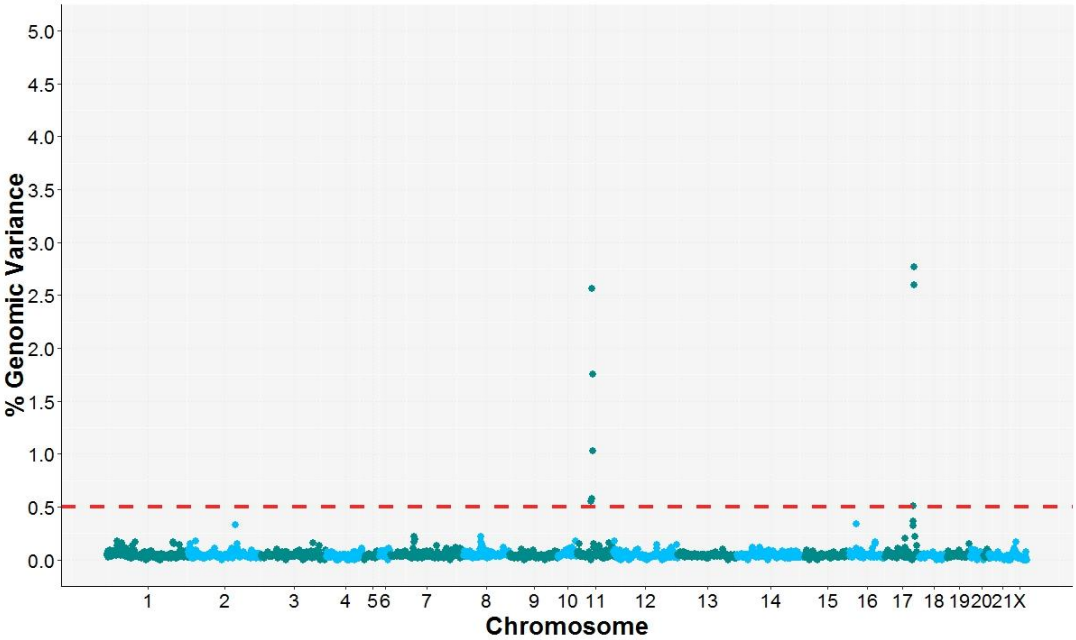


Figure 4. Manhattan plot for implanted embryos (IE) using the percentage of genomic variance explained by each non-overlapping one megabase window.

3.4.4.3. Ovulation rate

The results did not show a strong genomic association for this trait due to the low amount of genomic variance explained by each associated window. Moreover, none of the windows were consecutive. Two genomic windows on chromosome 9, window 996 and 993, only explained 1.13% (0.84 % without line) and 1.03% (0.94 % without line) of the genomic variance, respectively (Table 3). Overall, all genomic windows associated with OR accounted for 3.95% (with and without line) of the genomic variance. This result is in contrast to a swine GWAS that found three relevant genomic regions associated with OR explaining 51% of the genomic variance (Schneider et al., 2014). The sample size of their study was considerably greater than in our study, and the swine population had much higher LD and genomic variability. Moreover, in our study animals came from a divergent selection experiment for UC, whose trait had a moderate (0.56) genetic correlation with OR (Blasco et al., 2005). Additionally, the genomic windows associated with OR did not agree with the associated genomic region found for three litter size traits - TNB, NBA, and IE (Figure 5). These results are in concordance with the null correlated response in litter size for OR selection in rabbits and the low genetic correlation estimated between OR and litter size (Laborda et al., 2011).

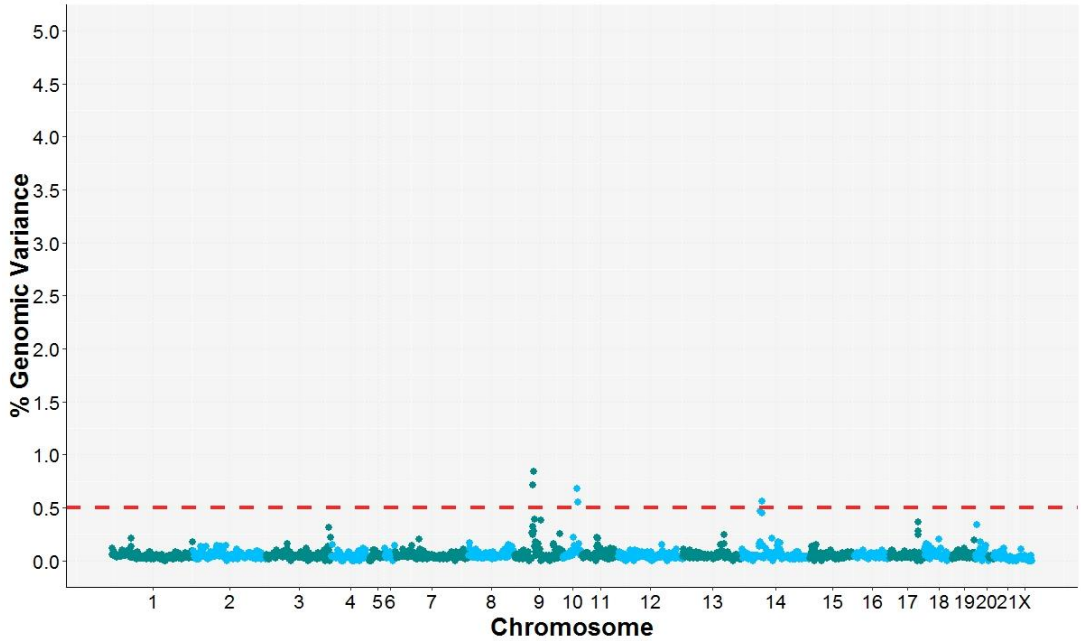


Figure 5. Manhattan plot for ovulation rate (OR) using the percentage of genomic variance explained by each non-overlapping one megabase window.

Table 3. Genomic windows associated with total number born (TNB), number born alive (NBA), implanted embryos (IE), and ovulation rate (OR) in rabbits.

Window ID	Chr	Position in Mb	Traits	%Var		#SNP	Genes
				+ Line	- Line		
993	9	42.0 - 43.0	OR	1.03	0.84	55	*CNTN3, 5S_rRNA*†*, U6*†
996	9	47.0 - 48.0	OR	1.13	0.94	52	*C4orf3, ENSOCUG00000021038, ENSOCUG00000002078, ENSOCUG00000025665, ENSOCUG00000023430*
1097	10	35.0 - 36.0	OR	0.68	0.64	80	*CALCR, U6*†*, ENSOCUG00000020017*†*, VPS50, HEPACAM2, SAMD9L, SAMD9, GINS2, ENSOCUG00000029687, CDK6*
1100	10	38.0 - 39.0	OR	0.55	0.89	72	*CDK14, CLDN12, GTPBP10, CFAP69, STEAP2, STEAP1*
1143	11	35.2 - 35.9	IE	0.79	0.25	55	*FAM173B, CCT5, CMBL, MARCH6, ROPN1L, ANKRD33B, ENSOCUG00000010666*
1144	11	36.0 - 37.0		0.89	0.45	76	*CTNND2, 5S_rRNA*†*, ENSOCUG00000027984*
1145	11	37.0 - 38.0		3.83	1.46	85	*U6*†
1146	11	38.0 - 39.0		1.71	0.62	66	*DNAH5, ENSOCUG00000025796*†
1147	11	39.0 - 40.0		2.77	1.33	71	*TRIO, FAM105A, OTULIN, ANKH, 5S_rRNA*†*, U6*†
1535	14	51.0 - 52.0	OR	0.56	0.64	83	*TIPARP, SNORA65*†*, LEKR1, U2*†*, CCNL1, VEPH1, PTX3, SNORD90*†

Window ID	Chr	Position in Mb	Traits	%Var		#SNP	Genes
				+ Line	- Line		
1902	17	69.1 - 70.0	IE	0.51	0.87	59	*VCPKMT, SOS2, L2HGDH, ATP5S, CDKL1, MAP4K5, ATL1, SAV1, NIN, ABHD12B, PYGL,*
1903	17	70.0 - 71.0	TNB	0.50	0.36	51	*TRIM9, TMX1, FRMD6, GNG2, ENSOCUG00000014681, NID2, SCARNA23*†
1904	17	71.0 - 72.0		0.54	0.30	61	*PTGDR, PTGER2, TXNDC16, GPR137C, ERO1A, PSMC6, STYX, FERMT2, DDHD1, 7SK*†*, ENSOCUG00000007858*
1905	17	72.0 - 73.0	TNB	2.76	18.72	66	*PNRC2, BMP4, 5S_RNA*†*, U4*†*, snoU13*†
			NBA	0.53	3.52		
			IE	2.77	16.95		
1906	17	73.1 - 73.3	TNB	3.56	20.10	16	*CDKN3, GMFB, CGRRF1, SAMD4A*
			NBA	0.74	6.84		
			IE	2.60	15.28		

Window ID: window identification; Chr: chromosome; Position in Mb: position of the genomic window in megabases on the OryCun2.0 corresponding chromosome. This indicates the position from the first to the last SNP within each window: %Var: percentage of genomic variance accounted for by the genomic window; + Line: including the line effect in the model; - Line: without the line effect in the model; #SNP: number of SNPs into the window; Genes: annotated genes in the window portion delimited by the SNPs included in the window. The pseudogenes are not included in this table. † Non-coding genomic DNA.

3.4.5. Associated SNPs in genomic regions

The Bayes factor criteria showed only relevant SNP associations for IE and TNB. These associated SNPs map to chromosome 11 for IE (Figure 6), and chromosome 17 for TNB and IE (Figure 6 and 7). The highest Bayes factor was for a SNP on chromosome 17, associated with TNB under the model without the line effect. The total number of SNPs between the two traits in chromosome 17 was 14 (five in the window 1905 and nine in the window 1906) (Table 4). This corroborated the remarkable importance of this genomic region on chromosome 17 as a putative QTL. However, the PPAs of SNPs within the putative QTL were low (0.04 to 0.15), which is expected with the small sample size used in our study. Stephens & Balding (2009) pointed out that PPAs have the advantage of being not very sensitive on sample size, power and number of tested SNPs. Despite that, they showed that small sample size can give low PPAs with real associations even under several Bayesian approaches based on different priors, according to their supplementary material. In our study, the putative QTL on chromosome 17 was consistent across the analyses of GWAS, under window and SNP association criteria, allele frequencies and linkage disequilibrium analyses (as shown below). All associated SNPs had an overall MAF above 0.28. Moreover, the associated SNPs for both TNB and IE showed an even higher value of MAF (from 0.33 to 0.49). The allele frequencies in the control population for these associated SNPs were intermediate (0.43 - 0.45), whilst they were higher for the low UC line (0.64 and 0.75) and very low (0.05) for the high UC line. We assumed that all of these SNPs were associated with the traits (TNB and IE) due to strong LD with their causal variants since selection could have modified the allelic frequencies of the SNPs associated with the causal variants. In this case, the joint analysis of the divergent selection would have led to intermediate frequencies, increasing the SNP detection power (López de Maturana et al., 2014; Kessner & Novembre, 2015). Thus, our experiment has been valuable for revealing novel QTLs associated with litter size traits in rabbits.

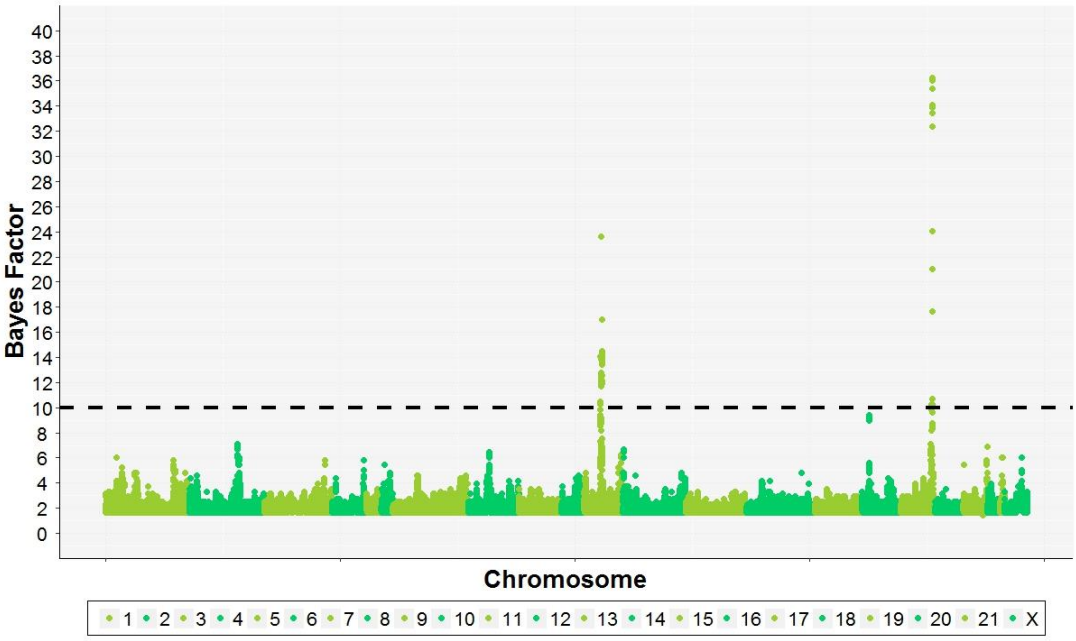


Figure 6. Manhattan plot for implanted embryos (IE) using the Bayes factors by each SNP along the rabbit chromosomes.

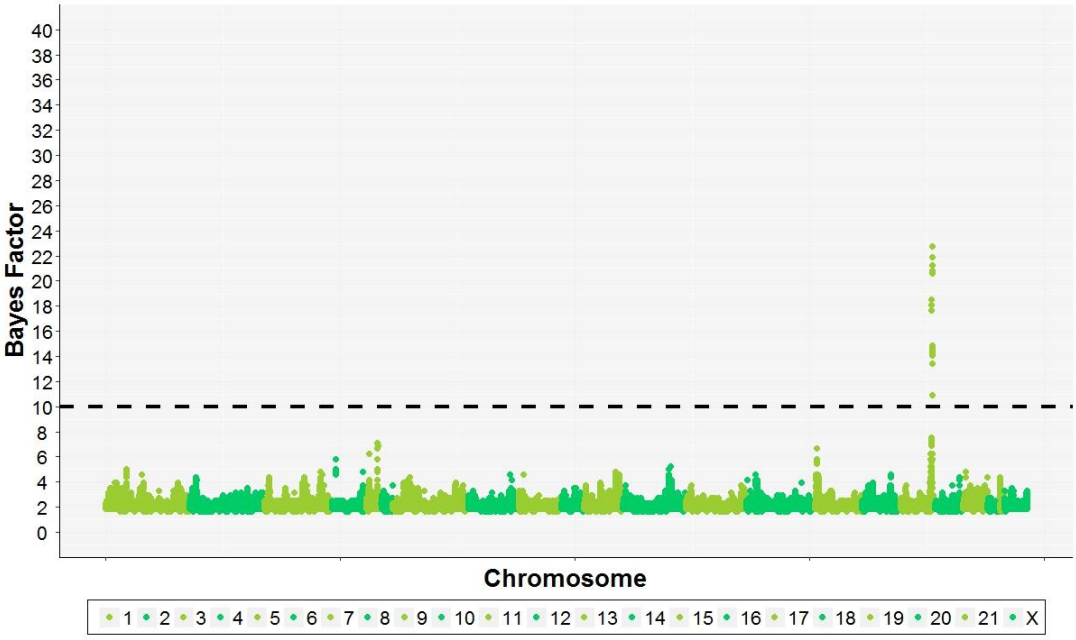


Figure 7. Manhattan plot for total number born (TNB) using the Bayes factors by each SNP along the rabbit chromosomes.

Table 4. Relevant polymorphisms (SNPs) for total number born (TNB) and implanted embryos (IE).

Window ID	SNP name	Chr	Position in Mb	Bayes Factor				Allele Reference	MAF
				+ Line		- Line			
				IE	TNB	IE	TNB		
1905	Affx-151926619	17	72.14	10.04	4.38	11.47	7.67	C	0.30
	Affx-151908415	17	72.15	10.25	4.17	11.98	7.67	T	0.30
	Affx-151904115	17	72.15	10.25	4.17	10.83	8.30	G	0.30
	Affx-151825298	17	72.23	8.15	18.49	4.51	24.84	A	0.50
	Affx-151870244	17	72.23	8.15	18.06	4.64	21.60	C	0.50
	Affx-152013420	17	72.25	8.15	17.64	4.64	26.14	A	0.50
	Affx-151801784	17	72.93	36.01	21.88	97.63	110.23	G	0.33
	Affx-151957551	17	72.95	36.23	20.61	94.44	101.12	T	0.33
	Affx-151955776	17	72.97	32.35	20.61	95.60	110.09	A	0.33
	Affx-151991400	17	72.98	34.07	21.25	94.16	122.62	G	0.33
	Affx-151972019	17	72.99	17.64	13.41	30.06	41.56	G	0.34
1906	Affx-151858851	17	73.10	21.03	10.88	40.09	31.63	G	0.34
	Affx-151802659	17	73.11	33.43	20.61	97.49	114.54	A	0.33
	Affx-151975417	17	73.11	35.37	22.73	97.49	114.69	T	0.33
	Affx-151955414	17	73.13	33.86	21.25	93.87	115.13	T	0.33
	Affx-151943719	17	73.13	24.01	20.82	57.08	70.32	T	0.33
	Affx-151913508	17	73.17	8.57	14.26	5.02	20.44	G	0.49
	Affx-151985483	17	73.18	10.04	14.26	6.15	22.89	T	0.49
	Affx-151933136	17	73.20	8.78	14.89	5.27	20.05	G	0.49

Window ID	SNP name	Chr	Position in Mb	Bayes Factor				Allele Reference	MAF
				+ Line		- Line			
				IE	TNB	IE	TNB		
1906	Affx-151974640	17	73.21	8.78	14.89	5.14	19.79	C	0.49
	Affx-151983535	17	73.22	10.04	14.89	5.90	23.80	T	0.49
	Affx-151823935	17	73.22	8.36	14.05	5.02	16.71	T	0.49
	Affx-151860280	17	73.24	8.57	14.47	5.14	21.08	G	0.49
	Affx-151999537	17	73.25	8.78	14.68	5.02	19.41	A	0.49
	Affx-151995315	17	73.27	10.25	14.47	6.40	23.80	T	0.49
	Affx-151909593	17	73.28	10.04	14.89	5.90	32.81	T	0.49

Window ID: window identification; Chr: chromosome; Position in Mb, position of the genomic window in megabases on the OryCun2.0 corresponding chromosome; + Line: including the line effect in the model; -Line: without the line effect in the model; MAF: minor allele frequency. The threshold value of Bayes factor was 10.

3.4.6. Linkage disequilibrium analysis

We assessed the LD in the consecutive associated genomic windows on chromosome 11 and 17. The genomic regions associated with IE (chromosome 11) showed a strong LD block amongst the windows 1145, 1146 and 1147. This block was more evident in the low UC line. This suggests that this QTL could have been under higher selection pressure for low UC than for high UC, in agreement with the asymmetric response estimated using the UC lines and the cryopreserved control population. This latter study showed the selection response was higher in the low UC line (Mocé et al., 2005; Santacreu et al., 2005). The SNPs that overreached the threshold for IE are indicated with black points in Figure 8. Most of them are mapped in the LD block made up by the three windows (1145, 1146 and 1147). This result is in contrast to the genomic region associated with TNB, NBA, and IE on chromosome 17 displaying several short LD blocks. Most of the associated SNPs within this QTL were in the window 1906 (Figure 9). This window presents a steady LD block within the control population ($r^2 > 0.8$). This would indicate that the UC selection formed new LD blocks from a large one in the control population.

In our study, both LD and GWAS results support the idea that QTL on chromosome 17 had a major impact on the divergent selection experiment. This hypothesis of an important QTL for litter size in the UC lines is supported by the great response at the second generation, half of the estimated response in this divergent selection experiment, as we said previously (Argente et al., 2003; Blasco et al., 2005).

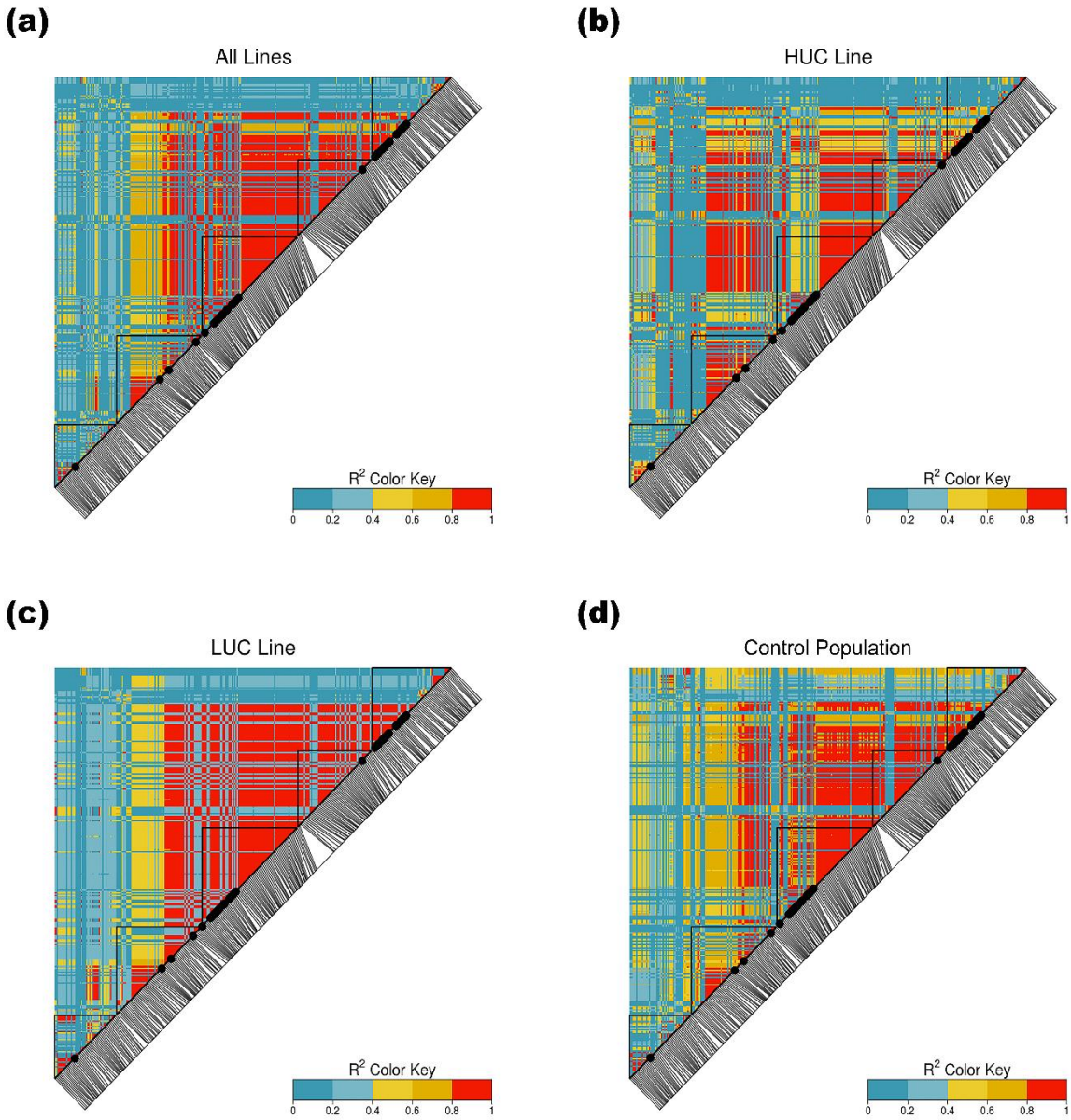


Figure 8. Linkage disequilibrium plot of chromosome 11 (35.2 – 40.0 Mb). Physical length is 4756 kb and contains a total of 353 SNPs. The black triangle stands for each one of five associated windows for implanted embryos. The black points are the 38 associated SNP. The colour red is the R-squared from 0.8 to 1.0 (strong LD). The computation was performed using data from **(a)** all lines, **(b)** HUC (high UC line), **(c)** LUC (low UC line) and **(d)** control population.

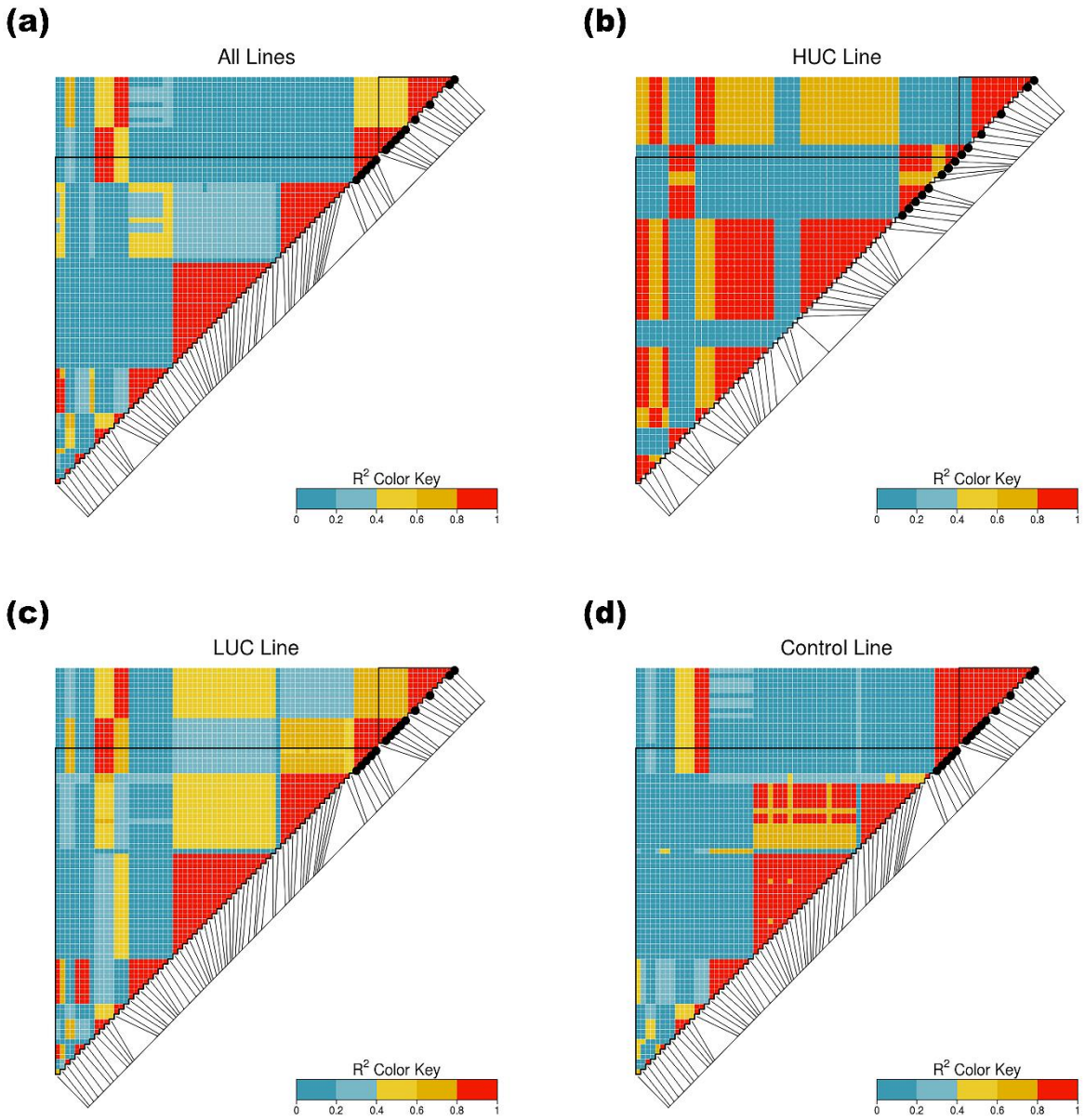


Figure 9. Linkage disequilibrium plot of chromosome 17 (72.0 – 73.2 Mb). Physical length is 1278 kb and contains a total of 82 SNPs. The black triangle stands for each one of two associated windows for total number born, number born alive and implanted embryos. The black points are the 14 associated SNP for total number born and implanted embryos. The colour red is the R-squared from 0.8 to 1.0 (strong LD). The computation was performed using data from **(a)** all lines, **(b)** HUC (high UC line), **(c)** LUC (low UC line) and **(d)** control population.

3.4.7. Gene search and functional annotations

The associated genomic regions disclose 72 coding and noncoding genes (additional file 1: Table S1); nine of them located on the genomic region associated with TNB, NBA and IE (chromosome 17) (Table 3). The top five results of the functional annotation analysis, using the genes in putative QTLs, are shown in Table 5. The human, mice and rabbit functional annotations from DAVID databases gave similar results. Therefore, we described these results using the annotated rabbit genes to subsequently perform a detailed functional seeking for each gene. The most relevant functions were linked to terms such as activity prostanoid receptors, cellular response to prostaglandin, negative regulation of striated muscle tissue development, carbohydrate derivative binding, and cyclin-dependent protein kinase activity. The genes related to reproductive processes and associated with TNB were *PTGDR, PTGER2, BMP4, STYX,* and *CDKN3*. The *PTGDR* and *PTGER2* belong to the prostaglandins receptor family which are essential for the adequate performance of uterus; mainly prostaglandin F receptor that presents underlying functions over the female reproductive cycle in mammals (Blesson & Sahlin, 2014). Also, a severe deficiency in the *PTGER2* genetic expression decreases fertilization and generates defects in cumulus expansion (Matzuk & Lamb, 2002). Otherwise, *PTGDR* gene presents an important role in the differentiation of germ and Sertoli cells of the embryonic testis in males (Rossitto et al., 2014). Genes of the transforming growth factor-β superfamily, including *BMP4*, are involved in follicular growth and development in mammals (Al-Samerria et al., 2015) avoiding the apoptosis of oocytes through regulation of both *Sohlh2* and *c-ki* (Ding et al., 2013). Nevertheless, the *BMP4* gene showed no association with OR, but it was associated with TNB and IE in our study. *BMP4* has been also implicated in trophoblast development, implantation, and placentation in humans (Li & Parast, 2014). *CDKN3* gene is related to inhibition and reduction of choline, particularly in the neural progenitor cells of the foetal hippocampus, producing cellular apoptosis (Zeisel, 2011). Moreover, the reduction of *STYX* expression disrupts spermatid development (Matzuk & Lamb, 2002). The 1903 window on chromosome 17, associated only with TNB, contains the *ERO1A* gene. This gene did not show a functional annotation directly related to reproductive processes but was identified as overexpressed between the UC lines in a previous study (Ballester et al., 2013). Moreover, it is the precursor of

the *ER1L* transcript, which is related to redox homeostasis and oxidative protein folding in the endoplasmic reticulum (Konno et al., 2015).

Regarding genes associated with IE, *BMP4* and *CDKN3* genes (chromosome 17) are annotated to embryo development processes in mice (Goggolidou et al., 2013). In chromosome 11, we found the *CCT5* gene related to sperm quality in bulls (Yathish et al., 2017). Finally, the genes annotated for OR did not have a direct relationship with this trait or the female reproductive physiology.

Previous candidate gene studies, using the UC divergent lines, showed genes associated with reproductive traits such as progesterone receptor (*PGR*) associated with IE (Peiró et al., 2008), tissue inhibitor of metalloproteinases 1 (*TIMP1*) associated with number of embryos (Argente et al., 2010) and oviduct glycoprotein 1 (*OVGP1*) associated with TNB (Merchán et al., 2009). However, our study did not identify associated genomic regions close to these genes.

In general, the candidate genes found in our study are different from those identified in GWAS for OR, TNB and NBA in swine (Onteru et al., 2012; Schneider et al., 2014; Bergfelder-Drüing et al., 2015). The main associations in these studies did not overlap amongst litter size traits. However, Schneider et al. (2012) found overlapping genomic windows for TNB, NBA, NBD and average piglet birth weight in swine; similar to the novel putative QTL found on chromosome 17.

Table 5. Top five functional enrichment from the analyses performed through DAVID online web.

Category	Term's CODE	Term	Genes	P-Value	Fold Enrichment	Bonferroni
Total Number Born (TNB)						
INTERPRO	IPR008365	Prostanoid receptor	*PTGER2, PTGDR*	7.38E-03	256.07	4.09E-01
GOTERM_MF_ALL	GO:0004955	prostaglandin receptor activity	*PTGER2, PTGDR*	1.10E-02	169.06	5.68E-01
GOTERM_BP_ALL	GO:0071379	cellular response to prostaglandin stimulus	*PTGER2, PTGDR*	1.75E-02	106.86	1.00E+00
GOTERM_BP_ALL	GO:0034694	response to prostaglandin	*PTGER2, PTGDR*	2.22E-02	83.96	1.00E+00
KEGG_PATHWAY	ocu05200	Pathways in cancer	*GNG2, BMP4, PTGER2*	3.94E-02	77.55	6.19E-01
Implanted Embryos (IE)						
INTERPRO	IPR023235	FAM105	*FAM105A, OTULIN*	3.83E-03	503.31	2.81E-01
GOTERM_BP_FAT	GO:0045843	negative regulation of striated muscle tissue development	*BMP4, SAV1*	2.92E-02	63.90	1.00E+00
GOTERM_BP_FAT	GO:0048635	negative regulation of muscle organ development	*BMP4, SAV2*	3.09E-02	60.35	1.00E+00
GOTERM_MF_FAT	GO:0097367	carbohydrate derivative binding	*CCT5, PYGL, BMP4, ATL1, MAP4K5, CDKL1, TRIO*	4.89E-02	2.37	9.95E-01
GOTERM_BP_FAT	GO:0060428	lung epithelium development	*BMP4, SAV1*	5.09E-02	36.21	1.00E+00
Ovulation Rate (OR)						
GOTERM_MF_ALL	GO:0097472	cyclin-dependent protein kinase activity	*CDK14, CDK6*	4.82E-03	362.29	2.97E-01

Category	Term's CODE	Term	Genes	P-Value	Fold Enrichment	Bonferroni
GOTERM_MF_ALL	GO:0004693	cyclin-dependent protein serine/threonine kinase activity	*CDK14, CDK7*	4.82E-03	362.29	2.97E-01
INTERPRO	IPR013130	Ferric reductase transmembrane component-like domain	*STEAP2, STEAP1*	1.13E-02	165.86	3.78E-01
KEGG_PATHWAY	ocu04978	Mineral absorption	*STEAP2, STEAP2*	2.95E-02	55.93	4.33E-01
UP_KEYWORDS		Cyclin	*CCNL1, CDK6*	3.14E-02	59.31	5.64E-01

3.5. Conclusions

Our study reveals associations between genomic regions and TNB, NBA, IE, OR. Two consecutive genomic windows on chromosome 17 were associated with three traits (TNB, NBA, and IE), and accounted for a meaningful percentage of the genomic variance for TNB, indicating that this genomic region could contain remarkable causal variants for litter size traits in rabbits. In addition, a genomic region on chromosome 11 appears particularly important for IE. The associated genomic regions harboured 72 genes. However, few of these genes were profiled as physiological candidate genes due to their link to reproductive processes (*i.e.*, *BMP4, PTDGR, PTGER2, STYX, and CDKN3)*. In summary, our results disclosed new putative QTLs for TNB and IE, likely responsible for the large divergent response obtained in the first two generations of selection. However, these results must be validated in independent maternal rabbit lines before being used in breeding programs. This study is the first GWAS for reproductive traits in rabbits and provides a starting point to disentangle the genetic basis of litter size traits in rabbits.

3.6. Abbreviations

ES: Embryo survival; FS: Foetal survival; GO: Gene ontology; GV: genomic variance; GWAS: Genome-wide association study; IE: Implanted embryos; LD: Linkage disequilibrium; MAF: Minor allele frequency; NBA: Number born alive; NBD: Number born dead; OR: Ovulation rate; PPA: the posterior probability of association; PS: Prenatal survival; QC: Quality control; QTL: Quantitative trait loci; SNP: Single nucleotide polymorphism; TNB: Total number born; UC: Uterine capacity, ULO: unilaterally ovariectomized.

3.7. References

Aken, B. L., S. Ayling, D. Barrell, L. Clarke, V. Curwen et al., 2016. "The Ensembl gene annotation system." *Database (Oxford),* 2016: baw093.

Al-Samerria, S., I. Al-Ali, J. R. McFarlane, and G. Almahbobi, 2015. "The impact of passive immunisation against BMPRIB and BMP4 on follicle development and ovulation in mice." *Reproduction,* 149(5): 403–411.

Argente, M. J., A. Blasco, J. A. Ortega, C. S. Haley, and P. M. Visscher, 2003. "Analyses for the presence of a major gene affecting uterine capacity in unilaterally ovariectomized rabbits." *Genetics,* 163(3): 1061–1068.

Argente, M. J., M. Merchán, R. Peiró, M. L. García, M. a. Santacreu et al., 2010. "Candidate gene analysis for reproductive traits in two lines of rabbits divergently selected for uterine capacity." *Journal of Animal Science,* 88(3): 828–836.

Argente, M. J., M. A. Santacreu, A. Climent, G. Bolet, and A. Blasco, 1997. "Divergent selection for uterine capacity in rabbits." *Journal of Animal Science,* 75(9): 2350–2354.

Ashburner, M., C. A. Ball, J. A. Blake, D. Botstein, H. Butler et al., 2000. "Gene Ontology: tool for the unification of biology." *Nature Genetics,* 25(1): 25–29.

Badawy, A. Y., R. Peiró, A. Blasco, and M. A. Santacreu, 2018. "Correlated responses on litter size traits and survival traits after two-stage selection for ovulation rate and litter size in rabbits." *Animal,* 13(3): 453-459.

Ballester, M., A. Castelló, R. Peiró, M. J. Argente, M. A. Santacreu et al., 2013. "Identification of differentially expressed genes in the oviduct of two rabbit lines divergently selected for uterine capacity using suppression subtractive hybridization." *Animal Genetics,* 44(3): 296–304.

Bergfelder-Drüing, S., C. Grosse-Brinkhaus, B. Lind, M. Erbe, K. Schellander et al., 2015. "A genome-wide association study in large white and landrace pig populations for number piglets born alive." *PLoS ONE,* 10(3): e0117468.

Blasco, A., M. J. Argente, C. S. Haley, and M. A. Santacreu, 1994. "Relationships between components of litter size in unilaterally ovariectomized and intact rabbit does." *Journal of Animal Science,* 72(12): 3066–3072.

Blasco, A., J. A. Ortega, A. Climent, and M. A. Santacreu, 2005. "Divergent selection for uterine capacity in rabbits. I. Genetic parameters and response to selection." *Journal of Animal Science,* 83(10): 2297–2302.

Blasco, A., and R. N. Pena, 2018. "Current Status of Genomic Maps: Genomic Selection/GBV in Livestock", pp. 61–80 in *Animal Biotechnology 2*, edited by H. Niemann and C. Wrenzycki. Springer International Publishing, Cham.

Blesson, C. S., and L. Sahlin, 2014. "Prostaglandin E and F receptors in the uterus." *Receptors & Clinical Investigation,* 1: e115

Borg, I., and P. J. F. Groenen, 2005. "Modern Multidimensional Scaling: Theory and Applications." Springer International Publishing.

Browning, B. L., and S. R. Browning, 2009. "A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals." *American Journal of Human Genetics,* 84(2): 210–223.

Carneiro, M., S. Afonso, A. Geraldes, H. Garreau, G. Bolet et al., 2011. "The genetic structure of domestic rabbits." *Molecular Biology and Evolution,* 28(6): 1801–1816.

Carneiro, M., C. J. Rubin, F. Di Palma, F. W. Albert, J. Alföldi et al., 2014b. "Rabbit genome analysis reveals a polygenic basis for phenotypic change during domestication." *Science*, 345(6200): 1074–1079.

Cartuche, L., M. Pascual, E. A. Gómez, and A. Blasco, 2014. "Economic weights in rabbit meat production." *World Rabbit Science,* 22(3): 165–177.

Cesar, A. S., L. C. Regitano, G. B. Mourão, R. R. Tullio, D. P. Lanna et al., 2014. "Genome-wide association study for intramuscular fat deposition and composition in Nellore cattle." *BMC Genetics*, 15(1): 39.

Christenson, R. K., K. A. Leymaster, and L. D. Young, 1987. "Justification of unilateral hysterectomy-ovariectomy as a model to evaluate uterine capacity in swine." *Journal of Animal Science,* 65(3): 738–744.

Ding, X., X. Zhang, Y. Mu, Y. Li, and J. Hao, 2013. "Effects of BMP4/SMAD signaling pathway on mouse primordial follicle growth and survival via up-regulation of Sohlh2 and c-kit." *Molecular Reproduction and Development* 80(1): 70–78.

Elmaghraby, M. M. A., and S. Z. Elkholya, 2010. "Characterizing Litter Losses in Purebred New Zealand White Rabbits." *Lucrări Ştiinţifice - Universitatea de Ştiințe Agricole şi Medicină Veterinară Iaşi,* 54: 304–310.

European Economic Community, 1998. "Council Directive 98/58/EC of 20 July 1998 concerning the protection of animals kept for farming purposes." *Official Journal L* 221(8): 8.

Fan, B., Z. Q. Du, D. M. Gorbach, and M. F. Rothschild, 2010. "Development and application of high-density SNP arrays in genomic studies of domestic animals." *Asian-Australasian Journal of Animal Science,* 23(7): 833–847.

García, M. L., and M. Baselga, 2002. "Genetic Response to Selection for Reproductive Performance in a Maternal Line of Rabbits." *World Rabbit Science,* 10(2): 71–76.

Garrick, D. J., and R. L. Fernando, 2013. "Implementing a QTL Detection Study (GWAS) Using Genomic Prediction Methodology", pp. 275–298 in *Genome-*

Wide Association Studies and Genomic Prediction, edited by C. Gondro, J. van der Werf, and B. Hayes. Humana Press, Totowa, NJ.

Goggolidou, P., S. Soneji, N. Powles-Glover, D. Williams, S. Sethi et al., 2013. "A chronological expression profile of gene activity during embryonic mouse brain development." *Mammalian Genome,* 24(11-12): 459–472.

Jiao, X., B. T. Sherman, D. W. Huang, R. Stephens, M. W. Baseler et al., 2012. "DAVID-WS: A stateful web service to facilitate gene/protein list analysis." *Bioinformatics,* 28(13): 1805–1806.

Kass, R. E., and A. E. Raftery, 1995. "Bayes Factors." *Journal of the American Statistical Association,* 90(430): 773-795.

Kessner, D., and J. Novembre, 2015. "Power analysis of artificial selection experiments using efficient whole genome simulation of quantitative traits." *Genetics,* 199(4): 991–1005.

Khalil, M. H., and A. M. Al-Saef, 2008. "Methods, criteria, techniques and genetic responses for rabbit selection: A review", pp. 1-22 in *Proc. 9th World Rabbit Congress*, Verona, Italy.

Konno, T., E. Pinho Melo, C. Lopes, I. Mehmeti, S. Lenzen et al., 2015. "ERO1-independent production of H 2 O 2 within the endoplasmic reticulum fuels Prdx4-mediated oxidative protein folding." *Journal of Cell Biology,* 211(2): 253–259.

Laborda, P., M. L. Mocé, A. Blasco, and M. A. Santacreu, 2012. "Selection for ovulation rate in rabbits: Genetic parameters and correlated responses on survival rates." *Journal of Animal Science,* 90(2): 439–446.

Laborda, P., M. L. Mocé, M. A. Santacreu, and A. Blasco, 2011. "Selection for ovulation rate in rabbits: Genetic parameters, direct response, and correlated response on litter size." *Journal of Animal Science,* 89(10): 2981–2987.

Lehermeier, C., V. Wimmer, T. Albrecht, H.-J. Auinger, D. Gianola et al., 2013. "Sensitivity to prior specification in Bayesian genome-based prediction models." *Statistical Application in Genetics and Molecular Biology,* 12(3): 375–391.

Li, Y., and M. M. Parast, 2014. "BMP4 regulation of human trophoblast development." *International Journal of Developmental Biology,* 58(2-4): 239–246.

López de Maturana, E., N. Ibáñez-Escriche, Ó. González-Recio, G. Marenne, H. Mehrban et al., 2014. "Next generation modeling in GWAS: comparing different genetic architectures." *Human Genetics,* 133(10): 1235–1253.

Marras, G., A. Rossoni, H. Schwarzenbacher, S. Biffani, F. Biscarini et al., 2017. "zanardi: an open-source pipeline for multiple-species genomic analysis of SNP array data." *Animal Genetics,* 48(1): 121.

Matzuk, M. M., and D. J. Lamb, 2002. "Genetic dissection of mammalian fertility pathways." *Nature Cell Biology & Nature Medicine,* 8: S40–S49.

Merchán, M., R. Peiró, M. J. Argente, M. A. Santacreu, M. L. García et al., 2009. "Analysis of the oviductal glycoprotein 1 polymorphisms and their effects on components of litter size in rabbits." *Animal Genetics,* 40(5): 756–758.

Mocé, M. L., M. A. Santacreu, A. Climent, and A. Blasco, 2005. "Divergent selection for uterine capacity in rabbits III. Responses in uterine capacity and its components estimated with a cryopreserved control population." *Journal of Animal Science,* 83(10): 2308–2312.

Mocé, M. L., M. A. Santacreu, A. Climent, and A. Blasco, 2004. "The effect of divergent selection for uterine capacity on fetal and placental development at term in rabbits : Maternal and embryonic genetic effects." *Journal of Animal Science,* 82(4): 1046–1052.

Onteru, S. K., B. Fan, Z. Q. Du, D. J. Garrick, K. J. Stalder et al., 2012. "A whole-genome association study for pig reproductive traits." *Animal Genetics,* 43(1): 18–26.

Onteru, S., D. M. Gorbach, J. M. Young, D. J. Garrick, J. C. M. Dekkers et al., 2013. "Whole genome association studies of residual feed intake and related traits in the pig." *PLoS ONE,* 8(6): e61756.

Peiró, R., M. Merchán, M. A. Santacreu, M. J. Argente, M. L. García et al., 2008. "Identification of single-nucleotide polymorphism in the progesterone receptor gene and its association with reproductive traits in rabbits." *Genetics,* 180(3): 1699–1705.

Piles, M., M. L. García, O. Rafel, J. Ramon, and M. Baselga, 2006. "Genetics of litter size in three maternal lines of rabbits: Repeatability versus multiple-trait models." *Journal of Animal Science,* 84(9): 2309–2315.

Quinton, V. M., J. W. Wilton, J. A. Robinson, and P. K. Mathur, 2006. "Economic weights for sow productivity traits in nucleus pig populations." *Livestock Science,* 99(1): 69–77.

Ragab, M., J. P. Sánchez, C. Mínguez, J. S. Vicente, and M. Baselga, 2014. "Litter size components in a full diallel cross of four maternal lines of rabbits." *Journal of Animal Science,* 92(8): 3231–3236.

Ros-Freixedes, R., S. Gol, R. N. Pena, M. Tor, N. Ibáñez-Escriche et al., 2016. "Genome-wide association study singles out SCD and LEPR as the two main loci influencing intramuscular fat content and fatty acid composition in Duroc pigs." *PLoS ONE,* 11(3): e0152496.

Rosenbloom, K. R., J. Armstrong, G. P. Barber, J. Casper, H. Clawson et al., 2015. "The UCSC Genome Browser database: 2015 update." *Nucleic Acids Research,* 43(D1): D670–D681.

Rossitto, M., S. Ujjan, F. Poulat, and B. Boizet-Bonhoure, 2014. "Multiple roles of the prostaglandin D2 signaling pathway in reproduction." *Reproduction,* 149(1): R49–R58.

Santacreu, M. A., M. L. Mocé, A. Climent, and A. Blasco, 2005. "Divergent selection for uterine capacity in rabbits II. Correlated Response in Litter Size and its components estimated with a cryopreserved control population." *Journal of Animal Science,* 83(10): 2303–2307.

Schneider, J. F., D. J. Nonneman, R. T. Wiedmann, J. L. Vallet, and G. A. Rohrer, 2014. "Genomewide association and identification of candidate genes for ovulation rate in swine." *Journal of Animal Science,* 92(9): 3792–3803.

Schneider, J. F., L. A. Rempel, W. M. Snelling, R. T. Wiedmann, D. J. Nonneman et al., 2012. "Genome-wide association study of swine farrowing traits. Part II: Bayesian analysis of marker data." *Journal of Animal Science,* 90(10): 3360–3367.

Shin, J. H., S. Blay, B. McNeney, and J. Graham, 2006. "LDheatmap : An R Function for Graphical Display of Pairwise Linkage Disequilibria between Single Nucleotide Polymorphisms." *Journal of Statistical Software,* 16(3): 1-10.

Stelzer, G., N. Rosen, I. Plaschkes, S. Zimmerman, M. Twik et al., 2016. "The GeneCards suite: from gene data mining to disease genome sequence analyses." *Current Protocols in Bioinformatics*, 54(1): 1.30.30-1.30.33

Stephens, M., and D. J. Balding, 2009. "Bayesian statistical methods for genetic association studies." *Nature Reviews Genetics,* 10(10): 681–690.

Toosi, A., R. L. Fernando, and J. C. M. Dekkers, 2018. "Genome-wide mapping of quantitative trait loci in admixed populations using mixed linear model

and Bayesian multiple regression analysis." *Genetics Selection Evolution,* 50(1): 32.

Yathish, H. M., S. Kumar, P. P. Dubey, R. P. Modi, R. Chaudhary et al., 2017. "Profiling of sperm gene transcripts in crossbred (Bos taurus x Bos indicus) bulls." *Animal Reproduction Science,* 177: 25–34.

Zeisel, S. H., 2011. "The supply of choline is important for fetal progenitor cells." *Seminars in Cell & Developmental Biology*, 22(6): 624–628.

Ziadi, C., M. L. Mocé, P. Laborda, A. Blasco, and M. A. Santacreu, 2013. "Genetic selection for ovulation rate and litter size in rabbits: Estimation of genetic parameters and direct and correlated responses." *Journal of Animal Science,* 91(7): 3113–3120.

3.8. Additional Files

Additional file 1: Table S1. Annotated genes in the genomic regions associated with litter size traits.

Gene Name	Chromosome	Rabbit Gene Description
5S_rRNA	9, 11, 17	5S ribosomal RNA
7SK	17	7SK RNA
ABHD12B	17	abhydrolase domain containing 12B
ANKH	11	ANKH inorganic pyrophosphate transport regulator
ANKRD33B	11	ankyrin repeat domain 33B
ATL1	17	atlastin GTPase 1
ATP5S	17	ATP synthase, H+ transporting, mitochondrial Fo complex, subunit s (factor B)
BMP4	17	bone morphogenetic protein 4
C4orf3	9	chromosome 4 open reading frame 3
CALCR	10	calcitonin receptor
CCNL1	14	cyclin L1
CCT5	11	chaperonin containing TCP1, subunit 5 (epsilon)
CDK14	10	cyclin-dependent kinase 14
CDK6	10	cyclin-dependent kinase 6
CDKL1	17	cyclin-dependent kinase-like 1 (CDC2-related kinase)
CDKN3	17	cyclin-dependent kinase inhibitor 3
CFAP69	10	cilia and flagella associated protein 69
CGRRF1	17	cell growth regulator with ring finger domain 1
CLDN12	10	claudin-12
CMBL	11	carboxymethylenebutenolidase homolog (Pseudomonas)
CNTN3	9	contactin 3 (plasmacytoma associated)
CTNND2	11	catenin (cadherin-associated protein), delta 2
DDHD1	17	DDHD domain containing 1
DNAH5	11	dynein, axonemal, heavy chain 5
ERO1A	17	endoplasmic reticulum oxidoreductase alpha
FAM105A	11	family with sequence similarity 105, member A
FAM173B	11	family with sequence similarity 173, member B
FERMT2	17	fermitin family member 2
FRMD6	17	FERM domain containing 6
GINS2	10	GINS complex subunit 2 (Psf2 homolog)
GMFB	17	glia maturation factor, beta
GNG2	17	guanine nucleotide binding protein (G protein), gamma 2
GPR137C	17	G protein-coupled receptor 137C
GTPBP10	10	GTP-binding protein 10 (putative)
HEPACAM2	10	HEPACAM family member 2

Gene Name	Chromosome	Rabbit Gene Description
L2HGDH	17	L-2-hydroxyglutarate dehydrogenase
LEKR1	14	Leucine, Glutamate and Lysine Rich 1
MAP4K5	17	mitogen-activated protein kinase 5
MARCH6	11	membrane-associated ring finger (C3HC4) 6, E3 ubiquitin protein ligase
NID2	17	nidogen 2 (osteonidogen)
NIN	17	ninein (GSK3B interacting protein)
OTULIN	11	OTU deubiquitinase with linear linkage specificity
PNRC2	17	proline-rich nuclear receptor coactivator 2
PSMC6	17	proteasome (prosome, macropain) 26S subunit, ATPase, 6
PTGDR	17	prostaglandin D2 receptor (DP)
PTGER2	17	prostaglandin E receptor 2 (subtype EP2), 53kDa
PTX3	14	pentraxin 3, long
PYGL	17	phosphorylase, glycogen, liver
ROPN1L	11	rhophilin associated tail protein 1-like
SAMD4A	17	sterile alpha motif domain containing 4A
SAMD9	10	sterile alpha motif domain containing 9
SAMD9L	10	sterile alpha motif domain containing 9-like
SAV1	17	salvador family WW domain containing protein 1
SCARNA23	17	Small Cajal body specific RNA 23
SNORA65	14	Small nucleolar RNA SNORA65
SNORD90	14	Small nucleolar RNA SNORD90
snoU13	17	Small nucleolar RNA U13
SOS2	17	son of sevenless homolog 2 (Drosophila)
STEAP1	10	metalloreductase STEAP1
STEAP2	10	Oryctolagus cuniculus STEAP family member 2, metalloreductase (STEAP2), mRNA
STYX	17	serine/threonine/tyrosine interacting protein
TIPARP	14	TCDD-inducible poly(ADP-ribose) polymerase
TMX1	17	thioredoxin-related transmembrane protein 1
TRIM9	17	tripartite motif containing 9
TRIO	11	trio Rho guanine nucleotide exchange factor
TXNDC16	17	thioredoxin domain containing 16
U2	14	U2 spliceosomal RNA
U4	17	U4 spliceosomal RNA
U6	9, 10, 11	U6 spliceosomal RNA
VCPKMT	17	valosin containing protein lysine (K) methyltransferase
VEPH1	14	ventricular zone expressed PH domain-containing 1
VPS50	10	VPS50 EARP/GARPII complex subunit

CHAPTER FOUR

4. GENOMIC REGIONS INFLUENCING INTRAMUSCULAR FAT IN DIVERGENTLY SELECTED RABBIT LINES

Authors: Bolívar S. Sosa-Madrid[1], Pilar Hernández[1], Agustín Blasco[1], Chris S. Haley[2,3], Luca Fontanesi[4], María A. Santacreu[1], Ramona N. Pena[5], Pau Navarro[2], and Noelia Ibáñez-Escriche[1].

Institutional affiliations

[1] Institute for Animal Science and Technology, Universitat Politècnica de València, 46022 Valencia, Spain.

[2] MRC Human Genetics Unit, MRC Institute of Genetics and Molecular Medicine, University of Edinburgh, Crewe Road, Edinburgh EH4 2XU, United Kingdom.

[3] Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Midlothian EH25 9RG, United Kingdom.

[4] Department of Agricultural and Food Sciences, Division of Animal Sciences, University of Bologna, 40127 Bologna, Italy.

[5] Departament de Ciència Animal, Universitat de Lleida–Agrotecnio Centre, E-25198 Lleida, Catalonia, Spain.

The content of this chapter has been published in *Animal Genetics*.

DOI: https://doi.org/10.1111/age.12873

4.1. Abstract

Intramuscular fat (IMF) is one of the main meat quality traits for breeding programmes in livestock species. The main objective of this study was to identify genomic regions associated with IMF content comparing two rabbit populations divergently selected for this trait, and to generate a list of putative candidate genes. Animals were genotyped using the Affymetrix Axiom OrcunSNP Array (200k). After quality control, the data involved 477 animals and 93,540 SNPs. Two methods were used in this research: single marker regressions with the data adjusted by genomic relatedness, and a Bayesian multiple marker regression. Associated genomic regions were located on the rabbit chromosomes (OCU) OCU1, OCU8 and OCU13. The highest value for the percentage of the genomic variance explained by a genomic region was found in two consecutive genomic windows on OCU8 (7.34%). Genes in the associated regions of OCU1 and OCU8 presented biological functions related to the control of adipose cell function, lipid binding, transportation and localisation (*APOLD1*, *PLBD1*, *PDE6H*, *GPRC5D* and *GPRC5A*) and lipid metabolic processes (*MTMR2*). The *EWSR1* gene, underlying the OCU13 region, is linked to the development of brown adipocytes. The findings suggest that there is a large component of polygenic effect behind the differences in IMF content in these two lines, as the variance explained by most of the windows was low. The genomic regions of OCU1, OCU8 and OCU13 revealed novel candidate genes. Further studies would be needed to validate the associations and explore their possible application in selection programmes.

Keyword: *divergent selection, intramuscular fat, genome-wide association study, meat quality, rabbits.*

4.2. Background

Intramuscular fat (IMF) contributes to improve organoleptic properties and sensory attributes of the meat, as demanded by consumers (Hocquette et al., 2010). Hence, a large number of studies have investigated the genetic factors controlling IMF content in meat and their implications for several species, *e.g.* in beef cattle (Sapp et al., 2002; Garrick, 2011; Ochsner et al., 2017), swine (McLaren & Schultz, 1992; Gao et al., 2007), sheep (Hopkins et al., 2011; Mortimer et al., 2014) and goats (Peña et al., 2011). Following these studies,

IMF has emerged as one of the most important meat quality parameters to be improved and in a few cases it has been included in breeding programmes (Gotoh et al., 2018; Pannier et al., 2018).

Moderate-to-high heritability and large variability have been reported for livestock IMF traits, which argue for a good potential for improving meat quality through genetic selection. IMF heritability is around 0.53 in swine (Ros-Freixedes et al., 2016), 0.38 in cattle (Mateescu et al., 2015), 0.48 in sheep (Mortimer et al., 2014) and 0.54 in rabbit (Martínez-Álvaro et al., 2016). Important limitations to IMF selection are IMF being recorded mainly at slaughter and the phenotyping process is costly. In this context, genetic marker selection based on quantitative trait locus (QTL) with high or moderate effect size could overcome some of these limitations.

At genomic level, studies carried out in beef cattle suggest that IMF could be influenced by a large number of genes (Strucken et al., 2017). Nevertheless, studies in Japanese Black cattle have reported genomic markers with large effects on IMF or marbling score around the *SCD, FASN, AKIRIN2, EDG1* and *RPL27A* genes (Gotoh et al., 2014; Sukegawa et al., 2014). Genomic markers on the genes *SCD* and *FASN* have been incorporated into a breeding programme of this breed to select elite sires (Gotoh et al., 2018). In swine, similarly to beef cattle, the results of experiments associating genetic markers with IMF are hardly conclusive with regards to the magnitude and importance of discovered associations (Pena et al., 2016). However, traits correlated to IMF such as fatty acid profiles have shown a noteworthy QTL on chromosome 14 in a Duroc commercial line (Uemoto et al., 2012; Ros-Freixedes et al., 2016). So far, IMF appears as a troublesome trait for mapping studies in livestock species, owing to either the lack of validation in the results or the insufficient power to detect genetic causal variants. Thus, genomic studies to understand the genetic control of IMF are still needed.

The rabbit has been shown to be an excellent animal model for other livestock species (Miller et al., 2014). Further, the recent availability of a high-density SNP (Single Nucleotide Polymorphism) array has facilitated the performance of genomic studies. At the Universitat Politècnica de València, a successful divergent selection experiment for IMF has been carried out (Martínez-Álvaro et

al., 2016). The developed rabbit lines were kept in the same environment and selection criteria only differ for the IMF selection objective. Selection could have modified SNP frequencies in opposite directions, leading to intermediate allelic frequencies when both lines are jointly considered. This could increase the detection power of associated loci in a genome-wide association study (GWAS) based on this experimental design.

The aim of this study was to carry out GWASs using these divergently selected rabbit lines to identify genomic regions associated with IMF and generate a list of putative candidate genes affecting this trait. Two different methods (single marker regression, SMR, and Bayesian multiple marker regression, BMMR) were applied to confirm the identified relevant genomic regions.

4.3. Material and Methods

4.3.1. Ethical statement

All experimental procedures were approved by the Ethical Committee of the Universitat Politècnica de València, according to Council Directives 98/58/EC (European Economic Community, 1998).

4.3.2. Animals and phenotypes

The animals of this study came from two rabbit lines divergently selected for IMF during nine generations at the Universitat Politècnica de València. The base population was composed of 83 does and 13 males from a synthetic rabbit line (Zomeño et al., 2013). The selection criterion was IMF content collected in two full-siblings of the first parity. The selection of the males was within sire family, avoiding mating between cousins to control inbreeding. At the 9th generation, high-IMF line consisted of 55 does and 10 males, and the low-IMF line consisted of 61 does and 10 males. Over all animals, the mean was 1.09 grams of IMF per 100g of *Longissimus thoracis et lumborum* (LTH) muscle, after adjusting data by systematic effects (parity order, line, month-season and sex) and a common litter random effect. The high-IMF line had a mean of 1.27 g/100g of LTH with 0.21 standard deviations, and the low-IMF line had a mean of 0.83 g/100g of LTH with 0.07 standard deviations. Details about the IMF divergent selection experiment can be found in Martínez-Álvaro *et al.* (2016). The selection response

was around 3.1 standard deviations at 9th generation, calculated as the difference between lines. The phenotypic difference between lines was 41% of the mean of the base population.

The rabbits were brought up jointly from 33 days at weaning until slaughter under the same handling and feeding conditions. After 9 weeks from birth, the rabbits were slaughtered following a fasting period of 4 hours. Carcasses were chilled 24 hours at 2.5 °C after slaughter and dissected to obtain a sample of the left LTH muscle for each animal. These samples were minced, frozen, lyophilised and milled. The IMF data were obtained using near-infrared spectroscopy (model 5000; FOSS NIRSystems Inc., Hilleroed, Denmark) (Zomeño et al., 2013; Martínez-Álvaro et al., 2016). In the last generation, 729 samples of the left LTH muscle of each animal were collected and IMF measured to compute the IMF selection response, and 480 rabbits were chosen from groups of an average size of four siblings per doe (dam) for the GWAS.

4.3.3. Genotypes and quality control

Obliquus abdominis muscle specimens (~50 grams), obtained after slaughter of the animals, were used for DNA extraction using a standard protocol (Green et al., 2012). A total of 480 individuals were genotyped using the Affymetrix Axiom OrcunSNP Array (Affymetrix, Inc. Santa Clara, CA, USA) at the "Centro Nacional de Genotipado" (CeGen) – Universidad de Santiago de Compostela. The SNP array contains 199,692 genetic molecular markers. The quality control was performed by Axiom Analysis Suite v. 3.0.1.4 and ZANARDI (Marras et al., 2017). The SNPs with a call rate of at least 0.95, MAF of at least 0.03 and a known autosomal chromosome position according to OryCun2.0 assembly (Carneiro et al., 2014) were used in the analyses. Furthermore, animals missing more than 3% of marker genotypes, or failing a Mendelian inheritance test, were excluded. The remaining missing genotypes were imputed by the software BEAGLE v4.0 (Browning & Browning, 2016). The SNPs with an imputation quality score $R^2 > 0.75$ were included. After filtering, the data included 477 animals (240 from the high-IMF line and 237 from the low-IMF line) and 93,540 SNPs. In addition, the SNP density was described in this research because the rabbit SNP array is new (Blasco & Pena, 2018).

4.3.4. Genome-wide association study

Prior to performing the GWAS, we performed a multidimensional scaling analysis to evaluate the population structure in our genomic data. The method treats the distances as Euclidean distances and preserves the original distance metric, between points, as well as possible (Borg & Groenen, 2005). The command cmdscale() from R package stats was used to implement this analysis (R Core Team, 2018).

Two methods were employed in this study: a frequentist and a Bayesian. Both methods included the mean and the systematic effects in the model: month-season (five levels), sex (two levels), order-parity (three levels), and line (two levels). The inclusion of a common litter random effect in the model was evaluated due to the importance of this effect in previous studies of IMF in rabbits (Martínez-Álvaro et al., 2016). Inclusion of this effect did not affect GWAS results (not shown), hence for simplicity we excluded this effect in the GWAS.

Single marker regression (SMR) with the data adjusted by genomic relatedness: The analysis was implemented using a family-based score test for association (FASTA). The SNP effects were evaluated with FASTA based on a polygenic-lineal mixed model that included the genomic kinship matrix to explain relatedness in the sampled population (Chen & Abecasis, 2007). The model equation was:

$$\boldsymbol{y} = \mathbf{1}\boldsymbol{\mu} + \boldsymbol{Xb} + \beta\boldsymbol{g} + \boldsymbol{Zu} + \boldsymbol{e}$$

where, $\boldsymbol{y}$ is the vector of IMF phenotypes, $\mathbf{1}$ is a vector of ones, $\boldsymbol{\mu}$ is the trait mean, $\boldsymbol{X}$ is the design matrix for the systematic effects, $\boldsymbol{b}$ is the vector of systematic effects, β is the substitution effect for a particular SNP, $\boldsymbol{g}$ is the vector of genotypes for each SNP denoted as the number of reference alleles for a particular SNP (0, 1 or 2), $\boldsymbol{Z}$ is the design matrix for random polygenetic effects, $\boldsymbol{u}$ is the vector of random polygenic effects with a normal distribution $N(0, \boldsymbol{G} \cdot \sigma_u^2)$, $\boldsymbol{e}$ is the vector of random residual effects with a normal distribution $N(0, \boldsymbol{I} \cdot \sigma_e^2)$. σ_u^2 is the genomic variance and $\boldsymbol{G}$ is the genomic kinship matrix computed using the genomic data by the method of Astle &

Balding (2009). The identity matrix was denoted as $\boldsymbol{I}$ and σ_e^2 is the residual variance. The implementation of the association analysis was performed using R software package GenABEL (Aulchenko et al., 2007). Furthermore, we utilized a genomic control method to avoid inflation in the statistic test. We calculated the lambda parameter that indicates the excess of false positives in the results. When its application is needed, the regression factor lambda (λ) corrects the observed p-values leading to new p-values for every assessed SNP (Aulchenko et al., 2007). In this research, we used two thresholds; a LD-adjusted Bonferroni (8.12×10^{-6}) calculated for 10-Mb LD blocks according to LD analysis implemented in PLINK, and also, a suggestive threshold of 1×10^{-4} due to the high relatedness of the samples (Lander & Kruglyak, 1995; Sahana et al., 2011; Do et al., 2018). As Bonferroni is a very conservative method, we also implemented the suggestive threshold because it is less stringent as the samples from animals with high relatedness would have genomic segments of LD larger than in human (Wang et al., 2016b; Schmid & Bennewitz, 2017). Therefore, the number of independent sites could be overestimated causing false-negative results if SNP density is not large enough to adjust Bonferroni by LD (Spencer et al., 2009; Do et al., 2014).

Bayesian multiple marker regression (BMMR): This method is more robust to population structure than SMR approaches (Toosi *et al.*, 2018). However, the line effect would correct for potential biases that might be derived from the family data structures in the investigated rabbit populations. Thus, the line effect was remained in the BMMR model. The parameters were estimated with the following Bayes B model (Cesar et al., 2014; Ros-Freixedes et al., 2016):

$$\boldsymbol{y} = \boldsymbol{1\mu} + \boldsymbol{Xb} + \sum_{j=1}^{k} \boldsymbol{z_j}\ \alpha_j\ \delta_j + \boldsymbol{e}$$

where $\boldsymbol{y}$, , $\boldsymbol{X}$, $\boldsymbol{b}$ and $\boldsymbol{e}$ are the same parameters that in frequentist method shown above. $\boldsymbol{z_j}$ is the vector including the genotypic covariate for each SNP or locus j (0, 1 or 2); α_j is the random substitution effect for SNP j; δ_j is the random 0/1 variable that represents presence (δ_j = 1 with probability 1-π) or

absence (δ_j = 0 with probability π) of SNPs in the model for a given iteration. The value of π is defined as the proportion of SNPs with zero effects in the model. The value of π in our study was 0.9988, which means that between 100 and 200 SNP markers have nonzero effects for every iteration. The parameters of the model were estimated with marginal posterior distributions using Markov chain Monte Carlo (MCMC). After some exploratory analysis, a total of 825,000 iterations were performed, with a burn-in period of 225,000 iterations. Only one sample every 60 iterations was saved to avoid the high correlation between consecutive samples. The GenSel® v. 4.90 software (Garrick & Fernando, 2013) was used for the GWAS analysis. The relevance of the association was assessed using two criteria, the Bayes factor (Stephens & Balding, 2009; Ros-Freixedes et al., 2016), and the percentage of the genomic variance explained for non-overlapping genomic windows of 1 megabase, calculated by marginal posterior density. The genomic windows were defined for each chromosome and according the OryCun2.0 rabbit genome assembly (Carneiro et al., 2014). In our study, 1,999 genomic windows were defined. Those windows accounting for at least 1.0% of the total genomic variance were considerate as important to continue with the subsequent analysis (Cesar et al., 2014). This threshold was 20 times greater than the average genomic variance explained by a window (0.05%). We also considered the consecutive windows that explained at least 0.5% of genomic variance having a strong linkage disequilibrium between them (Ros-Freixedes et al., 2016) as SNPs associated with a causal variant can be located between consecutive windows and the estimated effect of association could be divided among these windows, hindering the detection of a genomic region (Beissinger et al., 2015).

In this study, we integrated the results from both frequentist and Bayesian methods to define the relevance of associations. This was established by the following procedure: first, we drew all genomic windows that overcame the condition expressed in the above paragraph. Then, the genomic windows harbouring SNP above or around 20 Bayes factor (Kass & Raftery, 1995) were extracted and considered as relevant genomic windows. These SNPs reaching at least one of thresholds, either suggestive or Bayes factor thresholds, were denoted as relevant polymorphisms. Finally, the genomic regions having relevant associations were chosen for functional gene analysis.

In addition, the three main important polymorphisms within relevant genomic regions were tested according to genotypes using contrasts by frequentist statistic. This test was carried out within IMF line in order to evaluate the statistical differences amongst genotypes of SNPs. To do that, a general linear model was implemented using R software (R Core Team, 2018).

4.3.5. Linkage disequilibrium and functional gene analysis

To evaluate the number of independent sites across the rabbit genome, a computation of LD for blocks was performed. The PLINK software was utilised to identify LD blocks (Purcell et al., 2007). The number of independent sites was calculated every 0.5, 1, 5, 10 and 20 Mb (genomic distance) across the whole rabbit genome. The LD-adjusted Bonferroni threshold used in this study was calculated using the number of independent sites for 10 Mb as the number of independent sites barely changed between 10 to 20 Mb. LD blocks were examined in the associated genomic regions through the Haploview software (Barrett et al., 2005). In order to visualise the genes into the relevant genomic regions (+/- 500 kb of associated SNP), we initially used the program UCSC Genome Browser (https://genome.ucsc.edu/cgi-bin/hgGateway). The gene annotations were determined using Ensembl Genes 96 Database in software BioMart (Aken et al., 2016). The functional enrichment and metabolic pathways analysis were finally performed with "Database for Annotation, Visualization and Integrated Discovery" (DAVID) v 6.8 (Jiao et al., 2012) and "Enrichr" (Kuleshov et al., 2016). The computation for the functional analyses was carried out using the parameters recommended by the authors. In addition, the search of annotated functions for each gene was performed individually using the database of all annotated functions from Ensembl and DAVID.

4.4. Results

4.4.1. Genomic data

A total of 93,540 autosomal SNPs with known chromosomal positions were retained after filtering for minor allele frequency (MAF) and call rate (see details in Materials and Methods). The number of retained SNPs on each of the 21 rabbit autosomes is shown in Table 1. The average physical distance between these SNPs was 22.61 kb. The average SNP number within 1-Mb windows was

46. One extended genomic region on OCU14 (54-65 Mb) did not contain any SNPs.

Table 1. Allocation of SNPs after quality control and average distance amongst contiguous SNPs on every chromosome.

OCU	Number of SNPs	Percentage of SNPs in OCU*	Average Distance (kb)	Chromosome Size (Mb)
1	9288	63%	20.98	194.85
2	7856	58%	22.19	174.33
3	7006	59%	22.22	155.69
4	3895	58%	23.47	91.39
5	1721	67%	21.84	37.99
6	1222	63%	22.48	27.50
7	7626	57%	22.78	176.68
8	5075	57%	22.03	111.80
9	5136	57%	22.58	116.25
10	2318	61%	19.38	48.00
11	3827	56%	22.81	87.55
12	7116	60%	21.83	155.35
13	5945	56%	24.11	143.36
14	5687	45%	28.81	163.90
15	4657	55%	22.71	109.05
16	3962	62%	21.32	84.48
17	3836	59%	21.94	85.01
18	3102	64%	21.45	69.80
19	2574	64%	21.00	57.28
20	1224	51%	24.66	33.19
21	467	55%	26.56	15.58
Total	93540	47%		

* The proportion of SNPs after quality control divided by number total of SNPs into OCU (rabbit chromosome) from the rabbit SNP array.

4.4.2. GWAS for intramuscular fat

Figure 1 reports a multidimensional scaling (MDS) plot obtained using the genotyped SNPs on the rabbits of the two divergent IMF lines. A strong structure separating the high- and low-IMF lines is clearly evident. Therefore, a line effect was included in the models. In addition, a polygenic effect was also included in

the SMR to adjust this model owing to the plausible effects derived from family-data structures, considering a genomic kinship matrix. After this correction, the calculated lambda parameter was 1.065, indicating that the correction of bias derived from the population structure was not enough. Hence, we also implemented the correction by the lambda parameter in the SMR analysis. Note that the first and second components of MDS accounted for 29.26% and 3.26% of genomic variance, respectively (Figure 1).

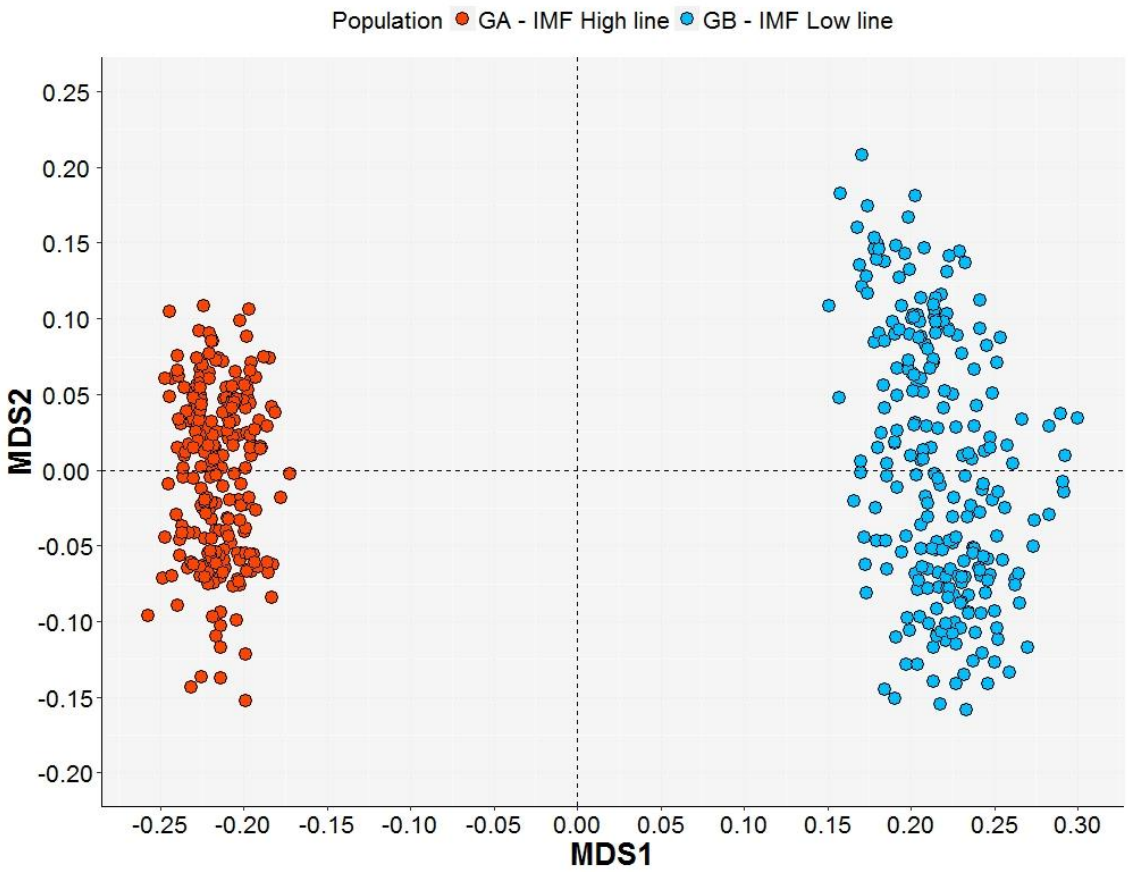


Figure 1. Multidimensional scaling plot of genomic data. The first component (MDS1) explained 29.26 % of the genomic variance and the second component (MDS2) explained 3.26 % of the genomic variance.

Two methods were used in this research: SMR with the data adjusted by genomic relatedness, and a BMMR (Bayes B method). We employed the term of "relevant" in order to denote those SNPs and genomic windows that we considered as true positive associations. In this research, we understand the GWAS as an exploratory analysis, which works as a mechanism for deriving promising genomic regions associated with IMF, and retrieving annotated rabbit genes. Table 2 shows the SNPs and genomic windows associated with IMF

according to the procedure for defining the relevant associations (see details in Materials and Methods). For both methods, the associated SNPs and genomic windows were located on OCU8 and OCU13. The two genomic windows on OCU13 (2Mb), containing ten relevant SNPs, accounted together for 1.30% of the total genomic variance. On OCU8, ten relevant polymorphisms showed the lowest p-values for the SMR method, and had high Bayes factors for the BMMR method (Figure 2). The two genomic windows containing these relevant polymorphisms accounted for 7.34% of the genomic variance. In addition, a genomic window on OCU1 was found associated with IMF by BMMR, explaining 2.03% of the genomic variance. The associated SNPs in this latter genomic window presented values close to the Bayes factor threshold, but these SNPs were distant from the p-value (suggestive) threshold for SMR method.

Regarding the LD analysis, we found that in our data the rabbit genome could be divided in 2,338 LD blocks and 6,158 independent sites, with the longest LD blocks having a maximum length of 10 Mb. The associated SNPs on OCU13 and on OCU8 displayed a high LD within chromosomal region (Figure 3). The associated genomic region on OCU13 (window 1380 and 1381) holds two LD blocks. The second LD block (of 1506 kb) included almost the two-associated windows (Additional file 1: Figure S1). The associated genomic region on OCU8 (window 841 and 842) presented just one block of 1945 kb, containing both windows (Additional file 2: Figure S2).

Table 2. Relevant polymorphisms (SNPs) and genomic windows associated with intramuscular fat.

SNP name	OCU	Position (bp)	P-value	Bayes Factor	Window		MAF
					Name	% Variance	
Affx-151793092	1	121151928	$1.10x10^{-3}$	15.95	118	2.03	0.24
Affx-151803947	1	121280205	$1.10x10^{-3}$	19.59			0.24
Affx-151888965	1	121308004	$1.10x10^{-3}$	16.03			0.25
Affx-151956200	8	14893810	$3.51x10^{-4}$	19.51	831	1.21	0.31
Affx-151962168	8	14913105	$3.51x10^{-4}$	24.86			0.32
Affx-151945237	8	14939285	$3.51x10^{-4}$	28.58			0.31
Affx-151973204	8	14972879	$1.83x10^{-4}$	18.38			0.31
Affx-151800097	8	25087426	$2.13x10^{-6}$	21.78	841	6.20	0.16
Affx-151900210	8	25227502	$3.33x10^{-6}$	44.73			0.16
Affx-151917268	8	25262821	$2.13x10^{-6}$	20.64			0.16
Affx-151813008	8	25268392	$2.13x10^{-6}$	22.57			0.16
Affx-151795704	8	25467177	$3.12x10^{-6}$	20.99			0.16
Affx-151972842	8	25643667	$2.06x10^{-6}$	24.15			0.16
Affx-151964185	8	25732369	$2.06x10^{-6}$	21.78			0.16
Affx-152000638	8	25751303	$2.06x10^{-6}$	21.17			0.16
Affx-151808634	8	25863739	$2.06x10^{-6}$	23.27			0.16
Affx-151853378	8	25874631	$2.12x10^{-6}$	21.25			0.16

SNP name	OCU	Position (bp)	P-value	Bayes Factor	Window		MAF
					Name	% Variance	
Affx-151824236	8	26115758	2.66×10^{-3}	21.87	842	1.14	0.16
Affx-151867012	13	84307591	7.14×10^{-5}	11.73	1380	0.79	0.09
Affx-151824373	13	84431723	7.14×10^{-5}	10.62			0.09
Affx-151874466	13	84447172	8.45×10^{-5}	11.90			0.09
Affx-151883028	13	84453332	7.14×10^{-5}	11.73			0.09
Affx-151801561	13	84537466	7.14×10^{-5}	25.39			0.09
Affx-151841215	13	84723427	2.20×10^{-5}	25.39			0.09
Affx-151846540	13	84738337	2.20×10^{-5}	26.98			0.09
Affx-151790364	13	84751504	2.23×10^{-5}	25.30			0.09
Affx-151939801	13	85316544	3.40×10^{-4}	43.81	1381	0.51	0.08
Affx-151937959	13	85333053	6.31×10^{-6}	15.69			0.09

OCU = rabbit chromosome, bp = base pair, % Variance = percentage of genomic variance explained by window.

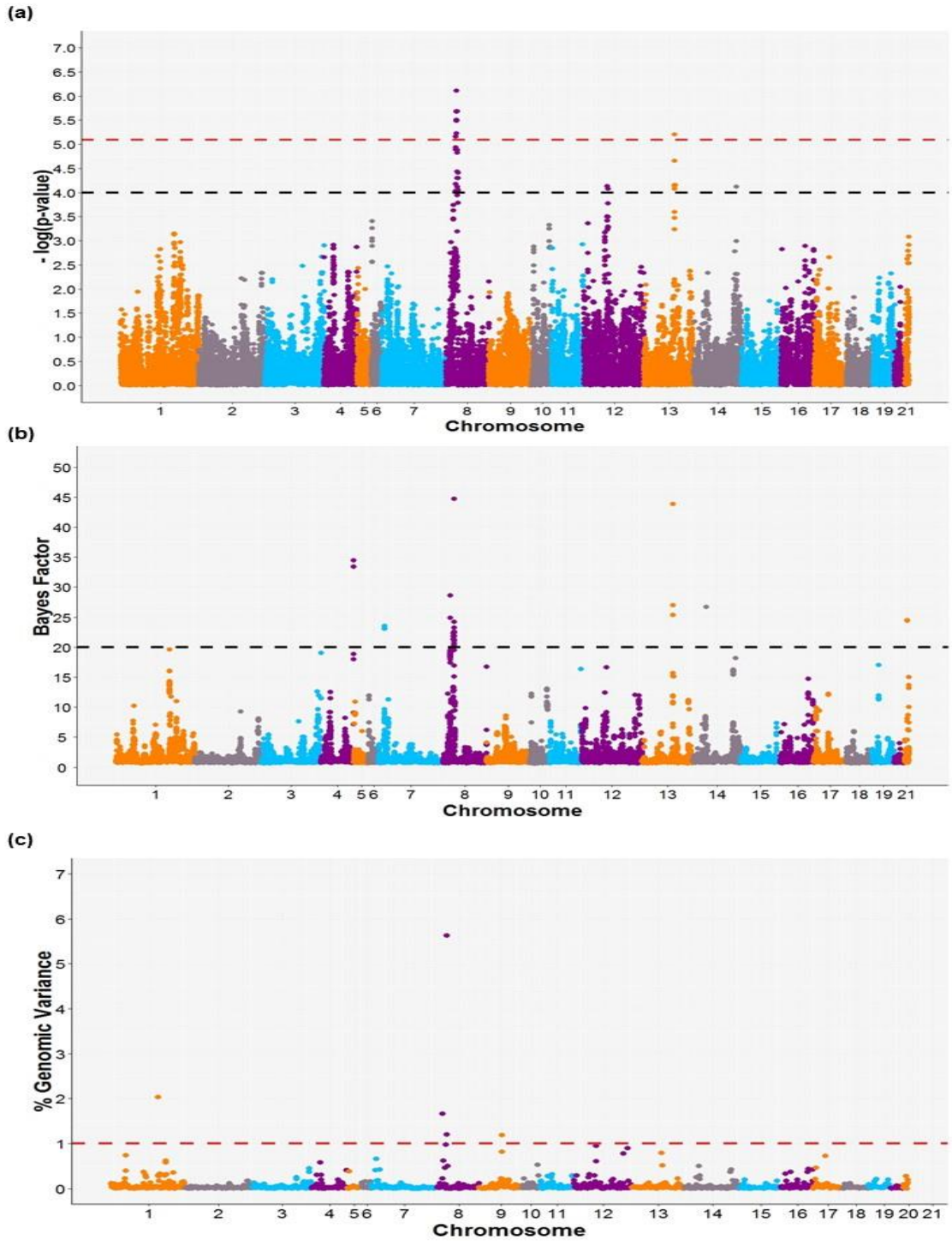


Figure 2. Manhattan plot for each model. **(a)** Single marker regression fitted by genomic relationship. The –log (p-value) thresholds are 5.09 (LD-Bonferroni – red dashed line) and 4.0 (suggestive – black dashed line) **(b)** The Bayes Factor (BF) for each SNP for the Bayesian multi-marker regression model. The black dashed line indicates the BF threshold of 20 **(c)** The percentage of genomic variance explained by each non-overlapping one megabase window for the Bayesian Multi marker regression model (threshold of 1% - red dashed line).

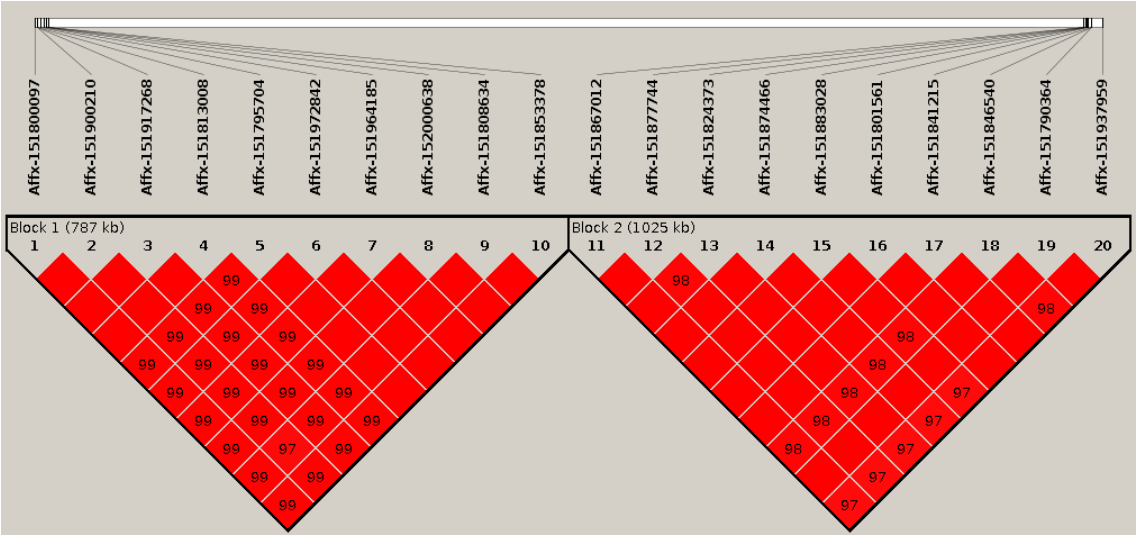


Figure 3. Linkage disequilibrium blocks from main relevant associated polymorphisms. Block 1 includes SNPs (1-10) on chromosome 8 in 24.59 - 26.95 Mb, and block 2 includes SNPs (11-20) on chromosome 13 in 83.81 - 86.00 Mb.

After the previous analysis (GWAS and LD), four relevant genomic regions were used to continue the searching of putative candidate genes based on the functional annotation analysis (Table 3). In these regions, we also tested the IMF differences between genotypes within line. Most of the SNPs tested presented statistical differences between one of the homozygous genotypes and the others genotypes within the high-IMF line. In the low-IMF line, except in region located 14.01-15.47 Mb in OCU8, these SNPs were not segregating (Additional file 3: Figure S3).

4.4.3. Functional annotation analysis and putative candidate genes.

The final objective of our study was to generate a list of putative candidate genes, in order to guide further research for investigating the genetic determination of IMF content. Overall, 46 genes are annotated to the four relevant genomic regions (Additional file 4: Tables S1).

Only three genes (two non-coding-protein genes and one protein-coding gene) mapped to the genomic region on OCU13 (Table 3). Among them stands out a novel annotated gene with Ensembl gene ID: *ENSOCUG00000027270* (84.56 Mb), which is linked to metal ion binding in rabbits. The genes located on the

genomic region on OCU8 were those showing a clearer relationship to lipid metabolism pathways. The "apolipoprotein L domain containing 1" gene *(APOLD1)* shows functions related to lipid binding, transportation, and localization. The "phospholipase B domain containing 1" (*PLBD1)* and "phosphodiesterase 6H" *(PDE6H)* genes show functions linked to hydrolase activity (phospholipases) and lipid metabolic processes. In human, several functional annotations, including sphingolipid signalling pathway, have been found for the "K-RAS proto-oncogene, GTPase" *(KRAS)* gene. Moreover, two members of the retinol-induced G protein-coupled protein receptors also stand out in OCU8: "G protein-coupled receptor class C group 5 member D" *(GPRC5D)* and "G protein-coupled receptor class C group 5 member A" *(GPRC5A)* (Table 3). On OCU1, the "myotubularin related protein 2" *(MTMR2)* gene displays biological functions linked to lipid metabolic processes. In addition to the biological and molecular functional annotations, a list of pathways that include these genes was generated from DAVID, the KEGG and Wiki pathways databases (Additional file 5: Table S2).

4.5. Discussion

Knowledge and understanding of control mechanisms of IMF content would be useful in the meat industry. Thus, GWAS was performed in order to identify genomic regions associated with IMF content in rabbits owing to the increasing importance of meat quality in livestock for consumers (Hocquette et al., 2010; Pena et al., 2016; Strucken et al., 2017).

Following GWAS detection power studies (Spencer et al., 2009; Visscher et al., 2017), the distribution of SNPs (after quality control) across the rabbit genome in our data was suitable for GWAS analysis in livestock, given the LD and SNP density (Fan et al., 2010; Zhang et al., 2012). For instance, LD blocks having distance of 98 kb show $r^2 = 0.5$ as a measure of LD within rabbit breeds (Carneiro et al., 2011). This would indicate that the 93,540 SNP having an average distance of 22.61 kb between SNPs can be useful for discovering true associations amongst SNPs and the causal variants of IMF.

Table 3. Summary of relevant genomic regions associated with intramuscular fat, and annotated rabbit genes.

CLUSTER	OCU	Position (bp)		Number of genes	Annotated rabbit gene
		Start	End		
1	1	120,651,928	121,986,803	9	*MAML2, MTMR2, CEP57, FAM76B, ENSOCUG00000025632*, SESN3, ENDOD1, KDM4D, CWC15*
2	8	14,014,437	15,472,879	9	*RASSF8, LMNTD1, RF00001, KRAS, ETFRF1, CASC1, LRMP, BCAT1, ENSOCUG00000021067**
3	8	24,587,426	26,948,204	25	*PDE6H, ARHGDIB, ERP27, MGP, ART4, SMCO3, ENSOCUG00000017177*, H2AFJ, HIST4H4, GUCY2C, PLBD1, ATF7IP, ENSOCUG00000017095*, ENSOCUG00000021765*, GRIN2B, RF00411, ENSOCUG00000021882*, EMP1, GSG1, FAM234B, HEBP1, GPRC5D, GPRC5A, DDX47, APOLD1*
4	13	83,807,591	85,998,108	3	*RF00026, ENSOCUG00000027270*, RF00001*

CLUSTER = denotes the genomic region, OCU = rabbit chromosome, bp = base pair.
*Novel genes are named with their Ensembl gene ID.

A challenge in GWAS analysis is the impact of confounding factors in the results. To avoid problems owing to population structure, we fit the genomic kinship matrix (Sul et al., 2018). The obtained λ value of 1.065 shows that this was almost enough to correct the population stratification effect. The purpose of implementing two methods was to corroborate the presence of associations between genomic windows or SNPs with IMF. The causal variants of moderate-to-high effect size can be detected by both methods in GWAS analyses when polymorphisms present high LD with these causal variants (López de Maturana et al., 2014). SNPs on OCU13 and OCU8 were found to be associated with IMF for both frequentist and Bayesian methods. However, the two associated windows on OCU13 (window 1380 and 1381) explained the low percentage of genomic variance (<1%). In addition, the LD block containing the most important SNPs on OCU13 covered a short physical distance and was uneven with regard to LD within this block (Additional file 1: Figure S1). This indicates that in this area of the genome a selective sweep process might not have been produced by divergent selection, since short-term selection increases LD and the expected length of the LD block that contains an important causal variant (Vitti et al., 2013). In addition, the reference alleles of these associated SNPs presented low allelic frequencies (close to zero) for the low-IMF line. The MAF value of the reference SNPs was also very low (<0.09) in both, low- and high-IMF lines (Table 2). All SNPs were fixed or near fixation in the low-IMF line, therefore the association of these SNPs with IMF was uncovered given their segregation in the high-IMF line. This could affect the association detection power even when the sample size is large (López de Maturana et al., 2014). For instance, if SNPs associated with the causal variants present a low MAF, the effects and association can be underestimated generating false-negative results.

In contrast, the associated region on OCU8 in 24.59 - 26.95 Mb explained a larger percentage of genomic variance between both associated windows (7.34%). Moreover, this region presented a strong and long linkage disequilibrium block between windows 841 and 842, which could imply a selective sweep process owing to divergent selection (additional file 2: Figure S2). The MAF values of the SNPs in this region were higher than on OCU13, reaching a maximum value of 0.16 (Table 2). Most of SNPs in OCU8 were fixed or near fixation in the low-IMF line. It seems that the causative variants and their surrounding SNPs would be at low frequency in the base population. This might

explain the fixation of SNPs in the low-IMF line and their segregation in the high-IMF line at the 9th generation. Therefore, this genomic region showed more evidence than the region on OCU13 for considering it as an important association driving the control mechanism of IMF. Finally, another potentially interesting genomic region was identified on OCU1. This region explained 2.03% of the IMF genomic variance, although the SNPs show –log (p-values) or Bayes factors below thresholds (Figure 2). This suggests that the association of these SNPs could be better captured by the method that considers the percentage of variance explained by the windows instead of evaluating each SNP individually. In addition, these SNPs present MAF values around 0.24 (0.48 for high-IMF line and close to zero for low-IMF line), which might suggest that the differences might be a consequence of the divergent selection process.

This is the first GWAS study for IMF in rabbits. Therefore, comparisons within rabbits are limited to previous candidate gene studies. In this sense, as Migdał et al. (2018), we did not find association between *FABP4* (OCU3) candidate gene and IMF. Our results are not in agreement with the studies for *FTO* (OCU5) (Zhang et al., 2013), *CAST* (OCU11) (Wang et al., 2016a) and *MYPN* (OCU18) (Wang et al., 2017) which found associations in two, one and one SNP within gene, respectively (p-values between 0.032 and 0.044). However, these associations should be taken with caution as the significance threshold was more liberal (p-value < 0.05, without applying correction for multiple testing) than in our GWAS study (p-value < $1x10^{-4}$). In agreement with GWAS studies for IMF in swine, our results suggest that there is a large polygenic component influencing the trait (Pena et al., 2016; Ros-Freixedes et al., 2016; Won et al., 2017). However, our results also showed important genomic regions associated with IMF. Especially in OCU8, a region of 2 Mb explains a notable percentage of the genomic variance (7.34%) in comparison with other GWAS studies for IMF (Cesar et al., 2014; Pena et al., 2016).

Several genes related to lipid metabolism (on OCU1, OCU8 and OCU13) were found in the associated regions. In OCU13, orthologues of a novel gene (Ensembl gene ID: *ENSOCUG00000027270*) have been reported in other species. In rabbits, there are not functional annotations related to lipid metabolism or intramuscular fat linked to this gene. However, in humans and mice this gene is known as *EWS* or *EWSR1*, and regulates the genetic expression of the

transcription factor "Y-Box Binding Protein 1" gene (*YBX1*). This transcription factors activates the expression of the gene *BMP7* ("Bone Morphogenetic protein 7") which in turn promotes the development of brown adipocytes (Wang & Seale, 2016)

The genomic regions on OCU8 contained the genes with the most important biological functions. Hence, the genes on this region can be considered as candidates for further research, given this window explains a large percentage of the IMF genomic variance (7.34%). In particular, *APOLD1, PLBD1, PDE6H* and *GPRC5A* were involved in functions of lipid transport, localisation and binding or in the control of adipose cell function. Two of these genes (*PLBD1* and, *PDE6H*) participated in the catabolism of phospholipids, which are the major component of cell membranes and have important implication in adipocyte hypertrophy (Chaves et al., 2011; Aloulou et al., 2012). As a result, *PLBD1* has been related to lipid catabolic processes, skeletal muscle weight and body mass index in mice (Lionikas et al., 2012; Nyima et al., 2016) and humans (Wahl et al., 2017). In addition, *KRAS* (OCU8) was associated with the control of fat deposition in chicken (Claire D'Andre et al., 2013) and was involved in sphingolipid signalling pathway. In humans, this gene was related to abnormal lipid metabolism in therapy of pancreatic cancer (Swierczynski et al., 2014). Another promising gene is *GPRC5A*, also known as *RAI3*, which is a key factor in repressing the differentiation of adipocytes in humans (Jin et al., 2017). This gene encodes for a member of the G-coupled proteins, a large family including over 800 receptors, amongst them, the olfactory receptors. *GRPC5A* belongs to a small subfamily of 4 members that are activated by retinol, the bioactive version of vitamin A. Although the role of *GPRC5A* is not well characterized at present, initial investigation report a link with lung cancer, and also as a negative regulator or with adipogenesis (Song et al., 2019). Given the dual role of retinol during the adipogenesis (a positive regulator of pre-adipocyte hyperplasia but a negative regulator of final maturation; see Wang et al., 2016c), *GRPC5A* rises as an interesting gene to mediate the inhibitory effect of retinoids in adipogenesis (Amisten et al., 2017).

In addition, *MTMR2* (OCU1) was linked to the metabolic process of lipids. This gene has been proposed as functional candidate gene for IMF in GWAS and

signature of selection studies in a Duroc population selected for IMF (Kim et al., 2015).

4.6. Conclusions and Implications

This is the first GWAS study for IMF in rabbits and hence provides a benchmark for continuing research in the field. Our findings support that four genomic regions (on OCU1, OCU8 and OCU13) influence IMF content. The genomic variance explained by these associated regions is important although no major causal variants seem to segregate in the analysed rabbit populations. Therefore, according to what we observed in these divergently selected lines, it seems that IMF content is mainly driven by a polygenetic effect. In addition, we identified some candidate genes on the associated genomic regions of OCU13 (*EWSR1*), OCU8 (*APOLD1*, *PLBD1*, *PDE6H*, *GPRC5A, KRAS*), and OCU1 *(MTMR2*) related to IMF. Nevertheless, further research would be necessary in order to corroborate these results; for instance, a genotype refinement or sequencing of promoter and exonic regions of the candidate genes and its validation in independent populations of rabbits. Our results could be important for further studies to discover polymorphisms that can assist IMF genetic improvement.

4.7. Abbreviations

BMMR: Bayesian multi marker regression; FASTA: family-based score test association; MDS: multidimensional scaling; GWAS: genome-wide association study; LD: linkage disequilibrium; LTH: *longissimus thoracis et lumborum*; MAF: minor allele frequency; SMR: single marker regression; OCU: rabbit chromosome; QTL: quantitative trait loci; IMF: intramuscular fat; SNP: single nucleotide polymorphism.

4.8. Availability of data and material

The datasets used and analysed in the current study are available from the Figshare Repository (https://doi.org/10.6084/m9.figshare.9934058.v1).

4.9. References

Aken, B. L., S. Ayling, D. Barrell, L. Clarke, V. Curwen et al., 2016. "The Ensembl gene annotation system." *Database (Oxford),* 2016: baw093.

Aloulou, A., Y. Ben Ali, S. Bezzine, Y. Gargouri, and M. H. Gelb, 2012. "Phospholipases: An Overview", pp. 63–85 in *Lipases and Phospholipases: Methods and Protocols,* edited by G. Sandoval. Humana Press, Totowa, NJ.

Amisten, S., I. Mohammad Al-Amily, A. Soni, R. Hawkes, P. Atanes et al., 2017. "Anti-diabetic action of all-trans retinoic acid and the orphan G protein coupled receptor GPRC5C in pancreatic beta-cells." *Endocrine Journal,* 64(3): 325–338.

Astle, W., and D. J. Balding, 2009. "Population structure and cryptic relatedness in genetic association studies." *Statistical Science,* 24(4): 451–471.

Aulchenko, Y. S., S. Ripke, A. Isaacs, and C. M. van Duijn, 2007. "GenABEL: An R library for genome-wide association analysis." *Bioinformatics,* 23(10): 1294–1296.

Barrett, J. C., B. Fry, J. D. Maller, and M. J. Daly, 2005. "Haploview: Analysis and visualization of LD and haplotype maps." *Bioinformatics,* 21(2): 263–265.

Beissinger, T. M., G. J. Rosa, S. M. Kaeppler, D. Gianola, and N. de Leon, 2015. "Defining window-boundaries for genomic analyses using smoothing spline techniques." *Genetics Selection Evolution,* 47(1): 30.

Blasco, A., and R. N. Pena, 2018. "Current Status of Genomic Maps: Genomic Selection/GBV in Livestock", pp. 61–80 in *Animal Biotechnology 2*, edited by H. Niemann and C. Wrenzycki. Springer International Publishing, Cham.

Borg, I., and P. J. F. Groenen, 2005. "Modern Multidimensional Scaling: Theory and Applications." Springer International Publishing.

Browning, B. L., and S. R. Browning, 2016. "Genotype imputation with millions of reference samples." *American Journal of Human Genetics,* 98(1): 116–126.

Carneiro, M., S. Afonso, A. Geraldes, H. Garreau, G. Bolet et al., 2011. "The genetic structure of domestic rabbits." *Molecular Biology and Evolution,* 28(6): 1801–1816.

Carneiro, M., C. J. Rubin, F. Di Palma, F. W. Albert, J. Alföldi et al., 2014. "Rabbit genome analysis reveals a polygenic basis for phenotypic change during domestication." *Science*, 345(6200): 1074–1079.

Cesar, A. S., L. C. Regitano, G. B. Mourão, R. R. Tullio, D. P. Lanna et al., 2014. "Genome-wide association study for intramuscular fat deposition and composition in Nellore cattle." *BMC Genetics*, 15(1): 39.

Chaves, V. E., D. Frasson, and N. H. Kawashita, 2011. "Several agents and pathways regulate lipolysis in adipocytes." *Biochimie,* 93(10): 1631–1640.

Chen, W. M., and G. R. Abecasis, 2007. "Family-Based Association Tests for Genomewide Association Scans." *American Journal of Human Genetics,* 81(5): 913–926.

Claire D'Andre, H., W. Paul, X. Shen, X. Jia, R. Zhang et al., 2013. "Identification and characterization of genes that control fat deposition in chickens." *Journal of Animal Science Biotechnology,* 4(1): 43.

Do, D. N., F. S. Schenkel, F. Miglior, X. Zhao, and E. M. Ibeagha-Awemu, 2018. "Genome wide association study identifies novel potential candidate genes for bovine milk cholesterol content." *Scientific Reports,* 8(1): 13239.

Do, D. N., A. B. Strathe, T. Ostersen, S. D. Pant, and H. N. Kadarmideen, 2014. "Genome-wide association and pathway analysis of feed efficiency in pigs reveal candidate genes and pathways for residual feed intake." *Frontiers in Genetics,* 5: 307.

European Economic Community, 1998. "Council Directive 98/58/EC of 20 July 1998 concerning the protection of animals kept for farming purposes." *Official Journal L* 221(8): 8.

Fan, B., Z. Q. Du, D. M. Gorbach, and M. F. Rothschild, 2010. "Development and application of high-density SNP arrays in genomic studies of domestic animals." *Asian-Australasian Journal of Animal Science,* 23(7): 833–847.

Gao, Y., R. Zhang, X. Hu, and N. Li, 2007. "Application of genomic technologies to the improvement of meat quality of farm animals." *Meat Science,* 77(1): 36–45.

Garrick, D. J., 2011. "The nature, scope and impact of genomic prediction in beef cattle in the United States." *Genetics Selection Evolution,* 43: 17.

Garrick, D. J., and R. L. Fernando, 2013. "Implementing a QTL Detection Study (GWAS) Using Genomic Prediction Methodology", pp. 275–298 in *Genome-Wide Association Studies and Genomic Prediction,* edited by C. Gondro, J. van der Werf, and B. Hayes. Humana Press, Totowa, NJ.

Gotoh, T., T. Nishimura, K. Kuchida, and H. Mannen, 2018. "The Japanese Wagyu beef industry: Current situation and future prospects - A review." *Asian-Australasian Journal of Animal Science,* 31(7): 933–950.

Gotoh, T., H. Takahashi, T. Nishimura, K. Kuchida, and H. Mannen, 2014. "Meat produced by Japanese Black cattle and Wagyu." *Animal Frontiers,* 4(4): 46–54.

Green, M. R., J. Sambrook, and J. Sambrook, 2012. "Molecular cloning : a laboratory manual." 4th Edition. Cold Spring Harbor Laboratory Press, N.Y.

Hocquette, J. F., F. Gondret, E. Baéza, F. Médale, C. Jurie et al., 2010. "Intramuscular fat content in meat-producing animals: development, genetic and nutritional control, and identification of putative markers." *Animal,* 4(2): 303–319.

Hopkins, D. L., N. M. Fogarty, and S. I. Mortimer, 2011. "Genetic related effects on sheep meat quality." *Small Ruminant Research,* 101(1-3): 160–172.

Jiao, X., B. T. Sherman, D. W. Huang, R. Stephens, M. W. Baseler et al., 2012. "DAVID-WS: A stateful web service to facilitate gene/protein list analysis." *Bioinformatics,* 28(13): 1805–1806.

Jin, C., W. Wang, Y. Liu, and Y. Zhou, 2017. "RAI3 knockdown promotes adipogenic differentiation of human adipose-derived stem cells by decreasing β-catenin levels." *Biochemical and Biophysical Research Communications,* 493(1): 618–624.

Kass, R. E., and A. E. Raftery, 1995. "Bayes Factors." *Journal of the American Statistical Association,* 90(430): 773-795.

Kim, E. S., R. Ros-Freixedes, R. N. Pena, T. J. Baas, J. Estany et al., 2015. "Identification of signatures of selection for intramuscular fat and backfat thickness in two Duroc populations." *Journal of Animal Science,* 93(7): 3292–3302.

Kuleshov, M. V., M. R. Jones, A. D. Rouillard, N. F. Fernandez, Q. Duan et al., 2016. "Enrichr: a comprehensive gene set enrichment analysis web server 2016 update." *Nucleic Acids Research,* 44(W1): W90–W97.

Lander, E., and L. Kruglyak, 1995. "Genetic dissection of complex traits: Guidelines for interpreting and reporting linkage results." *Nature Genetics,* 11(3): 241–247.

Lionikas, A., C. Meharg, J. M. J. Derry, A. Ratkevicius, A. M. Carroll et al., 2012. "Resolving candidate genes of mouse skeletal muscle QTL via RNA-Seq and expression network analyses." *BMC Genomics,* 13(1): 592.

López de Maturana, E., N. Ibáñez-Escriche, Ó. González-Recio, G. Marenne, H. Mehrban et al., 2014. "Next generation modeling in GWAS: comparing different genetic architectures." *Human Genetics,* 133(10): 1235–1253.

Marras, G., A. Rossoni, H. Schwarzenbacher, S. Biffani, F. Biscarini et al., 2017. "zanardi: an open-source pipeline for multiple-species genomic analysis of SNP array data." *Animal Genetics,* 48(1): 121.

Martínez-Álvaro, M., P. Hernández, and A. Blasco, 2016. "Divergent selection on intramuscular fat in rabbits: Responses to selection and genetic parameters." *Journal of Animal Science*, 94(12): 4993–5003.

Mateescu, R. G., D. J. Garrick, A. J. Garmyn, D. L. Vanoverbeke, G. G. Mafi et al., 2015. "Genetic parameters for sensory traits in longissimus muscle and their associations with tenderness, marbling score, and intramuscular fat in Angus cattle." *Journal of Animal Science*, 93(1): 21–27.

McLaren, D. G., and C. M. Schultz, 1992. "Genetic Selection to Improve the Quality and Composition of Pigs." *In 45th Reciprocal Meat Conferences Proceedings,* Colorado State University. 115-121

Migdał, Ł., K. Kozioł, S. Pałka, W. Migdał, A. Otwinowska-Mindur et al., 2018. "Single nucleotide polymorphisms within rabbits (Oryctolagus cuniculus) fatty acids binding protein 4 (FABP4) are associated with meat quality traits." *Livestock Science,* 210: 21–24.

Miller, I., C. Rogel-Gaillard, D. Spina, L. Fontanesi, and A. M. de Almeida, 2014. "The Rabbit as an Experimental and Production Animal: From Genomics to Proteomics." *Current Protein and Peptide Science,* 15(2): 134–145.

Mortimer, S. I., J. H. J. van der Werf, R. H. Jacob, D. L. Hopkins, L. Pannier et al., 2014. "Genetic parameters for meat quality traits of Australian lamb meat." *Meat Science,* 96(2): 1016–1024.

Nyima, T., M. Müller, G. J. E. J. Hooiveld, M. J. Morine, and M. Scotti, 2016. "Nonlinear transcriptomic response to dietary fat intake in the small intestine of C57BL/6J mice." *BMC Genomics,* 17(1): 106.

Ochsner, K. P., M. D. Macneil, R. M. Lewis, and M. L. Spangler, 2017. "Economic selection index development for beefmaster cattle I: Terminal breeding objective." *Journal of Animal Science*, 95(3): 1063–1070.

Pannier, L., G. E. Gardner, R. A. O'Reilly, and D. W. Pethick, 2018. "Factors affecting lamb eating quality and the potential for their integration into an MSA sheepmeat grading model." *Meat Science,* 144: 43–52.

Peña, F., M. Juárez, A. Bonvillani, P. García, O. Polvillo et al., 2011. "Muscle and genotype effects on fatty acid composition of goat kid intramuscular fat." *Italian Journal of Animal Science,* 10(3): e40.

Pena, R. N., R. Ros-Freixedes, M. Tor, and J. Estany, 2016. "Genetic marker discovery in complex traits: A field example on fat content and composition in pigs." *International Journal of Molecular Sciences,* 17(12): 2100.

Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira et al., 2007. "PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses." *American Journal of Human Genetics,* 81: 559–575.

R Core Team, 2018. "R: A Language and Environment for Statistical Computing."

Ros-Freixedes, R., S. Gol, R. N. Pena, M. Tor, N. Ibáñez-Escriche et al., 2016. "Genome-wide association study singles out SCD and LEPR as the two main loci influencing intramuscular fat content and fatty acid composition in Duroc pigs." *PLoS ONE,* 11(3): e0152496.

Sahana, G., B. Guldbrandtsen, and M. S. Lund, 2011. "Genome-wide association study for calving traits in Danish and Swedish Holstein cattle." *Journal of Dairy Science,* 94(1): 479–486.

Sapp, R. L., J. K. Bertrand, T. D. Pringle, and D. E. Wilson, 2002. "Effects of selection for ultrasound intramuscular fat percentage in Angus bulls on carcass traits of progeny." *Journal of Animal Science*, 80(8): 2017–2022.

Schmid, M., and J. Bennewitz, 2017. "Invited review: Genome-wide association analysis for quantitative traits in livestock – a selective review of statistical models and experimental designs." *Archives Animal Breeding,* 60(3): 335–346.

Song, H., B. Sun, Y. Liao, D. Xu, W. Guo et al., 2019. "GPRC5A deficiency leads to dysregulated MDM2 via activated EGFR signaling for lung tumor development." *International Journal of Cancer,* 144(4): 777–787.

Spencer, C. C. A., Z. Su, P. Donnelly, and J. Marchini, 2009. "Designing genome-wide association studies: sample size, power, imputation, and the choice of genotyping chip." *PLoS Genetics,* 5(5): e1000477.

Stephens, M., and D. J. Balding, 2009. "Bayesian statistical methods for genetic association studies." *Nature Reviews Genetics,* 10(10): 681–690.

Strucken, E. M., H. A. Al-Mamun, S. de las Heras-Saldana, M. N. Bedhane, D. Lim et al., 2017. "Finding the marble - The polygenic architecture of intramuscular fat." *Journal of Animal Breeding and Genomics,* 1(2): 69–76.

Sukegawa, S., T. Miyake, T. Ibi, Y. Takahagi, H. Murakami et al., 2014. "Multiple marker effects of single nucleotide polymorphisms in three genes, AKIRIN2, EDG1 and RPL27A, for marbling development in Japanese Black cattle." *Animal Science Journal* 85(3): 193–197.

Sul, J. H., L. S. Martin, and E. Eskin, 2018. "Population structure in genetic studies: Confounding factors and mixed models." *PLoS Genetics,* 14(12): e1007309.

Swierczynski, J., A. Hebanowska, and T. Sledzinski, 2014. "Role of abnormal lipid metabolism in development, progression, diagnosis and therapy of pancreatic cancer." *World Journal of Gastroenterology,* 20(9): 2279–2303.

Toosi, A., R. L. Fernando, and J. C. M. Dekkers, 2018. "Genome-wide mapping of quantitative trait loci in admixed populations using mixed linear model and Bayesian multiple regression analysis." *Genetics Selection Evolution,* 50(1): 32.

Uemoto, Y., H. Nakano, T. Kikuchi, S. Sato, M. Ishida et al., 2012. "Fine mapping of porcine SSC14 QTL and SCD gene effects on fatty acid composition and melting point of fat in a Duroc purebred population." *Animal Genetics,* 43(2): 225–228.

Visscher, P. M., N. R. Wray, Q. Zhang, P. Sklar, M. I. McCarthy et al., 2017. "10 years of GWAS discovery: Biology, function, and translation." *American Journal of Human Genetics,* 101(1): 5–22.

Vitti, J. J., S. R. Grossman, and P. C. Sabeti, 2013. "Detecting Natural Selection in Genomic Data." *Annual Review of Genetics,* 47: 97–120.

Wahl, S., A. Drong, B. Lehne, M. Loh, W. R. Scott et al., 2017. "Epigenome-wide association study of body mass index and the adverse outcomes of adiposity." *Nature,* 541(7635): 81–86.

Wang, J., M. A. Elzo, L. Yan, S. Chen, X. Jia et al., 2016a. "A single nucleotide polymorphism in CAST gene is associated with meat quality traits in rabbits." *Animal Science Papers and Reports,* 34(3): 269–278.

Wang, W., and P. Seale, 2016. "Control of brown and beige fat development." *Nature Reviews Molecular Cell Biology,* 17(11): 691–702.

Wang, J., Y. Shi, M. A. Elzo, Y. Su, X. Jia et al., 2017. "Myopalladin gene polymorphism is associated with rabbit meat quality traits." *Italian Journal of Animal Science,* 16(3): 400–404.

Wang, X., N. R. Tucker, G. Rizki, R. Mills, P. H. Krijger et al., 2016b. "Discovery and validation of sub-threshold genome-wide association study loci using epigenomic signatures." *Elife,* 5: e10557.

Wang, B., Q. Yang, C. L. Harris, M. L. Nelson, J. R. Busboom et al., 2016c. "Nutrigenomic regulation of adipose tissue development — role of retinoic acid: A review." *Meat Science,* 120: 100–106.

Won, S., J. Jung, E. Park, and H. B. Kim, 2017. "Identification of genes related to intramuscular fat content of pig using genome-wide association study." *Asian-Australasian Journal of Animal Science,* 31(2): 157–162.

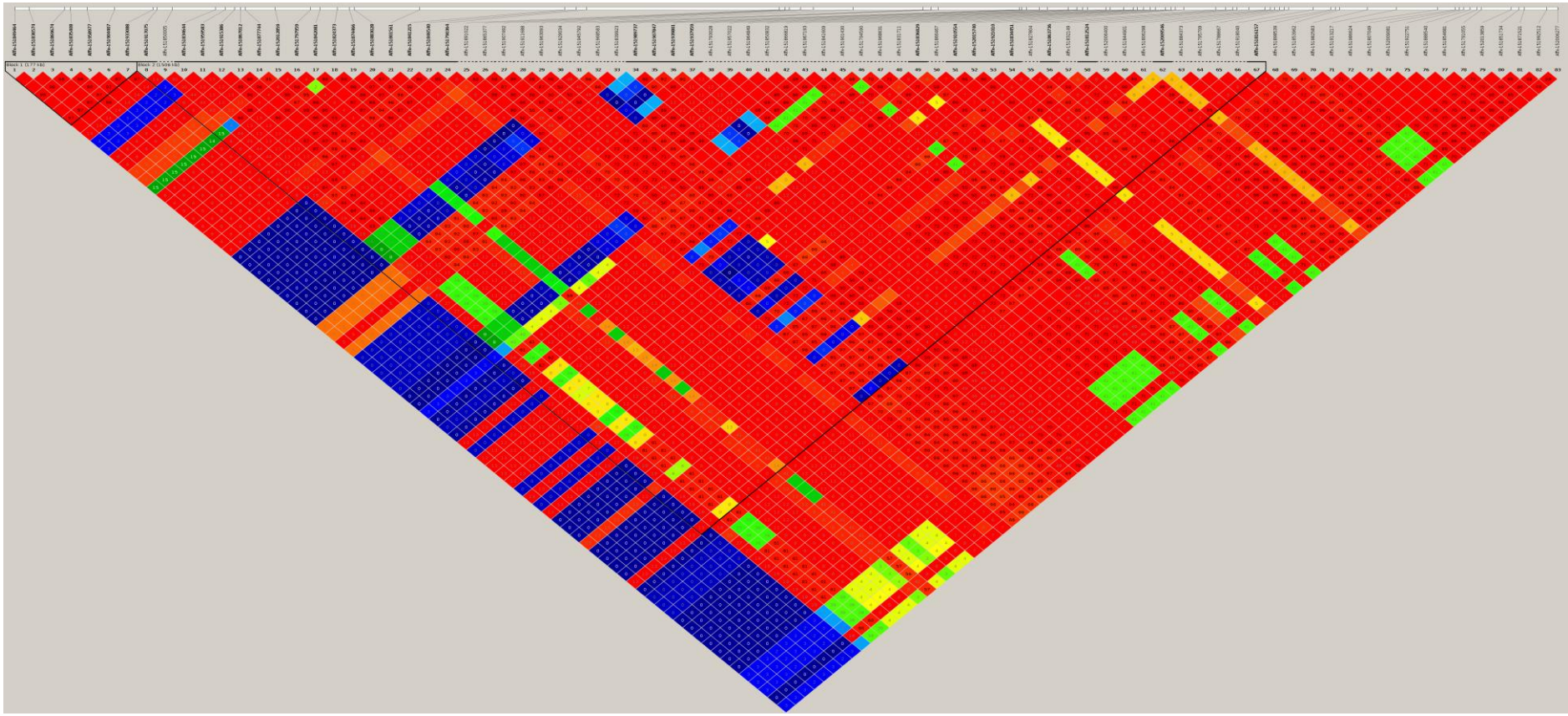
Zhang, G. W., L. Gao, S. Y. Chen, X. B. Zhao, Y. F. Tian et al., 2013. "Single nucleotide polymorphisms in the FTO gene and their association with growth and meat quality traits in rabbits." *Gene,* 527(2): 553–557.

Zhang, H., Z. Wang, S. Wang, and H. Li, 2012. "Progress of genome wide association study in domestic animals." *Journal of Animal Science Biotechnology,* 3(1): 26.

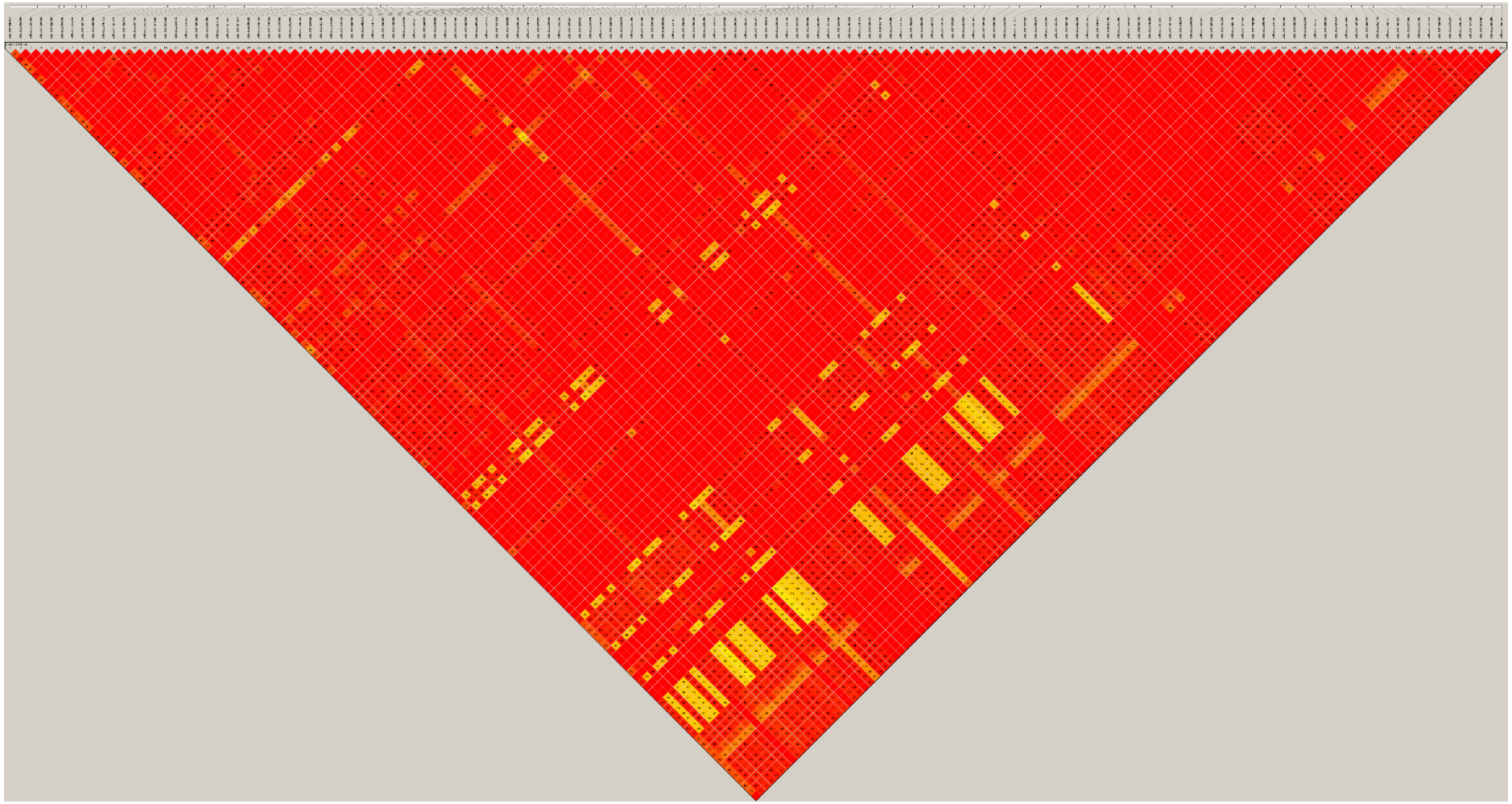
Zomeño, C., P. Hernandez, and A. Blasco, 2013. "Divergent selection for intramuscular fat content in rabbits. I. Direct response to selection." *Journal of Animal Science,* 91(9): 4526–4531.

4.10. Additional Files

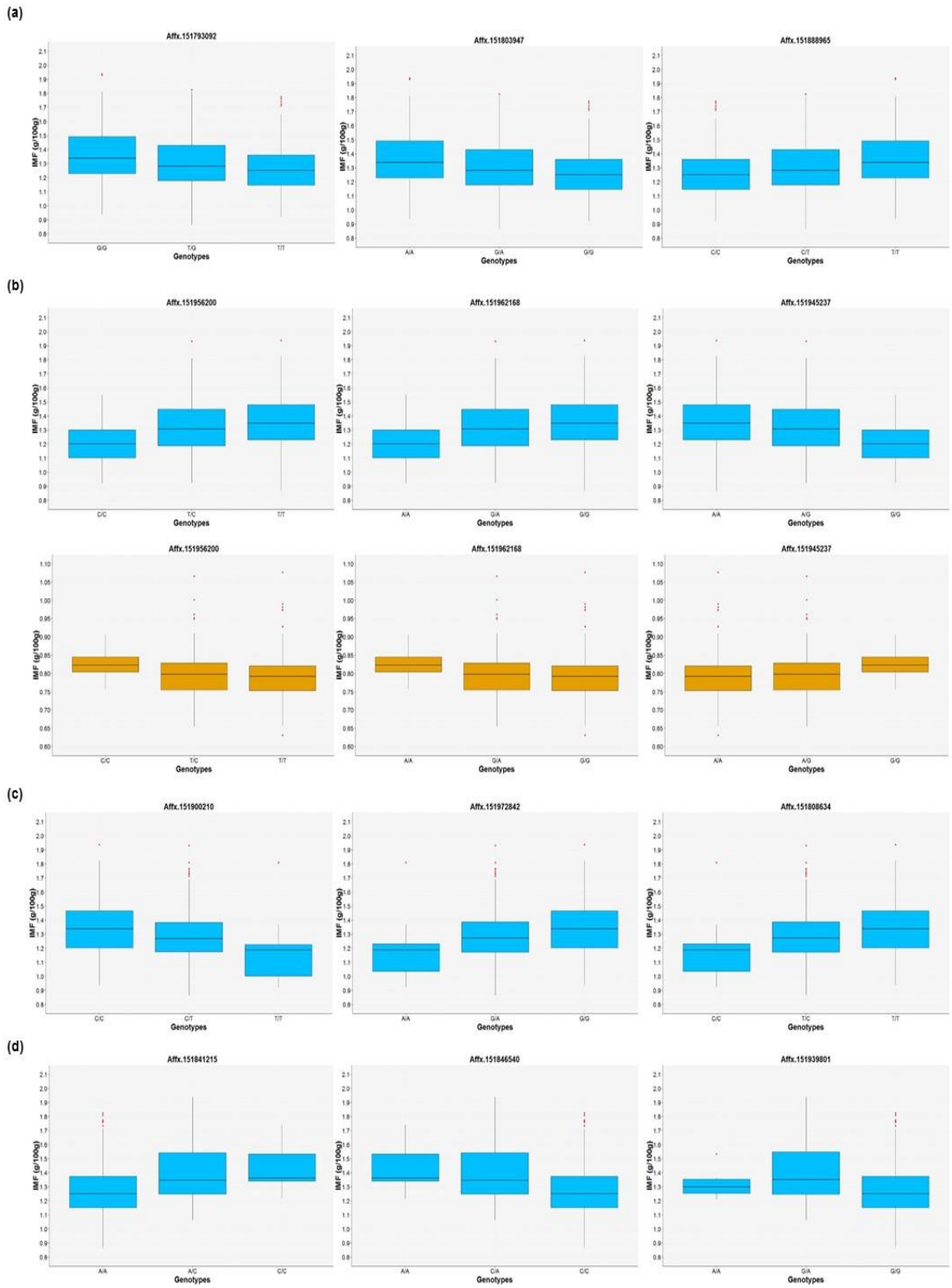
Additional supporting information may be found online in the Supporting Information section at the end of the article.



Additional file 1: Figure S1. Linkage disequilibrium (LD) block of the associated genomic region on OCU13. The window 1380 and 1381 display a shared LD block of 1506 kb. This block includes 18 SNPs of the first window and 42 SNPs from the second window. The red colour indicates a high LD and the blue colour indicate a low LD.



Additional file 2: Figure S2. Linkage disequilibrium (LD) block of an associated genomic region on OCU8. The window 841 and 842 display a shared LD block of 1945 kb. The block includes 145 SNPs. The red colour indicates a high LD.



Additional file 3: Figure S3. Assessment of genotypes for the three relevant SNPs within genomic regions associated with intramuscular fat. The light blue colour denotes IMF high line and orange colour denotes IMF low line. Boxplots of polymorphisms (SNPs) in **(a)** OCU1 (120.65 – 121.99 Mb), **(b)** OCU8 (14.01 - 15.47 Mb), **(c)** OCU8 (24.59 - 26.95 Mb), and **(d)** OCU13 (83.81 - 86.00 Mb). The SNPs in the regions (a), (c), and (d) displayed minor allele frequencies (MAF) below 0.03 (close to zero) within IMF low line precluding their assessment.

Additional file 4: Table S1. Genes found in the genomic regions associated with intramuscular fat.

Gene stable ID	Rabbit gene name	OCU	Gene type	Gene description
ENSOCUG00000010820	*MAML2*	1	protein coding	mastermind like transcriptional coactivator 2 [Source:NCBI gene;Acc:100345650]
ENSOCUG00000014568	*MTMR2*	1	protein coding	myotubularin related protein 2 [Source:NCBI gene;Acc:100352381]
ENSOCUG00000014557	*CEP57*	1	protein coding	centrosomal protein 57 [Source:NCBI gene;Acc:100353145]
ENSOCUG00000014549	*FAM76B*	1	protein coding	family with sequence similarity 76 member B [Source:NCBI gene;Acc:100354406]
ENSOCUG00000025632	*SUCLA2**	1	protein coding	succinate-CoA ligase ADP-forming beta subunit [Source:NCBI gene;Acc:105868516
ENSOCUG00000026640	*SESN3*	1	protein coding	sestrin 3 [Source:NCBI gene;Acc:100354919]
ENSOCUG00000014040	*ENDOD1*	1	protein coding	endonuclease domain containing 1 [Source:NCBI gene;Acc:100355178]
ENSOCUG00000001631	*KDM4D*	1	protein coding	lysine-specific demethylase 4D [Source:NCBI gene;Acc:100348204]
ENSOCUG00000001629	*CWC15*	1	protein coding	CWC15 spliceosome associated protein homolog [Source:NCBI gene;Acc:100355678]
ENSOCUG00000010737	*RASSF8*	8	protein coding	Ras association domain family member 8 [Source:NCBI gene;Acc:100357319]
ENSOCUG00000006725	*LMNTD1*	8	protein coding	lamin tail domain containing 1 [Source:NCBI gene;Acc:100357834]
ENSOCUG00000028888	*RF00001*	8	rRNA	
ENSOCUG00000012106	*KRAS*	8	protein coding	KRAS proto-oncogene, GTPase [Source:NCBI gene;Acc:100347487]
ENSOCUG00000012099	*ETFRF1*	8	protein coding	electron transfer flavoprotein regulatory factor 1 [Source:NCBI gene;Acc:100347981]
ENSOCUG00000015984	*CASC1*	8	protein coding	cancer susceptibility 1 [Source:NCBI gene;Acc:100358611]

Gene stable ID	Rabbit gene name	OCU	Gene type	Gene description
ENSOCUG00000015956	*LRMP*	8	protein coding	lymphoid restricted membrane protein [Source:NCBI gene;Acc:100358873]
ENSOCUG00000003963	*BCAT1*	8	protein coding	branched chain amino acid transaminase 1 [Source:NCBI gene;Acc:100337981]
ENSOCUG00000021067		8	protein coding	
ENSOCUG00000006014	*PDE6H*	8	protein coding	phosphodiesterase 6H [Source:NCBI gene;Acc:100345162]
ENSOCUG00000006010	*ARHGDIB*	8	protein coding	Rho GDP dissociation inhibitor beta [Source:NCBI gene;Acc:100345676]
ENSOCUG00000006003	*ERP27*	8	protein coding	endoplasmic reticulum protein 27 [Source:NCBI gene;Acc:100347486]
ENSOCUG00000016964	*MGP*	8	protein coding	matrix Gla protein [Source:NCBI gene;Acc:100008989]
ENSOCUG00000017191	*ART4*	8	protein coding	ADP-ribosyltransferase 4 (Dombrock blood group) [Source:NCBI gene;Acc:100008862]
ENSOCUG00000023339	*SMCO3*	8	protein coding	single-pass membrane protein with coiled-coil domains 3 [Source:NCBI gene;Acc:100348987]
ENSOCUG00000017177		8	protein coding	WW domain binding protein 11 [Source:NCBI gene;Acc:100349748]
ENSOCUG00000027663	*H2AFJ*	8	protein coding	histone H2A.J [Source:NCBI gene;Acc:100350997]
ENSOCUG00000017172	*HIST4H4*	8	protein coding	histone H4 [Source:NCBI gene;Acc:100351746]
ENSOCUG00000008162	*GUCY2C*	8	protein coding	guanylate cyclase 2C [Source:NCBI gene;Acc:100008740]
ENSOCUG00000026751	*PLBD1*	8	protein coding	phospholipase B domain containing 1 [Source:NCBI gene;Acc:100347232]
ENSOCUG00000012361	*ATF7IP*	8	protein coding	activating transcription factor 7 interacting protein [Source:NCBI gene;Acc:100353017]
ENSOCUG00000017095		8	protein coding	

Gene stable ID	Rabbit gene name	OCU	Gene type	Gene description
ENSOCUG00000021765		8	miRNA	
ENSOCUG00000015111	*GRIN2B*	8	protein coding	glutamate ionotropic receptor NMDA type subunit 2B [Source:NCBI gene;Acc:100353266]
ENSOCUG00000027899	*RF00411*	8	snoRNA	
ENSOCUG00000021882		8	miRNA	
ENSOCUG00000022231	*EMP1*	8	protein coding	epithelial membrane protein 1 [Source:NCBI gene;Acc:100009209]
ENSOCUG00000012837	*GSG1*	8	protein coding	germ cell associated 1 [Source:NCBI gene;Acc:100348479]
ENSOCUG00000012826	*FAM234B*	8	protein coding	family with sequence similarity 234 member B [Source:NCBI gene;Acc:100348737]
ENSOCUG00000015344	*HEBP1*	8	protein coding	heme binding protein 1 [Source:NCBI gene;Acc:100353525]
ENSOCUG00000015331	*GPRC5D*	8	protein coding	G protein-coupled receptor class C group 5 member D [Source:NCBI gene;Acc:100349237]
ENSOCUG00000016556	*GPRC5A*	8	protein coding	G protein-coupled receptor class C group 5 member A [Source:NCBI gene;Acc:100349493]
ENSOCUG00000014390	*DDX47*	8	protein coding	DEAD-box helicase 47 [Source:NCBI gene;Acc:100354028]
ENSOCUG00000025481	*APOLD1*	8	protein coding	apolipoprotein L domain containing 1 [Source:NCBI gene;Acc:100354429]
ENSOCUG00000023584	*RF00026*	13	snRNA	
ENSOCUG00000027270	*EWSR1**	13	protein coding	EWS RNA binding protein 1 [Source:HGNC Symbol;Acc:HGNC:3508
ENSOCUG00000028459	*RF00001*	13	rRNA	

* Genes annotated to the human orthologue.

Additional file 5: Table S2. Functions of genes identified in this study through Enrichr and DAVID.

Results of Enrich

DataBase	Orthologue Genes (rabbit chromosome)	Code	Term (Homo sapiens - Hs or Mus musculus - Mm)	Combined Score
KEGG	*H2AFJ(8), HIST4H4(8), KRAS(8), GRIN2B(8)*	HSA05034	Alcoholism	13.2227
KEGG	*H2AFJ(8), HIST4H4(8), GRIN2B(8)*	HSA05322	Systemic lupus erythematosus	9.3380
KEGG	*KRAS(8), GRIN2B(8)*	HSA04720	Long-term potentiation	8.1546
KEGG	*SESN3(1), KRAS(8)*	HSA04211	Longevity regulating pathway - mammal	6.9843
KEGG	*ARHGDIB(8), KRAS(8)*	HSA04722	Neurotrophin signaling pathway	5.9656
KEGG	*GUCY2C(8), PDE6H(8)*	HSA00230	Purine metabolism	4.1724
KEGG	*HIST1H4(8), KRAS(8)*	HSA05203	Viral carcinogenesis	4.0400
KEGG	*KRAS(8), GRIN2B(8)*	HSA04015	Rap1 signaling pathway	3.7396
KEGG	*KRAS(8)*	HSA05216	Thyroid cancer	3.6327
KEGG	*KRAS(8), GRIN2B(8)*	HSA04014	Ras signaling pathway	3.3951
KEGG	*KRAS(8)*	HSA04320	Dorso-ventral axis formation	3.2936
KEGG	*BCAT1(8)*	HSA01210	2-Oxocarboxylic acid metabolism	3.1575
KEGG	*SUCLA2(1)*	HSA00640	Propanoate metabolism	2.8686
KEGG	*KRAS(8)*	HSA04960	Aldosterone-regulated sodium reabsorption	2.8658
KEGG	*GRIN2B(8)*	HSA05033	Nicotine addiction	2.6715
KEGG	*SUCLA2(1)*	HSA00020	Citrate cycle (TCA cycle)	2.5982
KEGG	*KRAS(8)*	HSA05219	Bladder cancer	2.5727
KEGG	*ARHGDIB(8)*	HSA04962	Vasopressin-regulated water reabsorption	2.5377
KEGG	*BCAT1(8)*	HSA00770	Pantothenate and CoA biosynthesis	2.4263
KEGG	*BCAT1(8)*	HSA00270	Cysteine and methionine metabolism	2.1672
KEGG	*BCAT1(8)*	HSA00280	Valine, leucine and isoleucine degradation	2.1457

DataBase	Orthologue Genes (rabbit chromosome)	Code	Term (Homo sapiens - Hs or Mus musculus - Mm)	Combined Score
KEGG	*KRAS(8)*	HSA05223	Non-small cell lung cancer	2.1353
KEGG	*KRAS(8)*	HSA05213	Endometrial cancer	2.0689
KEGG	*KRAS(8)*	HSA05221	Acute myeloid leukemia	1.9737
KEGG	*MAML2(1)*	HSA04330	Notch signaling pathway	1.9012
KEGG	*GRIN2B(8)*	HSA05030	Cocaine addiction	1.8983
KEGG	*KRAS(8)*	HSA04370	VEGF signaling pathway	1.8632
KEGG	*KRAS(8)*	HSA04730	Long-term depression	1.8226
KEGG	*GRIN2B(8)*	HSA05014	Amyotrophic lateral sclerosis (ALS)	1.7835
KEGG	*KRAS(8)*	HSA05214	Glioma	1.7789
KEGG	*KRAS(8)*	HSA04213	Longevity regulating pathway - multiple species	1.6955
KEGG	*KRAS(8)*	HSA05210	Colorectal cancer	1.6879
KEGG	*KRAS(8)*	HSA05212	Pancreatic cancer	1.5734
KEGG	*KRAS(8)*	HSA04664	Fc epsilon RI signaling pathway	1.4557
KEGG	*GRIN2B(8)*	HSA05031	Amphetamine addiction	1.4339
KEGG	*KRAS(8)*	HSA05211	Renal cell carcinoma	1.4012
KEGG	*KRAS(8)*	HSA05230	Central carbon metabolism in cancer	1.2870
KEGG	*KRAS(8)*	HSA05218	Melanoma	1.1559
KEGG	*MTMR2(1)*	HSA00562	Inositol phosphate metabolism	1.1470
KEGG	*KRAS(8)*	HSA04917	Prolactin signaling pathway	1.1305
KEGG	*KRAS(8)*	HSA04662	B cell receptor signaling pathway	1.1145
KEGG	*KRAS(8)*	HSA05220	Chronic myeloid leukemia	1.0157
KEGG	*KRAS(8)*	HSA04012	ErbB signaling pathway	0.9771
KEGG	*KRAS(8)*	HSA05215	Prostate cancer	0.9488
KEGG	*SESN3(1)*	HSA04115	p53 signaling pathway	0.9110
KEGG	*KRAS(8)*	HSA04912	GnRH signaling pathway	0.9075

DataBase	Orthologue Genes (rabbit chromosome)	Code	Term (Homo sapiens - Hs or Mus musculus - Mm)	Combined Score
KEGG	*GRIN2B(8)*	HSA04713	Circadian entrainment	0.8412
KEGG	*KRAS(8)*	HSA04914	Progesterone-mediated oocyte maturation	0.8343
KEGG	*KRAS(8)*	HSA04933	AGE-RAGE signaling pathway in diabetic complications	0.8268
KEGG	*KRAS(8)*	HSA04540	Gap junction	0.7618
KEGG	*KRAS(8)*	HSA04915	Estrogen signaling pathway	0.7467
KEGG	*MTMR2(1)*	HSA04070	Phosphatidylinositol signaling system	0.7236
KEGG	*KRAS(8)*	HSA04916	Melanogenesis	0.7150
KEGG	*BCAT1(8)*	HSA01230	Biosynthesis of amino acids	0.6944
KEGG	*KRAS(8)*	HSA04725	Cholinergic synapse	0.6615
KEGG	*KRAS(8)*	HSA05231	Choline metabolism in cancer	0.6256
KEGG	*KRAS(8)*	HSA04660	T cell receptor signaling pathway	0.5350
KEGG	*GRIN2B(8)*	HSA04724	Glutamatergic synapse	0.4508
KEGG	*KRAS(8)*	HSA04726	Serotonergic synapse	0.4456
KEGG	*KRAS(8)*	HSA04919	Thyroid hormone signaling pathway	0.3467
KEGG	*SUCLA2(1)*	HSA01200	Carbon metabolism	0.2997
KEGG	*KRAS(8)*	HSA04071	Sphingolipid signaling pathway	0.2862
KEGG	*GRIN2B(8)*	HSA04728	Dopaminergic synapse	0.2763
KEGG	*KRAS(8)*	HSA04068	FoxO signaling pathway	0.2549
KEGG	*KRAS(8)*	HSA04360	Axon guidance	0.2415
KEGG	*KRAS(8)*	HSA05160	Hepatitis C	0.2171
KEGG	*CWC15(1)*	HSA03040	Spliceosome	0.2155
KEGG	*KRAS(8)*	HSA04210	Apoptosis	0.1997
KEGG	*KRAS(8)*	HSA04650	Natural killer cell mediated cytotoxicity	0.1814
KEGG	*KRAS(8)*	HSA04910	Insulin signaling pathway	0.1240
KEGG	*KRAS(8)*	HSA05161	Hepatitis B	0.0906

DataBase	Orthologue Genes (rabbit chromosome)	Code	Term (Homo sapiens - Hs or Mus musculus - Mm)	Combined Score
KEGG	*KRAS(8)*	HSA04072	Phospholipase D signaling pathway	0.0746
KEGG	*KRAS(8)*	HSA04550	Signaling pathways regulating pluripotency of stem cells	0.0705
KEGG	*KRAS(8)*	HSA04530	Tight junction	0.0672
KEGG	*KRAS(8)*	HSA04921	Oxytocin signaling pathway	0.0419
KEGG	*GRIN2B(8)*	HSA05016	Huntington's disease	-0.0322
KEGG	*GRIN2B(8)*	HSA05010	Alzheimer's disease	-0.0384
KEGG	*EWSR1(13)*	HSA05202	Transcriptional misregulation in cancer	-0.0416
KEGG	*KRAS(8)*	HSA05205	Proteoglycans in cancer	-0.0534
KEGG	*KRAS(8)*	HSA04010	MAPK signaling pathway	-0.0684
KEGG	*KRAS(8)*	HSA04062	Chemokine signaling pathway	-0.0688
KEGG	*KRAS(8)*	HSA04810	Regulation of actin cytoskeleton	-0.0735
KEGG	*MTMR2(1), SUCLA2(1), BCAT1(8)*	HSA01100	Metabolic pathways	-0.0780
KEGG	*KRAS(8)*	HSA04151	PI3K-Akt signaling pathway	-0.0915
KEGG	*GRIN2B(8)*	HSA04024	cAMP signaling pathway	-0.0923
KEGG	*KRAS(8)*	HSA05206	MicroRNAs in cancer	-0.0947
KEGG	*GRIN2B(8)*	HSA04080	Neuroactive ligand-receptor interaction	-0.0954
KEGG	*KRAS(8)*	HSA05166	HTLV-I infection	-0.0981
KEGG	*KRAS(8)*	HSA05200	Pathways in cancer	-0.1120
WIKIPATHWAY	*GPRC5A(8), GPRC5D(8)*	WP2369	GPCRs, Class C Metabotropic glutamate, pheromone (Hs)	15.6094
WIKIPATHWAY	*GPRC5A(8), GPRC5D(8)*	WP80	GPCRs, Class C Metabotropic glutamate, pheromone (Mm)	14.6975
WIKIPATHWAY	*MGP(8)*	WP501	NOTCH1 regulation of human endothelial cell calcification (Hs)	6.1611
WIKIPATHWAY	*GUCY2C(8), PDE6H(8)*	WP207	Purine metabolism (Mm)	5.3406

DataBase	Orthologue Genes (rabbit chromosome)	Code	Term (Homo sapiens - Hs or Mus musculus - Mm)	Combined Score
WIKIPATHWAY	*KRAS(8)*	WP327	Serotonin Receptor 2 and ELK-SRF/GATA4 signaling (Hs)	5.0556
WIKIPATHWAY	*GPRC5A(8), KRAS(8)*	WP1396	Integrated Pancreatic Cancer Pathway (Hs)	4.8393
WIKIPATHWAY	*KRAS(8)*	WP1270	Nanoparticle-mediated activation of receptor signaling (Hs)	4.6529
WIKIPATHWAY	*HIST4H4(8)*	WP2185	Gastric Cancer Network 1 (Hs)	4.2209
WIKIPATHWAY	*KRAS(8)*	WP474	RalA downstream regulated genes (Hs)	4.1832
WIKIPATHWAY	*KRAS(8)*	WP2380	MAPK Cascade (Hs)	4.0027
WIKIPATHWAY	*KRAS(8)*	WP3413	MAPK Cascade (Mm)	3.8582
WIKIPATHWAY	*SUCLA2(1)*	WP2643	TCA Cycle (Mm)	3.6417
WIKIPATHWAY	*KRAS(8)*	WP732	Extracellular vesicle-mediated signaling in recipient cells (Hs)	3.4936
WIKIPATHWAY	*KRAS(8)*	WP2361	IL-5 Signaling Pathway (Hs)	3.2793
WIKIPATHWAY	*GRIN2B(8)*	WP422	Hypothetical Network for Drug Addiction (Mm)	3.2223
WIKIPATHWAY	*ARHGDIB(8)*	WP251	FAS pathway and Stress induction of HSP regulation (Mm)	3.1941
WIKIPATHWAY	*GRIN2B(8)*	WP189	Hypothetical Network for Drug Addiction (Hs)	3.1572
WIKIPATHWAY	*ARHGDIB(8)*	WP2377	G13 Signaling Pathway (Mm)	3.0049
WIKIPATHWAY	*GPRC5A(8), GPRC5D(8)*	WP434	Non-odorant GPCRs (Mm)	2.9910
WIKIPATHWAY	*ARHGDIB(8)*	WP2870	G13 Signaling Pathway (Hs)	2.9422
WIKIPATHWAY	*ARHGDIB(8)*	WP127	FAS pathway and Stress induction of HSP regulation (Hs)	2.8848
WIKIPATHWAY	*KRAS(8)*	WP1246	ErbB Signaling Pathway (Hs)	2.7410
WIKIPATHWAY	*GRIN2B(8)*	WP666	Splicing factor NOVA regulated synpatic proteins (Mm)	2.7154
WIKIPATHWAY	*KRAS(8)*	WP571	Aryl Hydrocarbon Receptor (Hs)	2.6332
WIKIPATHWAY	*MAML2(1)*	WP1253	Notch Signaling Pathway (Hs)	2.4221

DataBase	Orthologue Genes (rabbit chromosome)	Code	Term (Homo sapiens - Hs or Mus musculus - Mm)	Combined Score
WIKIPATHWAY	*MGP(8)*	WP2290	Endochondral Ossification (Mm)	2.3491
WIKIPATHWAY	*KRAS(8)*	WP298	Oncostatin M Signaling Pathway (Hs)	2.2680
WIKIPATHWAY	*MGP(8)*	WP524	Endochondral Ossification (Hs)	2.1843
WIKIPATHWAY	*KRAS(8)*	WP314	IL-5 Signaling Pathway (Mm)	2.1637
WIKIPATHWAY	*KRAS(8)*	WP619	Signaling Pathways in Glioblastoma (Hs)	2.1494
WIKIPATHWAY	*KRAS(8)*	WP673	Rac1/Pak1/p38/MMP-2 pathway (Hs)	2.0524
WIKIPATHWAY	*HIST4H4(8)*	WP61	Histone Modifications (Hs)	2.0351
WIKIPATHWAY	*SESN3(1)*	WP2586	p53 signaling (Mm)	2.0285
WIKIPATHWAY	*GRIN2B(8)*	WP1983	Alzheimers Disease (Mm)	1.8911
WIKIPATHWAY	*KRAS(8)*	WP2374	DNA Damage Response (only ATM dependent) (Hs)	1.6906
WIKIPATHWAY	*BCAT1(8)*	WP151	Amino Acid metabolism (Mm)	1.6726
WIKIPATHWAY	*KRAS(8)*	WP2261	IL-3 Signaling Pathway (Mm)	1.6243
WIKIPATHWAY	*KRAS(8)*	WP3303	G Protein Signaling Pathways (Mm)	1.6024
WIKIPATHWAY	*KRAS(8)*	WP455	TNF alpha Signaling Pathway (Hs)	1.5752
WIKIPATHWAY	*KRAS(8)*	WP2902	G Protein Signaling Pathways (Hs)	1.4993
WIKIPATHWAY	*GRIN2B(8)*	WP2075	BDNF signaling pathway (Hs)	1.3118
WIKIPATHWAY	*GRIN2B(8)*	WP69	Alzheimers Disease (Hs)	1.2732
WIKIPATHWAY	*KRAS(8)*	WP662	Regulation of Actin Cytoskeleton (Hs)	1.1733
WIKIPATHWAY	*KRAS(8)*	WP232	Regulation of Actin Cytoskeleton (Mm)	1.1447
WIKIPATHWAY	*KRAS(8)*	WP231	Integrated Breast Cancer Pathway (Hs)	1.1232
WIKIPATHWAY	*KRAS(8)*	WP373	MAPK signaling pathway (Mm)	1.0490
WIKIPATHWAY	*KRAS(8)*	WP710	EGF/EGFR Signaling Pathway (Hs)	0.9870
WIKIPATHWAY	*KRAS(8)*	WP295	Chemokine signaling pathway (Mm)	0.9853
WIKIPATHWAY	*KRAS(8)*	WP35	MAPK Signaling Pathway (Hs)	0.9724
WIKIPATHWAY	*KRAS(8)*	WP2271	EGFR1 Signaling Pathway (Mm)	0.9416

DataBase	Orthologue Genes (rabbit chromosome)	Code	Term (Homo sapiens - Hs or Mus musculus - Mm)	Combined Score
WIKIPATHWAY	*HEBP1(8)*	WP111	Circadian rythm related genes (Hs)	0.7912
WIKIPATHWAY	*KRAS(8)*	WP2064	Focal Adhesion-PI3K-Akt-mTOR-signaling pathway (Mm)	0.6032
WIKIPATHWAY	*GRIN2B(8)*	WP2059	XPodNet - protein-protein interactions in the podocyte expanded by STRING (Mm)	0.1172

Functions search from DAVID analysis

Category	Term	Genes	P-Value	Fold Enrichment	FDR
GOTERM_BP_DIRECT	GO:0008380~RNA splicing	ENSOCUG00000014390, ENSOCUG00000017177	3.56E-02	52.5842	30.0541
INTERPRO	IPR004031:PMP-22/EMP/MP20/Claudin	ENSOCUG00000012837, ENSOCUG00000022231	7.22E-02	25.8108	53.3161
INTERPRO	IPR017978:GPCR, family 3, C-terminal	ENSOCUG00000016556, ENSOCUG00000015331	8.82E-02	20.9713	60.8533
KEGG_PATHWAY	ocu05322:Systemic lupus erythematosus	ENSOCUG00000015111, ENSOCUG00000027663, ENSOCUG00000017172	1.73E-02	13.3007	13.9990
KEGG_PATHWAY	ocu05034:Alcoholism	ENSOCUG00000015111, ENSOCUG00000027663, ENSOCUG00000017172	2.64E-02	10.6098	20.6850
GOTERM_BP_ALL	GO:0008380~RNA splicing	ENSOCUG00000014390, ENSOCUG00000017177, ENSOCUG00000001629	3.36E-02	9.7951	39.0736

Category	Term	Genes	P-Value	Fold Enrichment	FDR
GOTERM_BP_FAT	GO:0008380~RNA splicing	ENSOCUG00000014390, ENSOCUG00000017177, ENSOCUG00000001629	3.50E-02	9.5853	39.5966
UP_KEYWORDS	Coiled coil	ENSOCUG00000012361, ENSOCUG00000010820, ENSOCUG00000017177, ENSOCUG00000014557, ENSOCUG00000015984, ENSOCUG00000001629, ENSOCUG00000014568, ENSOCUG00000026640, ENSOCUG00000015956, ENSOCUG00000010737	1.59E-02	2.3665	13.6936

CHAPTER FIVE

5. THE EFFECT OF DIVERGENT SELECTION FOR INTRAMUSCULAR FAT IN THE DOMESTIC RABBIT GENOME.

Authors: B. S. Sosa-Madrid[1], L. Varona[2], A. Blasco[1], P. Hernández[1], C. Casto-Rebollo[1], and N. Ibáñez-Escriche[1].

Institutional affiliations

[1] Institute for Animal Science and Technology, Universitat Politècnica de València, 46022 Valencia, Spain.

[2] The AgriFood Institute of Aragon (IA2). Department of Anatomy, Embryology and Genetics, Universidad de Zaragoza, 50013 Saragossa, Spain.

The content of this chapter has been accepted in *Animal*.

5.1. Abstract

An experiment of divergent selection for intramuscular fat was carried out at Universitat Politècnica de València. The high response of selection in intramuscular fat, after nine generations of selection, and a multidimensional scaling analysis showed a high degree of genomic differentiation between the two divergent populations. Therefore, local genomic differences could link genomic regions, encompassing selective sweeps, to the trait used as selection criterion. In this sense, the aim of this study was to identify genomic regions related to intramuscular fat through three methods for detection of selection signatures, and to generate a list of candidate genes. The methods implemented in this study were: Wright's fixation index (F_{st}), cross population – composite likelihood ratio (XP-CLR) and cross population – extended haplotype homozygosity (XP-EHH). Genomic data came from the 9th generation of the two populations divergently selected, 237 from low line and 240 from high line. A high SNP-density array, Affymetrix Axiom OrcunSNP Array (around 200k SNPs), was used for genotyping samples. Several genomic regions distributed along chromosomes were identified as signatures of selection (SNPs having a value above cut-off of 1%) within each method. In contrast, eight genomic regions, harbouring 80 SNPs (OCU1, OCU3, OCU6, OCU7, OCU16, and OCU17), were identified by at least two methods and none by the three methods. In general, our results suggest that intramuscular fat selection influenced multiple genomic regions which can be a consequence of either only selection effect or the combined effect of selection and genetic drift. In addition, 73 genes were retrieved from the eight selection signatures. After functional and enrichment analyses, the main genes into the selection signatures linked to energy, fatty acids, carbohydrates and lipid metabolic processes were *ACER2, PLIN2, DENND4C, RPS6, RRAGA* (OCU1), *ST8SIA6, VIM* (OCU16), *RORA, GANC* and *PLA2G4B* (OCU17). This genomic scan is the first study using rabbits from a divergent selection experiment. Our results pointed out a large polygenic component of the studied trait. Besides, promising positional candidate genes would be analysed in further studies in order to bear out their contributions to this trait and their feasible implications for rabbit breeding programs.

Keyword: *genome scan, genomic divergence, lagomorph, meat quality, selection signatures.*

5.2. Implications

Intramuscular fat content is an essential factor in meat quality because affects nutritional, sensory and technological properties of meat; such as tenderness, flavour and juiciness of meat. In this study we applied methods of selection signatures to identify genomic regions modified by a divergent selection experiment for intramuscular fat in rabbits. Results revealed several selection signatures across the rabbit genome with genes linked to lipid metabolism. These findings will help to increase our understanding of intramuscular fat genomic regulation and could be used to apply in genomic evaluation programs for intramuscular fat.

5.3. Background

Selection and mutation trigger shifts in the genome architecture of traits, gathering the history of particular populations at a genomic level (Oleksyk et al., 2010). Genomic regions harbouring genes influenced by a selective process can be detected by the methods for the identification of selection signatures. These methods can be categorized in four groups depending on the assumptions behind the null hypothesis of absence of selection (Qanbari and Simianer, 2014) based on: (i) classical analyses of genetic variability (*e.g.* F_{st} – Wright's fixation index, π - nucleotide diversity); (ii) reduction of local variation in genomic regions (*e.g.* ROH – run of homozygosity); (iii) modification of allelic frequency spectrum (*e.g.* TD – Tajima's D, Fay and Wu H test, CLR – composite likelihood ratio); (iv) and linkage disequilibrium decay (*e.g.* |iHS| – integrated haplotype score, EHH – extended haplotype homozygosity, varLD – variation of linkage disequilibrium); see reviews by Oleksyk et al., 2010, and Qanbari & Simianer, 2014. The choice of methods depends on the type of selective events, timescale of selective events, the density of the genotyping data and the number of populations available for each particular study. A combination of methods for selection signatures can provide a clearer evidence of the genomic regions considered as selection signatures (Utsunomiya et al., 2013).

The identification of genomic regions containing genes affected by natural and artificial selection can be a difficult task, because selection is a complex phenomenon involving a potentially large number of traits (Mallick et al., 2009). Conversely, populations derived from divergent selection experiments for one

trait provide a valuable biological material for detecting those signatures, as the genetic divergence between them is linked to one particular trait (Qanbari and Simianer, 2014). In this sense, several studies of divergent selection were used to detect genomic regions associated with selection events in poultry: for body weight (Johansson et al., 2010), feather pecking behaviour (Grams et al., 2015), and antibody response (Lillie et al., 2017); and pigs: for intramuscular fat (Kim et al., 2015), and feed efficiency (Mauch et al., 2018).

In rabbits, an experiment of divergent selection for intramuscular fat was carried out at the Universitat Politècnica de València attaining a high selection response (Martínez-Álvaro et al., 2016). The genomic information from these two rabbit lines establishes an outstanding material to disentangle the genetic architecture of intramuscular fat content through the implementation of genome-wide scan studies for the detection of selection signatures.

The aim of this study was to identify genomic regions using three methods to detect selection signatures that exploit the genomic information from divergent populations and based on distinct hypotheses. The first is the fixation index (F_{st}) (Qanbari and Simianer, 2014), based on conventional genetic differentation; the second is the XP-CLR (Chen et al., 2010), which analyses the modifications on the allele frequency spectrum; and the last one is the XP-EHH (Sabeti et al., 2007), focused on the differences on the extension of linkage disequilibrium between populations. The final objective was to generate a list of potential candidate genes associated with intramuscular fat content.

5.4. Material and Methods

5.4.1. Ethical statement

The animal manipulations were approved by the Ethical Committee of the Universitat Politècnica de València, according to Council Directives 98/58/EC (European Economic Community, 1998).

5.4.2. Animals, genotyping data and quality control

The two rabbit lines selected divergently for intramuscular fat came from a synthetic line (base population) reared at Universitat Politècnica de València (Zomeño et al., 2013). Each line was composed of 8 to 10 sires and 40 to 60 does per generation. Further details of the divergent selection experiment for intramuscular fat are presented in Martínez-Álvaro et al. (2016). After nine generations of selection, the response was 3.1 phenotypic standard deviations (41% of the mean from the base population), estimated as the phenotypic difference between the two divergently selected lines (Sosa-Madrid et al., 2019). Muscle samples were collected for genotyping. A total of 480 individual rabbits (240 from each line) at 9th generation were genotyped with the Affymetrix Axiom OrcunSNP Array (around 200k SNPs). In addition, we genotyped 96 ancestors at 8th generation (10 sires and 38 dams by each line). Quality control of the SNP data was performed using *"Axiom Analysis Suite v. 3.0.1.4"* by using the following criteria: (i) individual call rate > 0.97; (ii) SNP call rate > 0.95; (iii) SNP minor allele frequency (MAF) > 0.05; (iv) and only autosomal SNPs with known positions were used. A MAF threshold of 0.05 was chosen in order to control the rate of false positive selection signatures and the effect of genotyping errors on the results. After filtering, we imputed the missing genotypes and inferred haplotype phases using population and genealogical information with *FImpute* (Sargolzaei et al., 2014). The final dataset consisted of 89,968 genotyped SNPs from 477 rabbits (240 from the high and 237 from the low lines, respectively).

5.4.3. Divergence between lines

At first, a multidimensional scaling (MDS) analysis with all genomic data was carried out to corroborate the divergence between lines. The command cmdscale() from *R* package *stats* was implemented for the MDS analysis. In addition, linkage disequilibrium was computed as Pearson's squared correlation coefficient (r^2) across the rabbit genome using *PLINK* (Purcell et al., 2007).

5.4.4. Detection of signatures of selection

The data were analysed using the following methods for the detection of selection signatures, taking advantage of selection in the two divergent lines after nine generations of selection:

5.4.4.1. Fixation Index (F_{st})

The F_{st} was computed for each SNP as:

$$F_{st} = \frac{(H_e - H_o)}{H_e}$$

where H_e and H_o are the expected and observed heterozygosity, respectively. The F_{st} values were clustered over sliding windows of predefined size (250kb, 500kb and 1000kb) surrounding every SNP. The F_{st} normalization was carried out in order to correct the F_{st} values due to the heterogeneous distribution of SNPs, after quality control, along the rabbit genome. The equation used was:

$$normalized_\mathrm{Fst} = \frac{\bar{X}_J - \mu}{S/\sqrt{n_j}}$$

This is based on the number of SNP within window j: n_j , the standard deviation using all data: S, the deviation from F_{st} average of a given window j: $\bar{X}_J$, and the F_{st} total mean: μ (Beissinger et al., 2015).

5.4.4.2. Cross Population Composite Likelihood Ratio Test (XP-CLR)

The XP-CLR method computes the likelihood ratio of selection signatures by comparing the spatial distribution of allele frequencies in an observed window to the frequency spectrum of the whole genome between two populations (Chen et al., 2010). The high line was used as the objective population and the low line was used as the reference population. In this analysis, XP-CLR software (http://genetics.med.harvard.edu/reich/Reich_Lab/Software.htm) was employed to compute the XP-CLR. After several exploratory analyses on XP-CLR score and its parameters, we defined a grid size of 2000 base pairs, sliding window size of three levels (250 kb, 500 kb and 1000 kb), maximum number of SNPs within a window 200, and a correlation value between two adjacent SNPs weighted with a cut-off of $r^2 > 0.95$ (author's recommendation). A shortcoming of composite likelihood ratio based-methods is that the correlation of marginal likelihood terms in the composite likelihood function is ignored. Thus, these methods overestimate the amount of information that is available in the data,

which can prompt false-positive signals of selection. To control for this issue, the XP-CLR method assigns weights to each marginal likelihood function in proportion to their statistical independence from all of the others. When $r^2 > 0.95$, CLR scores for two SNPs are down-weighted. After performing the analyses for every level of sliding window, the XP-CLR score for every SNP was chosen as the value of the nearest grid to each SNP. More details of the parameters of this method are described in Chen et al. (2010).

5.4.4.3. Cross Population Extended Haplotype Homozygosity Test (XP - EHH)

The extended haplotype homozygosity (EHH) profiles are defined as the probability that two randomly chosen haplotypes are identical by descent for the entire interval from the core region to a given point. The XP-EHH test compares the integrated EHH profiles between two populations around the same SNP, detecting ongoing selection or nearly fixed sites (overrepresented haplotypes) unveiling selection in one population (Sabeti et al., 2007). As in the XP-CLR score, we defined the high line as the objective population and low line as the reference population. First, we calculated the integrated haplotype score (*IHH*) for both lines. Then, the statistic was calculated at each SNP position as:

$$XP\text{-}EHH_{high\text{-}low} = \ln\left(IHH_{high} / IHH_{low}\right)$$

in which $XP\text{-}EHH_{high\text{-}low}$ is the XP-EHH between the high and low intramuscular fat lines, IHH_{high} is the integrated haplotype score of the high line, and IHH_{low} is the integrated haplotype score of the low line (Sabeti et al., 2007; Qanbari & Simianer, 2014). The maximum of extended haplotype was restricted in 250 kb, 500 kb and 1000 kb in order to compare with the other methods and to evaluate the sensitivity of the methods to the window size. The computation of XP-EHH score was carried out using selscan software (Szpiech & Hernandez, 2014), and finally, normalization of these data was performed since recombination rates vary widely across the rabbit genome within and between populations. This normalization was carried out setting all such log-ratios have zero mean and unit variance. Thus, the EHH statistic can be interpreted as a measure of selection solely after appropriate normalization for genome-wide difference in

haplotype length between populations as the distribution of recombination sites and the recombination rate are not steady (Sabeti et al., 2007).

5.4.5. Enrichment analysis of functional annotation, and gene ontology (GO) terms

In this study, we used a cut-off of 1% (the 99th percentile of all values) for every method. In order to determine the genomic regions of interest for searching genes and functional annotations, we used the physical position of the SNPs exceeding the cut-off (+/- 250 kb) in at least two methods. This distance criterion was chosen based on the results of the relationship between window size and the threes methods of selection signatures used in this study. We considered that criterion for searching genes under the assumption that an outstanding signature of selection must be detected in at least two methods, showing better evidence in this way and taking into account that the methods use different null hypotheses of absence of selection (Kim et al., 2015; González-Rodríguez et al., 2016). Besides, we considered a cut-off of 1% in order to reduce false positives in the detection of selection signatures (Mallick et al., 2009).

The genomic regions for each rabbit chromosome (OCU) were defined based on the OryCun2.0 rabbit genome assembly (Carneiro et al., 2014b). The genes comprised within those genomic regions were identified using *BIOMART* (Ensemble 98; https://www.ensembl.org/index.html) and then, the functional annotation and gene ontology (GO) terms enrichment analysis was performed using Enrichr (Kuleshov et al., 2016). The gene functional analysis was carried out using the rabbit and mouse annotation databases. Parameters recommended by the developers of *Enrichr* (*i.e.* p-value < 0.05 for Fisher exact test, and a high combined score, greater than 70) were used for the identification of genes in this study. We focused on the genes related to biological functions (GO terms) of energy metabolism and lipid metabolism.

5.5. Results and Discussion

The average physical distance between SNPs was 23.51 kb after quality control. However, the density of the SNPs through the rabbit genomic map was heterogeneous (Figure 1). The average of SNP density by each 1-Mb window was

41.87 with a standard deviation of 22.36, ranging between zero and 93. The OCU14, OCU20 and OCU21 had a low density with the average number of SNPs 32.29, 35.56 and 28.38 per Mb, respectively. On OCU14, two large gaps without any SNP marker was found (54.0 - 64.0 Mb and 89.0 - 95.0 Mb). Despite the gaps, the SNP density used in our study was in line with other studies (Gurgul et al., 2018; Ma et al., 2019). The heterogeneity of SNP density confirmed the need of data normalisation for the methods of selection signatures.

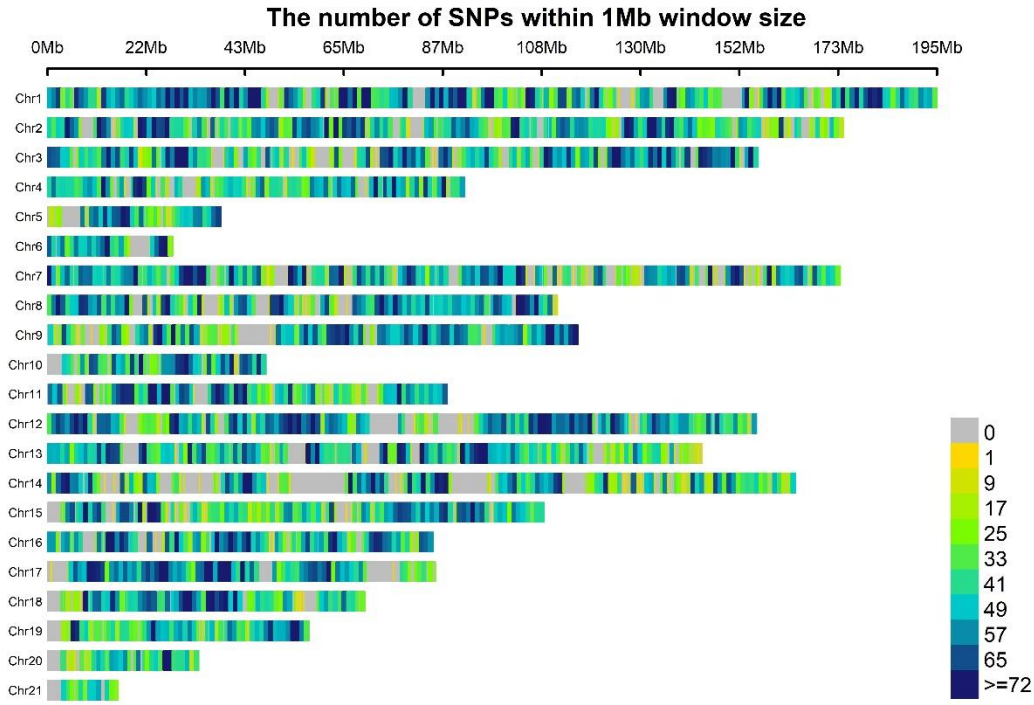


Figure 1. Single nucleotide polymorphism (SNP) density within 1-Mb window size using 89,968 SNPs after quality control for the intramuscular fat genomic data in rabbits.

The linkage disequilibrium was very high, with estimated r^2 values of 0.81, 0.68 and 0.52 at physical distances of 40 kb, 250 kb and 1000 kb, respectively. The number of linkage disequilibrium blocks was 2309 encompassing 89346 SNPs and showing a distance of 1 Mb for the longest linkage disequilibrium block. Moreover, the MDS displayed a noticeable genomic differentiation between the individuals (8th and 9th generation) from the high and low lines; 17.59 % and 2.75 % of variance explained by two first dimensions, respectively (Figure 2). As

expected, this analysis also corroborated the close relationship between the parents (8th generation) and their offspring (9th generation) in both lines.

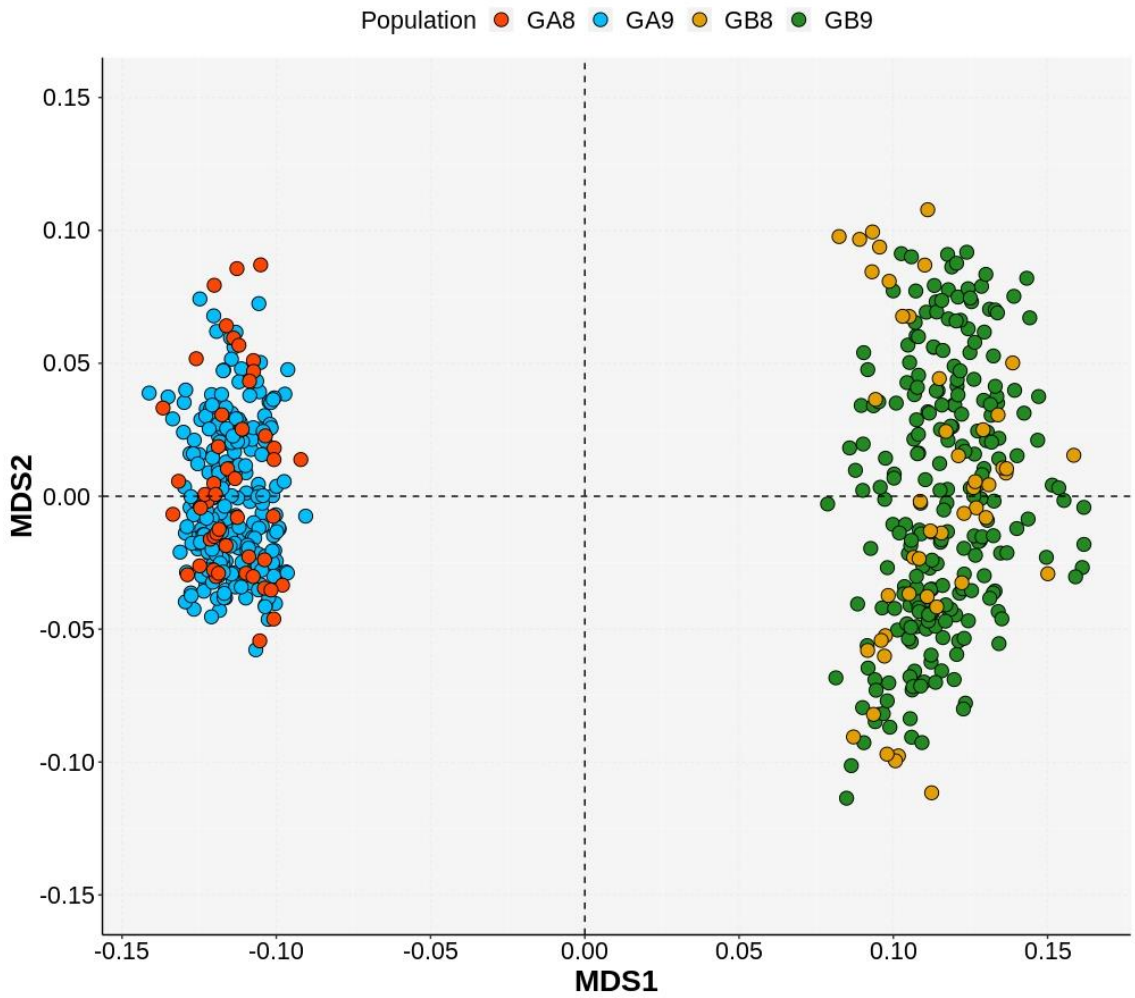


Figure 2. Multidimensional scaling plot of intramuscular fat genomic data in rabbits. The plot displays the first component (MDS1), the second component (MDS2) and intramuscular fat lines: high line at 8th generation (GA8 in red), high line at 9th generation (GA9 in light blue), low line at 8th generation (GB8 in yellow), high line at 9th generation (GB9 in dark green).

5.5.1. Genome-wide scan for populations from divergent selection

The identification of potential signatures of selection was performed with three methods (F_{st}, XP-CLR and XP-EHH) and three window sizes (250 kb, 500 kb and 1000 kb). The correlations between window sizes within every method were very high for F_{st} method (> 0.92) and lower for the XP-CLR and XP-EHH methods, especially between 250 kb and 1000 kb (Table 1). However, all correlations

within the methods were high, between 0.72 and 0.97, confirming that the results were robust to variations in window size. Thus, hereinafter we will refer exclusively to the results obtained with a window size of 500 kb. As expected, the correlations of the results between methods were very low, showing a mean of 0.08 using absolute values (Table 1). These correlations agreed with a genome scan study using pigs divergently selected for intramuscular fat, in which the correlations were less of 0.12 between methods of selection signatures (iHS, F_{st} and Rsb; Kim et al., 2015). It can be explained because every method entails a distinct hypothesis, (Qanbari and Simianer, 2014; González-Rodríguez et al., 2016), capturing different selection signals depending on the timescale of selective events (Utsunomiya et al., 2013). Hence, we decided to analyse the results separately and with those SNPs exceeding a cut-off of 1% in at least two methods for establishing selection signatures and to search for candidate genes.

The average of 500 kb windows for F_{st} (non-normalised), XP-CLR (non-normalised) and XP-EHH (in absolute value) were 0.0973, 1.7228 and 0.5878, respectively. The average of F_{st} by computing each SNP was 0.10. This average of F_{st} was higher than the results obtained in other studies between several populations of domestic European rabbit, F_{st} = 0.08 (Carneiro et al., 2014a), and also, from an experiment of divergent selection for uterine capacity in rabbits, F_{st} = 0.05 (Sosa-Madrid et al., 2017). To our knowledge, no comparison can be made for XP-EHH and XP-CLR because until date this is the first study using these methods for detecting signatures of selection in rabbits.

Table 1. Correlations between the levels of window size within methods of selection signatures in rabbits.

	XP-CLR 250kb	**XP-CLR 500kb**	**XP-CLR 1000kb**	**XP-EHH 250kb**	**XP-EHH 500kb**	**XP-EHH 1000kb**	**F_{st} 250kb**	**F_{st} 500kb**	**F_{st} 1000kb**
XP-CLR 250kb	1	0.8798	0.8142	-0.0284	-0.0298	-0.0437	0.1804	0.1697	0.1545
XP-CLR 500kb		1	0.9587	-0.0427	-0.0463	-0.0634	0.1804	0.1731	0.1571
XP-CLR 1000kb			1	-0.0442	-0.0486	-0.0669	0.1787	0.1726	0.1572
XP-EHH 250kb				1	0.8951	0.7189	-0.0195	-0.0226	-0.0238
XP-EHH 500kb					1	0.8229	-0.0259	-0.0283	-0.0302
XP-EHH 1000kb						1	-0.0403	-0.0424	-0.0463
F_{st} 250kb							1	0.9719	0.9215
F_{st} 500kb								1	0.9648
F_{st} 1000kb									1

XP-CLR = cross population – composite likelihood ratio. XP-EHH = cross population – extended haplotype homozygosity. F_{st} = Wright's fixation index.

The results of the genome-wide scan are shown by Manhattan plots in Figure 3. The results of each method individually showed several chromosomes with SNPs exceeding the cut-off of 1%. However, the joint results of the three methods did not evidence a genomic region clearly linked to a signature of selection. None of SNPs had values exceeding the cut-off of 1% in all methods (Figure 4). In contrast, several SNPs (80) associated with selection signatures overlapped between at least two methods. The overlapping results between XP-CLR and F_{st} were greater than the others, harbouring SNPs in OCU16 (31 SNPs) and OCU17 (24 SNPs); see Table 2. It can be explained because F_{st} and XP-CLR are based in allele frequencies and could detect dramatic shifts of opposite symmetrical allele frequencies for the SNPs located in the vicinity of an important causative variant. Conversely, XP-EHH is based on haplotype lengths comparison and was designed to compare a selected population with a reference population (non-selected; see Sabeti et al., 2007). Then, if the extension of a selected haplotype occurs in both lines with similar strength, the XP-EHH power detection could be lower than when regions were selected in one of the divergent lines but not in the another.

In total, eight genomic regions of the rabbit genome were identified as selection signatures (Table 2). This number is low in comparison with most signature selection studies (13 – 224) using populations of divergent artificial selection in poultry for body weight (Johansson et al., 2010), antibody response (Grams et al., 2015) and feather pecking behaviour (Lillie et al., 2017); and in a divergent selection experiment for intramuscular fat and backfat thickness in pigs (15 regions; Kim et al., 2015). Nevertheless, most of these studies identified selection signatures using only one method, unlike the criterion of our study.

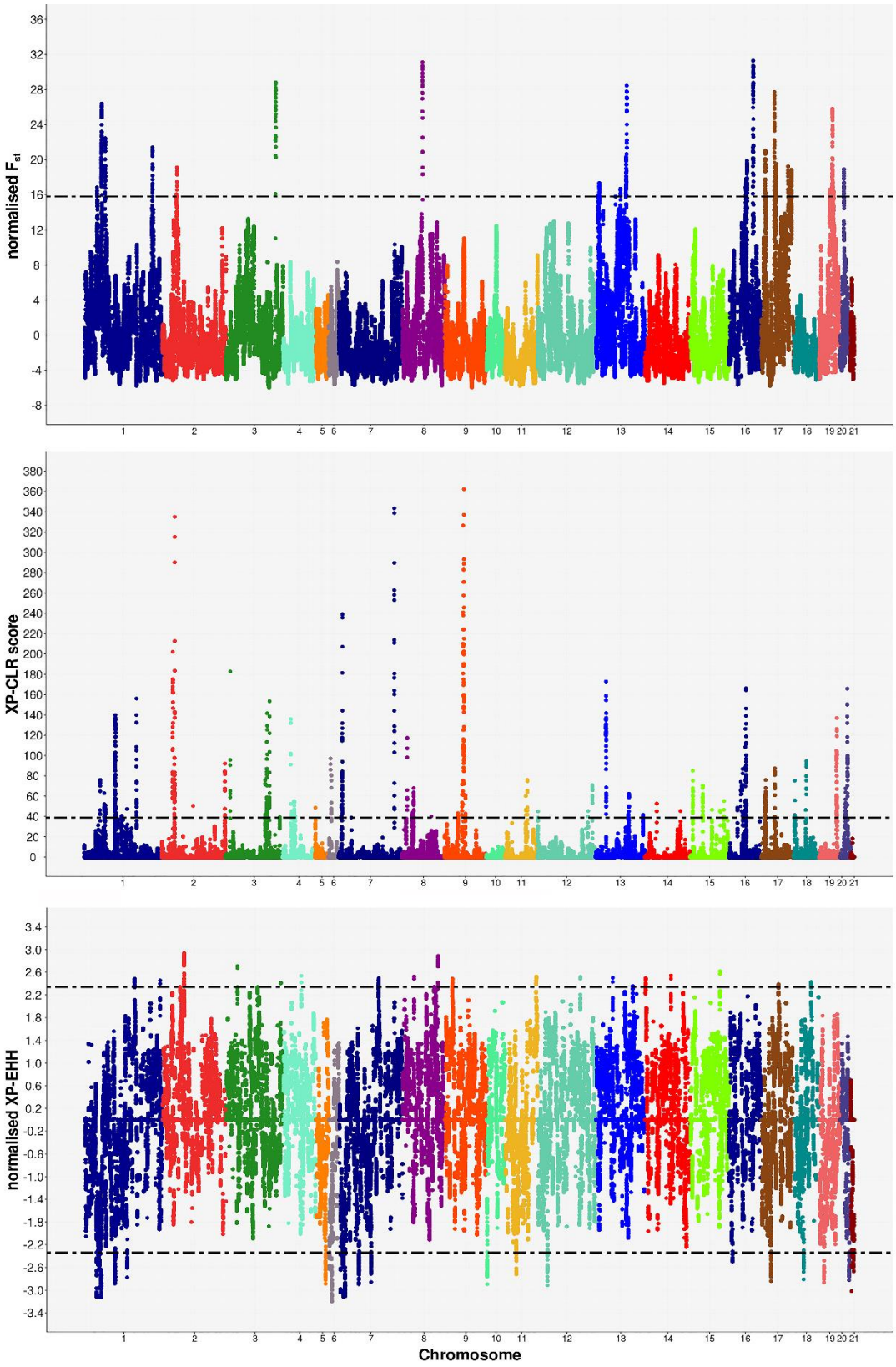


Figure 3. Manhattan plot of 500 kb window for every method of selection signatures in rabbits: normalised Wright's fixation index, F_{st} (top); cross population composite likelihood ratio test, XP-CLR score (middle); and normalised cross population extended haplotype homozygosity test, XP-EHH (bottom). The dashed line denotes the cut-off 1% (F_{st}: 15.81, XP-CLR: 38.94, and XP-EHH: +/- 2.34).

Table 2. SNP name (SNP_ID), rabbit chromosome (OCU), cluster (genomic region), SNP physical position in megabase and values for three methods of selection signatures based on detection in at least two methods using a cut-off of 1%.

SNP_ID	OCU	CLUSTER	Physical position	XP-CLR[1]	XP-EHH[2]	F_{st}[3]
Affx-151788669	1	1	34.40	**42.80**	**-2.78**	3.25
Affx-151841835	1		34.42	**46.32**	**-2.72**	4.05
Affx-151948493	1		34.42	**42.35**	**-2.72**	4.05
Affx-151981842	1		34.53	**70.13**	**-2.89**	7.58
Affx-151800050	1		34.55	**74.74**	**-2.88**	8.12
Affx-151888128	1		34.56	**42.29**	**-2.89**	8.10
Affx-151808312	1		34.57	**72.84**	**-2.88**	9.02
Affx-151996305	1		34.59	**76.09**	**-2.87**	9.25
Affx-151796600	1		34.60	**64.85**	**-2.86**	9.33
Affx-151996963	3	2	148.58	**39.42**	**2.41**	-1.34
Affx-151940966	6	3	6.58	**68.34**	**-2.89**	1.05
Affx-151916999	6		6.59	**47.82**	**-2.89**	1.11
Affx-151906393	6		6.60	**80.26**	**-2.88**	1.17
Affx-151850643	6		6.62	**75.42**	**-3.11**	1.62
Affx-152006617	6		6.63	**80.80**	**-3.07**	1.62
Affx-151909107	6		6.65	**53.68**	**-2.98**	1.86
Affx-151858638	7	4	7.85	**62.05**	**-2.48**	0.64
Affx-151988414	7		7.87	**48.46**	**-2.53**	1.11
Affx-151901134	7		7.89	**83.02**	**-2.54**	1.26
Affx-151884578	7		7.89	**80.85**	**-2.50**	1.26
Affx-151968222	7		7.91	**82.74**	**-2.53**	1.26
Affx-151923372	7		7.92	**82.83**	**-2.55**	1.55
Affx-151832398	7		7.95	**60.81**	**-2.41**	2.09
Affx-151887243	7		7.96	**78.92**	**-2.41**	2.09
Affx-151798377	7		7.99	**121.74**	**-2.38**	2.65
Affx-152002624	16	5	44.14	**56.15**	0.57	**16.01**
Affx-151964090	16		44.16	**76.86**	0.58	**16.05**
Affx-151954735	16		44.18	**56.19**	0.57	**16.63**
Affx-152011401	16		44.21	**49.99**	0.58	**16.44**
Affx-151994299	16		44.25	**76.42**	0.58	**16.82**
Affx-151935006	16		44.27	**44.46**	0.58	**16.87**
Affx-151934731	16		44.28	**39.12**	0.58	**16.91**
Affx-151916386	16		44.31	**41.23**	0.58	**17.23**
Affx-151945660	16		44.31	**49.81**	0.58	**17.05**
Affx-151892655	16		44.33	**85.41**	0.58	**16.87**
Affx-151923274	16		44.34	**139.09**	0.57	**16.92**
Affx-151981680	16		44.34	**164.39**	0.57	**16.92**
Affx-151922936	16		44.37	**166.24**	0.57	**16.42**
Affx-152012312	16		44.37	**146.40**	0.85	**16.23**
Affx-151904619	16		44.38	**111.16**	0.85	**16.83**
Affx-151947283	16		44.39	**78.58**	0.85	**16.83**
Affx-151892171	16		44.41	**47.52**	0.85	**17.38**

Affx-151999419	16		44.42	**50.66**	0.71	**17.38**
Affx-151831673	16		44.46	**53.22**	0.74	**16.96**
Affx-151877806	16		44.47	**52.99**	0.74	**17.19**
Affx-151933718	16		44.48	**52.90**	0.74	**17.41**
Affx-151961515	16		44.50	**51.22**	0.73	**17.85**
Affx-152004824	16		44.51	**50.14**	0.73	**17.67**
Affx-152008187	16		44.52	**74.51**	0.73	**17.46**
Affx-151886887	16		44.54	**93.91**	0.74	**17.25**
Affx-151900728	16		44.55	**100.41**	0.32	**17.25**
Affx-151875439	16		44.56	**96.80**	0.31	**17.25**
Affx-151786498	16		44.57	**118.82**	0.30	**16.81**
Affx-151820958	16		44.59	**131.41**	0.32	**16.86**
Affx-151942314	16		44.61	**112.63**	0.31	**16.40**
Affx-151797733	16		44.63	**100.38**	0.30	**15.93**
Affx-151854426	17		11.32	**47.27**	-1.68	**17.79**
Affx-152017855	17		11.45	**41.36**	-1.71	**20.90**
Affx-151809007	17		11.46	**40.34**	-1.71	**20.65**
Affx-151945077	17		11.49	**44.26**	-1.71	**20.79**
Affx-151827750	17		11.50	**45.09**	-1.71	**21.05**
Affx-151813388	17		11.51	**49.33**	-1.70	**20.79**
Affx-151897106	17		11.51	**60.52**	-1.71	**20.69**
Affx-151970040	17	6	11.52	**61.40**	-1.69	**20.95**
Affx-151854218	17		11.53	**60.99**	-1.69	**20.69**
Affx-151999939	17		11.54	**58.80**	-1.69	**20.70**
Affx-151872016	17		11.56	**56.37**	-1.69	**20.06**
Affx-151809616	17		11.57	**55.06**	-1.67	**19.42**
Affx-151800782	17		11.58	**54.38**	-1.69	**19.06**
Affx-151992875	17		11.59	**53.07**	-1.64	**18.71**
Affx-151953403	17		11.62	**46.73**	-1.64	**16.22**
Affx-151860917	17	7	29.59	**39.64**	-0.22	**19.14**
Affx-151984545	17		29.64	**39.42**	-0.40	**17.52**
Affx-151841455	17		30.42	**80.81**	-0.16	**15.89**
Affx-152009920	17		30.45	**39.82**	-0.27	**16.57**
Affx-151819416	17		30.55	**84.32**	0.06	**16.43**
Affx-151905376	17	8	30.56	**87.35**	-0.13	**16.43**
Affx-151933923	17		30.57	**83.99**	-0.12	**16.15**
Affx-151909639	17		30.57	**81.33**	0.74	**16.45**
Affx-151912729	17		30.59	**71.87**	0.79	**15.99**

SNP = Single nucleotide polymorphism. XP-CLR[1] = cross population composite likelihood ratio test. XP-EHH[2] = cross population extended haplotype homozygosity test. These values are normalised and F_{st}[3] = fixation index. The bold data and green cells indicate the values exceeding a cut-off of 1% (XP-CLR: 38.94, and XP-EHH: +/- 2.34, F_{st}: 15.81).

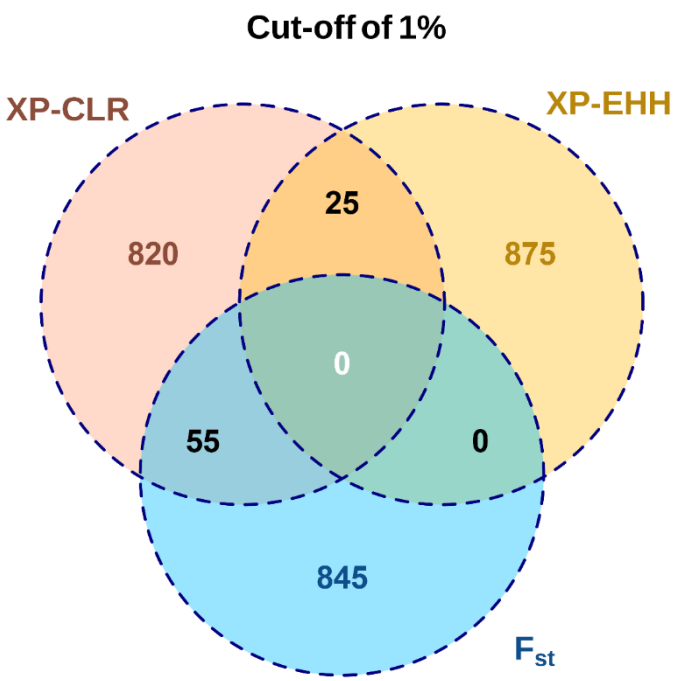


Figure 4. Venn diagram of methods of selection signatures in rabbits: cross population composite likelihood ratio test (XP-CLR), cross population extended haplotype homozygosity test (XP-EHH), and Wright's fixation index (F_{st}).

The selection signatures identified in our study can be a consequence of the selection of a polygenic trait with a high heritability such as the intramuscular fat (Martínez-Álvaro et al., 2016), or due to the further effect of genetic drift. The last hypothesis could be plausible because of the reduced number of sires used in the first generations, the mating structure (a female was mated with the same sire during its production life; Zomeño et al., 2013), and the increase of two sire families in the last generations in each line. In many ways, detection of selective sweeps in smaller populations is more difficult than in large populations as extensive drift can obscure and weaken the selection signatures (Mallick et al., 2009; Johansson et al., 2010). However, genetic drift would generate random shifts of allelic frequency across the rabbit genome and our results showed the existence of consecutives SNPs with high scores of selection signatures within methods which is a direct evidence of selection. For instance, cluster 5 in OCU16 presented high scores of normalised F_{st} (up to 17.85, Table 2) and a substantial length (487.49 kb) encompassing 31 SNPs. They could be identified because we employed windows for detecting selection signatures instead of evaluating each SNP of the rabbit array. In addition, under the F_{st} method and

a cut-off of 1%, a cluster in OCU13 encompassing 45 SNPs was identified by this study in a relevant genomic region (83.8 – 86.0 Mb) associated with intramuscular fat in rabbits, according to a genome-wide association study (GWAS) using the two lines of divergent selection (Sosa-Madrid et al., 2019). This region showed SNPs with high normalised F_{st}, reaching values up to 20.33 (0.51 as F_{st} mean of 500 kb windows), albeit with only one of these SNPs, Affx-151937959, agreed with the relevant SNPs reported by GWAS (Sosa-Madrid et al., 2019). This SNP showed a low MAF (0.09), but the surrounding SNPs presented very high MAF (up to 0.44). The methods of selection signatures can validate GWAS results assuming the presence of major genes affecting a selected trait. Otherwise, these methods would reveal new associated genomic regions, unlike GWAS results, when the selected trait has a large polygenic component influencing several genomic regions (Qanbari and Simianer, 2014).

On the other hand, divergent selection for intramuscular fat did not lead to fixation of alternate alleles of any of the SNPs studied. The selection response was very high (3.1 SD), hence, we expected some SNPs associated with causal variants had their alternate alleles fixed o nearby fixation in one of the opposite divergent lines (*e.g.* frequencies in high line: 0 A / 1 T, and in low line: 1 A / 0 T). These SNPs would show MAF values of 0.5 using all samples (both lines). However, the SNPs did not show both conditions. All these results would suggest several soft selective sweeps caused by short-term divergent selection of intramuscular fat instead of few hard selective sweeps, controlling this trait (Oleksyk et al., 2010).

5.5.2. Underlying selected genes and gene ontology terms for divergent selection

Potential candidate genes were explored within the genomic regions identified as signatures of selection using a cut-off of 1%. The number of genes disclosed for each method were 579, 443, and 368 for XP-CLR, XP-EHH and F_{st}, respectively (see Additional file 1: Table S1). From these genes, 73 were detected by at least two methods of selection signatures. These genes were grouped in 63 protein coding and 10 non-protein coding genes (see Additional file 2: Table S2). The results of the first ten biological processes of the GO term enrichment analysis through *Enrichr* are presented in Additional file 3: Table S3. The GO

term enrichment analysis did not identified pathways related to biological processes affecting the intramuscular fat.

Table 3. Biological processes of highlighted genes identified by at least two methods of selection signatures in rabbits for cut-off of 1%.

Biological Process	Highlighted genes
Insulin signalling pathways	*DENND4C*
Lipid droplets and storage	*VIM, PLIN2, ACER2*
mTORC signalling pathways	*RRAGA, RPS6*
Carbohydrate metabolism process	*RORA, ST8SIA6, GANC*
Lipid metabolic process	*RORA, ACER2, PLA2G4B*
Regulation of adipocytes differentiation	*RORA*
Phospholipase activity	*PLA2G4B*
Processes related to intramuscular fat*	*PLIN2, SLC24A2, RTF1, RORA*

* Based on genomic and gene expression studies of intramuscular fat.

A deep search of biological functions for the 73 genes disclosed 12 promising candidate genes related to lipid and carbohydrate metabolisms which are important pathways to modulate the intramuscular fat (Table 3). Genes involved in the lipid metabolism were alkaline ceramidase 2 (*ACER2*), Perilipin 2 (*PLIN2*), Vitemin (*VIM*), Ras related GTP binding A (*RRGA*), ribosomal protein S6 (*RPS6*), RNA Polymerase-Associated Protein RTF1 Homolog (*RTF1*), solute-carrier gene family 24 member 2 (*SLC24A2*) and phospholipase A2 group IVB (*PLA2G4B*). From these, it is worth to highlight *ACER2* (OCU1), *VIM* (OCU16) and *PLIN2* (OCU1), which are tightly related to lipid droplets and storage, being crucial in disease such as obesity, diabetes and atherosclerosis. *ACER2* encoding ceramidases and broken-down ceramides to sphingosine and free fatty acids at alkaline pH. *VIM* can cause an excessive endosomal cholesterol accumulation due to an imbalance of its iterations with other proteins (Walter et al., 2009). *PLIN2* bears an essential role over long-chain fatty acid transport. Genomic

studies reported *PLIN2* associations with intramuscular fat content (Gandolfi et al., 2011) and its composition in pigs (Gol et al., 2016). Moreover, gene expression studies for intramuscular fat in pig identified differentially expressed genes such as *RTF1* in OCU17 (Damon et al., 2012) and *SLC24A2* in OCU1 (Li et al., 2010). *RRGA* and *RPS6* in OCU1 could stimulate the lipogenesis and the lipid accumulation via activation of the mTOR signalling pathways (Wipperman et al., 2019). *PLA2G4B* in OCU17 is linked to phospholipid catabolic processes because of its phospholipase A2 (PL2) activity. This enzyme has been widely studied using knockout and transgenic mice, showing to be important for the fatty acid pathway, *e.g.* for oleic acid (Aloulou et al., 2012).

Phenotypic variation of intramuscular fat between divergent lines could also be due to differences in regulation of lipid and carbohydrates (glycogen) metabolisms. This latter is important for intramuscular fat as the glycolytic products could be used to synthesize fatty acids, being incorporated into cholesterol esters, triacylglycerol, and phospholipids in hepatocytes, increasing the lipid stores (Rui, 2014). Genes involved in the carbohydrate metabolism were RAR related orphan receptor A (*RORA*), glucosidae alpha neutral C (*GANC*), ST8 alpha-N-acetyl-neuraminide alpha-2,8-sialytransferase 6 (*ST8SIA6*) and DENN domain containing 4C (*DENND4C*). The regulation of differentiating pre-adipocytes by retinoic acid is controlled by *RORA* in OCU1, bearing a crucial role in triglyceride (lipids) / glucose homeostasis and various immune functions. The *RORA* functions are tightly related to hepatosteatosis, obesity, and insulin resistance. Besides, *RORA* was identified by genomic studies in Nellore (Cesar et al., 2014) and Chinese Wagyu cattle (Wang et al., 2019) having extreme phenotypes of intramuscular fat composition and marbling. Hence, we presented *RORA* as the principal candidate gen for further studies. In addition, other genes as *GANC* in OCU17 is involved in the hydrolysis of glycogen; and *ST8SIA6* in OCU16 is important in the pathways of oligosaccharide metabolic process and carbohydrate biosynthetic process. *DENND4C* in OCU1 could be indirectly modulate the intramuscular fat content through control of glucose transport in response to insulin. However, the specific functions of these genes over the intramuscular fat remains unknown. Further analyses would be needed to corroborate the relationships between these genes (their polymorphisms) and the intramuscular fat in rabbits.

5.6. Conclusions

In conclusion, a large number of genomic regions were identified within each method of selection signatures. A total number of 80 SNPs and 73 genes were detected using selection signatures exceeding cut-off of 1% at least two of the methods; XP-CLR, XP-EHH, and F_{st}. General biological functions were retrieved from enrichment analysis. However, genes such as *ACER2*, *PLIN2* (OCU1), *ST8SIA6*, *VIM* (OCU16), *RORA*, *GANC* and *PLA2G4B* (OCU17) linked to energy metabolism, carbohydrates metabolism, and lipid metabolism were identified as candidate genes to explain the differences in intramuscular fat observed between the divergent lines. The findings of the current study suggest that the intramuscular fat in rabbits is influenced by a large polygenic component.

5.7. Abbreviations

MAF: minor allele frequency; OCU: rabbit chromosome; MDS: multidimensional scaling; XP-CLR: cross population composite likelihood ratio test; XP-EHH: cross population extended haplotype homozygosity test; F_{st}: Wrigth's fixation index; GO: gen ontology; EHH: extended haplotype homozygosity; SNP: single nucleotide polymorphism; GWAS: genome-wide association study; *ACER2*: alkaline ceramidase 2; *PLIN2*: Perilipin 2; *VIM*: Vitemin; *RRGA*: Ras related GTP binding A; *RPS6*: ribosomal protein S6; *RTF1*: RNA Polymerase-Associated Protein RTF1 Homolog, *SLC24A2*: solute-carrier gene family 24 member 2; *PLA2G4B*: phospholipase A2 group IVB; *RORA*: RAR related orphan receptor A; *GANC*: glucosidae alpha neutral C; *ST8SIA6*: ST8 alpha-N-acetyl-neuraminide alpha-2,8-sialytransferase 6; *DENND4C*: DENN domain containing 4C.

5.8. References

Aloulou, A., Y. Ben Ali, S. Bezzine, Y. Gargouri, and M. H. Gelb, 2012. "Phospholipases: An Overview", pp. 63–85 in *Lipases and Phospholipases: Methods and Protocols,* edited by G. Sandoval. Humana Press, Totowa, NJ.

Beissinger, T. M., G. J. Rosa, S. M. Kaeppler, D. Gianola, and N. de Leon, 2015. "Defining window-boundaries for genomic analyses using smoothing spline techniques." *Genetics Selection Evolution,* 47(1): 30.

Carneiro, M., F. W. Albert, S. Afonso, R. J. Pereira, H. Burbano et al., 2014a. "The genomic architecture of population divergence between subspecies of the European Rabbit." *PLoS Genetics*, 10(8): e1003519.

Carneiro, M., C. J. Rubin, F. Di Palma, F. W. Albert, J. Alföldi et al., 2014b. "Rabbit genome analysis reveals a polygenic basis for phenotypic change during domestication." *Science*, 345(6200): 1074–1079.

Cesar, A. S., L. C. Regitano, G. B. Mourão, R. R. Tullio, D. P. Lanna et al., 2014. "Genome-wide association study for intramuscular fat deposition and composition in Nellore cattle." *BMC Genetics*, 15(1): 39.

Chen, H., N. Patterson, and D. Reich, 2010. "Population differentiation as a test for selective sweeps." *Genome Research*, 20(3): 393–402.

Damon, M., J. Wyszynska-Koko, A. Vincent, F. Hérault and B. Lebret, 2012. "Comparison of Muscle Transcriptome between Pigs with Divergent Meat Quality Phenotypes Identifies Genes Related to Muscle Metabolism and Structure." *PLoS ONE*, 7(3): e33763.

European Economic Community, 1998. "Council Directive 98/58/EC of 20 July 1998 concerning the protection of animals kept for farming purposes." *Official Journal L* 221(8): 8.

Gandolfi, G., M. Mazzoni, P. Zambonelli, G. Lalatta-Costerbosa, A. Tronca et al., 2011. "Perilipin 1 and perilipin 2 protein localization and gene expression study in skeletal muscles of European cross-breed pigs with different intramuscular fat contents." *Meat Science,* 88(4): 631–637.

Gol, S., R. Ros-Freixedes, P. Zambonelli, M. Tor, R. N. Pena, et al., 2016. "Relationship between perilipin genes polymorphisms and growth, carcass and meat quality traits in pigs." *Journal of Animal Breeding and Genetics,* 133(1): 24–30.

González-Rodríguez, A., S. Munilla, E. F. Mouresan, J. J. Cañas-Álvarez, C. Díaz et al., 2016. "On the performance of tests for the detection of signatures of selection: a case study with the Spanish autochthonous beef cattle populations." *Genetics Selection Evolution*, 48(1): 81.

Grams, V., R. Wellmann, S. Preuß, M. A. Grashorn, J. B. Kjaer et al., 2015. "Genetic parameters and signatures of selection in two divergent laying hen lines selected for feather pecking behaviour." *Genetics Selection Evolution*, 47(1): 77.

Gurgul, A., I. Jasielczuk, K. Ropka-Molik, E. Semik-Gurgul, K. Pawlina-Tyszko et al., 2018. "A genome-wide detection of selection signatures in conserved

and commercial pig breeds maintained in Poland." *BMC Genetics,* 19(1): 95.

Johansson, A. M., M. E. Pettersson, P. B. Siegel, and Ö. Carlborg, 2010. "Genome-Wide Effects of Long-Term Divergent Selection." *PLoS Genetics,* 6(11): e1001188.

Jump, D. B., D. Botolin, Y. Wang, J. Xu, O. Demeure et al., 2008. "Docosahexaenoic acid (DHA) and hepatic gene transcription." *Chemistry and Physics of Lipids,* 153(1): 3–13.

Kim, E. S., R. Ros-Freixedes, R. N. Pena, T. J. Baas, J. Estany et al., 2015. "Identification of signatures of selection for intramuscular fat and backfat thickness in two Duroc populations." *Journal of Animal Science,* 93(7): 3292–3302.

Kuleshov, M. V., M. R. Jones, A. D. Rouillard, N. F. Fernandez, Q. Duan et al., 2016. "Enrichr: a comprehensive gene set enrichment analysis web server 2016 update." *Nucleic Acids Research,* 44(W1): W90–W97.

Li, X., C.-K. Lee, B.-H. Choi, T.-H. Kim, J.-J. Kim et al., 2010. "Quantitative gene expression analysis on chromosome 6 between Korean native pigs and Yorkshire breeds for fat deposition." *Genes & Genomics,* 32(4): 385–393.

Lillie, M., Z. Sheng, C. F. Honaker, B. J. Dorshorst, C. M. Ashwell et al., 2017. "Genome-wide standing variation facilitates long-term response to bidirectional selection for antibody response in chickens." *BMC Genomics*, 18(1): 99.

Ma, H., S. Zhang, K. Zhang, H. Zhan, X. Peng et al., 2019. "Identifying selection signatures for backfat thickness in Yorkshire pigs highlights new regions affecting fat metabolism." *Genes,* 10(4): 254.

Mallick, S., S. Gnerre, P. Muller, and D. Reich, 2009. "The difficulty of avoiding false positives in genome scans for natural selection." *Genome Research*, 19(5): 922–933.

Martínez-Álvaro, M., P. Hernández, and A. Blasco, 2016. "Divergent selection on intramuscular fat in rabbits: Responses to selection and genetic parameters." *Journal of Animal Science*, 94(12): 4993–5003.

Mauch, E., B. Servin, H. Gilbert, and J. Dekkers, 2018. "Signatures of selection in two independent populations of pigs divergently selected for feed efficiency." *Animal Industry Report,* AS 664(1): AS 3274.

Oleksyk, T. K., M. W. Smith, and S. J. O'Brien, 2010. "Genome-wide scans for footprints of natural selection." *Philosophical Transactions of the Royal Society B: Biological Science*, 365(1537): 185–205.

Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira et al., 2007. "PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses." *American Journal of Human Genetics,* 81: 559–575.

Qanbari, S., and H. Simianer, 2014. "Mapping signatures of positive selection in the genome of livestock." *Livestock Science,* 166: 133–143.

Rui, L., 2014. "Energy metabolism in the liver." *Comprehensive Physiology*, 176(5): 177–197.

Sabeti, P. C., P. Varilly, B. Fry, J. Lohmueller, E. Hostetter et al., 2007. "Genome-wide detection and characterization of positive selection in human populations." *Nature,* 449(7164): 913–918.

Sargolzaei, M., J. P. Chesnais, and F. S. Schenkel, 2014. "A new approach for efficient genotype imputation using information from relatives." *BMC Genomics,* 15(1): 478.

Sosa-Madrid, B. S., P. Hernández, A. Blasco, C. H. Haley, L. Fontanesi et al., 2019. "Genomic regions influencing intramuscular fat in divergently selected rabbit lines." *Animal Genetics,* doi: https://doi.org/10.1111/age.12873, Published by John Wiley & Sons Ltd 7 November 2019.

Sosa-Madrid, B. S., N. Ibáñez-Escriche, M. A. Santacreu, L. Varona, and A. Blasco, 2017. "Huellas de selección en un experimento de selección divergente para capacidad uterina en conejo." In Proceedings of the XVII *Jornadas sobre Producción Animal*, 30-31 May 2017, Zaragoza, Spain, pp. 558–560.

Szpiech, Z. A., and R. D. Hernandez, 2014. "selscan: an efficient multithreaded program to perform EHH-based scans for positive selection." *Molecular Biological Evolution,* 31(10): 2824–2827.

Utsunomiya, Y. T., A. M. Pérez O'Brien, T. S. Sonstegard, C. P. Van Tassell, A. S. do Carmo et al., 2013. "Detecting loci under recent positive selection in dairy and beef cattle by combining different genome-wide scan methods." *PLoS ONE*, 8(5):e64280.

Walter, M., F. W. Chen, F. Tamari, R. Wang and Y. A. Ioannou, 2009. "Endosomal lipid accumulation in NPC1 leads to inhibition of PKC, hypophosphorylation of vimentin and Rab9 entrapment." *Biology of the Cell*, 101(3): 141–153.

Wang, Z., H. Ma, L. Xu, B. Zhu, Y. Liu, F. Bordbar et al., 2019. "Genome-wide scan identifies selection signatures in Chinese Wagyu cattle using a high-density SNP array." *Animals,* 9(6): 296.

Wipperman M. F., D. C. Montrose, A. M. Gotto and D. P. Hajjar, 2019. "Mammalian Target of Rapamycin." *The American Journal of Pathology*, 189(3): 492–501.

Zomeño, C., P. Hernandez, and A. Blasco, 2013. "Divergent selection for intramuscular fat content in rabbits. I. Direct response to selection." *Journal of Animal Science,* 91(9): 4526–4531.

5.9. Additional Files

Additional file 1: Table S1. Top five GO terms of biological processes using cut-off of 1% for each method of selection signatures in rabbits.

Method	XP-CLR				
Genes	579				
Name	**GO term**	**P-value**	**Odds ratio**	**Combined Score**	**Genes**
ribosomal small subunit export from nucleus	GO:0000056	6.74E-03	15.26	76.27	*LTV1; RPS15*
ventricular compact myocardium morphogenesis	GO:0003223	6.74E-03	15.26	76.27	*DSP; BMPR1A*
regulation of anion channel activity	GO:0010359	6.74E-03	15.26	76.27	*TCAF2; TCAF1*
maintenance of protein location in extracellular region	GO:0071694	6.74E-03	15.26	76.27	*FBN2; NBL1*
glutamate catabolic process	GO:0006538	6.74E-03	15.26	76.27	*GLUD1; GLUD2*
Method	XP-EHH				
Genes	443				
Name	**GO term**	**P-value**	**Odds ratio**	**Combined Score**	**Genes**
regulation of osteoclast development	GO:2001204	4.16E-03	19.55	107.20	*NOTCH2; FBN1*
detection of chemical stimulus involved in sensory perception	GO:0050907	1.90E-07	6.83	105.77	*OR4A8; OR4C12; OR4C45; OR4C11; OR4A47; OR4P4; OR4C15; OR2A4; OR4A15; OR4C13; OR4S2; OR4C46*

aminoglycoside antibiotic metabolic process	GO:0030647	7.58E-03	14.66	71.58	*AKR1B10; AKR1B1*
daunorubicin metabolic process	GO:0044597	7.58E-03	14.66	71.58	*AKR1B10; AKR1B1*
positive regulation of TORC1 signalling	GO:1904263	7.58E-03	14.66	71.58	*RRAGA; LARS*
Method	normalised-F_{st}				
Genes	368				
Term	**GO term**	**P-value**	**Odds ratio**	**Combined Score**	**Genes**
citrate metabolic process	GO:0006101	3.43E-03	21.57	122.47	*STAT5A; STAT5B*
interleukin-15-mediated signalling pathway	GO:0035723	9.31E-04	14.94	104.24	*STAT5A; STAT5B; STAT3*
cellular response to interleukin-15	GO:0071350	9.31E-04	14.94	104.24	*STAT5A; STAT5B; STAT3*
oxaloacetate metabolic process	GO:0006107	4.75E-03	18.49	98.94	*STAT5A; STAT5B*
valine metabolic process	GO:0006573	4.75E-03	18.49	98.94	*STAT5A; STAT5B*

Gene Ontology (GO) knowledgebase is the world's largest source of information on the functions of genes. Odds ratios are used to compare the relative odds of the occurrence of the outcome of interest given exposure to the variable of interest. The odds ratio is calculated using the analysed genes and GO terms of the datable. XP-CLR = cross population – composite likelihood ratio. XP-EHH = cross population – extended haplotype homozygosity. F_{st} = Wright's fixation index. TORC1 = mammalian target of rapamycin complex 1.

Additional file 2: Table S2. Annotated genes in the selection signatures of intramuscular fat in rabbits.

Gene stable ID	Gene description	Gene name	Chromosome/ scaffold name	Gene start (bp)	Gene end (bp)	Gene type
ENSOCUG00000002462	solute carrier family 24 member 2 [Source:HGNC Symbol;Acc:HGNC:10976]	SLC24A2	1	34121351	34381731	protein coding
ENSOCUG00000012439	alkaline ceramidase 2 [Source:HGNC Symbol;Acc:HGNC:23675]	ACER2	1	34454721	34501675	protein coding
ENSOCUG00000025352	40S ribosomal protein S4, X isoform-like [Source:NCBI gene;Acc:100338112]	Rps4x	1	34520162	34520953	protein coding
ENSOCUG00000003336	ribosomal protein S6 [Source:HGNC Symbol;Acc:HGNC:10429]	RPS6	1	34527129	34530485	protein coding
ENSOCUG00000003320	DENN domain containing 4C [Source:HGNC Symbol;Acc:HGNC:26079]	DENND4C	1	34532623	34658797	protein coding
ENSOCUG00000023552			1	34565805	34565914	snoRNA
ENSOCUG00000037713			1	34659452	34697576	lncRNA
ENSOCUG00000012244	perilipin 2 [Source:HGNC Symbol;Acc:HGNC:248]	PLIN2	1	34734595	34759320	protein coding
ENSOCUG00000003507	HAUS augmin like complex subunit 6 [Source:HGNC Symbol;Acc:HGNC:25948]	HAUS6	1	34761804	34814946	protein coding
ENSOCUG00000028955	small Cajal body-specific RNA 8 [Source:HGNC Symbol;Acc:HGNC:32564]	SCARNA8	1	34804610	34804740	scaRNA
ENSOCUG00000003506	Ras related GTP binding A [Source:HGNC Symbol;Acc:HGNC:16963]	RRAGA	1	34823569	34824510	protein coding
ENSOCUG00000012362	stabilizer of axonemal microtubules 1 [Source:HGNC Symbol;Acc:HGNC:28566]	SAXO1	1	34838189	34990237	protein coding
ENSOCUG00000009754	leucine rich repeat containing 6 [Source:HGNC Symbol;Acc:HGNC:16725]	LRRC6	3	148347222	148453432	protein coding

Gene stable ID	Gene description	Gene name	Chromosome/ scaffold name	Gene start (bp)	Gene end (bp)	Gene type
ENSOCUG00000034557			3	148473001	148476496	lncRNA
ENSOCUG00000009761	transmembrane protein 71 [Source:HGNC Symbol;Acc:HGNC:26572]	TMEM71	3	148493203	148534309	protein coding
ENSOCUG00000009768	PHD finger protein 20 like 1 [Source:HGNC Symbol;Acc:HGNC:24280]	PHF20L1	3	148550576	148623229	protein coding
ENSOCUG00000034085			3	148659556	148661451	protein coding
ENSOCUG00000028722			3	148715973	148716105	snoRNA
ENSOCUG00000002999	thyroglobulin [Source:HGNC Symbol;Acc:HGNC:11764]	TG	3	148726297	148867431	protein coding
ENSOCUG00000017684	Src like adaptor [Source:HGNC Symbol;Acc:HGNC:10902]	SLA	3	148781918	148840524	protein coding
ENSOCUG00000021148	shisa family member 9 [Source:HGNC Symbol;Acc:HGNC:37231]	SHISA9	6	6114954	6432579	protein coding
ENSOCUG00000038543			6	6437084	6533361	lncRNA
ENSOCUG00000031076			6	6690408	6711660	lncRNA
ENSOCUG00000003406	thiamin pyrophosphokinase 1 [Source:HGNC Symbol;Acc:HGNC:17358]	TPK1	7	7412937	7795934	protein coding
ENSOCUG00000001027	NOBOX oogenesis homeobox [Source:HGNC Symbol;Acc:HGNC:22448]	NOBOX	7	7838974	7844349	protein coding
ENSOCUG00000025879	Rho guanine nucleotide exchange factor 5 [Source:HGNC Symbol;Acc:HGNC:13209]	ARHGEF5	7	7861292	7885252	protein coding

Gene stable ID	Gene description	Gene name	Chromosome/ scaffold name	Gene start (bp)	Gene end (bp)	Gene type
ENSOCUG00000010368	olfactory receptor 2A1/2A42 [Source:NCBI gene;Acc:100356551]	OR2A1	7	7929276	7930091	protein coding
ENSOCUG00000021687	olfactory receptor 13 [Source:NCBI gene;Acc:100356800]	Olfr13	7	7939208	7940140	protein coding
ENSOCUG00000006436	kinetochore protein Spc25-like [Source:NCBI gene;Acc:100357315]		7	7973609	7974289	protein coding
ENSOCUG00000034348			7	7984992	7986230	protein coding
ENSOCUG00000034276	neuroblastoma breakpoint family member 9 [Source:NCBI gene;Acc:103348625]		7	7989864	8006238	protein coding
ENSOCUG00000025612	olfactory receptor family 2 subfamily A member 14 [Source:HGNC Symbol;Acc:HGNC:15084]	OR2A14	7	8013085	8014011	protein coding
ENSOCUG00000032840			7	8017830	8039557	lncRNA
ENSOCUG00000022708			7	8040429	8041372	protein coding
ENSOCUG00000021482			7	8069582	8079336	protein coding
ENSOCUG00000036352			7	8094274	8095485	lncRNA
ENSOCUG00000021016	olfactory receptor family 6 subfamily B member 1 [Source:HGNC Symbol;Acc:HGNC:8354]	OR6B1	7	8129584	8133013	protein coding
ENSOCUG00000001333	olfactory receptor 2F1-like [Source:NCBI gene;Acc:100357829]	Olfr451-ps1	7	8174630	8295709	protein coding
ENSOCUG00000034930			7	8201202	8201957	protein coding

Gene stable ID	Gene description	Gene name	Chromosome/ scaffold name	Gene start (bp)	Gene end (bp)	Gene type
ENSOCUG00000007240	ST8 alpha-N-acetyl-neuraminide alpha-2,8-sialyltransferase 6 [Source:HGNC Symbol;Acc:HGNC:23317]	ST8SIA6	16	43862449	44016871	protein coding
ENSOCUG00000009222	vimentin [Source:HGNC Symbol;Acc:HGNC:12692]	VIM	16	44091745	44099432	protein coding
ENSOCUG00000007235	tRNA aspartic acid methyltransferase 1 [Source:HGNC Symbol;Acc:HGNC:2977]	TRDMT1	16	44124048	44183627	protein coding
ENSOCUG00000029667	cubilin [Source:HGNC Symbol;Acc:HGNC:2548]	CUBN	16	44213239	44491445	protein coding
ENSOCUG00000030399	Ras suppressor protein 1 [Source:HGNC Symbol;Acc:HGNC:10464]	RSU1	16	44502655	44709091	protein coding
ENSOCUG00000001991	complement C1q like 3 [Source:HGNC Symbol;Acc:HGNC:19359]	C1QL3	16	44748685	44780163	protein coding
ENSOCUG00000001985	phosphotriesterase related [Source:HGNC Symbol;Acc:HGNC:9590]	PTER	16	44783349	44860212	protein coding
ENSOCUG00000012900	RAR related orphan receptor A [Source:HGNC Symbol;Acc:HGNC:10258]	RORA	17	11702755	11891409	protein coding
ENSOCUG00000011046	tau tubulin kinase 2 [Source:HGNC Symbol;Acc:HGNC:19141]	TTBK2	17	29360750	29472877	protein coding
ENSOCUG00000037370			17	29475638	29475829	protein coding
ENSOCUG00000011031	codanin 1 [Source:HGNC Symbol;Acc:HGNC:1713]	CDAN1	17	29480384	29491285	protein coding
ENSOCUG00000011018	StAR related lipid transfer domain containing 9	STARD9	17	29495468	29615866	protein coding
ENSOCUG00000011004	HAUS augmin like complex subunit 2 [Source:HGNC Symbol;Acc:HGNC:25530]	HAUS2	17	29621687	29642371	protein coding

Gene stable ID	Gene description	Gene name	Chromosome/ scaffold name	Gene start (bp)	Gene end (bp)	Gene type
ENSOCUG00000010999	leucine rich repeat containing 57 [Source:NCBI gene;Acc:100342146]	Lrrc57	17	29642835	29647754	protein coding
ENSOCUG00000010995	synaptosome associated protein 23 [Source:NCBI gene;Acc:100008776]	SNAP23	17	29652558	29670490	protein coding
ENSOCUG00000016382	zinc finger protein 106 [Source:HGNC Symbol;Acc:HGNC:12886]	ZNF106	17	29719749	29777468	protein coding
ENSOCUG00000024481	calpain 3 [Source:NCBI gene;Acc:100008726]	CAPN3	17	29781670	29828844	protein coding
ENSOCUG00000011176	glucosidase alpha, neutral C [Source:NCBI gene;Acc:100343672]	Ganc	17	29836357	29914449	protein coding
ENSOCUG00000015759	EH domain containing 4 [Source:HGNC Symbol;Acc:HGNC:3245]	EHD4	17	30179178	30253268	protein coding
ENSOCUG00000011997	spectrin beta, non-erythrocytic 5 [Source:HGNC Symbol;Acc:HGNC:15680]	SPTBN5	17	30257315	30296213	protein coding
ENSOCUG00000011984	phospholipase A2 group IVB [Source:NCBI gene;Acc:100533118]	PLA2G4B	17	30299312	30307757	protein coding
ENSOCUG00000035726	jumonji domain containing 7 [Source:NCBI gene;Acc:100341549]	JMJD7	17	30308417	30313754	protein coding
ENSOCUG00000011969	mitogen-activated protein kinase binding protein 1 [Source:HGNC Symbol;Acc:HGNC:29536]	MAPKBP1	17	30315910	30365469	protein coding
ENSOCUG00000006505	MAX dimerization protein MGA [Source:HGNC Symbol;Acc:HGNC:14010]	MGA	17	30372251	30449268	protein coding
ENSOCUG00000006229	TYRO3 protein tyrosine kinase [Source:HGNC Symbol;Acc:HGNC:12446]	TYRO3	17	30568857	30582861	protein coding
ENSOCUG00000007409	RNA polymerase II associated protein 1 [Source:HGNC Symbol;Acc:HGNC:24567]	RPAP1	17	30601127	30614693	protein coding

Gene stable ID	Gene description	Gene name	Chromosome/ scaffold name	Gene start (bp)	Gene end (bp)	Gene type
ENSOCUG00000026867	leukocyte receptor tyrosine kinase [Source:HGNC Symbol;Acc:HGNC:6721]	LTK	17	30616839	30624998	protein coding
ENSOCUG00000023580	inositol-trisphosphate 3-kinase A [Source:HGNC Symbol;Acc:HGNC:6178]	ITPKA	17	30625711	30627736	protein coding
ENSOCUG00000000278	RTF1 homolog, Paf1/RNA polymerase II complex component [Source:HGNC Symbol;Acc:HGNC:28996]	RTF1	17	30643777	30693176	protein coding
ENSOCUG00000004206	NADH:ubiquinone oxidoreductase complex assembly factor 1 [Source:HGNC Symbol;Acc:HGNC:18828]	NDUFAF1	17	30719275	30731356	protein coding
ENSOCUG00000003277	nucleolar and spindle associated protein 1 [Source:NCBI gene;Acc:100349779]	Nusap1	17	30737649	30774531	protein coding
ENSOCUG00000003274	Opa interacting protein 5 [Source:HGNC Symbol;Acc:HGNC:20300]	OIP5	17	30774845	30791085	protein coding
ENSOCUG00000013832	calcineurin like EF-hand protein 1 [Source:HGNC Symbol;Acc:HGNC:17433]	CHP1	17	30813272	30868942	protein coding
ENSOCUG00000028255	U2 spliceosomal RNA [Source:RFAM;Acc:RF00004]	U2	17	30826486	30826562	snRNA

Mouse and human orthologues (>80% querry genes identical to target gen of this species)

Additional file 3: Table S3. First ten go ontology (GO) terms of biological processes using genes located in genomic regions that exceeded a cut-off of 1% in at least two methods of selection signatures in rabbits.

GO term's CODE	Term	P-value	Odds Ratio	Combined Score	Annotated Mouse Genes
GO:0072528	pyrimidine-containing compound biosynthetic process	1.67E-02	59.52	243.65	*TPK1*
GO:0070417	cellular response to cold	1.67E-02	59.52	243.65	*SAXO1*
GO:0010665	regulation of cardiac muscle cell apoptotic process	1.67E-02	59.52	243.65	*LTK*
GO:0044003	modification by symbiont of host morphology or physiology	1.94E-02	51.02	201.04	*RRAGA*
GO:1904526	regulation of microtubule binding	1.94E-02	51.02	201.04	*TTBK2*
GO:2000483	negative regulation of interleukin-8 secretion	1.94E-02	51.02	201.04	*MAPKBP1*
GO:0042723	thiamine-containing compound metabolic process	1.94E-02	51.02	201.04	*TPK1*
GO:0033629	negative regulation of cell adhesion mediated by integrin	1.94E-02	51.02	201.04	*ACER2*
GO:0060292	long term synaptic depression	1.94E-02	51.02	201.04	*SLC24A2*
GO:0060850	regulation of transcription involved in cell fate commitment	1.94E-02	51.02	201.04	*RORA*

Gene Ontology (GO) knowledgebase is the world's largest source of information on the functions of genes. Odds ratios are used to compare the relative odds of the occurrence of the outcome of interest given exposure to the variable of interest. The odds ratio is calculated using the analysed genes and GO terms of the datable.

CHAPTER SIX

6. GENERAL DISCUSSION

This thesis is focused on the identification of genetic markers, genomic regions and genes associated with litter size traits and intramuscular fat (IMF). The samples came from two independent experiments of divergent selection in rabbits for uterine capacity (UC) (Blasco et al., 2005) and for IMF (Martínez-Álvaro et al., 2016), respectively. Divergent selection experiments increase the detection power, according to simulation studies (Kessner & Novembre, 2015; Lou et al., 2019). However, it is rare to find most causative variants explaining a great part of the variability of a given trait through a unique genomic analysis, *e.g.* genome wide association study – GWAS (López de Maturana et al., 2014; Schmid & Bennewitz, 2017; Georges et al., 2019). Besides, the identification of genetic markers and genes depends on the genomic architecture of a trait: number and effect size of causative variants, interaction effects amongst these variants, and distribution of genetic markers and causative variants across the rabbit genome (Lou et al., 2019; Georges et al., 2019). Thus, we used both GWAS and genome scans studies approaches to identify relevant genetic markers and genes associated with each analyzed trait in this thesis. We initially proposed three scenarios for our divergent selection experiments, taking into account different genomic architectures:

(a) The trait presents *few causative variants of large effect* (less than five) and some causative variants with small effect (between 50 and 100). The divergent selection caused shifts of allelic frequencies in the causative variants with large effect. The opposite alleles of *most of these causative variants* are fixed or nearby fixation *in each line* of divergent selection (one allele in the high line and another allele in the low line).

(b) The trait presents *several causative variants of small effect* (greater than 100) and no causative variants with large effect. The divergent selection caused shifts of allelic frequencies in *most causative variants with small effect*. The opposite alleles of these causatives variants are fixed or nearby fixation *in each line* of divergent selection.

(c) The trait presents *several causative variants of small effect* (greater than 100) and no causative variants with large effect. The divergent selection caused noticeable shifts of allelic frequencies in *a few causative variants*

with small effect, since the opposite alleles of these causative variants are fixed or nearby fixation *in one line but not in the other one.*

If genetic makers have high linkage disequilibrium with causative variants and present minor allele frequencies (MAF) near 0.5, the identification of genetic markers located near the causative variants would be easy in the first two scenarios, *(a)* and *(b)* (López de Maturana et al., 2014); but not in the third scenario even though studies would use a large sample size. This happens because genetic markers close to causative variants would be undetectable, since their shifts of allele frequencies are negligible. Moreover, the variation of a given trait is hypothetically linked to different causative variants in each line, and consequently to different genetic markers.

One of the main issues of GWAS analysis is the "missing heritability". UC lines showed that the estimates of genomic heritabilities including line effect were similar to the heritabilities calculated by Blasco et al. (2005). Therefore, there were not 'missing heritabilities'. Conversely, this genetic parameter was higher in the model excluding line effect (*e.g.* 0.34 and 0.30 for TNB and IE, respectively), suggesting an overestimation of the heritability. This might be the result of using a small sample of highly related animals for traits with a few major genes in agreement with the relevant genomic region identified on rabbit chromosome (OCU) 17 at 70.0 - 73.3 Mb. This region was associated with three little size traits: implanted embryos (IE), total number born (TNB) and number born alive (NBA). The genomic variances explained by this region were up to 32%, 39% and 10% for IE, TNB, and NBA, respectively, under a model excluding the line effect. Thus, this genomic region was considered as a novel reproductive QTL in rabbits, since shows an important pleiotropic effect. The main retrieved genes in this QTL were *PNRC2, BMP4, CDKN3, GMFB, CGRRF1,* and *SAMD4A*. The genes within this novel QTL have not been previously reported in GWAS for reproductive traits in pigs (Onteru et al., 2013; Bergfelder-Drüing et al., 2015; Guo et al., 2016). Genes of bone morphogenetic protein (*BMP*) family, including *BMP4* gene from our GWAS for UC, are linked to reproductive traits in pig (Hunter et al., 2005), sheep (Demars et al., 2013), mice and human (Shimasaki et al., 2004). The findings of this study suggest that the divergent selection agreed with the scenario *(a)*, in which few causative variants of large effect present MAF close to 0.5. Besides, this can be supported by the great response

at the second generation, half of the estimated selection response in UC divergent selection experiment (Blasco et al., 2005). A study focused on the presence of a feasible mayor gene segregating in these UC lines showed positive results (Argente et al., 2003). Nevertheless, our results must be interpreted with caution because of the small sample size, triggering higher sampling noise and increasing the false positives rate. Furthermore, these results came from phenotypic records collected at second parity; therefore, our findings may not be extrapolated to other parities. However, the genetics correlation between the second and subsequent parities are high, being greater than 0.84, according to an experiment in three maternal rabbit lines (Piles et al., 2006).

Further genomics studies on QTL in OCU17 would be necessary in order to corroborate the effect size and identify causative variants using rabbit commercial populations; for instance, the refining of the genomic region by genome sequencing, gene expressions, and local association studies (Ioannidis et al., 2009; Schmid & Bennewitz, 2017). In addition, a preliminary study of selection signatures using UC lines has been carried out. According to this analysis, the SNPs in the associated region in OCU17 presented F_{st} up to 0.39, being their values over five times standard deviation (SD) of F_{st} using both UC lines (Sosa-Madrid et al., 2017). Nevertheless, the small sample size hindered reliable results from several methods of selection signatures, leading to non-conclusive results with these methods.

A selection response of 3.10 SD was attained from the divergent selection experiment for IMF at 9th generation. Our initial hypothesis was that the high genetic divergence between the IMF lines was the result of the selection of few causative variants of large effects, the scenario *(a)*, or several causative variants of small effects linked to detectable and nearby SNPs, the scenario *(b)*. In some livestock, such as pigs and beef cattle, the selection against fat depot have been carried out directly and indirectly by correlated traits in the breeding programs, *e.g.* the selection for backfat thickness or for feed efficiency in pigs (Hermesch, 2004); and for marbling or for 12th-rib fat in beef cattle (Ochsner et al., 2017). Therefore, if there were major genes for IMF, these genes would be fixed or nearby fixation. On the contrary, in rabbits, the idea of few mayor genes for the IMF, the scenario *(a)*, would be feasible, since this species had never been selected for meat quality traits or for any trait having a high correlation between

it and IMF (Martínez-Álvaro et al., 2016). Besides, we assumed a high detection power because of using of animals from the IMF divergent selection. These animals are an exceptional material for genomic analyses (Kessner & Novembre, 2015; Lou et al., 2019). In the case of scenario *(b)*, we would detect several SNPs associated with causative variants of small effect if this scenario would be suitable for the IMF divergent selection. Results of GWAS for IMF disclosed ten SNPs in OCU8 and ten SNPs in OCU13 with relevant p-values and Bayes factors. Nevertheless, none of these genomic regions explained over 10% of IMF genomic variance. In addition, SNPs within associated genomic regions had low MAF. Overall, an allele of SNP was fixed in low IMF line and this same allele was in high frequency in high IMF line. This finding rejects the hypotheses based on scenarios *(a)* and *(b)*. Thus, the results suggest the divergent selection of IMF experiment produced shifts of allele frequency in different SNPs and different causative variants for each IMF line, represented by the scenario *(c)*.

Regarding the study of selection signatures, multiple genomics regions are considered as selection signatures from each method (F_{st}, XP-EHH and XP-CLR). The results showed only one selection signature in OCU13 that agreed with the GWAS for IMF, under F_{st} method with cut-off of 1%. Nonetheless, the eight selection signatures identified by at least two methods are not in agreement with the physical positions of genomic regions associated with IMF and reported by the GWAS for IMF. Therefore, the IMF selection signatures bear out the assumption of scenario *(c)*, meaning that IMF in rabbits is a complex trait having a large polygenic component with their causative variants barely detectable. That scenario was similar to the Fisher's infinitesimal model (Georges et al., 2019). Besides, the hypothesis of scenario *(c)* would be supported by a constant genetic progress, around 5% per generations of selection (Martínez-Álvaro et al., 2016). Taking into account that, we cannot suit the scenario *(b)* to our results as the GWAS detected SNPs having low MAF which are unlike the selection signatures. The conclusion of IMF having a large polygenic component is similar to the conclusions derived from genomic analyses for IMF experiments in sheep (Duijvesteijn et al., 2018), pigs (Pena et al., 2016), and bovine (Strucken et al., 2017). In addition, the absence of detectable major genes (causative variants) generates few agreements between results from GWAS and methods of selection signatures. Kim et al. (2015) reported low correlations between GWAS and F_{st} (0.33), iHS in the high IMF line (-0.01), iHS in the low IMF line (0.01), and Rsb

(0.01) using divergently selected lines in pigs. Therefore, the absence of major genes would explain our results for genomic analyses of IMF in rabbits.

Our research study shed light on how the rabbit genome is linked to the trait variability in the two divergent selection experiments. The findings of this thesis showed that litter size and intramuscular fat are undoubtedly complex traits. Thus, it would be necessary more evidence derived from other omics, such as transcriptomics, proteomics, metabolomics, and metagenomics; in order to comprehend control mechanisms and complex networks driving these traits in rabbits.

References

Argente, M. J., A. Blasco, J. A. Ortega, C. S. Haley, and P. M. Visscher, 2003. "Analyses for the presence of a major gene affecting uterine capacity in unilaterally ovariectomized rabbits." *Genetics,* 163(3): 1061–1068.

Bergfelder-Drüing, S., C. Grosse-Brinkhaus, B. Lind, M. Erbe, K. Schellander et al., 2015. "A genome-wide association study in large white and landrace pig populations for number piglets born alive." *PLoS ONE,* 10(3): e0117468.

Blasco, A., J. A. Ortega, A. Climent, and M. A. Santacreu, 2005. "Divergent selection for uterine capacity in rabbits. I. Genetic parameters and response to selection." *Journal of Animal Science,* 83(10): 2297–2302.

Demars, J., S. Fabre, J. Sarry, R. Rossetti, H. Gilbert et al., 2013. "Genome-wide association studies identify two novel BMP15 mutations responsible for an atypical hyperprolificacy phenotype in sheep." *PLoS Genetics*, 9(4): e1003482.

Duijvesteijn, N., S. Bolormaa, C. Gondro, S. Clark, M. Khansefid et al., 2018. "Genome-wide association study of meat quality traits using whole-genome sequence data in a multi-breed sheep population." *Proceeding of World Congress on Genetics Applied to Livestock Production,* Volume Methods and Tools – GWAS, 257.

Georges, M., C. Charlier, and B. Hayes, 2019. "Harnessing genomic information for livestock improvement." *Nature Reviews Genetics,* 20(3): 135–156.

Guo, X., G. Su, O. F. Christensen, L. Janss, and M. S. Lund, 2016. "Genome-wide association analyses using a Bayesian approach for litter size and

piglet mortality in Danish Landrace and Yorkshire pigs." *BMC Genomics,* 17(1): 468.

Hermesch, S., 2004. "Genetic improvement of lean meat growth and feed efficiency in pigs." Australian Journal of Experimental Agriculture, 44 (5): 383.

Hunter, M. G., V. Brankin, R. L. Quinn, E. M. Ferguson, S. A. Edwards et al., 2005. "Oocyte–somatic cell–endocrine interactions in pigs." *Domestic Animal Endocrinology,* 29(2): 371–384.

Ioannidis, J. P. A., G. Thomas, and M. J. Daly, 2009. "Validating, augmenting and refining genome-wide association signals." *Nature Reviews Genetics,* 10(5): 318–329.

Kim, E. S., R. Ros-Freixedes, R. N. Pena, T. J. Baas, J. Estany et al., 2015. "Identification of signatures of selection for intramuscular fat and backfat thickness in two Duroc populations." *Journal of Animal Science,* 93(7): 3292–3302.

Kessner, D., and J. Novembre, 2015. "Power analysis of artificial selection experiments using efficient whole genome simulation of quantitative traits." *Genetics,* 199(4): 991–1005.

López de Maturana, E., N. Ibáñez-Escriche, Ó. González-Recio, G. Marenne, H. Mehrban et al., 2014. "Next generation modeling in GWAS: comparing different genetic architectures." *Human Genetics,* 133(10): 1235–1253.

Lou, R. N., N. O. Therkildsen, and P. W. Messer, 2019. "The effects of quantitative trait architecture on detection power in artificial selection experiments." BioRxiv 672683.

Martínez-Álvaro, M., P. Hernández, and A. Blasco, 2016. "Divergent selection on intramuscular fat in rabbits: Responses to selection and genetic parameters." *Journal of Animal Science*, 94(12): 4993–5003.

Ochsner, K. P., M. D. Macneil, R. M. Lewis, and M. L. Spangler, 2017. "Economic selection index development for beefmaster cattle I: Terminal breeding objective." Journal of Animal Science, 95(3): 1063–1070.

Onteru, S., B. Fan, Z. Q. Du, D. J. Garrick, K. J. Stalder et al., 2012. "A whole-genome association study for pig reproductive traits." *Animal Genetics,* 43(1): 18–26.

Pena, R. N., R. Ros-Freixedes, M. Tor, and J. Estany, 2016. "Genetic marker discovery in complex traits: A field example on fat content and composition in pigs." *International Journal of Molecular Sciences,* 17(12): 2100.

Piles, M., M. L. García, O. Rafel, J. Ramon, and M. Baselga, 2006. "Genetics of litter size in three maternal lines of rabbits: Repeatability versus multiple-trait models." *Journal of Animal Science,* 84(9): 2309–2315.

Schmid, M., and J. Bennewitz, 2017. "Invited review: Genome-wide association analysis for quantitative traits in livestock – a selective review of statistical models and experimental designs." *Archives Animal Breeding,* 60(3): 335–346.

Shimasaki, S., R. K. Moore, F. Otsuka, and G. F. Erickson, 2004. "The bone morphogenetic protein system in mammalian reproduction." *Endocrine Reviews,* 25(1): 72–101.

Sosa-Madrid, B. S., N. Ibáñez-Escriche, M. A. Santacreu, L. Varona, and A. Blasco, 2017. "Huellas de selección en un experimento de selección divergente para capacidad uterina en conejo." In Proceedings of the XVII *Jornadas sobre Producción Animal*, 30-31 May 2017, Zaragoza, Spain, pp. 558–560.

Strucken, E. M., H. A. Al-Mamun, S. de las Heras-Saldana, M. N. Bedhane, D. Lim et al., 2017. "Finding the marble - The polygenic architecture of intramuscular fat." *Journal of Animal Breeding and Genomics,* 1(2): 69–76.

CHAPTER SEVEN

7. CONCLUSIONS AND IMPLICATIONS

- The genomic analyses of this thesis used animals from two successful experiments of divergent selection in rabbits; however, the detection power for genomic analyses relies strongly on genomic architectures.

- Genome wide association studies (GWAS) for litter size traits disclosed a promising QTL in the rabbit chromosome (OCU) 17 at 70.0 - 73.3 Mb. The intermediate frequencies of SNPs within this QTL can be a result of the divergent selection for uterine capacity. This genomic region presented an important pleiotropic effect for implanted embryos, total number born and number born alive. Nonetheless, this QTL would be validated in maternal rabbit commercial lines.

- Main genomic regions associated with intramuscular fat (IMF) were in OCU8 (24.6 - 26.9 Mb) and in OCU13 (83.8 - 86.0 Mb). The findings suggest that this trait has a large polygenic component due to the small genomic variance accounted for by genomic regions.

- Genomic scans studies using IMF lines revealed multiple genomic regions considered as selection signatures when methods were individually analysed. Despite some genes presented function directly related to energy, carbohydrates and lipid metabolisms, none of the selection signatures identified by at least two methods were in the regions reported by GWAS. This corroborates that IMF has a large polygenic component.

- From both results, GWAS and selection signatures studies, we suggest that the IMF divergent selection affected different causative variants and different SNPs in each IMF line.

- Regarding the implications, the results of this thesis do not still have implications for maternal rabbit breeding programs. No particular SNP was found explaining a large part of the genomic variance of litter size traits.

- GWAS and genome scan studies do not work as methods for accomplishing relevant information for rabbit breeding programs. These methods must be used as an exploratory analysis.