Universidad de León

Facultad de Veterinaria

Departamento de Producción Animal

Programa de Doctorado

Ciencias Veterinarias y de los

Alimentos

Use of genomic tools for the study of the genetic basis of resistance to gastrointestinal nematode infections in adult sheep

"*Utilización de herramientas genómicas para el estudio de la base genética de la resistencia a las infecciones por nematodos gastrointestinales en ovejas adultas*"

Praveen Krishna Chitneedi

León, 2020

Supervisors:

Dr. Beatriz Gutiérrez Gil

Prof. Juan José Arranz Santos

Departamento de Producción Animal, Facultad de Veterinaria,

Universidad de León, Campus de Vegazana s/n, León 24071, Spain.

The research work included in this PhD Thesis memory has been by supported by the LE248U14 project funded by the Castilla and León Regional Government (Junta de Castilla y León) and the AGL2015-66035-R project funded by the Spanish Ministry of Economy and Competitiveness (MINECO).

Praveen krishna Chitneedi is a grateful grantee funded by a predoctoral fellowship from the Castilla and León Regional Government (Junta de Castilla y León; ORDEN EDU/310/2015 de 10 de abril) and the European Social Fund.

“Arise awake and stop not until the goal is achieved”

Swami Vivekananda

Acknowledgements

Firstly, I would like to express my sincere gratitude to my research guide and supervisor Dr. Beatriz Gutiérrez-Gil and supervisor Prof. Juan Jose Arranz for the continuous support of my PhD study and related research, for their patience, motivation, and immense knowledge and have given me more power and spirit to excel in the research writing. Your advice on both research as well as on my career have been invaluable.

I am also grateful to María Martínez-Valladares, Elsa García-Gámez, Marina Atlija, Aroa Suárez-Vega and Cristina Esteban Blanco for providing me the initial research materials to carry out my PhD research. I am also pleased to say thank you to the super computation institute (SCAYLE) for providing me the required computational resources to complete my research goals on time.

I would like to extend my sincere thanks to Prof. Christa Kühn for giving me an opportunity to work with her group at Institute of Genome Biology, Leibniz Institute for Farm Animal Biology (FBN) in Germany and also the fellow group members in the institute who supported me during my short stay. Thank you so much for teaching me and providing me the required resources to carry out a successful short stay at your institute.

I would also like to thank Junta Castilla y León and European social fund for their financial support to my PhD research and FAANG Cost action committee for the grant to short research stay at FBN, Germany, without their financial support I otherwise would not have been able to develop my scientific discoveries.

I thank my fellow lab mates and the staff at the Animal production department, University of León for the stimulating discussions and for all the fun we have had in the last six years. I cannot list all the names here, but you are always on my mind. Your friendship made my research work a wonderful experience.

There are no words to express my gratitude to my parents (Udaya Bhaskara Rao and Surya Kumari), for your constant moral support and encouragement to fulfil my personal and professional goals and providing me the comfort to explore unknow territories in my life.

I would like to extend my acknowledgements to my sister Sravani, brother-in -law Chaitanya, my two little nephews, my relatives and friends in India and abroad for their warm wishes and affection, which made me who I am today.

Finally, I must acknowledge with tremendous and deep thanks to my wife Geeta Gayatri, who has been by my side throughout this PhD, living every single minute of it, to give me positive energy to go beyond my limits with your love, patience, support what help me to complete this dissertation journey.

All of them and those I haven't named who have done their bit to make these years of work, Thank you very much!

Table of Contents

Índice de contenidos

1. Research approach and objectives

The present doctoral Thesis in the field of the genetics underlying gastrointestinal nematode (GIN) resistance in sheep has been carried out in the research group of Animal Breeding and Genetics of the University of León (also known as MEGA-ULE group, from *Mejora Genetica Animal*). The study has been carried out using animals from the Spanish Churra breed, which is an autochthonous dairy sheep breed from the region of Castilla y León, in the north west region of Spain. The National Association of Breeders of Selected Churra sheep (ANCHE) has routinely collaborated for many years with the MEGA-ULE group, which has traditionally been dedicated to the study of the genetic basis of traits of economic interest in dairy sheep, initially mainly focused on the study of milk production traits, at both the quantitative level, to estimate genetic parameters and develop appropriate models for genetic selection (Othmane *et al.* 2002) and also at the molecular level, to identify QTL regions influencing milk production traits (Gutierrez-Gil *et al.* 2009; García-Gámez *et al.* 2012a) or study candidate genes (García-Gámez *et al.* 2011). This global approach has also been applied to other functional traits of economic interest in dairy sheep such as morphology traits (Fernández *et al.* 1997; Gutiérrez-Gil *et al.* 2011) or disease resistance traits, mainly mastitis resistance (Gutiérrez-Gil *et al.* 2007) and genetic resistance to gastrointestinal nematodes (GIN) (Gutiérrez-Gil *et al.* 2009, 2010). In this section, I try to briefly present the previous research efforts of the MEGA-ULE group in relation to the study of the GIN infections and the existing scientific scenario in the field of genomic research at the start of this project, both are major elements shaping the final content of this Thesis.

Considering the relevant impact that GIN infections have in grazing dairy sheep populations in the region of Castilla y León (Spain), the MEGA-ULE group has focused some of their research efforts to the study of the genetic factors influencing resistance to GIN in sheep. This has enabled the estimation of genetic parameters of indicator traits of GIN infection, such as faecal egg count (FEC), or serum levels of immunoglobulin A (IgA) and pepsinogen (Pep), in a commercial population of Spanish Churra sheep (Gutierrez-Gil *et al.* 2010) and the detection of quantitative trait locus (QTL) associated to the parasite resistance/susceptibility status in this breed. The initial studies on this regard were based on about 180 microsatellite markers (Gutiérrez-Gil *et al.* 2009) and later defined and extended based on the analysis of thousands of Single Nucleotide Polymorphisms (SNP) through the use of the medium density 50K SNP-chip (Atlija *et*

al. 2016), a genomic tool directly derived from the information generated by the sheep genome sequencing project (International Sheep Genomics Consortium *et al.* 2010).

The initial QTL results reported based on classical linkage analysis (LA) with 181 microsatellite markers included four QTL for the FEC trait on ovine chromosomes (OAR)1, 6, 10 and 14 and one QTL for the IgA trait on OAR1 (Gutiérrez-Gil *et al.* 2009). The most significant genetic effect detected in that study was the genome-wide significant QTL for FEC on OAR6. The later analysis performed in a different commercial population of Churra sheep with 43,613 autosomic SNP markers (after quality control), allowed the identification of some of the previously detected QTL and some new significant QTL/associations for FEC and IgA, by applying different statistical gene mapping approaches: classical LA, combined linkage disequilibrium and linkage analysis (LDLA) and also genome-wide association analysis (GWA) analysis (Atlija *et al.* 2016). Interestingly, the LA and LD&LA, replicated the QTL detected by Gutierrez-Gil *et al.* (2009) in OAR6 for FEC and allowed the refinement of its confidence interval from 100.6 to 88.1 cM.

The medium density ovine 50K-chip used in that study is a commercial SNP-array (Illumina Infinium Ovine SNP50 BeadChip) that includes over 54,000 SNP markers distributed evenly across the sheep genome. This SNP-array has been proven to be a useful genomic tool to map the causal mutation of traits with simple mendelian inheritance in sheep such as several genetic diseases (Becker *et al.* 2010; Suaŕez-Vega *et al.* 2013; Suárez-Vega *et al.* 2015a). However, in the case of gene mapping studies related to complex traits, which are controlled by many genetic effects, each of them explaining a small proportion of variance, the potential of this medium-density SNP-chip offers limited ability to identify causal mutations. For example, from all the traits analysed by the MEGA-ULE group in Churra sheep, related to milk production, udder morphology, resistance to mastitis and to GIN infections, only a causal mutation of a QTL influencing milk protein and fat percentage was detected in the *LALBA* gene (García-Gámez *et al.* 2012b). This limitation to identify causal variants, similarly than the limited accuracy of genomic selection predictions, is related to the extent of linkage disequilibrium (LD) that persists between causal variants and the density of markers available to detect it. In sheep, in general, and specially in some breeds, including Churra, the extent of LD is short (Kijas *et al*. 2014). With the aim of providing a genomic tool to explore the behaviour of the LD extend over short chromosomal distances (Kijas *et al.* 2014), the International Sheep

Genomic Consortium (ISGC) developed the ovine HD BeadChip. This SNP-chip offers a platform carrying assays for over 600,000 loci with an average spacing of around 5 kb (J. Mc Ewan, unpublished). Because the genotyping of this tool across an entire population is expensive, its combination with strategies of genotype imputation (Li *et al.* 2009b) may offer good opportunities to reach increased power for QTL mapping studies and also obtain more accurate predictions in sheep populations.

In addition, this project Thesis has largely been modelled by the impact that second generation sequencing technologies or next generation sequencing (NGS) have had in the field of livestock genomics after the progressive decrease of its price since 2003, when the human genome project was finished, and especially considering the drastic price decrease observed in the last decade (Schuster 2008; Ghosh *et al.* 2018). Today massive parallel sequencing allows the global study of structural DNA variants, through Whole Genome Sequencing (WG-seq), the overall analysis of transcriptome profiles through RNA Sequencing (RNA-seq) to compare gene expression profiles, identify new transcripts, isoforms or regulatory features such as miRNA or long non-coding RNAs (lncRNA) (Costa-Silva *et al.* 2017; Stark *et al.* 2019). Also in the last years the study of epigenetic regulatory mechanisms of gene expression can be massively dissected through the application of ChIP-seq and Whole Genome Bysfulfate Sequencing (WGBS) (Kernaleguen *et al.* 2018). Trying to identify genomic solutions to real problems of dairy sheep populations, the first research projects of the MEGA-ULE group exploiting NGS were focused on the study of classic dairy phenotypes. For example, through the dynamic analysis of the milk somatic cell transcriptome by using RNA-seq in two dairy sheep breeds (Suárez-Vega *et al.* 2015b). Also WG-seq has been exploited by this group to perform high-resolution mapping within the confidence region of a QTL influencing the somatic cell score (SCS), as an indicator trait of subclinical mastitis (Gutiérrez-Gil *et al.* 2018). The efforts of the group to undertake these studies, in relation to the development of free-software based bioinformatic workflows and the development of computational facilities (e.g. MEGA-ULE own servers and collaboration with the SCAYLE super-computation centre) are also major basis for the development of the work presented in this Thesis.

Considering all this, the present Thesis was planned with the global objective of using genomic tools for the study of the genetic basis of resistance to gastrointestinal nematode infections in adult sheep.

To reach this global objective in an efficient way, and following the established official research plan for this research project, two specific objectives have been defined:

1. To refine and high-resolution mapping of QTL influencing GIN resistance in adult sheep. Imputation of SNP-chip genotypes and whole genome sequencing will be exploited to identify QTL and potential causal mutations influencing indicator traits of GIN resistance traits in a commercial population of dairy sheep exposed to GIN natural infection. Also, HD-chip imputed genotypes will be used to update LD estimations and other population parameters (*Ne*, *F*, ROH extent).
2. Perform a global transcriptomic study of target abomasal tissues after an experimental challenge with *Teladorsagia circumcincta* in adult ewes previously classified as resistant and susceptible. RNA-seq datasets will be analysed to identify differential transcript expression patterns of genes and long non-coding RNAs (lnc-RNA) and also to identify potential causal mutations related to GIN resistance.

2. Introduction

2.1. Dairy sheep and its socio-economic importance in Castilla y León, Spain

Domestication of preferred wild animals as livestock to fulfil our needs for food, clothing and farming can be considered as a crucial stage in human evolution from early human into modern human being, as it changed our lifestyle from hunter-gatherers to farmers (Mignon-Grasteau *et al.* 2005; Hartung, 2013) Among livestock species, the small ruminants, sheep and goats were one of the first domesticated animals for milk, meat and wool. The early domestication of sheep and goat dates back to 9000 BC near middle east (Pedrosa *et al.* 2005). Sheep is a versatile animal and can adopt to different climatic conditions. In sheep, today there are different breeds and sub-breeds, due to multiple domestication events for several generations in Eurasia and in African regions (Chessa *et al.* 2009).

Spain accounts for around 27% of sheep milk delivered to industry in the European Union for the year 2016 according to the Ministry of Agriculture, Food and Environment, Spain (MAGRAMA). Among which, the major proportion of products delivery comes from the north-western autonomous community of Castilla y León. Dairy sheep are very important livestock in Castilla y León. Around 53.7% of the milk production of dairy sheep in Spain is produced in the region of Castilla y León according to MAGRAMA (2016). In this region, sheep milk is traditionally used for cheese and yogurt production. Assaf and Churra are two popular dairy sheep breeds in Castilla y León. The Assaf is a non-local breed introduced in León province during 1970´s from Israel (Legaz *et al.* 2008). The Churra sheep is autochthonous to this region. It is one of the primitive breeds of the Iberian Peninsula and its existence dates back to Medieval times (Grau-Sologestoa, 2015). The breeding program for Churra sheep is actively carried out by ANCHE (National Association of Breeders of Selected Churra sheep), which was created in 1973 with the objective of the genealogical control, production control and genetic improvement of this breed for milk production and also the production of suckling lamb meat. ANCHE initiated the breeding program of Churra sheep in 1986 with technical support advice from the Department of Animal Production of the University of León (de la Fuente *et al.* 1995). The original objective of this breeding program was to improve the milk yield by maintaining its natural adaptability to the harsh environmental conditions. In few years, the program included also as selection objective the improvement of the milk composition (milk protein and fat contents) to maximize cheese yield. Morphology

traits, including udder and body conformation, and number of born lambs were later considered in the year 2006 with the collection of the first lambs as future progeny test sires.

2.2 Impact of GIN infection in sheep and control strategies

GIN infection is one of the major health related issues faced by the sheep farming industry worldwide, which incurs major economic losses for the farmers in the form of the cost associated with anthelmintic treatment and the negative effect in the overall productivity, fertility and survival of sheep with a range of health issues (Dominik, 2005; Mavrot *et al.* 2015). The prevalence of nematode infection is mostly dependent on the optimal conditions for the development of nematode species in the sheep rearing regions (O'Connor *et al.* 2006) and also requires sufficient humidity to facilitates the free-living stages of the nematode life cycle (Martínez-Valladares *et al.* 2013b). The seasonal changes also effect the prevalence of nematodes in a given region. The Castilla y León region, with has a Mediterranean climate with continental influences, is characterised by long cold winters and warm summers, which favours the prevalence of GIN nematodes like *Teladorsagia circumcincta* (Martínez-Valladares *et al.* 2013b). Moreover, the prevalence of GIN infections in this region was previously reported and found to play a significant role in the overall well-being of the sheep industry of Castilla y León (Martínez-González *et al.* 1998; Martínez-Valladares *et al.* 2013b). Classically, the control of GIN infections in sheep populations used to be carried out using anthelmintic drugs, but GIN resistance against anthelmintic drugs gradually appeared and in sheep flocks, including those included in the region of Castilla y León (Martínez-Valladares *et al.* 2013a). The alternative approaches to tackle this issue includes, pasture management, vaccination and developing resistant to GINs through selective breeding for successive generations (Sayers and Sweeney, 2005). Among different alternatives to anthelmintics, selection and breeding for resistance to nematode infection is the most favourable approach for control against nematode infection, as it is more sustainable compared to other approaches and it also reduces both direct and indirect costs associated with the eradication of nematode infection (Dominik, 2005). However, the resistance against nematode infection in sheep is a complex (quantitative) phenotype to measure and also to introduce in a sheep breeding program (Dominik, 2005). In other sheep populations where

the selection for nematode resistance has been carried out, the most commonly used indicator trait has been FEC. The heritability of this traits varies from low to moderate (h^2 range: 0.1-0.3) (Gutiérrez-Gil *et al.* 2010; Mirkena *et al.* 2010; Mpetile *et al.* 2015; Brown and Fogarty, 2017; Ngere *et al.* 2018). The FEC trait can be measured after natural or artificial infection and the measurement highly varies with the seasons and also with the change in climatic conditions leading to over or under FEC estimates. Thus, we can consider that classical genetic selection in relation to GIN infection in sheep is inefficient as the routine measure of the FEC trait, which shows a limited heritability, is time consuming and expensive. Because of that, many authors have suggested that the molecular marker information to increase the resistance of animals to GIN infection would be a more efficient approach (Dominik, 2005; Davies *et al.* 2006; Krawczyk and Słota, 2009).

2.3 The evolution of sheep genomics

Genomics can be defined as an interdisciplinary field of science that focuses on the structure, function, evolution, mapping and editing of the genomes of different species. Genomics is divided into two major branches, structural genomics, which focuses on the study of DNA sequence variations and their effects, and functional genomics, which studies the expression of genes and the functional regulation of genomes.

In relation to the field of domestic animal breeding like sheep, genomics as a whole, aims to identify genetic variations, patterns of gene expression and regulatory mechanisms of the genome that have a direct influence on the traits of economic interest. While some of the economic interest traits are single gene-controlled traits (e.g. halothane hypersensitivity, wool colour, etc.), most of them are complex traits and are controlled by numerous genes or Quantitative Trait Loci or QTLs (e.g. milk production, milk and meat composition, yield traits, morphological traits and disease resistance traits).

At the time of the initial studies focused on QTL mapping in livestock species (Georges *et al.* 1995; Nezer *et al.* 1999), mainly based on linkage analysis (LA) and microsatellite markers, the purpose of QTL identification for complex traits was the application of molecular information to selection programs, through the strategy known as Marker Assisted Selection (MAS). In MAS, the information from markers linked to QTLs was suggested to be used in addition to the information provided by the pedigree and

phenotypes. Ideally, to efficiently exploit the molecular information derived from a QTL it would be necessary the identification of the specific gene and mutation responsible for the genetic effect initially mapped as a QTL, known as QTN (Quantitative Trait Nucleotide). The application of this specific information in the improvement programmes constitutes the strategy known as Gene Assisted Selection (GAS). Practical examples of GAS include the use of the *DGAT1* gene genotype in New Zealand dairy cattle (Spelman *et al.* 2002) and the routinely genotyping of polymorphisms of the ovine *PRNP* gene to increase resistance to Scrapie disease in the European Union, as regulated by the EC Directive 2003/100/EC (European Commission 2003). However, considering the large number of studies classical QTL mapping based on LA and the use microsatellites showed a limited power to identify causative mutations.

The high level of competitiveness for increased sequence yields that took place during the last years of the Human Genome project facilitated the development of the Next Generation Sequencing (NGS) technologies (Schuster, 2008). In the case of sheep, a first virtual draft of the sheep genome was made by taking advantage of bacterial artificial chromosome end sequences, a sparse marker map, and the sequences of cow, dog, and human genomes (Dalrymple *et al.* 2007). Subsequent work based on this virtual genome and accelerated by NGS technologies derived into the original first draft of the sheep reference genome (International Sheep Genomics Consortium *et al.* 2010). The advances in sequencing technologies allowed the development, of high-performance genotyping platforms in domestic species like sheep known as arrays or SNP chips, which are genomic tools that include thousands of SNPs spaced at more or less constant distances and that provide dense coverage of the entire genome. In addition to increase the gene mapping potential, the availability of SNP chips has allowed the implementation of Genomic Selection (GS) proposed in 2001 by Meuwissen *et al.* (2001) which is a variant of MAS based on genotyping information of animal.

The NGS technologies have also allowed the characterization and quantification of a complete range of "omic" sciences, such as genomics, transcriptomics and epigenomics, with which we try to understand the mechanisms and functions of the genome as a whole entity. Depending on the type of project, we can choose a different type of NGS technology application: (i) the genome sequencing, or whole genome sequencing (WG-seq), and whole exome sequencing, are used for the identification of causal variants, rare variants or the study of the metagenome of a given sample; (ii) the sequencing of the

RNA or transcriptome (RNA-seq) of a tissue is useful for the evaluation of the expression levels of all the genes that are expressed at a given physiological state, or in response to a given treatment or challenge (e.g. experimental infection, nutritional challenge), as well as for the identification of genes and gene networks that are differentially expressed between two conditions; and (iii) the study of the heritable changes in gene expression and other genomic functions without alteration of the DNA sequence can be studied also through global technologies such as chromatin immunoprecipitation sequencing (ChIP-seq) or Whole Genome Bisulfite Sequencing (WGB-seq) .

2.3.1 Structural genomics in sheep breeding

The study of genomes at a structural level is based on the analysis of genetic markers for the identification of variants associated with phenotypes of interest. The major types of markers that have been used in gene mapping studies in animal breeding are microsatellite and SNP (single nucleotide polymorphism) markers.

Microsatellites (1 to 10 nucleotides) are subcategories of tandem repeats (TRs) that, together with the predominant interspersed repeats make up genomic repetitive regions (Vieira *et al.* 2016). Based on microsatellite markers, in the last decade of the twentieth century medium density linkage maps were developed. Microsatellite type markers were the basis for the first linkage maps in most of the domestic species such as: hen (Bumstead and Palyga, 1992), cow (Barendse *et al.* 1994), pig (Ellegren *et al.* 1994), sheep (Crawford *et al.* 1995) and goat (Vaiman *et al.* 1996). These linkage maps were a fundamental tool for the first QTL detection projects, in both plants and animal species, based on linkage analysis (LA), which studies the segregation of chromosomal fragments labelled by markers within populations with a family-based structure.

For the different livestock species, the information derived from the corresponding Sequencing Genome and HapMap projects allowed the identification of thousands of SNPs (Single Nucleotide Polymorphisms) across the genome (Nicholas and Hobbs, 2014). SNP markers are variations of a single nucleotide in the genomic DNA. Most frequently, SNPs are bi-allelic, a characteristic that makes them highly suitable for large-scale genotyping. The advances in sequencing technologies allowed the development, of high-performance genotyping platforms in domestic species like sheep known as arrays or SNP chips, which are genomic tools that include thousands of SNPs spaced at more or

less constant distances and that provide dense coverage of the entire genome. The first chips marketed were those known as medium density SNP-chips (about 50, 000-60, 000 markers, 50-60K) and years later, high density SNP-chips were manufactured (including up to 600, 000-800, 000 SNPs). Today commercial SNP-chips for different species are available, e.g. cow (50K, 800K), chicken (60K, 500K), goat (50K), sheep (50K, 700K), pig (60K, 600K), horse (50K) and dog (170K). These genotyping platforms are widely developed with the llumina Infinium and Affymetrix Axiom technologies among others. The bioinformatics tool SNPchiMp v.3 shows a large part of the commercial chips for farm species (http://bioinformatics.tecnoparco.org/SNPchimp/). There is also the possibility of developing specific SNP-chips (custom chips), in which both the density of markers and the markers themselves are chosen according to the specific objective (e.g. paternity testing, genomic selection through imputation, etc) for which the SNP-chip is developed.

To analyse the generated genotypes with the SNP-chips, the available reference genome allows to determine the position of the markers on the basis of the reference sequence, which in turn constitutes the physical map. The association analysis at the genomic level is performed based on the physical map provided by the genome sequence itself. Unlike classical linkage studies the density of markers provided by massive genotyping platforms such as SNP-chips, allows genes to be mapped on the basis of population information, which is based on the linkage disequilibrium (LD), through association studies. Due to higher density of genotype information and being independent of family design, The, association studies at genomic level (GWA analysis, Genome-Wide Association analysis) emerged as a more efficient alternative to the classic mapping of QTLs (Andersson, 2009).

The coverage of genomes provided by SNP chips has also been used for the development of selective sweep mapping studies at a higher resolution that were provided by microsatellite markers. The estimation of population genetics parameters was used to identify selective sweeps (fixation index, heterozygosis, LD, haplotypic structure). In populations with divergent phenotypes the analysis of these large datasets of genetic markers has served to identify numerous signatures of selection in the genomes of pig (Rubin *et al.* 2012), cow (Gutiérrez-Gil *et al.* 2015), sheep (Gutiérrez-Gil *et al.* 2014) and dog (Vaysse *et al.* 2011)

As a further step, some sheep genomic studies have used WG-seq to perform a high-resolution analysis of the genetic variability within target regions previously identified based on the analysis of SNP-chip datasets. For instance, the analysis of the whole sequence information from a QTL segregating trio (father Qq, resistant offspring QQ and susceptible offspring qq) served to find a causal mutation related to resistance to mastitis resistance in the *SOCS2* gene in the Lacaune breed (Rupp *et al.* 2015). Using the same approach, Gutierrez-Gil *et al.* (2018) reported a list of potential functional variants for a QTL detected in Churra sheep in relation to the same phenotype, measured through the somatic cell count (SCC) indicator trait. By contrasting the genetic variability identified through WG-seq of the Churra and Australian Merino breeds in regions previously identified as selection sweeps, Gutiérrez-Gil *et al*. (2017) highlighted some mutations that could be associated to the intensive selection process the fine wool Merino lines have been subjected to QTL mapping analysis, (Rupp *et al.* 2015; Gutiérrez-Gil *et al.* 2018) identification of signatures of selection (Gutiérrez-Gil *et al.* 2017) and the analysis of trios segregation (father Qq, resistant son QQ and susceptible son qq), of a QTL identified in sheep served to find a causal mutation related to mastitis resistance in the *SOCS2* gene (Rupp *et al.* 2015).

In addition to increase the gene mapping potential, the availability of SNP chips has allowed the implementation of a variant of MAS, proposed in 2001 by Meuwissen *et al.* (2001), and called GS. The proposal of these authors is based on estimating the genetic value of the animals from markers distributed uniformly throughout the genome. For this purpose, the genetic effects of each marker are first estimated and then added to predict the global breeding value of the animal. This process involves two stages. The first, where the effects of the markers are estimated using a reference population (training population) that includes individuals with two types of information available: 1) genotypes for a large number of markers and 2) phenotypic measurements for the traits for which the GS is to be performed. In the second stage, the estimated effects are applied to candidates for selection from a wider commercial population. For these candidate individuals, genotypic information is available but phenotypic measures are not available (Boichard *et al.* 2016). To be efficient, this approach is very demanding, both in terms of the number of genotyped individuals and the number of markers analysed.

The main advantage of the GS is that candidates can be evaluated and therefore selected without their phenotypic information or that of their progeny. Thus, on the basis of

genotypic information, we could select animals after birth, or even select embryos. This determines that, for species with a long generation interval, e.g. cattle, the GS can increase the genetic progress based on the reduction of the generation interval. If the cost of genotyping is affordable for the species, the testing of a large number of candidates can be used to increase the intensity of selection and improve the utilization of available genetic resources.

The methods for applying GS are diverse. Given the large number of markers analysed, all the methods used consider the effects of markers as random effects, their values being distributed according to a given statistical distribution, which varies depending on the method. One of the most commonly used methods is the GBLUP (Genomic Best Linear Unbiased Prediction) method proposed by VanRaden *et al.* (2009), which is an extension of the polygenic BLUP method where the matrix of relationships is based on information from markers rather than the pedigree. As an alternative, Bayesian methods have been proposed. These methods give a high weight in the model to the SNPs estimated to be close to causal variants, and eliminate from the model some other SNPs that have no effect on the trait (Mrode and Thompson 2005; Gianola 2013).

Apart from its clear advantages, the GS has difficulties to be applied in commercial animal populations related to the need to have a large reference population, with thousands of animals for the correct estimation of effects and the need to "recalibrate" the effects of SNPs from time to time, i.e. to continue measuring phenotypes in reference populations. All these difficulties have been limited in Holstein cattle due to the special structure of this worldwide spread population, characterized by a small effective population size, a large extent of the linkage disequilibrium and a very high number of animals with phenotypes and genotypes available, initially in USA and Canada, and later in the rest of countries. Also, the high uniformity of farming systems and phenotype records in Hostein cattle are a help to maximize the advantages of GS. On the other hand, dairy sheep populations present important difficulties for the practical application of GS. These include small population sizes with very diverse population structure, relatively large effective population sizes associated to short extents of linkage disequilibrium, and a wide variety of farming system and measured phenotypes across Mediterranean countries, where this type of production occurs.

The application of GS requires the routine genotyping of many animals, which is one of the major economic limitations of its implementation in commercial sheep populations.

In order to reduce genotyping costs, the option of genotyping young candidates with a low density chip (e.g. 3K or 6K) has been proposed to infer or impute non-genotyped markers based on information generated in a reference population genotyped with a higher density panel (e.g. 50K or 500K). The advantages of genotype imputation also make it possible to share genotyping data between different populations, which can be of great help in improving the accuracy of genomic estimates, or also in increasing the statistical power of GWA-type studies. Imputation also allows genotypes generated with different marker panels to be combined, provided there is sufficient overlap between the markers of each panel (Sargolzaei *et al.* 2014). Several factors affect the accuracy of the imputed genotypes. Among them, the size and structure of the reference population and its genetic relationship with the animals in which the imputation will be performed, the distribution of the SNPs in the chromosomes, the structure of the genome of the analysed population or breed, dependent in turn on the effective size of the population, and the frequency of the alleles of the markers in the population (since alleles with a low minor allele frequency (MAF) present more difficulties for their imputation) (Ventura *et al.* 2016). In general, imputation methods are based on the construction of chromosomal phases from the genotypes of the reference population, with Beagle (Browning and Browning 2013) and FImpute (Sargolzaei *et al.* 2014) being two of the most used imputation programs in animal populations.

Genotype imputation, in addition to having been developed to infer genotypes between panels of different densities (from 3K to 50K, from 3 to 500K, or from 50K to 500K) (Hayes *et al.* 2012; VanRaden *et al.* 2013) has also been developed to be applied to the scenario in which the highest density of markers is given on the basis of genomic sequencing data or Whole Genome Sequencing (WG-seq) (Van Binsbergen *et al.* 2014). In this sense, the study of several imputation scenarios has shown that imputation in two steps, for example from a chip-50K to a chip-500K and then to WG-seq improves the precision of imputation with respect to imputation in a single stage (van Binsbergen *et al.* 2014).

2.3.2 Functional genomics in sheep breeding

While structural genomics focuses on the study of DNA sequence as something static, functional genomics is the part of genomics that focuses on dynamic aspects such as

transcription, translation and regulation of gene expression in a particular cell and under specific conditions. Classical studies of RNA expression, for example, using the Northern blot technique, were very tedious and slow to develop into an isolated gene. In 1995, a new technology, known as expression microchips or microarrays, was developed thanks to great advances in the development of computational tools and by adapting the microchips of the computer industry to the study of gene expression (Hamadeh and Afshari, 2000). Microarrays are a collection of cDNA probes of known sequence attached to a solid surface that are used for gene expression analysis. To do this, the RNA of a tissue or cells of interest is labelled with appropriate fluorochromes. After hybridization of the array-labelled RNA, the chip is scanned and the fluorescent signal at each point on the chip with a cDNA probe provides a measure of the expression of the corresponding gene. Their functioning basically consists of measuring the level of hybridization between the specific probe, and the target molecule, and they are generally indicated by fluorescence and through a computer analysis of the image, which indicates the level of expression of the corresponding gene (Bendixen *et al.* 2005). The numerous studies developed with microarrays in sheep provided a large amount of information at a functional level in relation to metabolic processes (Satterfield *et al.* 2009; Chang *et al.* 2019), nematode resistance (Diez-Tascón *et al.* 2005; Keane *et al.* 2006; Knight *et al.* 2011), immune responses (Galindo *et al.* 2008) and skin pigmentation (García-Gámez *et al.* 2011; Peñagaricano *et al.* 2012) among others. However, the technology of expression microarrays presented important limitations, such as the difficulty of statistical analysis, low reproducibility and, above all, the need to know the sequences of the genes to be analysed.

In recent years, based on the development of NGS technologies, high-throughput transcriptomic studies have become an essential component of functional genomics. The data generated by RNA-seq provide a snapshot of the expression profile of all genes in a particular tissue, as well as a view of gene functions related to a particular trait (Kadarmideen, 2014). This technology also gives us a more accurate measure of the expression profiles of transcripts and their isoforms than previous methods such as microarrays or RT-PCR (Wang *et al.* 2009). A detailed review of different applications of RNA-seq technology in the field of animal genomics is presented by Wickramasinghe *et al.* (2014). Examples of these applications in sheep include studies of the transcriptome of somatic cells in milk for the identification of genes differentially expressed throughout

lactation (Suárez-Vega *et al.* 2015b; Suarez-Vega *et al.* 2016) for the identification of variants in sheep milk transcriptome (Suárez-Vega *et al.* 2017), to characterize the wool fibre transcriptome in sheep (Zhang *et al.* 2017) and the identification of differentially expressed genes (DEGs) after GIN infection (Guo *et al.* 2016; McRae *et al.* 2016; El-Ashram *et al.* 2017), among other studies. RNA-seq data can also be subjected to specific analyses for the identification of new isoforms of RNA transcripts and the identification of lncRNA transcripts (Bush *et al.* 2018) the latter being elements of the genome that seem to play a fundamental role in functional regulation (Ransohoff *et al.* 2018).

In this regard it should be noted that the studies based on NGS data (WG-seq and RNA-seq) require bioinformatic tools and statistical methods suitable for processing, handling and analysis of the large amount of generated data. In the last years the field of Bioinformatics has experienced a tremendous development, being many of the applications developed in open source code, especially in R and Python languages, and being available for systems based on Linux (Gentleman *et al.* 2004; Magi *et al.* 2010).

2.4 Genomic studies in relation to the resistance to GIN infection in sheep

2.4.1. Gene mapping studies

Some of the initial studies in relation to the genetic basis underlying nematode resistance in sheep followed the candidate gene approach where genes known to be involved in the immune response in this kind of infections were sequenced to find polymorphisms, which were later tested for association with and indicator trait of GIN resistance, mainly the FEC phenotype. For instance, different studies assessed the role of polymorphisms genes included in Major Histocompatibility Complex (MHC), such as DRB1 (Schwaiger *et al.* 1995).

On the other hand, the first genome scans performed for GIN resistance in sheep with microsatellite markers, carried out at AgResearc (New Zealand) and CSIRO (Australia) through the analysis of half-sib pedigrees bred from sires derived from crosses of divergent lines selected over several generations for high and low resistance to nematode parasites (predominantly *Haemonchus contortus* and *Trichostrongylus colubriformis*),

identified QTL for GIN resistance in OAR3 and OA6, suggesting in both projects a large effect for the QTL located on the q arm of chromosome 3. Interferon gamma (IFN-γ), which is a cytokine that plays an important role in the regulation of the immune response to parasitic infection (Wikel 1997) was the identified as the most likely candidate gene for the mapped effect and different subsequent studies focused on the study of polymorphisms in the genomic region harbouring this gene (Crawford, A.M. and McEwan, J.C. 1998; Coltman *et al.* 2001; Sayers *et al.* 2005). Although the identification of the causal mutation in this gene has not been reached, this candidate gene is still one of the most important candidates studied in reference to GIN resistance in sheep (Patra *et al.* 2016).

Other early genome scan in this field were that reported by Beh *et al.* (2002) which was based on the analysis of family sires derived by crossing divergent lines of sheep selected for response to challenge with the intestinal parasitic nematode *Trichostrongylus colubriform.* Later, QTL mapping studies for GIN resistance traits were reported different sheep breeds: In half-sib Scottish Blackface lambs infected with multiple species of GINs (Davies *et al.* 2006); in Romneys sheep, which are FEC-based selection lines of outcross pedigrees of infected with multiple nematode (Crawford *et al.* 2006); in half-sib families of Merino sheep infected by *Haemonchus contortus* (Marshall *et al.* 2009); and also a commercial population of half-sib families of Spanish Churra dairy sheep after natural infection Gutiérrez-Gil *et al.* (2009). In general, these studies used FEC as indicator traits, although in the case of the study reported in Churra sheep, other traits such as the serum level of IgA and the serum level of Pepsinogen (as a measurement of the tissue damage produced by the infection in the abomasum) were also analysed (Gutiérrez-Gil *et al.* 2009). The results of the microsatellite-based genome scans related to GIN infection in sheep, as for classical production traits, can be found in the SheepQTLdb database (Hu *et al.* 2013), which compiles the results derived from numerous QTL detection projects in various domestic species (cow, pig, chicken, sheep, horse and rainbow trout).

In relation to the QTL reported by the MEGA-ULE group in the Churra sheep study (Gutiérrez-Gil *et al.* 2009), it is important to highlight that these results refer to adult animals, in particular of the Churra dairy sheep breed. This aspect differs from most of the rest of studies focused on the study of the genetic resistance to GIN in young animals. This particularity is based on the specific production system of dairy sheep breeds in

Spain where suckling lambs are sacrificed at the age of 21-28 days and have only been fed exclusively from maternal milk. This makes that adult ewes are the animals that suffer from the GIN infections, especially after lambing. The studies performed by this group in relation to parasite resistance have resulted from the collaboration with the research group of Parasitology from the Department of Animal Health at the ULE. Epidemiological studies by this second group to investigate the parasitism in dairy sheep in northwest Spain and the environmental factors that affect the prevalence and intensity of gastrointestinal infection under semi-intensive management conditions have shown that *T. circumcincta* is the species causing most of the clinical infections in dairy ewes of this geographical region (Martínez-Valladares *et al.* 2013b). These studies have also shown a high level of prevalence of anthelmintic resistance in sheep nematode parasites in the region of Castilla y León (Álvarez-Sánchez *et al.* 2006), which encourages the use of genetic breeding to increase the natural resistance of the animals to this kind of infections. Hence, within the European project *SheepMilkSafety* (Barillet *et al.* 2006), the MEGA-ULE group estimated genetic parameters for three indicator traits of GIN resistance in adult sheep, FEC, serum IgA and serum pepsinogen levels (Gutiérrez-Gil *et al.* 2010), and reported the identification of QTL for FEC and IgA on chromosomes OAR 1, 6, 10 and 14 (Gutiérrez-Gil *et al.* 2009). These studies were carried out using a commercial population of Spanish Churra sheep showing a structure of half-sib families which was appropriate to apply a daughter design classical linkage analysis. It should be noted that after the detection of QTL through a genome scan with sparse markers like microsatellites, it is necessary to carry out fine mapping studies, normally focused on the QTL identified with higher statistical support. This is because the analysis of a limited number of markers per chromosome determines a very low resolution of the QTL mapping process, this means that the estimated confidence intervals involve very large regions of the genome where further analyses are needed for the identification of the genuine causal gene. QTL fine mapping consists on the analysis of a larger number of markers in the region initially identified to harbour a QTL in order to reduce the wide confidence interval obtained in the initial mapping and, if possible, identify the QTL causal mutation. Previously to the availability of the reference genome sequence, classical fine mapping studies included the analysis of new microsatellite markers and the discovery of new markers (microsatellites and SNPs) in the target region by sequencing, and a very costly and time consuming process of gene cloning by exploiting comparative mapping (e.g. the gene cloning of the *DGAT1* gene for milk fat content in Holstein cattle

as reported by Grisart *et al.* (2002)). However, in reference to resistance to GIN infection in sheep none of the microsatellite-based studies reported in the scientific literature identified causal mutations.

The availability of a commercial medium density 50K-chip for sheep in 2009 opened new opportunities for the field of QTL mapping in this species as they may provide initial gene mapping results that are accurate enough to facilitate the assessment of a shorter list of potential candidate causal genes as a previous step to QTN identification. The analysis methodology of GWA using SNP-chip datasets has proved to be highly efficient for the identification of genomic regions underlying complex traits in humans (Visscher *et al.* 2017) and livestock species (Kemper and Goddard, 2012; Womack *et al.* 2012). In any case, the identification of the genuine causal mutations is still a challenging task, and the most recent studies in the QTL mapping field are also exploiting whole genome (Duijvesteijn *et al.* 2018; Al Kalaldeh *et al.* 2019a) or targeted sequencing (Periasamy *et al.* 2014; Estrada-Reyes *et al.* 2019a; b) to explore the genetic variations within the genomic regions containing the studied QTL.

A list of SNP-chip studies related to GIN resistance traits in sheep, most of them based on the 50K-chip, but some other exploiting the HD-chip (600K) or lower densities (12K SNP-chip), are summarized in Table 1. The results of some of these studies but not of all are included in SheepQTLdb (Hu *et al.* 2013), where they are classified as "Associations" to differentiate them from the regions identified with the classical QTL studies based on microsatellite markers. Altogether, the SNP-chip-based studies for GIN resistance traits have refined some of the wide QTL regions previously discovered with microsatellite markers and have also identified novel QTL regions. This is the case the of the study reported by Atlija *et al.* (2016), performed by the MEGA-ULE group in a commercial population of Spanish Churra sheep previously genotyped with the 50K-chip and where phenotypes for GIN resistance were collected. Interestingly, the LA analysis reported in this work replicated the OAR6 QTL for FEC previously identified with microsatellite markers by Gutierrez-Gil *et al.* (2009), reducing the confidence interval from an interval of 40 to 10 Mb. On the other hand, the use of LDLA and GWA analyses allowed the identification of new other QTLs segregating in the Churra sheep population (Atlija *et al.* 2016).

As previously indicated, the 16 most recent studies included among the list of 16 studies indicated in Table 1, one has exploited imputation of SNP-chip to WG-seq density (Al

Kalaldeh *et al.* 2019a) to refine QTL regions for GIN resistance in multi-breed sheep population whereas three other studies have used targeted re-sequencing to explore the genetic variants within genomic regions harbouring significant QTL and various candidate genes (Periasamy *et al.* 2014; Estrada-Reyes *et al.* 2019a; b). The study Al Kaladeh *et al.* (2019a) identified some genomic regions on OAR 2, 6, 18 and 24 associated to parasite resistance and the pathway analysis identified 13 genes within these significant regions (*SH3RF1, HERC2, MAP3K, CYFIP1, PTPN1, BIN1, HERC3, HERC5, HERC6, IBSP, SPP1, ISG20,* and *DET1*), which have various roles in innate and acquired immune response mechanisms, as well as cytokine signalling. The SNPs associated to *H. contortus* resistance in Florida Native sheep were identified in the study of Estrada-Reyes *et al.* (2019a) on OAR 1, 2, 3, 4, 6, 7, 11, 15, 18, 20, 24 and 26, which identified loci related to Th17, Treg and Th2 responses and identified potential markers to natural *H. contortus* exposure in genes *ITGA4, MUC15, TLR3, PCDH7, CFI, CXCL10, TNF, CCL26, STAT3, GPX2, IL2RB* and *STAT6*. In unrelated breeds across Asia, Europe and South America Periasamy *et al.* (2014) identified 41SNPs across 31 candidate genes in relation to nematode resistance in sheep to confirm a strong phylogeographic structure across sheep breed and balancing selection operating at SNP loci located within immune pathway genes.

Table 1: The list of high-resolution mapping performed for the detection of resistance to GIN infection in sheep

Study	Genotype data	Identifier of sheep QTL database	Trait	Type of infection	Study population	Methodology[1]
Atlija *et al.* (2016)	Ovine SNP 50k BeadChip	95607, 95608	LFEC	Natural infection	Adult Churra ewes	LA
Atlija *et al.* (2016)	Ovine SNP 50k BeadChip	95610, 95625, 95635, 95645, 95646	LFEC	Natural infection	Adult Churra ewes	LDLA
Benavides *et al.* (2015)	Ovine SNP 50k BeadChip	31314, 31315	Average FEC (AVFEC)	Natural challenge conditions (AVFEC)	Red Maasai x Dorper Backcross Population	GWA
Berton *et al.* (2017)	Ovine SNP12k BeadChip		FEC	Natural infection	Santa Inês breed	GWA
Garza Hernandez *et al.* (2018)	Ovine SNP 50k BeadChip	140303 - 140310	FEC	Natural infection	Texel sheep	GWA
Al Kalaldeh *et al.* (2019b)	Ovine SNP 600k HDBeadChip	178732 - 178769	Worm egg counts in multi breed sheep	Natural infection	Australian Sheep (various breeds)	GWA
Al Kaladeh *et al.* (2019a)	Ovine SNP 600k HDBeadChip		Worm egg counts in multi breed sheep	Natural infection	Australian Sheep (various breeds)	RHM (200 SNP window)
Al Kaladeh *et al.* (2019a)	Whole Genome Sequencing		Worm egg counts in multi breed sheep	Natural infection	Australian Sheep (various breeds)	RHM (250 Kbps window)
McRae *et al.* (2014)	Ovine SNP 50k BeadChip	125773 - 125802	FEC	Natural infection	Romney and Perendale	SSM (moving 5 SNP on average)

Periasamy *et al.* (2014)	Targeted re-sequencing	57493, 57499	FEC	Natural and Artificial challenge	22 breeds across Asia, Europe and South America	Candidate gene approach by target re-sequencing
Pickering *et al.* (2015)	Ovine SNP 50k BeadChip		FEC	Natural infection	Different New Zealand sheep breed	GWA
Riggio *et al.* (2013)	Ovine SNP 50k BeadChip		FEC	Natural infection	Scottish Blackface lambs	GWA
Riggio *et al.* (2013)	Ovine SNP 50k BeadChip		FEC	Natural infection	Scottish Blackface lambs	RHM (100 SNP window size)
Riggio *et al.* (2014)	Ovine SNP 50k BeadChip		FEC	Natural infection	Scottish Blackface, Sarda X Lacaune and Martinik Black-Belly X Romane)	RHM (100 SNP window size)
Estrada-Reyes *et al.* (2019a)	Targeted sequencing		FEC, packed cell volume (PCV), IgA, IgG, IgM and Average daily gain (ADG)	Natural and Artificial challenge	Dorper, Katahdin, and St. Croix	GWA
Estrada-Reyes et. al., (2019b)	Targeted sequencing	180503 - 180561	FEC, packed cell volume (PCV), IgA, IgG, IgM and Average daily gain (ADG)	Natural infection	Florida Native sheep	GWA

[1] *LA: Linkage Analysis; LDLA: Linkage Disequilibrium combined with Linkage Analysis; GWA: Genome-wide Association analysis; RHM: Regional Heritability Mapping; SSM: Selection Sweep Mapping*

2.5. Functional characterization of the mechanisms involved in the resistance to GIN infection in sheep.

2.5.1. Tools used for functional characterization of GIN resistance in sheep

In sheep, the functional characterization of QTLs (Dervishi *et al.* 2011; Sallé *et al.* 2014) and candidate genes (MacKinnon *et al.* 2009; Gossner *et al.* 2012; McRae *et al.* 2016) after GIN infection has served to identify upregulation or downregulation of some specific alleles and genes at a particular stage of infection. The study of these expression profiles will provide an insight into the physiology of host-parasite interaction and the possible mechanisms of host resistance involved at different parasitic stages. This provides new possibilities to control GIN parasite through vaccination or post infection medication or by identifying the mutations controlling GIN resistance in sheep, which may improve the ability to perform genetic selection directly by targeting the identified key genes. Different high-throughput technologies are providing a vast amount of information of new gene targets for parasite control. The RT-PCR was one of the preliminary technologies used to study gene expression profiles in studies focused on the resistance to GIN infection in sheep (Schallig 2000). This technology has evolved with the time from semi-quantitative assays to 'real-time' RT-PCR (Newton and Meeusen 2003). The RT-PCR was used to conduct studies to detect expression changes in specific molecules like intelectins, interleukins, galectins and specific genes known to be involved in disease resistance and susceptibility mechanisms (French *et al.* 2008; Zaros *et al.* 2010) and to identify differentially expressed genes in response to nematode infection in sheep (Ingham *et al.* 2008; Menzies *et al.* 2010). With the availability of high-throughput technologies, firstly microarrays and later RNA-seq, studies focused on a global functional characterization of GIN resistance in sheep have become possible. Although these studies are expensive and time consuming, they generate large amounts of data that inform about many significant genes and pathways involved in the immune response against GIN infection in sheep (Newton and Meeusen 2003; McRae *et al.* 2016). About these high-throughput gene expression experiments, a special care has to be taken at the time of defining the experiment design and performing the statistical analyses to obtain useful and reliable information. Due to the large scale of these experiments a small error may generate a vast amount of false positive results and validation of errors is also complex for these kinds of genomic approaches. Classically, the validation of the results

of microarray expression studies have been carried out with RT-PCR by selecting the highly significant or by randomly selecting few significant genes (Knight *et al.* 2011; Gossner *et al.* 2013; Yang *et al.* 2015) . This has been also in many cases the approach to validate results from RNA-seq studies (Guo *et al.* 2016; El-Ashram *et al.* 2018; Zhang *et al.* 2019) Validation of the results for these studies is also proven by comparing the obtained results with other similar studies. However, in the case of RNA-seq experiments some authors have shown that confirmation with RT-PCR is not needed because RNA-seq technology boasts a general high level of data reproducibility across lanes and flow-cells which reduces the need of technical replication within these experiments (Marioni *et al.* 2008).Thus, we believe that the biological replicates included in our analysis are enough to ensure the validity of the RNA-seq results, without depending on RT-PCR confirmation.

2.5.2. Expression studies in sheep for resistance to GIN

In the past two decades several gene expression studies were carried out to detect genes responsible for resistance and susceptibility to GIN infection in sheep. From Table 2, which shows a list of such expression studies performed with RT-PCR, microarrays and RNA-seq technologies, it can be observed that the initial expression studies were carried out with RT-PCR. These preliminary RT-PCR studies were focused on the mechanism of the immunological response by considering specific immune cells like eosinophils (Balic *et al.* 2006) and molecules like immunoglobulins (Ig) (White *et al.* 2001) and mammalian chitinases (Knight *et al.* 2007). Later, this technique was used to study the genetic bases of ovine resistance to GIN infection by identifying DEGs (Ingham *et al.* 2008) and to study the expression of specific selected genes (French *et al.* 2008; Zaros *et al.* 2010). The availability of microarrays made possible the parallel measurement of the expression of multiple genes responsible for resistance to GIN in sheep. As shown in Table 2, multiple studies were conducted with microarrays to find DEGs in sheep after GIN infection (Diez-Tascón *et al.* 2005; Keane *et al.* 2006; Rowe *et al.* 2008, 2009; Andronicos *et al.* 2010). In some of these microarray studies, the candidate gene identified as DEGs were subsequently validated using RT-PCR (Knight *et al.* 2010, 2011; Gossner *et al.* 2013; Yang *et al.* 2015). In recent years, expression studies in relation to GIN resistance in sheep have been carried out by using RNA-seq, mainly to identify and

study DEGs (Table 2). The identification of the number of DEGs has increased multi-fold with RNA-seq studies. In most of these RNA-seq studies, the identified DEGs were also validated by RT-PCR. On other hand, the interpretation of results for these global transcriptomic studies is generally based on the results of gene enrichment analyses that are used to identify the biological pathways enriched by the identified DEGs.

Globally, the expression studies performed in relation to GIN resistance in sheep, as summarized Table 2, identified multiple DEGs. Some of these DEGS were commonly identified across different studies irrespectively of the breed, experimental design and the technology used. For example, genes that belong to the intelectin family, such as *ITLN1* and *ITLN2*, were commonly identified in two independent microarray studies (Rowe *et al.* 2009; Knight *et al.* 2011) In these studies the studied animals were infected with different nematode species (*H. contortus* and *T. circumcincta* respectively) and they adopted different methodology for the collection and preparation of samples. Moreover, the differential expression of intelectin genes in relation to GIN resistance in sheep has also been identified in another independent RT-PCR study (French *et al.* (2008) and a RNA-seq-base study (El-Ashram *et al.* 2017). Likewise, galectins genes (French *et al.* 2008; Knight *et al.* 2011; McRae *et al.* 2016) matrix metalloproteinase genes (Knight *et al.* 2011; Guo *et al.* 2016) interleukins (French *et al.* 2008; Rowe *et al.* 2009) and other immune related genes were also commonly identified as DEGS by many different studies based on different technologies (see details in Table 2). Furthermore, some metabolic and immunological pathways like Cytokine signalling (Colditz 2003; Ingham *et al.* 2008; Al Kalaldeh *et al.* 2019b), Toll-like receptor signalling (Ingham *et al.* 2008; Benavides *et al.* 2016; Al Kalaldeh *et al.* 2019b) and Extracellular matrix (Ingham *et al.* 2008; Nagaraj *et al.* 2012; Benavides *et al.* 2016) were commonly identified among different studies. Overall, all these reported gene expression studies have been useful to identify the possible common functional mechanisms underlying GIN resistance in sheep across different nematode infections. The identification of these candidate genes involved in the immune response to GIN infection could be also a first step towards the identification of genetic variants within those genes or their regulatory genomic regions that could serve as genetic markers to be used as genomic information to integrate in classical or genomic selection programs of commercial populations considering the resistance to GIN infection as a secondary functional trait to consider in selection indexes.

Table 2 The list expression studies related to GIN in sheep and the techniques used in each study

Animal age / breed	Infection status	Tissue used for study	Nematode species	Technique	Study
Lambs Merino Sheep	Experimental infection (Resistance vs susceptible)	Abomasal tissues	*H. contortus*	RNA-seq, validation RT-PCR	Zhang *et al.* (2019)
One-month old lambs	Experimental infection (Ex vivo infection)	Ovine abomasal explants	*H. contortus*	RNA-seq, validation RT-PCR	El-Ashram *et al.* (2018)
Two-month-old sheep	Experimental infection	Fundic abomasal samples	*H. contortus*	RNA-seq, validation RT-PCR	El-Ashram *et al.* (2017)
Lambs Scottish Blackface	Experimental infection (Susceptible sheep)	Lymph node	*T. circumcincta*	RNA-seq, validation RT-PCR	McRae *et al.* (2016)
1-year old Canaria Hair Breed (CHB) and Canaria Sheep (CS)	Experimental infection (Infected CHB and CS)	Abomasal mucosa	*H. contortus*	RNA-seq, validation RT-PCR	Guo *et al.* (2016)
Texel and Suffolk lambs	Natural infection	Abomasal lymph node	*N. battus*	RNA-seq, validation RT-PCR	Ahmed *et al.* (2015)
six-month-old female Hu sheep	Experimental infection	T lymphocytes	*H. contortus*	Microarray, validation RT-PCR	Yang *et al.* (2015)
Lambs Scottish Blackface	Experimental infection (Resistant vs Control)	Lymph node	*T. circumcincta*	Microarray, validation RT-PCR	Gossner *et al.* (2013)

Yearling sheep	Experimental infection (Immune day 5 vs Naïve day 5 and Naive day 5 vs Naive day 0)	Abomasal mucosa	*T. circumcincta*	Microarray, RT-PCR validation	Knight *et al.* (2011)
Naïve female Romney lambs	Experimental infection	Migratory cells in afferent lymph	*T. colubriformis*	Microarray, RT-PCR validation	Knight *et al.* (2010)
Lambs	Experimental infection (Primary challenge vs tertiary challenge)	Lymph node	*T. colubriformis*	Microarray, qPCR validation	Andronicos *et al.* (2010)
6 months old Merino-cross wethers	Experimental infection	Abomasal mucosa	*H. contortus*	Sequential microarray	Rowe *et al.* (2009)
Merino-cross wethers aged 6 months	Experimental infection	Abomasal mucosa	*H. contortus*	Microarray	Rowe *et al.* (2008)
Resistant and susceptible Perendale sheep lines	Natural infection (Resistance vs susceptible)	Duodenum	*Gastrointestinal parasites*	Microarray	Keane *et al.* (2006)
Resistant and susceptible Perendale sheep lines	Natural infection (Resistance vs susceptible)	Duodenum	*Gastrointestinal parasites*	Microarray	Diez-Tascón *et al.* (2005)
Lambs (Scottish Blackface x Leicester)	Experimental infection (Infected sheep)	Abomasum, Lymph node	*T. circumcincta*	Semi-quantitative RT-PCR	French *et al.* (2009)

Brazilian Somalis sheep	Natural infection (Resistance vs susceptible)	Abomasum, abomasal lymph nodes and small intestine	Gastrointestinal nematode	RT-PCR	Zaros *et al.* (2010)
Scottish Blackface-cross yearling sheep	Experimental infection (unchallenged naïve (unv), challenged naïve (cnv), unchallenged previously infected (upi) and challenged previously infected (cpi))	Abomasal fold	*T. circumcincta*	RT-PCR	French *et al.* (2008)
Lambs	Experimental infection (Resistant sheep)	Abomasal mucosa	*T. colubriformis, H. contortus*	RT-PCR	Ingham *et al.* (2008)
Scottish Blackface cross yearling lambs	Unchallenged naïve (worm-free), challenged naïve sheep which had no previous exposure to nematodes, and challenged previously infected sheep	Abomasal fold and gastric lymph node samples	*T. circumcincta*	RT-PCR	Knight *et al.* (2007)
Merino wethers (2–3 years old)	Experimental infection	Abomasal fold	*H. contortus*	RT-PCR	Balic *et al.* (2006)
Shed-reared lambs	Experimental infection	Abomasal lymph node	*H. contortus*	RT-PCR	White *et al.* (2001)

3. Materials and Methods

As this PhD Thesis is compendium of different research studies, the materials and methods used for different studies were available in the corresponding research articles and conference papers.

All the research articles and conference papers related to this PhD Thesis can be found in “Results” section.

4. Results

LIST OF PUBLICATIONS

Objective 1

Study 1: Praveen Krishna Chitneedi, Juan-Jose Arranz, Aroa Suárez-Vega, Elsa García-Gámez, and Beatriz Gutiérrez-Gil. (2017). Estimations of linkage disequilibrium, effective population size and ROH-based inbreeding coefficients in Spanish Churra sheep using imputed high-density SNP genotypes. Animal genetics, *48*(4), pp.436-446. *DOI: https://doi.org/10.1111/age.12564*

Study 2: Praveen Krishna Chitneedi, Beatriz Gutiérrez-Gil and Juan-Jose Arranz. (2017). Low density chip design and genomic imputation from low density (LD-Chip) SNP chip to medium (50K-Chip) and high (HD-Chip) density in sheep. XVII Jornadas sobre Producción Animal, Zaragoza, España, 30 y 31 de mayo de 2017, pp.522-524. *https://www.cabdirect.org/cabdirect/abstract/20173312869* (17th National Conference on Animal Production**,** Zaragoza, Spain, 30th-31st of May 2017.)

Study 3: Praveen Krishna Chitneedi, Marina Atlija, Juan-Jose Arranz and Beatriz Gutiérrez-Gil. (2015). GWAS analysis for gastrointestinal nematodes resistance traits using imputed high density chip genotypes in sheep. XVI Jornadas sobre Producción Animal, Zaragoza, España, 19 y 20 de mayo de 2015, pp.471-473. *https://www.cabdirect.org/cabdirect/abstract/20163059168* (16th National Conference on Animal Production**,** Zaragoza, Spain, 19th-20th of May 2015.)

Study 4: Praveen Krishna Chitneedi, Beatriz Gutiérrez-Gil, Cristian Esteban-Blanco, Marina Atlija, and Juan-Jose Arranz (2020). Detection and high-resolution study of QTLs underlying resistance to gastrointestinal nematode infection in adult sheep by using imputed HD SNP-chip genotypes and whole-genome sequencing. "Manuscript under preparation".

Objective 2

Study 5: Praveen Krishna Chitneedi, Juan-Jose Arranz, Aroa Suárez-Vega, María Martinez-Valladares and Beatriz Gutiérrez-Gil. (2017). Preliminary differential transcriptomic analysis of abomasal mucosa from resistant and susceptible sheep to gastrointestinal nematodes (GINs) after an experimental infection with *T.*

circumcincta. 36th International Society for Animal Genetics Conference (ISAG), 16th-21st of July 2017, Dublin, Ireland. *https://buleria.unileon.es/handle/10612/6845*

Study 6: Praveen Krishna Chitneedi, Aroa Suárez-Vega, María Martinez-Valladares, Juan-Jose Arranz and Beatriz Gutiérrez-Gil (2018). Exploring the mechanisms of resistance to Teladorsagia circumcincta infection in sheep through transcriptome analysis of abomasal mucosa and abomasal lymph nodes. Veterinary research, 49(1), pp.39. *DOI: https://doi.org/10.1186/s13567-018-0534-x*

Study 7: Praveen Krishna Chitneedi, Aroa Suárez-Vega., María Martinez-Valladares, Juan-Jose Arranz and Beatriz Gutiérrez-Gil (2018). Variant discovery in genes identified as differentially expressed genes between the abomasal lymph node transcriptome of resistant and susceptible adult sheep to *Teladorsagia circumcincta* infection. XIX Reunión Nacional de Mejora Genetica Animal, 14-15 de junio, 2018, León, España. *https://buleria.unileon.es/handle/10612/8285* (19th National Animal Breeding Meeting, 14th -15th of June 2018, León, Spain.)

Study 8: Praveen Krishna Chitneedi, Juan-Jose Arranz, Aroa Suárez-Vega, María Martinez-Valladares and Beatriz Gutiérrez-Gil. (2020). Identification of potential functional variants underlying ovine resistance to gastrointestinal nematode infection by using RNA-Seq. Animal genetics. [E-pub ahead of print 2020 Jan 3]. *DOI: https://doi.org/10.1111/age.12894*

Study 9: Praveen Krishna Chitneedi, Christa Kühn, Rosemarie Weikard, Juan-Jose Arranz, María Martinez-Valladares, and Beatriz Gutiérrez-Gil (2019). Detection of long-noncoding RNAs from the differential transcriptomic analysis of abomasal lymph node from resistant and susceptible sheep to gastrointestinal nematode (*T. circumcincta*) after an experimental infection. 37th International Society for Animal Genetics (ISAG) conference, 7th- 12th of July 2019, Lleida, Spain.

Study 1

ANIMAL GENETICS Immunogenetics, Molecular Genetics and Functional Genomics

Estimations of linkage disequilibrium, effective population size and ROH-based inbreeding coefficients in Spanish Churra sheep using imputed high-density SNP genotypes

Praveen Krishna Chitneedi, Juan-Jose Arranz, Aroa Suárez-Vega, Elsa García-Gámez, and Beatriz Gutiérrez-Gil.

Departamento de Producción Animal, Facultad de Veterinaria, Universidad de León, Campus de Vegazana s/n, León 24071, Spain.

Animal genetics, 48(4), pp.436-446. https://doi.org/10.1111/age.12564

Licence no: 4753020667396

ANIMAL GENETICS Immunogenetics, Molecular Genetics and Functional Genomics

doi: 10.1111/age.12564

Estimations of linkage disequilibrium, effective population size and ROH-based inbreeding coefficients in Spanish Churra sheep using imputed high-density SNP genotypes

P. K. Chitneedi, J. J. Arranz, A. Suarez-Vega, E. García-Gámez and B. Gutiérrez-Gil

Departamento de Producción Animal, Facultad de Veterinaria, Universidad de León, Campus de Vegazana s/n, León 24071, Spain.

Summary

In this study, the availability of the Ovine HD SNP BeadChip (HD-chip) and the development of an imputation strategy provided an opportunity to further investigate the extent of linkage disequilibrium (LD) at short distances in the genome of the Spanish Churra dairy sheep breed. A population of 1686 animals, including 16 rams and their half-sib daughters, previously genotyped for the 50K-chip, was imputed to the HD-chip density based on a reference population of 335 individuals. After assessing the imputation accuracy for BEAGLE v4.0 (0.922) and FIMPUTE v2.2 (0.921) using a cross-validation approach, the imputed HD-chip genotypes obtained with BEAGLE were used to update the estimates of LD and effective population size for the studied population. The imputed genotypes were also used to assess the degree of homozygosity by calculating runs of homozygosity and to obtain genomic-based inbreeding coefficients. The updated LD estimations provided evidence that the extent of LD in Churra sheep is even shorter than that reported based on the 50K-chip and is one of the shortest extents compared with other sheep breeds. Through different comparisons we have also assessed the impact of imputation on LD and effective population size estimates. The inbreeding coefficient, considering the total length of the run of homozygosity, showed an average estimate (0.0404) lower than the critical level. Overall, the improved accuracy of the updated LD estimates suggests that the HD-chip, combined with an imputation strategy, offers a powerful tool that will increase the opportunities to identify genuine marker-phenotype associations and to successfully implement genomic selection in Churra sheep.

Keywords HD-chip, imputation, LD extent, runs of homozygosity, SNPs

Introduction

The accuracy and power of both genome-wide association studies (GWAS) and genomic selection are dependent on the extent of linkage disequilibrium (LD) and the density of the markers used to exploit it. The extent of LD varies between populations because it is influenced by their evolutionary history and effective population size (N_e) (Slatkin 2008). In sheep, the Sheep HapMap project, based on analysis of the Ovine SNP50 BeadChip (50K-chip) across a collection of 74 ovine breeds from all over the world (Kijas *et al.* 2014), showed that LD in sheep extended over much shorter distances than in cattle and was much shorter than in Holstein cattle (Raadsma 2010). Among the sheep breeds analysed, Spanish Churra ($n = 120$) was reported to show one of the most remarkable decays of LD with the distance between markers and the lowest (0.8%) genome coverage of haplotype blocks, compared with other breeds with intermediate coverage (Milk Lacaune, 0.9%; Rambouillet, 1%; Australian Suffolk, 1.03%; Australian Pool Dorset, 2.94%) and the very high haplotype block coverage shown by Soay sheep (21.84%) (Raadsma 2010). This short extent of LD in Churra sheep appears to be linked to the high level of diversity shown by Churra compared with other breeds from southern and Mediterranean Europe (Kijas *et al.* 2012).

A detailed assessment of the LD extension in Churra sheep and estimations of effective population (N_e) and inbreeding levels based on analysis of the 50K-chip in a

Address for correspondence

B. Gutiérrez-Gil, Departamento de Producción Animal, Facultad de Veterinaria, Universidad de León, Campus de Vegazana s/n, León 24071, Spain.
E-mail: beatriz.gutierrez@unileon.es

Accepted for publication 14 March 2017

larger set of animals ($n = 1686$) considered in the Sheep HapMap project was reported by our group (García-Gámez *et al.* 2012). This work supported the previous results about the short extent of LD in Churra sheep when compared with other populations (LD decay curve intermediate between those recently reported for Barbados Blackbelly and Australian Suffolk; Kijas *et al.* 2014). However, as indicated by these authors, the precision of the LD estimates based on the 50K-chip is limited over the shortest marker distances due to the small number of marker pairs separated by short distances, with only approximately 0.045% of the marker pairs analysed in the Churra genome showing a distance shorter than 40 kb (García-Gámez *et al.* 2012).

The availability of the ovine High Density Bead Chip (HD-chip), which allows for the analysis of more than 600 000 SNP markers, provides an increased power for GWAS-based studies and genomic predictions in sheep populations. This genomic tool also offers the opportunity to perform high resolution estimations of the LD extension at the shortest marker distances, although its cost for a large population remains expensive. This constraint could be overcome by the implementation of genotyping imputation methods (Hayes *et al.* 2012), in which, stated simply, a subset of individuals genotyped for two SNP panels of different densities (medium and high, in our case) are used as a reference population to infer the genotypes of a larger population from the low to the high marker density using the ancestral relationship and allelic frequencies in the population genotyped at the high density. Imputation accuracies from 50K-chip to HD-chip and whole-genome densities have been estimated in three Australian sheep breeds to be greater than 80% using imputation methods based on population-wide LD without considering pedigree information (Hayes *et al.* 2012).

In this context, through the present study we have firstly assessed the imputation accuracy of two alternative available imputation packages, BEAGLE v4.0 (Browning & Browning 2009, 2013) and FIMPUTE_v2.2 (Sargolzaei *et al.* 2012, 2014), in inferring HD-chip genotypes for the population of 1686 Churra ewes previously analysed with the 50K-chip (García-Gámez *et al.* 2012) using a subset of animals genotyped with the HD-chip as a reference. Secondly, we have used the most accurately imputed HD dataset to update LD extension, N_e and genetic diversity estimations in Churra sheep. In addition, we present here an estimation of the degree of homozygosity based on the HD-chip-imputed dataset by calculating runs of homozygosity (ROHs), an informative parameter that estimates the length of homozygous stretches in each individual. In addition to updating the classical estimates of the inbreeding level in this population based on pedigree and the HD-chip molecular information, we have followed Gomez-Raya *et al.* (2015) to estimate additional inbreeding coefficients that consider the synteny and the length of the identified ROHs.

Materials and methods

SNP genotyping, diversity parameters and quality control

This study comprised two resource populations. The first comprised a total of 1686 animals from a commercial population of Spanish Churra dairy sheep distributed in 16 half-sib families, including 1670 daughters and the corresponding 16 sires, which were genotyped with the ovine 50K-chip. The average family size was 105 daughters per ram (ranging from 29 to 277 animals per half-sib family). The raw genotypes of this dataset had previously been subjected to a two-step quality filtering: first, per individual (call rate > 10%) and second, per marker (call rate > 5%; minor allele frequency > 0.05, Hardy-Weinberg equilibrium P-value <0.00001) (Atlija *et al.* 2016). The files resulting from this quality control (QC), which included a total of 43 613 markers for 1686 individuals are available in the Dryad repository (1686IND50kdata.map and 1686IND50kdata.ped files; https://doi.org/10.5061/dryad.pr7gh). The second resource population was genotyped with the HD-chip and comprised 96 Churra sires widely used in artificial insemination and a subset of animals belonging to the first resource population, including the 16 rams and 224 ewes, distributed as 14 daughters per half-sib family. For this HD-chip raw dataset of 336 Churra individuals, we first estimated genetic diversity parameters following Kijas *et al.* (2014) as described in detail in Appendix S1 (Section 1.1). Later, the two-steps of the QC described above were also applied to the HD-chip genotypes (see details in Appendix S1, Section 1.1). The resulting dataset from this QC, which includes 335 animals and 490 940 markers, is also available at the Dryad repository (335INDHDdata.map and 335INDHDdata.ped files; https://doi.org/10.5061/dryad.pr7gh). For these two resource populations, the markers positions were based on the ovine genome assembly Oar_v3.1. (http://www.ensembl.org/Ovis_aries/Info/Index).

Linkage disequilibrium and sample population estimations for the HD-chip imputed dataset

An assessment of the imputation accuracy from the 50K-chip to the HD-chip density for our Churra sheep population was performed with two different software, BEAGLE v4.0 (Browning & Browning 2009, 2013) and FIMPUTE v2.2 (Sargolzaei *et al.* 2014), following a cross-validation methodology described in Appendix S1 (Section 1.2). The distribution of the markers showing an error rate higher than 0.3 (mis-imputed markers) was also assessed and a mis-imputation hot-spot were defined when at least five mis-imputed markers were found within a 2-Mb long interval.

Using the HD-chip imputed dataset showing the highest imputation accuracy, updated LD and N_e estimations were performed for the whole population under study following the procedures previously implemented for the 50K-chip by García-Gámez *et al.* (2012) and summarized in detail in Appendix S1 (Section 1.3). To assess how imputation could influence the accuracy of the LD and N_e estimations, the following were also obtained for two additional datasets: (i) the HD-chip genotypes of the 335 included in the reference population and (ii) the imputed HD-chip genotypes of the 1,686 animals under study after removing the markers included in the regions identified as mis-imputation hot-spots. The estimates obtained for these HD-chip related datasets were compared with those previously reported based on the 50K-chip (García-Gámez *et al.* 2012).

Runs of homozygosity and inbreeding coefficients

Runs of homozygosity were identified in our imputed dataset using PLINK (Purcell *et al.* 2007), following the approach and criteria described in detail in Appendix S1 (Section 1.4). The inbreeding level of the resource population was estimated based on the HD-chip-imputed genotypes following two general approaches. First, the inbreeding coefficient was calculated based on the observed vs. expected number of homozygous genotypes by implementing the *--het* option in PLINK (F_{Mol}). Second, inbreeding coefficients were estimated considering the ROH estimates. This was undertaken based on the total ROH content, considering the distribution of the length of a ROH as previously detected (F_{ROH}) and, also, considering the distribution of the lengths of ROHs as a random variable with an associated distribution or probability density, as suggested by Gomez-Raya *et al.* (2015). For the last approach, using the R source code provided as additional material by these authors, we estimated inbreeding coefficients based on (i) the quantiles of the distribution of the length of ROH ($F_{\mathrm{ROH\text{-}Q}}$), and based on (ii) fitting an exponential distribution to the ROH-length distribution and using the mean of the exponential density ($F_{ROH-E}{}^{m}$). For calculating all of the inbreeding coefficients, the individuals analysed were grouped based on their year of birth (from 2001 to 2008), and the mean value for the three aforementioned parameters were estimated. After estimating the inbreeding coefficients, they were standardized to fall between 0 and 1 using the approach suggested by Gomez-Raya *et al.* (2015). In addition to these marker-based inbreeding estimates, we also had the estimates of inbreeding based exclusively on the pedigree information previously presented by García-Gámez *et al.* (2012) (F_{PED}) available for the studied Churra sheep population. To compare these coefficients, and considering that some of the genomic-based inbreeding coefficients did not fall within the range of 0–1, we used the standardization procedure suggested by Gomez-Raya *et al.* (2015).

Results and Discussion

Diversity parameters for the HD-chip genotypes

The diversity parameters obtained for the raw HD-chip dataset in the reference population are provided in Appendix S1 (Section 2.1). After quality control, the HD-Chip dataset included 335 animals and 490 940 markers. The distribution of the MAF for the HD-Chip SNPs after QC is shown in Fig. S1.

Imputation accuracy assessment

The average concordance obtained across the 26 autosomal chromosomes, was 92.2% for BEAGLE and 92.1% for FIMPUTE (see individual estimates per chromosome in Table S1). A detailed description and discussion of these results is provided in Appendix S1 (Section 2.2). Considering the 7.8% of genotypes mis-imputed by BEAGLE (the software offering the higher accuracy), a total of 384 markers across the genome reached the 0.3 imputation error rate (Fig. S2). In general these markers were randomly distributed across the genome with the exception of seven mis-imputation hot-spot regions identified according our criteria on chromosomes 2, 3, 9, 16, 17, 23 and 25 (Table S2, Fig. S2). Whereas in general the number of mis-imputed markers in these hot-spot regions was low (range: 5–8), a region on OAR25 (1.703–2.665 Mb) showed a remarkable high number ($n = 73$) of mis-imputed markers (see Table S2). This hot-spot region of imputation errors on OAR25 could be related to the low quality of the reference genome ensemble in that region or to a high recombination rate region, as high recombination rates hinder the accuracy of haplotype phasing and genotype imputation (Weng *et al.* 2014). In any case, and in addition to the expected improvements of quality of the sheep reference genome, it would be advisable that additional markers be added in an updated version of the 50K-chip for that specific region of OAR25 to increase the accuracy of imputation approaches that could be performed in the coming future in different sheep populations.

LD and N_e estimations in the Churra genome based on the imputed HD-chip genotypes

Due to its slightly higher imputation accuracy, and despite its longer computational time, BEAGLE was used to obtain the imputed HD-chip genotypes for the whole study population across the 26 autosomes without using the masking strategy implemented for the accuracy estimation.

These imputed genotypes were considered for studying the high density-based haplotype structure of the Churra sheep genome. A total of 1,670 genotyped animals with 819 869 800 pairs of imputed SNPs were used to re-estimate the LD parameters in our Churra resource

Table 1 Mean linkage disequilibrium (LD) estimated among synthetic SNPs at different physical distances based on the HD-chip genotypes imputed in the Churra population analyzed in this study.

Distance	*n*	r^2	SD r^2	D'	SD D'
<10 kb	900 938	0.2754	0.3044	0.8602	0.2622
10–20 kb	1 047 956	0.1934	0.2460	0.7453	0.31850
20–40 kb	2 009 443	0.1463	0.2031	0.6686	0.3331
40–60 kb	2 017 826	0.1131	0.1666	0.6031	0.3347
60–100 kb	4 017 750	0.0884	0.1347	0.5501	0.3289
100–200 kb	9 991 072	0.0664	0.1006	0.4978	0.3174
200–500 kb	29 784 956	0.0532	0.0779	0.4604	0.3060
500 kb–1 Mb	49 256 569	0.0468	0.0679	0.4371	0.2976
1–2 Mb	97 395 757	0.0406	0.0594	0.4098	0.2869
2–5 Mb	284 771 507	0.0308	0.0463	0.3601	0.2659
5–10 Mb	451 518 459	0.0210	0.0323	0.2999	0.2367
10–20 Mb	819 664 122	0.0136	0.0215	0.2437	0.2068
20–50 Mb	1 833 041 774	0.0080	0.0127	0.1904	0.1792
50 Mb	1 525 242 214	0.0061	0.0095	0.1688	0.1687

n, number of SNP marker pairs included in each of the physical distance bins defined; r^2 and SD r^2, mean and standard deviation of r^2 estimated for each of the physical distance bins defined; $^{(d,e)}D'$ and SD D', mean and standard deviation of D' estimated for each of the physical distance bins defined.

population. As shown in Table 1, where the mean r^2 and D' values estimated for the different marker distance bins are provided, the two LD parameters showed a decreasing trend with increased physical distance between SNPs. For markers up to 10 kb apart, the average r^2 was 0.275 ($D' = 0.860$), and at more than 50 Mb inter-marker distance, the average r^2 was 0.006 ($D' = 0.169$) (Fig. 1; HDImp dataset). The persistence of LD over short distances showed pronounced LD decay, with average r^2 values of 0.21, 0.17 and 0.14 for markers separated by 10, 20 and 30 kb respectively (Fig. 2a, and see zoom-in view in Fig. 2b). The half-length of LD (r^2) was 0.024 (Fig. S3), which corresponded to a distance between SNPs of 1.03 Mb. The LD pattern based on the HD-chip imputed genotypes was compared with the LD estimates reported for the same population based on the 50K-chip genotypes given by García-Gámez *et al.* (2012) and also with the estimates obtained based on the HD-chip reference dataset and the imputed genotyped after the removal of the mis-imputation hot-spot SNPs (Table S3; Fig. 1).

The overall pattern of the LD (r^2) estimates across 1-Mb-interval SNP-binning categories (Fig. S3) obtained here for the imputed HD-chip dataset was very similar to that presented for the 50K-chip in the same population (García-Gámez *et al.* 2012). Through the graphic comparison between the LD patterns presented in Fig. 1 for these two datasets (HDImp vs. 50K) we can see that, when expressed in r^2 units, the LD estimates at shorter distances are lower for the HD-chip imputed dataset than for the 50K-chip dataset, whereas at longer distances there are not remarkable differences between the two datasets estimates. However, when expressed in D' units, the estimates of the imputed dataset were slightly greater across all marker distances than those based on the 50K-chip (Table 1, Fig. 1). The difference in the pattern between D' and r^2 estimates can be explained by the different way in which these LD measurements depend on the marginal allele frequencies (Pritchard & Przeworski 2001). As explained by Evans & Cardon (2005), r^2 is less influenced by a ceiling

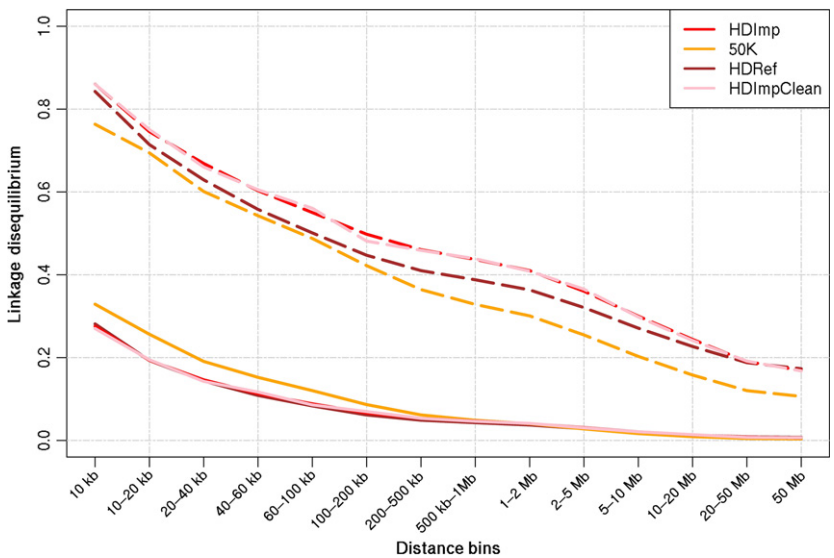


Figure 1 Average linkage disequilibrium (LD) as a function of genomic distance between markers estimated for Churra sheep based on the four datasets compared in the present study: 50K, the 50K-chip dataset analyzed by García-Gámez *et al.* (2012); HDImp, the 50K- to HD-chip imputed dataset obtained with BEAGLE in the present study; HDRef, the HD-chip reference population dataset; HDImpClean, the HD-chip imputed dataset after removal of the mis-imputed SNPs included in the identified mis-imputation hot-spot regions. The LD values (*y*-axis), provided as D' (dashed lines) and r^2 (solid lines), are plotted against inter-marker distance bins (*x*-axis). For each case, the total number of marker pairs were assigned according to their physical distance into 14 categories: <10, 10–20, 20–40, 40–60, 60–100, 200–500 kb, 0.5–1, 1–2, 2–5, 5–10, 10–20, 20–50 and >50 Mb.

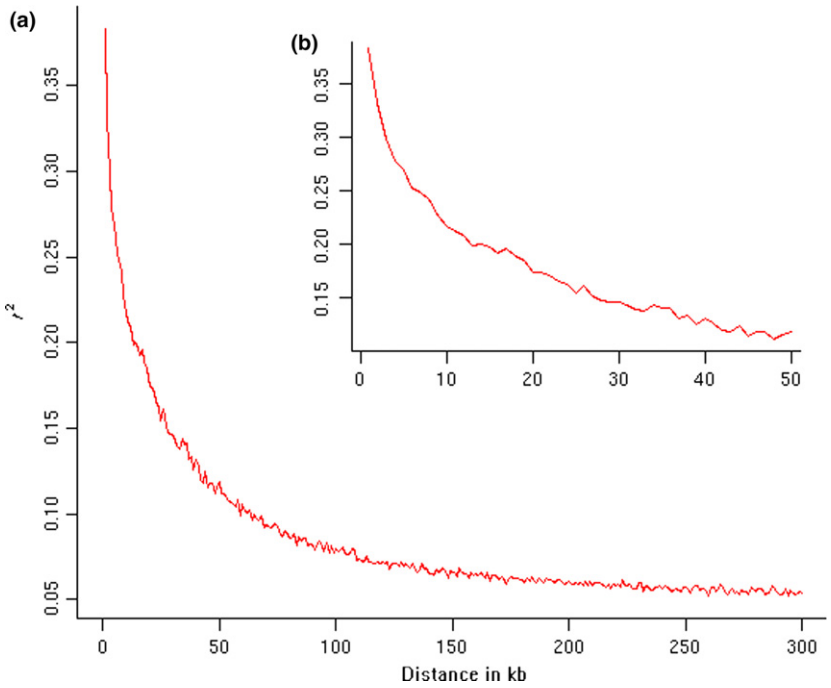


Figure 2 Linkage disequilibrium, expressed as r^2, calculated in the Churra sheep population based on the imputed genotypes of the HD-chip (490 940 SNPs), and after being averaged in 1-kb bins (a) plotted for short genomic distances (up to 300 kb) and (b) a zoomed view providing a detailed view of the 0–50 kb.

effect. Based on Fig. 1, r^2 appears as a more appropriate measurement of LD than D' to assess the gain in accuracy provided by a higher marker density. The lower r^2 estimates in the shorter distances observed for the imputed HD-chip dataset when compared with the 50K-chip dataset are in agreement with the substantially lower estimate for the half-length of LD estimated in this work ($r^2 = 0.024$ for a marker distance of 1.03 Mb) when compared with that reported by García-Gámez *et al.* (2012) ($r^2 = 0.033$ for a marker distance 2.5 Mb). Globally, these results show that the HD-chip provides a substantially improved definition for LD estimations at short distances and that the Churra sheep genome showed an even a sharper decline in LD than that previously estimated based on the 50K-chip.

The comparison of the LD estimates based on the imputed HD-chip dataset with those obtained based directly on the HD-chip reference population dataset (HDImp vs. HDRef in Fig. 1; Table 1 vs. Table S3) showed very similar values when expressed in r^2 units, although the D' estimates were slightly lower when they were based on the 335 animals of the reference population than in the whole population imputed genotypes. Hence, imputation does not appear to introduce relevant inaccuracy in the LD estimations and the r^2 estimates are not affected by the mis-imputation errors. This probably is due to the fact that when selecting the reference population to consider in this work we followed the same structure as the global population, indicating the major importance that the structure of reference populations may have in imputation-based approaches. Regarding the LD, estimates expressed in D' units may be overestimated due to loss of detection of some rare recombinants linked to the implicit imputation error rate. From the other comparison, we see that the removal of the 109 markers included in the seven mis-imputation hot-spot regions did not have any impact on the LD estimates (HDImp vs. HDImpClean in Fig. 1; Table 1 vs. Table S3), suggesting that for the level of imputation accuracy reported here and the general random distribution of the mis-imputed genotypes, the imputation process did not introduce any relevant bias on the accuracy of the LD estimates.

Because D' correlates poorly between populations (Evans & Cardon 2005), we have used the LD estimates expressed in r^2 units to compare the extension of LD between Churra and other sheep populations analyzed with the HD-chip by by Kijas *et al.* (2014). These comparisons showed that the estimated LD r^2 values at 10 kb (0.2107) and 70 kb (0.088) in Churra sheep were slightly higher than in Merino and lower than in Poll Dorset, Suffolk and Border Leicester (Kijas *et al.* 2014). By plotting the updated HD-chip based estimates for Churra against the LD curve presented by these authors for the 50K-chip density, the decay of LD at the shortest distances in Churra would be very close to that shown by Qezel sheep. These observations seem to be directly related to the high level of genetic diversity reported for Churra based on the 50K-chip Sheep HapMap project analysis (Kijas *et al.* 2012) and updated here based on the HD-chip (Appendix S1; Section 2.1), which show a diversity level very close to that reported for the Merino breeds, which are among the most diverse sheep breeds presented in previous studies (Kijas *et al.* 2012, 2014). Even for the genetic distance (D) parameter, our estimates suggested a slightly greater genetically divergence for Churra (0.268) than for Merino (0.262–0.266; Kijas *et al.* 2014). This is in agreement with the high levels of genetic diversity reported

for ovine Mediterranean breeds, likely related to the first migrations of Neolithic communities and their animals from the domestication centre (Kijas *et al.* 2012).

In any case, for the aim of detecting QTL–SNP associations, it is clear that in a breed such as Churra, which has a low extent of LD, the HD-chip is a powerful tool that will increase the opportunities to identify genuine associations that could not be detected with the 50K-chip. Considering the distinction between 'usable' and 'detectable' LD, as suggested by Du *et al.* (2007), and considering that, in continuous traits, the power to detect a causal mutation will depend on the size effect of the QTL, the allele frequencies in the population, and so on, we could follow these authors and use the following formula in case–control studies to estimate the required sample size to genotype the causal mutation: N/r^2, where N is the sample size needed for genotyping the causal mutation (Du *et al.* 2007). Hence, when considering that, for Churra sheep, the LD values at an average distance between adjacent markers for the 50K-chip and the HD-chip densities were $r^2 = 0.1018$ (at 60 kb) and $r^2 = 0.3759$ (at 6 kb) respectively, we could expect that, for the same population size, the HD-chip sample size (N) would yield a significant increase in statistical power to detect genes controlling complex traits. With reference to genomic predictions, some studies have suggested the need for r^2 values greater than 0.2 (Meuwissen *et al.* 2001; Calus *et al.* 2008). Hence, based on the average LD estimates mentioned above for the two chips, it seems that, in Churra sheep, genomic selection could be implemented efficiently if based only on the HD-chip density.

Effective population size

The N_e estimates for the four considered datasets represented in Fig. S4 show the same global pattern and reflect the important decrease in population size that occurred 250 to 25 generations ago and the slower decrease after the start of the breeding scheme in 1986 (de la Fuente *et al.* 1995). The confidence intervals of the N_e estimates obtained with the HD-chip imputed dataset are also represented in this figure and provided in Table S4. When comparing the estimates between the 50K-chip [considering the estimates for the 1 cM ~ 1 Mb ratio reported by García-Gámez *et al.* (2012)] with the HD-related datasets (Table S4, Fig. S4), we see very similar values for the range between 10 and 25 generations ago, whereas the most recent estimates based on the 50K-chip (one to five generations ago) appear to be over-estimated as result of the lack of identification of some crossing-overs due to the lower marker density. On the contrary, the 50K-chip N_e estimates for the most distant generations are lower than those estimated for the high density datasets, probably as result of the low inaccuracy of the 50K-chip to estimate the extent of LD at the shortest distances. The lower and more accurate estimate of inbreeding reported here for one generation ago, $N_e = 83$, is substantially lower than are the estimates based on the 50K-chip (García-Gámez *et al.* 2012) ($N_e = 128$ as an average of different conversion ratios; $N_e = 159$ based on the 1 cM ~ 1 Mb ratio) and is also lower than the threshold estimated to ensure the viability in the long term of an animal population ($N_e = 100$) (Meuwissen 2009). This updated information highlights the need for considering inbreeding control as a major point of the breeding programme of the dairy sheep Churra population. Compared with the results of the HapMap population based on the 50K-chip (Kijas *et al.* 2012), 50 generations ago, when Churra had a $N_e = 600$, the updated estimate with the HD-chip for this time point was similar although slightly lower, $N_e = 513$. This estimate is intermediate between that reported for Merino ($N_e = 833$) and those values reported for other breeds showing a greater extent of LD and lower genetic diversity (Border Leicester, $N_e = 242$; Polled Dorset, $N_e = 318$) (Kijas *et al.* 2014). When comparing the N_e estimates based on different HD-chip datasets, we see that the removal of the mis-imputation hot-spot SNPs do not have any impact across any of the time points considered (HDImp vs. HDImpClean in Fig. S4). However, we see that the global error rate introduced by imputation (HDImp vs. HDImpRef in Fig. S4) tends to decrease the N_e estimates for the most distant generations, whereas imputation did not have any effect on the most recent estimates. This appears to correlate with the lower estimates of LD in D' units observed for the HD-chip reference population dataset when compared with the estimates of imputed HD-chip genotypes (Fig. 1) and the influence of the low accuracy to estimate N_e in very distant time points.

Detection of ROHs in the Churra genome

The chromosome showing the highest number of ROHs was OAR2 (14 025 ROHs) (Fig. 3), with a significantly higher number of ROHs than for OAR1 and OAR3, which had similar lengths. Our ROH analyses showed that, on average, approximately 7% of the whole genome was under ROH segments, with the highest coverage shown by OAR24 (10.4%) and OAR26 (9.7%). The chromosomes showing the lowest coverage of ROHs were OAR1 (4.6%) and OAR7 (4.9%) (Fig. 3). The estimation of the total length of ROHs per individual (Fig. S5) showed that 98% of the individuals had a total ROH length greater than 50 Mb, approximately 70% of the individuals had a total ROH length greater than 100 Mb and 0.007% had a total ROH length greater than 500 Mb. Based on the distribution of ROHs in the different bins (Table 2), roughly 88% of the ROHs had lengths of less than 5 Mb, and 0.008% of the ROHs spanned more than 20 Mb. This distribution of ROHs observed for Churra sheep across the different chromosomes shows a similar pattern to that described by Al-Mamun *et al.* (2015) and focused on both pure and cross-sheep breeds with genotypes for the 50K-chip.

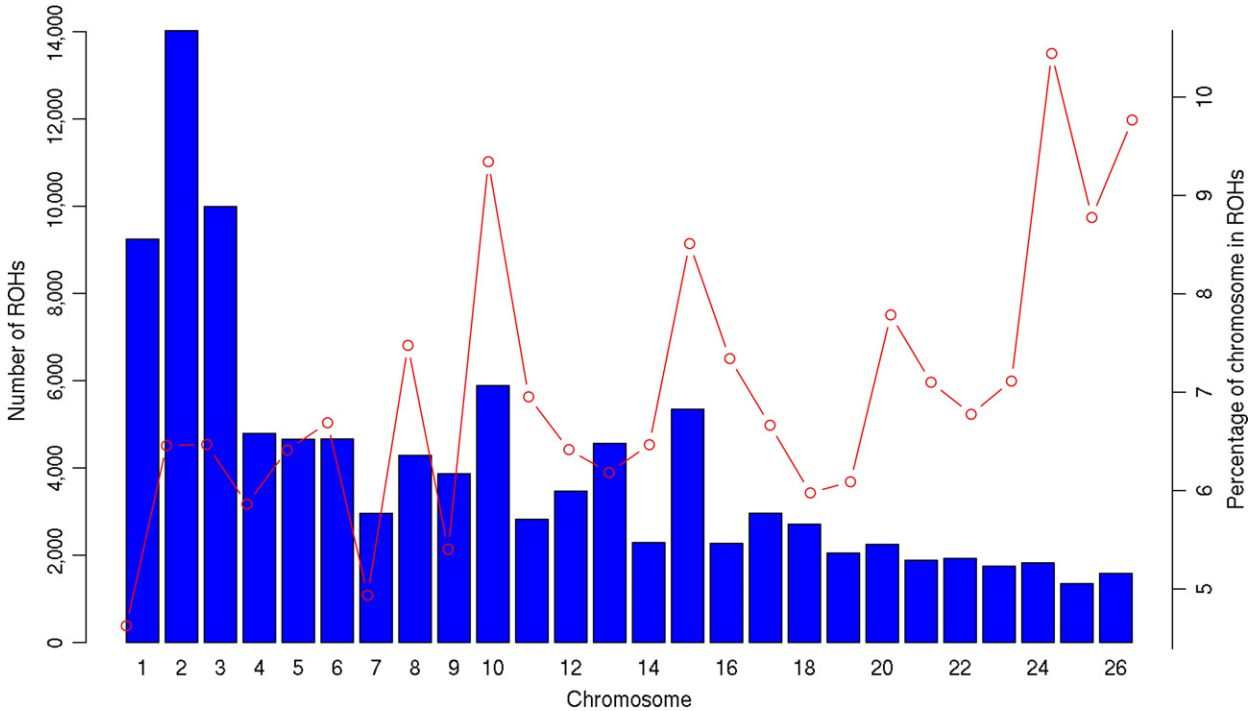


Figure 3 The number of runs of homozygosity (ROH) across the 26 autosomal ovine chromosomes, represented by blue bars, identified in all 1,670 daughters of the analyzed Churra population. The line plot represents the proportion of ROHs with respect to the total length of each chromosome, which was estimated by calculating the mean ROH across the 1,670 considered individuals in each chromosome and then estimating it as a percentage across the whole chromosomal length.

However, the number of ROH segments was much higher in our estimation, which could be attributed to the higher genotyping density of our dataset. The distribution of the sum of ROHs per individual (Fig. S5) also showed a similar distribution to that of pure breeds reported by Al-Mamun *et al.* (2015). All of the individuals had ROHs greater than 1 Mb and approximately 99% of individuals had ROHs longer than 5 Mb, which are higher estimates than those reported for Border Leicester, Merino and Poll Dorset pure breeds, for which approximately 88% of the animals had at least one ROH longer than 5 Mb (Al-Mamun *et al.* 2015).

Considering the ROHs that were common to different individuals, a total of 12 238 ROH consensus regions were identified in our population, with the number of occurrences per region ranging from 133 animals (7.89%) on OAR25 to 1,057 animals (62.69%) on OAR8 (Table 3). The OAR8 region involving the largest number of occurrences was found within the 31.9–32.1 Mb interval, followed by the ROH segment located between 12.24 and 12.25 Mb on chromosome 2 (950 occurrences, 56%). Other chromosomes had either shorter consensus regions or lower occurrences. Interestingly, the OAR8 region including the ROH region with the largest number of occurrences (Table 3) overlaps with a region identified as under selection (related to marker *s20065*, with position 32 159 065 bp according to the updated reference genome Oar_v3.1) in the analysis of the Sheep HapMap project dataset reported by Kijas *et al.* (2012). According to SheepQTLdb (http://www.animalgenome.org/cgi-bin/QTLdb/OA/index), this region is related to QTL for gastrointestinal nematode resistance (Crawford *et al.* 2006) and fat carcass content (Cavanagh *et al.* 2010). This region includes two genes: *BVES* (*blood vessel epicardial substance*), which is expressed in cardiac and skeletal muscle and could play an important role in the development of these tissues, and *LIN28B* (*lin-28 homolog B*), which is a suppressor of microRNA biogenesis and has been reported as the first genetic determinant regulating the timing of human pubertal growth and development (Ong *et al.* 2009).

Inbreeding coefficient estimation using the HD-chip genotypes

The inbreeding coefficients estimated based on the HD-chip reference genotypes for the animals born from 2001 to 2008 showed mean values of $F_{\mathrm{Mol}} = -0.0228$ (SD = 0.011), $F_{\mathrm{ROH}} = 0.042$ (SD = 0.019), $F_{\mathrm{ROH\text{-}Q}} = -94.127$ (SD = 43.595) and $F_{ROH-E}{}^{m} = -0.375$ (SD = 0.239). The distributions of the genomic inbreeding coefficients estimated here and also the inbreeding coefficient based on the pedigree estimated by García-Gámez *et al.* (2012) are shown in Fig. S6a.

Considering the average values of the standardized inbreeding coefficient estimates for the complete population

Table 2 Mean number of runs of homozygosity (ROHs) grouped in four length bins (in Mb) identified per chromosome based on HD-chip imputed data of the Churra sheep population analyzed here.

Chromosome	<5 Mb	5–10 Mb	10–20 Mb	>20 Mb
1	8 139	719	315	73
2	12 782	855	286	101
3	8 678	897	316	100
4	4 199	386	145	62
5	4 079	407	131	43
6	3 998	401	201	67
7	2 634	191	100	35
8	3 727	368	166	27
9	3 524	217	104	24
10	5 304	379	150	57
11	2 536	232	48	8
12	3 106	222	113	28
13	4 194	253	103	13
14	2 067	148	53	22
15	4 771	433	117	27
16	1 927	232	95	17
17	2 620	217	95	32
18	2 466	159	63	23
19	1 842	152	44	13
20	1 977	191	75	5
21	1 672	158	46	9
22	1 757	110	40	21
23	1 475	188	58	27
24	1 559	211	52	4
25	1 165	124	46	13
26	1 334	194	44	12
Total ROHs	93 532	8044	3006	863

(Fig. S6b) and according to the year of birth, from 2001 to 2008 (Fig. 4), overall we observed that all of the genomic-based coefficients resulted in higher estimates of inbreeding than did the pedigree-based coefficient. The overall profile of F_{Mol} suggested a higher inbreeding level than that of F_{ROH}, whereas the two inbreeding coefficients, considering the lengths of the ROHs—$F_{\mathrm{ROH\text{-}Q}}$ and $F_{ROH-E}{}^{m}$—which according to Gomez-Raya *et al.* (2015) are related to recent inbreeding, suggested the highest estimates of inbreeding. In any case, the general profiles of the genomic-based inbreeding estimates obtained across the years considered (Fig. 4) showed the same general pattern, with a slow and continuous increase in the inbreeding level from 2001 to 2008, during which 2003 showed a more pronounced increase, which could be related to the introduction in 2003 of udder and body morphology traits (de la Fuente *et al.* 2011) as selection objectives in the breeding programme. There were also some differences depending on the estimation approach implemented. Hence, there was a decrease in the inbreeding estimates detected in 2002 by F_{Mol} and F_{ROH} but not by the two coefficients related to recent inbreeding, $F_{\mathrm{ROH\text{-}Q}}$ and $F_{ROH-E}{}^{m}$. Also, in 2006 there was a certain peak in inbreeding detected by the three ROH-related coefficients but not by F_{Ped} or F_{Mol}, which could be related to the selection performed for scrapie-resistant haplotypes at the *PRNP* locus based on the requirements of the European Commission (2003).

As previously shown by García-Gámez *et al.* (2012) based on the use of the 50K-chip, our comparison of inbreeding coefficients indicate that the genomic-based estimates are more reliable than are the pedigree-based coefficient. In both cases, by using the 50K-chip or HD-chip genotypes, we could identify inbreeding that was not identified on the basis of the available pedigree. Using the same approach to estimate these genomic-based estimates, the higher accuracy provided by the HD-chip results in higher inbreeding estimates than with the estimates based on the 50K-chip (Fig. 4, F_{Mol} vs. F_{Mol}-50K). Based on the HD-chip imputed dataset, we also estimated F_{ROH} to exploit the information from the ROH analysis previously presented. In contrast to F_{Mol}, this coefficient ignores fragments consisting of a single or a few contiguous homozygous SNPs. In our Churra population, comparison of the standardized estimates for these two coefficients showed that the estimates of inbreeding were lower when the total ROH length was considered in the estimations. Hence, as a general estimation of inbreeding in our resource population, we believe that the average estimate of $F_{\mathrm{ROH}} = 0.042$ (SD = 0.019) could be the most appropriate. Even based on a subset of animals from the selection nucleus, this estimate was less than the critical level of 0.0625 suggested by Li *et al.* (2009) considering the inbreeding level obtained when mating cousins, which suggests that the inbreeding levels in Churra sheep are under-controlled in the general population. In any case, and considering the reduced updated N_e estimation reported here for this population, we think that if genotypic information would be routinely available in the population, the estimation of F_{ROH} per individual flock could be an advisable strategy for the early identification of negative effects that are directly related to inbreeding depression. This strategy could avoid the appearance of inherited diseases, such as the two recently reported in Churra sheep flocks (Pérez *et al.* 2013; Benavides *et al.* 2015). Remarkably, the causal mutations of these two disorders were identified on the basis of the ROH-mapping strategy (Suárez-Vega *et al.* 2013, 2015).

As suggested by Gomez-Raya *et al.* (2015), ROH information could be further exploited by considering the length of ROH fragments. ROHs due to recent inbreeding will tend to be longer because there has been little opportunity for recombination to break up segments that are identical by descent (Kirin *et al.* 2010). In our population, the two coefficients related to recent inbreeding—$F_{\mathrm{ROH\text{-}Q}}$ and $F_{ROH-E}{}^{m}$ coefficients—showed, after standardization, higher inbreeding estimates than did the general F_{ROH} (Fig. 4). This could be easily explained by the active selection scheme of the Churra sheep population. In any case, as noted earlier, the inbreeding levels estimated here did not show a risk to the genetic diversity levels of the population, which according to the present study, were quite high compared with other sheep populations.

Table 3 Runs of homozygosity (ROHs) consensus regions identified in this work with the largest number of occurrences per chromosome (Chr.).

Chr.	No. of ROH occurrences	SNP1_id	SNP1_bp	SNP2_id	SNP2_bp	Length (kb)
1	523	oar3_OAR1_251059482	251 059 482	oar3_OAR1_251059482	25 105 9482	0
2	950	oar3_OAR2_122498848	122 498 848	oar2_130852040_X.1	122 524 397	25.549
3	901	oar3_OAR3_183058396	183 058 396	oar3_OAR3_183070877	183 070 877	12.481
4	564	oar3_OAR4_48589193	48 589 193	oar3_OAR4_48660356	48 660 356	71.163
5	705	oar3_OAR5_38301098	38 301 098	oar3_OAR5_38310097	38 310 097	8.999
6	597	oar3_OAR6_6198787	6 198 787	oar3_OAR6_6215153	6 215 153	16.366
7	208	oar3_OAR7_50405139	50 405 139	oar3_OAR7_50405139	50 405 139	0
8	1057	oar3_OAR8_31903148	31 903 148	s20065.1	32 159 065	255.917
9	490	oar3_OAR9_77387147	77 387 147	oar3_OAR9_77790278	77 790 278	403.131
10	642	oar3_OAR10_7290758	7 290 758	oar3_OAR10_7290758	7 290 758	0
11	296	oar3_OAR11_32805104	32 805 104	s28097.1	33 120 184	315.08
12	492	oar3_OAR12_78544792	78 544 792	oar3_OAR12_78676654	78 676 654	131.862
13	503	oar3_OAR13_42509642	42 509 642	oar3_OAR13_42644675	42 644 675	135.033
14	295	oar3_OAR14_34447593	34 447 593	oar3_OAR14_34467539	34 467 539	19.946
15	940	oar3_OAR15_927699	927 699	oar3_OAR15_927699	927 699	0
16	179	oar3_OAR16_32283318	32 283 318	oar3_OAR16_32383673	32 383 673	100.355
17	278	oar17_591689.1	472 363	oar3_OAR17_519983	519 983	47.62
18	253	s36862.1	65 975 442	oar3_OAR18_66305732	66 305 732	330.29
19	269	oar3_OAR19_496065	496 065	DU264531_279.1	619 479	123.414
20	430	oar3_OAR20_707588	707 588	oar3_OAR20_722921	722 921	15.333
21	468	oar3_OAR21_18147099	18 147 099	oar3_OAR21_18241100	18 241 100	94.001
22	385	oar3_OAR22_50317587	50 317 587	oar3_OAR22_50326245	50 326 245	8.658
23	259	oar3_OAR23_43345164	43 345 164	OAR23_45926303.1	43 397 227	52.063
24	210	oar3_OAR24_457915	457 915	oar3_OAR24_571267	571 267	113.352
25	133	oar3_OAR25_19160982	19 160 982	oar3_OAR25_19505987	19 505 987	345.005
26	211	oar3_OAR26_39558587	39 558 587	oar3_OAR26_39558587	39 558 587	0

The positions in bp according the reference genome (Oar_v3.1) for the start marker (SNP1) and end marker (SNP2) are indicated here, together with the length (kb) defined by the ROH interval.

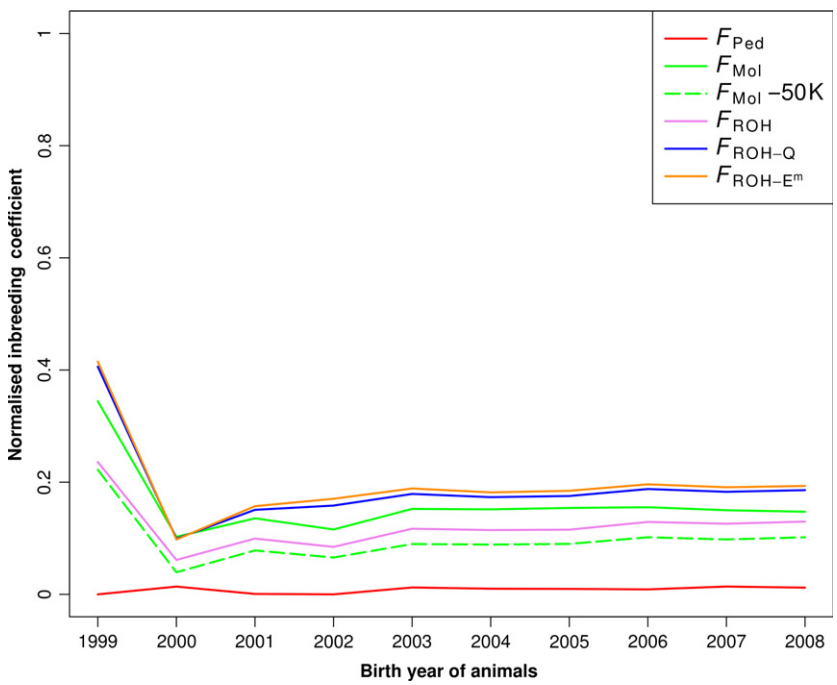


Figure 4 Average normalized inbreeding coefficients estimated for animals of the Churra sheep population analysed here born between 1999 and 2008, including the pedigree-based coefficient and four additional genomic-based coefficients based on the imputed HD-chip genotypes generated in this study. Inbreeding coefficients abbreviations: $F_{PED,}$ pedigree-based inbreeding coefficient calculated for this population as described by García-Gámez *et al.* (2012); $F_{Mol,}$ inbreeding coefficient based the observed vs. expected number of homozygous genotypes; F_{ROH}, inbreeding coefficient based on the total ROH content; F_{ROH-Q}, inbreeding coefficient based on the quantiles of the distribution of the length of ROH; $F_{ROH-E}{}^{m}$, inbreeding coefficient based on fitting an exponential distribution to the ROH-length distribution and using the mean of the exponential density.

Conclusions

The imputation strategy of 50K-chip to HD-chip genotypes optimized here appears to be appropriate for future use in Churra sheep. The updated information regarding the extent of LD in the Churra based on the HD-chip density will be considered in the future design of a genomic selection-based programme and in the planning of marker-phenotype association studies. We have also assessed the impact of imputation on LD and N_e estimates. Considering that the extent of LD at short marker distances is even less than initially estimated based on the 50K-chip, the Churra sheep dairy population analysed here appears to show one of the lowest LD extension levels described in sheep, which is correlated with the high genetic diversity levels reported for this breed. The identification of ROHs described herein not only provide information about specific regions that could be related to selection events but also has been exploited to improve the accuracy of inbreeding coefficient estimates and to assess the impact of the current active breeding programme on inbreeding.

Acknowledgements

P. K. Chitneedi is a grateful grantee of fellowship funded by the Junta de Castilla y León Government and the European Social Fund. B. Gutiérrez-Gil is funded through the Spanish Ramón y Cajal Programme (RYC-2012-10230) from the Spanish Ministry of Economy and Competitiveness (MINECO). Financial support was received from project AGL2015-66035-R funded by MINECO and co-funded by the European Regional Development Fund. The support and availability to the computing facilities of the Foundation of Supercomputing Center of Castile and León (FCSCL) (http://www.fcsc.es) is greatly acknowledged.

Conflict of interest

The authors declare they have no conflict of interests in this research.

References

Al-Mamun H.A., Clark S.A., Kwan P. & Gondro C. (2015) Genome-wide linkage disequilibrium and genetic diversity in five populations of Australian domestic sheep. *Genetics Selection Evolution* **47**, 90.

Atlija M., Arranz J.J., Valladares M.M. & Gil B.G. (2016) Detection and replication of QTL underlying resistance to gastrointestinal nematodes in adult sheep using the ovine 50K SNP array. *Genetics Selection Evolution* **48**, 4.

Benavides J., Suárez-Vega A., Gutiérrez J., de la Hera S., Fuertes M., Delgado L., Fernández M., Ferreras Mdel C., Arranz J.J. & Pérez V. (2015) Generalized severe junctional epidermolysis bullosa with congenital absence of skin in Churra lambs. *Veterinary Dermatology*, **26**, 367–73, e82–3.

Browning B.L. & Browning S.R. (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**, 210–23.

Browning B.L. & Browning S.R. (2013) Improving the accuracy and efficiency of identity-by-descent detection in population data. *Genetics* **194**, 459–71.

Calus M.P., Meuwissen T.H., Roos A.P.W. & Veerkamp R.F. (2008) Accuracy of genomic selection using different methods to define haplotypes. *Genetics* **178**, 553–61.

Cavanagh C.R., Jonas E., Hobbs M., Thomson P.C., Tammen I. & Raadsma H.W. (2010) Mapping quantitative trait loci (QTL) in sheep. III. QTL for carcass composition traits derived from CT scans and aligned with a meta-assembly for sheep and cattle carcass QTL. *Genetics, Selection, Evolution* **42**, 36.

Crawford A.M., Paterson K.A., Dodds K.G. *et al.* (2006) Discovery of quantitative trait loci for resistance to parasitic nematode infection in sheep: I. Analysis of outcross pedigrees. *BMC Genomics* **7**, 178.

Du F.X., Clutter A.C. & Lohuis M.M. (2007) Characterizing linkage disequilibrium in pig populations. *International Journal of Biological Sciences* **3**, 166–78.

European Commission (2003) 2003/100/EC: Commission decision of 13 February 2003 laying down minimum requirements for the establishment of breeding programmes for resistance to transmissible spongiform encephalopathies in sheep. *Official Journal of the European Union* **41**, 41–5.

Evans D.M. & Cardon L.R. (2005) A comparison of linkage disequilibrium patterns and estimated population recombination rates across multiple populations. *American Journal of Human Genetics* **76**, 681–7.

de la Fuente L.F., Baro J.A. & San Primitivo F. (1995) Breeding programme for the Spanish Churra sheep breed. *Cahiers Options Méditerranéennes* **11**, 165–72.

de la Fuente L.F., Gonzalo C., Sánchez J.P., Rodríguez R., Carriedo J.A. & San Primitivo F. (2011) Genetic parameters of the linear body conformation traits and genetic correlations with udder traits, milk yield and composition, and somatic cell count in dairy ewes. *Canadian Journal of Animal Science* **91**, 585–91.

García-Gámez E., Sahana G., Gutiérrez-Gil B. & Arranz J.J. (2012) Linkage disequilibrium and inbreeding estimation in Spanish Churra sheep. *BMC Genetics* **13**, 43.

Gomez-Raya L., Rodríguez C., Barragán C. & Silió L. (2015) Genomic inbreeding coefficients based on the distribution of the length of runs of homozygosity in a closed line of Iberian pigs. *Genetics, Selection, Evolution* **47**, 81.

Hayes B.J., Bowman P.J., Daetwyler H.D., Kijas J.W. & van der Werf J.H.J. (2012) Accuracy of genotype imputation in sheep breeds. *Animal Genetics* **43**, 72–80.

Kijas J.W., Lenstra J.A., Hayes B. *et al.* (2012) Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. *PLoS Biology* **10**, e1001258.

Kijas J.W., Porto-Neto L., Dominik S., Reverter A., Bunch R., McCulloch R., Hayes B.J., Brauning R. & McEwan J. (2014) Linkage disequilibrium over short physical distances measured in sheep using a high-density SNP chip. *Animal Genetics* **45**, 754–7.

Kirin M., McQuillan R., Franklin C.S., Campbell H., Mckeigue P.M. & Wilson J.F. (2010) Genomic runs of homozygosity record population history and consanguinity. *PLoS One* **5**, 1–7.

Li M.H., Strandén I. & Kantanen J. (2009) Genetic diversity and pedigree analysis of the Finnsheep breed. *Journal of Animal Science* **87**, 1598–605.

Meuwissen T. (2009) Genetic management of small populations: a review. *Acta Agriculturae Scandinavica, Section A—Animal Science* **59**, 71–9.

Meuwissen T.H., Hayes B.J. & Goddard M.E. (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics* **157**, 1819–29.

Ong K.K., Elks C.E., Li S. *et al.* (2009) Genetic variation in *LIN28B* is associated with the timing of puberty. *Nature Genetics* **41**, 729–33.

Pérez V., Suárez-Vega A., Fuertes M., Benavides J., Delgado L., Ferreras M.C. & Arranz J.J. (2013) Hereditary lissencephaly and cerebellar hypoplasia in Churra lambs. *BMC Veterinary Research* **9**, 156.

Pritchard J.K. & Przeworski M. (2001) Linkage disequilibrium in humans: models and data. *American Journal of Human Genetics* **69**, 1–14.

Purcell S., Neale B., Todd-Brown K. *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics* **81**, 559–75.

Raadsma H. (2010) LD analyses in the sheep HapMap project using an ovine 50K SNP array linkage disequilibrium–LD. *Proceedings of the Plant and Animal Genome Conference* 9th–13th January, 2010. San Diego, USA.

Sargolzaei M., Chesnais J.P. & Schenkel F. (2012) *Efficient Combined Family and Population Imputation in Large Data Sets*. GEB Open Industry Session, Saint Hyacynthe, QC, Canada. October 30, 2012.

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

Slatkin M. (2008) Linkage disequilibrium—understanding the evolutionary past and mapping the medical future. *Nature Reviews Genetics* **9**, 477–85.

Suárez-Vega A., Gutiérrez-Gil B., Cuchillo-Ibáñez I., Saez-Valero J., Pérez V., García-Gámez E., Benavides J. & Arranz J.J. (2013) Identification of a 31-bp deletion in the *RELN* gene causing lissencephaly with cerebellar hypoplasia in sheep. *PLoS One* **8**, 1–12.

Suárez-Vega A., Gutiérrez-Gil B., Benavides J., Perez V., Tosser-Klopp G., Klopp C., Keennel S.J. & Arranz J.J. (2015) Combining GWAS and RNA-seq approaches for detection of the causal mutation for hereditary Junctional epidermolysis bullosa in sheep. *PLoS One* **10**, e0126416.

Weng Z.-Q., Saatchi M., Schnabel R.D., Taylor J.F. & Garrick D.J. (2014) Recombination locations and rates in beef cattle assessed from parent-offspring pairs. *Genetics Selection, Evolution* **46**, 34.

Supporting information

Additional supporting information may be found online in the supporting information tab for this article:

Figure S1 Distribution of minor allele frequency (MAF) estimates for the markers included in the ovine HD-chip based on the original genotypes of the 335 Churra individuals genotyped for this chip in this study.

Figure S2 Manhattan plot showing the imputation error rate with each data point representing the error rate of each SNP in the imputed HD genotype data across the 26 autosomal chromosomes after imputing the genotypes from 50k to HD-chip using BEAGLE V4.0.

Figure S3 Graphical representation of the pair-wise LD (r^2) estimates for the 490 940 imputed SNP markers considered in this work plotted against the distance between marker pairs by averaging the inter-marker distance into 1-Mb bins.

Figure S4 The effective population size (N_e) calculated for the past 250 generations (T) for Churra sheep based on the four datasets compared in the present study: 50K, the 50K-chip dataset analyzed by García-Gámez *et al.* (2012); HDImp, the 50K- to HD-chip imputed dataset obtained with BEAGLE in the present study, HDRef, the HD-chip reference population dataset; HDImpClean, the HD-chip imputed dataset after removal of the mis-imputed SNPs included in the identified mis-imputation hot-spot regions.

Figure S5 Total length of ROH across all the ovine autosomes per each individual.

Figure S6 Histograms of frequencies of the inbreeding coefficients estimated in the Churra population analyzed here, including the pedigree-based coefficient and four additional genomic-based coefficients based on the imputed HD-chip genotypes generated in this study.

Table S1 Estimated accuracy of the imputation performed from the 50K-chip genotypes to HD-chip genotypes in the studied population for the 26 autosomal chromosomes using BEAGLE v4.0 and FIMPUTE v2.2 software by applying a cross-validation approach with a total of 14 iterations.

Table S2 Identification of error hot-spot regions defined by the inclusion of at least five markers showing an imputation error rate higher than 0.3 (based on BEAGLE v4.0 imputed genotypes) within a 2-Mb long interval.

Table S3 Linkage disequilibrium (LD) estimates in r^2 and D' units estimated based on different datasets considered in the present study to be compared with the estimates reported in the main text of this study to evaluate the effect of the imputation and the mis-imputed markers on LD estimations.

Table S4 Effective population size (N_e) estimates obtained based on the different datasets compared in this study: 50K-chip dataset (taken from García-Gámez *et al.* 2012), HD-chip genuine dataset, imputed HD-chip dataset and Imputed HD-chip after removal of error hot spot SNPs.

Appendix S1 Supplemental materials, methods and results.

Study 2

Low density chip design and genomic imputation from low density (LD-Chip) SNP chip to medium (50K-Chip) and high (HD-Chip) density in sheep

Praveen Krishna Chitneedi, Beatriz Gutiérrez-Gil and Juan-Jose Arranz.

Departamento de Producción Animal, Facultad de Veterinaria, Universidad de León, Campus de Vegazana s/n, León 24071, Spain.

XVII Jornadas sobre Producción Animal, Zaragoza, España, 30 y 31 de mayo de 2017, pp.522-524**.** https://www.cabdirect.org/cabdirect/abstract/20173312869 (17th National Conference on Animal Production**,** Zaragoza, Spain, 30th-31st of May 2017.)

Evaluación de la imputación de genotipos desde un chip de baja densidad (3K) a media (Chip-50K) y alta (Chip-HD) densidad en el ganado ovino

Chitneedi[1], P.K., A., Gutiérrez-Gil, B. y Arranz, J.J.
[1]Departamento de Producción Animal, Facultad de Veterinaria, Universidad de León, 24071 León. pchi@unileon.es

INTRODUCCIÓN

En el ganado ovino la aplicación de la selección genómica depende, en gran medida, del precio asociado al genotipado de los animales. Por ello, la posibilidad de tener un chip con un coste adaptado al valor individual de los animales podría hacer la selección genómica más atractiva para el sector ovino. Sin embargo, el descenso del precio iría encaminado a la utilización de un chip con una baja densidad de marcadores, lo que podría comprometer la precisión de la selección genómica. La imputación de genotipos podría ofrecer una solución práctica y eficiente a este problema, reduciendo el coste total asociado con el genotipado y, al mismo tiempo, mantener una fiabilidad aceptable de las estimaciones genómicas (Erbe et al., 2012). La imputación es una aproximación *in-silico* por la que para una población genotipada con un chip de baja o media densidad se infieren los genotipos faltantes de un chip de más alta densidad en base a la información de una población de referencia que ha sido genotipada con la máxima densidad de marcadores considerada. Distintos autores han probado la eficiencia de la imputación genómica, con una precisión que en ganado ovino varía entre el 71% y el 98% (Hayes et al., 2012; Ventura et al., 2016).

La aproximación de imputación más eficiente se basaría en el genotipado de alta densidad sólo para los animales con mayor contribución genética en la población, como los fundadores o los antecesores, mientras que el resto de la población se genotiparía con un chip de densidad más baja para luego inferir los genotipos faltantes (Cleveland y Hickey, 2013). El tamaño de la población de referencia, la fiabilidad de la información referente al pedigrí, la relación entre la población a imputar y la población de referencia, así como el tipo de información utilizada por el software seleccionado para el análisis, desempeñan un papel crítico para obtener genotipos imputados con alta precisión (Hozé et al., 2013). En este estudio, hemos evaluado el potencial y la precisión de un chip de baja densidad (Chip-LowDensity, Chip-LD) virtual, diseñado *in silico,* para realizar imputación de genotipos a dos niveles superiores de densidades genotípicas, densidad media (Chip-50K) y alta densidad (Chip-500K o Chip-HighDensity, Chip-HD).

MATERIAL Y MÉTODOS

Población de estudio: Este trabajo se basa en una población comercial de ganado ovino lechero con un total de 1.775 animales de del Núcleo de Selección de ANCHE. Un total de 1.686 animales (16 machos y 1.670 hijas de los mismos) han sido genotipados con el Chip-50K (37.400 SNPs, tras el control de calidad). Un subconjunto de 240 animales de esa población (los 16 machos y 14 hijas de cada familia), además de 94 de los machos con más hijas en el Núcleo de Selección se genotiparon con el Chip-HD (490.940 SNPs, tras el control de calidad).

Chip-LD: El chip de baja densidad a evaluar, Chip-LD, se diseñó mediante la selección de un subconjunto de los marcadores comunes entre los chips ovinos de media (Illumina Ovine SNP50 BeadChip) y de alta densidad (Ovine Infinium® HD SNP BeadChip). Los criterios de selección fueron la localización de los marcadores a lo largo de los 26 autosomas ovinos y su grado de informatividad medida mediante la frecuencia para el alelo menos frecuente (MAF > 0,3). La distancia media entre los marcadores seleccionados para el Chip-LD fue de aproximadamente de 1 Mb y el MAF medio aproximadamente 0,40.

Imputación de genotipos: El software utilizado para la imputación ha sido Beagle_v4.0 (Browning et al., 2008). Los datos de entrada para este análisis incluyeron los datos de genotipado del Chip-LD, los datos del pedigrí de la población y los datos de genotipado del chip de densidad superior en cada caso, el Chip-50K o el Chip-HD. La evaluación de la precisión de imputación llevada a cabo se ha basado en una estrategia de validación cruzada en la que se utilizaron dos porcentajes de enmascaramiento para los animales de la población de referencia, del 10% (10 iteraciones) y 30% (4 iteraciones).

RESULTADOS Y DISCUSIÓN

En base a los criterios establecidos para el diseño del Chip-LD *in silico*, un total de 2,935 fueron incluidos en el panel de baja densidad a evaluar en el presente trabajo. El rendimiento del software utilizado fue apropiado, con un rango de tiempo de ejecución de 3-30 minutos (para los cromosomas OAR26 y OAR1, respectivamente, y para la imputación de Chip-LD a Chip-50K y Chip-LD a Chip-HD).

Considerando los 26 cromosomas ovinos, el promedio de la precisión de imputación de los genotipos del Chip-50K a partir del Chip-LD para resultó ser muy similar para los dos niveles considerados de enmascaramiento, el 10% y el 30%, con valores del 93,53% (±0,0086) y 93,58% (±0,0081), respectivamente. Para la imputación del Chip-LD al Chip-HD, la precisión de la imputación fue del 88,3% (± 0,021) para el nivel de enmascaramiento del 10% de la población de referencia y del 86,52% (± 0,012) cuando se enmascaró el 30% de la población. La Figura 1 muestra el promedio, para las iteraciones realizadas, de la tasa de error en la imputación para cada uno de los marcadores imputados en los cuatro escenarios considerados. La menor precisión de imputación observada cuando la imputación de genotipos se hizo del Chip-LD al Chip-HD, comparando con la imputación al Chip-50K (aprox. 5-7% más baja), se podría atribuir a que el porcentaje de SNPs que tienen que ser imputados es mayor que cuando la imputación se hace en dos pasos y también al menor tamaño de la población de referencia (335 *vs* 1.680 animales) disponible para la imputación a la densidad superior. En ganado vacuno, otros autores han sugerido que, en el caso de la imputación desde la baja a la alta densidad, la mejor estrategia sería la imputación en dos pasos, inicialmente a densidad media y posteriormente a la mayor densidad (Berry et.al., 2014). En cualquier caso, una población de referencia de mayor tamaño aumentaría considerablemente la precisión de la imputación (Hozé et al., 2013).

Comparando los resultados para los dos niveles de enmascaramiento considerados en la población de referencia, del 10% y 30%, vemos que este factor no tiene influencia para la imputación del Chip-LD al Chip-50K, mientras que en la imputación del Chip-LD al Chip-HD, la tasa de error de imputación fue superior en un 2% cuando el enmascaramiento afectó al 30% de la población de referencia. Esto se podría atribuir al menor número de iteraciones en los que se basa la estimación para este nivel de enmascaramiento, y/o al mayor número de genotipos a inferir para el mismo tamaño de población de referencia.

Con respecto a la distribución de la tasa de errores de imputación a lo largo de los cromosomas, a pesar de ser bastante homogénea para todos los escenarios, se observó que los marcadores localizados en los extremos de cada cromosoma muestran una tasa de error considerablemente más alta que la media del genoma (ver Figura 1). La identificación de esta condensación de errores en los extremos cromosómicos es relevante y se podría resolver mediante la redistribución de los marcadores en el diseño del Chip-LD, tratando de aumentar la densidad de marcadores en los extremos de los cromosomas. Hay que señalar además que, para la imputación a la densidad más alta, la tasa de errores fue excepcionalmente alta en el extremo proximal del cromosoma 25, llegando casi al 60%. Esto puede ser debido a la inadecuada asignación de SNPs en el Chip-HD, o a la mala calidad del borrador del genoma ovino de referencia en esa región específica.

Como principal conclusión del estudio podemos destacar que la imputación es una potente herramienta para inferir genotipos en poblaciones comerciales de ganado ovino, ofreciendo una tasa de errores relativamente baja, lo que es muy ventajoso en el ahorro de costes de genotipado. El coste de esta estrategia podría ser equivalente en costes al control de paternidad con marcadores microsatélites y, para aquellas poblaciones que utilicen la selección genómica, podría obtenerse con el coste de genotipado de un chip de baja densidad una estimación inicial del valor genético genómico que podría ser utilizado para decidir con qué animales se realiza la reposición.

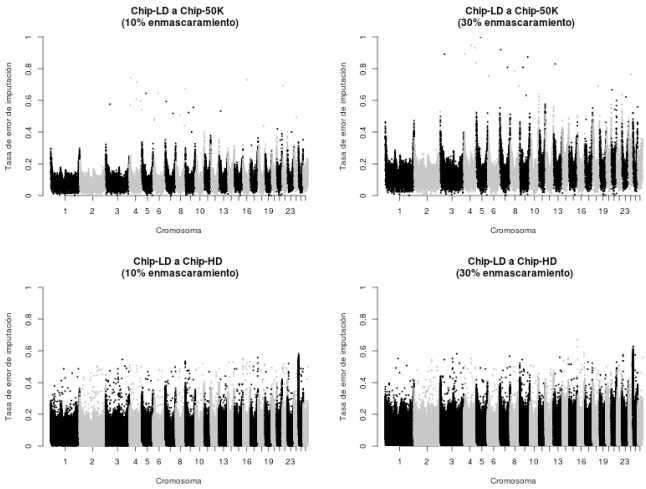


Figura 1. *Distribución de la tasa de errores de imputación a lo largo de los 26 cromosomas ovinos en los cuatro escenarios considerados: (i) Imputación del Chip-LD al Chip-50K con el 10% (a) y el 30% (b) de enmascaramiento; (ii) Imputación del Chip-50K al Chip-HD con el 10% (c) y el 30% (d) de animales enmascarados.*

REFERENCIAS BIBLIOGRÁFICAS

• Berry, D.P. et al. 2014. J. Animal. Breed Genet. 131: 165-172. • Browning B.L. & Browning S.R. 2009. Am. J. Hum. Genet. 84: 210-223. • Cleveland, M.A. & Hickey, J.M. 2013. J. Animal Sci. 91: 3583-3592. • Erbe, M., et al. 2012. J. Dairy Sci. 95: 4114-4129. • Hayes, B. et al. 2012. Animal Gen. 43: 72-80. • Hozé, C. et al. 2013. Genetics Sel. Evol. 45: 33 • Purcell, S. et al. 2007. Am. J. Hum. Genet. 81: 559-575. • Ventura, R.V. et al. 2016. Genet. Sel. Evol. 48: 71.

Agradecimientos: Trabajo financiado por el proyecto AGL2012-34437 del Ministerio de Economía y Competitividad España (MINECO). P.K. Chitneedi es beneficiario de una beca predoctoral de la Junta de Castilla y León y el Fondo Social Europeo. B. Gutiérrez-Gil es investigadora contratada del programa "Ramón y Cajal" del MINECO (RYC-2012-10230).

LOW DENSITY CHIP DESIGN AND GENOMIC IMPUTATION FROM LOW DENSITY (LD-Chip) SNP CHIP TO MEDIUM (50K-Chip) AND HIGH (HD-Chip) DENSITY IN SHEEP

ABSTRACT: In this study we designed an *in silico* low density chip (LD-Chip) for a half-sib commercial population of the Selection Nucleus of Churra sheep and investigated the potential of this genomic tool to impute genotypes to medium (50K-Chip) and high (HD-Chip) marker densities. Based on the available genotypes for the ovine Chip-50K (n = 1,686) and HD-Chip (335 individuals) a total of 2,935 markers common to the two panels and evenly distributed across the sheep genome were included in the *in silico* Chip-LD. The average error rate for the LD-Chip to 50K-Chip imputation was around 93.5% and ranged from 88,3% to 86,52% for the LD-Chip to HD-Chip imputation (depending on the masking level considered, 10% and 30%, respectively). For all the scenarios, a higher imputation error rate than the average was found at both ends of all the chromosomes. For the imputation to the HD-Chip density, a very high error rate was identified at the proximal end of chromosome 25. Globally, and although the identified limitations need to be solved, the *in silico* LD-Chip described here, together with the implemented imputation strategy, appear as an appropriate and affordable approach to obtain, in addition to the basic genetic tests (e.g. paternity), preliminary estimations of genomic breeding values that could guide breeding decisions.

Keywords*:* Churra sheep, SNP-Chip, imputation, accuracy.

Study 3

GWAS analysis for gastrointestinal nematodes resistance traits using imputed high density chip genotypes in sheep

Praveen Krishna Chitneedi, Marina Atlija, Juan-Jose Arranz and Beatriz Gutiérrez-Gil.

Departamento de Producción Animal, Facultad de Veterinaria, Universidad de León, Campus de Vegazana s/n, León 24071, Spain.

XVI Jornadas sobre Producción Animal, 19 y 20 de mayo de 2015, Zaragoza, España. pp.471-473. https://www.cabdirect.org/cabdirect/abstract/20163059168 (16th National Conference on Animal Production**,** Zaragoza, Spain, 19th-20th of May 2015.)

ANÁLISIS GWAS PARA RESISTENCIA A NEMATODOS GASTROINTESTINALES MEDIANTE LA IMPUTATIÓN DE GENOTIPOS DEL CHIP DE ALTA DENSIDAD EN EL GANADO OVINO

Chitneedi, P.K., Atlija, M., Arranz, J.J. y Gutiérrez-Gil, B.
Departamento de Producción Animal, Facultad de Veterinaria, Universidad de León, 24071 León. Email: pchi@unileon.es

INTRODUCCIÓN

Las infecciones por nematodos gastrointestinales (GIN) constituyen un grave problema para las explotaciones ovinas basadas en sistemas extensivos o semi-extensivos, tanto a nivel sanitario y de bienestar animal como a nivel de salud pública (posible riesgo de zoonosis y de residuos de fármacos antiparasitarios). Existe, por tanto, la necesidad de identificar métodos alternativos eficaces frente al método clásico de control basado exclusivamente en el uso de fármacos antiparasitarios. A raíz de diversos estudios que pusieron de manifiesto el componente genético de parte de la variación observada en la resistencia de las ovejas a los parásitos internos (revisado por Raadsma et al., 1997), distintos autores han propuesto la selección genética, combinada con otros métodos, como una estrategia eficaz de control de los GIN. En este contexto, el objetivo de este estudio es la identificación de regiones genómicas asociadas a Resistencia a GIN en una población comercial de Ganado ovino de raza Churra mediante un barrido genómico de alta densidad basado en un protocolo de imputación de los genotipos del *Illumina Ovine HD BeadChip* (HD-chip) en una población inicialmente genotipada con el *Illumina Ovine SNP50 BeadChip* (50K-chip), en base a un subconjunto de animales genotipados con el HD-chip utilizados como referencia. Tras la evaluación de la fiabilidad del proceso de imputación, y en base a los genotipos imputados se ha realizado un estudio de asociación a nivel genómico (GWAS, *Genome-wide Association Study*) con un carácter relacionado con los niveles de infección por GIN, los niveles séricos de Inmunoglobulina A (IgA) frente a *Teladorsagia circumcinta (L-IV)*.

MATERIAL Y MÉTODOS

La población objeto de estudio es una población comercial de ganado ovino lechero de raza Churra que presenta la estructura de una población de 16 familias de medio-hermanas, incluyendo los 16 machos cabeza de pedigrí y 1.670 hijas, todas ellas pertenecientes a rebaños del Núcleo de Selección de Anche (ANCHE). Esta población base ha sido previamente genotipada con el chip de 50K SNPs. Tras el control de calidad realizado según García-Gámez et al. (2012) quedaron disponibles para análisis un total de 43.613 SNPs localizados en los cromosomas autosómicos.

La población de referencia utilizada para la imputación incluyó 240 animales de la población base, con 16 machos y 14 hijas de cada uno, que fueron genotipados con el HD-chip. Inicialmente se obtuvieron genotipos para un total de 606.006 marcadores SNPs, que tras el control de calidad se redujeron a 492.767 SNPs autosómicos, entre los cuales se incluyen la mayoría de los marcadores del 50K-chip. El proceso de imputación de los genotipos del HD-chip ausentes en los animales de la población genotipada con el 50K-chip se realizó con el software *Beagle* 3.3.2 (Browning & Browning et al., 2008). Los genotipos resultantes fueron convertidos a formato *Plink* (Purcell et al., 2007) con el programa fcGENE (Roshyara, et al., 2014). La imputación realizada con *Beagle* se basa en información poblacional (desequilibrio de ligamiento y frecuencias alélicas), obviándose en este caso la información proporcionada por el pedigrí de la población. Es por ello que en algunos casos, se pueden estimar genotipos incorrectamente o no realizar la imputación. Por ello, se estimó la fiabilidad de la imputación en base a la comparación de los genotipos imputados en las hijas de la población de referencia con sus genotipos directamente obtenidos con el HD-chip (genotipos imputados *vs* genotipos genuinos). Con el fin de evitar sesgos en esta evaluación, se realizaron 14 iteraciones de imputación, enmascarando en cada una de ellas el genotipo de 16 hijas, una de cada familia analizada.

Para un subconjunto de 533 animales de esta población se obtuvieron muestras de sangre para la obtención de medidas fenotípicas de IgA en suero sanguíneo, siendo los genotipos de estos animales los que se utilizaron para el análisis de asociación realizado posteriormente. Tras la normalización de los valores de IgA mediante la transformación box-cox, se estimaron las *Yield Deviations* (datos brutos corregidos para el efecto “rebaño”). El

análisis GWAS se realizó con *ProbABLE* (Aulchenko et al., 2010), un software específicamente desarrollado para realizar análisis GWA a partir de genotipos imputados. Siguiendo el protocolo sugerido por los desarrolladores de este programa, se realizó un primer análisis en *GenABLE* (Aulchenko et al., 2007) con el fin de estimar la matriz de varianzas y convarianzas, lo que proporciona un test de asociación entre el fenotipo a estudiar y los marcadores genéticos en las muestras de individuos relacionados. El segundo paso del análisis de asociación se realizó con *ProbABLE*, utilizando como datos genotípicos para el análisis, los archivos de probabilidad de imputación correcta para el SNP obtenidos en el proceso de imputación de *Beagle*. Los umbrales de significación análisis GWA se determinaron aplicando la corrección de Bonferroni, considerando el número de marcadores analizados para cada cromosoma (umbral 5% *chromowome-wise*) y para todo el genoma (umbral 5% *genome-wise*). Para calcular el número de marcadores "independientes" en cada caso, teniendo en cuenta el desequilibrio de ligamiento existente, se utilizó el programa *SimpleM* (Gao et al., 2008), asumiendo un valor de 0,995 para el parámetro PCA. Todos los análisis, imputación y GWA, se realizaron en un potente servidor Intel (procesador de 2.50GHz, 8 cores y 16GB RAM) en el CPU del Centro de Supercomputación de Castilla y León.

RESULTADOS Y DISCUSIÓN

La precisión de la imputación realizada por *Beagle* se estimó usando la opción de concordancia de *Plink* (merge-mode 7) en los individuos genotipados para el 50K-chip y el HD-chip. Debido al alto tiempo de computación exigido en el proceso de imputación, la estimación de la fiabilidad se realizó en base a los resultados obtenidos en seis cromosomas de distinta longitud (cromosomas 1, 4, 10, 16, 20, 26). La concordancia promedio estimada a partir de las 14 iteraciones de imputación realizadas fue 90,85%, siendo muy similar entre los seis cromosomas analizados (rango: 90-92 %, SD promedio: ± 1,55%). Ha de tenerse en cuenta que esta estimación está afectada por la reducción de la potencia en el proceso de imputación debido al enmascaramiento de los animales utilizados como referencia en cada iteración. Así, cuando el proceso se realizó sin eliminar ningún animal de la población de referencia, la precisión de la imputación fue del 94%. Por ello, podemos considerar que la fiabilidad real proporcionada por Beagle en la población de Churra considerada varía entre 91-94%, suficientemente alta como para utilizar los genotipos imputados para posteriores análisis de asociación con caracteres de interés productivo. La precisión estimada en este estudio está dentro del rango de precisión de otros estudios realizados en oveja (83-93%, según la raza) estimado por (Hayes et al., 2012) utilizando *Beagle*. En el ganado vacuno, por lo general, los niveles de precisión obtenidos con este programa es más alta. Hozé et al. (2013) estimaron una precisión del 99% tras los análisis realizados en 16 razas de ganado vacuno francés. En ganado vacuno Holstein Chino (Weng et al., 2013) las estimaciones variaron entre el 90-98%, dependiendo de la proporción de individuos de referencia eliminados en las iteraciones.

El análisis de asociación con el fenotipo IgA realizado con *ProbABLE* identificó un único SNP significativo al nivel 5% *chromosome-wise* localizado en el cromosoma 15 (p_c-value corregido = 5,63), mientras que no se detectó ninguna asociación significativa al nivel 5% *genome-wise*. El marcador que mostró evidencia de influir los niveles de IgA es el oar3_OAR15_24870525, y se encuentra localizado en la posición 24.870.525 pb del OAR15. El efecto de substitución alélica estimado para este SNP fue de 0,297 ± 0,063 unidades de las YDs analizadas (0,382 desviaciones estándar del fenotipo). De acuerdo a la base de datos *Sheep QTLdb* (http://www.animalgenome.org/cgi-bin/QTLdb/OA/index), la posición de este QTL se encuentra dentro del intervalo de confianza estimado (18,3-30,5 Mb) para un QTL con influencia sobre Fecal egg count (FEC), el clásico carácter indicador de los niveles de infección por GINs en una población de retrocruzamiento Red Maasai x Dorper (Silva et al., 2012). La falta de resultados significativos se puede explicar por la escasa potencia del estudio debido al limitado número de animales analizados. Para explotar la información de la estructura de diseño hija de la población en estudio, de forma complementaria a este análisis tipo GWA, se realizarán futuros análisis basados en análisis de ligamiento (LA) y la combinación de desequilibrio de ligamiento con análisis de ligamiento (LDLA).

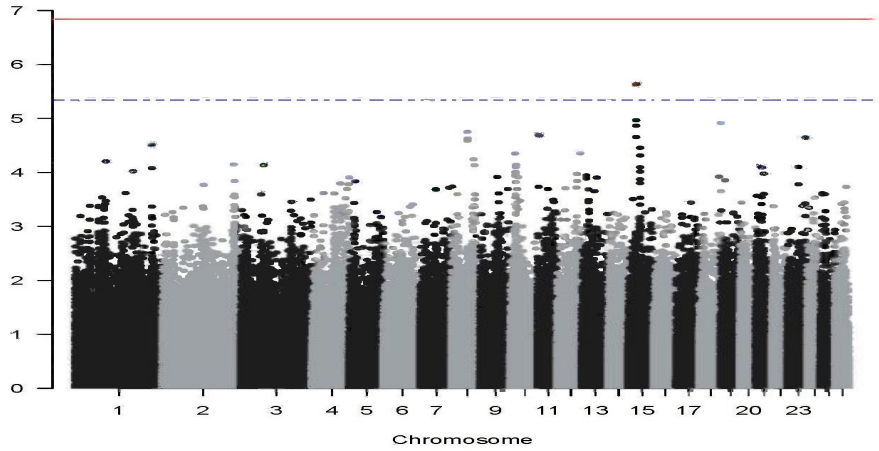


***Figura 1**. Resultado del análisis GWAS basado en los genotipos del HD-chip (700K) imputados con Beagle en la población comercial de ganado ovino analizada.* Para el carácter analizado, IgA, se representan los valores *log(1/P). Las líneas horizontales representan el umbral del 5% chromosome-wise promedio para los 26 autosomas (azul), y el umbral del 5% genome-wise (rojo);* obtenidos tras la corrección de Bonferroni aplicada *(Gao et al., 2011).*

REFERENCIAS BIBLIOGRÁFICAS

• Atlija et al. (2015). Anim. Genet. submitted • Aulchenko et al. 2007. Bioinformatics 23, 1294-1296. • Aulchenko et al. 2010. Bioinformatics 11,134. • Gao et al., 2011. Genet. Epidemiol. 35, 154-158. • Hayes et al. 2012. Anim. Genet. 43, 72-80. • Hozé et al. 2013. Genet. Sel. Evol. 45, 33. • Purcell et al. 2007. Am. J. Hum. Genet. 81, 559-575. • Raadsma et al. 1997. The genetics of sheep. 199-224. (ed. L Piper and A Ruvinsky) • Roshyara et al. 2014. BMC Genet. 15, 88 • Browning & Browning, 2009. Am. J. Hum. Genet. 84, 210-223. • Silva et al. 2012 Anim. Genet. 43, 63-71. CABI Publishing. • Weng et al. 2013. Animal 7, 729-735.

Agradecimientos: Este trabajo ha sido financiado por el proyecto Europeo *NematodeSystemHealth* del programa Marie Curie-ITN (Ref. FP7-PEOPLE-2010-ITN #264639) y el proyecto LE245A12-2 financiado por la Junta de Castilla y León. B. Gutierrez-Gil es investigadora contratada a través del programa "Ramón y Cajal" del Ministerio de Economía y Competitividad.

GWAS ANALYSIS FOR GASTROINTESTINAL NEMATODES RESISTANCE TRAITS USING IMPUTED HIGH DENSITY CHIP GENOTYPES IN SHEEP

ABSTRACT: The aim of this study was to identify genomic regions influencing the serum levels of Immunoglobulin A (IgA), an indicator trait of resistance to gastrointestinal nematode (GIN) infections in Spanish Churra dairy sheep. With this aim, we performed a Genome-wide Association Study (GWAS) based on imputed genotypes for the Ovine High Density (HD)-chip for a population of 1,686 animals belonging to 16 different half-sib families. The whole population, including the 16 sires and their daughters had available genotypes for the *Ovine 50K-chip*. The imputation process was based on the 240 animals of the resource population that had been genotyped for the HD-chip. We estimated the accuracy of imputation, using only the genotypes available for the reference population based on a masking strategy designed to avoid any bias in the accuracy estimation. The GWAS identified a significant SNP at the 5% chromosome-wise level, located on OAR15. As complementary approaches, future analyses will exploit the linkage analysis and combined linkage disequilibrium with linkage analysis to exploit the half-sib structure of the studied resource population.

Keywords*:* sheep, parasite resistance, imputation, GWAS

Study 4

Detection and high-resolution study of QTLs underlying resistance to gastrointestinal nematode infection in adult sheep by using imputed HD SNP-chip genotypes and whole-genome sequencing

Praveen Krishna Chitneedi, Beatriz Gutiérrez-Gil, Cristina Esteban-Blanco, Marina Atlija and Juan-Jose Arranz.

Departamento de Producción Animal, Facultad de Veterinaria, Universidad de León, Campus de Vegazana s/n, León 24071, Spain.

Manuscript under preparation

Detection and high-resolution study of QTLs underlying resistance to gastrointestinal nematode infection in adult sheep by using imputed HD SNP-chip genotypes and whole-genome sequencing

Chitneedi, P.K., Gutiérrez-Gil, B., Esteban-Blanco, Atlija, M., and Arranz, J.J.
Departamento de Producción Animal, Facultad de Veterinaria, Universidad de León, 24071
León. Email: pchi@unileon.es

Abstract

Gastrointestinal nematode (GIN) infections are one of the major health issues for grazing sheep populations. Because ovine resistance to the GIN infection is a highly complex trait, the identification of genes influencing this trait could enhance the efficiency of classical selection towards population with increased GIN resistance. Our group has previously performed a search of quantitative trait loci (QTL) for GIN resistance in a commercial population of 1696 Churra adult ewes based on a medium-density SNP-array (50K-chip) and records of two GIN resistance indicator traits (Atlija et al., 2016). In this study, the imputed 50K-chip to high density (HD) chip (600K) genotypes generated for that population (Chitneedi et al., 2017) have been used to refine the confidence intervals (CIs) of the previously identified QTL by performing a classical linkage analysis (LA). In addition, the new refined CI of the most promising QTL region previously reported on chromosome 6 (OAR6) for feacal egg count (FEC) has been studied at high-resolution by performing a detailed analysis of whole genome sequencing (WGSeq) datasets of a segregating trio for that QTL (which includes the *Qq* sire and 2 daughters with extremely divergent phenotypes for the FEC trait in concordance with their QTL inferred genotypes, QQ and qq). Considering the refined CI of the significant QTL The LA HD genome scan performed in the studied population for the FEC and IgA traits positional candidate genes related to the immune response were identified as potential functional candidate genes. In addition, the analysis of the WGSeq segregating trio dataset identified a total 433 high quality variants within the OAR6 CI QTL region (88.2-88.3 Mb), 357 of which were intragenic concordant variants. Based on a variant functional annotation analysis, this study provides a list of variants that could be considered as potentially underlying GIN resistance in sheep. Before translating these results into practical breeding programs, future research efforts should confirm the role of the candidate genes and mutations highlighted by this study in relation to GIN resistance through their analysis in commercial populations with available phenotypes for GIN resistance indicator traits.

Introduction

Sheep resistance to GIN is a largely complex phenotype. The practical implementation of breeding programs addressing the enhancement of GIN resistance is hampered by the cost and difficulties to routinely measure indicator traits, such as the faecal egg count (FEC) or serum levels of immunoglobulin A (IgA). Because the use of genetic markers associated to these indicator traits could be used to increase the efficiency of such breeding schemes, several research projects have been undertaken with the aim

of identifying genomic regions associated to GIN resistance across different sheep populations with a total of 168 QTL and 28 significant associations for parasite resistance traits being currently included in the SheepQTLdb (Hu et al., 2013). Studies in dairy cattle have shown that genomic predictions can be improved by including information of sequence variants that are casual or close to causal mutations according to previous QTL or genome-wide association (GWA) studies (Brøndum et al., 2015; Veerkamp et al., 2016). This supports the value of studies aiming the identification of potential causal mutations related to GIN resistance in sheep populations.

Here, we present a follow-up study of the QTL identification for indicator traits of GIN resistance in a population of Spanish Churra sheep adult ewes presented by Atlija et al., (2016) based on a medium-density SNP Chip (50K-chip). First, a higher density genome scan for QTL detection has been performed for the same population and phenotypes using the imputed genotypes for a high-density Chip (HD-chip, with more than 600.000 SNP markers) based on a previous work of our research group (Chitneedi et al., 2017). Then, the confidence interval (CI) refined here for the most significant QTL region detected by Atlija et al., (2016), located on OAR6 for the FEC trait, was subjected to a further high-density analysis by exploiting whole genome sequencing of a segregating trio. The potential role of the list of mutations identified here as potentially associated to GIN resistance should be later evaluated through their genotyping in commercial sheep populations with available measurements for indicator traits.

Materials and methods

Animals with phenotypic data for GIN resistance indicator traits

The population of animals and phenotypes studied were previously analyzed by Atlija et al., (2016). Briefly, the population under study included 518 Spanish Churra ewes distributed in 14 half-sib families, and the corresponding 14 sires (532 animals). The available indicator traits of GIN resistance were single measurements obtained for the analyzed ewes for FEC and serum IgA levels. For the analyses presented here, we have used the logarithmic transformation of FEC (LFEC) and the box-cox transformation of the IgA levels (IgA_t) already considered by Atlija et al., (2016).

Genomic Imputation from 50K-chip to HD-chip

All the animals with phenotypic measurements for GIN resistance traits considered in this study belonged to a larger population of 1,686 individuals which had been genotyped with the medium density chip, Illumina OvineSNP50 BeadChip (50K-chip). This population had also available imputed genotypes for the Ovine High Density Bead Chip (HD-chip) (Kijas et al., 2014) derived from Chitneedi et al., (2017). Considering the imputation probabilities provided by BEAGLE_v4.0 (Browning and Browning, 2013) for the imputed genotypes (490,940 markers for the 1686 animals), we filtered those genotypes showing an imputation probability higher than 0.85, using the plink *--vcf-min-gp 0.85* option. Finally, from the total population of 1,686, we extracted the imputed genotypes corresponding to the 532 animals for which LFEC and IgA_t records were available.

QTL mapping LA analyses

Using the QTLMap_v9.2 software (Filangi et al., 2010), a linkage analysis (LA, *calcul option 4*) across the 26 ovine autosomal chromosomes was performed for the 532

animals under study for GIN traits by using their HD-chip imputed genotypes and considering a mapping step interval of 0.01 cM. The phenotypes considered for QTL mapping were the Yield deviations (YD) for transformed indicator traits, LFEC and IgA_t, which had been previous estimated following a multivariate animal model by Atlija et al. (2016). Using QTLMap, we calculated the QTL mapping significance thresholds at the chromosome-wise significance level through a total of 1000 permutations (at 0.01 cM steps). Genome-wise significance thresholds were based on the chromosome-wise significance threshold by correcting for the total number of chromosomes under analysis. For the significant QTL detected, the peak likelihood ratio test (LRT) values were converted to logarithm odds ratio (LOD) values (Beraldi et al., 2007) and confidence intervals (CI) for the QTL locations were estimated by the widely used 1-LOD drop-off method (Lander and Botstein, 1989). The proportion of phenotypic variance that was explained by the significant QTL detected by LA was calculated based on the corresponding LOD values using the formula (Doerge, 2010)

$$\sigma_p = 1 - 10^{-\frac{2}{n}LOD}.$$

For the significant QTL identified in the across-family analysis, within-family LA analyses were performed with the imputed HD-chip genotypes to identify the corresponding QTL segregating families.

Concordance with other QTL studies and identification of functional candidate genes within the QTL CIs

The possible concordance of the significant QTL detected here with previously associated results in the literature was assessed by extracting from SheepQTLdb (Hu et al., 2013) all the GIN-related QTL and associations included in the genomic intervals corresponding to the refined CI QTL regions. Correspondence was checked considering the CI interval annotated for QTL regions, and a a 250 Kb interval centred on the originally significant SNP marker for the single point significant associations annotated in the database. The positional candidate genes overlapping the confidence intervals of the identified QTL in the current study were also extracted by using a web-based tool BioMart based on the Ensembl release 94 base using sheep reference genome assembly Oar v3.1 (http://www.ensembl.org/biomart/martview/). The discovered positional candidate genes were assessed as putative functional candidate genes in relation to immune response by using a list of 5,029 genes known to be related to the immune response, based on the public databases IRIS (Kelley et al., 2005) and ImmPort (Bhattacharya et al., 2014), both available at the InnateDB (Breuer et al., 2013).

Selection of segregating trios for further study the OAR6 QTL

The OAR6 QTL region was selected for further study as that was the most significant QTL effect identified by Atlija et al. (2016), which at the same time replicated the most significant QTL for FEC reported by Gutierrez-Gil et al (2009) in a different population of Churra sheep based on a microsatellite-based genome scan. The family to be subjected to WGSeq analysis, family 7, was selected based on the significant segregating familiess identified by within-family LA analysis performed with the HD-chip imputed genotypes. Although Atlija et al. (2016) had identified two segregating families for the OAR6 QTL, only one significant family was identified here based on the newly estimated significance thresholds. The selection of the daughters to be included in the sequenced trio was based on marker phases reconstruction for the across-family CI interval using the

out_phases and *out_phases_offspring* options of QTLMap (Le Roy et al., 2013). To assign the QTL alleles, Q and q, to the corresponding paternal phase at the QTL region, the *out_pded* option of this software was used to estimate the transmission marginal probabilities for all of the animals of the selected family for the across-family QTL peak position. By combining the estimated QTL effect of each sire and the inheritance probabilities of daughter phases with extreme divergence (close or equal to zero or one), we characterized the resistance and susceptible associated haplotypes. The sire haplotype associated to ewes with higher positive LFEC yield deviation values was denoted as *q* allele (increased susceptibility to nematode infection) whereas the sire haplotype linked to negative LFEC yield deviation values was denoted as *Q* (increased resistance to nematode infection). Based on this criterion for each daughter, we identified the daughters inheriting either the *Q* or *q* sire alleles at the target CI QTL region. By ranking the daughters based on LFEC yield deviation values, we selected two homozygous daughters with extreme LFEC phenotypes consistent with the QTL effect. Hence, we identified, for the considered family *QQ* daughter showing an extreme low LFEC value and a *qq* daughter showing an extreme high LFEC within the family phenotypic yield deviation values.

Whole-genome sequencing and variant annotation

The three DNA samples of the trio selected were subjected to WGSeq by using the paired-end Illumina technology in an Illumina HiSeq 2000 sequencer. For the sequencing data generated the variant identification analysis was performed by following the bioinformatic workflow implemented by Gutiérrez-Gil et al. (2017), based on the use of two variant calling software, GATK (REF) and Samtools (REF). The list of variants commonly identified by these two prediction software was extracted using BCFtools utilities (Narasimhan et al., 2016) to produce a final VCF file of 'high quality' variants (SNP and indels). These variants were filtered to focus on the targeted QTL region (considering both the across-family and the within-family estimated CIs) and for concordance with the QTL segregation pattern (heterozygous for the sire and alternative homozygous for the daughters: 01/00/11 or 01/11/00).

After these filtering steps of the high quality variants, we performed a functional annotation analysis with the Variant Effect Predictor software (McLaren et al., 2016), based on the Ovis_aries_v3.1 Ensembl build 94 to identy the variants with a potential functional impact (stop codon, missense, etc.). To assess the potential biological consequences of these variants, the SIFT tool of VEP was used to classify the variants as 'tolerated or deleterious'. By comparing the list of genes harboring the variants with a potential biological impact with a list of 5,029 immune related genes of IRIS (Kelley et al., 2005) and ImmPort (Bhattacharya et al., 2014) databases which are available at InnateDB (Breuer et al., 2013).

Results

Genome scan with imputed HD-chip genotypes

After the quality control of the raw HD-chip genotypes obtained for a subset of 240 animals and a set of 96 additional IA sires from animals from the global studied population, genotypes for a total of 490,940 SNPs and 335 animals were available as reference dataset to perform the 50K-chip to HD-chip imputation of genotypes for all the population, as described in detail in Chitneedi et al. (2017). As result of the

imputation process, genotypes for a total of 490,791 SNPs across the 26 sheep autosomes and with an imputation probability higher than 0.85 were used for LA analysis**.** The across family LA identified one significant QTL for $IgA_{t,}$ located on OAR22 (peak at 3.59 cM), and two QTL for LFEC, on OAR6 (peak at 88.21 cM) and OAR8 (peak at 2.15 cM), as summarized in Table 1 and graphically presented in Figure 1. The obtained likelihood ratio test (LRT) values and also the location of these QTLs were similar to those previously reported with the 50K-chip by Atlija et al. (2019). Interestingly, for the three QTL the length of the CI estimated based on the new HD-scan was shorter than in the previous study, showing a refinement of around 99% and 87.2% for the OAR6 and OAR8 QTL influencing FEC, and a refinement of around 20% for the OAR22 QTL influencing IgA_t (see Table 1 for details). The results of the within-family analysed performed for each of the significant QTL detected in the across-family analysis identified a single significant family for each of the QTL regions (see Table 1 for details). The within-family analysis peaks were close to earlier detected peaks with the 50K-chip dataset. The proportion of phenotypic variance explained by the detected was 0.122 % and 0.127 % for the OAR6 and OAR8 FEC QTL, respectively, and 0.10 % for the IgA_t QTL on OAR22.

Correspondence with previously reported QTL and identification of functional candidates

The list of QTLs/associated previously reported that showed overlapped with the refined estimated CI intervals of the three significant QTL is shown in Table 2. Four QTL/associations included in the SheepQTLdb for the trait type 'parasite resistance' overlapped with QTL regions identified in the current study. On the other hand, 10 positional candidate genes were extracted from sheep reference genome for the three refined CI regions; two of these genes, *RASSF6* and *COL12A1,* which were included in the OAR6 and OAR8, respectively, were identified as immune-related genes.

Selection of trio animals based on IBD and phenotypic values

For the single segregating family identified for the OAR6 FEC QTL, Family 7 (Sire_2976), and after haplotype reconstruction of the markers at the within-family QTL peak (97.1-97.6 cM), we classified the daughters of the segregating sire in two groups according the inherited QTL allele, Q or q. Then, considering also the phenotypic information (LFEC yield deviation values) we selected two alternative homozygous ewes showing extreme phenotypic values in concordance with inherited sire QTL allele. Hence, considering the positive/negative sign of QTL effect estimated for Sire_2967 (family 7), the 77 daughters that belonged to this family were distributed as 37 daughters harboring the *q* type allele-haplotype (IBD value close to zero) and 40 daughters harboring the *Q* allele-haplotype (IBD value close to one). Within the *Q* daughters the LFEC yield deviation ranged from -40,563 to 26,612 with an average standard deviation of -7377.63. Out of the 77 daughters, 9 daughters were homozygous for the *QQ* haplotype showing the longest homozygous phase covering the within-family CI region (97.1-97.6 cM) and one daughter (5088) was homozygous for the *qq* haplotype. Hence, daughters 5759 (*QQ*) and 5444 (*qq*), located at ranking positions 23/77 (YD LFEC = -13,262) and 72/77 (YD LFEC = 22,642) according to increasing LFEC yield deviation values, were selected to be subjected to WGSeq together with Sire_2976.

Variants detected within the confidence intervals of across and within family 7 LA

Considering the three sequence DNA samples, the average number of raw reads per sample were 266,247,042. After trimming the reads, around 93% of reads passed the subsequent quality control. From these, 92% of the reads were mapped against the sheep reference genome Oar_v3.1. After genomic region filtering, considering the across-family estimated CI of the OAR6 QTL (88.2- 88.3 cM) a total of 433 high quality variants (430 SNPs and 3 indels) of which 357 were intragenic variants. Most of these variants, 410 were concordant with the QTL segregation pattern. Among the QTL concordant intragenic coding variants we found seven missense and 13 synonymous mutations. The list of the 7 missense variants is shown in Table 3. Four missense mutations were in the *AFM* gene region and a missense variant was located in the immune-related *RASSF6* gene. The list only includes a missense deleterious variant located at 6:88252392-88252392 (rs420795500), within an non-annotated gene (ENSOARG00000014280).

In addition, we also studied the genetic variants identified within the CI estimated in the within-family analysis of Family 7. In this interval, 97.1-97.6 cM, a total of 2,650 high quality variants were identified (2597 SNPs and 53 indels), 1,436 of which were intragenic variants. The variant annotation of the 2,078 QTL concordant variants identified in the within-family CI region showed13 synonymous variants in the *SEC31A* and *LIN54* genes but none missense or deleterious mutations.

Discussion

This study has proven that the mapping accuracy of QTL identification analyses can be increased by using imputed HD-chip genotypes, leading to an important reduction of the estimated QTL CIs. It should be noted that after the classical detection of QTL by means of a genome scans with sparse markers like microsatellites, it was necessary to carry out fine mapping studies, normally focused on the QTL identified with greater statistical support to pin point the causal mutation. In relation to GIN resistance, this approach has not achieved the identification of the detection of causal mutations of QTL for previously identified resistance to GIN in sheep (Moreno et al., 2014; Atlija et al., 2016). Currently, the availability of next generation sequencing technologies like WGSeq and targeted re-sequencing technologies have been also used to identify and further analyze QTL regions responsible for resistance to nematode infection in sheep (Periasamy et al., 2014; Al Kalaldeh et al., 2019; Estrada-Reyes et al., 2019a; b).

In the current study, by performing LA using an imputed HD-chip genotype dataset, we increased the mapping accuracy for three QTL regions previously identified for indicator traits of resistance to GIN infection in adult sheep and significantly refined their estimated CIs (from 20 to 99% of reduction of length for the QTL CI). As expected when analyzing the same population than that previously studied, we have identified the same QTL that those reported by Atlija et al. (2016). Despite the use of a higher density genotypic dataset, no new QTLs underlying the studied phenotypes were encountered. Among all the refined QTL with imputed HD-chip genotypes, we achieved the highest refinement of the across-family estimated CI for the LFEC QTL located on OAR6. In the case of the across family LA, the refined CI intervals (88.2- 88.3 cM) for the OAR6 QTL was within the previously detected CI interval (80.8-91.4 cM) identified with the 50K-chip genotypes (Atlija et al., 2016). But for the within family LA, for the CI region estimated for Family 7 (97.1-97.6 Mb) was slightly far away from the previous CI

estimated at the within-family level (80.4-94.8 cM) (Atlija et al., 2016). Further study should be addressed to better understand the nature of these mapping discrepancies. may suggest that across-family LA can take advantage of higher density genotype data But in case of smaller population there is a tendency for the QTL to be located toward the center of the chromosome while using higher density genotype data (Li et al., 2010). Thus, the CI region of 50k chip for within family analysis better represents the QTL region than HD chip CI region. We found five QTL related to nematode infection phenotypes reported in SheepQTLdb overlapping with the estimated refined CIs in the current study (Table 2): two QTLs on OAR6 (QTL:95607 and QTL:95613) and two on OAR8 (QTL:95608 and QTL:95617) related to the trait type "Parasite resistance" and one QTL on OAR22 related to trait type "Immune Capacity" (QTL:95609). All these five QTLs were identified on adult Churra sheep with the 50K-chip genotypes and no QTL from other studies, mainly carried out in young animals, overlapped with the regions reported here. This suggests that the QTL regions reported here for Churra sheep might be, as previously suggested by Atlija et al. (2016), exclusively related to GIN resistance in adult sheep. This could be confirmed by conducting future similar studies with adult sheep of other sheep breeds.

On the other hand, the high-resolution study of the genetic variability identified in the OAR6 LFEC QTL region highlighted, after the corresponding filtering steps, genetic variants in the *RASSF6* and *AFM* genes as potential candidate mutations to underlie this QTL effect. The gene *RASSF6,* which is included in the InnateDB database (Breuer et al., 2013), is involved in tumor suppression in humans and has been also found to dictate the degree of inflammatory response against the respiratory syncytial virus (Allen et al., 2007). Note that in the present study a missense variant was identified as a potential functional variant within the *RASSF6* gene. On the other hand, the gene *AFM* belongs to the albumin gene family. Another *ALB* gene, which also belongs to albumin family, was previously found to show a kind of association with the 50K-chip genotypic data in both LA and LDLA analyses presented by Atlija et al., (2016). The albumin protein was found to be highly expressed in duodenum of sheep resistant to gastrointestinal nematodes (Keane et al., 2006). After variant annotation, we found four missense variants in the *AFM* gene. These missense mutations might be significant with regards to nematode infection in adult sheep. Further research on this gene is needed to confirm its possible involvement in relation to nematode resistance in sheep. One missense, was deleterious (Table 3), although no studies have been reported in relation to nematode infection and the *ENSOARG00000014280* gene. Four of the tolerated missense variants were found in the *AFM* gene region and one tolerated missense variant was found within the *RASSF6* gene (Table 3). We found six positional candidate genes in the CI region of OAR8 (Table 1), out of which the gene *COL12A1* is included in InnateDB (Breuer et al., 2013). The gene *COL12A1* encodes the type XII collagen, which is associated with type I collagen. In infected mice against schistosomiasis pathogenesis, the type I collagen production has been found, as a Th2-type response cytokine, to be driven by IL-13 (Chiaramonte et al., 1999).

Conclusion

The current LA-based HD-genome scan performed for the LFEC and IgA_t indicator traits by using an imputed HD-chip dataset, has enabled to refine the QTL CIs estimated by our previously reported study based on the 50K-chip genotypes. No new QTLs were

detected for the studied traits. For all the refined QTL regions, we present a list of positional and functional candidate genes two of which were immune-related genes, and therefore potential functional variants. Additionally, by performing WGSeq of a segregating trio and a subsequent variant annotation analysis for the OAR6 QTL influencing LFEC, this study provides a list of promising candidate genes and genetic variants that may be considered as underlying casual mutations of the targeted OAR6 QTL effect. With the availability of WGSeq datasets for a large number of animals from the Churra sheep reference population, future studies should focus on the imputation of genotypes from HD-chip genotypes to WGSeq in order to further identify the polymorphisms directly related to GIN resistance in Churra sheep.

References

Allen NPC, Donninger H, Vos MD, Eckfeld K, Hesson L, Gordon L, Birrer MJ, Latif F and Clark GJ 2007. RASSF6 is a novel member of the RASSF family of tumor suppressors. Oncogene 26, 6203–6211.

Atlija M, Arranz J-JJ, Martinez-Valladares M and Gutiérrez-Gil B 2016. Detection and replication of QTL underlying resistance to gastrointestinal nematodes in adult sheep using the ovine 50K SNP array. Genetics Selection Evolution 48, 4.

Beraldi D, McRae AF, Gratten J, Slate J, Visscher PM and Pemberton JM 2007. Mapping quantitative trait loci underlying fitness-related traits in a free-living sheep population. Evolution.

Bhattacharya S, Andorf S, Gomes L, Dunn P, Schaefer H, Pontius J, Berger P, Desborough V, Smith T, Campbell J, Thomson E, Monteiro R, Guimaraes P, Walters B, Wiser J and Butte AJ 2014. ImmPort: Disseminating data to the public for the future of immunology. Immunologic Research 58, 234–239.

Breuer K, Foroushani AK, Laird MR, Chen C, Sribnaia A, Lo R, Winsor GL, Hancock REW, Brinkman FSL and Lynn DJ 2013. InnateDB: Systems biology of innate immunity and beyond - Recent updates and continuing curation. Nucleic Acids Research.

Brøndum RF, Su G, Janss L, Sahana G, Guldbrandtsen B, Boichard D and Lund MS 2015. Quantitative trait loci markers derived from whole genome sequence data increases the reliability of genomic prediction. Journal of Dairy Science.

Browning BL and Browning SR 2013. Improving the accuracy and efficiency of identity-by-descent detection in population data. Genetics 194, 459–471.

Chiaramonte MG, Donaldson DD, Cheever AW and Wynn TA 1999. An IL-13 inhibitor blocks the development of hepatic fibrosis during a T-helper type 2-dominated inflammatory response. Journal of Clinical Investigation 104, 777–785.

Chitneedi PK, Arranz JJ, Suarez-Vega A, García-Gámez E and Gutiérrez-Gil B 2017. Estimations of linkage disequilibrium, effective population size and ROH-based inbreeding coefficients in Spanish Churra sheep using imputed high-density SNP genotypes. Animal Genetics 48, 436–446.

Doerge RW 2010. A Guide to QTL Mapping with R/qtl by BROMAN, K. W. and SEN, S. Biometrics 66, 658–659.

Estrada-Reyes ZM, Rae O, Postley C, Jiménez Medrano MB, Gutiérrez JDL and Mateescu RG 2019a. Association study reveals Th17, Treg and Th2 loci related to resistance to Haemonchus contortus in Florida Native sheep. Journal of Animal Science.

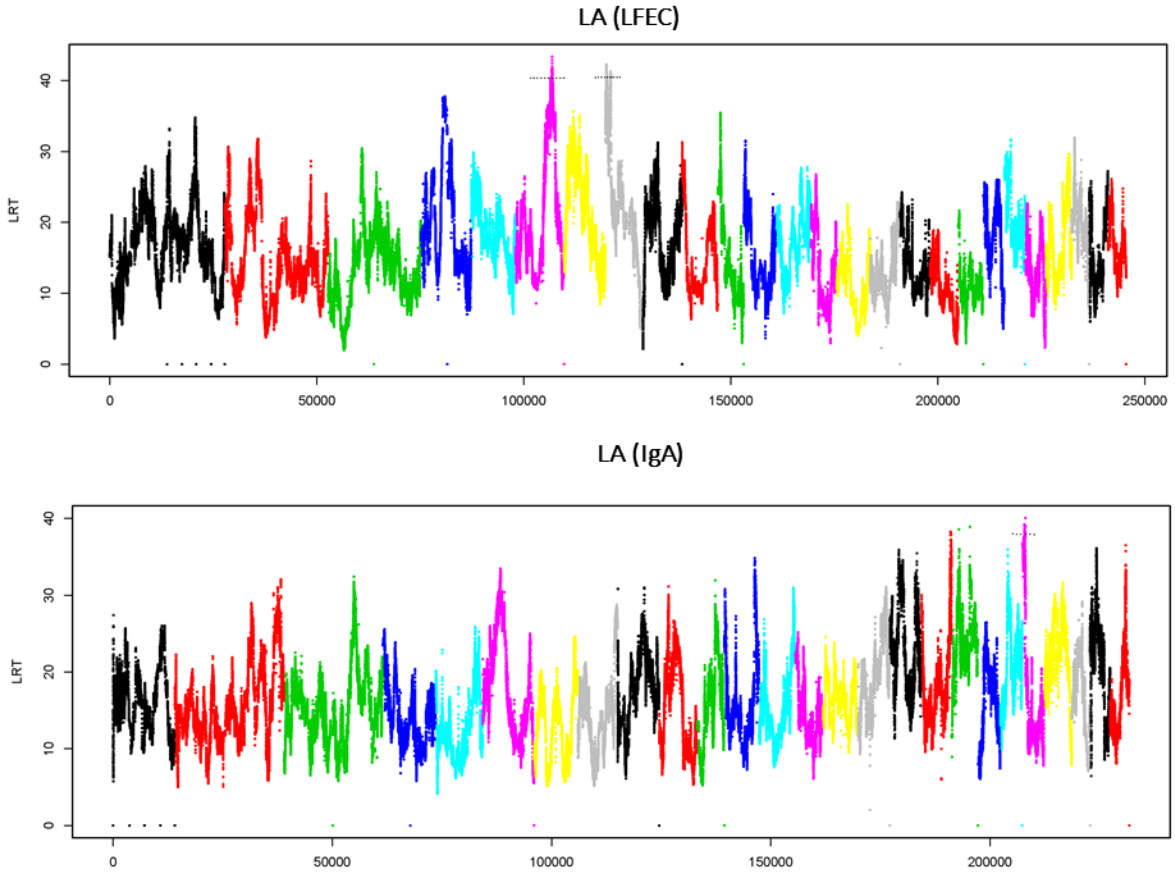
Estrada-Reyes ZM, Tsukahara Y, Goetsch AL, Gipson TA, Sahlu T, Puchala R and Mateescu

RG 2019b. Association analysis of immune response loci related to Haemonchus contortus exposure in sheep and goats using a targeted approach. Livestock Science.
Filangi O, Moreno C, Gilbert H, Legarra A, Le Roy P EJ 2010. QTLMap, a software for QTL detection in outbred populations. In In Proceedings of the 9th World Congress on Genetics Applied to Livestock Production, pp. 1–6. Leipzig.
Gutiérrez-Gil B, Pérez J, Álvarez L, Martínez-Valladares M, de la Fuente L-FF, Bayán Y, Meana A, Primitivo FS, Rojo-Vzquez FA, Arranz J-JJ, Bayón Y, Meana A, Primitivo FS, Rojo-Vázquez F-A and Arranz J-JJ 2009. Quantitative trait loci for resistance to trichostrongylid infection in Spanish Churra sheep. Genetics Selection Evolution 41, 46.
Hu ZL, Park CA, Wu XL and Reecy JM 2013. Animal QTLdb: An improved database tool for livestock animal QTL/association data dissemination in the post-genome era. Nucleic Acids Research 41, 871–879.
Al Kalaldeh M, Gibson J, Duijvesteijn N, Daetwyler HD, MacLeod I, Moghaddar N, Lee SH and van der Werf JHJ 2019. Using imputed whole-genome sequence data to improve the accuracy of genomic prediction for parasite resistance in Australian sheep. Genetics, selection, evolution : GSE.
Keane OM, Zadissa A, Wilson T, Hyndman DL, Greer GJ, Baird DB, McCulloch AF, Crawford AM and McEwan JC 2006. Gene expression profiling of naive sheep genetically resistant and susceptible to gastrointestinal nematodes. BMC Genomics 7, 42.
Kelley J, De Bono B and Trowsdale J 2005. IRIS: A database surveying known human immune system genes. Genomics 85, 503–511.
Kijas JW, Porto-Neto L, Dominik S, Reverter A, Bunch R, McCulloch R, Hayes BJ, Brauning R and McEwan J 2014. Linkage disequilibrium over short physical distances measured in sheep using a high-density SNP chip. Animal genetics 45, 754–7.
Lander ES and Botstein S 1989. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121, 185.
Le Roy P, Elsen JM, Gilbert H, Moreno C, Legarra A, Filangi O 2013. QTLMap 0.9.6 User's guide 1–56 https://forge-dga.jouy.inra.fr/attachments/download/2502/qtlmapV0.9.6.pdf, Accessed 21st Nov 2016
Li H, Hearne S, Bänziger M, Li Z and Wang J 2010. Statistical properties of QTL linkage mapping in biparental genetic populations. Heredity.
McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, Flicek P and Cunningham F 2016. The Ensembl Variant Effect Predictor. Genome Biology.
Moreno C, Jacquiet P, Bouvier F, Cortet J, Blanchard-Letort A, Guégnard F, François D, Bourdillon Y, Grisez C, Prevot F, Averadere A, Demars J, Sarry J, Stella A, Woloszyn F, Canale-Tabet K, Cabaret J, Tosser-Klopp G and Salle G 2014. Validation of QTL affecting resistance to nematodes in Sheep identified in a Back-Cross design in a pure breed population TT -. 10th World Congress of Genetics Applied to Livestock Production Genetic.
Narasimhan V, Danecek P, Scally A, Xue Y, Tyler-Smith C and Durbin R 2016. BCFtools/RoH: A hidden Markov model approach for detecting autozygosity from next-generation sequencing data. Bioinformatics.
Periasamy K, Pichler R, Poli M, Cristel S, Cetrá B, Medus D, Basar M, Thiruvenkadan AK, Ramasamy S, Ellahi MB, Mohammed F, Teneva A, Shamsuddin M, Podesta MG and Diallo A 2014. Candidate gene approach for parasite resistance in sheep - Variation in immune pathway genes and association with fecal egg count. PLoS ONE.

Veerkamp RF, Bouwman AC, Schrooten C and Calus MPL 2016. Genomic prediction using preselected DNA variants from a GWAS with whole-genome sequence data in Holstein-Friesian cattle. Genetics Selection Evolution.

Figures

Figure 1 Graphical representation of linkage analysis (LA) genome scans with imputed HD-chip genotypes for two indicator traits of parasite resistance for the following analyzed traits: *LFEC* Log-transformed faecal egg count and IgA_t Box-Cox-transformed optical density ratio (ODR) values of immunoglobulin A activity. Likelihood ratio test (LRT) values obtained across the 26 ovine autosomes are represented. For those chromosomes that harbour significant QTL, the horizontal lines indicate the 5 % chromosome-wise significance threshold for LA (a, b).



Tables

Table 1 Characterization of the significant chromosomal-wise QTLs detected for GIN resistance traits by the linkage analysis (LA) performed in the Churra population analyzed by Atlija et al. (2016) by using imputed HD-chip genotypes.

	Across family analysis						Within- family analysis		
Trait[1]	**OAR[2]**	**P_c-value[3]**	**Position of max LRT (cM)[4]**	**CI (cM)[5] TGI (Mb)[6]**	**Percentage of refinement in the QTL CI region[7]**	**Candidate gene identified[8]**	**Segregating family identifier**	**Position of max LRT (cM)[9]**	**CI (cM)[10]**
FEC	6	<0.05	88.21	88.2-88.3	99.05%	***RASSF6***, *AFM*	Family 7	97.1	97.1-97.6
	8	<0.05	2.15	1.7-2.4	87.20%	***COL12A1***, *TMEM30A, FILIP1, SENP6, U6*	Family 4	31.5	30.9-32.7
IgA_t	22	<0.05	3.59	0.0-4.4	20%	*UBE2D1, CISD1, IPMK*	Family 8	0.65	0-0.67

[1] Analyzed traits: LFEC log-transformed faecal egg count, IgAt Box-Cox-transformed optical density ratio (ODR) values of immunoglobulin A activity.
[2] OAR ovine chromosome.
[3] Pc-value Chromosome-wise significance P-value established through 1000 permutation analysis.
[4, 9] Position of the chromosome (in centi Morgans) at which the maximum likelihood ratio test of the LA is reached in the analysis involving the 14 half-sib families included in this work (across-family analysis) or the individual analysis of the segregating families (those showing a Pc-value <0.05 in the within-family analysis), respectively.
[5, 10] CI Confidence interval (in cM) estimated from the position of the max LRT for the across-family analysis and the within-family analyses, respectively, following the 1-LOD-drop-off method.
[6] TGI Target genomic interval (Mb) defined as the corresponding genomic region, according to the sheep reference genome assembly Oar_v3.1, to the CI estimated for the LA significant QTL.
[7] The percentage of refinement achieved by current replication study using imputed genotype data from 50K-chip to HD-chip compared to previous study Atlija et al. (2016) [8] based on 50k-chip genotype data.
[8]The genes indicated in bold were reported in immune database InnateDB.

Table 2 Correspondence of the sheep QTL regions identified in the present study for resistance to GIN infection in adult sheep with previously reported genetic effects (QTL and associations) influencing health related traits in sheep, based on SheepQTLdb (https://www.animalgenome.org/cgi-bin/QTLdb/OA/summary).

Target Genomic Interval	Type of analysis	Trait abbreviation	QTL ID (Sheep QTLdb)[a]	Trait Name	Chromosome: bp	Trait Class	Trait type[b]	Authors
OAR6: 88200000-88300000	QTL	FECGEN	QTL:95607	Fecal egg count	Chr.6:80866917-91374479 bp	Trait: Health	Parasite Resistance	Atlija et al. 2016
	QTL	FECGEN	QTL:95613	Fecal egg count	Chr.6:85018309-90229318 bp	Trait: Health	Parasite Resistance	Atlija et al. 2016
OAR8:1700000-2400000	QTL	FECGEN	QTL:95608	Fecal egg count	Chr.8:962426-3433729 bp	Trait: Health	Parasite Resistance	Atlija et al. 2016
	QTL	FECGEN	QTL:95617	Fecal egg count	Chr.8:37510-12848380 bp	Trait: Health	Parasite Resistance	Atlija et al. 2016
OAR22:0.0-4400000	QTL	IGA	QTL:95609	Immunoglobulin A level	Chr.22:259746-5806152 bp	Trait: Health	Immune Capacity	Atlija et al. 2016

[a] QTL identifier (ID): unique identified number of the QTL/associations annotated in SheepQTLdb.

Table 3 Functional characterization of the intragenic missense variants detected by whole genome sequencing of the analysed segregating trio which mapped within the across-family estimated confidence interval and showed concordance with the segregation pattern of the OAR6 QTL influencing the LFEC trait.

OAR[a]	Ref-allele	Alt-allele	Location	Annotation	Impact[b]	Gene symbol[c]	Ensembl Gene ID	Existing Variation	SIFT
6	G	A	6:88202978-88202978	Missense variant	Mod	*AFM*	ENSOARG00000014129	rs407636292	Tolerated (0.45)
6	G	A	6:88202978-88202978	Missense variant	Mod	*AFM*	ENSOARG00000014129	rs407636292	Tolerated (0.34)
6	A	G	6:88206674-88206674	Missense variant	Mod	*AFM*	ENSOARG00000014129	rs424283558	Tolerated (1)
6	A	G	6:88206674-88206674	Missense variant	Mod	*AFM*	ENSOARG00000014129	rs424283558	Tolerated (1)
6	C	T	6:88295388-88295388	Missense variant	Mod	***RASSF6***	ENSOARG00000014363	rs428754250	Tolerated (0.27)
6	G	T	6:88252392-88252392	Missense variant	Mod		ENSOARG00000014280 (Bovine Orthologous: ENSBTAG00000049436)	rs420795500	Deleterious (0.02)
6	G	A	6:88255946-88255946	Missense variant	Mod		ENSOARG00000014280 (Bovine Orthologous: ENSBTAG00000049436)	rs406087055	Tolerated (0.22)

[a] OAR ovine chromosome.

[b] Mod- Moderate Impact

[c]The genes indicated in bold were identified as immune-related genes due to its inclusion in the InnateDB database (https://www.innatedb.com/).

Study 5

Preliminary differential transcriptomic analysis of abomasal mucosa from resistant and susceptible sheep to gastrointestinal nematodes (GINs) after an experimental infection with *T. circumcincta*.

Praveen Krishna Chitneedi[1], Juan-Jose Arranz[1], Aroa Suárez-Vega[1], María Martinez-Valladares[2,3] and Beatriz Gutiérrez-Gil[1].

[1]Departamento de Producción Animal, Facultad de Veterinaria, Universidad de León, León, Spain; [2]Departamento de Sanidad Animal, Facultad de Veterinaria, Universidad de León, León, Spain; [3]Instituto de Ganadería de Montaña, CSIC-ULE, León, Spain.

36th International Society for Animal Genetics Conference (ISAG), 16th-21st July, 2017, Dublin, Ireland. https://buleria.unileon.es/handle/10612/6845

Preliminary differential transcriptomic analysis of abomasal mucosa from resistant and susceptible sheep to gastrointestinal nematodes (GINs) after an experimental infection with *T. circumcincta*.

P. K. Chitneedi*[1], J. J. Arranz[1], A. Suarez-Vega[1], M. Martínez-Valladares[2,3], B. Gutiérrez-Gil[1];
[1]Departamento de Producción Animal, Facultad de Veterinaria, Universidad de León, León, Spain;,
[2]Departamento de Sanidad Animal, Facultad de Veterinaria, Universidad de León, León, Spain;,
[3]Instituto de Ganadería de Montaña, CSIC-ULE, León, Spain

This study presents the preliminary results of a differential transcriptomic analysis between abomasal mucosa samples obtained from two groups of sheep classified as resistant or susceptible to gastrointestinal nematodes (GINs) after an experimental infection with *T. circumcincta.* Initially based on the faecal egg counts (FEC) of 119 Spanish Churra ewes sampled in a flock, 10 animals with the highest FEC values and 14 with the lowest FEC values were included in the study. After a first experimental infection (EI1) 18 of these animals with *T. circumcincta*, and based on the accumulative FEC estimated from day 14 to 31 post-infection, eight ewes were classified as 'Susceptible' and seven ewes as 'Resistant'. After an antihelminthic treatment, a second experimental infection (EI2) of these animals was performed. On day 7 after EI2 the animals were sacrificed and abomasal mucosa tissue samples were collected. The total mRNA extracted from these samples was were later sequenced using an Illumina Hi-SEqn 2000 sequencer by generating 'paired-end' reads of 75 bp, with a depth of 30M reads. The bioinformatics work flow analysis included the assessment of raw sequence data quality using FastQC, the alignment against the sheep reference genome (Oar_v.3.1), the quantification and normalization of gene expression performed with Cufflinks and finally the differential expression analysis performed with two different R-based packages, DESEqn 2 and edgeR. DESEqn 2 identified one differentially expressed (DE) gene, *SYT8* ($P\text{-}value_{adj}$ = 0.020) whereas the edgeR analysis found 18 DE genes among which *SYT8* was the second highly expressed gene (FDR_{adj} = 0.0023). *SYT8* is involved in trafficking and exocytosis of secretory vesicles in non-neuronal tissues. Many of the genes highlighted by the edge R are related to muscular excitation and contraction whereas one belongs to the Major Histocompatibility Complex. Because one of the proposed mechanisms of parasite resistance in sheep is the increase in peristalsis, some of the genes highlighted by the edgeR analysis may be further investigated in future studies.

KEYWORDS
sheep and related species
RNA-seq
transcriptome
complex trait
bioinformatics tools

Study 6

Exploring the mechanisms of resistance to Teladorsagia circumcincta infection in sheep through transcriptome analysis of abomasal mucosa and abomasal lymph nodes

Praveen Krishna Chitneedi[1], Aroa Suárez-Vega[1], María Martinez-Valladares[2,3], Juan-Jose Arranz[1] and Beatriz Gutiérrez-Gil[1].

[1]Departamento de Producción Animal, Facultad de Veterinaria, Universidad de León, León, Spain; [2]Departamento de Sanidad Animal, Facultad de Veterinaria, Universidad de León, León, Spain; [3]Instituto de Ganadería de Montaña, CSIC-ULE, León, Spain.

Veterinary research, 49(1), p.39. https://doi.org/10.1186/s13567-018-0534-x

Chitneedi *et al. Vet Res (2018) 49:39*
https://doi.org/10.1186/s13567-018-0534-x

VETERINARY RESEARCH

RESEARCH ARTICLE Open Access

Exploring the mechanisms of resistance to *Teladorsagia circumcincta* infection in sheep through transcriptome analysis of abomasal mucosa and abomasal lymph nodes

Praveen K. Chitneedi[1], Aroa Suárez-Vega[1], María Martínez-Valladares[1,2,3], Juan José Arranz[1] and Beatriz Gutiérrez-Gil[1*]

Abstract

The present study exploited the RNA-seq technology to analyze the transcriptome of target tissues affected by the *Teladorsagia circumcincta* infection in two groups of adult ewes showing different statuses against gastrointestinal nematode (GIN) infection with the aim of identifying genes linked to GIN infection resistance in sheep. For this, based on the accumulated faecal egg count of 18 adult Churra ewes subjected to a first experimental infection with *T. circumcincta*, six ewes were classified as resistant and six others as susceptible to the infection. These 12 animals were dewormed and infected again. After humanitarian sacrifice of these 12 animals at day 7 post-infection, RNA samples were obtained from abomasal mucosa and lymph node tissues and RNA-Seq datasets were generated using an Illumina HiSeq 2000 sequencer. The distribution of the genes based on their expression level were very similar among the two different tissues and conditions. The differential expression analysis performed with two software (DESeq and EdgeR) only identified common differentially expressed genes (DEGs), a total of 106, in the lymph node samples which were considered as GIN-activated. The enrichment analysis performed for these GIN-activated genes identified some pathways related to cytokine-mediated immune response and the *PPARG* signaling pathway as well as disease terms related to inflammation and gastro-intestinal diseases as enriched. A systematic comparison with the results of previous studies confirmed the involvement of genes such as *ITLN2*, *CLAC1* and *galectins*, in the immune mechanism activated against *T. circumcincta* in resistant sheep.

Introduction

In Spain, the dairy sheep production of indigenous sheep breeds is based on grazing livestock systems where gastrointestinal nematode (GIN) infections pose a major health problem to adult ewes and cause important economic losses [1]. Ovine resistance to the GIN infection is a highly complex character [1] and identification of genes influencing increased resistance to GIN infection would be of interest to enhance the efficiency of selection in commercial flocks through the use of molecular information. Hence, several studies have tried to identify QTL influencing indicator traits for GIN infection, such as fecal egg count (FEC), serum levels of IgA and pepsinogen. However, due to the major economic impact of GIN in young animals in countries such as the UK or Australia, most of these QTL studies are focused on lambs [2–5] whereas a limited number of studies have focused on adult sheep [6, 7] or combined data from aged ewes and lambs [1]. In a study carried out with adult sheep, Atlija et al. [7] suggested that some of the QTL identified that did not show overlapping with previous studies in lambs could be related to specific mechanisms of the immune response that is activated in adult animals. This would support the theory suggested by Stear et al. [8] that the genetic variation in FEC in lambs is

*Correspondence: beatriz.gutierrez@unileon.es
[1] Departamento de Producción Animal, Facultad de Veterinaria, Universidad de León, Campus de Vegazana s/n, 24071 León, Spain
Full list of author information is available at the end of the article

predominantly a consequence of genetic variation in worm length and hence worm fecundity whereas, in contrast, mature sheep may be able to regulate not only fecundity, but also worm number.

As an alternative method to identify genes linked to GIN resistance, studies based on gene expression analysis can identify those genes whose expression may differ between animals and the gene pathways activated depending on the status against infection. In sheep infected by GIN, these kinds of studies based on expression analysis by RT-PCR of specific candidate genes [9–11], the use of microarrays [12, 13] or recently on the RNA-Seq technology [14, 15], have been mainly focused on young animals. For example, for lambs infected with GIN, different studies have shown that the animal status is dependent on the different type of host immune response that is activated against the nematodes in the abomasal (gastric) mucosa and abomasal lymph node [16, 17]. The *T. circumcincta* infection in lambs has been shown to trigger significant Th2 cytokine changes in mucosa with increased mucosa production of eosinophilia, mastocytosis and neutrophils. Upon GIN infection, the intelectins (*ITLN1*, *ITLN2* and *ITLN3*) [9, 10] and the interleukin (*IL-3*, *IL-4*, *IL-5* and *IL-13*) [18–20] transcripts have been consistently found to be upregulated to induce Th2 response in tissues like abomasal mucosa and lymph nodes of infected lambs whereas this expression was absent in naïve sheep. Irrespective of the breed and infective nematode species, some conserved gene expression responses were identified in relation to early inflammation in resistant lambs and in relation to a chronic inflammatory state in susceptible lambs [11].

In order to gain knowledge on the different gene expression responses activated in adult sheep between resistant and susceptible animals against *T. circumcincta* infection, the present study presents a comparative analysis of RNA-Seq datasets obtained from abomasal mucosa and abomasal lymph node obtained from two groups of adult Churra sheep previously classified as resistant and susceptible against this GIN infection. RNA-Seq has been shown as a powerful deep-sequencing technology that can help to elucidate previously inaccessible complexities underlying gene expression responses related to complex quantitative traits such as the resistance to nematode infection [21]. Hence, the study of complete target tissue transcriptomes presented here tries to provide a global picture of the different mechanisms activated as response to infection by GIN in adult sheep.

Materials and methods

Animals and experimental infections

Faecal samples of adult dairy ewes from four Churra dairy sheep flocks reared under semi-intensive management and belonging to the ANCHE breeders' association (national association of Spanish Churra sheep breeders) were collected to initially assess the GIN infection levels after natural infection. Based on individual FEC, the farm showing the largest range for FEC was selected for additional sampling. In the selected flock, faecal samples were collected from a total of 119 adult ewes 3 months after the last deworming treatment. Based on the individual FEC measures, a total of 18 sheep (age range 6–8 years old) showing the most extreme FEC values were selected for our study. A first experimental infection (EI1) was performed based on a single oral administration of 50 000 *T. circumcincta* third stage larvae (L3) on these animals and after a treatment with one oral dose of ivermectin (0.2 mg/kg bw, Orame®, Merial, Spain). After this infection, collection of faeces was performed every 2 days, starting from day 14 to day 31 post-infection to calculate the accumulated FEC from each animal. Based on these values, six sheep were classified as susceptible and six sheep as resistant to infection by *T. circumcincta*. One month after the EI1, all these 12 selected ewes were treated with moxidectin by subcutaneous injection (0.2 mg/kg bw, Cydectin®, Zoetis, Spain) and 3 weeks later were exposed to a second experimental infection (EI2) with a single oral dose of 70 000 *T. circumcincta* L3. At day 7 after EI2, the animals were sacrificed by an intravenous injection with a lethal dose of 20 mL per sheep of sodium pentobarbital (Dolethal®, Vetoquinol, Spain). At necropsy, abomasal mucosa and lymph node samples for all animals were immediately collected in RNAlater™ Stabilization Reagent (Sigma-Aldrich, St. Louis, MO, USA). These samples were stored overnight at 4 °C and then frozen at −80 °C.

RNA extraction, sequencing and bioinformatics analysis

mRNA were extracted from the abomasal mucosa and abomasal lymph node samples from the six resistant and six susceptible selected animals, using the Absolutely RNA miRNA Kit from Agilent (La Jolla, CA, USA). RNA integrity (RIN value) was analyzed using the agilent 2100 bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). Considering both tissue samples the RIN values of the RNA samples ranged between 6.7 and 7.8. The preparation of the libraries and subsequent sequencing was performed with an Illumina HiSeq sequencer 2000, generating stranded paired-end reads of 75 bp with a depth of 30 million reads.

The read quality of each sample was assessed using FastQC V_0.11.5 software [22]. The high quality read samples were later aligned against the ovine reference genome (Oar_v3.1) using the alignment software STAR_v2.5.2b [23]. We used the option *–outSAMtype BAM Unsorted* to obtain an unsorted bam file instead of the

sam file obtained by default with the STAR software. The unsorted aligned reads were indexed and sorted by read names using Samtools_v1.3.1 [24]. The sorted reads were counted for features (genes) in each sample using the HTSeq-count software [25] with the intersection-strict mode and reverse stranded option and using the information of the reference sheep genome annotation (Oar_v3.1.88.gtf) available at the FTP Download-Ensemble (release 88). This produces a feature list (gene list) with number of raw counts for each sample.

Before performing the differential gene expression analysis, we tried to quantify the abundance of all the annotated genes for each abomasal lymph node and abomasal mucosa sample analyzed. The gene expression levels were normalized by library size and gene length by calculating fragments per kilobase of exon per million fragments mapped (FPKM) with the RSEM software package [26] using the ensemble genome annotation (Oar_v3.1) as a reference. Initially the reference sheep genome annotation (Oar_v3.1.88.gtf) was preprocessed using the option *rsem-prepare-reference (–star)*. After that, we estimated the gene expression levels for each sample using the *rsem-calculate-expression (–star)* option of the RSEM program. The genes with at least 0.01 FPKM in each sample were considered as expressed and the mean gene expression across susceptible and resistant samples of each tissue were classified as high (> 500 FPKM), medium (10–500 FPKM) and low (< 10 FPKM) expressed genes based on their FPKM values.

The differential expression analysis was performed using two R based packages, EdgeR [27] and DESeq [28], with the raw counts from each sample obtained from HTSeq count. In these two software, the differential analysis of count data was performed in a similar way, but following different approaches to estimate normalization and dispersion. DESeq is, in comparison, less powerful but EdgeR is more sensitive to outliers [29]. Thus, after performing differential expression analysis between the resistant and the susceptible sample groups individually with the two software, only the genes commonly identified as differentially expressed genes (DEGs) by both EdgeR (FDR < 0.05) and DESeq (adjusted $P < 0.05$) were considered as "GIN-activated" DEGs. This double analysis is expected to reduce the presence of false positive results from our analysis. Based on log fold change (Log_2FC) values, the DEGs were further classified as up-regulated in each group. These gene lists were subjected to three types of gene-set enrichment analysis [gene ontology (GO) analysis, KEGG pathway analysis and disease association analysis] using the web-based tool WebGestalt [30]. For these analyses, the human genome was considered as a reference and the parameters considered were the default statistical method Hypergeometric, multiple test adjustment with the BH method (Benjamini–Hochberg FDR), the significant level of adj$P < 0.05$. In addition, for a term to be considered significantly enriched a minimum of five genes were required for GO and disease association analyses, and a minimum of three genes were required for the KEGG pathway analysis.

Results

Infected sheep status confirmation

Based on the accumulated FEC estimation carried out on the 18 animals subjected to EI1, six animals were classified as "susceptible" (Chu6, Chur9, Chu11, Chu14, Chu17 and Chu19) and six as "resistant" (Chu1, Chu2, Chu7, Chu8, Chu15 and Chu21). Two of the individuals that were classified as "resistant" had zero counts throughout the control phase (Chu7, Chu15). The remaining five animals showed an uncertain profile between the two defined categories and therefore were discarded from further analysis. The mean accumulated FEC in the susceptible group was 5594 ± 2661 eggs per gram (epg) and 308 ± 338 epg in the resistant group.

Gene expression level

A total of 24 mRNA samples, 12 from abomasal mucosa and 12 from abomasal lymph node, were considered for massive parallel sequencing. The FASTQC analysis showed that the quality of the sequenced reads for all the samples in two tissues under analysis were of high quality and thus, no trimming was performed. Across all the samples on average, around 72.3 and 79.2% of reads were uniquely mapped against the reference genome Oar_v3.1, for abomasal mucosa and abomasal lymph node tissues respectively. The aligned reads were sorted and counted against the list of annotated genes from the sheep reference genome Oar_v3.1. Out of the 27 054 annotated genes in this reference genome, on an average, around 16 210 (59%) genes in abomasal mucosa tissue sample data and 16 808 (62%) genes in abomasal lymph node tissue samples had at least one raw read count. The heatmap plot including the raw read count data of all abomasal mucosa and lymph node samples shows a clear distinction between both types of tissues (Figure 1). However, the total number of expressed genes was very similar among the two tissues and conditions, with an average of 15 627 ± 136 and 15 708 ± 154 genes expressed in abomasal mucosa and lymph node respectively (see Additional file 1). The distribution of the expressed genes across the different expression levels considered was also very similar in the two tissues and conditions (~1.31% of the genes classified as highly expressed genes; ~43% of the genes showing an intermediate expression level; ~55% of the genes with a low expression level). In the abomasal mucosa 179 and 180 genes were identified as highly

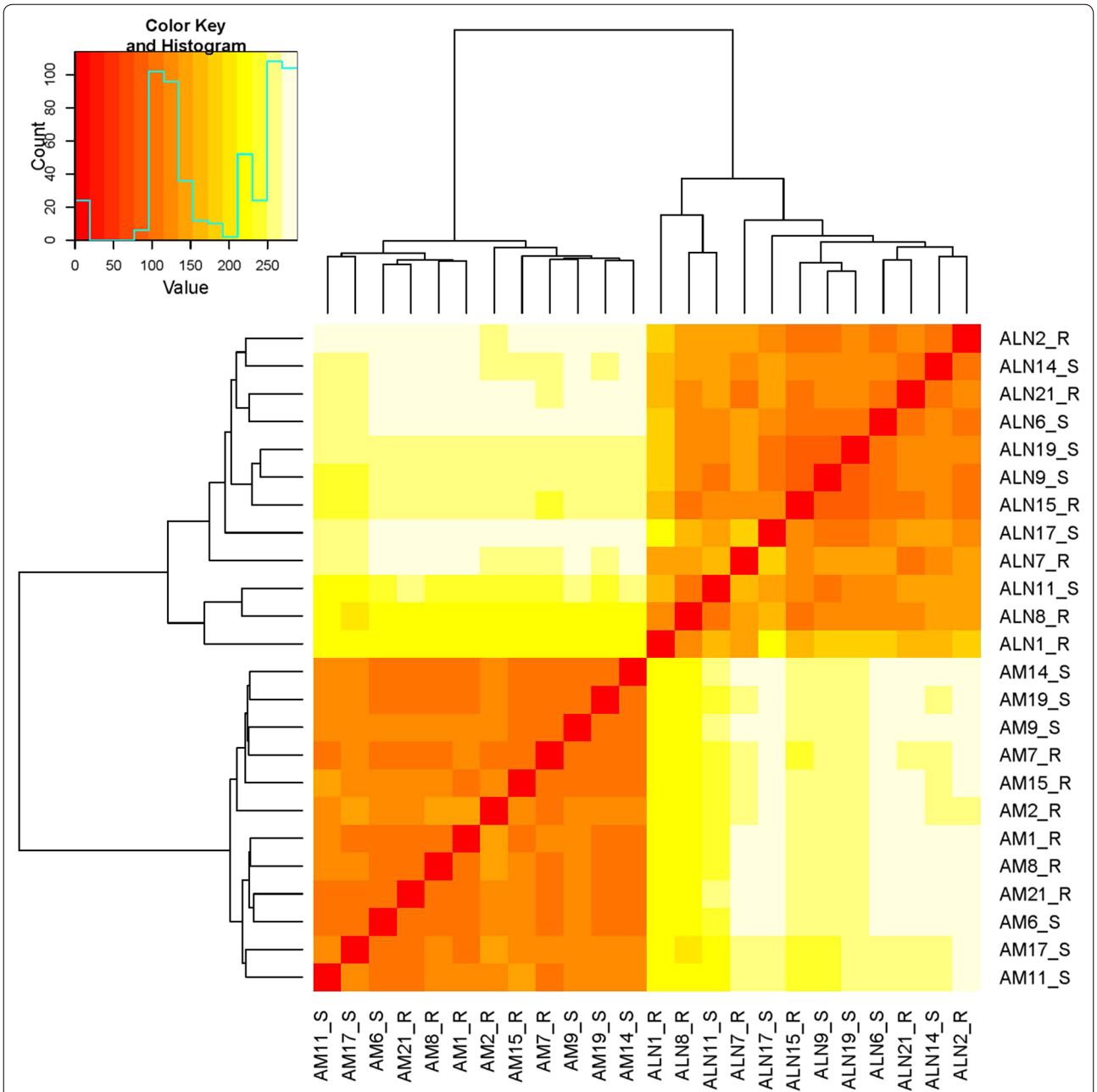


Figure 1 Heatmap plot and hierarchical clustering of all the genes expressed in abomasal mucosa (AM) and abomasal lymph node (ALN) samples analyzed herein. The plot includes the heatmap for samples of both tissues from 12 adult sheep subjected to an experimental infection with *T. circumcincta* and previously classified as resistant ("R") and susceptible ("S") based on a previous experimental infection with *T. circumcincta*.

expressed genes for the resistant and susceptible samples respectively. These highly expressed genes involved approximately 1.1% of the total FPKM in each group and were considered core genes (179 common core genes for the two compared groups, 0 specific core genes for the resistant group and 1 specific core gene for the susceptible group). In the lymph node 251 and 214 genes were identified as highly expressed genes for the resistant and susceptible samples, respectively (201 genes identified as core genes in the two conditions, 50 specific core genes of the resistant group and 13 specific core genes of the susceptible group). Considering the two tissues, we found 123 common genes in the highly expressed category (> 500 FPKM), 56 highly expressed genes specific to

abomasal mucosa and 78 highly expressed genes specific to abomasal lymph node tissues (Additional file 2). By performing GO enrichment analysis with these common and specific highly expressed category genes, we found that the enriched terms resulting from the genes common to both tissues were related to basic physiology like *translation termination*, *translational elongation*, *mRNA catabolic process*, *RNA binding*, etc. (Additional file 3). The GO analysis for the core genes specific to abomasal mucosa highlighted terms such as *digestion*, *ATPase activity*, *hydrolase activity*, *ATP hydrolysis coupled proton transport* (Additional file 4), whereas in the analysis of the core genes specific to abomasal lymph node some of the enriched terms were *muscle contraction*, *collagen fibril organization*, *protein binding*, *extra cellular matrix*, etc. (Additional file 5). Within each tissue, the heatmap plot did not show a clear clustering of the samples based on the two contrasting groups, resistant versus susceptible samples. However, from the clustering obtained for the lymph node samples there were two resistant and two susceptible samples showing the most divergent expression pattern (Additional file 6).

After performing the differential expression analysis with the abomasal mucosa RNA-Seq data, we found 33 DEGs (adjusted $P < 0.05$) using DESeq and no significant DEGs with EdgeR (FDR < 0.05) (Additional file 7). Out of these 33 significant genes, eight were up-regulated in resistant sheep and 25 were up-regulated in susceptible sheep. The differential expression analysis of the abomasal lymph node samples with DESeq (adjusted $P < 0.05$) and EdgeR (FDR < 0.05) showed, respectively, a total of 261 and 125 DEGs. Among them, 106 genes were commonly identified as DEGs by both software and were considered as GIN-activated DEGs (Figure 2; Additional file 8). Out of these 106 common GIN-activated genes, 71 were up-regulated in resistant sheep and 35 were up-regulated in susceptible sheep. Note that for the intelectin gene, *ITLN*, a survey of orthologous genes indicates that this transcript corresponds to the bovine and human *ITLN2* gene.

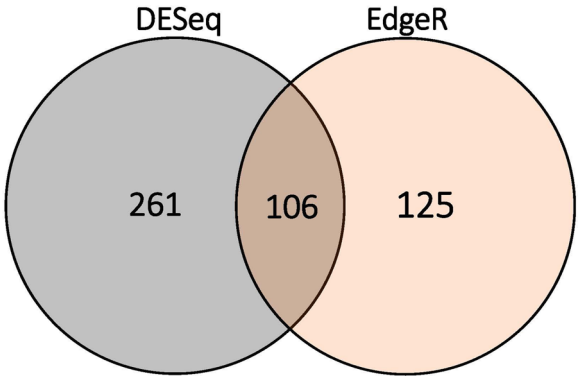


Figure 2 Venn diagrams showing the total number of DEGs of abomasal lymph node samples with DESeq and EdgeR programs. The Venn diagram shows the total number of DEGs expressed with DESeq, EdgeR and the common DEGs between DESeq and EdgeR (GIN-activated genes) identified in the analysis of the 12 abomasal lymph node samples.

The GO enrichment analysis performed for the GIN-activated DEGs with upregulated expression in resistant sheep shows nine significant terms in the *biological process* database, two of them related to cytokine response (*cytokine-mediated signaling pathway*, *cellular response to cytokine stimulus*), eight significant terms in the *cellular component* database and no significant term in the *molecular function* database (Additional file 9). On the contrary, for the GIN-activated DEGs showing upregulated expression in susceptible sheep, only three significant terms included in the *cellular component* database were identified (Additional file 9). The other two enrichment analyses performed, the KEGG pathway and disease association analyses, only identified significant terms for the GIN-activated DEGs showing upregulated expression in resistant sheep (Table 1). Among the three significant terms identified in the KEGG pathway analysis, we consider worth mentioning the *PPAR signaling pathway*, whereas the four significant terms identified in the disease association analysis were clearly related to the studied phenotype (*gastrointestinal neoplasms*, *intestinal/gastrointestinal diseases*, *inflammation*) (Table 1).

Discussion

The mechanisms of sheep resistance to GIN infections involve complex immune responses. In relation to the specific infection with *T. circumcincta*, several studies have previously been carried out in lambs showing that, as a local immune response to *T. circumcincta* infection, the levels of IgA and eosinophils were increased in the abomasal tissues of infected lambs [31–34]. Based on cDNA microarray-based studies, several up-regulated transcripts were found in abomasal tissues of *T. circumcincta* infected lambs and confirmed the activation of the Th2-type immune response in mucosa tissue, including eosinophilia and mastocytosis [35–37]. In this study, we tried to decipher the immune mechanisms activated in adult sheep during infection with *T. circumcincta* through the transcriptome analysis of the two main tissues targeted by the infection, the abomasal mucosa and the abomasal lymph node.

Although our initial purpose was to classify the animals to be studied as resistant and susceptible based on natural infection, the lack of homogeneity regarding the infection level when working with animals in pasture determined the need to perform a first experimental infection (EI1)

to ensure similar infection levels among all the animals included in the study.

The results of the differential expression analyses performed in the two tissues studied suggest that at day 7 post-infection there is not a clear differential gene expression response in the abomasal mucosa (due to the lack of DEGs identified by EdgeR). However, a differential response in the abomasal lymph node was clearly observed, based on the identification of 106 genes commonly identified as DEGs by the two software. We acknowledge that the results from the differential expression analyses performed with EdgeR and DESeq may include a similar fraction of false negative, as they both rely on a negative binomial model and use the false discovery rate procedure [38] to adjust for multiple testing. In this regard, Zhang et al. [39] suggested that taking the intersection of DEGs from two or more tools is recommended if the number of false positives is a major concern in the study. Following this, we considered for further analyses those genes that were identified as DEGs by the two implemented methods, reducing the initial number of DEGs from 261 and 125, respectively for DESeq and EdgeR, to 106 genes defined as GIN-activated DEGs.

Hence, focusing on the 106 GIN-activated DEGs identified in the complete analysis of the lymph node samples, some of the significant terms identified by the enrichment analyses showed a clear correspondence with immune response mechanisms. In particular, the two GO terms related to cytokines highlighted by the GO analysis of genes upregulated in resistant ewes (*cytokine-mediated signaling pathway* and *cellular response to cytokine stimulus*) were related to the same five genes (*PALM3*, *DUOX2*, *PPARG*, *PF4*, *IL5RA*) (Table 1). The role of *PPARG*, *DUOX2* and *IL5RA* genes in relation to the immune response has been previously reported in different nematode infection studies [11, 20, 40]. In our study, the *PPARG* gene also supported the identification of the *PPAR* signaling pathway in the KEGG analysis, together with the *RXRG* and *AQP7* genes. The *PPARG* gene was also linked to many of the significant related terms highlighted by the disease association analysis (*gastrointestinal neoplasms*, *intestinal diseases*, *inflammation* and *gastrointestinal diseases*, Table 1). The *IL5RA* gene also supported the identification of the term *inflammation* as enriched in the disease association analysis (Table 1).

Some of the DEGs identified in our study have been previously identified as DEGs by other sheep gene expression analysis in relation to gastrointestinal nematode infection. To help assess the level of correspondence of our results with other studies, we summarize the results of our comparative literature survey in Table 2. We show that 10 out of the 106 GIN-activated genes identified in our study have been previously reported to show a modified expression due to GIN infection: *ITLN*, *LYZ*, *LOC443162* (*galectin 14*), *LGALS4* (*galectin 4*), *CLCA1*, *ALPL*, *PDZK1IP1*, *PPARG*, *KRT5*, *IL5RA* (genes highlighted in bold font in Table 2). Note that for the gene annotated as *ITLN* in the sheep genome, the ortholog analysis clearly shows that it corresponds to the

Table 1 Results from enrichment analyses performed on the 71 GIN-activated DEGs up-regulated in resistant sheep identified for abomasal lymph node samples

Database	Name	Nb. of genes	ID	Gene symbol	Statistics**
KEGG pathway analysis	Glycine, serine and threonine metabolism	3	260	*ALAS2*, *PSPH*, *GLYCTK*	C = 32; O = 3; E = 0.03; R = 103.67 rawP = 3.33e−06; adjP = 9.99e−06
	PPAR signaling pathway	3	3320	*RXRG*, *PPARG*, *AQP7*	C = 70; O = 3; E = 0.06; R = 47.39; rawP = 3.59e−05; a djP = 5.39e−05
	Metabolic pathways	7	1100	*CMBL*, *ALAS2*, *CKMT1A*, *PSPH*, *GLYCTK*, *ALPL*, *TST*	C = 1130; O = 7; E = 1.02; R = 6.85; rawP = 6.15e−05; adjP = 6.15e−05
Disease association analysis	Gastrointestinal neoplasms	5	DB_ ID:PA444257	*KRT20*, *PGC*, *B4GALNT2*, *PPARG*, *LGALS4*	C = 354; O = 5; E = 0.32; R = 15.62 rawP = 1.66e−05; adjP = 0.0002
	Intestinal diseases	5	DB_ ID:PA444632	*KRT20*, *PGC*, *PPARG*, *SLC22A4*, *LGALS4*	C = 331; O = 5; E = 0.30; R = 16.70 rawP = 1.20e−05; adjP = 0.0002
	Inflammation	5	DB_ ID:PA444620	*SFTPD*, *PPARG*, *PF4*, *CLCA1*, *IL5RA*	C = 435; O = 5; E = 0.39; R = 12.71 rawP = 4.43e−05; adjP = 0.0003
	Gastrointestinal diseases	5	DB_ ID:PA444256	*KRT20*, *PGC*, *PPARG*, *SLC22A4*, *LGALS4*	C = 413; O = 5; E = 0.37; R = 13.39 rawP = 3.46e−05; adjP = 0.0003

**C: the number of reference genes in the category, O: the number of genes in the gene set and also in the category, E: the expected number in the category, R: ratio of enrichment, rawP: *P* value from hypergeometric test, adjP: *P* value adjusted by the multiple test adjustment.

Table 2 Gastro-intestinal nematode infection studies that have overlapping genes with our study

Animal age/breed	Infection status	Tissue used for study	*Nematode species*	*Genes*[a]	Technique	Study
Yearling Scottish Blackface sheep	Challenged naive sheep	Abomasal mucosa	*T. circumcincta*	***ITLN2***, ***IL4***, ***galectin 14***	RT-PCR, Western blot	[9]
Yearling sheep	Immune day 5 vs naïve day 5	Abomasal mucosa	*T. circumcincta*	***CLCA1***, ***PDZK1IP1***	cDNA microarray, RT-PCR, QT-PCR	[37]
Yearling sheep	Naive sheep	Abomasal mucosa	*T. circumcincta*	***ITLN2***, *ITLN1*, *ITLN3* ***galectin 4***, *galectin 1*	cDNA microarray, RT-PCR, QT-PCR	[37]
Yearling sheep	Naive day 5 vs naive day 0	Abomasal mucosa	*T. circumcincta*	***LYZ***, *MMP13*	cDNA microarray, RT-PCR, QT-PCR	[37]
6 months old Merino-cross wethers	Infected sheep	Abomasal mucosa	*H. contortus*	***ITLN2***, ***CLCA1***, *interleukins*	Sequential microarray (across all arrays)	[12]
6 months old Merino-cross wethers	Infected sheep	Abomasal mucosa	*H. contortus*	***ALPL***, ***PDZK1IP1***	Sequential microarray (day 22 vs day 3 biopsies)	[12]
Lambs	Primary challenge vs Tertiary challenge	Lymph node	*T. colubriformis*	***PPARG***, ***KRT5***, *SLC31A2*, *KRT18*	micro array data, QT PCR	[13]
Lambs	Resistant sheep	Abomasal mucosa	*T. colubriformis*, *H. contortus*	*DUOX1*, *IL2RA*, *IL10*	RT-PCR	[11]
Adult sheep	Immune sheep	Abomasal mucosa	*T. circumcincta*	***LYZ***, ***ITLN2***, *ITLN3*, *CLCA*, *KRT10*, *KRT8*, *KRT19*	SDS-PAGE and Shotgun proteomics	[44]
Lambs Scottish Blackface	Susceptible sheep	Lymph node	*T. circumcincta*	*SLC30A2*, ***galectin 14***	RNA-seq	[14]
1 year old canaria hair breed (CHB) and canaria sheep (CS)	Infected CHB	Abomasal mucosa	*H. contortus*	*galectin 15*, *IL5*, ***ALPL***, *MMP1*, *MMP11*, *MMP14*, *MMP2*	RNA-seq, RT-PCR	[15]
1 year old canaria hair breed (CHB) and canaria sheep (CS)	Infected CHB and CS	Abomasal mucosa	*H. contortus*	*SLC2A3*, *IL1RL1*	RNA-seq, RT-PCR	[15]
Lambs Scottish Blackface	Resistant vs Control	Lymph node	*T. circumcincta*	***IL5RA***, *IL13*, *IL13RA2*, *IL1RL1*, *IL4*, *SLC9A4*	cDNA microarray, RT-PCR	[20]
Lambs (Scottish Blackface x Leicester)	Infected sheep	Abomasum, lymph node	*T. circumcincta*	***ITLN2***, *ITLN1*, *ITLN3*	Semi-quantitative RT-PCR	[10]

[a] List of genes whose expression level is affected by GIN-infection. Those highlighted in bold font were also identified as GIN-activated differentially expressed genes in our study.

ITLN2 gene, which has been identified by other studies as activated by the GIN infection. The rest of genes presented in Table 2 not highlighted in bold font are genes that belong to the same family as some of the GIN-activated genes reported here.

Among the list of ten genes commonly identified by our study and other authors to be responsive to GIN infection, *PPARG*, *LYZ*, and *IL5RA* are directly related to inflammatory response. *PPARG* encodes for the *peroxisome proliferator activated receptor gamma*, which is a ligand activated transcription factor that regulates adipocyte differentiation and glucose homeostasis, but it has also been recognized as playing a key role in the immune response through its ability to inhibit the expression of inflammatory cytokines and to direct the differentiation of immune cells towards anti-inflammatory phenotypes [41]. In our study the *PPARG* gene was found to be upregulated in resistant ewes compared with susceptible ewes. A modified expression pattern of the *PPARG* gene in relation to the infection by *T. colubriformis* and *H. contortus* in sheep has already been reported by Andronicos et al. [13] (Table 2). A study carried out on naïve Perendale lambs suggested this gene plays a role in coordinately regulating genes more highly expressed in the intestine of the susceptible lambs [35]. In mice, high expression of the *PPARG* gene was found in response to nematode infection and the mice lacking the *PPARG* gene were unable to mount protective immune response to nematode infection. Hence, *PPARG* was suggested as a factor that drives type 2 responses in worm infection [40].

The interleukin gene *IL5RA* was also a GIN-activated gene up-regulated in resistant sheep. This gene supported the enriched GO terms *cytokine-mediated signaling pathway* and *cellular response to cytokine stimulus* and the disease association related term *Inflammation*. This gene was also found to have an increased expression

in resistant Scottish Blackface lambs to *T. circumcincta* by Gossner et al. [20] (Table 2). IL5RA is required for the biological activities of IL5 to promote eosinophil-mediated activation and recruitment into tissues in acute inflammatory responses [42]. Increased levels of eosinophils have been classically linked to resistance to *Trichostrongyle* parasites in sheep. Hence, eosinophilia has been suggested as a marker of resistance to *T. circumcincta* in Scottish Blackface lambs [34], whereas a more pronounced eosinophilia has been documented in animals bred for increased resistance to *T. colubriformis* infection [43].

The *LYZ* gene encodes lysozyme, a protein with antibacterial activity. In our study this gene was the most highly up-regulated GIN-activated gene in susceptible sheep. This observation agrees with the work of Knight et al. [37] who reported, in the abomasal mucosa, down-regulation of members of the gastric lysozyme family (*LYZ 1A*, *2A*, *3A* and *4A*) in immune versus naïve sheep at days 2 and 5 post-challenge with *T. circumcincta* (Table 2). Gastric lysozyme genes are highly expressed in the ovine abomasum and are thought to act as a major digestive enzyme of the peptidoglycan cell walls of bacteria entering from the rumen, functioning at low pH. Other studies, however, have found genes of the lysozyme family to be up-regulated in abomasal epithelial extracts from previously infected sheep versus naïve sheep [37, 44] (Table 2). Also the alterations in lysozyme production have been suggested to contribute to the resulting nutritional loss seen in infected animals [37].

Another GIN-activated gene up-regulated in resistant sheep based on our study, *CLCA1*, has also been identified as up-regulated in immune lambs in the study reported by Knight et al. [37] and in the *H. contortus* challenged yearling lambs analyzed by Rowe et al. [12]. The encoded protein of this gene is thought to act as a multifunctional signaling protein, including an early modulator of immune responses by regulation of cytokines [45]. Proteins of the CLCA family may contribute to parasite expulsion by being responsible for mucus hydration across the gut epithelium and smooth muscle contraction [12] (Table 2).

Another transcript related to mucous cells, *ITLN*, was the third most highly up-regulated in resistant sheep in our study. The intelectin 1 and 2 are protein coding genes related to carbohydrate binging. Other studies have already reported an increased expression of this gene in relation to the infection response of lambs to *T. circumcincta* [9, 37] and *H. contortus* infection [12]. The early expression of *ITLN* post-challenge in immune yearling sheep compared with naïve yearling sheep was suggested as a protective role and it may also alter the characteristics of mucus leading to worm entrapment [9] (Table 2). Up-regulation of *ITLN* has also been reported in resistant mice in response to *Trichuris muris* infection [46]. The expression of *ITLN1* and *ITLN3* was found in lymph node tissues in response to *T. circumcincta* infection in Scottish Blackface x Leicester lambs but no protein expression was found on immunohistochemistry [10] (Table 2). In cattle, the expression of *ITLN2* was reported in abomasal mucosa tissue of resistant 12 month old Angus cattle [47]. In murine models the intelectin transcript plays a key role in the expression of *IL-25*, *IL23* and has been found to amplify type 2 immune response in asthma and atrophic dermatitis conditions in humans [48]. Overall, our study supports the role of some genes such as *ITLN2*, *CLCA1* in adult sheep GIN resistance. These two genes have been previously reported to be increased in immune lambs and to be up-regulated in resistant adult sheep. Whereas the expression of these genes in lambs was found to be affected by infection in abomasal mucosa, our study did not identify an altered pattern in that tissue but only in the lymph node.

Three genes identified in our study as GIN-activated, and up-regulated in resistant sheep, belong to the galectin family, *LOC443162* (Gal-14), *LGALS4* (Gal-4) and *LOC101102156* (Gal-9). Galectins mediate innate and adaptive immune functions by modulating the activity of complement receptor 3, macrophage and dendrocyte adhesion to lymphocytes [49]. The expression of *Gal-14* was found in adult sheep after exposure to an allergen (house dust mites) and may be responsible for eosinophil function and inflammation due to allergy [50]. There was a significant up-regulation of *Gal-4* in challenged naïve yearling lambs after infection with *T. circumcincta* L3 [37]. The expression of *Gal-14* was maximum at day 10 post-challenge in yearling Scottish Blackface lambs previously infected with *T. circumcincta* [9]. In Scottish Blackface lambs, the differential expression of the *Gal-14* gene was reported in abomasal lymph node of resistant animals after 14-day post-infection with *T. circumcincta* [14] (Table 2). In addition, a role of galectins has been widely reported in reference to different host parasite interactions and these proteins appear to be responsible for adhesion of pathogens to host cells and host adaptive immunity [51].

Our list of GIN-activated genes also includes some genes belonging to the same gene families reported in other relevant GIN studies. Some of these include genes from the matrix metallopeptidase (MMP) family (*MMP28*) and solute carrier family genes (*SLC22A4* and *SLC25A34*, *SLC38A2*). The role of these genes in inflammatory diseases has been previously described. Hence, different matrix metallopeptidase family genes were related to inflammatory response in humans, in particular the *MMP28* gene was responsible for altering

inflammatory response in mice [52]. Also, the expression of other SLC family members were reported in response to nematode infection in both susceptible and resistant sheep and variants in the *SLC22A4* gene have been associated to Inflammatory Bowel disease in humans [53]. Whereas, high expression of *SLC30A2* in susceptible sheep in response to *T. circumcincta* infection has been reported in Scottish Blackface lambs [14], *H. contortus* infection shows a significant impact on the expression of the *SLC2A3* gene in Canary sheep breeds [15] (Table 2).

It is noteworthy that there is a list of about 86 genes identified as GIN-activated in our study that have not been reported or that do not belong to gene families considered as responsive to GIN infection in previous studies (Additional file 10). Interestingly, for some of these genes we found connection with the immune response (*PGC*, *SFTPD*, *TUBA4A*, *SST*, *BPIFB1*, *PF4*, *B4GALNT2*, *JCHAIN*, *AQP7*, *KLHL25*, *NEDD4*, *ANO6*). Because most of the previously reported studies considered in our comparative survey (Table 2) are focused on lambs, we think that the genes reported here in Additional file 10 might indicate genes that are specifically activated in adult sheep and not in lambs.

Other genes identified as DEGs by other studies have not been found as GIN-activated genes in the present work. Hence, genes belonging to the chemokine (*CCL*, *CXCL*) and the collagen families (*COL9A2*, *COL6A5*), or the interferon gamma gene (*IFNG*), etc. which are reported as DEGs in abomasal lymph node in resistant lambs as a response to *T. circumcincta* infection [14, 20] were not found as DEGs in our study. This may be due to differential activation of the immune response to *T. circumcincta* infection in adult sheep compared with young animals although some similar pathways were activated. Also the differences among the different studies regarding the species responsible for gastrointestinal parasite infection, the different experimental approaches (natural vs artificial), the post-infection sample extraction timings, or the environmental conditions (dry vs humid climates) are other major factors that could explain discrepancies related to the specific genes activated during GIN infection.

Also some of these genes may have been identified as DEGs in our study by one of the two DE analyses performed. Some genes such as *CENPN*, *FABP4*, *HSH2D*, *KIF2C*, *KIFC1*, *MMP1*, *NDC80*, *NEK2*, *PKMYT1*, *SFN*, *SPAG5*, *UBE2C*, *UHRF1*, which have been reported as GIN related by other studies were only identified as DEGs by DESeq but not by EdgeR. Hence, the approach implemented in our paper to avoid false positives, may have determined a restrictive threshold and we may have lost some important genes in our final list of GIN-activated genes.

Our study provides a global picture of the changes that occur in the transcriptome of target tissues, abomasal mucosa and abomasal lymph node, as a response to *T. circumcincta* infection in resistant and susceptible adult sheep. Whereas at day 7 post-infection, we did not find a differential response between the two compared groups in abomasal mucosa, a total of 106 genes were identified to show a distinct pattern between the two contrasted groups in the lymph node samples. The comparative study of our results with the available literature has shown remarkable coincidences for some of these genes with other gene expression studies related to GIN infection in lambs. Hence, the expression of genes such as *ITLN2*, *CLCA1*, *galectin 14*, etc. appear to be consistently affected by nematode infection, in both lambs and adult sheep. The differential expression of some immune-related genes reported in the present study as a response to *T. circumcincta* and that are not coincident with previous studies focused on lamb animals could indicate immune mechanisms that are specifically activated in adult animals (e.g. *PGC*, *SFTPD*, *TUBA4A*, *SST*, *BPIFB1*, *PF4*, *B4GALNT2*, *JCHAIN*, *AQP7*, *KLHL25*, *NEDD4*, *ANO6*). The RNA-Seq technology has shown here to be an appropriate platform to investigate the molecular mechanisms underlying the immune response to nematode infection in adult sheep. Future studies will focus on the genetic variability of the GIN-activated genes reported here with the aim of identifying potential candidate mutations that could be directly implemented in selection programs to increase GIN resistance in commercial sheep populations.

Additional files

Additional file 1. Gene expression levels in the two tissues and conditions studied. Distribution of gene expression levels for the genes expressed in the transcriptome of abomasal mucosa and abomasal lymph node samples for the two groups of animals compared in the present study.

Additional file 2. Venn diagrams showing the number of genes identified highly expressed genes (> 500 FPKM) by the analysis of the abomasal mucosa and abomasal lymph node transcriptomes. The gene expression levels were normalized by library size and gene length by calculating Fragments Per Kilobase of Exon Per Million Fragments Mapped (FPKM). A total number of 123 genes were identified as highly expressed in the two tissues, whereas 56 and 78 genes were highly expressed specifically in the abomasal mucosa and the abomasal lymph node samples respectively.

Additional file 3. Gene-set enrichment analysis (GO) for the highly expressed genes in both tissues studied. Significant terms from the Gene Ontology (GO) enrichment analysis performed for the genes identified as highly expressed genes (≥ 500 FPKM) in both tissues studied, abomasal mucosa and abomasal lymph node tissue.

Additional file 4. Gene-set enrichment analysis (GO) for the highly expressed genes in abomasal mucosa tissue. Significant terms from the Gene Ontology (GO) enrichment analysis performed with WebGestalt for the genes identified as highly expressed (≥ 500 FPKM) specifically in abomasal mucosa.

Additional file 5. Gene-set enrichment analysis (GO) for the highly expressed genes in lymph node tissue. Significant terms from the Gene Ontology (GO) enrichment analysis performed with WebGestalt for the genes identified as highly expressed (≥ 500 FPKM) specifically in lymph node.

Additional file 6. Heatmap plot of the lymph node transcriptome of six resistant and six susceptible adult ewes based on raw read counts. The heatmap plot of raw RNA read counts did not show a clear differentiation between the two groups of samples (Resistant vs Susceptible). But the most divergent clusters of two resistant (ALN7_R and ALN1_R) and two susceptible (ALN19_S and ALN9_S) samples can be observed.

Additional file 7. List of genes identified as significantly differentially expressed genes (DEGs) by the DESeq analysis for the abomasal mucosa samples. The log2FoldChange value obtained from the DESeq analysis is given for the genes identified as up-regulated in the resistant group and for the genes identified as up-regulated in the susceptible group.

Additional file 8. List of DEGs identified as GIN-activated genes for the abomasal lymph node samples (Resistant and Susceptible ewes to *T. circumcincta* infection). The log2FoldChange value obtained from the EdgeR and DESeq analyses are given for the genes identified as GIN-activated in the resistant group and for the genes identified as GIN-activated in the susceptible group.

Additional file 9.Significant terms from the Gene Ontology (GO) enrichment analysis performed with WebGestalt for the genes identified as GIN-activated in abomasal lymph node tissue. The results of the GO enrichment analysis are provided separately for the genes identified as up-regulated in the resistant and susceptible groups.

Additional file 10. Novel candidate genes for GIN resistance response presented in this work. List of DEGs identified as GIN-activated genes in the analysis of the abomasal lymph node samples that have not been previously reported as GIN related genes and that could be related to the immune mechanisms specifically activated in adult sheep. The log2FoldChange obtained in the analyses performed with EdgeR and DESeq are provided for each gene.

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
JJA and BGG conceived and designed the study. MMV designed and performed the experimental infections. ASV optimized the bioinformatics workflow and PKC performed the bioinformatics analyses. The manuscript was written by PKC and critically reviewed by JJA, BGG, MMV and ASV. All authors read and approved the final manuscript.

Acknowledgements
Financial support for this project was received from the LE248U14 project of Junta de Castilla and León Government. P. K. Chitneedi is funded by a predoctoral fellowship from the Junta de Castilla and León Government and the European Social Fund. B Gutiérrez-Gil is funded by the "Ramón y Cajal" Programme (RYC-2012-10230) from the Spanish Ministry of Economy, Industry and Competitiveness (MINECO). M. Martínez-Valladares is also funded by the "Ramón y Cajal" Programme (RYC-2015-18368) from MINECO.

Ethics approval and consent to participate
The whole experiment was carried out according to the current National Spanish legislation on the protection of animals used in experimentation (Royal Decree 53/2013) and after obtaining the positive report from the subcommittee for experimentation and animal welfare of the University of León (OEBA), and the approval of the competent body of the regional government, Junta de Castilla y León.

Author details
[1] Departamento de Producción Animal, Facultad de Veterinaria, Universidad de León, Campus de Vegazana s/n, 24071 León, Spain. [2] Departamento de Sanidad Animal, Facultad de Veterinaria, Universidad de León, Campus de Vegazana s/n, 24071 León, Spain. [3] Instituto de Ganadería de Montaña, CSIC-Universidad de León, 24346, Grulleros, León, Spain.

Publisher's Note

Received: 28 December 2017 Accepted: 28 February 2018
Published online: 27 April 2018

References
1. Riggio V, Pong-Wong R, Sallé G, Usai MG, Casu S, Moreno CR, Matika O, Bishop SC (2014) A joint analysis to identify loci underlying variation in nematode resistance in three European sheep populations. J Anim Breed Genet 131:426–436
2. Silva MVB, Sonstegard TS, Hanotte O, Mugambi JM, Garcia JF, Nagda S, Gibson JP, Iraqi FA, McClintock AE, Kemp SJ, Boettcher PJ, Malek M, Van Tassell CP, Baker RL (2012) Identification of quantitative trait loci affecting resistance to gastrointestinal parasites in a double backcross population of Red Maasai and Dorper sheep. Anim Genet 43:63–71
3. McRae KM, McEwan JC, Dodds KG, Gemmell NJ (2014) Signatures of selection in sheep bred for resistance or susceptibility to gastrointestinal nematodes. BMC Genomics 15:637
4. Riggio V, Matika O, Pong-Wong R, Stear MJ, Bishop SC (2013) Genome-wide association and regional heritability mapping to identify loci underlying variation in nematode resistance and body weight in Scottish Blackface lambs. Heredity 110:420–429
5. Sallé G, Jacquiet P, Gruner L, Cortet J, Sauvé C, Prévot F, Grisez C, Bergeaud JP, Schibler L, Tircazes A, François D, Pery C, Bouvier F, Thouly JC, Brunel JC, Legarra A, Elsen JM, Bouix J, Rupp R, Moreno CR (2012) A genome scan for QTL affecting resistance to *haemonchus contortus* in sheep. J Anim Sci 90:4690–4705
6. Gutiérrez-Gil B, Pérez J, Álvarez L, Martínez-Valladares M, de la Fuente LF, Bayón Y, Meana A, San Primitivo F, Rojo-Vázquez FA, Arranz JJ (2009) Quantitative trait loci for resistance to trichostrongylid infection in Spanish Churra sheep. Genet Sel Evol 41:46
7. Atlija M, Arranz J-J, Martinez-Valladares M, Gutiérrez-Gil B (2016) Detection and replication of QTL underlying resistance to gastrointestinal nematodes in adult sheep using the ovine 50K SNP array. Genet Sel Evol 48:4
8. Stear MJ, Strain S, Bishop SC (1999) Mechanisms underlying resistance to nematode infection. Int J Parasitol 29:51–56
9. French AT, Knight PA, Smith WD, Brown JK, Craig NM, Pate JA, Miller HR, Pemberton AD (2008) Up-regulation of intelectin in sheep after infection with *Teladorsagia circumcincta*. Int J Parasitol 38:467–475
10. French AT, Knight PA, David Smith W, Pate JA, Miller HR, Pemberton AD (2009) Expression of three intelectins in sheep and response to a Th2 environment. Vet Res 40:53
11. Ingham A, Reverter A, Windon R, Hunt P, Menzies M (2008) Gastrointestinal nematode challenge induces some conserved gene expression changes in the gut mucosa of genetically resistant sheep. Int J Parasitol 38:431–442
12. Rowe A, Gondro C, Emery D, Sangster N (2009) Sequential microarray to identify timing of molecular responses to *Haemonchus contortus* infection in sheep. Vet Parasitol 161:76–87
13. Andronicos N, Hunt P, Windon R (2010) Expression of genes in gastrointestinal and lymphatic tissues during parasite infection in sheep genetically resistant or susceptible to *Trichostrongylus colubriformis* and *Haemonchus contortus*. Int J Parasitol 40:417–429
14. McRae KM, Good B, Hanrahan JP, McCabe MS, Cormican P, Sweeney T, O'Connell MJ, Keane OM (2016) Transcriptional profiling of the ovine abomasal lymph node reveals a role for timing of the immune response in gastrointestinal nematode resistance. Vet Parasitol 224:96–108
15. Guo Z, González JF, Hernandez JN, McNeilly TN, Corripio-Miyar Y, Frew D, Morrison T, Yu P, Li RW (2016) Possible mechanisms of host resistance to *Haemonchus contortus* infection in sheep breeds native to the Canary Islands. Sci Rep 6:26200

16. Muñoz-Guzmán MA, Cuenca-Verde C, Valdivia-Anda G, Cuéllar-Ordaz JA, Alba-Hurtado F (2012) Differential immune response between fundic and pyloric abomasal regions upon experimental ovine infection with *Haemonchus contortus*. Vet Parasitol 185:175–180
17. Terefe G, Lacroux C, Andréoletti O, Grisez C, Prevot F, Bergeaud JP, Penicaud J, Rouillon V, Gruner L, Brunel JC, François D, Bouix J, Dorchies P, Jacquiet P (2007) Immune response to *Haemonchus contortus* infection in susceptible (INRA 401) and resistant (Barbados Black Belly) breeds of lambs. Parasite Immunol 29:415–424
18. Lacroux C, Nguyen THC, Andreoletti O, Prevot F, Grisez C, Bergeaud JP, Gruner L, Brunel JC, Francois D, Dorchies P, Jacquiet P (2006) *Haemonchus contortus* (Nematoda: Trichostrongylidae) infection in lambs elicits an unequivocal Th2 immune response. Vet Res 37:607–622
19. Pernthaner A, Cole SA, Morrison L, Hein WR (2005) Increased expression of interleukin-5 (IL-5), IL-13, and tumor necrosis factor alpha genes in intestinal lymph cells of sheep selected for enhanced resistance to nematodes during infection with *Trichostrongylus colubriformis*. Infect Immun 73:2175–2183
20. Gossner A, Wilkie H, Joshi A, Hopkins J (2013) Exploring the abomasal lymph node transcriptome for genes associated with resistance to the sheep nematode *Teladorsagia circumcincta*. Vet Res 44:68
21. Costa V, Aprile M, Esposito R, Ciccodicola A (2013) RNA-Seq and human complex diseases: recent accomplishments and future perspectives. Eur J Hum Genet 21:134–142
22. Bioinformatics B (2011) FastQC A quality control tool for high throughput sequence data. UK Babraham Inst, Cambridge
23. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR (2013) STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29:15–21
24. Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R (2009) The Sequence alignment/map (SAM) format and SAMtools. Bioinformatics 25:2078–2079
25. Anders S, Pyl PT, Huber W (2015) HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics 31:166–169
26. Li B, Dewey CN (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinform 12:323
27. Robinson MD, McCarthy DJ, Smyth GK (2009) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26:139–140
28. Anders S, Huber W (2010) Differential expression analysis for sequence count data. Genome Biol 11:R106
29. Anders S, McCarthy DJ, Chen Y, Okoniewski M, Smyth GK, Huber W, Robinson MD (2013) Count-based differential expression analysis of RNA sequencing data using R and Bioconductor. Nat Protoc 8:1765–1786
30. Wang J, Duncan D, Shi Z, Zhang B (2013) WEB-based gene set analysis toolkit (WebGestalt): update 2013. Nucleic Acids Res 41:W77–W83
31. Martínez-Valladares M, Vara-Del Río MP, Cruz-Rojo MA, Rojo-Vázquez FA (2005) Genetic resistance to *Teladorsagia circumcincta*: IgA and parameters at slaughter in Churra sheep. Parasite Immunol 27:213–218
32. Strain SA, Bishop SC, Henderson NG, Kerr A, McKellar QA, Mitchell S, Stear MJ (2002) The genetic control of IgA activity against *Teladorsagia circumcincta* and its association with parasite resistance in naturally infected sheep. Parasitology 124:545–552
33. Balic A, Bowles VM, Liu YS, Meeusen ENT (2003) Local Immune responses in sensitized sheep following challenge infection with *Teladorsagia circumcincta*. Parasite Immunol 25:375–381
34. Henderson NG, Stear MJ (2006) Eosinophil and IgA responses in sheep infected with *Teladorsagia circumcincta*. Vet Immunol Immunopathol 112:62–66
35. Keane O, Zadissa A, Wilson T, Hyndman DL, Greer GJ, Baird DB, McCulloch AF, Crawford AM, McEwan JC (2006) Gene expression profiling of naive sheep genetically resistant and susceptible to gastrointestinal nematodes. BMC Genomics 7:42
36. Craig NM, Miller HRP, Smith WD, Knight PA (2007) Cytokine expression in naïve and previously infected lambs after challenge with *Teladorsagia circumcincta*. Vet Immunol Immunopathol 120:47–54
37. Knight PA, Griffith SE, Pemberton AD, Pate JM, Guarneri L, Anderson K, Talbot RT, Smith S, Waddington D, Fell M, Archibald AL, Burgess STG, Smith DW, Miller HRP, Morrison IW (2011) Novel gene expression responses in the ovine abomasal mucosa to infection with the gastric nematode *Teladorsagia circumcincta*. Vet Res 49:78
38. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B 57:289–300
39. Zhang ZH, Jhaveri DJ, Marshall VM, Bauer DC, Edson J, Narayanan RK, Robinson GJ, Lundberg AE, Bartlett PF, Wray NR, Zhao QY (2014) A comparative study of techniques for differential expression analysis on RNA-seq data. PLoS One 9:e103207
40. Chen T, Tibbitt CA, Feng X, Stark JM, Rohrbeck L, Rausch L, Sedimbi SK, Karlsson MC, Lambrecht BN, Hedestam GB, Hendriks RW, Chambers BJ, Nylén S, Coquet JM (2017) PPAR-γ promotes type 2 immune responses in allergy and nematode infection. Sci Immunol 2:eaal5196
41. Martin H (2010) Role of PPAR-gamma in inflammation. Prospects for therapeutic intervention by food components. Mutat Res 690:57–63
42. Pease JE, Williams TJ (2001) Eotaxin and asthma. Curr Opin Pharmacol 1:248–253
43. Woolaston RR, Manueli P, Eady SJ, Barger IA, Le Jambre LF, Banks DJ, Windon RG (1996) The value of circulating eosinophil count as a selection criterion for resistance of sheep to trichostrongyle parasites. Int J Parasitol 26:123–126
44. Athanasiadou S, Pemberton A, Jackson F, Inglis N, Miller HR, Thévenod F, Mackellar A, Huntley JF (2008) Proteomic approach to identify candidate effector molecules during the in vitro immune exclusion of infective *Teladorsagia circumcincta* in the abomasum of sheep. Vet Res 39:58
45. Erickson NA, Mundhenk L, Giovannini S, Glauben R, Heimesaat MM, Gruber AD (2016) Role of goblet cell protein CLCA1 in murine DSS colitis. J Inflamm 13:5
46. Datta R, DeSchoolmeester ML, Hedeler C, Paton NW, Brass AM, Else KJ (2005) Identification of novel genes in intestinal tissue that are regulated after infection with an intestinal nematode parasite. Infect Immun 73:4025–4033
47. Li RW, Rinaldi M, Capuco AV (2011) Characterization of the abomasal transcriptome for mechanisms of resistance to gastrointestinal nematodes in cattle. Vet Res 42:114
48. Yi L, Cheng D, Zhang K, Huo X, Mo Y, Shi H, Di H, Zou Y, Zhang H, Zhao J, Xu Y, Erle DJ, Zhen G (2017) Intelectin contributes to allergen-induced IL-25, IL-33, and TSLP expression and type 2 response in asthma and atopic dermatitis. Mucosal Immunol 10:1491–1503
49. Vasta GR, Quesenberry M, Ahmed H, O'Leary N (1999) C-type lectins and galectins mediate innate and adaptive immune functions: their roles in the complement activation pathway. Dev Comp Immunol 23:401–420
50. Dunphy JL, Barcham GJ, Bischof RJ, Young AR, Nash A, Meeusen EN (2002) Isolation and characterization of a novel eosinophil-specific galectin released into the lungs in response to allergen challenge. J Biol Chem 277:14916–14924
51. Vasta GR (2009) Roles of galectins in infection. Nat Rev Microbiol 7:424–438
52. Manicone AM, McGuire JK (2008) Matrix metalloproteinases as modulators of inflammation. Semin Cell Dev Biol 19:34–41
53. McCann MJ, Johnston S, Reilly K, Men X, Burgess EJ, Perry NB, Roy NC (2014) The effect of turmeric (*Curcuma longa*) extract on the functionality of the solute carrier protein 22 A4 (SLC22A4) and interleukin-10 (IL-10) variants associated with inflammatory bowel disease. Nutrients 6:4178–4190

Study 7

Variant discovery in genes identified as differentially expressed genes between the abomasal lymph node transcriptome of resistant and susceptible adult sheep to *Teladorsagia circumcincta* infection

Praveen Krishna Chitneedi[1], Aroa Suárez-Vega[1]., María Martinez-Valladares[2,3], Juan-Jose Arranz[1] and Beatriz Gutiérrez-Gil[1]

[1]Departamento de Producción Animal, Facultad de Veterinaria, Universidad de León, León, Spain; [2]Departamento de Sanidad Animal, Facultad de Veterinaria, Universidad de León, León, Spain; [3]Instituto de Ganadería de Montaña, CSIC-ULE, León, Spain.

XIX Reunión Nacional de Mejora Genetica Animal, 14-15 de junio, 2018, León. https://buleria.unileon.es/handle/10612/8285 (19th National Animal Breeding Meeting, 14th -15th of June 2018, León, Spain.)

VARIANT DISCOVERY IN GENES IDENTIFIED AS DIFFERENTIALLY EXPRESSED GENES BETWEEN THE ABOMASAL LYMPH NODE TRANSCRIPTOME OF RESISTANT AND SUSCEPTIBLE ADULT SHEEP TO *TELADORSAGIA CIRCUMCINCTA* INFECTION

Chitneedi[1], P.K., Suárez-Vega[1], A., Martínez Valladares[2,3], M., Arranz[1] J.J. y Gutiérrez-Gil[1], B.

[1]Departamento de Producción Animal, Facultad de Veterinaria, Universidad de León, 24071 León. [2]Departamento de Sanidad Animal, Facultad de Veterinaria, Universidad de León, Campus de Vegazana s/n, 24071 León, Spain. [3] Instituto de Ganadería de Montaña, CSIC-Universidad de León, 24346, Grulleros, León, Spain pchi@unileon.es

INTRODUCTION

Gastrointestinal nematode infections are one of the major health issues facing grazing sheep populations and it incurs on major economic losses for sheep breeders. The resistance/susceptibility trait appears to be a highly complex trait (Behnke et al. 2003; Dominik 2005). In sheep resistance to nematode infection shows a moderate level of heritability (range 0.3-0.6) (Stear et al. 2001). Several QTL mapping studies have tried to identify genomic regions and mutations that influence resistance to nematode infection (Atlija et al. 2016; Coltman et al. 2001; Gutiérrez-Gil et al. 2009; Sayers et al. 2005), although the detection of causal mutations for this trait is still a challenge for the research community. The recently available RNA-seq technology provides the opportunity to extract high-throughput transcriptome data from a specific tissue to perform gene quantification, differential gene expression and detection of variants (SNPs and indels), which could be assessed as potential causal mutations (Hudson, Dalrymple, and Reverter 2012). A previous study of our research group has identified a list of 106 differential expression genes (DEGs) based on RNA-Seq dataset obtained from the abomasal lymph nodes of 12 adult sheep, previously classified as resistant or susceptible to GIN infection based on an artificial infection with *T. circumcincta* larvae Chitneedi et al. (2018). In the present study we present a detailed study of the variants mapping within the list of DEGs previously reported in that study. Thus, the present study provides a list of functionally relevant variants that could underlie the genetic control of resistance/susceptibility to *T. circumcincta* in adult sheep.

MATERIALS AND METHODS

Experimental infection: The animals included in the present study were 12 adult ewes of Churra sheep from a commercial flock of Churra dairy breed reared under a semi-intensive management and belonging to the National Association of Spanish Churra sheep breeders (ANCHE). These animals were subjected to two experimental infections with *T. circumcincta* larvae, as described in detail by Chitneedi et al. (2018).

RNA sequencing: From the six resistant and the six susceptible sheep included in this study, mRNA was extracted from abomasal lymph nodes, using the Absolutely RNA miRNA Kit from Agilent (La Jolla, CA, USA). RNA integrity (RIN value) was analyzed using the Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). The average RIN values of the RNA samples was 7.2 ranging between 6.7 and 7.5. After library preparation, the subsequent sequencing was performed with an Illumina HiSeq sequencer 2000, generating stranded paired-end reads of 75 base pairs with a depth of 30 million reads.

Bioinformatics analysis: The read quality of the 12 abomasal lymph node samples was determined using FastQC_v0.11.5 (Bioinformatics 2011). All the samples were aligned against the ovine genome assembly v.3.1 (Oar_v3.1) using the STAR_v2.5.3a aligner (Dobin et al. 2013). After alignment, editing read groups and marking duplicate reads were carried out with Picard_v2.9.0 (Broad Institute 2016). Then, the pre-processing of aligned reads, including read trimming, realignment and recalibration was performed with GATK_v3.7 (Van der Auwera et al. 2013). The variant calling analysis was individually performed using two different software, GATK_v3.7 and SAMTOOLS_v1.4 (Wysoker et al. 2009). Later, quality filters were applied independently to each of the resulting VCF files. The variants detected by Samtools were filtered with SnpSift (QUAL > 30) whereas for the GATK-detected variants the GATK specific filtering recommendations were followed (DP > 5 & QUAL > 30 & MQ > 40.0 & QD > 5.0 & FS < 60.0). After that, variants commonly identified by the two independent software were extracted with BCFtools_v1.4 ("isec" option) (Wysoker et al. 2009) and were considered as high-quality variants.

With the aim of characterizing transcriptome variants that may underlie sheep resistance to GIN, we used SnpEff ("–fi" option) followed by a bed file with the coordinates of the 106 DEGs reported by Chitneedi et al. (2018) to select the variants included in the studied genes. The variants extracted from the DEG coordinates were individually annotated for functional consequences with the software VEP_v90 (McLaren et al. 2016) and SnpEff_v4.3 (Cingolani et al. 2012). Later, we selected, for each subset, those predicted by the two annotation analyses to have relevant functional consequences (classified as HIGH or MODERATE impact). The selected variants were annotated with the online web tool VEP (https://www.ensembl.org/Tools/VEP) and the amino acid substitution effects on protein function were predicted using the SIFT algorithm (Kumar, Henikoff, and Ng 2009).

RESULTS AND DISCUSSION

The pipeline followed for variant discovery with the transcriptome data of abomasal lymph node is shown in Figure 1. Based on the FastQC estimates, all the 12 samples analyzed were of high quality; thus, no trimming was performed. After aligning against the ovine reference genome (Oar_v3.1) around 80.27% paired reads per sample were uniquely mapped against the reference genome and around 8.5 % paired reads were mapped to multiple loci. After performing the variant calling and variant filtering, we found 1,326,960 common variants considered as high quality variants across the whole genome. From these high-quality variants, we extracted 6,168 variants (6,104 SNPs, 30 insertions and 34 deletions) mapping within the 106 DEGs. Among these, 332 variants (329 SNPs, 3 insertions and 2 deletions) were predicted to be of relevant functional impact ("High" and "Moderate"). The functional annotation of these variants with the VEP, which included 109 novel variants revealed a total of 471 functional consequences distributed across 60 genes. From these, the SIFT algorithm detected 50 deleterious variants most were missense consequence and two were splice region variants. These deleterious variants were included in a total of 15 genes. Four of these genes, *BPIFB1* (Zhou et al. 2017), *KRT20* (Sen et al. 2012), *SLC38A2* (Carter 2012) and FNDC1 (Sigdel et al. 2015) have been found to be associated with the immune response. Another gene harboring a missense deleterious variant was *MMP28*, which has been reported as responsible for cell proliferation in response to skin injury (Saarialho-Kere et al. 2002). Other genes harboring "High" impact variants were *LGALS4* (splice acceptor variant), *SLC38A2* (stop lost), *ASIC3* (stop gained and splice acceptor variant), and *SULF1* (frameshift variant). Future studies focusing on the

variations in these gene regions and pathway analysis combining these genes may provide information about potential causal mutations related to resistance against GIN infection in sheep. In addition, the deleterious variants which are located in the non-annotated gene regions were equally important for further investigation.

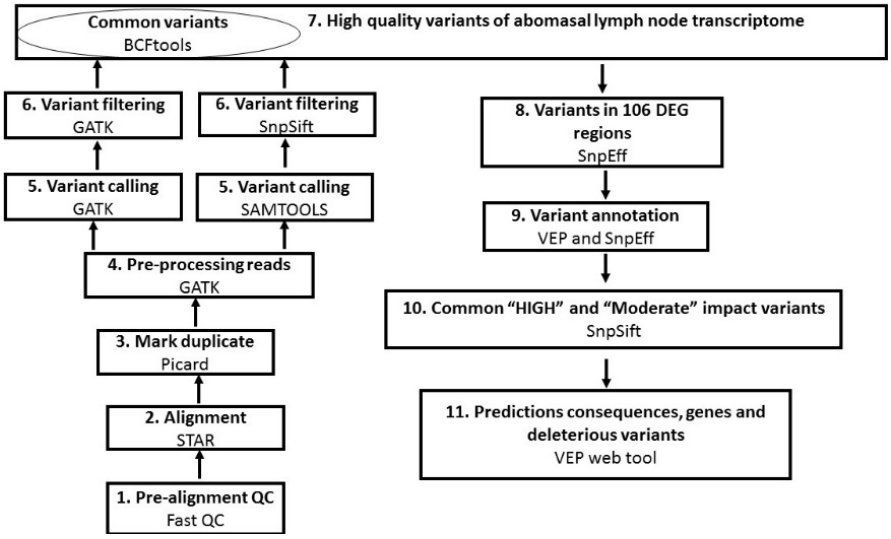


Figure 1. *The pipeline shows the steps followed and tools used to detect variants and the consequences related to nematode infection in sheep found in DEG regions.*

REFERENCES

•Atlija, M, et al. 2016. Genet. Sel. Evol 48: 4. • Van der A et al. 2013. Cur Prot in Bioinfo (SUPL.43). • Behnke, J M et al. 2003. J. helmi 77: 99–110. • Bioinformatics, Babraham. 2011. UK: Babraham Institute. • Broad Institute. 2016. Picard Tools. • Carter, A. M. 2012. Phys Rev 92: 1543–76. • Chitneedi, P K et al. 2018. Vet. Res 49: 39. • Churko et al. 2013. Circu. Res 112: 1613–23. • Cingolani, P et al. 2012. Fly 6: 80–92. • Coltman, D.W. et al. 2001. Paras 122: 571–82. • Dobin, Alexander et al. 2013. Bioinfo 29: 15–21. • Dominik, S. 2005. Genet. Sel. Evol 37 Sup 1: S83–96. • Gutiérrez-Gil, B et al. 2009. Genet. Sel. Evol 41: 46. • Hudson, N J. et al. 2012. BMC Geno 13. • Kumar, P. et al. 2009. N. Prot 4: 1073–82. • McLaren, W. et al. 2016. G. Biology 17. • Saarialho-Kere, U et al. 2002. J. I. Derm 119: 14–21. • Sayers, G. et al. 2005. Paras 131: 403–9. • Sen, A. et al. 2012. P. N. A. Sci 109: 20667–72. • Sigdel, K R et al. 2015. J. Imm. Res 2015. • Stear, M. J. et al. 2001. R Vet Sci 71: 1–7. • Wysoker, A et al. 2009 Bioinfo 25: 2078–79. • Zhou et al. 2017. Open Medicine (Poland) 12: 299–307. •

Acknowledgements: Financial support for this project was received from the LE248U14 project of Junta de Castilla and León Government. P. K. Chitneedi is funded by a predoctoral fellowship from the Junta de Castilla and León Government and the European Social Fund. B Gutiérrez-Gil is funded by the "Ramón y Cajal" Programme (RYC-2012-10230) from the Spanish Ministry of Economy, Industry and Competitiveness (MINECO). M. Martínez-Valladares is also funded by the "Ramón y Cajal" Programme (RYC-2015-18368) from MINECO.

Study 8

ANIMAL GENETICS Immunogenetics, Molecular Genetics and Functional Genomics

Identification of potential functional variants underlying ovine resistance to gastrointestinal nematode infection by using RNA-Seq

Praveen Krishna Chitneedi[1], Juan-Jose Arranz[1], Aroa Suárez-Vega[1], María Martinez-Valladares[2,3] and Beatriz Gutiérrez-Gil[1]

[1]Departamento de Producción Animal, Facultad de Veterinaria, Universidad de León, León, Spain; [2]Departamento de Sanidad Animal, Facultad de Veterinaria, Universidad de León, León, Spain; [3]Instituto de Ganadería de Montaña, CSIC-ULE, León, Spain.

Animal genetics DOI: doi.org/10.1111/age.12894

Licence no: 4753021165438

ANIMAL GENETICS Immunogenetics, Molecular Genetics and Functional Genomics

FULL PAPER

doi: 10.1111/age.12894

Identification of potential functional variants underlying ovine resistance to gastrointestinal nematode infection by using RNA-Seq

P. K. Chitneedi*, J. J. Arranz*, A. Suárez-Vega*, M. Martínez-Valladares[†,‡] and B. Gutiérrez-Gil*

*Departamento de Producción Animal, Facultad de Veterinaria, Universidad de León, Campus de Vegazana s/n, 24071, León, Spain. [†]Departamento de Sanidad Animal, Facultad de Veterinaria, Universidad de León, Campus de Vegazana s/n, 24071, León, Spain. [‡]Instituto de Ganadería de Montaña, CSIC-ULE, 24346, Grulleros, León, Spain.

Summary

In dairy sheep flocks from Mediterranean countries, replacement and adult ewes are the animals most affected by gastrointestinal nematode (GIN) infections. In this study, we have exploited the information derived from an RNA-Seq experiment with the aim of identifying potential causal mutations related to GIN resistance in sheep. Considering the RNA-Seq samples from 12 ewes previously classified as six resistant and six susceptible animals to experimental infection by *Teladorsagia circumcincta*, we performed a variant calling analysis pipeline using two different types of software, GATK version 3.7 and SAMTOOLS version 1.4. The variants commonly identified by the two packages (high-quality variants) within two types of target regions – (i) QTL regions previously reported in sheep for parasite resistance based on SNP-chip or sequencing technology studies and (ii) functional candidate genes selected from gene expression studies related to GIN resistance in sheep – were further characterised to identify mutations with a potential functional impact. Among the genes harbouring these potential functional variants (930 and 553 respectively for the two types of regions), we identified 111 immune-related genes in the QTL regions and 132 immune-related genes from the initially selected candidate genes. For these immune-related genes harbouring potential functional variants, the enrichment analyses performed highlighted significant GO terms related to apoptosis, adhesion and inflammatory response, in relation to the QTL related variants, and significant disease-related terms such as inflammation, adhesion and necrosis, in relation to the initial candidate gene list. Overall, the study provides a valuable list of potential causal mutations that could be considered as candidate causal mutations in relation to GIN resistance in sheep. Future studies should assess the role of these suggested mutations with the aim of identifying genetic markers that could be directly implemented in sheep breeding programmes considering not only production traits, but also functional traits such as resistance to GIN infections.

Keywords abomasal lymph node, adult sheep, genetic variant, parasite resistance, RNA-Seq, transcriptome

Introduction

Gastrointestinal nematode (GIN) infections are one of the major health issues of grazing sheep populations and incur major economic losses for sheep breeders. The host genetic variation underlying resistance/susceptibility to nematode infections is a topic of interest, as the resistance/susceptibility trait appears to be a highly complex trait related to the different immune responses that are activated in sheep, depending on multiple factors such as the breed, age and immune status of the host (Behnke *et al.* 2003; Dominik 2005). Additionally, the nematode species responsible for the infection is a factor that explains variation in the activated immune response (Dominik 2005). Although the reported estimates show a wide range of variation, different studies have shown that sheep resistance to nematode infection, generally estimated for the faecal egg count (FEC) as an indicator trait, is a moderately heritable trait

Address for Correspondence

Beatriz Gutiérrez-Gil, Departamento de Producción Animal, Facultad de Veterinaria, Universidad de León, Campus de Vegazana s/n, 24071 León, Spain. E-mail: beatriz.gutierrez@unileon.es

Accepted for publication 28 November 2019

(h^2 range 0.1–0.3) and responds to selection (Gutiérrez-Gil *et al.* 2010; Mirkena *et al.* 2010; Mpetile *et al.* 2015; Brown & Fogarty 2017; Ngere *et al.* 2018). Hence, several sheep studies have tried to identify genomic regions, QTL, and mutations that influence resistance to nematode infection. Initially, these studies were based on sparse maps of microsatellites and the use of classical linkage analysis (LA) on family-structured populations (Coltman *et al.* 2001; Beraldi *et al.* 2007; Gutiérrez-Gil *et al.* 2009). In the last years, the higher marker density offered by SNP arrays, including medium- and high-density chips, allows substantial improvement of mapping accuracy and redefinition of shorter confidence intervals, in many cases by exploiting LD population information by implementing GWASs (e.g. Al Kalaldeh et al. 2019a; Atlija *et al.* 2016; Benavides *et al.* 2015). The information from those analysis can be complementary to that provided by the analysis of the SNP-chip information with other classical methods such as LA, LD combined with LA (Atlija *et al.,* 2016), selection sweep mapping (McRae *et al.* 2014) or regional heritability mapping (Riggio *et al.* 2013).

Several of these studies have used FEC as an indicator trait of resistance (Coltman *et al.* 2001; Gutiérrez-Gil *et al.* 2009; McRae *et al.* 2014; Atlija *et al.* 2016). For this trait the SheepQTL database (Hu *et al.* 2015) currently includes a total of 124 classical QTL (based on microsatellite marker studies) and 48 associations (based on the analysis of dense SNP-chip datasets). The physiological complexity of disease resistance, and the large diversity of the experimental designs of these studies (addressing different sheep breeds, parasite species, type of infections – natural or experimental), may explain the limited consistency between them and their lack of power to pinpoint causal mutations with precision. The type of immune response determined by the age of the host, innate or acquired (Grencis *et al.* 2014), may also explain the lack of consistency between some of these studies. In this regard, whereas many of the mapping studies for GIN resistance have studied young animals, our research group has focused efforts in this field on adult dairy ewes (Gutierrez-Gil et al. 2009; Atlija *et al.* 2016). The reason for this is because in the Spanish northwest region of Castilla y León the production system of dairy sheep flocks is based on the milk from adult ewes and the sale of suckling lambs fed exclusively on maternal milk. Hence, in this production system replacement ewes and adult sheep are the animals subjected to the direct effects of helminth infections (García-Pérez et al. 2002). The severe economic losses that these infections may cause in these flocks (Wolstenholme *et al.* 2004) highlight the need to undertake specific studies, within the field of the genetic basis of GIN resistance in sheep, in adult animals.

Over the last decade, the availability of high-throughput sequencing technology has created large datasets of high-resolution genomic data to identify novel variants and mutations related to different complex traits (Churko *et al.* 2013). The analysis of whole genome/transcriptome sequencing datasets may provide valuable information for QTL characterisation as the complete genetic variability (at the DNA sequence level, or the transcriptomic level) within a target region can be interrogated for concordance with the QTL genotypes, increasing the chances of detecting the causal underlying mutation or markers in strong LD with the causative mutation (Rupp *et al.* 2015). In this sense, a recent study related to parasite resistance in sheep has extended, by exploiting genotype imputation, the information from SNP arrays towards the WGS density (Al Kalaldeh et al. 2019b). In the case of RNA-Seq datasets, in addition to the classical gene quantification and differential gene expression analyses, they can also be analysed for the detection of variants (SNPs and indels) and search for potential causal mutations at a lower cost than that of WGS (Wang *et al.* 2009).

In this study, we have exploited the information derived from an RNA-Seq experiment with the aim of identifying potential causal mutations related to GIN resistance in sheep. For that, after the general characterisation of the transcriptomic variants of the abomasal lymph node of 12 adult Churra sheep after an experimental infection with *Teladorsagia circumcincta* (Chitneedi *et al.* 2018), we performed a detailed analysis of the variants identified in critical genome regions, related to QTL and candidate genes for parasite resistance. Hence, we provide here a list of potential functional variants within immune-related genes that could be considered, e.g. in future confirmation studies, as candidate causative mutations to underlie the genetic control of resistance/susceptibility to *T. circumcincta* in adult sheep.

Material and methods

Animals, experimental infections and RNA sequencing

The animals included in the present study were 12 adult ewes of Churra sheep and the experiment and criteria for their selection have been previously described in detail by Chitneedi *et al.* (2018). Briefly, a commercial flock of Churra dairy breed reared under a semi-intensive system and naturally infected by GIN was sampled for the analysis of FEC (global flock phenotypic values: average = 25.59 eggs per gram, epg; SD = 60.31). FEC measurements were determined by the modified McMaster method (MAFF 1986). The detection limit for this technique was 15 epg. This flock was selected from a group of six tested flocks because, owing to a more intensive management system, the other flocks did not show evidence of natural GIN infection (average FEC values in the rest of flocks were close to zero). From the selected flock we selected two groups of ewes (age range 4–10 years old) showing extreme phenotypic FEC values: a group of 11 animals showing low FEC values (range 0–15 epg) and a group of eight animals

showing high FEC values (range 90–225 epg). These animals were translated to the experimental farm of the Mountain Livestock Institute, León where, after an anti-helmintic treatment (oral dose of ivermectin), they were subjected to a first experimental infection (EI1) with one dose of 50 000 *T. circumcincta* third-stage larvae (L3). After the EI1, collection of faeces was performed every 2 days, starting from days 14–31 post-infection to calculate the accumulated FEC from each animal according to Vercruysse *et al.* (1993). Based on these values, six sheep (age range 4–10 years; mean age 7.6) were classified as susceptible (range 2310–9666 epg; average 5594 ± 2661 epg) and six other sheep (age range 4–10 years; mean age 7) were classified as resistant (range 0–915 epg; average 308 ± 338 epg) to *T. circumcincta* experimental infection. One month after the EI1, the 12 ewes under study were subjected again to an anthelmintic treatment (subcutaneous injection of moxidectin) and 3 weeks later were exposed to a second experimental infection (EI2) with one dose of 70 000 *T. circumcincta* L3 (see Chitneedi *et al.* 2018 for further details). Note that during the whole study all sheep were maintained indoors so they were not exposed to natural infection with other GIN species. Also, the efficacy of the anthelmintic treatments was confirmed before both experimental infections by coprological analysis at 2 weeks post-treatment. At day 7 after EI2, the animals were sacrificed and, at necropsy, abomasal mucosa and lymph node samples for all animals were immediately collected in RNAlater™ Stabilization Reagent (Sigma-Aldrich, St Louis, MO, USA). The Absolutely RNA miRNA Kit from Agilent (La Jolla, CA, USA) was used for mRNA extraction for the two tissues and the assessment of the RNA integrity with the Agilent 2100 bioanalyser (Agilent Technologies, Santa Clara, CA, USA) showed RIN values ranging between 6.7 and 7.5. The preparation of the libraries and subsequent sequencing were performed with an Illumina HiSeq sequencer 2000, generating stranded paired-end reads of 75 bp with a depth of 30 million reads. Because the previous work on the dataset here analysed did not identify any confirmed differentially expressed genes (DEGs) in the abomasal mucosa samples (after combining the results of the analyses performed with two different software) (Chitneedi *et al.* 2018), the present variant calling study is focused on the analysis of the lymph node RNA-Seq dataset.

Bioinformatics analysis workflow

The workflow from quality control of raw samples to extract genes harbouring potential variants is shown in Fig. S1. The read quality of the raw dataset was determined using FastQC version 0.11 (Andrews 2010), and the samples were aligned against the Ovine Genome Assembly version 3.1 (Oar_v3.1) (Ensembl Database 2012) using star version 2.5.3a (Dobin *et al.* 2013). Later, the marking of duplicate reads was carried out with Picard version 2.9.0 (Broad Institute 2018) and the pre-processing of the aligned reads with gatk version 3.7 (Van der Auwera *et al.* 2013). The variant calling analysis was individually performed using two different types of software, gatk version 3.7 and Samtools version 1.4 (Wysoker *et al.* 2009). Later, quality filters were applied independently to each of the resulting variant call format files. The variants detected by Samtools were filtered with SnpSift (QUAL > 30) (Cingolani *et al.* 2012), whereas for the other dataset, the gatk filtering tool was used following the gatk-specific recommendations (DP > 5, QUAL > 30, MQ > 40.0, QD > 5.0 and FS < 60.0). After that, the variants commonly identified by the two independent software were extracted with BCFtools version 1.4 ('isec') (Wysoker *et al.* 2009) and were considered as high-quality variants. All of these variants were annotated for functional consequences with the vep version 90 (McLaren *et al.* 2016) software using the default analysis option which uses the annotation information (Ensembl transcripts) provided by the reference genome (Ovis_aries version 3.1). For the input variants, vep returned detailed annotation for effects on transcripts, proteins and regulatory regions (McLaren *et al.* 2016), including the predicted effects of coding non-synonymous variants on protein function obtained through the sift algorithm integrated as a tool in the vep annotation (Ng & Henikoff 2003).

Considering the results of the whole transcriptome annotation, and with the aim of identifying variants that could underlie sheep resistance to *T. circumcincta*, we performed the following filtering steps. Firstly, we used SnpEff (*–fi* option) (Cingolani *et al.* 2012) to extract the variants included in a selection of two types of previously reported target regions: (i) QTL for the FEC trait; and (ii) expression-related candidate genes for GIN resistance in sheep. Owing to the limited mapping accuracy of genome scans based on microsatellite markers, the selection of reported QTL related to parasite resistance was focused on studies based on the use of the SNP-Chip and next-generation sequencing technologies. After a deep review of the recent literature in the field, a total of 230 significant genomic regions reported by a list 11 different studies (including significant SNP associations detected by GWAS, or regions detected by LA, LD combined with LA, selection sweep mapping or regional heritability mapping), were considered for extraction of genetic variability from our RNA-Seq samples. A detailed characterisation of these regions, all of them referred to from now on as QTL, is provided in Table S1. Note that for the single point significant associations, reported in general by GWAS analyses, the target interval considered in this study was defined as a 250 kb interval centred on the originally significant SNP marker (following the criteria previously applied by Atlija *et al.* 2016); for the rest of regions the genomic interval provided in the original publication was considered here (Table S1). In relation to the second type of target regions considered for extraction of genetic

variability, we selected here a list of 1892 genes previously reported as highly expressed or DEGs after GIN infection in sheep by 12 different studies in the field. These studies include our DE analysis of the RNA-Seq dataset analysed here, and 11 other gene expression studies whose experimental design (age; breed; infection status: tissue used for study; nematode species; analysis technique) is summarised in Table 1. The final list of 1892 unique genes considered as target regions for variant extraction, including their corresponding coordinates, is presented in Table S2. These genes will be referred to from now on as candidate GIN-activated genes.

Secondly, within the considered target regions, we selected for further consideration as potential functional variants all of the variants classified by VEP as 'high impact' variants (those inferred to have a disruptive impact on the protein, probably causing protein truncation or loss of function or triggering nonsense mediated decay) and, among the 'moderate impact' variants (inferred to have a non-disruptive variant that might change protein effectiveness), we also selected those missense variants that were predicted as 'deleterious' by the SIFT tool (note that the variants predicted as 'deleterious low confidence' were not extracted). Considering that the final aim of this study was the characterisation of the transcriptome variants that may be of special interest in relation to parasite resistance, the list of genes harbouring the variants with relevant functional consequences identified across the whole transcriptome was compared with a list of 5029 genes known to be related to the immune response, based on the public databases IRIS (Kelley *et al.* 2005) and ImmPort (Bhattacharya *et al.* 2014), both available at the Innate database (Breuer *et al.* 2013). Finally, to provide a global picture of the biological functions of the unique immune-related genes

Table 1 List of studies related to the study of gene expression related to gastrointestinal nematode (GIN) infection in sheep from which the candidate GIN-activated genes considered in the present study were selected.

Animal age/breed	Infection status	Tissue used for study	Nematode species	No. of genes[1]	Technique	Study
Yearling sheep	Experimental infection (immune day 5 vs naive day 5 and naive day 5 vs naive day 0)	Abomasal mucosa	*Teladorsagia circumcincta*	32	cDNA microarray, RT-PCR, QT-PCR	Knight *et al.* 2011
6-Month-old Merino-cross wethers	Experimental infection	Abomasal mucosa	*Haemonchus contortus*	72	Sequential microarray	Rowe *et al.* 2009
Lambs	Experimental infection (primary challenge vs tertiary challenge)	Lymph node	*Trichostrongylus colubriformis*	55	Microarray data, QT PCR	Andronicos *et al.* 2010
Lambs	Experimental infection (resistant sheep)	Abomasal mucosa	*Trichostrongylus colubriformis, Haemonchus contortus*	63	RT-PCR	Ingham *et al.* 2008
Adult sheep	Natural infection (immune sheep)	Abomasal mucosa	*Teladorsagia circumcincta*	14	SDS-PAGE and Shotgun proteomics	Athanasiadou *et al.* 2008
Lambs, Scottish Blackface	Experimental infection (susceptible sheep)	Lymph node	*Teladorsagia circumcincta*	250	RNA-seq	McRae *et al.* 2016
1-Year-old Canaria Hair Breed (CHB) and Canaria Sheep (CS)	Experimental infection (infected CHB and CS)	Abomasal mucosa	*Haemonchus contortus*	555	RNA-seq, RT-PCR	Guo *et al.* 2016
Lambs, Scottish Blackface	Experimental infection (resistant vs control)	Lymph node	*Teladorsagia circumcincta*	341	cDNA microarray, RT-PCR	Gossner *et al.* 2013
Lambs, Scottish Blackface × Leicester	Experimental infection (infected sheep)	Abomasum, lymph node	*Teladorsagia circumcincta*	32	Semi-quantitative RT-PCR	French *et al.* 2009
Lambs, Merino Sheep	Experimental infection (resistance vs susceptible)	Abomasal tissues	*Haemonchus contortus*	596	RNA-seq	Zhang *et al.* 2019
1-Month-old lambs	Experimental infection (*ex vivo* infection)	Ovine abomasal explants	*Haemonchus contortus*	193	RNA-seq, RT-PCR	El-Ashram *et al.* 2018
Adult Churra Sheep	Experimental infection (resistance vs susceptible)	Abomasal lymph node	*Teladorsagia circumcincta*	106	RNA-seq	Chitneedi *et al.* 2018

[1]The total number of genes selected from each study.

harbouring the mutations with a potential biological impact identified within the target QTL, we performed a GO enrichment analysis with the web-based software WebGestalt (Wang *et al.* 2013). This software was used also to perform a disease association enrichment analysis with the candidate GIN-activated genes included in the IRIS and ImmPort databases and that harboured mutations with a potential biological impact. For these enrichment analyses, we used the human genome as reference and the following parameters: default statistical method (hypergeometric); minimum number of genes included in the term = 10 (GO enrichment analysis); minimum number of genes included in the term = 17 (disease association analysis); multiple test adjustment, BH method (Benjamini–Hochberg FDR); and significance level = adj$P < 0.01$.

Results

Variant detection and functional annotation

Based on the FastQC estimates, all 12 samples were of high quality; thus, no trimming was performed. The descriptive statistics of the raw and aligned reads, considering the 12 lymph node samples under study, are provided in Table 2. Considering SNPs, insertions and deletions, the number of common variants detected by GATK and Samtools packages that were considered as high-quality variants was 1 326 960 (Table 3). (Note: the corresponding variant call format file with the high-quality variants has been submitted to the European Variation Archive, https://www.ebi.ac.uk/eva/, 'Analysis ERZ1030229' within 'Project PRJEB33693'.) Of them, a total of 1 086 508 variants (81.8%) were mutations already described in the SNP databse (version ovis_ariesRelease91), whereas 240 452 were novel variants, without an associated *rs* number (Table 3). The variant density per Mb (Fig. 1a), when aligned against the sheep genome, showed a quite uniform distribution with an average variant density of 498 variants/Mb. The regions showing the highest variant density were located on chromosome 20 (OAR20) (range 2280–4370 variants/Mb). Some other regions located on OAR6, OAR11, OAR14, OAR17, OAR20 and OAR22 also showed more than 2000 variants/Mb. (Fig. 1a). The distribution of the variant density across the two types of target regions is also shown in Fig. 1b and c. Interestingly, some of the regions showing the highest variant density across the genome previously mentioned, those located on OAR20, OAR24 and OAR14, were coincident with QTL effects described for GIN resistance in sheep. (Fig. 1b). On the other hand, it can be seen that the variant density identified within the targeted GIN-activated genes was much lower than in the QTL regions (Fig. 1c).

From the high-quality variants, we extracted 224 191 variants mapping within 230 QTL regions related to resistance of sheep to parasites and 98 576 variants mapping within 1892 candidate GIN-activated genes. A total of 16 184 genetic variants were common between the QTL and the considered candidate genes (Table 3). The results of the annotation of these variants across the whole transcriptome, and also across the considered QTL and candidate gene regions, are summarised in Table S3. The location of the variants regarding the annotated genomic regions grouped in general categories (coding, UTRs, introns, non-coding RNA and intergenic) is shown in Fig. 2 for the variants detected across the whole transcriptome, and across the two types of target regions. For the whole transcriptome and the QTL regions, the distribution of the variants was similar, with the percentages of intronic, intergenic variants and non-coding variants (about 40, 24 and 27% respectively) being much higher than the percentage of coding variants (about 6%). The differences in the distribution observed for the variants detected within the candidate GIN-activated genes are obviously determined by the lack of intergenic variation.

Table 2 Descriptive statistics of the 12 abomasal lymph node RNA-Seq samples analysed in the present study.

Descriptive statistics	Reads
Average paired reads	45 574 815
Average read length	152 bp
Uniquely mapped read against Oar version 3.1	36 622 137 (80.27%)
Paired reads mapped to multiple loci	3 807 018 (8.5%)
Variants detected	
GATK	2 054 157
Samtools	1 798 196
Common variants	1 326 960

Based on the predicted impact on the protein function provided by the annotation analysis, the majority of the variants identified across the transcriptome were modifiers (93.39%), whereas the percentages of variants classified as low, moderate and high impact variants were 4.24, 2.29 and 0.06% (Table S3). As it is shown in Fig. 3, these proportions were similar for variants detected within the considered QTL regions (93.00, 4.35, 2.57 and 0.06%, for modifier, low-, moderate- and high-impact variants respectively), whereas for the candidate GIN-activated genes the percentage of variants included within the modifier category was slightly lower (88.86%) and slightly higher for each of the rest of the categories (7.59, 3.45 and 0.08%, for low-, moderate- and high-impact respectively).

Focusing on the variants that we have considered as potential functional variants within the target regions, we identified 186 and 105 high-impact variants, and 744 and 448 missense deleterious variants within the QTL and the candidate GIN-activated gene regions respectively. The percentage of those variants that were novel within each category ranged between 32 and 18% (Table 3). A full characterisation of the annotation of these variants ('high

Table 3 Distribution of the types of variants and of the variants considered in this work of potential functional impact (those detected as variant annotation for the high-quality variants detected across the whole abomasal lymph node transcriptome, the QTL and candidate GIN-activated genes considered in the present study).

		Abomasal lymph node transcriptome	Parasite resistance QTL regions	Candidate GIN-activated genes	Common variants (QTL and candidate GIN-activated genes)[2]
Types of high-quality variants[1]					
Transcriptome variants	SNPs	1 311 850	221 654	97 524	16 000
	Insertions	6960	1153	508	81
	Deletions	8150	1384	544	103
	Total	1 326 960	224 191	98 576	16 184
Novel variants	SNPs	237 058	36 492	8824	1428
	Insertions	1485	235	95	17
	Deletions	1909	307	103	19
	Total	240 452	37 034	9022	1464
Variants considered of functional biological impact					
Transcriptome variants	'High impact' variants	1066	186	105	18
	Missense deleterious variants	4076	744	448	71
Novel variants	'High impact' variants	343	52	26	5
	Missense deleterious variants	1008	138	86	14

The number of novel variants (those without an associated *rs* number) for each category and group is also indicated.
[1]The variants commonly identified by both software packages, GATK and SAMTOOLS, were considered as high-quality variants.
[2]Common variants identified between parasite resistance QTL and the 1892 genes considered as candidate GIN-activated genes in this work (based on a literature review of expression analysis studies in relation to parasite resistance in sheep).

impact' and missense deleterious variants) is provided in Table S4 for QTL regions and in Table S5 for the candidate genes. These potential functional variants within the QTL regions were distributed across a total of 424 genes, 111 of which were included in the considered reference databases of immune-related genes (highlighted in yellow in Table S4, Table 4). On the other hand, the 553 potential functional variants located within the candidate GIN-activated genes were distributed across a total of 336 genes, 132 of which were included in the considered reference databases of immune-related genes (highlighted in yellow in Tables S5 and 4).

The enrichment analyses performed with the 111 immune-related genes harbouring mutations with potential biological impact within the considered QTL regions identified a total of 41 significant GO terms from the biological process database. These terms were grouped, at the most superior hierarchical level, under the following five significant terms indicated in Table S6: *regulation of apoptotic process, cell adhesion, cell surface receptor signalling pathway, inflammatory response* and *enzyme linked receptor protein signalling pathway*. In addition, for the list of 132 immune-related GIN-activated genes harbouring mutations with potential biological impact, the disease association enrichment analysis highlighted five disease terms (*inflammation, adhesion, necrosis, immune system diseases* and *infection*) (Table S7).

Discussion

The use of next-generation sequencing facilitates the challenge of identifying genomic variation across genomes, which is a crucial step for unravelling the relationship between genotype and phenotype in both human and livestock species (Piskol *et al.* 2013). Here we consider the RNA-Seq dataset previously generated by our research group to identify DEGs in the transcriptome of the abomasal lymph node samples collected from 12 sheep (previously classified as six resistant and six susceptible) subjected to an experimental infection of *T. circumcincta* (Chitneedi *et al.* 2018) to study the genetic variation across the lymph node transcriptome at the time of an active immune response to GIN infection (7 days post-infection). Although the sensitivity to identify a variant and the coverage at a portion of the genome with RNA-Seq data is much lower than that of WGS or whole-exome sequencing (Piskol *et al.* 2013), variant calling with RNA-Seq data is less expensive and a proper experimental design can overcome this shortcoming. In our case, we adopted an analysis workflow previously described to identify variants across the sheep milk transcriptome (Suárez-Vega *et al.* 2017). The analysis of the genetic variability of the studied animals was carried out by considering the six resistant and six susceptible sheep as a single group, as it has been shown that a higher number

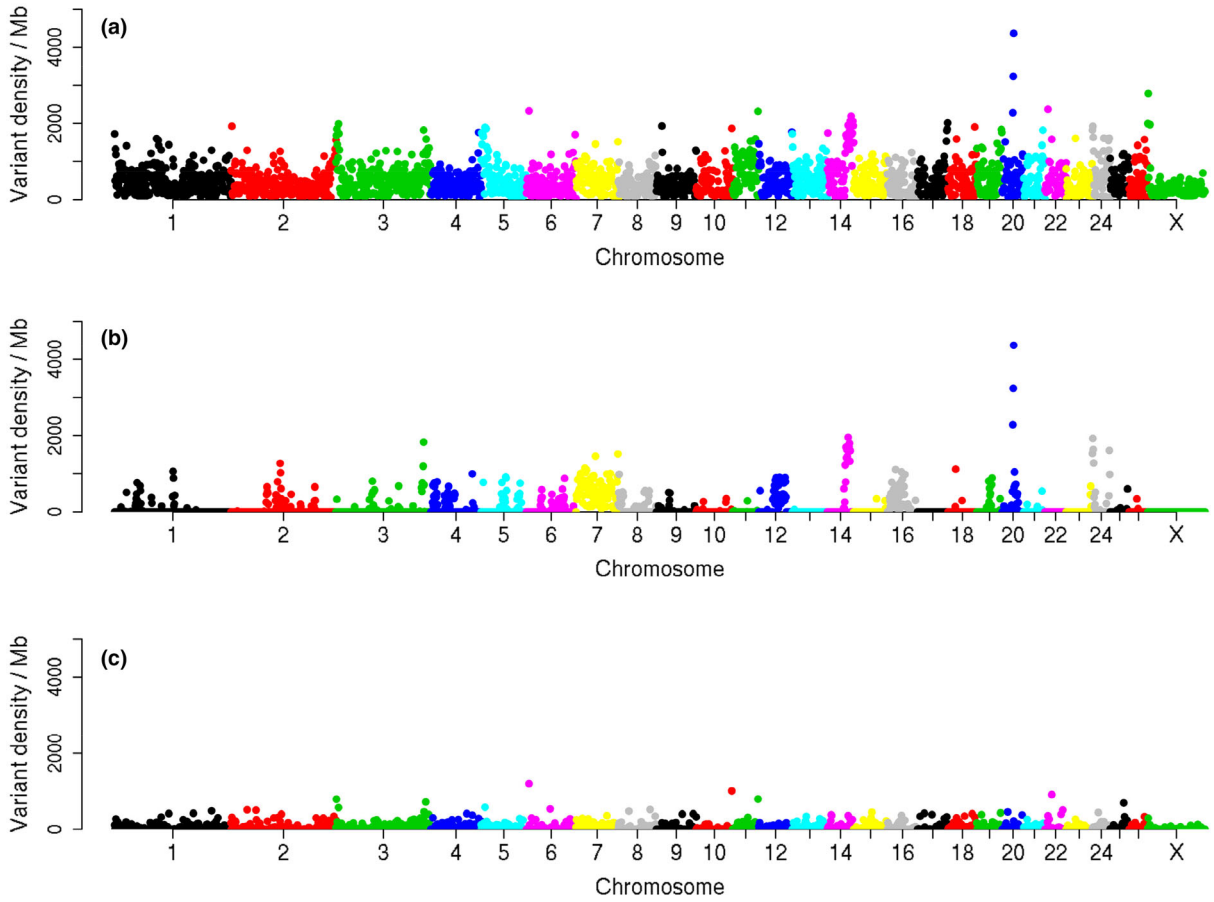


Figure 1 Manhattan plots showing the distribution of transcriptomic variants across the sheep reference genome (Oar_v3.1). a) Distribution of the high-quality variants identified from the analysis of the whole abomasal lymph node transcriptome of the 12 animals considered in the present study. b) Distribution of the subset of the high-quality variants identified from the analysis of the whole abomasal lymph node transcriptome that were mapped within the list of 230 previously reported QTL regions considered. c) Distribution of the subset of the high-quality variants identified from the analysis of the whole abomasal lymph node transcriptome that were mapped within candidate GIN-activated gene regions. Each dot in the plot indicates the number of variants per Mb (Mega base) and their position across the length of each chromosome.

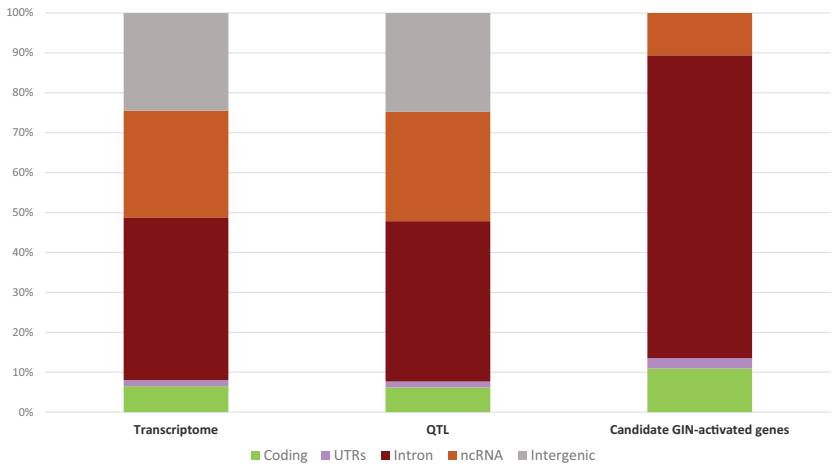


Figure 2 Distribution of variants in the annotated genomic regions. The location of the variants for the annotated genomic regions grouped in general categories (coding, UTRs, introns, non-coding RNA and intergenic) for the whole transcriptome, QTL and candidate GIN-activated gene regions is here indicated.

of samples considered during variant calling analysis results in a higher coverage and confidence in variant calls (Piskol *et al.* 2013). To ensure the quality of the variants identified, we performed the variant calling analysis with two different software packages, GATK and SAMTOOLS, such as in Gutiérrez-Gil *et al.* (2018).

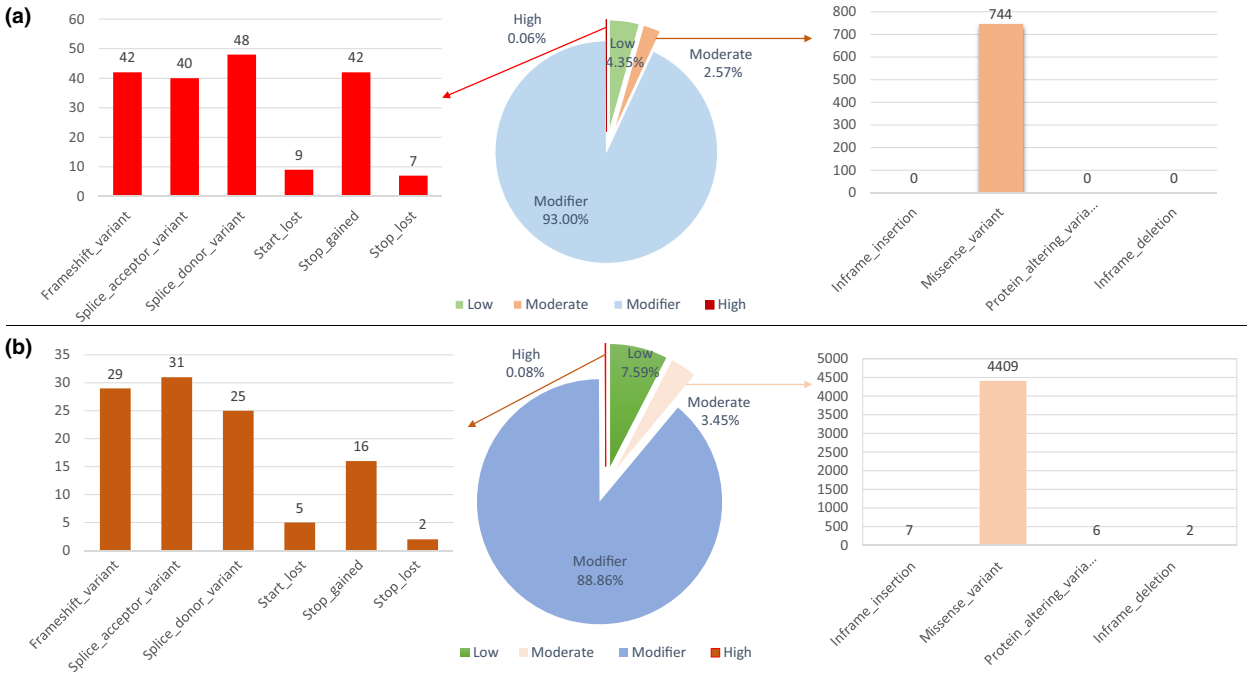


Figure 3 Pie charts showing the distribution of functional consequences of annotated variants identified within the two types of target regions considered. For each of the two target regions studied here, QTL and GIN-activated genes (a and b respectively), the annotation characterisation obtained with the VEP_90 software for the 'high impact' (left hand side) and the 'moderate impact' (right hand side) variants detected in the present study are presented.

Table 4 Distribution of the variants with a potential functional impact located within the considered list of QTL and candidate GIN-activated gene regions.

	Parasite resistance QTL	Candidate GIN-activated genes
Potential functional impact variants ('high impact' and moderate impact missense deleterious)		
Variants	930	553
Variant consequences	930	580
Number of genes harbouring potential functional impact mutations	424	336
Number of immune-related genes harbouring potential functional impact mutations	111	132

Regarding the results of the variant annotation reported here, the large percentage of the variants identified in intergenic regions, introns or non-coding RNA (~83%) agrees with the percentage of variants identified within non-coding regions through the transcriptomic analysis of the milk somatic cells in sheep (~71%) (Suárez-Vega *et al.* 2017). Indeed, the distribution pattern of variants across different genomic regions (introns, non-coding RNA, coding, intergenic and UTRs) shown in Fig. 2 is similar to that reported in humans by Piskol *et al.* (2013) for RNA-Seq. As indicated by these authors, the large number of intronic variants in our RNA-Seq analysis (40% for the transcriptome and QTL regions, and 75% for the candidate gene regions) can be explained by (i) the fact that poly(A)+ RNA-capturing protocols can capture premature mRNAs that still contain introns or (ii) the presence of one type of alternative spliced RNAs, intron retention; in any case, introns compose a much larger fraction and are more variable than exons (Piskol *et al.* 2013). Alternatively, the identification of variants in intergenic regions may indicate the presence of non-coding RNAs that have not yet been annotated in the sheep reference genome and which are increasingly being accepted as one of the factors contributing to the complex physiology and architecture of mammalian species (Dhamija & Menon 2018). These observations support the need to further improve the annotation of the reference genome.

The systematic search of transcriptome variants and the filtering steps implemented here have allowed us to highlight potential genetic variants that could be considered as potential causal variants of phenotypic effects in relation to *T. circumcincta* sheep resistance. For the first of the two types of target regions, Table S8 provides a link between the 36 'high impact' and the 142 missense deleterious variants harboured by immune-related genes mapping within the 230 QTL considered in this work. From this list, we can extract valuable information, especially for those QTL for which a limited number of potential causal variants are

suggested. Hence, the 16 QTL for which only one or two potential causal variants have been identified through our transcriptomic approach may be suitable for further studies to assess the role of those specific mutations in relation to the previously mapped QTL effects (Table S9). Among the 17 genes indicated in Table S9, we would highlight the mutations located in the genes *IFI44*, *COL3A1*, *PLXNC1*, *SEMA3A*, *CD74*, *MAP1S*, *CSF1R*, *CCNDBP1*, *BNIP2*, *IL20RA*, *PRKDC*, *KLHL21*, *LAMC2*, *AEN*, *NUMBL*, *TCIRG1* and *HRH1*, as they were identified as single candidate casual mutations for 15 QTL reported for parasite resistance (Table S4).

Regarding the specific five terms highlighted by the GO enrichment analysis of the genes located in QTL regions harbouring the list of mutations with potential biological impact (Table S6), *cell surface receptor signalling pathway* was the significant enriched term including the highest number of genes (32). During GIN infection in sheep, the cytokine-mediated multifactorial cellular signalling has been found to guide the cell-mediated localised response (Sykes & Coop 2001). In *T. circumcincta*-infected sheep, the secretion of Th2 cytokines was associated with resistance and the activation of Th1/Th7 cytokines was linked to susceptibility (Gossner *et al.* 2012, 2013). Through further linking each target QTL region with the corresponding enriched terms supported by the genes harbouring the potential causal variants (e.g. *cell adhesion, cell surface signalling pathway*, etc.) our results might help to identify the biological function related to the role of the QTL effect in relation to the host response to GIN resistance (see last column in Table S8 and Table S9). Globally, this list of candidate mutations in relation to previously identified QTL may be of great interest to design future studies aiming to identify genetic markers that could be directly used in genetic selection to improve parasite resistance in sheep. Ideally, the confirmation of the relationship of these mutations with the corresponding QTL should be tested in the population used to map the genetic effect. If this is not possible, general association analyses in large commercial populations could be also an appropriate approach to test the link of these mutations to indicator traits of parasite resistance across different sheep breeds and in relation to different parasite species.

Among the five disease terms identified as significantly enriched by the analysis performed for the 132 immune-related GIN-activated genes harbouring potential causal variants, *Adhesion* was the term including the highest number of genes (25) (Table S7). Adhesion molecules, such as those coded by the genes indicated in Table S7 (e.g. *LAMA4*, *SLC3A2*, *ITGA8*) play a key role in cell–cell and cell–extracellular matrix interactions, which are necessary for proliferation and differentiation of both eosinophils and other inflammatory cells during their activation at inflammatory sites (Stevenson *et al.* 2001). Cell adhesion and cell matrix adhesion have been reported as important biological processes involved in the development of resistance to a primary infection with *Haemonchus contortus* in Merino sheep (Zhang *et al.* 2019), whereas cell adhesion molecules pathways (*SELP*, *NLGN2*, *L1CAM*, *ITGA4*) were identified as highly expressed in goats resistant to GIN infection (Bhuiyan *et al.* 2017). In relation to the term *inflammation*, Khan & Collins (2004) found that the development of inflammation in the intestinal mucosal surface was considered as an early immune response to GIN infection in sheep. In this sense, Amarante *et al.* (2007), based on the study of three sheep breeds naturally infected with *Trichostrongylus colubriformis*, suggested that the inflammatory cells impair the parasite's establishment, development and survival. Finally, it has been shown that at the infection site *Trichostrongylus* larvae induce intestinal epithelial cell necrosis which determines the release of interleukin 33 (IL33) protein (Andronicos *et al.* 2012). For the gene coding that protein, *IL33* (which supported the *Necrosis* enriched term), our filtering approach identified a single potential functional variant (*rs410259751*, Table S5) that should be considered as candidate for future confirmation of genetic markers underlying the host response to GIN infection in sheep. Considering the list of genes that enriched the mentioned disease-related terms and for which a single potential functional variant was identified through the variant annotation reported in this study, we highlight a list of 67 mutations that could be considered as GIN resistance candidate mutations in future studies (see last column of Table S7).

In conclusion, this study provides a transcriptomic approach to identify potential functional variants underlying the resistance to *T. circumcincta* in experimentally infected adult sheep through the analysis of the abomasal lymph node RNA-Seq dataset extracted from both resistant and susceptible sheep. In relation to parasite resistance QTL regions previously reported, our analyses identified some important potential causal mutations harboured by immune-related genes. We identified few other candidate causal mutations in the genomic regions of candidate GIN-activated genes. Future studies should assess the role of the potential causal mutations discovered in this study with the aim of identifying genetic markers that could be directly implemented in sheep breeding programmes designed to combat nematode infections.

Acknowledgements

Financial support for this project was received from the LE248U14 project of Junta de Castilla and León Government. The research was also partially funded by the Spanish Ministry of Economy and Competitiveness (AGL2015-66035-R project, co-funded by the European Regional Development Fund). P. K. Chitneedi is funded by a predoctoral fellowship from the Junta de Castilla and León Government and the European Social Fund. M. Martínez-

 doi: 10.1111/age.12894

Valladares is also funded by the "Ramón y Cajal" Programme (RYC-2015-18368) from MINECO.

References

Al Kalaldeh M., Gibson J., Lee S.H., Gondro C., van der Werf J.H. (2019a) Detection of genomic regions underlying resistance to gastrointestinal parasites in Australian sheep. *Genetics Selection Evolution* **51**, 37.

Al Kalaldeh M., Gibson J., Duijvesteijn N., Daetwyler H.D., MacLeod I., Moghaddar N., Lee S.H., van der Werf J.H. (2019b) Using imputed whole-genome sequence data to improve the accuracy of genomic prediction for parasite resistance in Australian sheep. *Genetics Selection Evolution* **51**, 32.

Amarante A.F.T., Raquel A.R., Patrizia A.B. (2007) Relationship of intestinal histology with the resistance to *Trichostrongylus colubriformis* infection in three breeds of sheep. *Pesquisa Veterinária Brasileira* **27**, 43–8.

Andrews S. (2010) *FastQC: a quality control tool for high throughput sequence data*. Bioinformatics B.

Andronicos N, Hunt P, Windon R (2010) Expression of genes in gastrointestinal and lymphatic tissues during parasite infection in sheep genetically resistant or susceptible to *Trichostrongylus colubriformis* and *Haemonchus contortus*. *International Journal for Parasitology* **40**, 417–29.

Andronicos N.M., McNally J., Kotze A.C., Hunt P.W., Ingham A. (2012) *Trichostrongylus colubriformis* larvae induce necrosis and release of IL33 from intestinal epithelial cells in vitro: implications for gastrointestinal nematode vaccine design. *International Journal for Parasitology* **42**, 295–304.

Athanasiadou S., Pemberton A., Jackson F., Inglis N., Miller H.R., Thévenod F., Mackellar A., Huntley J.F. (2008) Proteomic approach to identify candidate effector molecules during the in vitro immune exclusion of infective *Teladorsagia circumcincta* in the abomasum of sheep. *Veterinary Research* **39**, 1.

Atlija M., Arranz J.J., Martinez-Valladares M., Gutiérrez-Gil B. (2016) Detection and replication of QTL underlying resistance to gastrointestinal nematodes in adult sheep using the ovine 50K SNP array. *Genetics Selection Evolution* **48**, 4.

Van der Auwera G.A., Carneiro M.O., Hartl C. *et al.* (2013) From fastQ data to high-confidence variant calls: The genome analysis toolkit best practices pipeline. *Current Protocols in Bioinformatics* **43**, 11–10.

Behnke J.M., Iraqi F., Menge D., Baker R.L., Gibson J., Wakelin D. (2003) Chasing the genes that control resistance to gastrointestinal nematodes. *Journal of Helminthology* **77**, 99–110.

Benavides M.V., Sonstegard T.S., Kemp S., Mugambi J.M., Gibson J.P., Baker R.L., Hanotte O., Marshall K., Van Tassell C. (2015) Identification of novel loci associated with gastrointestinal parasite resistance in a Red Maasai x Dorper backcross population. *PLoS ONE* **10**, e0122797.

Beraldi D., McRae A.F., Gratten J., Pilkington J.G., Slate J., Visscher P.M., Pemberton J.M. (2007) Quantitative trait loci (QTL) mapping of resistance to strongyles and coccidia in the free-living Soay sheep (*Ovis aries*). *International Journal for Parasitology* **37**, 121–9.

Bhattacharya S., Andorf S., Gomes L. *et al.* (2014) ImmPort: Disseminating data to the public for the future of immunology. *Immunologic Research* **58**, 234–39.

Bhuiyan A., Li J., Wu Z., Ni P., Adetula A., Wang H., Du X. (2017) Exploring the genetic resistance to gastrointestinal nematodes infection in goat using RNA-sequencing. *International Journal of Molecular Sciences* **18**, 751.

Breuer K., Foroushani A.K., Laird M.R., Chen C., Sribnaia A., Lo R., Winsor G.L., Hancock R.E.W., Brinkman F.S.L., Lynn D.J. (2013) InnateDB: Systems biology of innate immunity and beyond - Recent updates and continuing curation. *Nucleic Acids Research* **41**, D1228–33.

Broad Institute G. repository, (2018) *Picard tools*.

Brown D.J., Fogarty N.M. (2017) Genetic relationships between internal parasite resistance and production traits in Merino sheep. *Animal Production Science* **57**, 209–15.

Chitneedi P.K., Suárez-Vega A., Martínez-Valladares M., Arranz J.J., Gutiérrez-Gil B. (2018) Exploring the mechanisms of resistance to *Teladorsagia circumcincta* infection in sheep through transcriptome analysis of abomasal mucosa and abomasal lymph nodes. *Veterinary Research* **49**, 39.

Churko J.M., Mantalas G.L., Snyder M.P., Wu J.C. (2013) Overview of high throughput sequencing technologies to elucidate molecular pathways in cardiovascular diseases. *Circulation Research* **112**, 1613–23.

Cingolani P., Platts A., Wang L.L. *et al.* (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly* **6**, 80–92.

Coltman D.W., Wilson K., Pilkington J.G., Stear M.J., Pemberton J.M. (2001) A microsatellite polymorphism in the gamma interferon gene is associated with resistancto gastrointestinal nematodes in a naturally-parasitized population of Soay sheep. *Parasitology* **122**, 571–82.

Ensembl Database (2012) *International Sheep Genome Consortium*. Ovis aries Oar_v3.1, INSDC Assembly.

Dhamija S., Menon M.B. (2018) Non-coding transcript variants of protein-coding genes–what are they good for? *RNA Biology* **15**, 1025–31.

Dobin A., Davis C.A., Schlesinger F., Drenkow J., Zaleski C., Jha S., Batut P., Chaisson M., Gingeras T.R. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21.

Dominik S. (2005) Quantitative trait loci for internal nematode resistance in sheep: a review. *Genetics Selection Evolution* **37**, S83.

El-Ashram S., Li C., Abouhajer F. *et al.* (2018) An ex vivo abomasal ovine model to study the immediate immune response in the context of *Haemonchus contortus* larval-stage. *Veterinary Parasitology* **254**, 105–13.

French A.T., Knight P.A., Smith W.D., Pate J.A., Miller H.R., Pemberton A.D. (2009) Expression of three intelectins in sheep and response to a Th2 environment. *Veterinary Research* **40**, 1–13.

García-Pérez A.l., Hurtado A., Oregui L.M., Juste R.A. (2002) Effects of a second annual strategic anthelmintic treatment in dairy sheep in Northern Spain. *Small Ruminant Research* **43**, 121–6.

Gossner A.G., Venturina V.M., Shaw D.J., Pemberton J.M., Hopkins J. (2012) Relationship between susceptibility of Blackface sheep to *Teladorsagia circumcincta* infection and an inflammatory mucosal T cell response. *Veterinary Research* **44**, 68.

Gossner A., Wilkie H., Joshi A., Hopkins J. (2013) Exploring the abomasal lymph node transcriptome for genes associated with resistance to the sheep nematode *Teladorsagia circumcincta*. *Veterinary Research* **43**, 26.

Grencis R.K., Humphreys N.E., Bancroft A.J. (2014) Immunity to gastrointestinal nematodes: mechanisms and myths. *Immunological reviews* **260**, 183–205.

Guo Z., González J.F., Hernandez J.N., McNeilly T.N., Corripio-Miyar Y., Frew D., Morrison T., Yu P., Li R.W. (2016) Possible mechanisms of host resistance to *Haemonchus contortus* infection in sheep breeds native to the Canary Islands. *Scientific Reports* **6**, 26200.

Gutiérrez-Gil B., Pérez J., Álvarez L., Martínez-Valladares M., La Fuente L.F., De Bayán Y., Meana A., Primitivo F.S., Rojo-Vzquez F.A., Arranz J.J. (2009) Quantitative trait loci for resistance to trichostrongylid infection in Spanish Churra sheep. *Genetics Selection Evolution* **41**, 46.

Gutiérrez-Gil B., Pérez J., De la Fuente L.F., Meana A., Martínez-Valladares M., San Primitivo F., Rojo-Vázquez F.A., Arranz J.J. (2010) Genetic parameters for resistance to trichostrongylid infection in dairy sheep. *Animal* **4**, 505–12.

Gutiérrez-Gil B., Esteban-Blanco C., Suarez-Vega A., Arranz J.J. (2018) Detection of quantitative trait loci and putative causal variants affecting somatic cell score in dairy sheep by using a 50K SNP-Chip and whole genome sequencing. *Journal of Dairy Science* **101**, 9072–88.

Hu Z.-L., Park C.A., Reecy J.M. (2015) Developmental progress and current status of the Animal QTLdb. *Nucleic Acids Research* **44**, D827–33.

Ingham A., Reverter A., Windon R., Hunt P., Menzies M. (2008) Gastrointestinal nematode challenge induces some conserved gene expression changes in the gut mucosa of genetically resistant sheep. *International Journal for Parasitology* **38**, 431–42.

Kelley J., de Bono B., & Trowsdale J. (2005) IRIS: A database surveying known human immune system genes. *Genomics* **85**, 503–11.

Khan W.I., Collins S.M. (2004) Immune-mediated alteration in gut physiology and its role in host defence in nematode infection. *Parasite Immunology* **26**, 319–26.

Knight P.A., Griffith S.E., Pemberton A.D. *et al.* (2011) Novel gene expression responses in the ovine abomasal mucosa to infection with the gastric nematode Teladorsagia circumcincta. *Veterinary Research* **42**, 78.

MAFF (Ministry of Agriculture, Fisheries and Food) (1986) *Manual of veterinary parasitological laboratory techniques*, 3rd edn. London: GB, p 1986.

McLaren W., Gil L., Hunt S.E., Riat H.S., Ritchie G.R.S., Thormann A., Flicek P., Cunningham F. (2016) The Ensembl Variant Effect Predictor. *Genome Biology* **17**, 122.

McRae K.M., McEwan J.C., Dodds K.G., Gemmell N.J. (2014) Signatures of selection in sheep bred for resistance or susceptibility to gastrointestinal nematodes. *BMC Genomics* **15**, 637.

McRae K.M., Good B., Hanrahan J.P., McCabe M.S., Cormican P., Sweeney T., O'Connell M.J., Keane O.M. (2016) Transcriptional profiling of the ovine abomasal lymph node reveals a role for timing of the immune response in gastrointestinal nematode resistance. *Veterinary Parasitology* **224**, 96–108.

Mirkena T., Duguma G., Haile A., Tibbo M., Okeyo A.M., Wurzinger M., Sölkner J. (2010) Genetics of adaptation in domestic farm animals: A review. *Livestock Science* **132**, 1–12.

Mpetile Z., Cloete S.W.P., Kruger A.C.M., Dzama K. (2015) Environmental and genetic factors affecting faecal worm egg counts in Merinos divergently selected for reproduction. *South African Journal of Animal Science* **45**, 510–20.

Ng P.C., Henikoff S. (2003) SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Research* **31**, 3812–4.

Ngere L., Burke J.M., Morgan J.L.M., Miller J.E., Notter D.R. (2018) Genetic parameters for fecal egg counts and their relationship with body weights in Katahdin lambs. *Journal of Animal Science* **96**, 1590–99.

Piskol R., Ramaswami G., Li J.B. (2013) Reliable identification of genomic variants from RNA-seq data. *American Journal of Human Genetics* **93**, 641–51.

Riggio V., Matika O., Pong-Wong R., Stear M.J., Bishop S.C. (2013) Genomewide association and regional heritability mapping to identify loci underlying variation in nematode resistance and body weight in Scottish Blackface lambs. *Heredity* **110**, 420.

Rowe A., Gondro C., Emery D., Sangster N. (2009) Sequential microarray to identify timing of molecular responses to *Haemonchus contortus* infection in sheep. *Veterinary Parasitology* **161**, 76–87.

Rupp R., Senin P., Sarry J. *et al.* (2015) A point mutation in suppressor of cytokine signalling 2 (Socs2) increases the susceptibility to inflammation of the mammary gland while associated with higher body weight and size and higher milk production in a sheep model. *PLoS Genetics* **11**, e1005629.

Stevenson L.M., Colditz I.G., Lejambre L.F. (2001) Expression of cell surface adhesion molecules by peripheral blood eosinophils during *Trichostrongylus colubriformis* infection in sheep. *Immunology and Cell Biology* **79**, 240–4.

Suárez-Vega A., Gutiérrez-Gil B., Klopp C., Tosser-Klopp G., Arranz J.J. (2017) Variant discovery in the sheep milk transcriptome using RNA sequencing. *BMC Genomics* **18**, 170.

Sykes A.R., Coop R.L. (2001) Interaction between nutrition and gastrointestinal parasitism in sheep. *New Zealand Veterinary Journal* **49**, 222–6.

Vercruysse J., Dorny P., Hong C., Harris T.J., Hammet N.C., Smith D.G., Weatherley A.J. (1993) Efficacy of doramectin in the prevention of gastrointestinal nematode infections in grazing cattle. *Veterinary Parasitology* **49**, 51–9.

Wang Z., Gerstein M., Snyder M. (2009) RNA-Seq: A revolutionary tool for transcriptomics. *Nature Reviews Genetics* **10**, 57.

Wang J., Duncan D., Shi Z., Zhang B. (2013) WEB-based gene set analysis toolkit (WebGestalt): update 2013. *Nucleic acids research* **41**, W77–83.

Wolstenholme A.J., Fairweather I., Prichard R., von Samson-Himmelstjerna G., Sangster N.C. (2004) Drug resistance in veterinary helminths. *Trends in Parasitology* **20**, 469–76.

Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G., Durbin R. (2009) The Sequence alignment/map (SAM) format and SAMtools. *Bioinformatics* **25**, 2078–9.

Zhang R., Liu F., Hunt P., Li C., Zhang L., Ingham A., Li R.W. (2019) Transcriptome analysis unraveled potential mechanisms of resistance to Haemonchus contortus infection in Merino sheep populations bred for parasite resistance. *Veterinary Research* **50**, 7.

Supporting information

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

Figure S1. Bioinformatic pipeline used in the current study.

 doi: 10.1111/age.12894

Table S1. List of 230 QTL regions for the faecal egg count (FEC) trait previously reported by 11 different QTL studies.

Table S2. List of 1892 unique genes considered in the present study as candidate gastrointestinal nematode (GIN)-activated genes based on the literature review of expression studies related to parasite resistance to gastrointestinal nematode infections in sheep.

Table S3. Summary statistics of the identified variants in the abomasal lymph node transcriptome after variant annotation with the software VEP version 90.

Table S4. List of potential functional variants ('high impact' and missense deleterious variants) identified in the abomasal lymph node transcriptome by mapping the 230 QTL regions influencing GIN resistance in sheep that were considered here from 11 independent QTL mapping studies based on SNP-chips or sequencing technology.

Table S5. List of potential functional variants ('high impact' and missense deleterious) identified in the abomasal lymph node transcriptome that mapped within the genomic regions harbouring the 1892 genes considered as candidate GIN-activated genes in the present study.

Table S6. Top hierarchical significant terms identified by GO enrichment analysis (biological process database) performed with WEBGESTALT for the 111 immune-related genes harbouring potential functional variants in relation to the list of 230 QTL regions influencing GIN resistance in sheep considered here.

Table S7. Significant terms enriched identified by the disease association enrichment analysis performed with WEBGESTALT for the immune-related 132 genes harbouring potential functional variants in relation to the list of candidate GIN-activated genes considered in this work.

Table S8. List of the 230 target QTL regions considered in the current study and their correspondence with the variants identified as candidate causal mutations in relation to GIN resistance in sheep according to the filtering steps applied in the present work.

Table S9. Subset of Table S8 with the 16 QTL regions for which a maximum of two potential causal variants were detected.

Study 9

Detection of long-noncoding RNAs from the differential transcriptomic analysis of abomasal lymph node from resistant and susceptible sheep to gastrointestinal nematode (*T. circumcincta*) after an experimental infection

Praveen Krishna Chitneedi[1], Christa Kühn[2], Rosemarie Weikard[2], Juan-Jose Arranz[1], María Martinez-Valladares[3,4], and Beatriz Gutiérrez-Gil[1].

[1]Departamento de Producción Animal, Facultad de Veterinaria, Universidad de León, León, Spain; [2]Leibniz Institute for Farm Animal Biology (FBN), Institute of Genome Biology,Dummerstorf 18196, Germany; [3]Departamento de Sanidad Animal, Facultad de Veterinaria, Universidad de León, León, Spain; [4]Instituto de Ganadería de Montaña, CSIC-ULE, León, Spain.

Collaborative work funded by the FAANG-Europe COST Action short term scientific mission (STSM) (Ref: ECOST-STSM-Request-CA15112-43505)

37th International Society for Animal Genetics (ISAG) conference, 7th- 12th of July 2019, Lleida, Spain.

Detection of long-non coding RNAs from the differential transcriptomic analysis of abomasal lymph node from resistant and susceptible sheep to gastrointestinal nematode (*T. circumcincta*) after an experimental infection.

P.K. Chitneedi, C. Kühn, R. Weikard, J.J Arranz, M. Martínez-Valladares, and B. Gutiérrez-Gil

In this study, we detected long noncoding RNAs (lncRNAs) in the transcriptome of abomasal lymph node tissue, which was extracted from adult sheep after an experimental infection with *T. circumcincta*. After a first experimental infection based on the accumulative faecal egg counts, six ewes were classified as 'Susceptible' and six ewes as 'Resistant'. After second experimental infection, these animals were sacrificed and abomasal lymph node tissue samples were collected. The total polyA RNA extracted from these samples was sequenced using an Illumina Hi-SEq 2000 sequencer by generating 'paired-end' reads of 75 bp, with a depth of 30M reads per sample. Using standard pipeline with different bioinformatics tools, initially the quality of raw sequence data was assessed, followed by trimming adapter and low-quality sequences, aligning and sorting the obtained reads against the sheep reference genome (Oar_v.3.1) ensembl release 95. Then we detected potential transcripts by reference-guided transcript assembly, merging and transcript annotation and finally counted the number of transcripts in each sample in the annotated feature file. On average, the data contained 44,203 genes with 77,039 transcripts. This annotated feature file was used to detect the lncRNA with the FEELnc package. In total 9,105 lncRNA were detected. Among them 2,092 were classified as novel with gffcompare tool. After differential gene expression with DESeq2 between susceptible and resistant ewes, we found 440 genes differentially expressed (adj p-value < 0.05) with 158 up-regulated and 282 down regulated in resistant sheep. Out of 440 differentially expressed genes, 44 contain 63 lncRNA transcripts out of these 30 were novel. Finally, we performed sequence similarity search by aligning the sequence data of the 63 putative lncRNA transcripts against sheep reference genome using BLASTN software.

Key Words: sheep and related species, Non-coding RNA, Bioinformatics Tools, RNA-seq

5. Discussion

The development of anthelmintic resistance in GINs has made the use of anthelmintic treatments no longer a viable option to tackle the parasite infection in sheep. As an alternative and a sustainable approach, several authors have suggested that the selection and breeding sheep for resistance against GIN (Gray 1997; Besier and Love 2003; Sayers and Sweeney 2005). Classical selection of sheep for GIN resistance was performed in some populations mainly based on the FEC indicator trait (Douch *et al.* 1996; Dominik 2005). However, these approaches were found to be unattractive due to higher cost and time involved in recording the phenotypes. Thus, investigating the genetic bases of nematode resistance in sheep has been considered helpful to identify genomic information that could assist classical genetic breeding, or even genomic selection approaches to improve resistance to GIN in commercial sheep populations.

Nematode resistance is a complex heritable phenotype, which is determined by a large number of genes with small effects. In small ruminants like sheep identifying the genomic regions and the gene expressions in response to GIN infection will increase our knowledge on the different immune mechanisms activated in the host against nematode infection. On this regard, many of the genomic studies for GIN resistance have studied young animals whereas our research group has focused its efforts in this field on adult dairy ewes (Gutierrez-Gil *et al*. 2009; Atlija *et al.* 2016). The reason for this is because in the Spanish northwest region of Castilla y León the production system of dairy sheep flocks is based on the milk from adult ewes and the sale of suckling lambs fed exclusively on maternal milk. Hence, in this production system, replacement ewes and adult sheep are the animals subjected to the direct effects of the helminth infections (García-Pérez *et al*. 2002). Thus, this PhD thesis carried out in the MEGA-ULE research group is focused on the mapping of QTL influencing the immune response towards nematode infection in adult sheep and also the identification of gene expression profiles that could be associated to the resistance/susceptibility status of this disease also in adult ewes.

Concerning gene mapping studies, the MEGA-ULE group identified some QTL for GIN resistance-related traits by performing a genome scan based on the LA of microsatellite markers (Gutiérrez-Gil *et al.* 2009). Later, with the availability of the 50K-chip for sheep, some new QTL were identified using different analysis methodologies, such as LA, LDLA and GWAS (Atlija *et al.* 2016). In this thesis, the first specific objective is an extension to the work of the MEGA-ULE group in the field of gene mapping for GIN

resistance by inferring the available 50K-chip genotypes to the density of the HD (600K) SNP-chip and also by exploiting information extracted from whole-genome sequencing (WG-seq) to filter the potential casual mutations of a previously reported QTL. As a pre-mapping step and to check the efficiency of the newly available HD-chip the previous estimates of LD extent, *Ne* and inbreeding coefficient based on the 50K-chip (García-Gámez *et al.* 2012) were initially validated using the imputed HD-chip dataset. The higher marker density allowed, also, to present a new analysis of detection of ROHs (runs of homozygosity). Later, the imputed HD-chip genotypes were used to refine the location of some previously identified QTL regions by Atlija *et al.* (2016) using LA and GWAS approaches, and also add some novel QTL for GIN resistance. The replication of the most significant QTL reported by Atlija *et al*. (2016), for FEC on OAR6, served us to conduct a high-resolution study of the genetic variability of the correspondent confidence interval by using information derived from the WG-seq analysis of a segregating trio for that QTL.

The second specific objective of this thesis was focused on the transcriptomic analysis of target tissues affected by the GIN infection in adult sheep. With the availability of NGS technologies like RNA sequencing (RNA-seq), the functional characterization of GIN resistance in sheep has become possible at the genomic level. In this objective, the transcriptomic data from abomasal tissues was studied after an experimental GIN infection to identify DEGs, functional variants and lncRNA transcripts that may play a role with the host response to nematode infections in sheep and to identify possible immune mechanisms against nematode infection.

In the following sections, the results obtained about each of the objectives defined for this Thesis project are summarized and discussed to their implications with the study of the genetic basis of nematode infection in adult sheep.

5.1 Objective 1: Refinement and high-resolution mapping of QTL influencing GIN resistance in adult sheep

5.1.1 Genomic imputation

The first and essential step for performing high-density gene mapping studies for GIN resistance in the Churra population previously studied by Atlija *et al.* (2016) was the genotyping of a subset of individuals of that population with the HD-chip and the subsequent imputation of genotypes for the global population from the medium to the high genotypic density. Importantly, the selection of the animals to be genotyped with the HD-chip was made carefully to maintain the half-sib family structure of the global population (14 daughters of each half-sib family and the 16 sires of the global population).

Previously to the generation of imputed genotypes for the global population for gene mapping or other high-density analyses, it was necessary to identify the best software to implement imputation in the half-sib structure family under study. For that, we first designed a cross-validation strategy to assess Beagle, one of the most used imputation software in livestock species. We are considering the most updated available version of this software, Beagle_v3.3.2 (Browning and Browning 2009).

The cross-validation masking strategy was adopted to eliminate any bias in the estimation of accuracy. This process was followed by considering population structure of sub-populations genotyped with HD-chip, which has 16 half-sib families with 14 daughters each. Hence, the 50K-chip to HD-chip density imputation was repeated iteratively (16 times) by masking the original HD-chip genotypes of one daughter per family across all daughters in each round. After imputation, the original HD-chip genotypes of each masked daughter were compared with the corresponding imputed genotypes for accuracy estimation.

Initially, the genotypes of the resource population were imputed from 50K to HD-chip using and the accuracy of imputation was estimated using the Plink (Purcell *et al.* 2007) concordance command merge-mode 7. Due to the high computation time required for the imputation process, the concordance estimation was made based on the results obtained on six chromosomes of different lengths (chromosomes 1, 4, 10, 16, 20, 26). The average concordance across these six chromosomes after 14 iterations was 90.85%, (range: 90-92

%, SD average: ± 1,55%). The actual concordance might be a bit higher than these estimates as masking few individuals in each iteration might reduce the power of imputation and when the genotypes were imputed without masking the animals of the reference population, the imputation accuracy was 94%. Thus, we can consider that the actual concordance provided by Beagle_v3.3.2 in the Churra population considered varies between 91-94%, high enough to use the imputed genotypes for later analysis of association analysis (**Study 3**) with traits of productive interest.

Later, we decided to repeat the imputation of genotypes from 50K to HD-chip for the global Churra population of 1,686 individuals by using the newly available version of Beagle, Beagle_v4.0 (Browning and Browning 2013), which in contrast to the previous version uses pedigree information, and also another imputation software known as FImpute_v2.2 (Sargolzaei *et al.* 2012), which had been highlighted by different authors about genomic imputation in livestock species because of its low computational time compared with other software (Johnston *et al.* 2011; Ma *et al.* 2013; Sargolzaei *et al.* 2014). Hence, as reported in **Study 1**, following a similar cross-validation strategy to that used with Beagle_v3.3.2, we estimated the imputation accuracy for Beagle_v4.0 and FImpute_v2.2, but considering the results obtained for all the sheep autosomal chromosomes. The estimated imputation accuracies for Beagle_4.0 and FImpute were 92.2% and 92.1%, respectively. These accuracy estimates were within the range of those reported by Hayes *et al.* (2012) for Beagle_v3.3.2, which does not consider pedigree information, in relation to the imputation of genotypes from medium-density to high-density SNP-chip panels, and even to sequencing data (range: 0.83 for Merino to 0.93 for Border Leicester and White-Faced Suffolk/Poll Dorset, respectively). Considering the little increase of imputation accuracy estimated for Beagle_v4.0 (92.2%) compared with that estimated for Beagle_v3.3.2 (average of 90.85%), we consider that the use of the pedigree information, as implemented by Beagle_v4.0, is an advisable approach when performing genotypic imputation in our study population. It seems that for imputation in a half-sib population like the one analysed here, the highest proportion of information to correctly impute genotypes comes from the known phases of the sires. It must be considered that the masking strategy used to estimate the imputation accuracy is likely to underestimate the actual imputation accuracy due to the decreased size of the reference population. Hence, when considering the whole reference population when assessing the imputation accuracy for chromosome 26, the gain of imputation accuracy was

approximately 5% for both Beagle and FImpute. We also compared the computing time for the two considered software. The computational time required for imputing the genotypes of chromosome 26, involving 8,728 markers in the HD-chip density, was approximately 17 minutes in Beagle and 1 minute in FImpute. This time advantage observed in the shortest chromosome under study showed a remarkable increase compared to the computation time for chromosome 1, which included 55,659 markers in the HD-chip dataset (165 minutes for Beagle and 7 minutes for FImpute). The difference in computational time could be attributed to the different algorithmic architectures of the two programs. Due to its slightly higher imputation accuracy, and despite its longer computational time, Beagle_v4.0 was used to obtain the imputed HD-chip genotypes for the whole study population across the 26 autosomes without using the masking strategy implemented for the accuracy estimation. These imputed genotype dataset was used for estimating the extent of LD, ROH, *Ne* and inbreeding coefficient (**Study 1**) and also for the high-resolution gene mapping Linkage Analysis for indicator traits of parasite resistance in sheep (**Study 4**).

From the practical point of view, the relevance of **Study 1** is related to the practical applications that imputation of genotypes may have in commercial populations, especially in relation to the implementation of Genomic Selection. In sheep, the application of genomic selection depends largely on the price associated with genotyping the animals. Therefore, the possibility of having a SNP-chip with a cost adapted to the individual value of the animals could make genomic selection more attractive for sheep breeders. The development of a low-density SNP-chip and the establishment of an imputation strategy to a higher density chip could offer a practical and efficient solution to this problem, by reducing the total cost associated with genotyping while maintaining acceptable reliability of genomic estimates (Erbe *et al.* 2012). In this context, a subsidiary study (**Study 2**) was performed to design an *in silico* low-density SNP-chip (LD SNP-chip) to later perform imputation of genotypes to 50K and HD-chip densities. For this, and considering the SNP-chip genotypes available in the same half-sib Churra population of 1,686 animals, we implemented the imputation strategy based on Beagle_4.0 which had previously optimized in **Study 1**. Hence, the potential and accuracy of a virtual low-density chip were evaluated to perform genotype imputation to medium density (50K-chip) and high density (HD-chip). The designed LD SNP-chip included 2,935 markers common to the two panels and evenly distributed across the sheep genome. The strategy

followed to estimate the imputation accuracy was similar to **Study 1** with only a minor variation in the cross-validation procedures in which 10% and 30% of the reference population was masked in each iteration. The average imputation accuracy when imputing from LD to 50K-chip density was around 93.53% to 93.58% and when imputing from LD to HD SNP-chip, the imputation accuracy ranged from 88.3% to 86.52% (in both cases, depending on the masking level considered, 10% and 30%, respectively). Globally, and although the identified limitations need to be solved, the *in silico* LD SNP-chip described here, together with the implemented imputation strategy, provide an appropriate and affordable approach to obtain simultaneously: i) basic genetic information already analysed currently in commercial flocks (e.g. for parentage testing and *PRNP* gene genotypes) and, ii) also estimations of genomic breeding values that could guide breeding decisions and determine the first step of today´s sheep commercial flocks into the implementation of Genomic Selection.

5.1.2 Estimate genetic diversity parameters and update the extent of linkage disequilibrium (LD) and sample population estimations using the imputed HD-chip genotypes.

In **Study 1**, using the information obtained from the 336 individuals genotyped with the HD-chip (606,006 markers), we obtained the following gene diversity parameter estimations: gene diversity (H_E) = 0.321, proportion of polymorphic loci (P_N) = 0.964 and genetic distance (D) = 0.268. After quality control, the HD-chip dataset included 335 animals and 490,940 markers.

A total of 1,670 genotyped animals with 819,869,800 pairs of imputed SNPs were used to re-estimate the LD parameters in our Churra resource population. For markers up to 10 kb apart, the average r^2 was 0.275 ($D´$ = 0.860) and at more than 50 Mb inter-marker distance, the average r^2 was 0.006 ($D´$ = 0.169). The persistence of LD over short distances showed pronounced LD decay, with average r^2 values of 0.21, 0.17 and 0.14 for markers separated by 10, 20 and 30 kb respectively. The half-length of LD (r^2) was 0.024, which corresponded to a distance between SNPs of 1.03 Mb. The overall pattern of the LD (r^2) estimates across 1 Mb interval SNP-binning categories obtained here for the imputed HD-chip dataset was very similar to that presented for the 50K-chip in the

same population (García-Gámez *et al.* 2012). The LD estimates expressed in r^2 units were used to compare the extension of LD between Churra and other sheep populations analysed with the HD-chip by Kijas *et al.* (2014). These comparisons showed that the estimated LD r^2 values at 10 kb (0.2107) and 70 kb (0.088) in Churra sheep were slightly higher than in Merino and lower than in Poll Dorset, Suffolk and Border Leicester (Kijas *et al.* 2014). By plotting the updated HD-chip based estimates for Churra against the LD curve presented by these authors for the 50K-chip density, the decay of LD at the shortest distances in Churra would be very close to that of Qezel sheep. These observations seem to be directly related to the high level of genetic diversity reported for Churra based on the 50K-chip Sheep HapMap project analysis (Kijas *et al.* 2012) and updated here based on the HD-chip. Churra sheep breed shows a diversity level very close to that reported for the Merino breeds, which are among the most diverse sheep breeds presented in previous studies (Kijas *et al.* 2012, 2014). Even for the genetic distance ($D´$) parameter, our estimates suggested a slightly greater genetically divergence for Churra (0.268) than for Merino (0.262–0.266; Kijas *et al*. 2014). Globally, these results show that the HD-chip provides a substantially improved definition for LD estimations at short distances and that the Churra sheep genome showed an even a sharper decline in LD than that previously estimated based on the 50K-chip.

The *Ne* estimates reported in our study show the important decrease in population size that occurred 250 to 25 generations ago and the slighter decrease after the start of the breeding scheme in 1986 (de la Fuente *et al.* 1995). When comparing the estimates between the 50K-chip [considering the estimates for the 1 cM ~ 1 Mb ratio reported by (García-Gámez *et al.* 2012)] with the HD dataset-related estimates, we see very similar values for the range between 10 and 25 generations ago, whereas the most recent estimates based on the 50K-chip (one to five generations ago) appear to be over-estimated as result of the lack of identification of some crossing-overs due to the lower marker density. On the contrary, the 50K-chip *Ne* estimates for the most distant generations are lower than those estimated for the high-density datasets, probably as the result of the low inaccuracy of the 50K-chip to estimate the extent of LD at the shortest distances. The lower and more accurate estimate of inbreeding reported here for one generation ago, *Ne* = 83, is substantially lower than the estimates based on the 50K-chip (García-Gámez *et al.* 2012) (*Ne* = 128 as an average of different conversion ratios; *Ne* = 159 based on the 1 cM ~ 1 Mb ratio) and is also lower than the threshold estimated to ensure the viability in

the long term of an animal population (*Ne* = 100) (Meuwissen 2009) This updated information highlights the need for considering inbreeding control as a major point of the breeding program of the dairy sheep Churra population. Compared with the results of the HapMap population based on the 50K-chip (Kijas *et al.* 2012), 50 generations ago, when Churra had a *Ne* = 600, the updated estimate with the HD-chip for this time point was similar although slightly lower, *Ne* = 513. This estimate is intermediate between that reported for Merino (*Ne* = 833) and those values reported for other breeds showing a greater extent of LD and lower genetic diversity (Border Leicester, *Ne* = 242; Polled Dorset, *Ne* = 318) (Kijas *et al.* 2014). The imputation did not have any effect on the most recent estimates. This appears to correlate with the lower estimates of LD in *D´* units observed for the HD-chip reference population dataset when compared with the estimates of imputed HD-chip genotypes and the influence of the low accuracy to estimate *Ne* in very distant time points.

The chromosome showing the highest number of ROHs was OAR2 (14,025 ROHs) with a significantly higher number of ROHs than for OAR1 and OAR3, which had similar lengths. Our ROH analyses showed that, on average, approximately 7% of the whole genome was under ROH segments, with the highest coverage shown by OAR24 (10.4%) and OAR26 (9.7%). The chromosomes showing the lowest coverage of ROHs were OAR1 (4.6%) and OAR7 (4.9%). The estimation of the total length of ROHs per individual showed that 98% of the individuals had a total ROH length greater than 50 Mb, approximately 70% of the individuals had a total ROH length greater than 100 Mb and 0.007% had a total ROH length greater than 500 Mb. Based on the distribution of ROHs in the different bins roughly 88% of the ROHs had lengths of less than 5 Mb, and 0.008% of the ROHs spanned more than 20 Mb. This distribution of ROHs observed for Churra sheep across the different chromosomes shows a similar pattern to that described by Al-Mamun *et al.* (2015) and focused on both pure and cross-sheep breeds with genotypes for the 50K-chip. However, the number of ROH segments was much higher in our estimation, which could be attributed to the higher genotyping density of our dataset. The distribution of the sum of ROHs per individual also showed a similar distribution to that of pure breeds reported by Al-Mamun *et al.* (2015). All of the individuals had ROHs greater than 1 Mb and approximately 99% of individuals had ROHs longer than 5 Mb, which are higher estimates than those reported for Border Leicester, Merino and Poll Dorset pure breeds, for which approximately 88% of the animals had at least one ROH

longer than 5 Mb (Al-Mamun *et al.* 2015). Considering the ROHs that were common to different individuals, a total of 12,238 ROH consensus regions were identified in our population. Interestingly, the OAR8 region including the ROH region with the largest number of occurrences overlaps with a region identified as under selection (related to marker s20065, with position 32,159,065 bp according to the updated reference genome Oar_v3.1) in the analysis of the Sheep HapMap project dataset reported by Kijas *et al*. (2012). According to SheepQTLdb (Hu *et al.* 2016) this region is related to QTL for gastrointestinal nematode resistance (Crawford *et al.* 2006) and fat carcass content (Cavanagh *et al.* 2010). This region includes two genes: *BVES* (blood vessel epicardial substance), which is expressed in cardiac and skeletal muscle and could play an important role in the development of these tissues, and *LIN28B* (lin-28 homolog B), which is a suppressor of microRNA biogenesis and has been reported as the first genetic determinant regulating the timing of human pubertal growth and development (Ong *et al*. 2010).

The inbreeding coefficients estimated based on the HD-chip reference genotypes for the animals born from 2001 to 2008 showed mean values of F_{Mol} = -0.0228 (SD = 0.011), F_{ROH} = 0.042 (SD = 0.019), F_{ROH-Q} = -94.127 (SD = 43.595) and $F_{ROH-E}{}^{m}$ = -0.375 (SD = 0.239). Considering the average values of the standardized inbreeding coefficient estimates for the entire population and according to the year of birth, from 2001 to 2008, we observed that all of the genomic-based coefficients resulted in higher estimates of inbreeding than did the pedigree-based coefficient. The overall profile of F_{Mol} suggested a higher inbreeding level than that of F_{ROH}, whereas the two inbreeding coefficients, considering the lengths of the ROHs F_{ROH-Q} and $F_{ROH-E}{}^{m}$ which according to Gomez-Raya *et al.* (2015) are related to recent inbreeding, suggested the highest estimates of inbreeding. In any case, the general profiles of the genomic-based inbreeding estimates obtained across the years considered showed the same general pattern, with a slow and continuous increase in the inbreeding level from 2001 to 2008, during which 2003 showed a more pronounced increase, which could be related to the introduction in 2003 of udder and body morphology traits as selection objectives in the breeding program (de la Fuente *et al.* 2011). There were also some differences depending on the estimation approach implemented. Hence, there was a decrease in the inbreeding estimates detected in 2002 by F_{Mol} and F_{ROH} but not by the two coefficients related to recent inbreeding, F_{ROH-Q} and $F_{ROH-E}{}^{m}$. As previously shown by Garcia-Gamez *et al.* (2012), based on the use of the 50K-chip, our comparison of inbreeding coefficients, indicate that the genomic-based

estimates are more reliable than are the pedigree-based coefficient. In both cases, by using the 50K-chip or HD-chip genotypes, we could identify inbreeding that was not identified on the basis of the available pedigree. Using the same approach to estimate these genomic-based estimates, the higher accuracy provided by the HD-chip results in higher inbreeding estimates than with the estimates based on the 50K-chip. Based on the HD-chip imputed dataset, we also estimated F_{ROH} to exploit the information from the ROH analysis previously presented. In contrast to F_{Mol}, this coefficient ignores fragments consisting of a single or a few contiguous homozygous SNPs. In our Churra population, comparison of the standardized estimates for these two coefficients showed that the estimates of inbreeding were lower when the total ROH length was considered in the estimations. Hence, as a general estimation of inbreeding in our resource population, we believe that the average estimate of F_{ROH} = 0.042 (SD = 0.019) could be the most appropriate. Even based on a subset of animals from the selection nucleus, this estimate was less than the critical level of 0.0625 suggested by Li *et al.* (2009) considering the inbreeding level obtained when mating cousins, which suggests that the inbreeding levels in Churra sheep are under-controlled in the general population.

As suggested by Gomez-Raya *et al.* (2015) ROHs information could be further exploited by considering the length of ROH fragments. ROHs due to recent inbreeding will tend to be longer because there has been little opportunity for recombination to break up segments that are identical by descent (Kirin *et al.* 2010). In our population, the two coefficients related to recent inbreeding $F_{ROH\text{-}Q}$ and $F_{ROH\text{-}E^m}$ coefficients showed, after standardization, higher inbreeding estimates than did the general F_{ROH}. This could be easily explained by the active selection scheme of the Churra sheep population. In any case, as noted earlier, the inbreeding levels estimated in this study did not show a risk to the genetic diversity levels of the population, which according to the present study, were quite high compared with other sheep populations.

5.1.3 Gene-mapping analyses based on Genome-wise association (GWA) and Linkage Analysis for indicator traits of parasite resistance in sheep.

The imputed HD-chip genotypes obtained with Beagle_v3.3.2 for the whole population of Churra sheep analysed by Garcia-Gámez *et al.* (2012) were used to perform a high-

density GWA analysis about GIN resistance traits, as reported in **Study 3**. Considering that from the global population of 1670 ewes and their 16 sires, a total of 533 had phenotypic measurements for the indicator traits of FEC and IgA, we performed a GWA analysis for these two traits using the ProbaABLE (Aulchenko *et al.* 2010) software. This software a software specifically developed to perform GWA analysis from imputed genotypes. Following the protocol suggested by the developers of this program, the first analysis in GenABLE (Aulchenko *et al.* 2007) in order to estimate the matrix of variances and covariances, which provides a test of association between the phenotype and the genetic markers in the samples of related individuals. The second step of the association analysis was performed with ProbABLE, using as genotypic data for analysis, the correct imputation probability files for the SNP obtained after imputation using Beagle_v3.3.2. The GWA analysis for the IgA phenotype identified one single significant SNP at the 5% chromosome-wise level located on OAR15 (P_c-value corrected = 5.63), while no significant associations were detected for this trait at the 5% genome-wise level. The marker that showed evidence of influencing IgA levels is oar3_OAR15_24870525. The estimated allelic substitution effect for this SNP was 0.297 ± 0.063 units of the YDs analysed (0.382 phenotype standard deviations). According to the database Sheep QTLdb (Hu *et al.* 2016), the position of this QTL is within the estimated confidence interval (18.3-30.5 Mb) of a QTL influencing FEC previously reported in a backcrossing population Red Maasai x Dorper (Silva *et al.* 2012). The limitation of significant results in this GWA study can be explained by the low statistical power of the study due to the limited number of animals analysed. In order to exploit the half-sib structure of the population, in a complementary way to this GWAS-type analysis, we carried out additional analysis based on LA.

5.1.4 Replication of QTL detection with imputed HD-chip genotype and the high-resolution analysis of target QTL regions with Whole Genome Sequencing (WG-seq).

Study 4 included in this thesis focused on implementing a high-density QTL mapping for GIN resistance traits in the same population previously analysed by Atlija *et al.* (2016). For that, the imputed HD-chip genotypes generated in **Study 1** were analysed with the

QTLMap software, whose use for gene mapping in the considered Churra sheep population had been optimized in previous studies of the MEGA-ULE group (García-Gámez *et al.*, 2013; Atlija *et al.*, 2016). **Study 4** has proven that the mapping accuracy of QTL identification analyses can be substantially increased by using imputed HD-chip genotypes, with the highest reduction observed (up to 99% of refinement) for the CI length observed for the OAR6 QTL influencing LFEC, which was selected as the most promising QTL region-based the initial report of this effect by Gutierrez-Gil *et al.* (2009) and its replication by Atlija *et al.* (2016). This study has also shown that despite the use of a higher marker density, no new QTLs underlying the studied phenotypes were encountered, suggesting that for complex traits, in this population, the 50K-chip may be sufficient to initially identify QTL regions, whereas the HD-chip offers advantage about the mapping accuracy by highlighting a smaller interval where positional candidate gene need to be evaluated for functional relationship with the QTL phenotypic effect. However, in the case of the LFEC QTL on OAR6, and about the QTL peak location, the results obtained with the HD-chip genotypes showed certain discrepancies between the across-population and the within-family analyses, which difficulted the selection of the confidence interval to study in the subsequent high-resolution study based on WG-seq dataset of the selected segregating trio. Further study should be addressed to understand the nature of these mapping discrepancies better.

Concerning the concordance with other studies, we only found overlapping with the QTL regions reported by Atlija *et al.* (2016) in the previous study of the same population, and using the same analysis methodology, with the 50K-chip. The lack of concordance with other GIN resistance mapping studies carried out in lambs suggests, as previously suggested by Atlija *et al.* (2016), that these QTL regions may be exclusively related to GIN resistance in adult sheep. This could be confirmed by conducting future similar studies involving adult animals from other sheep breeds.

On the other hand, the high-resolution study of the genetic variability identified by WG-seq in the OAR6 LFEC QTL region highlighted, after the corresponding filtering steps, genetic variants in the *RASSF6* and *AFM* genes as potential candidate mutations to underlie this QTL effect. The gene *RASSF6,* which is included in the InnateDB database (Breuer *et al. 2013)*, is involved in tumour suppression in humans and has been also found to dictate the degree of inflammatory response against the respiratory syncytial virus (Allen *et al*. 2007). Note that in the present study, a missense variant was identified as a

potential functional variant within the *RASSF6* gene. On the other hand, the gene *AFM* belongs to the albumin gene family. After variant annotation, we found four missense variants in the *AFM* gene. These missense mutations might be significant with regards to nematode infection in adult sheep. In relation to the OAR8 QTL influencing LFEC, the refinement of the estimated CI provided by the HD-chip genotype scan allowed us to identify the *COL12A1* gene as the most promising positional candidate gene as it was identified as immune-related gene due to its inclusion in the InnateDB database (Breuer *et al.* 2013). Further research on this QTL region would be needed, for instance, based on the WG-seq of a specific segregating trio, to provide a list of potential causal mutations to be confirmed in commercial populations.

5.2 Objective 2: Transcriptomic study of resistance to GIN infections in adult sheep

In this objective, we tried to decipher the immune mechanisms activated in adult sheep during infection with *T. circumcincta* through the transcriptomic analysis of the two main tissues targeted by the infection, the abomasal mucosa and the abomasal lymph node. Based on the accumulated FEC estimation carried out on the Churra sheep subjected to the first experimental infection with *T. circumcincta*, six animals were classified as "susceptible" and six as "resistant". The mean accumulated FEC in the susceptible group was 5594 ± 2661 eggs per gram (epg) and 308 ± 338 epg in the resistant group. A total of 24 mRNA samples, 12 from abomasal mucosa and 12 from the abomasal lymph node, were considered for massively parallel sequencing.

5.2.1 The different gene expression responses activated in adult sheep between resistant and susceptible animals against *T. circumcincta* infection.

A preliminary study, differential transcriptomic analysis (**Study 5**) was carried out with abomasal mucosa samples obtained from two groups of twelve sheep in which six classified as "susceptible" and six as "resistant" to GINs after experimental infection with *T. circumcincta.* The bioinformatics workflow analysis included the assessment of raw

sequence data quality using FastQC, the alignment against the sheep reference genome (Oar_v.3.1), the quantification and normalization of gene expression performed with Cufflinks and finally the differential expression analysis performed with two different R-based packages, DESeq2 and edgeR. DESeq2 identified one differentially expressed (DE) gene, *SYT8* (P_{adj} = 0.020) whereas the edgeR analysis found 18 DE genes among which *SYT8* was the second highly expressed gene (FDR_{adj} = 0.0023). *SYT8* is involved in trafficking and exocytosis of secretory vesicles in non-neuronal tissues. Many of the genes highlighted by the edgeR are related to muscular excitation and contraction whereas one belongs to the Major Histocompatibility Complex. Because one of the proposed mechanisms of parasite resistance in sheep is the increase in peristalsis, some of the genes highlighted by the edgeR analysis may be further investigated in future studies.

The differential transcriptomic analysis (**Study 6**) was further carried with abomasal lymph node tissue in addition to abomasal mucosa extracted from GIN infected sheep. The FastQC analysis showed that the quality of the sequenced reads for all the samples in the two tissues under analysis was of high quality and thus, no trimming was performed. Across all the samples on average, around 72.3% and 79.2% of reads were uniquely mapped against the reference genome Oar_v3.1, for abomasal mucosa and abomasal lymph node tissues respectively. The aligned reads were sorted and counted against the list of annotated genes from the sheep reference genome Oar_v3.1. Out of the 27,054 annotated genes in this reference genome, on an average, around 16,210 (59%) genes in abomasal mucosa tissue sample data and 16,808 (62%) genes in abomasal lymph node tissue samples had at least one raw read count. The distribution of the expressed genes across the different expression levels considered was also very similar in the two tissues and conditions (~1.31% of the genes classified as highly expressed genes; ~43% of the genes showing an intermediate expression level; ~55% of the genes with a low expression level). In the abomasal mucosa, 179 and 180 genes were identified as highly expressed genes for the resistant and susceptible samples respectively. These highly expressed genes involved approximately 1.1% of the total FPKM in each group and were considered core genes (179 common core genes for the two compared groups, 0 specific core genes for the resistant group and 1 specific core gene for the susceptible group). In the lymph node, 251 and 214 genes were identified as highly expressed genes for the resistant and susceptible samples, respectively (201 genes identified as core genes in the two conditions, 50 specific core genes of the resistant group and 13 specific core genes

of the susceptible group). Considering the two tissues, we found 123 common genes in the highly expressed category (> 500 FPKM), 56 highly expressed genes specific to abomasal mucosa and 78 highly expressed genes specific to abomasal lymph node tissues. By performing GO enrichment analysis with these common and specific highly expressed category genes, we found that the enriched terms resulting from the genes common to both tissues were related to basic physiology like translation termination, translational elongation, mRNA catabolic process, RNA binding, etc. The GO analysis for the core genes specific to abomasal mucosa highlighted terms such as digestion, ATPase activity, hydrolase activity, ATP hydrolysis coupled proton transport. Whereas in the analysis of the core genes specific to the abomasal lymph node, some of the enriched terms were muscle contraction, collagen fibril organization, protein binding, extracellular matrix, etc.

After performing the differential expression analysis with the abomasal mucosa RNA-seq data, we found 33 DEGs (adjusted $P < 0.05$) using DESeq and no significant DEGs with EdgeR (FDR < 0.05). Out of these 33 significant genes, eight were up-regulated in resistant sheep, and 25 were up-regulated in susceptible sheep. The differential expression analysis of the abomasal lymph node with DESeq (adjusted $P < 0.05$) and EdgeR (FDR < 0.05) showed, respectively, a total of 261 and 125 DEGs. Among them, 106 genes were commonly identified as DEGs by both software and were considered as GIN-activated DEGs. Out of these 106 common GIN-activated genes, 71 were up-regulated in the resistant sheep and 35 were upregulated in susceptible animals. The GO enrichment analysis performed for the GIN-activated DEGs with upregulated expression in resistant sheep shows nine significant terms in the biological process database, two of them related to cytokine response (*cytokine-mediated signalling pathway, cellular response to cytokine stimulus*), eight significant terms in the cellular component database and no significant term in the molecular function database. On the contrary, for the GIN-activated DEGs showing upregulated expression in susceptible sheep, only three significant terms included in the cellular component database were identified. The two other enrichment analyses performed, the KEGG pathway and disease association analyses, only identified significant terms for the GIN-activated DEGs showing upregulated expression in resistant sheep. Among the three significant terms identified in the KEGG pathway analysis, we consider worth mentioning the *PPAR signalling pathway*, whereas the four significant terms identified in the disease association analysis

were related to the studied phenotype (*gastrointestinal neoplasms, intestinal/ gastrointestinal diseases, inflammation*).

The results of the differential expression analyses performed in the two tissues studied suggest that at day 7 post-infection there is not a clear differential gene expression response in the abomasal mucosa (due to the lack of DEGs identified by EdgeR). However, a differential response in the abomasal lymph node was observed, based on the identification of 106 genes commonly identified as DEGs by the two software. We acknowledge that the results from the differential expression analyses performed with EdgeR and DESeq may include a similar fraction of false negative, as they both rely on a negative binomial model and use the false discovery rate procedure (Benjamini and Hochberg 1995) to adjust for multiple testing. In this regard, Zhang *et al*. (2014) suggested that taking the intersection of DEGs from two or more tools is recommended if the number of false positives is a major concern in the study. Following this, we considered for further analyses those genes that were identified as DEGs by the two implemented methods, reducing the initial number of DEGs from 261 and 125, respectively for DESeq and EdgeR, to 106 genes defined as GIN-activated DEGs.

Hence, focusing on the 106 GIN-activated DEGs identified in the complete analysis of the lymph node samples, some of the significant terms identified by the enrichment analyses showed a clear correspondence with immune response mechanisms. In particular, the two GO terms related to cytokines highlighted by the GO analysis of genes upregulated in resistant ewes (*cytokine-mediated signalling pathway and cellular response to cytokine stimulus*) were related to the same five genes (*PALM3, DUOX2, PPARG, PF4, IL5RA*). The role of *PPARG, DUOX2* and *IL5RA* genes about the immune response have been previously reported in different nematode infection studies (Ingham *et al.* 2008; Gossner *et al.* 2013; Chen *et al.* 2017). In our study, the *PPARG* gene also supported the identification of the *PPAR* signalling pathway in the KEGG analysis, together with the *RXRG* and *AQP7* genes. The *PPARG* gene was also linked to many of the significantly related terms highlighted by the disease association analysis (gastrointestinal neoplasms, intestinal diseases, inflammation and gastrointestinal diseases). The *IL5RA* gene also supported the identification of the term inflammation as enriched in the disease association analysis. Some of the DEGs identified in our study have been previously identified as DEGs by other sheep gene expression analysis in relation to gastrointestinal nematode infection. The 10 out of the 106 GIN-activated genes

identified in our study have been previously reported to show a modified expression due to GIN infection: *ITLN, LYZ, LOC443162* (galectin 14), *LGALS4* (galectin 4), *CLCA1*, *ALPL, PDZK1IP1, PPARG, KRT5, IL5RA*. Note that for the gene annotated as *ITLN* in the sheep genome, the ortholog analysis clearly shows that it corresponds to the *ITLN2* gene, which has been identified by other studies as activated by the GIN infection. Among the list of ten genes commonly identified by our study and other authors to be responsive to GIN infection, *PPARG, LYZ,* and *IL5RA* are directly related to the inflammatory response. *PPARG* encodes for the peroxisome proliferator-activated receptor gamma, which is a ligand activated transcription factor that regulates adipocyte differentiation and glucose homeostasis, but it has also been recognized as playing a key role in the immune response through its ability to inhibit the expression of inflammatory cytokines and to direct the differentiation of immune cells towards anti-inflammatory phenotypes (Martin 2010). In our study, the *PPARG* gene was found to be upregulated in resistant ewes compared with susceptible ewes. A modified expression pattern of the *PPARG* gene with the infection by *T. colubriformis* and *H. contortus* in sheep has already been reported by Andronicos *et al.* (2010). A study carried out on naïve Perendale lambs suggested this gene plays a role in coordinately regulating genes more highly expressed in the intestine of the susceptible lambs (Keane *et al.* 2006). The interleukin gene *IL5RA* was also a GIN-activated gene up-regulated in resistant sheep. This gene supported the enriched GO terms cytokine-mediated signaling pathway and cellular response to cytokine stimulus and the disease association related term Inflammation. This gene was also found to have an increased expression in resistant Scottish Blackface lambs to *T. circumcincta* by Gossner *et al.* (2013). The *LYZ* gene encodes lysozyme, a protein with antibacterial activity. In our study this gene was the most highly up-regulated GIN-activated gene in susceptible sheep. This observation agrees with the work of Knight *et al.* (2011) who reported, in the abomasal mucosa, downregulation of members of the gastric lysozyme family (*LYZ 1A, 2A, 3A* and *4A*) in immune versus naïve sheep at days 2 and 5 post-challenge with *T. circumcincta*. Other studies, however, have found genes of the lysozyme family to be up-regulated in abomasal epithelial extracts from previously infected sheep versus naïve sheep (Athanasiadou *et al.* 2008; Knight *et al.* 2011). Another GIN-activated gene up-regulated in resistant sheep based on our study, *CLCA1*, has also been identified as up-regulated in immune lambs in the study reported by (Knight *et al.* 2011) and in the *H. contortus* challenged yearling lambs analyzed by Rowe *et al.* (2009). The encoded protein

of this gene is thought to act as a multifunctional signaling protein, including an early modulator of immune responses by regulation of cytokines (Erickson *et al.* 2016). Proteins of the CLCA family may contribute to parasite expulsion by being responsible for mucus hydration across the gut epithelium and smooth muscle contraction (Rowe *et al.* 2009). Another transcript related to mucous cells, *ITLN*, was the third most highly up-regulated in resistant sheep in our study. The intelectin 1 and 2 are protein-coding genes related to carbohydrate binging. Other studies have already reported an increased expression of this gene with the infection response of lambs to *T. circumcincta* (French *et al.* 2008; Knight *et al.* 2011) and *H. contortus* infection (Rowe *et al.* 2009). Overall, our study supports the role of some genes such as *ITLN2, CLCA1* in adult sheep GIN resistance. These two genes have been previously reported to be increased in immune lambs and to be up-regulated in resistant adult sheep. Whereas the expression of these genes in lambs was found to be affected by infection in abomasal mucosa, our study did not identify an altered pattern in that tissue but only in the lymph node. Three genes identified in our study as GIN-activated, and up-regulated in resistant sheep, belong to the galectin family, *LOC443162* (Gal-14), *LGALS4* (Gal-4) and *LOC101102156* (Gal-9). Galectins mediate innate and adaptive immune functions by modulating the activity of complement receptor 3, macrophage and dendrocyte adhesion to lymphocytes (Vasta *et al.* 1999). There was a significant up-regulation of Gal-4 in challenged naïve yearling lambs after infection with *T. circumcincta* L3 (Knight *et al.* 2011). The expression of Gal-14 was maximum at day 10 post-challenge in yearling Scottish Blackface lambs previously infected with *T. circumcincta* (French *et al.* 2008). Our list of GIN-activated genes also includes some genes belonging to the same gene families reported in other relevant GIN studies. Some of these include genes from the matrix metallopeptidase (MMP) family (*MMP28*) and solute carrier family genes (*SLC22A4, SLC25A34* and *SLC38A2*). The role of these genes in inflammatory diseases has been previously described. The high expression of *SLC30A2* in susceptible sheep in response to *T. circumcincta* infection has been reported in Scottish Blackface lambs (McRae *et al.* 2016), *H. contortus* infection shows a significant impact on the expression of the *SLC2A3* gene in Canary sheep breeds (Guo *et al.* 2016). It is noteworthy that there is a list of about 86 genes identified as GIN-activated in our study that have not been reported or that do not belong to gene families considered as responsive to GIN infection in previous studies. Interestingly, for some of these genes we found connection with the immune response (*PGC, SFTPD, TUBA4A, SST, BPIFB1, PF4, B4GALNT2, JCHAIN, AQP7, KLHL25,*

NEDD4, ANO6). We think that these DEGs might indicate genes that are specifically activated in adult sheep and not in lambs. Also, the differences among the different studies regarding the species responsible for gastrointestinal parasite infection, the different experimental approaches (natural vs artificial), the post-infection sample extraction timings, or the environmental conditions (dry vs humid climates) are other major factors that could explain discrepancies related to the specific genes activated during GIN infection.

5.2.2 Identification of potential causal mutations related to GIN resistance in adult sheep by exploiting RNA-seq.

Although the sensitivity to identify a variant and the coverage at a portion of the genome with RNA-seq data is much lower than that of whole genome or whole-exome sequencing (Piskol *et al.* 2013), variant calling with RNA-seq data is less expensive and a proper experimental design can overcome this shortcoming. In our case, we adopted an analysis workflow previously described to identify variants across the sheep milk transcriptome (Suárez-Vega *et al.* 2017) to perform a variant calling analysis in the lymph node RNA-seq dataset previously analysed for differential expression analysis. This analysis was focused only on abomasal lymph node tissue because our previous work had not identified any confirmed differentially expressed genes (DEGs) in the abomasal mucosa samples. The analysis of the genetic variability of the studied animals was carried out by considering the 6 resistant and 6 susceptible sheep as a single group, as it has been shown that a higher number of samples considered during variant calling analysis results in a higher coverage and confidence in variant calls (Piskol *et al.* 2013).

Initially, a preliminary genetic variant study (**Study 7**) was carried out in genes identified as DEGs between the abomasal lymph node transcriptome of six resistant and six susceptible adult sheep to *Teladorsagia circumcincta* infection. After aligning against the ovine reference genome (Oar_v3.1), around 80.27% paired reads per sample were uniquely mapped against the reference genome and around 8.5 % paired reads were mapped to multiple loci. After performing the variant calling and variant filtering, we found 1,326,960 common variants considered as high-quality variants across the whole genome. From these high-quality variants, we extracted 6,168 variants (6,104 SNPs, 30

insertions and 34 deletions) mapping within the 106 DEGs. Among these, 332 variants (329 SNPs, 3 insertions and 2 deletions) were predicted to be of relevant functional impact ("High" and "Moderate"). The functional annotation of these variants with the VEP, which included 109 novel variants revealed a total of 471 functional consequences distributed across 60 genes. From these, the SIFT algorithm detected 50 deleterious variants most were missense consequence and two were splice region variants. These deleterious variants were included in a total of 15 genes. Four of these genes, *BPIFB1* (Zhou *et al.* 2017), *KRT20* (Sen *et al.* 2012), *SLC38A2* (Carter 2012) (Carter 2012) and *FNDC1* (Sigdel *et al.* 2015) are associated with the immune response. Another gene harboring a missense deleterious variant was *MMP28*, which has been reported as responsible for cell proliferation in response to skin injury (Saarialho-Kere *et al.* 2002). Other genes harboring "High" impact variants were *LGALS4* (splice acceptor variant), *SLC38A2* (stop lost), *ASIC3* (stop gained and splice acceptor variant), and *SULF1* (frameshift variant).

The preliminary study was followed by a variant calling analysis (**Study 8**) across the whole transcriptome. In this study, the genetic variability of QTL and candidate gene regions related to GIN resistance was investigated by identifying the potential functional variants related to phenotypic variation for GIN resistance. To ensure the quality of the variants identified, we performed the variant calling analysis with two different software packages, GATK and Samtools, such as in Gutiérrez-Gil *et al.* (2018). Considering SNPs, insertions and deletions, the number of common variants detected by GATK and Samtools packages, which were considered as high-quality variants, was 1,326,960. Of them, a total of 1,086,508 variants (81.8%) were mutations already described in SNPdb (version ovis_ariesRelease91), whereas 240,452 were novel variants, without an associated *rs* number. From the high-quality variants, we extracted 224,191 variants mapping within 230 QTL regions related to the resistance of sheep to parasites and 98,576 variants mapping within 1,892 genes previously reported in the literature as differentially expressed genes about GIN resistance in sheep, which will be referred from now on as candidate GIN-activated genes. A total of 16,184 genetic variants were shared between the QTL regions and the considered candidate genes. For the whole transcriptome and the QTL regions, the distribution of the variants was very similar, with the percentages of intronic, intergenic variants and non-coding variants (about 40%, 24% and 27%, respectively) being much higher than the percentage of coding variants (about 6%). The

lack of intergenic variation determines the differences in the distribution observed for the variants detected within the candidate GIN-activated genes.

Based on the predicted impact on the protein function provided by the annotation analysis carried out, the majority of the variants identified across the transcriptome were modifiers (93.39%), whereas the percentage of variants classified as low, moderate and high impact variants were 4.24%, 2.29% and 0.06%. These proportions were very similar for variants detected within the considered QTL regions (93.00%, 4.35% 2.57% and 0.06%, for modifier, low, moderate, and high impact variants respectively), whereas for the candidate GIN-activated genes the percentage of variants included within the modifier category was slightly lower (88.86%) and slightly higher for each of the rest of categories (7.59%, 3.45% and 0.08%, for low, moderate, and high impact respectively).

We identified 186 and 105 high impact variants, and 744 and 448 missense deleterious variants within the QTL and the candidate GIN-activated genes regions, respectively. The percentage of those variants that were novel within each category ranged between 32% and 18%. These potential functional variants within the QTL regions were distributed across a total of 424 genes, 111 of which were included in the considered reference databases of immune-related genes. On the other hand, the 553 potential functional variants located within the candidate GIN-activated genes were distributed across a total of 336 genes, 132 of which were included in the considered reference databases of immune-related genes.

The enrichment analyses performed with the 111 immune-related genes harbouring mutations with potential biological impact within the considered QTL regions identified a total of 41 significant GO terms from the biological process database. These terms were grouped, at the most superior hierarchical level, under the following five significant terms: *regulation of apoptotic process, cell adhesion, cell surface receptor signalling pathway, inflammatory response* and *enzyme-linked receptor protein signalling pathway.*

In addition, for the list of 132 immune-related GIN-activated genes harbouring mutations with potential biological impact, the disease association enrichment analysis highlighted five disease terms (*Inflammation, Adhesion, Necrosis, Immune System Diseases* and *Infection*).

Regarding the results of the variant annotation reported here, the large percentage of the variants identified in intergenic regions, introns or non-coding RNA (approximately 83%) is in agreement with the percentage of variants identified within non-coding regions

through the transcriptomic analysis of the milk somatic cells in sheep (approximately 71%) (Suárez-Vega *et al*. 2017). Indeed, the distribution pattern of variants across different genomic regions (introns, non-coding RNA, coding, intergenic and UTRs) is very similar to that reported in humans by Piskol *et al*. (2013) for RNA-seq. As indicated by these authors, the large number of intronic variants in our RNA-seq analysis (40% for the transcriptome and QTL regions, and 75% for the candidate gene regions) can be explained by (i) the fact that poly(A)+ RNA-capturing protocols can capture premature mRNAs that still contain introns, or (ii) by the presence of one type of alternative spliced RNAs, intron retention; in any case, introns compose a much larger fraction and are more variable than exons (Piskol *et al.* 2013). Alternatively, the identification of variants in intergenic regions may indicate the presence of non-coding RNAs that have not yet been annotated in the sheep reference genome and which are increasingly being accepted as one of the factors contributing to the complex physiology and architecture of mammalian species (Dhamija and Menon 2018). These observations support the need to improve the annotation of the reference genome further.

The systematic search of transcriptome variants and the filtering steps implemented here have allowed us to highlight potential genetic variants that could be considered as potential causal variants of phenotypic effects in relation to *T. circumcincta* sheep resistance. We would highlight the mutations located in the genes *IFI44*, *COL3A1*, *PLXNC1*, *SEMA3A*, *CD74*, *MAP1S*, *CSF1R*, *CCNDBP1*, *BNIP2*, *IL20RA*, *PRKDC*, *KLHL21*, *LAMC2*, *AEN*, *NUMBL*, *TCIRG1* and *HRH1*, as they were identified as single candidate casual mutations for fifteen QTL reported for parasite resistance.

Regarding the specific five terms highlighted by the GO enrichment analysis of the genes located in QTL regions harbouring the list of mutations with potential biological impact, *cell surface receptor signalling pathway* was the significant enriched term including the highest number of genes (32). During GIN infection in sheep, the cytokine-mediated multifactorial cellular signalling has been found to guide the cell-mediated localized response (Sykes and Coop 2001). In *T. circumcincta*-infected sheep, the secretion of Th2 cytokines was associated with resistance and the activation of Th1/Th7 cytokines was linked to susceptibility (Gossner *et al*. 2012, 2013). Through further linking each target QTL region with the corresponding enriched terms supported by the genes harbouring the potential causal variants (e.g. *cell adhesion*, *cell surface signalling pathway*, etc.) our results might help to identify the biological function related to the role of the QTL effect with the host response to GIN resistance. Globally, the provided list of candidate

mutations in relation to previously identified QTL may be of great interest to design future studies aiming to identify genetic markers that could be directly used in genetic selection to improve parasite resistance in sheep. Ideally, the confirmation of the relationship of these mutations with the corresponding QTL should be tested in the population used to map the genetic effect. If not possible, general association analyses in large commercial populations could be also an appropriate approach to test the link of these mutations to indicator traits of parasite resistance across different sheep breeds and with different parasite species.

Among the five disease terms (*Inflammation, Adhesion, Necrosis, Immune System Diseases* and *Infection*) identified as significantly enriched by the analysis performed for the 132 immune-related GIN-activated genes harbouring potential causal variants, *Adhesion* was the term including the highest number of genes (25). Adhesion molecules play a key role in cell-cell and cell-extracellular matrix interactions, which are necessary for proliferation and differentiation of both eosinophils and other inflammatory cells during their activation at inflammatory sites (Stevenson *et al.* 2001). Cell adhesion and cell-matrix adhesion have been reported as important biological processes involved in the development of resistance to primary infection with *Haemonchus contortus* in Merino sheep (Zhang *et al*. 2019), whereas cell adhesion molecules pathways were identified as highly expressed in resistant goats to GIN infection (Bhuiyan *et al.* 2017). Concerning the term *Inflammation*, Khan and Collins (2004) found that the development of inflammation in the intestinal mucosal surface was considered as an early immune response to GIN infection in sheep. In this sense, Amarante *et al.* (2007), based on the study of three sheep breeds naturally infected with *Trichostrongylus colubriformis*, suggested that the inflammatory cells impair the parasite's establishment, development, and survival. Finally, it has been shown that, at the infection site, *Trichostrongylus* larvae induce intestinal epithelial cell necrosis which determines the release of Interleukin 33 (IL33) protein (Andronicos *et al.* 2012). For the gene coding that protein, *IL33* (which supported the *Necrosis* enriched term) our filtering approach identified a single potential functional variant (*rs410259751*) that should be considered as a candidate for future confirmation of genetic markers underlying the host response to GIN infection in sheep.

5.2.3 Identification of long noncoding RNAs related to GIN resistance in adult sheep after *T. circumcincta* infection

Long noncoding RNAs (lncRNA) are found to play multiple biological functions and their activity also varies with the developmental stage of cells and tissues and also under different disease states (Ma *et al.* 2012). It was found that the lncRNA function by regulating or interacting with other molecules like RNA, DNA and proteins (Ma *et al.* 2012; Schmitz *et al.* 2016). Thus, inferring their role will elucidate the possible mechanisms of complex biological processes related to different metabolic disorders and disease conditions. But compared to regular protein-coding genes, the sequence and secondary structures of lncRNA transcripts are usually not conserved (Mercer *et al.* 2009; Pang *et al.* 2009). This makes it challenging to investigate lncRNA function directly based on their physical structure. The recent availability of different bioinformatics tools like CNCI (Sun *et al.* 2013), PLEK (Li *et al.* 2014), PLAR (Hezroni *et al.* 2015) and FEELnc (Wucher *et al.* 2017) enabled to predict lncRNAs from unknown transcripts.

In the present study (**Study 9**) on the identification of lncRNAs in the abomasal lymph node from GIN sheep. After trimming the adapters, primers and poly (A) sequences for the RNA-seq datasets generated for the 12 samples, around 2.1 billion reads ranging from 128,603,512 to 222,630,451 million reads per sample were filtered for further analysis. On average, 89.8 % of the sequences were aligned against the Ovis_aries_v3.1-r95 reference genome. Finally, after performing annotation guided transcriptome assembly and merging all the 12 samples, 77,039 transcripts were assembled corresponding to 44,203 genes. After filtering out unwanted and protein-coding transcripts, the FEELnc program based on its potential coding score, we detected 9,105 transcripts as putative lncRNAs and 729 as mRNAs. Out of these 9,105 lncRNA transcripts, 2,092 lncRNAs transcripts were classified as novel by gffcompare software.

After excluding the outlier sample detected, we performed differential gene expression analysis and found 3,148 differentially expressed genes between 6 resistant and 5 susceptible sheep. Out of which 1,635 were upregulated, and 1,513 were downregulated in resistant sheep. By investigating the DEGs associated with the lncRNA transcripts, we found 683 lncRNA transcripts associated with 457 DEGs, out of which 263 lncRNA (153 DEGs) transcripts were considered as novel. After performing sequence similarity with

BLAST, 533 out of the 683 identified lncRNA transcripts were aligned with 99% similarity. Among them, 167 were novel transcripts and no overlapping genes were found on these novel transcripts. The rest 367 lncRNA transcripts included 301 overlapping genes. Some of the DE genes linked to lncRNA such as *IRF1* (Araujo *et al.* 2009), *CCL14* (Araujo *et al.* 2009; Zhang *et al.* 2019), *KRT8* (Athanasiadou *et al.* 2008) and *LGALS3* (Nagaraj *et al.* 2012) had previously been reported to influence GIN resistance in sheep and cattle. These are preliminary results that have to pave the way to reach a more in-depth knowledge of the complexity of the molecular mechanisms underlying GIN resistance in sheep.

5.3 Global Discussion

In this PhD thesis, we attempted to provide a better understanding of the genetic basis for gastrointestinal nematode (GIN) resistance in adult sheep by using the latest available genomic tools and techniques. All the studies included in this PhD memory were conducted on adult Churra sheep, which is an autochthonous dairy sheep breed of Castilla y León, region located in the northwest of Spain. Unlike most other studies focused in the genetics of GIN resistance, which was conducted on lambs, we investigated GIN resistance in adult sheep. The reason for this is based on the specific features of the dairy sheep industry in the Castilla y León region, which is a production system based on adult ewes and the sale of suckling lambs fed exclusively on maternal milk, where replacement ewes and adult sheep (especially after lambing) are the only animals subjected to the direct effects of helminth infections (García-Pérez *et al.* 2002).

At the technical level, firstly, the availability of the sheep HD-chip motivated the MEGA-ULE group to, further scan the sheep genome with a higher marker density, with the previous gene mapping studies performed by the MEGA_ULE group for GIN resistance based on sparse microsatellite linkage maps (Gutiérrez-Gil *et al.* 2009;) or the analysis of the medium density 50K-chip (Atlija *et al.* 2016). Secondly, the sharp reduction in the per-base cost associated to NGS technologies that happened after 2012 was exploited to open a new research line focused on the study of the functional genomic mechanisms controlling GIN resistance in sheep. Thus, by exploiting the latest available genomic tools in the field of sheep genomic research, the two main objectives of this PhD Thesis

memory were defined as i) Refine and high-resolution mapping of QTL influencing GIN resistance in adult sheep; ii) Global transcriptomic study of target abomasal tissues after an experimental challenge with *Teladorsagia circumcincta* in adult ewes previously classified as resistant and susceptible. The first objective of identification of the genomic regions responsible to GIN in sheep was achieved by using currently available genomic tools like the ovine HD-chip and WG-seq and the second objective focused on the study of transcriptomic expression profiles in GIN infected sheep was achieved by exploiting one of the most popular NGS technologies across all the life sciences studies, the RNA-seq technology. These two primary objectives are independent to each other and were carried out as parallel studies.

The *in silico* genomic imputation strategy optimized through the consecution of the first objective of this PhD Thesis, and described in detail in **Study 1**, facilitated to impute our base Churra population genotyped with the 50K-chip to the HD-chip density. By limiting the number of animals genotyped with the HD-chip, and imputing genotypes for the whole population, we could address a genome scan for QTL detection with an increased marker density than that achieved by Atlija *et al.* (2016) at an affordable genotyping cost. In order to validate the accuracy of the imputed genotypes, we adopted a cross-validation strategy by masking a few individuals of the reference population and repeating the accuracy estimation. The achieved estimated accuracy was 91%-93% across our Churra half-sib population with different software, Beagle_v3.3.2 and Beagle_v4.0, which was within the precision range of other sheep studies (83-93%, according to breed) as estimated by Hayes *et al.*, (2012).

In addition, the genotypes generated for the 335 animals of the Churra population under study analysed with the HD-chip, including the information for a total of 490,940 markers, allowed to present an update of gene diversity parameters for the Churra sheep, as reported in **Study 1**. The use of the imputed HD-chip genotypes obtained for the whole population was used to update previous estimates of LD, *Ne*, and inbreeding coefficient for the Churra sheep breed, which had been previously estimated by our group based on the 50K-chip genotypes (García-Gámez *et al.* 2012). The overall pattern of the LD (r^2) estimates obtained for the imputed HD-chip dataset was very similar to that presented based on the 50K-chip in the same Churra population. When expressed in r^2 units, the LD estimates at shorter distances are lower for the HD-chip imputed dataset than for the 50K-chip dataset, whereas at longer distances there are no remarkable differences between the

two datasets estimates. However, when expressed in *D´* units, the estimates of the imputed dataset were slightly higher across all marker distances than those based on the 50K-chip. When comparing the *Ne* estimates between the 50K-chip, by considering the estimates for the 1cM ~ 1Mb ratio reported by Garcia- Gamez *et al.* (2012), with the HD-related datasets, we observed very similar values for the range between 10 and 25 generations ago, whereas the most recent estimates based on the 50K-chip (one to five generations ago) appear to be over-estimated as result of the lack of identification of some crossing-overs due to the lower marker density. On the contrary, the 50K-chip *Ne* estimates for the most distant generations are lower than those estimated for the HD-chip datasets, probably as a result of the low inaccuracy of the 50K-chip to estimate the extent of LD at the shortest distances. As an added value, we also used the imputed HD-chip genotypes to study the distribution of ROHs across the different chromosomes of the Churra sheep genome. This distribution shows a similar pattern to that described by Al-Mamun *et al.* (2015), which was based on 50K-chip genotypes analysed in both pure and cross-sheep breeds. However, the number of ROH segments was much higher in Churra sheep, which could be attributed to the higher genotyping density of our dataset. Roughly, 88% of the ROHs across the 26 autosomal chromosomes had lengths of less than 5 Mb, and 0.008% of the ROHs spanned more than 20 Mb. The distribution of the sum of ROHs per individual also showed a similar distribution to that reported by Al-Mamun *et al.* (2015) for pure breeds. We observed that all the genomic-based inbreeding coefficients resulted in higher estimates of inbreeding than the pedigree-based coefficient. In summary, **Study 1** included in this PhD Thesis memory provides an updated picture of the structure of the Churra sheep genome that may serve to better assess the potential of future gene mapping studies undertaken in this population to identify causal mutations and also evaluate the possibilities that genomic selection could offer in this population characterized by a short extent of LD.

The development of an *in silico* low-density chip and the optimized imputation strategy to higher density chips presented in **Study 2** could offer a practical and efficient solution to apply genomic selection in Castilla y León dairy sheep commercial populations by reducing the total cost associated with genotyping while maintaining acceptable reliability of genomic estimates. The average estimated imputation accuracy for the LD to 50K-chip density (ranged between 93.53% to 93.58%, for a masking level of 10% and 30%, respectively) appear as an appropriate and affordable approach to offer to the

breeders in order to obtain the basic genetic tests (e.g. paternity, *PRNP* genotypes) and estimations of genomic breeding values that could guide breeding decisions at very early age of the animal. On the other hand, it seems that further methodological efforts should be done to better optimized the imputation from the LD to the HD-chip density (estimated accuracy ranging between 88.3% and 86.52%, for a masking level of 10% and 30%, respectively) before trying to implement this approach in a commercial population. In any case, **Study 2** shows an initial evaluation of providing a practical and cost-efficient solution for the sheep breeders' association assessing the possibility of initiating a way into genomic selection.

The GWAS with imputed HD-chip genotype data and with IgA phenotype identified one unique significant SNP at 5% chromosome-wise level located on chromosome 15 (p_c-value corrected = 5.63), while no significant association was detected at the 5% genome-wise level. The marker that showed evidence of influencing IgA levels is the oar3_OAR15_24870525, and is located at position 24,870,525 bp of the OAR15. According to the basis of data Sheep QTLdb (Hu *et al.* 2016), the position of this QTL is within the estimated confidence interval (18.3-30.5 Mb) for a QTL with influence on Faecal egg count (FEC), the classic indicator trait of the levels of GIN infection in a backcrossing population Red Maasai x Dorper (Silva *et al*. 2012). The low power of the study can explain the lack of more significant results in this GWAS due to the limited number of animals analysed.

Considering the family structure of our population, we also applied an LA-based HD-genome scan by using the QTLMap software with the imputed HD-chip genotypes, as reported in **Study 4**, which is currently under preparation to be submitted for publication. This analysis identified the same QTL that the 50K-based LA genome scan reported by Atlija *et al.* (2016), with no new QTL regions being identified. However, the higher marker density used for gene mapping in this work allowed to substantially reduce the CI of the detected QTL when compared with the Atlija *et al.* (2016) results, achieving a CI refinement ranging from 20 to 90%, depending on the QTL region considered. These reductions of the estimated QTL CI lengths achieved by the HD-chip genome scan, together with the high-resolution analysis of the variants located in the CI of the OAR6 QTL region, based on the WG-seq analysis of a segregating trio, following the previously optimized strategy by the MEGA-ULE group in Gutierrez-Gil *et al.* (2018), have allowed us to present a list of putative mutations that could be directly associated to GIN resistance

traits in adult sheep. However, to confirm the role of these mutations with GIN resistance, which has been shown to be an extremely complex phenotype, further research efforts are required by analyzing these mutations in commercial populations with available phenotypic data related to GIN resistance.

In addition, when comparing the different results of the two gene mapping studies included in this Thesis memory based on the imputed HD-chip genotypes, based on different methodological approaches, a GWAS (**Study 2**) and a LA genome scan, it is important to take into account that these two approaches are assuming a different analysis model, and exploiting different type of information. The GWAS analysis exploits population information, whereas the LA genome scan is highly affected by the within-family QTL segregation. Hence, we consider, in agreement with García-Gámez *et al.* (2013), that by performing these different analyses we have gained a global picture of all the segregating QTL effects present in the studied population for the GIN resistance indicator traits analyzed. Especially important to note, considering the high level of family structure shown by the specific population here analyzed, based on 14 half-sib families of dairy ewes, the significant value and relevance of the LA-based results in relation to the planning of future high-resolution analyses aiming the identification of the causal mutation or QTN of the genetic effect previously mapped, for example, through the segregating trio WG-seq-based approach presented herein.

As a parallel approach to the identification the genomic regions responsible for GIN resistance in sheep, and within the framework of the second objective defined for this PhD project, we also tried to decipher the immune mechanisms activated in adult sheep during infection with GIN *T. circumcincta* through the transcriptome analysis of the abomasal mucosa and the abomasal lymph node tissues targeted by the infection. Six animals were classified as “susceptible”, and six as “resistant” based on the accumulated FEC estimation carried out on the Churra sheep that were subjected to experimental infection with *T. circumcincta*. Three distinct expression analyses were performed with the available twelve sample transcriptome data of adult Churra sheep: (i) differential gene expression analysis (DEG), (ii) identification of functional variants and, (iii) detection of lncRNAs.

A preliminary differential transcriptomic analysis performed for the abomasal mucosa RNA-seq dataset generated for the 12 considered animals, which is described in **Study**

5, suggested that at 7 days post-infection, there is not a clear differential gene response in the abomasal mucosa between Resistant and Susceptible ewes. This conclusion was based on the lack of overlap between the results of the analyses performed with two software, DESeq2 and EdgeR. Because of that, **Study 6** was focused on reporting the results of the DE analyses performed for the abomasal lymph samples. In this tissue a differential gene response it was observed, based on the identification of 106 genes commonly identified as DEGs by the two mentioned software. These 106 genes commonly identified as DEGs by both software were considered as GIN-activated DEGs. Out of these 106 common GIN-activated genes, 71 were up-regulated in resistant sheep and 35 were upregulated in susceptible sheep. Additional enrichment analyses considering these DEG identified genes highlighted as significant GO terms some interesting terms about the immune response such as *cytokine-mediated signaling pathway, cellular response to cytokine stimulus),* and significant KEGG pathways, such as *PPAR signaling pathway,* which is known to have a role in the immune response (Chen *et al.* 2017). Interestingly, some of the genes identified as GIN-activated genes, which have not been previously identified as DEGs for GIN resistance by other studies (mainly focused on lambs). we found a connection with immune response molecular mechanisms (*PGC, SFTPD, TUBA4A*, *SST, BPIFB1, PF4, B4GALNT2, JCHAIN, AQP7, KLHL25, NEDD4, ANO6*). We think that these DEGs might indicate genes that are activated explicitly in adult sheep and not in lambs.

In addition, the information generated through the RNA-seq technology for the lymph node samples was used to identify the potential causal mutations related to GIN infection in sheep, by performing a variant calling analysis on the transcriptome samples. From the initial genetic variability (**Study 7**) identified across the whole the transcriptome, we later studied in detail by performing functional annotation analysis (**Study 8**), the variants included in the genomic previously identified as DEGs in relation to GIN resistance between the abomasal lymph node transcriptome of six resistant and six susceptible adult sheep to *T. circumcincta* infection. In **Study 7** we identified 1,326,960 high-quality variants across the whole genome. From these high-quality variants, we extracted 6,168 variants within the 106 DEGs identified in **Study 6**. Among these, 332 variants were predicted to have a potential relevant functional impact and included 50 deleterious variants located within 15 genes some of which were to associate with immune response. In addition, the genetic variability of previously reported QTL regions and other

functional candidate genes related to GIN resistance was investigated. Hence, **Study 7** provides a list of 930 and 553 potential functional variants within these two types of target regions, QTL and candidate genes, which were located within 111 and 132 immune-related genes, respectively for the two types of target regions. The enrichment analyses performed with these immune-related genes highlighted significant Gene Ontology terms related to apoptosis, adhesion and inflammatory response and significant disease-related terms such as *inflammation, adhesion* and *necrosis*. Some of these terms may help to better understand the immune mechanism underlying some of the considered QTL. Also in relation to this, this study offers a list of previously reported QTL for which a single potential functional variant was identified, and that could be assessed, by future research, as potential causal variants.

To elucidate the potential functional role of lncRNAs with GIN infection in adult sheep, the detection of lncRNAs was performed by implementing specific analyses on the transcriptomic data generated for abomasal lymph node tissue samples (**Study 9**). We performed a differential gene expression analysis again and identified the corresponding lncRNA transcripts related to the DEGs. The lncRNA transcripts were detected using FEELnc (Wucher *et al.* 2017) software. The differential gene expression analysis performed, in this case for 6 Resistant and 5 Susceptible samples, identified a total of 3,148 DEGs, whereas 683 lncRNA transcripts were associated with 457 DEGs. Focusing on the 367 lncRNA transcripts that were later found to include 301 overlapping genes, this work has highlighted a new list of genes that could be regulated by lncRNA , some of which (*IRF1, CCL14, KRT8,* and *LGALS3)* have been previously reported to influence o the host response against the nematode infection in sheep and cattle.

Globally, this PhD Thesis has identified and refined the location of significant genomic regions responsible for GIN resistance in sheep by performing the classical LA and LD based on a GWA study. Based on these two analyses we have provided a list of genetic variants and candidate genes that are potentially associated with GIN resistance indicator traits. Additionally, the transcriptomic expression studies quantified the level of gene expression in target tissues of infected sheep and provided a comparison of the gene expression profiles between Resistant and Susceptible ewes. The DE analysis has been later extended to identify lncRNAs associated with DEGs. This provides an almost whole picture of the differential expression of transcripts (coding and lncRNA) identified

between the two compared groups of animals based on the disease/susceptibility status defined by the previous *T. circuncinta* experimental infection to which the studied animals were subjected. In addition, the global transcriptomic research performed has allowed us to present a global map of the genetic variants identified in the the lymph node transcriptome of sheep after infection with *T. circuncinta*, and has provided a list of genetic variants that could be associated to GIN resistance in sheep by its location within QTL or GIN-activated genes. The results reported in this Thesis, and derived from the efforts made to successfully address the two major Objectives defined for this PhD project, may open a new and integrative research line based on the confirmation of the potential associations suggested here, with GIN resistance traits, for the lists of variants derived from both, the structural and the functional studies, included in this PhD memory. For that, for example, future research projects of the MEGA-ULE group could be built on the design of a custom SNP-chip including the genetic variants highlighted by both the gene mapping studies reported here and also by the variant calling analysis performed on the lymph node RNA-seq dataset. The analysis of this *custom* SNP-chip in commercial populations with available phenotypic data related to GIN resistance would serve to confirm the potential associations suggested here.

Based on all the above mentioned, this PhD Thesis has provided a step forward towards the identification of genes underlying the resistance to GIN infections in sheep in two senses: (i) by providing a deeper knowledge on the structural organization of the Churra sheep genome (through the analysis of the LD extent, and related analyses) and increasing the marker density of the previous genome scans performed by the MEGA-ULE group for GIN resistance, and (ii) by attempting, for the first time in adult sheep, a functional approach for the study of the genetic mechanisms underlying the GIN resistance phenotype in sheep. As added values, the optimized imputation strategy of genotypes reported here may define a first step towards the implementation of genomic selection in the commercial populations of Spanish dairy sheep breeds.

6. Conclusions

Conclusions:

1. The methodological strategy optimized here for imputation of genotypes from the 50K-chip to HD-chip marker densities appears to be appropriate (based on the reported imputation accuracy and the saving of genotyping costs) not only for research objectives but also for practical use in commercial populations. On this regard, the virtual low-density chip (LD-chip) here developed would offer an accurate and cost-efficient genomic tool to use in dairy sheep.
2. The HD gene mapping studies presented here for GIN resistance indicator traits confirmed and refined the location of three QTL regions previously identified.
3. The Whole Genome Sequencing analysis applied to a segregating trio for the most promising QTL, located on OAR6, has enabled to present a list of potentially functional variants associated with GIN resistance traits in adult sheep.
4. The whole-transcriptomic analysis by RNA-Seq in adult ewes identified as Susceptible and Resistant to GIN has revealed the differential expression of 106 genes by comparing the gene expression profile of the abomasal lymph node. Some of these genes have been suggested to show a connection in immune mechanisms explicitly activated in adult animals. This study has been extended to identify lncRNAs associated with DEGs between the two compared groups to provide a global picture of the different transcripts that may be related to GIN resistance in sheep.
5. The variant calling analyses for the abomasal lymph node RNA-seq dataset identified potential functional variants within two types of considered regions, previously reported QTL regions and functional candidate genes for parasite resistance in sheep.

7. Conclusiones

Conclusiones:

1. La estrategia metodológica optimizada para la imputación de genotipos del chip-50K al chip-HD parece ser apropiada (en base a la precisión de imputación y el ahorro de costes de genotipado) no sólo en relación a proyectos de investigación sino también para su uso práctico en poblaciones comerciales. En este sentido, el chip virtual de baja densidad (LD-chip) diseñado podría ofrecer una herramienta genómica precisa y eficiente económicamente para su uso en el ganado ovino lechero.
2. Los estudios de mapeo de alta densidad para caracteres relacionados con la resistencia a GINs han permitido confirmar y refinar la localización de tres regiones QTL previamente identificadas.
3. El análisis de secuenciación del genoma completo aplicado en un trío segregante para el QTL situado en OAR6, ha permitido identificar una lista de variantes funcionales potencialmente asociadas con caracteres de resistencia a GINs en ovejas adultas.
4. El análisis transcriptómico completo mediante RNA-seq en ovejas adultas identificadas como Susceptibles y Resistentes a GIN ha revelado la expresión diferencial de 106 genes al comparar el patrón de expresión génica en ganglio linfático abomasal. La falta de correspondencia de algunos de estos genes con estudios funcionales anteriores realizados en corderos podría sugerir la existencia de mecanismos inmunológicos propios de animales adultos. Este estudio se ha complementado mediante la identificación de lncRNA asociados a DEGs entre los dos grupos comparados a fin de ofrecer una imagen global de los genes implicados en la resistencia a las infecciones por GIN en ovejas.
5. La identificación de variantes en los datos de RNA-seq de los ganglios linfáticos abomasales nos ha permitido aportar una lista de polimorfismos con una potencial importancia funcional dentro de los dos tipos de regiones genómicas consideradas: QTLs previamente descritos y genes candidatos relacionados con la resistencia parasitaria en el ganado ovino.

9. Summary

The present PhD Thesis, which has been developed in the Animal Breeding research group of the University of León (MEGA-ULE group), was set out with the **global objective** to identify the genetic basis underlying resistance to gastrointestinal nematode (GIN) infection in adult sheep, by taking advantage of the latest available genomic technology such as the Illumina ovine High-Density Bead chip (HD-chip) and the cost-effective next-generation sequencing (NGS) technologies. This global objective was achievedby addressing two parallel **specific objectives**. The first specific objective was focused on a high-density mapping of genomic regions underlying resistance to GIN in sheep using imputed HD SNP genotypes in a population previously analysed with the medium density SNP-chip (50K-chip) and for which two indicator traits of parasite infection level were available. The second objective was focused on the functional characterization of genomic regions that are responsible for resistance to GIN in adult sheep using RNA sequencing (RNA-seq) data extracted from abomasal tissues (abomasal mucosa and abomasal lymph node tissues) of adult Churra sheep after experimental infection with GIN *T. circumcincta*.

Within the activities required to fulfil the first specific objective, the first step was the imputation of genotypes from the 50K-chip genotypes to the HD-chip density (600K), with a previous estimation of the accuracy of the imputation strategy utilized. The analysis was performed in a commercial population of Churra sheep with available genotypes for the medium density of 50K-chip. This population consisted of 16 half-sib families, with a total of 1,670 daughters. The 1,686 animals in the population were genotyped with a 50K-chip and a subset of 240 animals of that population, together with 96 artificial insemination sires, were also genotyped with a HD-chip (600K). Using the two genotype datasets, the 50K-chip genotypes of the global population were imputed to the HD-chip using the Beagle software. The average concordance obtained across different studies and scenarios for the 26 autosomal chromosomes, was around 90-92%, which is within the imputation accuracy range reported by other sheep genomic studies using Beagle (83-93%).

The results on the evaluation of different imputation strategies have been presented in **Study 1,** where the imputed HD-chip genotypes were used to update the estimates of LD, identify regions involving runs of homozygosity (ROH) and update the estimation of effective population size (*Ne*) for the studied population. The updated LD estimations provided evidence that the extent of LD in Churra sheep is even shorter than that reported

based on the 50K-chip and is one of the shortest extents compared with other sheep breeds. Through different comparisons, we have also assessed the impact of imputation on LD and *Ne* estimates. The inbreeding coefficient, considering the total length of the ROH, showed an average estimate (0.0404) lower than the critical level. In general, the higher accuracy of updated LD estimates suggests that the HD-chip, combined with an appropriate imputation strategy, offers a powerful tool for gene mapping studies or the implementation of genomic selection in sheep.

By taking advantage of the imputation strategy optimized, and considering the limitation of using genomic selection in dairy sheep due to the genotyping costs, **Study 2** reports the development of an *in silico* low-density chip (LD-chip) and the potential of this genomic tool to impute genotypes to medium (50K-chip) and high (HD-chip) marker densities. For that, we used the 50K and HD-chip genotypes available for the global Churra sheep population studied in **Study 1**. A total of 2,935 markers common to the two panels and evenly distributed across the sheep genome were included in the virtual LD-chip. The average accuracy for the LD-chip to 50K-chip imputation ranged from 93.53% to 93.58%, and for the LD-chip to HD-chip imputation ranged from 88.3% to 86.52% (depending on the masking level considered, 10% and 30%, respectively). Globally, and although certain identified limitations need to be solved, the *in silico* LD-chip described in **Study 2**, together with the implemented imputation strategy, appears to provide an appropriate approach to obtain information for basic genetic tests (e.g. paternity), and also estimations of genomic breeding values that could guide breeding decisions.

The last actitivity performed within the first objective was the performance of gene mapping studied based on the imputed HD-chip genotypes. First, we performed a GWA analysis based on the the imputed HD-chip genotypes obtained for the 532 daughters of the commercial population with available phenotypes for GIN resistance indicator traits. In **Study 3** we present the results of the GWA study performed for the indicator trait Immunoglobulin A serum levels (IgA), which identified a single significant SNP at the 5% chromosome-wise level, located on OAR15. To exploit the half-sib structure of the studied resource population, we later performed a genome scan based on Linkage analysis (LA), which is described in **Study 4** for two indicator traits of GIN resistance, IgA and faecal egg count (FEC). Also, in **Study 4**, the genetic variability of the new refined CIs for the most promising QTL, identified for FEC on chromosome 6 (OAR6), was studied at high resolution by using the information derived from the analysis of WG-seq datasets

of a segregating trio for that QTL (which included the *Qq* sire, and 2 daughters with extremely divergent phenotypes for the FEC trait in concordance with their QTL inferred QTL genotypes, *QQ* and *qq*). In this analysis, a total of 433 high-quality variants were identified within the OAR6 CI QTL region (88.2-88.3 Mb), 357 of which were intragenic concordant variants for the three animals analysed. Based on a variant functional annotation analysis, **Study 4** provides a list of variants that could be considered as potentially underlying GIN resistance in sheep.

The second objective related with the functional characterization of genomic regions responsible for GIN resistance in adult sheep was based on the global transcriptomic study through RNA sequencing of abomasal tissue samples obtained for 12 adult ewes of Churra sheep after performing a first experimental infection (EI1) with GIN *T. circumcincta.* Based on the individual accumulated FEC measures after EI1, six sheep were classified as susceptible, and six sheep were classified as resistant to GIN. Later the selected animals were exposed to a second experimental infection (EI2) with *T. circum*cincta. At day 7 after EI2, the animals were sacrificed and at necropsy, abomasal mucosa and lymph node samples for all animals were immediately collected and the mRNA extracted for two tissues, abomasal mucosa and abomasal lymph node.

Using the abomasal mucosal RNA-seq samples an initial preliminary differential transcriptomic analysis was performed between the abomasal mucosa samples obtained from two groups of sheep classified as resistant or susceptible to the experimental infection with *T. circumcincta* (**Study 5**). The differential expression analysis performed with two different R-based packages, DESeq2 and edgeR, did not show concordant results, suggesting that at 7 days post-infection, there is not a clear differential gene response in the abomasal mucosa between resistant and susceptible ewes. Because of that, **Study 6** was focused on reporting the results of the DE analyses performed for the abomasal lymph samples. For this tissue, the expression analysis performed with the two previously mentioned software identified a total of 106 common differentially expressed genes (DEGs), which were considered as GIN-activated. The enrichment analyses performed for these GIN-activated genes identified some immune related and the PPARG (Peroxisome proliferator-activated receptor gamma) signaling pathway as well as disease terms related to inflammation and gastro-intestinal diseases. A systematic comparison with the results of previous studies confirmed the involvement of genes such as *ITLN2*,

CLAC1 and galectins, in the immune mechanism activated against *T. circumcincta* in resistant sheep.

In addition, the information generated through the RNA-seq technology for the lymph node samples was used to identify potential causal mutations related to GIN infection in sheep, by performing a variant calling analysis on this transcriptome dataset. The initial study of the genetic variability identified throughout the transcriptome and in the 106 genes previously identified as GIN-activated, described in **Study 7**, was followed by a more extensive functional annotation analysis described in **Study 8**. In both studies variant calling analyses were performed with two different types of software, GATK_v3.7 and Samtools_v1.4 and the variants commonly identified by the two different software packages were considered "high-quality variants". In **Study 7** we identified 1,326,960 high-quality variants across the whole transcriptome. From these high-quality variants, we extracted 6,168 variants within the 106 DEGs identified in **Study 6**. In **Study 8**, in addition, the high-quality variants with a potential functional impact were identified for two types of target regions, (i) QTL regions previously reported in sheep for parasite resistance and (ii) functional candidate genes selected from gene expression studies related to GIN resistance in sheep. Overall, the study provides a valuable list of potential causal mutations that could be considered as candidate causal mutations concerning GIN resistance in sheep.

In **Study 9**, we continued exploiting the transcriptomic dataset obtained by RNA-seq for the abomasal lymph node tissue previously analysed to identify long noncoding RNAs (lncRNAs) that could be associated to GIN resistance in sheep. Using a specific bioinformatic pipeline for the identification of lncRNA based on the FEELnc package, the analysis identified 44,203 genes with 77,039 transcripts, 9,105 of which were identified as lncRNAs (2,092 of them were classified as *novel*). An additional differential gene expression analysis, identified a total of 3,148 DEGs, 457 of which were associated with 683 of the identified lncRNA transcripts. Focusing on the 367 lncRNA transcripts that were later found to include 301 overlapping genes, this work has highlighted a new list of GIN-activated genes that could be regulated by lncRNA, some of which (*IRF1, CCL14, KRT8,* and *LGALS3)* have been previously reported to influence on the host response against the nematode infection in sheep or cattle.

Based on all the above mentioned, this PhD Thesis has provided a step forward towards the identification of genes underlying the resistance to GIN infections in sheep in two senses: (i) by providing a deeper knowledge on the structural organization of the Churra sheep genome (through the analysis of the LD extent, and related analyses) and increasing the marker density of previous genome scans performed by the MEGA-ULE group for GIN resistance, and (ii) by addressing, for the first time in adult sheep, a functional approach for the study of the genetic mechanisms underlying the GIN resistance phenotype in sheep. As an added value, the optimized imputation strategy of genotypes reported here may define a first step towards the implementation of genomic selection in the commercial populations of Spanish dairy sheep breeds.

9. Resumen

La actual tesis doctoral, que se ha desarrollado en el grupo de investigación de mejora genética animal de la Universidad de León (grupo MEGA-ULE), tiene como **objetivo global** el estudio de la base genética subyacente a la resistencia a los nematodos gastrointestinales (GIN) en ovejas adultas. Para ello se han utilizado metodologías empleadas en la genómica, tales como el chip de genotipado de SNPs de alta densidad (HD) y las tecnologías se cuenciación conocidas como *next generation sequencing* (NGS). Para el cumplimiento el objetivo global se han abordado dos **objetivos específicos.** El primero de ellos tuvo por objeto realizar un mapeo de alta densidad de regiones con influencia sobre la resistencia a los GIN en ovejas mediante el análisis de genotipos imputados para el chip HD-chip en una población de raza Churra previamente analizada con el chip de media densidad (50K-chip) y para la cual estaban disponibles datos de dos fenotipos relacionados con la carga parasitaria. El segundo objetivo específico se centró en la caracterización funcional de regiones genómicas asociadas con la resistencia a los GIN en ovejas adultas, utilizando el anásis del transcriptioma completo de diferentes tejidos (ganglio linfático abomasal y mucosa del abomaso) después de la infección experimental con *T. circumcincta*, mediante la metodología RNA-seq.

Dentro de las actividades necesarias para cumplir el primer objetivo concreto, el primer paso fue la imputación de genotipos del chip de 50K a la densidad del chip de HD (600K), realizándo previamente una estimación de la precisión de la estrategia de imputación utilizada. La imputación de genotipos se realizó en una población comercial de ovejas de raza Churra con genotipos disponibles para la densidad media de 50K-chip. Esta población estaba formada por 16 familias de medio hermanas, con un total de 1.670 hijas. Los 1.686 animales de la población se genotiparon con un chip de 50K SNP y un subconjunto de 240 animales de esa población, más otros 96 machos de inseminación artificial se genotiparon, además, con un chip HD (600K). Utilizando los dos conjuntos de datos genotípicos disponibles, los genotipos del 50K-chip para la población global fueron imputados al chip HD utilizando el software Beagle. La concordancia media obtenida en diferentes estudios y escenarios para los 26 cromosomas autosómicos, fue de alrededor el 90-92%, que está dentro del rango de precisión de imputación reportado por otros estudios genómicos de ovejas utilizando Beagle (83-93%).

Los resultados sobre la evaluación de diferentes estrategias de imputación se han presentado en el **Estudio 1** donde los genotipos imputados para el chip de alta densidad

sirvieron para evaluar el nivel de desequilibrio ligamiento (LD) en el genoma de la raza Churra, identificar regiones conocidas como *runs of homozygosity* (ROH) y actualizar las estimaciones del tamaño efectivo *(Ne)* de la población con una mayor precisión aportada por la alta densidad de marcadores. Las estimaciones actualizadas de LD ponen de manifiesto que el grado de LD en la raza ovina Churra es aún más corto que el descrito en base al 50K-chip, siendo una de las extensiones más cortas descritas en razas ovinas. El coeficiente de consanguinidad, teniendo en cuenta la longitud total del ROH, mostró una estimación media (0,0404) inferior al nivel crítico. En general, la mayor precisión de las estimaciones actualizadas de LD sugiere que el HD-chip, combinado con una adecuada estrategia de imputación, ofrece una poderosa herramienta en estudios de mapeo genético o la aplicación de la selección genómica en ganado ovino.

Aprovechando la estrategia de imputación optimizada, y considerando la limitación del uso de la selección genómica en ovejas lecheras debido a los costos de genotipado, el **Estudio 2** describe el desarrollo *in silico* de un chip de baja densidad (LD-chip) y el potencial de esta herramienta genómica para imputar genotipos a media (50K-chip) y alta (HD-chip) densidad de marcadores. Para ello se utilizaron los genotipos del 50K-chip y HD-chip disponibles para la población global de ovejas Churra estudiada en el **Estudio 1.** Así, el LD-chip virtual diseñado incluyó un total de 2.935 marcadores comunes a los dos paneles y distribuidos uniformemente a través del genoma ovino. La precisión media para la imputación del LD-chip a 50K-chip osciló entre 93,53% y 93,58%, y para la imputación del LD-chip al HD-chip osciló entre 88,3% a 86,52% (dependiendo del nivel de enmascaramiento considerado, 10% y 30%, respectivamente). A nivel global, y aunque será necesario resolver ciertas limitaciones identificadas, el LD-chip descrito en el **Estudio 2**, junto con la sugerida estrategia de imputación, parecen aportar una aproximación apropiada para obtener información para pruebas genéticas básicas (por ejemplo, paternidad), y también para estimaciones de los valores de cría genómicos que podrían ser utilizados para guiar la toma de decisiones en relación a la selección de candidatos a la selección.

La última actividad del primer objetivo fue la realización de estudios de mapeo genético de alta resolución. En primero lugar se realizó un estudio GWA con los genotipos imputados del HD-chip obtenidos para las 532 hijas de la población comercial y sus fenotipos disponibles para caracteres indicadores de resistencia a GIN. En el **Estudio 3** se presentan los resultados del análisis GWAS para el caracter “niveles séricos de

Inmunoglobulina A" (IgA), que identificó unicamente un SNP significativo al nivel de signficación 5% *chromosome-wise*, localizado en OAR15. Posteriormente, con el fin de explotar la estructura de la población de medio hermanas considerada en este trabajo realizamos un escaneo del genoma basado en el análisis de ligamiento (LA), que se describe en el **Estudio 4** para los dos fenotipos indicadores de resistencia a la infección por GIN medidos en esta población, IgA y el recuento de huevos en heces (FEC). Este análisis ha permitido refinar los intervalos de confianza (CI) de los regiones QTL (del inglés *Quantitative Trait Loci*) previamente identificadas en oveja Churra para caracteres de resistencia a GIN en base al 50K-chip. Además, en el **Estudio 4**, la variabilidad genética del nuevo CI refinado para el QTL más prometedor, identificado en el cromosoma 6 (OAR6) para FEC, se estudió a alta resolución utilizando la información derivada de la WG-seq para un trío segregante para ese QTL (que incluyó el padre *Qq* y 2 hijas con fenotipos divergentes extremos para el rasgo FEC en concordancia con sus genotipos para el QTL, hija *QQ* e hija *qq*). En este estudio se han identificado un total de 433 variantes de alta calidad dentro de la región considerada en OAR6 (88,2-88,3 Mb), de las cuales 357 eran variantes intragénicas concordantes con el genotipo del QTL para los tres animales analizados. Basado en un análisis de anotación funcional de variantes, el **Estudio 4** proporciona una lista de variantes que podrían considerarse como posibles mutaciones asociadas con la resistencia ovina a las infecciones por GIN.

El segundo objetivo relacionado con la caracterización funcional de las regiones genómicas responsables de la resistencia a la GIN en ovejas adultas se basó en el estudio transcriptómico global a través de la secuenciación de RNA de muestras de tejido abomasal obtenidas para 12 ovejas adultas de raza Churra. Sobre la base de las medidas individuales de FEC en una infección experimental (EI1) con *T. circumcincta* , seis ovejas fueron clasificadas como susceptibles, y seis ovejas fueron clasificadas como resistentes a GIN. Posteriormente, estos animales fueron expuestos a una segunda infección experimental (EI2) con *T. circum*cincta. En el día 7 tras EI2, los animales fueron sacrificados y la necropsia permitió la recogida, para cada animal, de muestras de mucosa abomal y ganglio linfático abomasal, a partir de las cuales se obtuvieron muestras de mRNA que fueron secuenciadas mediante metodología RNA-seq.

Utilizando las muestras de RNA-seq de la mucosa abomasal se realizó un análisis preliminar de expresión diferencial comparando las muestras de mucosa abomasal de ovejas clasificadas como resistentes o susceptibles a la infección experimental con *T.*

circumcincta (**Estudio 5)**. El análisis de expresión diferencial mediante los softwares DESeq2 y edgeR no mostró resultados significativos comunes, sugiriendo que a los 7 días tras la infección con *T. circumcinta* no hay una respuesta diferencial clara en relación al perfil transcripcional en la mucosa abomasal de animales resistentes y susceptibles. Debido a eso, el **Estudio 6** se centró principalmente en describir los resultados del análisis DE realizado para las muestras de ganglio linfático abomasal. Para este tejido, el análisis de expresión realizado con los dos programas mencionados anteriormente identificó un total de 106 DEGs comunes, que fueron considerados como genes activados por GIN. Los análisis de enriquecimiento realizados para estos genes identificó como significativas algunas rutas relacionadas con la respuesta inmune mediada por citoquinas y la ruta de señalización de PPARG (Peroxisome proliferator-activated receptor gamma), además de términos de enfermedades relacionados con la inflamación y las enfermedades gastrointestinales. La comparación sistemática de los genes identificados como activados por GIN con resultados de estudios funcionales anteriores confirmó la implicación de genes como *ITLN2, CLAC1* y galectinas, en los mecanismos inmunes activados en ovejas resistentes frente a *T. circumcincta*.

Por otra parte, la información generada mediante la tecnología de RNA-seq para las muestras de ganglios linfáticos fue utilizada para identificar mutaciones con un posible impacto funcional en regiones previamente identificadas como asociadas a la resistencia parasitaria en el ganado ovino. Para ello, se realizó un análisis de identificación de variantes considerando conjuntamente todos los datos de transcriptoma de las muestras de ganglios linfáticos. El estudio inicial de la variabilidad genética identificada a lo largo de todo el transcriptoma y en los 106 genes previamente identificados como activados por GIN, descrito en **Estudio 7**, fue seguido por un análisis de anotación funcional más extenso descrito en el **Estudio 8.** En ambos estudios se realizó análisis de identificación de variantes con dos tipos diferentes de software, GATK_v3.7 y Samtools_v1.4 y las variantes comúnmente identificadas por los dos paquetes se consideraron "variantes de alta calidad". En el **Estudio 7** se identificaron 1.326.960 variantes de alta calidad en todo el transcriptoma. De estas variantes de alta calidad, se extrajeron un total de 6.168 variantes dentro de los 106 DEG identificados en el **Estudio 6.** En el **Estudio 8,** además, se identificaron las variantes de alta calidad con un posible impacto funcional en dos tipos de regiones de interés, i) regiones QTL previamente descritas en la oveja para caracteres de resistencia a parásitos; ii) genes candidatos funcionales seleccionados a partir de

estudios de expresión génica relacionados con la resistencia a las infecciones por GIN en ovejas. En general, este estudio proporciona una valiosa lista de posibles mutaciones causales que podrían considerarse como mutaciones causales candidatas relativas a la resistencia a las infecciones por GIN en ovejas.

En el **Estudio 9,** los datos de RNA-seq obtenidos para ganglios linfáticos abomasales fueron sometidos a un nuevo análisis con el objetivo de identificar RNA largos no codificantes (lncRNAs) que pudieran asociarse a la resistencia a los GIN en ovejas. Utilizando un flujo de análisis bioinformático específico para la detección de este tipo de transcrito basado en el paquete *FEELnc*, se identificaron 44.203 genes anotados y un total de 77.039 transcritos, de los cuales 9.105 fueron clasificados como lncRNAs (2.092 de los cuales fueron considerados como *novel*). Un análisis adicional de expresión difirencial permitió identificar un total de 3.148 DEGs, 457 de los cuales estaban asociados a 683 de los lncRNA previamente identificados. Considerando los 367 lncRNAs para los que se identificaron 301 genes solapantes, este trabajo proporciona una nueva lista de genes activados por GIN que podrían estar regulados por lncRNA. Para algunos de esos genes *(IRF1, CCL14, KRT8* y *LGALS3)* se ha descrito previamente una posible relación con la respuesta del huésped contra la infección por nematodos en ovejas o en ganado bovino.

En conjunto, esta Tesis doctoral representa un avance hacia el conocimiento de la base genética de la resistencia a las infecciones por GIN en ovejas en dos sentidos: i) proporcionando un conocimiento más profundo sobre la organización estructural del genoma de las ovejas de raza Churra (mediante el análisis de la extensión de la LD, etc) y aumentando la densidad de marcadores de barridos genómicos previamnete realizaods por el grupo MEGA-ULE para caracteres de resistencia a GIN, y también (ii) al abordar, por primera vez en ovejas adultas, un enfoque funcional a través del análisis global del transcriptoma de tejidos afectados por la infección por GIN, para el estudio de los mecanismos genéticos subyacentes al fenotipo de resistencia. Como valor añadido, la estrategia optimizada para la imputación de genotipos podría suponer un primer paso hacia la implementación de la selección genómica en las poblaciones comerciales de razas lecheras españolas.

References

Ahmed A. M., Good B., Hanrahan J. P., McGettigan P., Browne J., Keane O. M., Bahar B., Mehta J., Markey B., Lohan A., Sweeney T., (2015) Variation in the Ovine abomasal lymph node transcriptome between breeds known to differ in resistance to the gastrointestinal nematode. *PLoS ONE*.

Al-Mamun H. A., A Clark S., Kwan P., Gondro C., (2015) Genome-wide linkage disequilibrium and genetic diversity in five populations of Australian domestic sheep. *Genetics, selection, evolution : GSE* **47**, 90.

Allen N. P. C., Donninger H., Vos M. D., Eckfeld K., Hesson L., Gordon L., Birrer M. J., Latif F., Clark G. J., (2007) RASSF6 is a novel member of the RASSF family of tumor suppressors. *Oncogene* **26**, 6203–6211.

Álvarez-Sánchez M. A., Mainar-Jaime R. C., Pérez-García J., Rojo-Vázquez F. A., (2006) Resistance of Fasciola hepatica to triclabendazole and albendazole in sheep in Spain. *Veterinary Record*.

Amarante A. F. T., Rocha R. A., Bricarello P. A., (2007) Relationship elationship of intestinal histology with the resistance to Trichostrongylus richostrongylus colubriformis infection in three breeds of sheep sheep. *Pesquisa Veterinaria Brasileira*.

Anders S., Huber W., (2010) Differential expression analysis for sequence count data. *Genome biology* **11**, R106.

Andersson L., (2009) Genome-wide association analysis in domestic animals: A powerful approach for genetic dissection of trait loci. *Genetica*.

Andronicos N., Hunt P., Windon R., (2010) Expression of genes in gastrointestinal and lymphatic tissues during parasite infection in sheep genetically resistant or susceptible to *Trichostrongylus colubriformis* and *Haemonchus contortus*. *International Journal for Parasitology* **40**, 417–429.

Andronicos N. M., McNally J., Kotze A. C., Hunt P. W., Ingham A., (2012) Trichostrongylus colubriformis larvae induce necrosis and release of IL33 from intestinal epithelial cells in vitro: Implications for gastrointestinal nematode vaccine design. *International Journal for Parasitology*.

Araujo R. N., Padilha T., Zarlenga D., Sonstegard T., Connor E. E., Tassel C. Van, Lima W. S., Nascimento E., Gasbarre L. C., (2009) Use of a candidate gene array to delineate gene expression patterns in cattle selected for resistance or susceptibility to intestinal nematodes. *Veterinary Parasitology*.

Athanasiadou S., Pemberton A., Jackson F., Inglis N., Miller H. R. P. P., Thévenod F., Mackellar A., Huntley J. F., (2008) Proteomic approach to identify candidate effector molecules during the in vitro immune exclusion of infective *Teladorsagia circumcincta* in the abomasum of sheep. *Veterinary Research* **39**, .

Atlija M., Arranz J.-J. J., Martinez-Valladares M., Gutiérrez-Gil B., (2016) Detection and replication of QTL underlying resistance to gastrointestinal nematodes in adult sheep using the ovine 50K SNP array. *Genetics Selection Evolution* **48**, 4.

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

Aulchenko Y. S., Struchalin M. V, Duijn C. M. van, (2010) ProbABEL package for genome-wide association analysis of imputed data. *BMC bioinformatics* **11**, 134.

Balic A., Cunningham C. P., Meeusen E. N. T., (2006) Eosinophil interactions with *Haemonchus contortus* larvae in the ovine gastrointestinal tract. *Parasite Immunology* **28**, 107–115.

Barendse W., Armitage S. M., Kossarek L. M., Shalom A., Kirkpatrick B. W., Ryan A. M., Clayton D., Li L., Neibergs H. L., Zhang N., Grosse W. M., Weiss J., Creighton P., McCarthy F., Ron M., Teale A. J., Fries R., McGraw R. A., Moore S. S., Georges M., Soller M., Womack J. E., Hetzel D. J. S., (1994) A genetic linkage map of the bovine genome. *Nature Genetics*.

Barillet F, Arranz JJ, Carta A, Jacquiet P, Stear M B. S., (2006) Final Consolidated Report of the European Union Contract of Acronym GeneSheepSafety. *QTLK5CT-2000-00656*145.

Becker D., Tetens J., Brunner A., Bürstel D., Ganter M., Kijas J., Drögemüller C., (2010) Microphthalmia in Texel sheep is associated with a missense mutation in the paired-like homeodomain 3 (PITX3) gene. *PLoS ONE*.

Beh K. J., Hulme D. J., Callaghan M. J., Leish Z., Lenane I., Windon R. G., Maddox J. F., (2002) A genome scan for quantitative trait loci affecting resistance to Trichostrongylus colubriformis in sheep. *Animal Genetics*.

Benavides M. V., Sonstegard T. S., Tassell C. Van, (2016) Genomic Regions Associated with Sheep Resistance to Gastrointestinal Nematodes. *Trends in Parasitology* **32**, 470–480.

Bendixen C., Hedegaard J., Horn P., (2005) Functional genomics in farm animals - Microarray analysis. In: *Meat Science*,

Benjamini Y., Hochberg Y., (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B …* **57**, 289–300.

Besier R. B., Love S. C. J., (2003) Anthelmintic resistance in sheep nematodes in Australia: The need for new approaches. *Australian Journal of Experimental Agriculture*.

Bhuiyan A. A., Li J., Wu Z., Ni P., Adetula A. A., Wang H., Zhang C., Tang X., Bhuyan A. A., Zhao S., Du X., (2017) Exploring the genetic resistance to gastrointestinal nematodes infection in goat using RNA-sequencing. *International Journal of Molecular Sciences*.

Binsbergen R. Van, Bink M. C. A. M., Calus M. P. L., Eeuwijk F. A. Van, Hayes B. J., Hulsegge I., Veerkamp R. F., (2014) Accuracy of imputation to whole-genome sequence data in Holstein Friesian cattle. *Genetics Selection Evolution*.

Boichard D., Ducrocq V., Croiseau P., Fritz S., (2016) Genomic selection in domestic animals: Principles, applications and perspectives. *Comptes Rendus - Biologies*.

Breuer K., Foroushani A. K., Laird M. R., Chen C., Sribnaia A., Lo R., Winsor G. L., Hancock R. E. W., Brinkman F. S. L., Lynn D. J., (2013) InnateDB: Systems biology of innate immunity and beyond - Recent updates and continuing curation. *Nucleic Acids Research*.

Brown D. J., Fogarty N. M., (2017) Genetic relationships between internal parasite resistance and production traits in Merino sheep. *Animal Production Science*.

Browning B. L., Browning S. R., (2009) A Unified Approach to Genotype Imputation and Haplotype-Phase Inference for Large Data Sets of Trios and Unrelated Individuals. *The American Journal of Human Genetics* **84**, 210–223.

Browning B. L., Browning S. R., (2013) Improving the accuracy and efficiency of identity-by-descent detection in population data. *Genetics* **194**, 459–471.

Bumstead N., Palyga J., (1992) A preliminary linkage map of the chicken genome. *Genomics*.

Bush S. J., Muriuki C., McCulloch M. E. B., Farquhar I. L., Clark E. L., Hume D. A., (2018) Cross-species inference of long non-coding RNAs greatly expands the ruminant transcriptome. *Genetics Selection Evolution*.

Carter A. M., (2012) Evolution of Placental Function in Mammals: The Molecular Basis of Gas and Nutrient Transfer, Hormone Secretion, and Immune Responses. *Physiological Reviews* **92**, 1543–1576.

Cavanagh C. R., Jonas E., Hobbs M., Thomson P. C., Tammen I., Raadsma H. W., (2010) Mapping Quantitative Trait Loci (QTL) in sheep. III. QTL for carcass composition traits derived from CT scans and aligned with a meta-assembly for sheep and cattle carcass QTL. *Genetics, selection, evolution : GSE* **42**, 36.

Chang E. I., Zarate M. A., Arndt T. J., Richards E. M., Rabaglino M. B., Keller-Wood M., Wood C. E., (2019) Ketamine reduces inflammation pathways in the hypothalamus and hippocampus following transient hypoxia in the late-gestation fetal sheep. *Frontiers in Physiology*.

Chen T., Tibbitt C. A., Feng X., Stark J. M., Rohrbeck L., Rausch L., Sedimbi S. K., Karlsson M. C. I., Lambrecht B. N., Karlsson Hedestam G. B., Hendriks R. W., Chambers B. J., Nylén S., Coquet J. M., (2017) PPAR-γ promotes type 2 immune responses in allergy and nematode infection. *Science Immunology* **2**, eaal5196.

Chessa B., Pereira F., Arnaud F., Amorim A., Goyache F., Mainland I., Kao R. R., Pemberton J. M., Beraldi D., Stear M. J., Alberti A., Pittau M., Iannuzzi L., Banabazi M. H., Kazwala R. R., Zhang Y. P., Arranz J. J., Ali B. A., Wang Z., Uzun M., Dione M. M., Olsaker I., Holm L. E., Saarma U., Ahmad S., Marzanov N., Eythorsdottir E., Holland M. J., Paolo A. M., Bruford M. W., Kantanen J., Spencer T. E., Palmarini M., (2009) Revealing the history of sheep domestication using retrovirus integrations. *Science*.

Chiaramonte M. G., Donaldson D. D., Cheever A. W., Wynn T. A., (1999) An IL-13 inhibitor blocks the development of hepatic fibrosis during a T-helper type 2-dominated inflammatory response. *Journal of Clinical Investigation* **104**, 777–785.

Colditz I. G., (2003) Metabolic effects of host defence responses during gastrointestinal parasitism in sheep. *Australian Journal of Experimental Agriculture*.

Coltman D. W., Wilson K., Pilkington J. G., Stear M. J., Pemberton J. M., (2001) A microsatellite polymorphism in the gamma interferon gene is associated with resistance to gastrointestinal nematodes in a naturally-parasitized population of Soay sheep. *Parasitology*.

Costa-Silva J., Domingues D., Lopes F. M., (2017) RNA-Seq differential expression analysis: An extended review and a software tool. *PLoS ONE*.

Crawford, A.M. and McEwan, J.C., (1998) Identification of animals resistant to nematode parasite infection. 46.

Crawford A. M., Dodds K. G., Ede A. J., Pierson C. A., Montgomery G. W., Garmonsway H. G., Beattie A. E., Davies K., Maddox J. F., Kappes S. W., Stone R. T., Nguyen T. C., Penty J. M., Lord E. A., Broom J. E., Buitkamp J., Schwaiger W., Epplen J. T., Matthew P., (1995) An autosomal genetic linkage map of the sheep genome. *Genetics*.

Crawford A. M., Paterson K. a, Dodds K. G., Diez Tascon C., Williamson P. a, Roberts Thomson M., Bisset S. a, Beattie A. E., Greer G. J., Green R. S., Wheeler R., Shaw R. J., Knowler K., McEwan J. C., (2006) Discovery of quantitative trait loci for resistance to parasitic nematode infection in sheep: I. Analysis of outcross pedigrees. *BMC genomics* **7**, 178.

Dalrymple B. P., Kirkness E. F., Nefedov M., McWilliam S., Ratnakumar A., Barris W., Zhao S., Shetty J., Maddox J. F., O'Grady M., Nicholas F., Crawford A. M., Smith T., Jong P. J. de, McEwan J., Oddy V. H., Cockett N. E., (2007) Using comparative genomics to reorder the human genome sequence into a virtual sheep genome. *Genome Biology*.

Davies G., Stear M. J., Benothman M., Abuagob O., Kerr A., Mitchell S., Bishop S. C., (2006) Quantitative trait loci associated with parasitic infection in Scottish Blackface sheep. *Heredity*.

Dervishi E., Uriarte J., Valderrábano J., Calvo J. H., (2011) Structural and functional characterisation of the ovine interferon gamma (IFNG) gene: Its role in nematode resistance in Rasa Aragonesa ewes. *Veterinary Immunology and Immunopathology*.

Dhamija S., Menon M. B., (2018) Non-coding transcript variants of protein-coding genes–what are they good for? *RNA Biology*.

Diez-Tascón C., Keane O. M., Wilson T., Zadissa A., Hyndman D. L., Baird D. B., McEwan J. C., Crawford A. M., (2005) Microarray analysis of selection lines from outbred populations to identify genes involved with nematode parasite resistance in sheep. *Physiological Genomics*.

Dominik S., (2005) Quantitative trait loci for internal nematode resistance in sheep: a review. *Genetics Selection Evolution* **37 Suppl 1**, S83–S96.

Douch P. G. C., Green R. S., Morris C. A., Mcewan J. C., Windon R. G., (1996) Phenotypic markers for selection of nematode-resistant sheep. In: *International Journal for Parasitology*,

Duijvesteijn N., Bolormaa S., Gondro C., Clark S., Khansefid M., Moghaddar N., Swan A. A., Stothard P., Daetwyler H. D., Werf J. H. J. van der, MacLeod I., (2018) Genome-wide association study of meat quality traits using whole-genome sequence data in a multi-breed sheep population. *Proceedings of the World Congress on Genetics Applied to Livestock Production*.

El-Ashram S., Nasr I. Al, El-Kemary M., Mehmood R., Hu M., Suo X., (2017) Early and late gene expression profiles of the ovine mucosa in response to Haemonchus contortus infection employing Illumina RNA-seq technology. *Parasitology International*.

El-Ashram S., Li C., Abouhajer F., Mehmood R., Nasr I. Al, Zhang Y., Lu T., Yili D., Suo X., Haoji Z., Li Z., Huang S., (2018) An ex vivo abomasal ovine model to study the immediate immune response in the context of Haemonchus contortus larval-stage. *Veterinary Parasitology*.

Ellegren H., Chowdhary B. P., Johansson M., Marklund L., Fredholm M., Gustavsson I., Andersson L., (1994) A primary linkage map of the porcine genome reveals a low rate of genetic recombination. *Genetics*.

Erbe M., Hayes B. J., Matukumalli L. K., Goswami S., Bowman P. J., Reich C. M., Mason B. A., Goddard M. E., (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**, 4114–29.

Erickson N. A., Mundhenk L., Giovannini S., Glauben R., Heimesaat M. M., Gruber A. D., (2016) Role of goblet cell protein CLCA1 in murine DSS colitis. *Journal of Inflammation (United Kingdom)* **13**, .

Estrada-Reyes Z. M., Tsukahara Y., Goetsch A. L., Gipson T. A., Sahlu T., Puchala R., Mateescu R. G., (2019a) Association analysis of immune response loci related to Haemonchus contortus exposure in sheep and goats using a targeted approach. *Livestock Science*.

Estrada-Reyes Z. M., Rae O., Postley C., Jiménez Medrano M. B., Gutiérrez J. D. L., Mateescu R. G., (2019b) Association study reveals Th17, Treg and Th2 loci related to resistance to Haemonchus contortus in Florida Native sheep. *Journal of Animal Science*.

European Commission, (2003) Commission Decision of 13 February 2003 Laying Down Minimum Requirements for the Establishment of Breeding Programmes for Resistance to Transmissible Spongi- form Encephalopathies in Sheep (2003/100/EC). *Official Journal of the European Union*Off. J. L 041:41–45.

Fernández G., Baro J. A., La Fuente L. F. De, San Primitivo F., (1997) Genetic Parameters for Linear Udder Traits of Dairy Ewes. *Journal of Dairy Science*.

French A. T., Knight P. A., Smith W. D., Brown J. K., Craig N. M., Pate J. A., Miller H. R. P., Pemberton A. D., (2008) Up-regulation of intelectin in sheep after infection with Teladorsagia circumcincta. *International Journal for Parasitology* **38**, 467–475.

French A. T., Knight P. A., David Smith W., Pate J. A., Miller H. R. P., Pemberton A. D., (2009) Expression of three intelectins in sheep and response to a Th2 environment. *Veterinary Research* **40**, .

Galindo R. C., Ayoubi P., García-Pérez A. L., Naranjo V., Kocan K. M., Gortazar C., la Fuente J. de, (2008) Differential expression of inflammatory and immune response genes in sheep infected with Anaplasma phagocytophilum. *Veterinary Immunology and Immunopathology*.

García-Gámez E., Reverter A., Whan V., McWilliam S. M., Arranz J. J., International Sheep Genomics Consortium, Kijas J., (2011) Using regulatory and epistatic networks to extend the findings of a genome scan: Identifying the gene drivers of pigmentation in Merino sheep. *PLoS ONE*.

García-Gámez E., Sahana G., Gutiérrez-Gil B., Arranz J.-J. J., (2012a) Linkage disequilibrium and inbreeding estimation in Spanish Churra sheep. *BMC Genetics* **13**, 43.

García-Gámez E., Gutiérrez-Gil B., Sahana G., Sánchez J. P., Bayón Y., Arranz J. J., (2012b) GWA Analysis for Milk Production Traits in Dairy Sheep and Genetic Support for a QTN Influencing Milk Protein Percentage in the LALBA Gene. *PLoS ONE*.

García-Pérez A. L., Hurtado A., Oregui L. M., Juste R. A., (2002) Effects of a second annual strategic anthelmintic treatment in dairy sheep in Northern Spain. *Small Ruminant Research*.

Gentleman R. C., Carey V. J., Bates D. M., Bolstad B., Dettling M., Dudoit S., Ellis B., Gautier L., Ge Y., Gentry J., Hornik K., Hothorn T., Huber W., Iacus S., Irizarry R., Leisch F., Li C., Maechler M., Rossini A. J., Sawitzki G., Smith C., Smyth G., Tierney L., Yang J. Y. H., Zhang J., (2004) Bioconductor: Open software development for computational biology and bioinformatics. *Genome Biol* **5**, .

Georges M., Nielsen D., Mackinnon M., Mishra A., Okimoto R., Pasquino A. T., Sargeant L. S., Sorensen A., Steele M. R., Zhao X., Womack J. E., Hoeschele I., (1995) Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing. *Genetics*.

Ghosh M., Sharma N., Singh A. K., Gera M., Pulicherla K. K., Jeong D. K., (2018) Transformation of animal genomics by next-generation sequencing technologies: a decade of challenges and their impact on genetic architecture. *Critical Reviews in Biotechnology*.

Gianola D., (2013) Priors in whole-genome regression: The Bayesian alphabet returns. *Genetics*.

Gomez-Raya L., Rodríguez C., Barragán C., Silió L., (2015) Genomic inbreeding coefficients based on the distribution of the length of runs of homozygosity in a closed line of Iberian pigs. *Genetics, selection, evolution : GSE* **47**, 81.

Gossner A. G., Venturina V. M., Shaw D. J., Pemberton J. M., Hopkins J., (2012) Relationship between susceptibility of Blackface sheep to Teladorsagia circumcincta infection and an inflammatory mucosal T cell response. *Veterinary Research*.

Gossner A., Wilkie H., Joshi A., Hopkins J., (2013) Exploring the abomasal lymph node transcriptome for genes associated with resistance to the sheep nematode *Teladorsagia circumcincta*. *Veterinary Research* **44**, .

Grau-Sologestoa I., (2015) Livestock management in Spain from Roman to post-medieval times: A biometrical analysis of cattle, sheep/goat and pig. *Journal of Archaeological Science*.

Gray G. D., (1997) The use of genetically resistant sheep to control nematode parasitism. In: *Veterinary Parasitology*,

Grisart B., Coppieters W., Farnir F., Karim L., Ford C., Berzi P., Cambisano N., Mni M., Reid S., Simon P., Spelman R., Georges M., Snell R., (2002) Positional candidate cloning of a QTL in dairy cattle: Identification of a missense mutation in the bovine DGAT1 gene with major effect on milk yield and composition. *Genome Research*.

Guo Z., González J. F., Hernandez J. N., McNeilly T. N., Corripio-Miyar Y., Frew D., Morrison T., Yu P., Li R. W., (2016) Possible mechanisms of host resistance to Haemonchus contortus infection in sheep breeds native to the Canary Islands. *Scientific reports* **6**, .

Gutiérrez-Gil B., El-Zarei M. F., Bayón Y., Álvarez L., La Fuente L. F. De, San Primitivo F., Arranz J. J., (2007) Short communication: Detection of quantitative trait loci influencing somatic cell score in Spanish churra sheep. *Journal of Dairy Science*.

Gutiérrez-Gil B., Pérez J., Álvarez L., Martínez-Valladares M., la Fuente L.-F. F. de, Bayán Y., Meana A., Primitivo F. S., Rojo-Vzquez F. A., Arranz J.-J. J., Bayón Y., Meana A., Primitivo F. S., Rojo-Vázquez F.-A., Arranz J.-J. J., (2009) Quantitative trait loci for resistance to trichostrongylid infection in Spanish Churra sheep. *Genetics Selection Evolution* **41**, 46.

Gutiérrez-Gil B., Pérez J., La Fuente L. F. De, Meana A., Martínez-Valladares M., San Primitivo F., Rojo-Vázquez F. A., Arranz J. J., (2010) Genetic parameters for resistance to trichostrongylid infection in dairy sheep. *Animal* **4**, 505–512.

Gutiérrez-Gil B., Alvarez L., la Fuente L. F. de, Sanchez J. P., San Primitivo F., Arranz J. J., (2011) A genome scan for quantitative trait loci affecting body conformation traits in Spanish Churra dairy sheep. *Journal of Dairy Science*.

Gutiérrez-Gil B., Arranz J. J., Pong-Wong R., García-Gámez E., Kijas J., Wiener P., (2014) Application of selection mapping to identify genomic regions associated with dairy production in sheep. *PLoS ONE* **9**, .

Gutiérrez-Gil B., Arranz J. J., Wiener P., (2015) An interpretive review of selective sweep studies in Bos taurus cattle populations: Identification of unique and shared selection signals across breeds. *Frontiers in Genetics*.

Gutiérrez-Gil B., Esteban-Blanco C., Wiener P., Chitneedi P. K., Suarez-Vega A., Arranz J. J., (2017) High-resolution analysis of selection sweeps identified between fine-wool Merino and coarse-wool Churra sheep breeds. *Genetics Selection Evolution*.

Gutiérrez-Gil B., Esteban-Blanco C., Suarez-Vega A., Arranz J. J., (2018) Detection of quantitative trait loci and putative causal variants affecting somatic cell score in dairy sheep by using a 50K SNP-Chip and whole genome sequencing. *Journal of Dairy Science*.

Hamadeh H., Afshari C. A., (2000) Gene chips and functional genomics. *American Scientist*.

Hartung J., (2013) A short history of livestock production. In: *Livestock Housing: Modern Management to Ensure Optimal Health and Welfare of Farm Animals*,

Hayes B. J., Bowman P. J., Daetwyler H. D., Kijas J. W., Werf J. H. J. van der, (2012) Accuracy of genotype imputation in sheep breeds. *Animal Genetics* **43**, 72–80.

Hezroni H., Koppstein D., Schwartz M. G., Avrutin A., Bartel D. P., Ulitsky I., (2015) Principles of Long Noncoding RNA Evolution Derived from Direct Comparison of Transcriptomes in 17 Species. *Cell Reports*.

Hu Z. L., Park C. A., Wu X. L., Reecy J. M., (2013) Animal QTLdb: An improved database tool for livestock animal QTL/association data dissemination in the post-genome era. *Nucleic Acids Research* **41**, 871–879.

Hu Z.-L. L., Park C. A., Reecy J. M., (2016) Developmental progress and current status of the Animal QTLdb. *Nucleic Acids Research* **44**, D827–D833.

Ingham A., Reverter A., Windon R., Hunt P., Menzies M., (2008) Gastrointestinal nematode challenge induces some conserved gene expression changes in the gut mucosa of genetically resistant sheep. *International Journal for Parasitology* **38**, 431–442.

International Sheep Genomics Consortium T., Archibald A. L., Cockett N. E., Dalrymple B. P., Faraut T., Kijas J. W., Maddox J. F., McEwan J. C., Hutton Oddy V., Raadsma H. W., Wade C., Wang J., Wang W., Xun X., (2010) The sheep genome reference sequence: A work in progress. *Animal Genetics*.

Johnston J., Kistemaker G., Sullivan P. G., (2011) Comparison of Different Imputation Methods. *Interbull Bulletin*25–33.

Kadarmideen H. N., (2014) Genomics to systems biology in animal and veterinary sciences: Progress, lessons and opportunities. *Livestock Science*.

Kalaldeh M. Al, Gibson J., Duijvesteijn N., Daetwyler H. D., MacLeod I., Moghaddar N., Lee S. H., Werf J. H. J. van der, (2019a) Using imputed whole-genome sequence data to improve the accuracy of genomic prediction for parasite resistance in Australian sheep. *Genetics, selection, evolution : GSE*.

Kalaldeh M. Al, Gibson J., Lee S. H., Gondro C., Werf J. H. J. Van Der, (2019b) Detection of genomic regions underlying resistance to gastrointestinal parasites in Australian sheep. *Genetics Selection Evolution*.

Keane O. M., Zadissa A., Wilson T., Hyndman D. L., Greer G. J., Baird D. B., McCulloch A. F., Crawford A. M., McEwan J. C., (2006) Gene expression profiling of naive sheep genetically resistant and susceptible to gastrointestinal nematodes. *BMC Genomics* **7**, 42.

Kemper K. E., Goddard M. E., (2012) Understanding and predicting complex traits: Knowledge from cattle. *Human Molecular Genetics*.

Kernaleguen M., Daviaud C., Shen Y., Bonnet E., Renault V., Deleuze J. F., Mauger F., Tost J., (2018) Whole-genome bisulfite sequencing for the analysis of genome-wide dna methylation and hydroxymethylation patterns at single-nucleotide Resolution. In: *Methods in Molecular Biology*,

Khan W. I., Collins S. M., (2004) Immune-mediated alteration in gut physiology and its role in host defence in nematode infection. *Parasite Immunology*.

Kijas J. W., Lenstra J. a., Hayes B., Boitard S., Porto Neto L. R., San Cristobal M., Servin B., McCulloch R., Whan V., Gietzen K., Paiva S., Barendse W., Ciani E., Raadsma H., McEwan J., Dalrymple B., (2012) Genome-Wide Analysis of the World's Sheep Breeds Reveals High Levels of Historic Mixture and Strong Recent Selection. *PLoS Biology* **10**, e1001258.

Kijas J. W., Porto-Neto L., Dominik S., Reverter A., Bunch R., McCulloch R., Hayes B. J., Brauning R., McEwan J., (2014) Linkage disequilibrium over short physical distances measured in sheep using a high-density SNP chip. *Animal genetics* **45**, 754–7.

Kirin M., McQuillan R., Franklin C. S., Campbell H., Mckeigue P. M., Wilson J. F., (2010) Genomic runs of homozygosity record population history and consanguinity. *PLoS ONE* **5**, 1–7.

Knight P. A., Pate J., Smith W. D., Miller H. R. P., (2007) An ovine chitinase-like molecule, chitinase-3 like-1 (YKL-40), is upregulated in the abomasum in response to challenge with the gastrointestinal nematode, Teladorsagia circumcincta. *Veterinary Immunology and Immunopathology*.

Knight J. S., Baird D. B., Hein W. R., Pernthaner A., (2010) The gastrointestinal nematode Trichostrongylus colubriformis down-regulates immune gene expression in migratory cells in afferent lymph. *BMC Immunology*.

Knight P. A., Griffith S. E., Pemberton A. D., Pate J. M., Guarneri L., Anderson K., Talbot R. T., Smith S., Waddington D., Fell M., Archibald A. L., Burgess S. T., Smith D. W., Miller H. R., Morrison I. W., (2011) Novel gene expression responses in the ovine abomasal mucosa to infection with the gastric nematode *Teladorsagia circumcincta*. *Veterinary Research* **42**, .

Krawczyk a., Słota E., (2009) Genetic markers to gastrointestinal nematode resistance in sheep: a review. *Helminthologia* **46**, 3–8.

la Fuente L. F. de, Baro J. A., San Primitivo F., (1995) Breeding programme for the Spanish Churra sheep breed. *Cahiers Options Méditerranéennes*.

la Fuente L. F. de, Gonzalo C., Sánchez J. P., Rodríguez R., Carriedo J. A., San Primitivo F., (2011) Genetic parameters of the linear body conformation traits and genetic correlations with udder traits, milk yield and composition, and somatic cell count in dairy ewes. *Canadian Journal of Animal Science* **91**, 585–591.

Legaz E., Álvarez I., Royo L. J., Fernández I., Gutiérrez J. P., Goyache F., (2008) Genetic relationships between Spanish Assaf (Assaf.E) and Spanish native dairy sheep breeds. *Small Ruminant Research*.

Li M. H., Strandén I., Kantanen J., (2009a) Genetic diversity and pedigree analysis of the Finnsheep breed. *Journal of animal science*.

Li Y., Willer C., Sanna S., Abecasis G., (2009b) Genotype Imputation. *Annual Review of Genomics and Human Genetics*.

Li A., Zhang J., Zhou Z., (2014) PLEK: A tool for predicting long non-coding RNAs and messenger RNAs based on an improved k-mer scheme. *BMC Bioinformatics* **15**, .

Love M. I., Huber W., Anders S., (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*.

Ma H., Hao Y., Dong X., Gong Q., Chen J., Zhang J., Tian W., (2012) Molecular mechanisms and function prediction of long noncoding RNA. *The Scientific World Journal*.

Ma P., Brøndum R. F., Zhang Q., Lund M. S., Su G., (2013) Comparison of different methods for imputing genome-wide marker genotypes in Swedish and Finnish Red Cattle. *Journal of Dairy Science*.

MacKinnon K. M., Burton J. L., Zajac A. M., Notter D. R., (2009) Microarray analysis reveals difference in gene expression profiles of hair and wool sheep infected with Haemonchus contortus. *Veterinary Immunology and Immunopathology*.

Magi A., Benelli M., Gozzini A., Girolami F., Torricelli F., Brandi M. L., (2010) Bioinformatics for next generation sequencing data. *Genes*.

Marioni J. C., Mason C. E., Mane S. M., Stephens M., Gilad Y., (2008) RNA-seq: An assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res* **18**, .

Marshall K., Maddox J. F., Lee S. H., Zhang Y., Kahn L., Graser H. U., Gondro C., Walkden-Brown S. W., Werf J. H. J. Van Der, (2009) Genetic mapping of quantitative trait loci for resistance to haemonchus contortus in sheep. *Animal Genetics*.

Martin H., (2010) Role of PPAR-gamma in inflammation. Prospects for therapeutic intervention by food components. *Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis* **690**, 57–63.

Martínez-González B., Díez-Baños N., Rojo-Vázquez F. A., (1998) An epidemiological study of gastrointestinal parasitism in dairy sheep flocks in León (NW Spain). *Small Ruminant Research*.

Martínez-Valladares M., Robles-Pérez D., Martínez-Pérez J. M., Cordero-Pérez C., Famularo M. D. R., Fernández-Pato N., González-Lanza C., Castañón-Ordóñez L., Rojo-Vázquez F. A., (2013a) Prevalence of gastrointestinal nematodes and Fasciola hepatica in sheep in the northwest of Spain: Relation to climatic conditions and/or man-made environmental modifications. *Parasites and Vectors*.

Martínez-Valladares M., Martínez-Pérez J. M., Robles-Pérez D., Cordero-Pérez C., Famularo M. R., Fernández-Pato N., Castañón-Ordóñez L., Rojo-Vázquez F. A., (2013b) The present status of anthelmintic resistance in gastrointestinal nematode infections of sheep in the northwest of Spain by in vivo and in vitro techniques. *Veterinary Parasitology*.

Mavrot F., Hertzberg H., Torgerson P., (2015) Effect of gastro-intestinal nematode infection on sheep performance: A systematic review and meta-analysis. *Parasites and Vectors*.

McRae K. M., Good B., Hanrahan J. P., McCabe M. S., Cormican P., Sweeney T., O'Connell M. J., Keane O. M., (2016) Transcriptional profiling of the ovine abomasal lymph node reveals a role for timing of the immune response in gastrointestinal nematode resistance. *Veterinary Parasitology* **224**, 96–108.

Menzies M., Reverter A., Andronicos N., Hunt P., Windon R., Ingham A., (2010) Nematode challenge induces differential expression of oxidant, antioxidant and mucous genes down the longitudinal axis of the sheep gut. *Parasite Immunology* **32**, 36–46.

Mercer T. R., Dinger M. E., Mattick J. S., (2009) Long non-coding RNAs: Insights into functions. *Nature Reviews Genetics*.

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

Meuwissen T., (2009) Genetic management of small populations: A review. *Acta Agriculturae Scandinavica, Section A — Animal Science* **59**, 71–79.

Mignon-Grasteau S., Boissy A., Bouix J., Faure J. M., Fisher A. D., Hinch G. N., Jensen P., Neindre P. Le, Mormède P., Prunet P., Vandeputte M., Beaumont C., (2005) Genetics of adaptation and domestication in livestock. In: *Livestock Production Science*,

Mirkena T., Duguma G., Haile A., Tibbo M., Okeyo A. M., Wurzinger M., Sölkner J., (2010) Genetics of adaptation in domestic farm animals: A review. *Livestock Science*.

Mpetile Z., Cloete S. W. P., Kruger A. C. M., Dzama K., (2015) Environmental and genetic factors affecting faecal worm egg counts in Merinos divergently selected for reproduction. *South African Journal of Animal Sciences*.

Mrode R. A., Thompson R., (2005) *Linear models for the prediction of animal breeding values: Second Edition*.

Nagaraj S. H., Harsha H. C., Reverter A., Colgrave M. L., Sharma R., Andronicos N., Hunt P., Menzies M., Lees M. S., Sekhar N. R., Pandey A., Ingham A., (2012) Proteomic analysis of the abomasal mucosal response following infection by the

nematode, *Haemonchus contortus*, in genetically resistant and susceptible sheep. *Journal of Proteomics* **75**, 2141–2152.

Newton S. E., Meeusen E. N. T., (2003) Progress and new technologies for developing vaccines against gastrointestinal nematode parasites of sheep. *Parasite Immunology*.

Nezer C., Moreau L., Brouwers B., Coppieters W., Detilleux J., Hanset R., Karim L., Kvasz A., Leroy P., Georges M., (1999) An imprinted QTL with a major effect on muscle mass and fat deposition maps to the IGF2 locus in pig [1]. *Nature Genetics*.

Ngere L., Burke J. M., Morgan J. L. M., Miller J. E., Notter D. R., (2018) Genetic parameters for fecal egg counts and their relationship with body weights in katahdin lambs. *Journal of Animal Science*.

Nicholas F. W., Hobbs M., (2014) Mutation discovery for Mendelian traits in non-laboratory animals: A review of achievements up to 2012. *Animal Genetics*.

O'Connor L. J., Walkden-Brown S. W., Kahn L. P., (2006) Ecology of the free-living stages of major trichostrongylid parasites of sheep. *Veterinary Parasitology*.

Ong K. K., Elks C. E., Li S., Zhao J. H., Luan J., Lars B., Gillson C. J., Glaser B., Golding J., Hardy R., Khaw K., (2010) Europe PMC Funders Group Genetic variation in LIN28B is associated with the timing of puberty. *nature genetics* **41**, 729–733.

Othmane M. H., La Fuente L. F. De, Carriedo J. A., San Primitivo F., (2002) Heritability and genetic correlations of test day milk yield and composition, individual laboratory cheese yield, and somatic cell count for dairy ewes. *Journal of Dairy Science*.

Pang K. C., Dinger M. E., Mercer T. R., Malquori L., Grimmond S. M., Chen W., Mattick J. S., (2009) Genome-Wide Identification of Long Noncoding RNAs in CD8 + T Cells . *The Journal of Immunology*.

Patra G., Jas R., Ghosh J., Borthakur S. K., Paul A., (2016) Single nucleotide polymorphism and expression studies of the interferon gamma gene and its role against Haemonchus contortus in Garole and Sahabadi sheep. *Asian Pacific Journal of Tropical Disease*.

Pedrosa S., Uzun M., Arranz J. J., Gutiérrez-Gil B., San Primitivo F., Bayón Y., (2005) Evidence of three maternal lineages in near eastern sheep supporting multiple domestication events. *Proceedings of the Royal Society B: Biological Sciences*.

Peñagaricano F., Zorrilla P., Naya H., Robello C., Urioste J. I., (2012) Gene expression analysis identifies new candidate genes associated with the development of black skin spots in Corriedale sheep. *Journal of Applied Genetics*.

Periasamy K., Pichler R., Poli M., Cristel S., Cetrá B., Medus D., Basar M., Thiruvenkadan A. K., Ramasamy S., Ellahi M. B., Mohammed F., Teneva A., Shamsuddin M., Podesta M. G., Diallo A., (2014) Candidate gene approach for parasite resistance in sheep - Variation in immune pathway genes and association with fecal egg count. *PLoS ONE*.

Piskol R., Ramaswami G., Li J. B., (2013) Reliable identification of genomic variants from RNA-seq data. *American Journal of Human Genetics*.

Purcell S., Neale B., Todd-Brown K., Thomas L., Ferreira M. A. R., Bender D., Maller J., Sklar P., Bakker P. I. W. de, Daly M. J., Sham P. C., (2007) PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics* **81**, 559–575.

Ransohoff J. D., Wei Y., Khavari P. A., (2018) The functions and unique features of long intergenic non-coding RNA. *Nature Reviews Molecular Cell Biology*.

Robinson M. D., McCarthy D. J., Smyth G. K., (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140.

Rowe A., Gondro C., Emery D., Sangster N., (2008) Genomic analyses of Haemonchus contortus infection in sheep: Abomasal fistulation and two Haemonchus strains do not substantially confound host gene expression in microarrays. *Veterinary Parasitology*.

Rowe A., Gondro C., Emery D., Sangster N., (2009) Sequential microarray to identify timing of molecular responses to Haemonchus contortus infection in sheep. *Veterinary Parasitology* **161**, 76–87.

Rubin C. J., Megens H. J., Barrio A. M., Maqbool K., Sayyab S., Schwochow D., Wang C., Carlborg Ö., Jern P., Jørgensen C. B., Archibald A. L., Fredholm M., Groenen M. A. M., Andersson L., (2012) Strong signatures of selection in the domestic pig genome. *Proceedings of the National Academy of Sciences of the United States of America*.

Rupp R., Senin P., Sarry J., Allain C., Tasca C., Ligat L., Portes D., Woloszyn F., Bouchez O., Tabouret G., Lebastard M., Caubet C., Foucras G., Tosser-Klopp G., (2015) A Point Mutation in Suppressor of Cytokine Signalling 2 (Socs2) Increases the Susceptibility to Inflammation of the Mammary Gland while Associated with Higher Body Weight and Size and Higher Milk Production in a Sheep Model. *PLoS Genetics*.

Saarialho-Kere U., Kerkelä E., Jahkola T., Suomela S., Keski-Oja J., Lohi J., (2002) Epilysin (MMP-28) expression is associated with cell proliferation during epithelial repair. *Journal of Investigative Dermatology* **119**, 14–21.

Sallé G., Moreno C., Boitard S., Ruesche J., Tircazes-Secula A., Bouvier F., Aletru M., Weisbecker J. L., Prévot F., Bergeaud J. P., Trumel C., Grisez C., Liénard E., Jacquiet P., (2014) Functional investigation of a QTL affecting resistance to Haemonchus contortus in sheep. *Veterinary Research*.

Sargolzaei M., Chesnais J. P., Schenkel F., (2012) Efficient combined family and population imputation in large data sets. *Open Industry Session, October 30, 2012*1–10.

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

Satterfield M. C., Song G., Kochan K. J., Riggs P. K., Simmons R. M., Elsik C. G., Adelson D. L., Bazer F. W., Zhou H., Spencer T. E., (2009) Discovery of candidate

genes and pathways in the endometrium regulating ovine blastocyst growth and conceptus elongation. *Physiological Genomics*.

Sayers G., Sweeney T., (2005) Gastrointestinal nematode infection in sheep – a review of the alternatives to anthelmintics in parasite control. *Animal Health Research Reviews*.

Sayers G., Good B., Hanrahan J. P., Ryan M., Angles J. M., Sweeney T., (2005) Major Histocompatibility Complex DRB1 gene: Its role in nematode resistance in Suffolk and Texel sheep breeds. *Parasitology* **131**, 403–409.

SCHALLIG H. D. F. H., (2000) Immunological responses of sheep to Haemonchus contortus . *Parasitology*.

Schmitz S. U., Grote P., Herrmann B. G., (2016) Mechanisms of long noncoding RNA function in development and disease. *Cellular and Molecular Life Sciences*.

Schuster S. C., (2008) Next-generation sequencing transforms today's biology. *Nature Methods*.

Schwaiger F. W., Gostomski D., Stear M. J., Duncan J. L., McKellar Q. A., Epplen J. T., Buitkamp J., (1995) An ovine Major histocompatibility complex DRB1 allele is associated with low faecal egg counts following natural, predominantly Ostertagia circumcincta infection. *International Journal for Parasitology*.

Sen A., Rothenberg M. E., Mukherjee G., Feng N., Kalisky T., Nair N., Johnstone I. M., Clarke M. F., Greenberg H. B., (2012) Innate immune response to homologous rotavirus infection in the small intestinal villous epithelium at single-cell resolution. *Proceedings of the National Academy of Sciences* **109**, 20667–20672.

Sigdel K. R., Cheng A., Wang Y., Duan L., Zhang Y. L., (2015) The Emerging Functions of Long Noncoding RNA in Immune Cells: Autoimmune Diseases. *Journal of Immunology Research* **2015**, .

Silva M. V. B., Sonstegard T. S., Hanotte O., Mugambi J. M., Garcia J. F., Nagda S., Gibson J. P., Iraqi F. A., McClintock A. E., Kemp S. J., Boettcher P. J., Malek M., Tassell C. P. Van, Baker R. L., (2012) Identification of quantitative trait loci affecting resistance to gastrointestinal parasites in a double backcross population of Red Maasai and Dorper sheep. *Animal Genetics* **43**, 63–71.

Spelman R. J., Ford C. A., McElhinney P., Gregory G. C., Snell R. G., (2002) Characterization of the DGAT1 gene in the New Zealand dairy population. *Journal of Dairy Science*.

Stark R., Grzelak M., Hadfield J., (2019) RNA sequencing: the teenage years. *Nature Reviews Genetics*.

Stevenson L. M., Colditz I. G., LeJambre L. F., (2001) Expression of cell surface adhesion molecules by peripheral blood eosinophils during Trichostrongylus colubriformis infection in sheep. *Immunology and Cell Biology*.

Suarez-Vega A., Gutierrez-Gil B., Klopp C., Tosser-Klopp G., Arranz J. J., (2016) Comprehensive RNA-Seq profiling to evaluate lactating sheep mammary gland transcriptome. *Scientific Data*.

Suaŕez-Vega A., Gutieŕrez-Gil B., Cuchillo-Ibañez I., Saéz-Valero J., Pérez V., Garciá-Gámez E., Benavides J., Arranz J. J., (2013) Identification of a 31-bp deletion in the RELN gene causing lissencephaly with cerebellar hypoplasia in sheep. *PLoS ONE*.

Suárez-Vega A., Gutiérrez-Gil B., Benavides J., Perez V., Tosser-Klopp G., Klopp C., Keennel S. J., Arranz J. J., (2015a) Combining GWAS and RNA-seq approaches for detection of the causal mutation for hereditary Junctional epidermolysis bullosa in sheep. *PLoS ONE* **10**, .

Suárez-Vega A., Gutiérrez-Gil B., Klopp C., Robert-Granie C., Tosser-Klopp G., Arranz J. J., (2015b) Characterization and comparative analysis of the milk transcriptome in two dairy sheep breeds using RNA sequencing. *Scientific Reports*.

Suárez-Vega A., Gutiérrez-Gil B., Klopp C., Tosser-Klopp G., Arranz J. J., (2017) Variant discovery in the sheep milk transcriptome using RNA sequencing. *BMC Genomics*.

Sun L., Luo H., Bu D., Zhao G., Yu K., Zhang C., Liu Y., Chen R., Zhao Y., (2013) Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. *Nucleic Acids Research* **41**, .

Sykes A. R., Coop R. L., (2001) Interaction between nutrition and gastrointestinal parasitism in sheep. *New Zealand Veterinary Journal* **49**, 222–226.

Vaiman D., Schibler L., Bourgeois F., Oustry A., Amigues Y., Cribiu E. P., (1996) A genetic linkage map of the male goat genome. *Genetics*.

VanRaden P. M., Tassell C. P. Van, Wiggans G. R., Sonstegard T. S., Schnabel R. D., Taylor J. F., Schenkel F. S., (2009) Invited review: Reliability of genomic predictions for North American Holstein bulls. *Journal of Dairy Science*.

VanRaden P. M., Null D. J., Sargolzaei M., Wiggans G. R., Tooker M. E., Cole J. B., Sonstegard T. S., Connor E. E., Winters M., Kaam J. B. C. H. M. van, Valentini a, Doormaal B. J. Van, Faust M. a, Doak G. a, (2013) Genomic imputation and evaluation using high-density Holstein genotypes. *Journal of dairy science* **96**, 668–678.

Vasta G. R., Quesenberry M., Ahmed H., O'Leary N., (1999) C-type lectins and galectins mediate innate and adaptive immune functions: Their roles in the complement activation pathway. *Developmental and Comparative Immunology* **23**, 401–420.

Vaysse A., Ratnakumar A., Derrien T., Axelsson E., Pielberg G. R., Sigurdsson S., Fall T., Seppälä E. H., Hansen M. S. T., Lawley C. T., Karlsson E. K., Bannasch D., Vilà C., Lohi H., Galibert F., Fredholm M., Häggström J., Hedhammar Å., André C., Lindblad-Toh K., Hitte C., Webster M. T., (2011) Identification of genomic regions associated with phenotypic variation between dog breeds using selection mapping. *PLoS Genetics*.

Vieira M. L. C., Santini L., Diniz A. L., Munhoz C. de F., (2016) Microsatellite markers: What they mean and why they are so useful. *Genetics and Molecular Biology*.

Visscher P. M., Wray N. R., Zhang Q., Sklar P., McCarthy M. I., Brown M. A., Yang J., (2017) 10 Years of GWAS Discovery: Biology, Function, and Translation. *American Journal of Human Genetics* **101**, 5–22.

Wang Z., Gerstein M., Snyder M., (2009) RNA-Seq: A revolutionary tool for transcriptomics. *Nature Reviews Genetics* **10**, 57–63.

White G. P., Meeusen E. N. T., Newton S. E., (2001) A single-chain variable region immunoglobulin library from the abomasal lymph node of sheep infected with the gastrointestinal nematode parasite Haemonchus contortus. *Veterinary Immunology and Immunopathology*.

Wickramasinghe S., Cánovas A., Rincón G., Medrano J. F., (2014) RNA-Sequencing: A tool to explore new frontiers in animal genetics. *Livestock Science*.

Wikel S. K., (1997) Immunity to Parasites: How Parasitic Infections are Controlled. Derek Wakelin . *The Quarterly Review of Biology*.

Womack J. E., Jang H. J., Lee M. O., (2012) Genomics of complex traits. *Annals of the New York Academy of Sciences*.

Wucher V., Legeai F., Hédan B., Rizk G., Lagoutte L., Leeb T., Jagannathan V., Cadieu E., David A., Lohi H., Cirera S., Fredholm M., Botherel N., Leegwater P. A. J., Béguec C. Le, Fieten H., Johnson J., Alföldi J., André C., Lindblad-Toh K., Hitte C., Derrien T., (2017) FEELnc: A tool for long non-coding RNA annotation and its application to the dog transcriptome. *Nucleic Acids Research*.

Yang Y., Zhou Q. J., Chen X. Q., Yan B. L., Guo X. L., Zhang H. L., Du A. F., (2015) Profiling of differentially expressed genes in sheep T lymphocytes response to an artificial primary Haemonchus contortus infection. *Parasites and Vectors*.

Zaros L. G., Coutinho L. L., Sider L. H., Medeiros H. R. de, Neves M. R. M. das, Benvenuti C. L., Navarro A. M. do C., Vieira L. da S., (2010) Evaluation of reference genes for real-time PCR studies of Brazilian Somalis sheep infected by gastrointestinal nematodes. *Genetics and Molecular Biology*.

Zhang Z. H., Jhaveri D. J., Marshall V. M., Bauer D. C., Edson J., Narayanan R. K., Robinson G. J., Lundberg A. E., Bartlett P. F., Wray N. R., Zhao Q. Y., (2014) A comparative study of techniques for differential expression analysis on RNA-seq data. *PLoS ONE* **9**, .

Zhang L., Sun F., Jin H., Dalrymple B. P., Cao Y., Wei T., Vuocolo T., Zhang M., Piao Q., Ingham A. B., (2017) A comparison of transcriptomic patterns measured in the skin of Chinese fine and coarse wool sheep breeds. *Scientific Reports*.

Zhang R., Liu F., Hunt P., Li C., Zhang L., Ingham A., Li R. W., (2019) Transcriptome analysis unraveled potential mechanisms of resistance to Haemonchus contortus infection in Merino sheep populations bred for parasite resistance. *Veterinary Research*.

Zhou W., Duan Z., Yang B., Xiao C., (2017) The effective regulation of pro-and anti-inflammatory cytokines induced by combination of PA-MSHA and BPIFB1 in initiation of innate immune responses. *Open Medicine (Poland)* **12**, 299–307.