Genetics Selection Evolution

Article in Press

Claudin and Rab proteins are key molecular components involved in coccidiosis resistance in Portuguese Merino sheep

Endika Varela-Martínez, Ana Afonso, Dimitra Mainou, Fábio Teixeira, Telmo Nunes, Pedro Vieira, Inês Sarraguça, Cristina Martins, Natalia Campbell, Rafael Cordeiro Silva, Tiago Perloiro, Luís Madeira Carvalho, Ana Cristina Ferreira, Luís Telo Gama, Helga Waap & Andreia J. Amaral

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

Claudin and Rab proteins are key molecular components involved in coccidiosis resistance in Portuguese Merino Sheep

Endika Varela-Martínez[1], Ana Afonso[2,3,4], Dimitra Mainou[5], Fábio Teixeira[6,7,8], Telmo Nunes[9,10], Pedro Vieira[11], Inês Sarraguça[9,10], Cristina Martins[2], Natalia Campbell[2,6], Rafael Cordeiro da Silva[12], Tiago Perloiro[12], Luís Madeira de Carvalho[9,10], Ana Cristina Ferreira[2,11,13], Luís Telo da Gama[9,10], Helga Waap[2,9,10,$], Andreia J. Amaral[6,7,9,10,$,*]

[1]Department of Genetics, Physical Anthropology and Animal Physiology, Faculty of Science and Technology, University of the Basque Country (UPV/EHU), Bº Sarriena, Leioa, 48940, Spain.

[2]Instituto Nacional de Investigação Agrária e Veterinária (INIAV), Av. da República, Quinta do Marquês, 2780-157 Oeiras, Portugal

[3]Global Health and Tropical Medicine, GHTM, Associate Laboratory in Translation and Innovation Towards Global Health, LA-REAL, Instituto de Higiene e Medicina Tropical, IHMT, Universidade NOVA de Lisboa, UNL, Rua da Junqueira 100, 1349-008 Lisboa, Portugal

[4]Instituto Politécnico de Santarém, Escola Superior de Saúde, Complexo Andaluz, Apartado 279, 2001-904 Santarém

[5]Faculty of Veterinary Medicine, Aristotle University of Thessaloniki, Greece

[6]MED—Mediterranean Institute for Agriculture, Environment and Development and CHANGE – Global Change and Sustainability Institute, University of Évora, Polo da Mitra, Ap. 94, 7006-554, Évora, Portugal

[7]School of Science and Technology, University of Evora, Evora, Portugal

[8]Faculty of Veterinary Medicine, University José Eduardo dos Santos, Huambo, Angola

[9]CIISA - Centre for Interdisciplinary Research in Animal Health, Faculty of Veterinary Medicine, University of Lisbon, Avenida da Universidade Técnica, 1300-477 Lisbon, Portugal

[10]Associate Laboratory for Animal and Veterinary Sciences (AL4AnimalS), 1300-477 Lisbon, Portugal

[11]Faculdade de Medicina Veterinária, Universidade Lusófona de Humanidades e Tecnologias , Centro Universitário de Lisboa, Campo Grande, 1749-024 Lisboa, Portugal

[12]Associação Nacional de Criadores de Ovinos de Raça Merina (ANCORME), Évora, Portugal

[13]BioISI – Instituto de Biosistemas e Ciências Integrativas, Faculdade de Ciências, Universidade de Lisboa, 1749-016 Lisboa, Portugal

$Equal contributions

*Corresponding author

Email addresses:

EV-M: endika.varela@ehu.eus

AA: anajuliaafonso74@gmail.com

DM: dimimainou@gmail.com

FT: Fabio.teixeira@uevora.pt

TN: tnunes@fmv.ulisboa.pt

PV: pedro.nunes.vieira@gmail.com

IS: Ines.soares22@hotmail.com

CM: cristina.martins@iniav.pt

NC: natalia.campbell@iniav.pt

RCS: rsilva@ancorme.com

TP: tperloiro@ancorme.com

LMC: madeiradecarvalho@fmv.ulisboa.pt

ACF: cristina.ferreira@iniav.pt

LTG: ltgama@fmv.ulisboa.pt

HW: helga.waap@rivm.nl

AJA: andreia.fonseca@uevora.pt

Abstract

Background

Although coccidial infection is often asymptomatic in sheep, both clinical and subclinical forms of the disease are linked to considerable production losses, mainly in young lambs. Studies aiming to identify genetic markers for use in selection programs towards increasing genetic resistance to coccidiosis are lacking and have yet to be performed in Portuguese Merino sheep. The purpose of this study was to identify genomic regions associated with resistance to coccidiosis by conducting a genome-wide association study (GWAS) in Portuguese Merino sheep.

Results

From an initial population of 1,022 sheep having known phenotypic characteristics, 206 and 202 distinct animals were genotyped using 50K and 600K Single Nucleotide Polymorphism (SNP) arrays, respectively. After the 50K array was imputed using a 600K array as reference, an association analysis was performed for faecal oocyst counts (FOC). We identified 12 SNPs that were significantly associated with resistance by using a chromosome-wide significance threshold. The significant SNPs were related to *Ccser1, Thsd4, Eci1, Tnfrsf12a, Chrm3* and *Slc20a2* genes. We identified 80 candidate genes located in the proximity of the significant SNPs using the defined confidence regions. Two types of gene set enrichment analyses were performed. Enrichment based on the set of candidate genes, identified the terms virus receptor activity and exogenous protein binding to be enriched, both due to two claudins, *CLDN6* and *CLDN9*. Enrichment based on gene interactions, showed enrichment of terms related to transport vesicles, mainly due to the presence of Rab proteins.

Conclusions

Given the role that Rab and Claudins play in host-parasite relationships, these results suggest the existence of reliable markers associated with resistance to coccidiosis. These markers should be explored in future studies to further validate their use in marker assisted selection, with the goal of enhancing sustainability of the breed conservation-management program.

Background

Coccidiosis is a parasitic disease caused by protozoa of the genus *Eimeria* and poses a significant threat to the health and productivity of sheep flocks worldwide. The disease damages epithelial cells that transport nutrients and fluids from the intestines into the host. Infections can result in dysentery or in more severe infections, haemorrhaging of blood and plasma into the gut lumen, or bacterial infection. Clinical signs of coccidiosis are observed mainly in 4–8-week-old lambs housed with their dams, and 2 to 3 weeks after weaning. Moreover, following a change in diet, especially after entering feedlots, sheep of all ages can develop coccidiosis [1]. Besides clinical coccidiosis, subclinical infection can have a significant negative impact on sheep production, causing reduced growth rate, lowered feed efficiency, increased susceptibility to other diseases, reduced production and, although less common, increased mortality [2]. Costs for control of coccidiosis in sheep have been estimated to be $100 million per year in the United States alone [3].

Current control of coccidiosis in sheep often relies on anticoccidial drugs, with associated costs. Moreover, emergence of drug-resistant strains has reduced treatment efficacy and capacity to control outbreaks. No new drugs in this category have been approved for decades [4]. Overall, drug-resistance problems, as well as consumer's preference for natural meat products, have driven the search for more sustainable

parasite control strategies. Thus, a series of complementary measures that have already put in place to reduce anti-helminthic resistance in ruminants, such as rotational grazing, use of natural anti-parasitic compounds, flock management practices, or dietary supplementation have been proposed and may be useful in integrated coccidiosis control programs [5]. Along with these options, genetic resistance to gastrointestinal (GI) nematodes has been of increasing interest as an alternative parasite control approach in ruminants. Breeding sheep for genetic resistance to coccidia infection is a promising tool to manage coccidiosis in sheep, thus achieving higher sustainability of the production system. Furthermore, selective breeding may bring ethical benefits by reducing the number of animals exposed to disease [6].

Genetic selection of more resistant individuals relies on the existence of genetic variation [7], and previous studies show there is evidence of genetic variation among sheep in resistance to several species of GI parasites, including *Eimeria* [8,9]. Previous studies of commercial sheep have considered selection of sheep for enhanced resistance to GI parasites [10–12]. Such selection is considered feasible when sheep face natural parasite challenges [13] and breeding for resistance should be considered for introduction into broader control programs [14]. As for other parasite resistance traits, for coccidiosis, FOC is a trait with low heritability, estimated at 0.09 ($\pm$ 0.03 SE) in Scottish Blackface sheep [8].

Previous studies have proposed that genomic selection is particularly useful for polygenic, lowly heritable traits that have a complex genetic architecture [6,15]. With the recent dramatic decrease in sequencing and genotyping costs, GS is more readily accessible than ever and has enabled reliable assessment of genetic disease risks and prediction of genomic breeding values [16]. GS can be performed using either all SNPs

or only a subset of SNPs obtained from whole-genome sequencing (WGS) or a SNP array. In theory, the accuracy of GS should be higher when using all SNPs from WGS, but in practice this increase is rarely observed [17], and small increases or even a decrease in accuracy have been reported [18,19]. SNPs that are not informative generate a burden and multiple studies have reported higher GS accuracy when using subsets of SNPs that are associated with the phenotype [20].

A GWAS allows identification of SNPs that are associated with a phenotype. Moreover, identifying candidate genes and corresponding transcription factors (TFs) associated with significant SNPs provides insights into the complex genetic architecture of traits by highlighting relevant biological processes. Previous studies have shown that TFs of genes can be associated with important traits in several livestock species, including sheep [21].

GWAS for identification of SNPs that are significantly associated with GI parasites requires availability of records at the individual sheep level, which are laborious and costly to obtain. Many studies have focused on identification of SNPs associated with faecal nematode egg counts [22–26] but the number of such SNPs identified are few. GWAS studies in chickens with induced coccidiosis identified SNPs located in genomic regions involved in tissue proliferation and repair and in primary innate immune response [27,28]. GWAS studies focused on FOC are scarce in sheep.

We here present the first GWAS for FOC in Portuguese Merino using genotypes for ~550K SNPs and phenotype data for ~400 Portuguese Merino lambs. Portuguese Merino sheep are mostly bred in Southern Portugal and represent an important economic resource in this region. The Merino group includes two breeds known as Black Merino and White Merino, which differ in coat colour and have been raised

independently for several decades. Merino sheep are pastured all year round and constantly exposed to a wide range of GI parasites. We hypothesised that a GWAS study in naturally infected Portuguese Merino lambs in this region would allow identification of SNPs significantly associated with GI parasites and with associated biological processes.

Materials and Methods

Study design

The study was developed as part of project PTDC/CVT-CVT/28798/2017, whose primary goals were to study the epidemiology of GI parasites in Merino sheep and identify SNPs significantly associated with faecal egg counts (FEC). Sample collection was conducted in farms distributed in the regions of *Ribatejo* and *Alentejo* in southern Portugal. A total of 1,022 Merino lambs (354 Merino Black and 668 Merino White) were screened for parasites in the study, distributed across 35 farms, with on average 29.2 ±9.68 animals sampled per farm. Blood samples were taken by the Merino Producer's Association (ANCORME) as part of the regular parentage testing program for pedigree confirmation and animal registration. Blood was aspirated into S-Monovette® DNA Exact tubes and a subsample transferred to FTA cards. Faecal samples were collected directly from the rectal ampulla of each animal into plastic bags, individually identified, and stored refrigerated until laboratory analysis. Samples were collected from September 2019 to December 2021. The mean lamb age at sample collection ranged from 171.9 (±13.8 SD) to 340.8 days (±34.0 SD). Depending on age and farm, several animals were dewormed between 54 and 261 days before sample collection with antihelmintics commonly used in sheep (benzimidazoles and

macrocyclic lactones). See Additional file 1, Table S1, for a more detailed summary per farm.

Phenotypic measurements

Faecal egg and oocyst counts were performed by the Mini-FLOTAC method [29,30]. Five grams of each faecal sample were homogenised in 45 ml saturated NaCl solution in a glass beaker and the homogenate passed through a tea strainer into a 50 ml conical tube. For oocyst counting, 1 ml of the filtered homogenate was filled into one of the chambers of the Mini-FLOTAC device, resulting in a sensitivity of 10 oocysts per gram of faeces. All oocysts under the grid of the chamber were counted and multiplied by 10 to obtain the number of oocysts per gram. Levels of FOC infection were classified according to Yan et al. (2021) [26].

Genotyping and quality control

Since the project prioritised finding SNPs significantly associated with GI nematodes, our initial approach was to select 408 animals with extreme values of faecal egg count (75th and 25th percentile), considering all data and per farm distributions. Coccidia burdens in these animals were assumed to reflect a natural distribution of infection in sheep and no association between FEC and FOC was identified in this subsample of animals selected for genotyping (see Additional file 2: Fig. S1). DNA was extracted using the sbeadex Livestock DNA Purification Kit. (LGC). The GeneSeek® Genomic Profiler™ Ovine 50K Beadchip (Neogen, Lansing, MI, USA) was used to genotype 206 samples and the remaining 202 animals were genotyped with the Ovine Infinium HD 600K Beadchip (Illumina Inc., San Diego, CA, USA). The distribution of lambs by breed group and genotyping panel is presented in Table 1. The farms from the genotyped animals were derived were composed of either white or black merino. Thus,

breed and farm were not independent variables. The sampling of animals per farm was quite uneven, with the number of animals genotyped per farm averaging 10 ,ranged from 1 to 25, and included 15 farms with less than 10 animals.

Quality control of data on both genotype datasets was performed using PLINK v1.9 [32]. Animals with a call rate below 90% were removed. In addition, markers located in sex chromosomes, duplicated in the chip, or with one or more multi-character allele codes were filtered out. Furthermore, SNPs were removed if the missing call rates exceeded 10%, the minor allele frequency (MAF) was less than 1%, or the Hardy-Weinberg equilibrium p-value was less than 1E-6. SNP positions were based in the Oar_v3.1 of the sheep genome assembly [33]. Prior to imputation, the conform-gt program (included in Beagle) was used to verify that the markers shared between the target and reference VCF files had identical genomic positions and allele designations. Then, the 50K array was imputed using Beagle v5.4 [34], setting the quality approved genotypes obtained, using the 600K array as reference. The imputed 50K and 600K datasets were combined and a second step for genotype quality validation and filtering was performed using the same criteria as described above. Next, SNPs in linkage disequilibrium (LD) were discarded from the dataset using PLINK, using a window of 50 SNPs, a window shift of five SNPs, and a r^2 greater than 0.5. The dataset obtained at this stage was used in the following analyses.

Genome-wide Association Analysis

Prior to the GWAS, a principal component analysis (PCA) was performed to check how the imputed dataset projected into the PCA space computed for the reference panel, and to check patterns and stratification of the data. The PCA was performed on the reference 600K array and the imputed 50K array was projected into it.

Association analysis for FOC was performed using the genome-wide complex trait analysis (GCTA v1.94.1) software tool [35]. Association analysis for FEC is ongoing and will not be presented here. The mixed linear model used for FOC was:

$$\mathbf{y} = \mu + \mathbf{X_1}\boldsymbol{\beta_1} + \mathbf{X_2}\boldsymbol{\beta_2} + \mathbf{X_3}\boldsymbol{\beta_3} + \mathbf{X_4}\boldsymbol{\beta_4} + \mathbf{g} + \mathbf{f} + \mathbf{e} \qquad (1)$$

where $\mathbf{y}$ is a vector of FOC values, $\boldsymbol{\mu}$ is the mean term, $\boldsymbol{\beta_1}$, $\boldsymbol{\beta_2}$, and $\boldsymbol{\beta_3}$ are vectors of fixed effects, which included the linear effect of age (in days), sex, and season at sample collection, and $\boldsymbol{\beta_4}$ is the fixed additive effect of the SNP to be tested. $\mathbf{X_1}$, $\mathbf{X_2}$, $\mathbf{X_3}$, and $\mathbf{X_4}$ are the corresponding design matrices linking observations in $\mathbf{y}$ to these fixed effects. $\mathbf{g}$ is a vector of the polygenic genetic effects of the individuals, $\mathbf{f}$ is the random effect of farm, and $\mathbf{e}$ is a vector of residual effects. The residuals and total genetic effects were assumed to follow a normal distribution $N(0,\mathbf{I}\sigma_e^2)$ and $N(0,\mathbf{G}\sigma_u^2)$, respectively, where $\mathbf{G}$ is interpreted as the genomic relationship matrix between individuals [35].

Significance thresholds for GWAS have been usually calculated using a *Bonferroni* multiple test-correction, but such adjustments can be conservative due to influence of LD between genetic markers [36]. Thus, a chromosome-wise significance threshold was calculated by the method named "simple M", which was proposed by Gao et al. (2008) [32] and has been shown to be an efficient alternative when more computationally intensive approaches such as permutation are not feasible [38]. Briefly, the effective number of independent markers for each chromosome (M_{eff_C}) was calculated as the number of markers needed so that the eigenvalues derived from the composite LD correlation matrix explain 99.5% of the variation. Then, the *Bonferroni* correction formula was applied to calculate chromosome-wise significance as $\alpha_C=0.05/M_{eff\_C}$. The total number of independent markers in this study was 229,235. The set of markers with a *p*-value lower than chromosome-wise significance were used for further analyses.

The genomic inflation factor (λ) estimates the amount of inflation by comparison of observed test statistics to those expected under the hypothesis of no effect [39]. Chi-square test statistics were calculated from *p*-values assuming one degree of freedom and λ was calculated as the median of observed chi-squared test statistics divided by the expected median [40]. A value of the genomic inflation factor close to unity reflects no evidence of inflation.

Post GWAS analyses

Haplotype blocks were estimated using PLINK and those with significant SNPs from the GWAS analyses were used for further analyses. Coordinates of significant SNPs were updated from the Oar_v3.1 assembly to the ARS-UI_Ramb_v2.0 assembly [41] using SNP variant information from Ensembl. Haplotype blocks were updated using the UCSC LiftOver tool [42]. Those SNPs that were unable to be re-mapped or haplotype blocks whose coordinates were only partially transferred (e.g., fragmented or spanning multiple regions in the new assembly), were excluded to avoid inconsistencies in SNP composition or inter-marker distances. The Genomic Annotation in Livestock for positional candidate Loci (GALLO v1.5) R package was used for annotation of genes and Quantitative Trait Loci (QTLs) that were in proximity to candidate markers. The Animal QTLdb database (release 53) for sheep was used for the analysis [43]. Candidate genes were considered if their boundaries were within 100 kbp upstream or downstream of significant SNPs. The LD decay distance, defined as the physical distance at which the average LD (r²) between markers drops to approximately 0.2, has been reported to be similar among different sheep breeds—approximately 100 kbp—by Marina et al. (2021) [44]. This value was therefore used to define regions of confidence in the present study. In addition, the AnimalTFDB 4.0 database was used to check whether any TFs or co-transcription factors were associated with the genes located

within the confidence regions [45]. Enrichment analyses for gene ontology (GO) terms

were performed with the gprofiler2 R package [46] forthe candidate gene list composed

of genes within the confidence regions defined above, using the False Discovery Rate

(FDR) multiple testing correction method with an FDR cut-off value of 0.05.

Remaining parameters were left as default. Furthermore, the GeneMANIA (v3.5.3)

Cytoscape plugin [47] was used to identify the most related genes to the set of genes

located in the confidence regions. For that purpose, orthologous human genes were

searched with biomaRt [48] and the resulting list of orthologous genes were passed to

GeneMANIA and run with default parameters, which uses the query gene-based method

as weights, in which interactions among the query gene list are prioritized over genes

outside the list. Genes without an orthologous sheep gene were removed from the

network.

Overrepresentation of transcription factor binding sites (TFBSs) in the regulatory

regions of genes located within confidence regions was performed using the

Transcription Factor Matrix Explorer (TFM-explorer v2.0) program [49]. Since

transcription start sites (TSSs) are uncertain for some genes in the current assembly, the

sequences from 3,000 bp upstream to 300 bp downstream of genes'TSSs were extracted

[50], based on the ARS-UI_Ramb_v2.0 assembly. These sequences were used as input

for TFM-explorer, which was executed using all vertebrate weight matrices from the

JASPAR database [51]. An enrichment analysis for GO terms was performed with

gprofiler2 [46].

Results

Genotype quality control and imputation

After quality control of the genotypes of the 50K chip samples, 11 samples were removed due to a low call rate. From the initial set of 48,841 autosomal and bi-allelic SNPs, missing call rates greater than 10% were found for 1050 SNPs, a MAF lower than 1% was detected for 2,486 SNPs, and a Hardy-Weinberg equilibrium p-value less than 1E-6 was found for 340 SNPs. Thus, genotypes for 44,965 SNPs for 195 sheep remained for further analyses.

Following the same quality control criteria for genotypes from the 600K chip samples, one sample was removed due to low call rate. From the initial set of 577,205 autosomal and bi-allelic SNPs, 12,162 SNPs were removed due to missing call rate, 25,033 SNPs were filtered out due to MAF thresholds, and 2,291 SNPs were removed due to Hardy-Weinberg equilibrium. Finally, genotypes for 537,719 SNPs for 201 sheep remained for further analyses.

The 50K chip dataset was imputed using the 600K chip as reference. From the 44,965 SNPs in the 50K chip data, only 34,992 SNPs had corresponding markers in the 600K chip. After imputation was performed and both VCF files were combined, quality control was performed with the previous filtering criteria, resulting in 2,709 and 3,839 SNPs filtered-out due to Hardy-Weinberg equilibrium and MAF, respectively. The resulting dataset used for GWAS was composed of genotyped for 531,171 SNPs for 396 sheep.

Phenotype statistics and population structure

FOC is a measure of the intensity of coccidia infection. Coccidia prevalence was high, with 88.1% of the genotyped samples with a value greater than 0. The average FOC value was 923.7 (±1,768.2 SD), the highest value being 13,570 oocysts/gram of faeces. We had 47 sheep with no infection (FOC = 0 oocysts/gram); 307 with a mild infection (FOC < 1800 oocysts/gram) and an average FOC value of 517.1 (±434.5 SD); 29 with a moderate infection (FOC = 1800-6000 oocysts/gram) and an average value of 3,155.9 (±1,420.7 SD); and 13 with a severe infection (FOC > 6000 oocysts/gram) and an average value of 8,886.2 (±2,256.3 SD). Most samples showed a mild infection.

Prior to the GWAS analysis, a PCA was performed to investigate population stratification (see Additional file 3: Fig. S2). Each principal component explained a small portion of the variance in the reference array, with PC1 explaining 2.5% and PC2 explaining 1.8%. Genotype imputation had no structural effect on the data. Clustering by breed-group was observed.

SNPs associated with FOC based on GWAS and gene annotation

The selected GWAS model displayed a genomic inflation factor (λ) of 0.924 and included age, sex, and season as fixed effects and farm as random effect (see Additional file 4: Fig. S3). Other models were analysed such as removing age, sex, or season as fixed effects, which generated genomic inflation factors slightly lower than the selected model (see Additional file 5: Fig. S4). Since breed is correlated with farm, we evaluated whether including breed as a fixed effect in addition to age, sex, season (fixed), and farm (random) would improve the model fit (see Additional file 5: Fig. S4e). The *p*-values from the selected model and the other investigated models were highly correlated (see Additional file 6: Fig. S5) and inclusion of breed as a fixed effect had negligible

impact on SNP associations (Pearson correlation of $-\log_{10}(p)$ values between models = 0.995; genomic inflation factor λ = 0.9245 vs. 0.9237; see Additional file 6: Fig. S5e). Given the minimal differences, we retained the simpler model (age, sex, season fixed; farm random) for the final analyses.

GWAS allowed identification of 14 significant SNPs at a chromosome-wide significance level. Since two of those SNPs could not be re-mapped to the ARS-UI_Ramb_v2.0 assembly, we ended up with 12 significant markers (Table 2, Fig. 1). The chromosome-wide significant thresholds employed are shown in Additional file 7, Table S2. Box plots showing the effect estimates of significant SNPs are in Additional file 8, Fig. S6. Most significant SNPs were located within introns of known genes (namely: *Ccser1, Thsd4, Eci1, Tnfrsf12a, Chrm3 and Slc20a2*), one had a missense predicted consequence on a transcript from *Znf200,* and a few SNPs located downstream of non-coding genes. Two significant SNPs were located within an intergenic region.

We identified additional genes within the confidence region of each SNP (see methods). This strategy enabled identification of TFs for the gene list of *Jund, E4f1*, *Znf213*, *Znf200,* and *Znf263,* and co-transcription factors for *Pik3r2* and *Tle3*. Together with genes that overlapped with the significant SNPs, the list of candidate genes included 80 genes (see Additional file 9: Table S3).

Gene set enrichment analyses and Gene-TF networks

Our analyses only identified one previously defined QTL related to ear size [52], which overlapped the confidence region around the rs161410499 marker. The obtained list of 80 candidate genes was further used to investigate enrichment of molecular functions and gene-TF networks. Two types of enrichment analysis were performed. The first

analysis involved testing for enrichment in a gene list and allowed identification of two molecular functions, virus receptor activity (GO:0001618, FDR=0.0026) and exogenous protein binding (GO:0140272, FDR=0.0026). Both terms were enriched due to the presence of two claudin family members in the gene list, *CLDN6* and *CLDN9*, as well as a solute carrier family member, *SLC20A2*. The second analysis considers interactions between genes within the query list and other possible gene interactions available in the databases s (Fig. 2a). This analysis identified the presynapse (GO:0098793, FDR=0.045) cellular component GO term and vesicle related GO terms such as vesicle docking (GO:0048278, FDR=0.045) in the biological process category, and exocytic vesicle (GO:0070382, FDR=0.044) in the cellular component category (Fig. 2b). Figure 2a shows that 45 of the identified candidate genes were found to interact (co-expression, genetic relationship or sharing protein domains) and formed a single network cluster. The network contains six of seven genes associated with vesicle related functions, as shown in Fig. 2a. These genes are from the same gene family, *RAB*, which belong to the RAS superfamily of small GTPases [53]. The genes that displayed higher levels of interactions were *RAB3A* and *RAB26*. The network shown in Fig. 2a also shows interactions between genes and TFs, which included several zinc finger proteins, including ZNF200 (with a tolerated missense SNP) and co-transcription factor *PIK3L*.

Additionally, overrepresentation of TFBSs in regulatory regions of genes located within the identified confidence regions was assessed. This analysis showed overrepresentation of binding sites for TFs of *TFAP2A*, *NFKB1*, *INSM1*, *NK-κβ*, *RelA*, *EBF1*, *GABPA*, *PLAG1,* and *ELK4*, among others (see Additional file 10: Table S4). The enrichment analysis showed terms associated to DNA-binding transcription factor activity (GO:0003700, FDR=1.96281E-09) and virus receptor activity (GO:0001618, FDR=0.000873382) (Fig. 3a). The predicted interactions of *RAB3A*, *RAB26, CLDN6,*

and *CLDN9* with TFs were represented as a network, showing interactions with

NFKB1, NF-kappaB, and RELA (Fig. 3b).

Discussion

Portuguese Merino is a sheep breed of high economic value and these animals are

typically kept in extensive system. Debilitation of flocks from coccidiosis is prevalent

among these sheep and resistance to anticoccidial drugs is a major problem. Alternative

anticoccidial approaches are needed. In this regard, our study utilized GWAS to identify

genetic markers associated with the biological response of lambs to coccidial infection.

Given the sample size, results from the two merino types were combined to increase the

sample size and improved capability for GWAS detection of significant SNPs. When

black and white merino samples were analysed separately, much fewer significant SNPs

were identified, most likely due to the low sample size of each merino type.

In total, GWAS identified 12 SNPs that were significantly associated with FOC in

lambs naturally exposed to parasites. Seven of these SNPs are located within the introns

of genes, two are intergenic, two are located downstream of non-coding genes, and one

has a tolerated missense predicted consequence. By considering the region of

confidence of each SNP, our analyses retrieved additional candidate genes, reaching a

final list of 80 candidate genes. The enrichment analysis identified enrichment of genes

from the claudin family, which are integral components of tight junctions and regulate

paracellular permeability in epithelial tissues [54]. Forty-five of the 80 candidate genes

have different types of interactions reported in the literature, namely, co-expression,

sharing of protein domains, and physical interactions. The enrichment analysis that

integrates gene interactions identified enrichment of genes encoding Rab proteins,

which are members of the RAS superfamily of small GTPases. Rab proteins recruit

specific sets of effector proteins onto membranes in their GTP-bound state and coordinate consecutive stages of transport, such as vesicle formation, vesicle and organelle motility, and tethering of vesicles with target membranes [53,55].The significance of the role that these members of the claudin and the RAS superfamily of small GTPases might play in sheep susceptibility to coccidiosis is outlined below.

Coccidiosis in sheep is an intestinal infection caused by different species of intracellular protozoans of the genus *Eimeria*. Non-sporulated oocysts are excreted onto grazing areas in the faeces of infected sheep. Under optimal conditions of moisture, oxygen, and temperature (24° to 32°C), these non-infective stages require a minimum of two to five days to sporulate and become infective. Oocysts are usually unable to survive temperatures below -30°C or above 40°C, but this wide range of temperatures allows for sporulated and non-sporulated oocysts to maintain their viability for more than a year, under proper environmental conditions, posing a constant source of infection to grazing livestock. Sporulated oocysts are ingested by sheep and eight sporozoites are released from each oocyst during passage through the intestine due to the action of bile and trypsin. After invading enterocytes, the parasite undergoes a complex cycle of asexual (merogony) and sexual (gametogony) multiplication, resulting in the production of more oocysts, which are shed into the environment with the faeces. The entire life cycle within the host takes 2 to 3 weeks [56], during which the parasite needs to penetrate a diverse range of host tissue barriers, acquire nutrients, and evade host response. The intestinal barrier is the primary defence against harmful substances and pathogens, whose integrity and impermeability are regulated by tight junctions. Claudins that are in the list of candidate genes, such as *CLDN6* and *CLDN9,* are transmembrane proteins that are involved in maintaining the integrity of epithelial cell layers in the intestinal lining [57]. Claudins are integral components of tight junctions between epithelial cells,

contributing to the selective permeability of the intestinal barrier. We hypothesize that changes of expression of *CLDN6* or of *CLDN9* affects the permeability of the intestinal epithelium, allowing for pathogens to promote their translocation and invade the host [58]. In fact, infection with the protozoan parasite *Cryptosporidium parvum* results in a substantial downregulation of *CLDN4* in mouse [59]. Chickens infected with *Eimeria tenella* show upregulated expression of *CLDN2* and decreased expression of *CLDN3* and E-cadherin genes in the intestine [60]. A study on dairy calves infected with *Eimeria bovis* observed changes in the expression of CLDN4 in cecum [61]. Coccidiosis is also known to trigger inflammatory response in the intestine. Other studies have shown that increased inflammation can affect the expression of claudins, including *CLDN6* and CLDN9 [62].

The gene interaction network identified a hub composed only of RAB proteins, two of which, *RAB3A* and *RAB26,* are in the list of GWAS candidate genes and display the highest level of interactions in the network. Rab proteins are a family of small GTPases that play crucial roles in regulating vesicular trafficking and membrane dynamics within eukaryotic cells [63]. Previous studies have shown that apicomplexan parasites manipulate Rab GTPases to modulate cellular trafficking pathways, in order to move host vesicles near to their parasitophorous vacuole membrane, scavenge nutrients from host vesicles, and control host phagocytic activity [64]. For example, *Toxoplasma gondii* is auxotrophic for many metabolites and the parasite scavenges sphingolipids from the host Golgi by sequestering Golgi-derived RAB GTPases such as *Rab14*, *Rab30,* and *Rab43* into the parasitophorous vacuole [65]. *Plasmodium berghei* hijacks the Golgi structure of the hepatocyte and interaction with *Rab11a* is critical for the ability of the parasite to induce Golgi fragmentation, in order to increase nutrient acquisition surfaces [66]. In addition to their vesicle related functions, Rab GTPases

modulate innate immunity by regulating transmembrane signals' transduction [67]. For example, *Rab14* overexpression protects *Plasmodium spp.* sporozoites from degradation in macrophages, while silencing *Rab14* results in increased phagocytosis [68]. Our results showed that SNPs identified by the GWAS are located within genes involved in genetic interactions, in which *RAB3A* and *RAB26* play a crucial role, given their large number of interactions with other genes of the network, which include TFs such as *PIK3R2*, which functions as part of the PI3K pathway, regulating cell growth, proliferation, and survival. *RAB3A* and *PIK3R2* were found to be co-expressed in human cancers [69]. The involvement of *Rab* proteins, specifically *RAB3A* and *RAB26*, in the context of coccidiosis, has not been widely studied. Given the known roles of these proteins, both RAB3A and RAB26 proteins likely play essential roles in the intracellular trafficking of both the *Eimeria* parasite and host immune cells. Modulation of *RAB3A* expression, with its role in vesicle fusion and neurotransmitter release, is crucial for the parasite, since it allows parasite invasion and immune cell modulation, while modulation of *RAB26* expression by the parasite, with its broader vesicular trafficking role, may assist in parasite replication and immune response during *Eimeria* infection.

Additionally, identification of the gene-transcription factor network in our study showed enrichment for DNA-transcription factor activator activity and RNA polymerase II-specific molecular functions. In fact, within this network, we identified interactions of *NF-κB* with *RAB3A, CLDN6,* and *CLDN9*. *NF-κB* plays a critical role in mediating responses to a notable diversity of external *stimuli*. It is a significant orchestrator of immune response, including specific changes of NF-κB levels resulting from parasite infections [70]. Previous studies have shown enhanced NF-κB activity in mice infected with *Trichuris muris*, followed by increased levels of several cytokines [71].

Modulation of macrophage regulation by NF-κB was also observed during *Trichinella spiralis* and *Brugia malayi* infections [72,73]. Moreover, the *RelA* gene was also identified as a TF that interacts with *RAB3A and RAB26,* which encode one of five proteins that may dimerize to form NF-κB complexes [74].

Conclusions

This study has identified GWAS SNPs that maybe in LD with genes that have the potential to establish a mechanism for resistance to coccidia infection in sheep. Our hypothesis (Fig. 4) is that *NF-κB* and *RELA* activate transcription of *CLDN6* and *CLDN9*, which are required to tightening junctions to block parasite entry. Then, *CLDN6/9* can activate *RAB3A*, enhancing junctional vesicle recycling or stabilization, boosting barrier function. *RAB3A* can then can signal to *RAB26*, which activates immune-modulatory responses (possibly recruiting or activating local immune cells). Other TFs in the network (e.g., *PPARG, TFAP2A, ZNF143*) may modulate expression for homeostasis or fine-tune the response depending on infection stage. To date, there have been no programs for breeding of resistance to GI parasites in this breed. Given, the number of SNPs, their statistical significance and functional impact, we hypothesize that Portuguese Merino sheep have undergone natural selection, that increased allele frequencies of genotypes associated with tolerance to FOC. This hypothesis is further supported by the fact that a previous study using a larger GWAS sampling size of Blackface sheep, that did not find any significant association of SNPs with FOC [26].

In conclusion the identified markers may be useful as a basis to further exploit genetic resistance to GI parasites, and not only in Portuguese Merino sheep. We have identified a mechanism that should be further investigated. Validation of the predicted interactions may also enable the identification of novel therapeutic targets. In future studies, it will

be important to clarify the role of these SNPs in this physiological mechanism responsible for resistance or susceptibility to coccidiosis in sheep.

Declarations

Ethics approval and consent to participate

This study was approved by the Ethics Committee of INIAV, ORBEA (Orgão Responsável pelo Bem-Estar dos Animais) on 18/07/2018, Portugal.

Consent for publication

Not applicable.

Availability of data and material

The dataset has been deposited in the European Variation Archive (EVA) under the study accession number PRJEB79893.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

HW and LTG developed the grant proposal. HW coordinated the project's implementation and established the phenotype database. TP coordinated sample collection. AF, DM, PV, IS, LMC, ACF, CM, NC and RCS collected data. AJA prepared samples for genotyping. EV-M, AJA, FT and TN conceived, designed and performed data analysis. EV-M, AJA co-wrote the manuscript. All authors reviewed and provided suggestions to the manuscript.

Acknowledgements

References

1. Foreyt WJ. Coccidiosis and cryptosporidiosis in sheep and goats. Veterinary clinics of north America: Food Anim Pract. 1990;6:655–70.

2. Campbell S. Coccidiosis in ruminants: the inevitable infection. Vet Irel J. 2021;11:48–52.

3. Kaltungo BY, Musa IW. A review of some protozoan parasites causing infertility in farm animals. ISRN Trop Med. 2013;2013:782609.

4. Noack S, Chapman HD, Selzer PM. Anticoccidial drugs of the livestock industry. Parasitol Res. Parasitol Res; 2019;118:2009–26.

5. Rapiya M, Hawkins H, Muchenje V, Mupangwa J, Marufu M, Dzama K, et al. Rotational grazing approaches reduces external and internal parasite loads in cattle. Afr J Range Forage Sci. 2019;36:151–9.

6. Rupp R, Mucha S, Larroque H, McEwan J, Conington J. Genomic application in sheep and goat breeding. Anim Front. 2016;6:39–44.

7. Falconer DS. The inheritance of liability to certain diseases, estimated from the incidence among relatives. Ann Hum Genet. 1965;29:51–76.

8. Pacheco A, McNeilly TN, Banos G, Conington J. Genetic parameters of animal traits associated with coccidian and nematode parasite load and growth in Scottish Blackface sheep. Animal. 2021;15:100185.

9. Sechi S, Salaris S, Scala A, Rupp R, Moreno C, Bishop SC, et al. Estimation of (co)variance components of nematode parasites resistance and somatic cell count in dairy sheep. Ital J Anim Sci. 2009;8:156–8.

10. Windon RG. Genetic control of resistance to helminths in sheep. Vet Immunol Immunopathol. 1996;54:245–54.

11. McRae KM, McEwan JC, Dodds KG, Gemmell NJ. Signatures of selection in sheep bred for resistance or susceptibility to gastrointestinal nematodes. BMC Genomics. 2014;15:1–13.

12. Ferreira GF, Ciappesoni G, Castells D, Amarilho-Silveira F, Navajas EA, Giorello D, et al. Feed conversion efficiency in sheep genetically selected for resistance to gastrointestinal nematodes. Anim Prod Sci. 2021;61:754–60.

13. Bishop SC, Jackson F, Coop RL, Stear MJ. Genetic parameters for resistance to nematode infections in Texel lambs and their utility in breeding programmes. Anim Sci. 2004;78:185–94.

14. Nieuwoudt SW, Theron HE, Krüger LP. Genetic parameters for resistance to Haemonchus contortus in Merino sheep in South Africa. J S Afr Vet Assoc. 2002;73:4–7.

15. Calus MPL, Veerkamp RF. Accuracy of breeding values when using and ignoring the polygenic effect in genomic breeding value estimation with a marker density of one SNP per cM. J Anim Breed Genet. 2007;124:362–8.

16. Raben TG, Lello L, Widen E, Hsu SDH. From genotype to phenotype: Polygenic prediction of complex human traits. Methods Mol Biol. 2022;2467:421–46.

17. Druet T, Macleod IM, Hayes BJ. Toward genomic prediction from whole-genome sequence data: impact of sequencing design on genotype imputation and accuracy of predictions. Heredity (Edinb). 2014;112:39–47.

18. Ober U, Ayroles JF, Stone EA, Richards S, Zhu D, Gibbs RA, et al. Using whole-genome sequence data to predict quantitative trait phenotypes in Drosophila melanogaster. PLoS Genet. 2012;8:e1002685.

19. Veerkamp RF, Bouwman AC, Schrooten C, Calus MPL. Genomic prediction using preselected DNA variants from a GWAS with whole-genome sequence data in Holstein-Friesian cattle. Genet Sel Evol. 2016;48:95.

20. Ni G, Cavero D, Fangmann A, Erbe M, Simianer H. Whole-genome sequence-based genomic prediction in laying chickens with different genomic relationship matrices to account for genetic architecture. Genet Sel Evol. 2017;49:8.

21. Gebreselassie G, Berihulay H, Jiang L, Ma Y. Review on Genomic Regions and Candidate Genes Associated with Economically Important Production and Reproduction Traits in Sheep (Ovies aries). Animals (Basel). 2020;10:33.

22. Thorne JW, Redden R, Bowdridge SA, Becker GM, Stegemiller MR, Murdoch BM. Genome-wide analysis of sheep artificially or naturally infected with gastrointestinal nematodes. Genes (Basel). 2023;14:1342.

23. Carracelas B, Navajas EA, Vera B, Ciappesoni G. Genome-wide association study of parasite resistance to gastrointestinal nematodes in Corriedale sheep. Genes (Basel). 2022;13:1548.

24. Benavides MV, Sonstegard TS, Kemp S, Mugambi JM, Gibson JP, Baker RL, et al. Identification of novel loci associated with gastrointestinal parasite resistance in a Red Maasai x Dorper Backcross population. PLoS One. 2015;10:e0122797.

25. Niciura SCM, Benavides MV, Okino CH, Ibelli AMG, Minho AP, Esteves SN, et al. Genome-wide association study for Haemonchus contortus resistance in Morada Nova sheep. Pathogens. 2022;11:939.

26. Pacheco A, Banos G, Lambe N, McLaren A, McNeilly TN, Conington J. Genome-wide association studies of parasite resistance, productivity and immunology traits in Scottish Blackface sheep. Animal. 2024;18:101069.

27. Boulton K, Nolan MJ, Wu Z, Psifidi A, Riggio V, Harman K, et al. Phenotypic and genetic variation in the response of chickens to Eimeria tenella induced coccidiosis. Genet Sel Evol. 2018;50:63.

28. Hamzić E, Buitenhuis B, Hérault F, Hawken R, Abrahamsen MS, Servin B, et al. Genome-wide association study and biological pathway analysis of the Eimeria maxima response in broilers. Genet Sel Evol. 2015;47:91.

29. Cringoli G, Maurelli MP, Levecke B, Bosco A, Vercruysse J, Utzinger J, et al. The Mini-FLOTAC technique for the diagnosis of helminth and protozoan infections in humans and animals. Nat Protoc. 2017;12:1723–32.

30. Maurelli MP, Dourado Martins OM, Morgan ER, Charlier J, Cringoli G, Mateus TL, et al. A qualitative market analysis applied to mini-FLOTAC and fill-FLOTAC for diagnosis of helminth infections in ruminants. Front Vet Sci. 2020;7:580649.

31. Yan X, Liu M, He S, Tong T, Liu iyong, Ding K, et al. An epidemiological study of gastrointestinal nematode and Eimeria coccidia infections in different populations of Kazakh sheep. PLoS One. 2021;16:e0272760.

32. Chang CC, Chow CC, Tellier LCAM, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience. 2015;4:7.

33. Jiang Y, Xie M, Chen W, Talbot R, Maddox JF, Faraut T, et al. The sheep genome illuminates biology of the rumen and lipid metabolism. Science. 2014;344:1168–73.

34. Browning BL, Zhou Y, Browning SR. A one-penny imputed genome from next generation reference panels. Am J Hum Genet. 2018;103:338–48.

35. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: A Tool for Genome-wide Complex Trait Analysis. Am J Hum Genet. 2011;88:76.

36. Johnson RC, Nelson GW, Troyer JL, Lautenberger JA, Kessing BD, Winkler CA, et al. Accounting for multiple comparisons in a genome-wide association study (GWAS). BMC Genomics. 2010;11:724.

37. Gao X, Starmer J, Martin ER. A multiple testing correction method for genetic association studies using correlated single nucleotide polymorphisms. Genet Epidemiol. 2008;32:361–9.

38. Hendricks AE, Dupuis J, Logue MW, Myers RH, Lunetta KL. Correction for multiple testing in a gene region. Eur J Hum Genet. 2014;22:414–8.

39. van Iterson M, van Zwet EW, Heijmans BT, 't Hoen PAC, van Meurs J, Jansen R, et al. Controlling bias and inflation in epigenome- and transcriptome-wide association studies using the empirical null distribution. Genome Biol. 2017;18:19.

40. van den Berg S, Vandenplas J, van Eeuwijk FA, Lopes MS, Veerkamp RF. Significance testing and genomic inflation factor using high-density genotypes or whole-genome sequence data. J Anim Breed Genet. 2019;136:418–29.

41. Davenport KM, Bickhart DM, Worley K, Murali SC, Salavati M, Clark EL, et al. An improved ovine reference genome assembly to facilitate in-depth functional annotation of the sheep genome. Gigascience. 2022;11:giab096

42. Hinrichs AS, Karolchik D, Baertsch R, Barber GP, Bejerano G, Clawson H, et al. The UCSC Genome Browser Database: update 2006. Nucleic Acids Res. 2006;34:D590–8.

43. Hu ZL, Park CA, Reecy JM. Bringing the Animal QTLdb and CorrDB into the future: meeting new challenges and providing updated services. Nucleic Acids Res. 2022;50:D956–61.

44. Marina H, Pelayo R, Suárez-Vega A, Gutiérrez-Gil B, Esteban-Blanco C, Arranz JJ. Genome-wide association studies (GWAS) and post-GWAS analyses for technological traits in Assaf and Churra dairy breeds. J Dairy Sci. 2021;104:11850–66.

45. Shen WK, Chen SY, Gan ZQ, Zhang YZ, Yue T, Chen MM, et al. AnimalTFDB 4.0: a comprehensive animal transcription factor database updated with variation and expression annotations. Nucleic Acids Res. 2023;51:D39–45.

46. Peterson H, Kolberg L, Raudvere U, Kuzmin I, Vilo J. gprofiler2 -- an R package for gene list functional enrichment analysis and namespace conversion toolset g:Profiler. F1000Res. 2020;9:709.

47. Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P, et al. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. Nucleic Acids Res. 2010;38:W214–20.

48. Durinck S, Spellman PT, Birney E, Huber W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nat Protoc. 2009;4:1184–91.

49. Tonon L, Touzet H, Varré JS. TFM-Explorer: mining cis-regulatory regions in genomes. Nucleic Acids Res. 2010;38:W286–92.

50. Verardo LL, Silva FF, Lopes MS, Madsen O, Bastiaansen JWM, Knol EF, et al. Revealing new candidate genes for reproductive traits in pigs: Combining Bayesian GWAS and functional pathways. Genet Sel Evol. 2016;48:9.

51. Castro-Mondragon JA, Riudavets-Puig R, Rauluseviciute I, Berhanu Lemma R, Turchi L, Blanc-Mathieu R, et al. JASPAR 2022: the 9th release of the open-access database of transcription factor binding profiles. Nucleic Acids Res. 2022;50:D165–73.

52. Gao L, Xu SS, Yang JQ, Shen M, Li MH. Genome-wide association study reveals novel genes for the ear size in sheep (Ovis aries). Anim Genet. 2018;49:345–8.

53. Stenmark H, Olkkonen VM. The Rab GTPase family. Genome Biol. 2001;2:REVIEWS3007.

54. Dhawan P, Singh AB, Sharma A. Claudin Family of Proteins and Cancer: An Overview. J Oncol. 2010;2010:541957.

55. Zerial M, McBride H. Rab proteins as membrane organizers. Nat Rev Mol Cell Biol. 2001;2:107–17.

56. Pout DD. Coccidiosis of Lambs. I. Observations on the naturally acquired infection. Br Vet J. 1973;129:555–67.

57. Zhang Y, Zhu X, Yu X, Novák P, Gui Q, Yin K. Enhancing intestinal barrier efficiency: A novel metabolic diseases therapy. Front Nutr. 2023;10:1120168.

58. Paradis T, Bègue H, Basmaciyan L, Dalle F, Bon F. Tight junctions as a key for pathogens invasion in intestinal epithelial cells. Int J Mol Sci. 2021;22:2506.

59. Kumar A, Chatterjee I, Anbazhagan AN, Jayawardena D, Priyamvada S, Alrefai WA, et al. Cryptosporidium parvum disrupts intestinal epithelial barrier function via altering expression of key tight junction and adherens junction proteins. Cell Microbiol. 2018;20:e12830.

60. Pham HHS, Matsubayashi M, Tsuji N, Hatabu T. Relationship between Eimeria tenella associated-early clinical signs and molecular changes in the intestinal barrier function. Vet Immunol Immunopathol. 2021;240:110321.

61. Walker MP, Evock-Clover CM, Elsasser TH, Connor EE. Short communication: Glucagon-like peptide-2 and coccidiosis alter tight junction gene expression in the gastrointestinal tract of dairy calves. J Dairy Sci. 2015;98:3432–7.

62. Liu T, Li J, Yu L, Sun HX, Li J, Dong G, et al. Cross-species single-cell transcriptomic analysis reveals pre-gastrulation developmental differences among pigs, monkeys, and humans. Cell Discov. 2021;7:8.

63. Homma Y, Hiragi S, Fukuda M. Rab family of small GTPases: an updated view on their regulation and functions. FEBS 2021;288:36–55.

64. Paone S, Olivieri A. Role of host small GTPases in apicomplexan parasite infection. Microorganisms. 2022;10:1370.

65. Romano JD, Nolan SJ, Porter C, Ehrenman K, Hartman EJ, Hsia R ching, et al. The parasite Toxoplasma sequesters diverse Rab host vesicles within an intravacuolar network. J Cell Biol. 2017;216:4235–54.

66. de Niz M, Caldelari R, Kaiser G, Zuber B, Heo W Do, Heussler VT, et al. Hijacking of the host cell Golgi by Plasmodium berghei liver stage parasites. J Cell Sci. 2021;134:jcs252213.

67. Su C, Zheng C. When Rab GTPases meet innate immune signaling pathways. Cytokine Growth Factor Rev. 2021;59:95–100.

68. Coppens I, Romano JD. Sitting in the driver's seat: Manipulation of mammalian cell Rab GTPase functions by apicomplexan parasites. Biol Cell. 2020;112:187–95.

69. Bild AH, Yao G, Chang JT, Wang Q, Potti A, Chasse D, et al. Oncogenic pathway signatures in human cancers as a guide to targeted therapies. Nature. 2005;439:353–7.

70. Manning AM, Anderson DC. Transcription Factor NF-κB: An Emerging Regulator of Inflammation. In: Bristol JA, editor. Annual Reports in Medicinal Chemistry. Academic Press;1994. p. 235–44.

71. Cao S, Zhang X, Edwards JP, Mosser DM. NF-κB1 (p50) Homodimers differentially regulate pro- and anti-inflammatory cytokines in macrophages. J Biol Chem. 2006;281:26041–50.

72. Guan F, Jiang W, Bai Y, Hou X, Jiang C, Zhang C, et al. Purinergic P2X7 receptor mediates the elimination of trichinella spiralis by activating NF-κB/NLRP3/IL-1b pathway in macrophages. Infect Immun. 2021;89:e00683-20.

73. Sharma A, Sharma P, Ganga L, Satoeya N, Mishra S, Vishwakarma AL, et al. Infective larvae of Brugia malayi induce polarization of host macrophages that helps in immune evasion. Front Immunol. 2018;9:305756.

74. Lu T, Stark GR. NF-κB: Regulation by methylation. Cancer Res. 2015;75:3692–5.

Tables

Table 1 Distribution of animals genotyped using the 50K and 600K SNP Chips

	Merino White Males	Merino Black Males	Merino White Females	Merino Black Females
50K Chip	11	3	136	56
600k Chip	19	7	118	58

Table 2 Significant SNPs for Coccidia FOC.

SNP rsID	Chr	bp	A1	A2	Freq	p-value	Gene	Consequence type
rs411343808	2	63,607,030	G	A	0,0202	2,00E-06	ENSOARG00020032169	downstream
rs417924701	5	4,799,630	A	C	0,0354	8,44E-07		intergenic
rs422982057	6	34,899,299	G	A	0,0101	3,63E-07	CCSER1	intron
rs411549257	7	17,031,457	G	A	0.0467	2,12E-07	ENSOARG00020027093	downstream
rs417935219	7	18,191,375	A	G	0.0341	8,99E-09	THSD4	intron
rs414033502	21	8,293,837	A	C	0.0265	9,63E-06	ENSOARG00020036475	intron
rs415434348	24	2,294,317	G	A	0.0442	2,72E-06	ECI1	intron
rs425217825	24	2,937,586	A	G	0.0139	3,08E-06	ENSOARG00020026408[*1]	synonymous/intron
rs161410499	24	3,214,651	G	A	0.0366	3,26E-06	ZNF200	missense
rs413352396	25	5,145,396	A	G	0.0758	7,78E-06		intergenic
rs418654751	25	11,926,932	A	G	0.0177	5,92E-06	CHRM3	intron
rs418055519	26	36,653,829	A	G	0,0290	6,61E-06	SLC20A2	intron

Freq=frequency of allele A1; Gene=gene with a predicted consequence of the SNP in Ensembl; Consequence type=predicted consequence in Ensembl.

[*1]Orthologue to cow TNFRSF12A

Figures

Figure 1. **Manhattan plot showing genome-wide association results for fecal oocyst count (FOC) in sheep naturally infected with *Eimeria* (coccidia).**

Blue line represents the chromosome-wide significance thresholds.

Figure 2. Gene–gene interaction network and enrichment of cellular components and biological processes.

(a) Gene–gene interaction network. For visualization purposes, nodes without connections to other nodes were removed. Edge colour represents type of relationship between nodes: purple, co-expression; green, genetic interactions; orange, predicted; grey, shared protein domain; and red, physical interaction. Node size represents the degree (number of connections) of the node. Red diamond shaped nodes are genes classified as transcription and co-transcription factors in AnimalTFDB 4.0 database. Yellow nodes are genes from the transport vesicle GO term. Black nodes are genes located in confidence regions of significant SNPs, while grey nodes are genes that are predicted to be functionally related to the query genes. (b) Enrichment analysis based on the gene-gene interaction network. The bubble plot shows the enriched terms in the Y-axis and the gene ratio (gene ratio = number of genes from network/all genes included in the term). Bubble size and colour represents number of genes and significance (FDR) in the GO term, respectively.

Figure 3. Enrichment of transcription factor binding sites (TFBS) and predicted TF–gene interaction network for candidate genes within confidence regions.

(a) Enrichment results with genes located in the confidence regions and transcription factors with overrepresentation of TFBSs in their regulatory regions. (b) Network representing the enriched genes in the serine-type endopeptidase inhibitor activity GO term and transcription factors with predicted TFBSs in those gene

Figure 4. Proposed molecular mechanism underlying resistance to *Eimeria* (coccidia) infection in sheep.

Schematic representation of the hypothesized host cellular response to *Eimeria* infection. Activation of the *NF-κB* signaling pathway triggers the expression of immune and inflammatory mediators, including genes such as *RAB3A, CLDN6,* and *CLDN9,* which may strengthen epithelial barrier integrity and coordinate vesicle trafficking involved in immune signalling. Tight junction proteins (Claudins, Occludin, JAM) contribute to maintaining epithelial cohesion and preventing parasite invasion, thereby enhancing resistance to coccidial infection. Created in BioRender. Varela Martínez, E. (2025) https://BioRender.com/1uygi0q

Additional files

Additional file 1 Table S1

Format: .xlsx

Title:

Description: Information regarding the animals used in the study summarized by farm.

Additional file 2 Figure S1

Format: .jpg

Title:

Description: Scatter plot showing faecal egg counts (FEC) in the x-axis and faecal oocyst counts (FOC) in the y-axis for the genotyped samples.

Additional file 3 Figure S2

Format: .jpg

Title:

Description: Two first principal components of PCA analysis of the reference genotype data (600K chip). The genetic data of the imputed 50K chip was projected into it. The shape represents genotyping using either 50K of 600K SNP chip, while the color represents different variables. (a) Colored by farm; (b) Colored by breed.

Additional file 4 Figure S3

Format: .jpg

Title:

Description: Quantile-quantile plot of the model used in this study for coccidia FOC.

Additional file 5 Figure S4

Format: .jpg

Title:

Description: Quantile-quantile plot of the model with the following variables for coccidia FOC: (a) age, sex and farm (random); (b) sex, season and farm (random); (c) age, season and farm (random); (d) Age, sex; (e) breed, age, sex, season and farm (random); (f) age, sex, season, three first principal components and farm (random).

Additional file 6 Figure S5

Format: .jpg

Title:

Description: Comparison of obtained p-values from two models in GCTA. The x-axis represents –log10(p-value) of the model with age, sex and season as fixed effect and farm as random effect, while the y-axis represents the values of the model with: (a) age, sex and farm (random); (b) sex, season and farm (random); (c) age, season and farm (random); (d) Age, sex; (e) breed, age, sex, season and farm (random); (f) age, sex, season, three first principal components and farm (random).

Additional file 7 Table S2

Format: .xlsx

Title:

Description: Chromosome-wise significant thresholds calculated for the GWAS analysis for FOC. The thresholds are calculated as -log10(Meff_C), where Meff_C is the effective number of independent markers for each chromosome.

Additional file 8 Figure S6

Format: .pdf

Title:

Description: Boxplots showing genotype distribution for the significant SNPs identified in this study. The marker ID can be seen at the bottom. The genotype in the x-axis and FOC values in the y-axis. n=number of Merino sheep.

Additional file 9 Table S3

Format: .xlsx

Title:

Description: Genes located within confidence regions (100 Kb) of the significant markers identified in the GWAS analysis.

Additional file 10 Table S4

Format: .xlsx

Title:

Description: Results from overrepresentation of transcription factor binding sites (TFBSs) in the regulatory regions of genes located in the confidence regions by TFM-explorer.