Received: 26 July 2022 | Accepted: 19 April 2023

DOI: 10.1111/age.13328

RESEARCH ARTICLE

ANIMAL GENETICS WILEY

Univariate and multivariate genome-wide association studies for hematological traits in Murciano-Granadina goats

Martina Macri[1,2] | **Maria Gracia Luigi-Sierra**[3] | **Dailu Guan**[3] | **Juan Vicente Delgado**[2] | **Javier Fernández Alvarez**[4] | **Marcel Amills**[3,5] | **Amparo Martínez Martínez**[2]

[1]Animal Breeding Consulting S.L. Parque Científico Tecnológico de Córdoba, Córdoba, Spain

[2]Departamento de Genética, Universidad de Córdoba, Campus Universitario de Rabanales, Córdoba, Spain

[3]Centre for Research in Agricultural Genomics (CRAG), CSIC-IRTA-UAB-UB, Universitat Autònoma de Barcelona, Bellaterra, Spain

[4]Asociación Nacional de Criadores de Caprino de Raza Murciano-Granadina (CAPRIGRAN), Granada, Spain

[5]Departament de Ciència Animal i dels Aliments, Facultat de Veterinària, Universitat Autònoma de Barcelona, Bellaterra, Spain

Correspondence
Amparo Martínez Martínez, Departamento de Genética, Universidad de Córdoba, Córdoba, 14014, Spain.
Email: amparomartinezuco@gmail.com

Marcel Amills, Centre for Research in Agricultural Genomics (CRAG), CSIC-IRTA-UAB-UB, Universitat Autònoma de Barcelona, Bellaterra, 08193, Spain.
Email: marcel.amills@uab.cat

Funding information
CERCA Programme/Generalitat de Catalunya and of the Spanish Ministry of Economy and Competitivity for the Center of Excellence Severo Ochoa 2020-2023, Grant/Award Number: CEX2019-000902-S; Spanish Ministry of Economy and Competitivity and by "ESF Investing in your future", Grant/Award Number: BES-C-2017-079709

Abstract

Hematological traits are important indicators of health status, and they are frequently used as criteria for clinical diagnosis. In humans, the genomic architecture of blood traits has been investigated in depth and thousands of associations with genetic variants have been found. In contrast, the association between marker genotypes and the variation of hematological traits has not been investigated in goats yet. Herewith, we have recorded 12 hematological parameters in 882 Murciano-Granadina goats that were also genotyped with the Goat SNP50 BeadChip (Illumina). Performance of a univariate genome-wide association study (GWAS) made it possible to detect one genomic region on goat chromosome (CHI) 21 (19.2–19.5 Mb) associated, at the genome-wide level of significance, with 4 red blood cell traits. The three markers displaying the highest significances were rs268272996 (CHI21: 19225290 bp), rs268273004 (CHI21: 19565629 bp) and rs268239059 (CHI13: 9615190 bp). Consistently, a multivariate GWAS indicated that the rs268273004 marker on chromosome 21 is associated with seven blood cell traits. Interestingly, this marker maps close to the FA Complementation Group I (*FANCI*) gene (CHI21: 20021947–20077025 bp), which is functionally related to Fanconi anemia, a syndrome characterized by bone marrow failure, aplastic anemia, and congenital disorders. We have also uncovered additional chromosome-wide significant associations between genetic markers and erythrocyte and leukocyte traits in the univariate GWAS. These findings evidence that the phenotypic variation of hematological traits in goats is regulated, at least to some extent, by polygenic determinants distributed in multiple chromosomes.

KEYWORDS
caprine, GWAS, hemogram, SNP

INTRODUCTION

Blood cells can be classified in two fundamental populations: red blood cells (RBC), which regulate the transport of oxygen and iron and facilitate hemostasis (Astle et al., 2016); and white blood cells (WBC), which are involved in acquired and innate immunity as well as in the development of allergic and inflammatory responses (Buttari et al., 2015). Neutrophils (NE) are the most numerous white blood cells, and they have a prominent role in the elimination of pathogens through a variety of mechanisms including degranulation, phagocytosis, and the amplification of inflammation (Janeway Jr et al., 2001). Eosinophils (EO), along with mast cells and basophils (BA), control mechanisms associated with allergy (Janeway Jr et al., 2001). Moreover, they release mediators triggering inflammatory processes and killing microbes and parasites (Janeway Jr et al., 2001). Monocytes (MO) are produced by the bone marrow and are released into the bloodstream to reach sites of infection, increasing in size and becoming macrophages able to engulf and digest microbes (Janeway Jr et al., 2001). Finally, B and T lymphocytes (LYM) are the main effectors of adaptive immunity. B LYM express surface immunoglobulin receptors which can recognize specific antigen epitopes and they are involved in the humoral immune response, while, in the thymus, T LYM mature into helper and cytotoxic T LYM, which are the primary effectors for cell-mediated immunity (Janeway Jr et al., 2001).

Investigating the genetic component of hematological parameters in livestock could be useful to understand the basis of pathological conditions related with anemia, immunodeficiency, thrombosis, and cancer. Zhang et al. (2014) identified 161 significant SNPs, including 44 genome-wide significant SNPs, associated with 11 hematological traits in Sutai pigs, while haplotype analysis revealed 154 genome-wide significant SNPs associated with nine hematological traits. In the polled yak, Ma et al. (2019) found 43 significant SNPs associated with nine hematological traits and five loci were classified as candidate genes for hematological traits. More recently, Chinchilla-Vargas et al. (2020) recorded 15 blood parameters in 570 beef cattle and found that their heritabilities range from 0.11 to 0.60. Moreover, they detected a few genomic regions with strong effects on blood traits and several candidate genes of interest were identified (Chinchilla-Vargas et al., 2020). As a whole, these results evidence that hematological parameters in livestock have an important genetic component that can be dissected via genomic approaches.

The genetic basis of blood parameters has not been explored yet in goats. In the current work, we aimed to fill this gap by performing univariate and multivariate genome-wide association studies (GWAS) for 12 hematological traits recorded in 882 Murciano-Granadina goats.

MATERIALS AND METHODS

Sample collection and phenotype recording

From 2017 to 2018, blood samples were collected from 1000 Murciano-Granadina (MG) breed female goats distributed in 11 intensive and three semi-intensive smallholder farms located in the province of Granada (Spain), with an average herd size of 70 individuals. All the animals sampled are registered in the National Association of Murciano-Granadina Breed Goats (CAPRIGRAN). In this sampling, 102 and 705 goats are maternal and paternal half-sisters, respectively, plus 74 full-sister pairs. Most were born between 2015 and 2016, except ~40 individuals born between 2005 and 2014. Nearly all goats were evaluated during their first lactation, although 10% of animals were scored between the second and sixth lactation. Blood samples were taken by jugular venipuncture, placed in 2 mL vials containing EDTA, and stored at −20°C in a freezer in the Animal Breeding Consulting Laboratory at the University of Córdoba (Spain). By using an automated Abbott Cell-Dyn 3700 Hematology Analyzer (GMI Laboratory), the Garfia SL Veterinary Laboratory (https://laboratorioveterinario.vet/) provided measurements for the following hematological parameters:

1. *Erythrocyte traits*: red blood cell count (RBC), hematocrit level (HCT), hemoglobin content (HGB), mean corpuscular volume (MCV), mean corpuscular hemoglobin (MCH), mean corpuscular hemoglobin concentration (MCHC), and red cell distribution width (RDW).
2. *Leukocyte traits*: counts of leukocytes (WBC), NE, EO, MO, and LYM.

Genomic DNA extraction and genotyping

The purification of genomic DNA was achieved with a modified salting-out procedure as reported by Guan et al. (2020). Genomic DNA samples were genotyped with the Illumina GoatSNP50 BeadChip (Illumina Inc.) according to the instructions of the manufacturer. This chip contains 54 241 SNPs uniformly distributed across the goat genome with an average spacing between SNPs of approximately 40 kb (Tosser-Klopp et al., 2014). Genomic locations of the SNPs and cluster files were provided by the International Goat Genome Consortium (http://www.goatgenome.org/). Standard SNP genotype quality control procedures were performed using PLINK v 1.9 (Chang et al., 2015). Markers with a call rate of <0.90, minor allele frequency (MAF) lower than 0.05, mapping to sex chromosomes or yielding genotypes that deviated significantly from the Hardy Weinberg equilibrium ($p < 0.001$) were filtered out. Moreover, individuals with a genotype call rate below 0.95 were also

excluded. By inspecting the data, we also noticed that 119 individuals had measurements for traits MCV, HCT, and MCHC that were highly discordant from those obtained in the remaining 882 individuals. These 119 goats were removed from the data set. After applying these quality control criteria, 47 401 SNPs and 882 individuals were retained to perform subsequent analyses. To investigate the population structure, a principal component analysis (PCA) was carried out with PLINK v 1.9 (Chang et al., 2015) and results were displayed on the basis of the first 2 components of the PCA. The PCA based on hematological data was built with the prcomp function (https://cran.r-project.org/web/packages/geomorph/vignettes/geomorph.PCA.html).

Univariate GWAS for blood traits

We plotted the distribution for each of the hematological traits with R (R Core Team, 2017). To verify the normality of such distributions, we performed a Shapiro–Wilk test (Shapiro & Wilk, 1965), and subsequently those phenotypes that deviated significantly from normality (RDW, MCHC, MCH, MCV, leukocytes, MO, EO, NE, LYM) were rank-based transformed by using the package GenAbel from R (Aulchenko et al., 2007). We did so because we assessed that this transformation yields residuals that are normally distributed, as previously reported by Luigi-Sierra et al. (2020). Then, association analyses were performed with the Genome-wide Efficient Mixed-Model Association (GEMMA) v 0.98.1 package (Zhou & Stephens, 2012) by fitting the following univariate linear mixed model for each trait:

$$\mathbf{y} = \mathbf{W\alpha} + \mathbf{x\beta} + \mathbf{u} + \boldsymbol{\varepsilon}; \mathbf{u} \sim \mathrm{MVN}_n\left(\mathbf{0}, \lambda\tau^{-1}\mathbf{K}\right) \text{ and } \boldsymbol{\varepsilon} \sim \mathrm{MVN}_n\left(\mathbf{0}, \tau^{-1}\mathbf{I_n}\right)$$

where $\mathbf{y}$ is a n-vector of hematological phenotypes for n=882 individuals; $\mathbf{W}$ is a $n \times c$ (c=number of fixed factors) matrix including a column of 1 s and the fixed effects, i.e. farm (14 levels), year of birth (10 levels) and batch effect (12 levels), since samples were submitted to the hematology laboratory in 12 batches; $\boldsymbol{\alpha}$ is a c-vector of the corresponding fixed effects including the intercept; $\mathbf{x}$ is a n-vector of marker genotypes; $\boldsymbol{\beta}$ is the effect size of the marker (allele substitution effect); $\mathbf{u}$ is a n-vector of random individual effects with a normal distribution $\mathbf{u} \sim \mathrm{N}(\mathbf{0}, \lambda\tau^{-1}\mathbf{K})$, where τ^{-1} is the variance of the residual error, λ is the ratio between the two variance components; and K is a known $n \times n$ relatedness matrix derived from SNP genotypes. Finally, $\boldsymbol{\varepsilon}$ is a n-vector of errors (Zhou & Stephens, 2012) and $\mathrm{I_n}$ is a $n \times n$ identity matrix. $\mathrm{MVN_n}$ corresponds to the n-dimensional multivariate normal distribution. In this study, likelihood ratio tests were carried out for each SNP by contrasting the alternative hypothesis ($\mathrm{H_1}$: $\beta \neq 0$) against the null hypothesis ($\mathrm{H_0}$: $\beta = 0$). Moreover, population structure was corrected by considering the K relatedness matrix, which takes all genome-wide SNPs as a random effect. The p-values obtained for each association were corrected for multiple testing by using the false discovery rate method (Benjamini & Hochberg, 1995). Associations with a q-value below 0.05 were considered as significant. Manhattan plots were built with the 'qqman' R package (R Core Team, 2017). The proportion of phenotypic variance explained by a given SNP (PVE) was estimated using the formula reported by Shim et al. (2015):

$$\mathrm{PVE} = \frac{2\beta^2\ \mathrm{MAF}\ (1-\mathrm{MAF})}{2\beta^2\ \mathrm{MAF}\ (1-\mathrm{MAF}) + \left[\mathrm{se}(\beta)\right]^2\ 2\mathrm{N}\ \mathrm{MAF}\ (1-\mathrm{MAF})}$$

where β is the estimated effect size of the SNP variant, se(β) is the standard error of the β estimate, MAF is the minor allele SNP frequency, and N is the sample size. Lambda inflation factors were calculated with the P λ function available in the R package QCEWAS (Van der Most et al., 2017), while the quantile quantile (Q-Q) plots were built with the ggqqplot function (https://www.rdocumentation.org/packages/ggplot2/versions/3.4.0/topics/ggplot). Genes located ±1 Mb around the significant SNPs were retrieved by using the Biomart tool from Ensembl (Kinsella et al., 2011) and subsequently analyzed with Uniprot (UniProt Consortium, 2019) and David Bioinformatic Resources (Huang et al., 2009) to annotate their biological functions.

Multivariate GWAS for blood traits

Besides the univariate GWAS (u-GWAS), we also performed a multivariate GWAS (MV-GWAS) implemented in GEMMA (Zhou & Stephens, 2012). Two independent MV-GWAS were performed for red blood cell and white blood cell traits. The main reasons for performing a MV-GWAS is that when traits are genetically correlated, this method takes advantage of cross-trait covariance as a source of information (while in the u-GWAS this is ignored), reduces the multiple testing burden (a single association test is performed for multiple traits) and, in the presence of pleiotropy, it is more consistent with biology (Galesloot et al., 2014). By using GEMMA, we implemented a multivariate linear mixed model:

$$\mathbf{Y} = \mathbf{WA} + \mathbf{x\beta^T} + \mathbf{U} + \mathbf{E}; \mathbf{U} \sim \mathbf{MN_{n\times d}}\left(\mathbf{0}, \mathbf{K}, \mathbf{V_g}\right), \mathbf{E} \sim \mathbf{MN_{n\times d}}\left(\mathbf{0}, \mathbf{I_{n\times n}}, \mathbf{V_e}\right)$$

where $\mathbf{Y}$ is a $n \times d$ matrix of n (n=882) individuals and d (d=12) phenotypes; $\mathbf{W}$ is a $n \times c$ matrix of covariates (fixed effects, as defined in the u-GWAS); $\mathbf{A}$ is a $c \times d$ matrix of the corresponding coefficients including the intercept; $\mathbf{x}$ is a n-vector of marker genotypes; $\boldsymbol{\beta}$ is the effect size of the marker for the d phenotypes; $\mathbf{U}$ is a $n \times d$ matrix of random effects; $\mathbf{E}$ is a $n \times d$ matrix of errors; $\mathbf{K}$ is a known $n \times n$ relatedness matrix derived from SNP genotypes; $\mathbf{I_{n\times n}}$ is a $n \times n$ identity matrix; $\mathbf{V_g}$ is a $d \times d$ symmetric matrix of genetic variance component; $\mathbf{V_e}$ is a $d \times d$ symmetric matrix of environmental

variance component and $MN_{n\times d}$ **(0, V_1, V_2)** denotes the a $n\times d$ matrix normal distribution with mean 0, row covariance matrix $\mathbf{V_1}$ ($n\times n$), and column covariance matrix $\mathbf{V_2}$ ($d\times d$) (Zhou & Stephens, 2012). Here, the null hypothesis, which assumes that marker effect sizes for all phenotypes are **0** (where **0** is a d-vector of zeroes), is contrasted against the alternative hypothesis (H_1: $\beta\neq\mathbf{0}$). The implementation of the false discovery rate correction for multiple testing (Benjamini & Hochberg, 1995) and the construction of Manhattan and QQ plots and the calculation of λ inflation factors were carried out as reported in the u-GWAS.

RESULTS AND DISCUSSION

Phenotypic variation of blood traits in Murciano-Granadina goats

The PCAs based on 50 K genotypes (Figure S1) and on 12 hematological phenotypes (Figure S2) did not reveal a strong clustering of goats according to their farm of origin. In the PCA based on 50 K genotypes, goats from farms GFB and GFM were clearly differentiated from their counterparts. In contrast, individuals from the remaining farms were quite scattered throughout the PCA and it was not possible to capture any significant clustering pattern or population substructure based on farm assignment. The raw means, standard deviations, and coefficient of variation of the 12 hematological traits recorded in 882 Murciano-Granadina goats are described in Table 1, and the distribution of raw blood trait measurements is indicated in Figure S3. Raw values did not differ substantially from those reported by Newcomer et al. (2021). The majority of blood trait means displayed low standard deviations, except white blood cell counts (14 487 ± 3919 cells per μL) which were slightly above the range values defined by Newcomer et al. (2021) i.e., 4000–12 000 leukocytes per μL. In the same sense, monocyte (462 cells per μL) and eosinophil (300 cells per μL) counts obtained from MG goats were in the range values reported by Newcomer et al. (2021) i.e., 0–500 monocytes per μL and 0–900 eosinophils per μL. Univariate and MV-GWAS consistently detect one SNP close to the Fanconi Anemia Complementation Group I gene as highly associated with red cell phenotypes.

After performing the u-GWAS, we detected 3 SNPs displaying genome-wide significant associations with blood traits. One SNP (rs268273004) on goat chromosome CHI21 (19.56 Mb) displayed genome-wide significant associations with MCH, MCHC, HCT, and MCV, another SNP (rs268272996), mapping to CHI21 (19.22 Mb), was significantly associated with MCH and the third SNP (rs268239059) mapping to CHI13 (9.22 Mb) was significantly associated with HCT (Figures 1, S4, Table 2).

The MV-GWAS also made it possible to detect a genome-wide significant association between the rs268273004 SNP (CHI21: 19.56 Mb) and seven red blood cell traits i.e., RBC, MCV, HGB, MCH, HCT, MCHC, and

TABLE 1 Descriptive statistics of 12 hematological traits recorded in 882 Murciano-Granadina goats.

Trait	Abbreviation	Range	Mean±SD	CV
Red blood cell count ($\times10^6$/μL)	RBC	8.61–16.60	14.063 ± 0.920	0.065
Mean corpuscular volume (fL)	MCV	15.70–34.70	22.959 ± 1.695	0.073
Hemoglobin concentration (g/dL)	HGB	6.40–14.70	10.677 ± 1.050	0.098
Mean corpuscular hemoglobin (pg/cell)	MCH	5.59–11	7.604 ± 0.701	0.092
Hematocrit (%)	HCT	20.50–44.80	32.358 ± 3.854	0.119
Mean corpuscular hemoglobin concentration (g/dL)	MCHC	20.30–51.20	33.332 ± 4.168	0.125
Red cell distribution width (%)	RDW	25.90–47.60	36.795 ± 3.855	0.104
Leukocyte count (cells per μL)	WBC	4100–32 200	14 487 ± 3919	0.270
Neutrophil count (cell per μL)	NE	420–27 950	7677 ± 4299	0.559
Lymphocyte count (cells per μL)	LYM	450–17 150	6047 ± 3125	0.516
Monocyte count (cells per μL)	MO	10–6100	462 ± 560	1.213
Eosinophil count (cells per μL)	EO	0–2940	300 ± 396	1.322

Abbreviations: CV, coefficient of variation; SD, standard deviation.

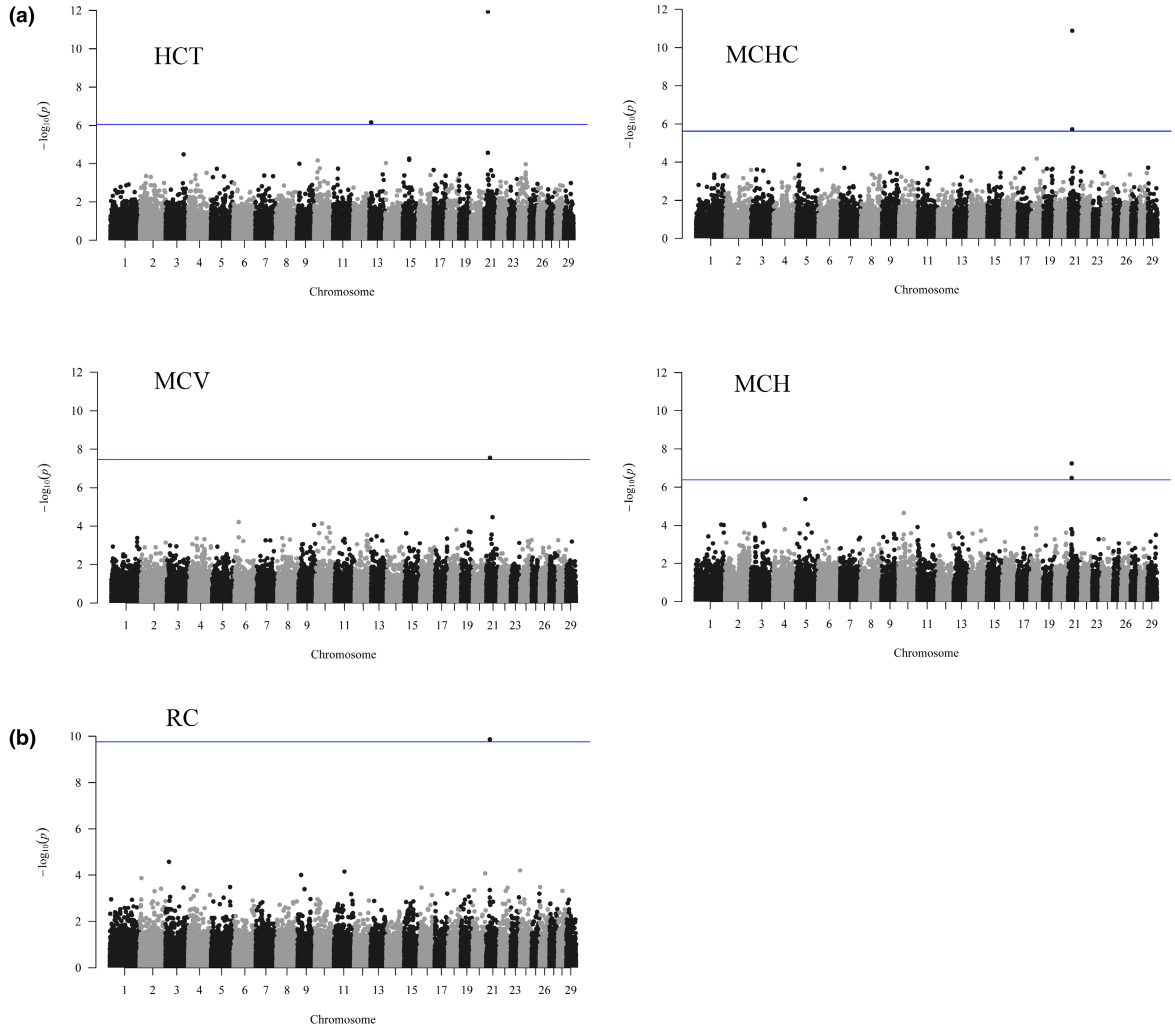

FIGURE 1 Manhattan plots displaying the results of the (a) univariate genome-wide association study for four hematological traits (HCT, hematocrit; MCHC, mean corpuscular hemoglobin concentration; MCH, mean corpuscular hemoglobin; MCV, mean corpuscular volume) and (b) the multivariate genome-wide association study for seven red cell (RC) phenotypes (HCT, hematocrit; HGB, hemoglobin; MCH, mean corpuscular hemoglobin; MCHC, mean corpuscular hemoglobin concentration; MCV, mean corpuscular volume; RBC, red blood cell count and RDW, red cell distribution width) in Murciano-Granadina goats. In the Manhattan plots, negative $\log_{10}$ *p*-values (y-axis) of the associations between SNPs and blood phenotypes are plotted against the genomic location of each SNP marker (x-axis). The blue line indicates the threshold of genome-wide significance after correction for multiple testing (*q*-value = 0.05).

RDW (Figures 1, S4, Table 3). The significance of this association for multiple traits (Table 3, *q*-value = 6.70×10^{-06}) was, in general, much higher than the individual significances obtained in the u-GWAS for each trait (Table 2). Moreover, three additional associations (with MCV, HGB, and RDW) were uncovered in the MV-GWAS when compared to the u-GWAS (Tables 2 and 4). These features are consistent with the increased statistical power of the MV-GWAS over the u-GWAS when traits under study are genetically correlated (Galesloot et al., 2014).

Close to SNPs rs268273004 and rs268272996, we detected the Fanconi Anemia Complementation Group I (*FANCI*) gene (CHI21: 22.15–22.22 Mb). The *FANCI* gene forms part of the Fanconi Anemia pathway implicated in the removal of interstrand DNA crosslinks to maintain genomic stability (Joo et al., 2011). Mutations in this gene can cause Fanconi anemia (Savage et al., 2015), a hereditary syndrome characterized by bone marrow failure, aplastic anemia, and congenital disorders (Sondalle et al., 2019). Knockout mice for the *FANCI* gene display hematopoietic defects as well as dwarfism and microphthalmia, with resemblance to typical Fanconi anemia symptoms (Dubois et al., 2019). Besides, Fanconi anemia patients display alterations of the cytoskeleton and redox state of erythrocytes (Malorni et al., 2000). We hypothesize that near or within the goat *FANCI* gene there might be at least one causal mutation with pleiotropic effects on diverse red blood cell traits but not leading to any observable pathology. Indeed, Tomaszowski et al. (2020) pointed out that homozygous knock-outs for single *FANC* genes only display mild and selective disease phenotypes rather than the full repertoire of Fanconi anemia symptoms, thus suggesting that polygenic gene inactivation is required for disease manifestation.

TABLE 2 Genome-wide significant associations between SNP markers and red blood cell parameters recorded in 882 Murciano-Granadina goats (univariate analysis).

Trait	Chr	rs	Pos	A1	MAF	β±SE	*p*-value	*q*-value	PVE
HCT	13	rs268239059	9 615 190	G	0.214	1.074±0.215	6.877×10^{-07}	1.035×10^{-06}	0.027
	21	rs268273004	19 565 629	G	0.204	1.420±0.240	7.903×10^{-09}	0.0002	0.038
MCHC	21	rs268272996	19 225 290	G	0.160	**−26.870±5.355**	8.068×10^{-07}	0.014	0.027
	21	rs268273004	19 565 629	G	0.204	**−32.254±4.605**	1.346×10^{-11}	2.036×10^{-07}	0.052
MCH	21	rs268272996	19 225 290	G	0.160	**−31.469±5.555**	3.210×10^{-08}	0.0005	0.035
	21	rs268273004	19 565 629	G	0.204	**−29.153±4.908**	6.687×10^{-09}	0.0002	0.038
MCV	21	rs268273004	19 565 629	G	0.204	**22.109±3.957**	2.792×10^{-08}	0.0014	0.034

Abbreviations: A1, minor allele; Chr, chromosome; HCT, hematocrit; MAF, minor allele frequency; MCH, mean corpuscular hemoglobin; MCHC, mean corpuscular hemoglobin concentration; MCV, mean corpuscular volume; Pos, position in base pairs; *p*-value, raw *p*-values; PVE, proportion of variance in phenotype explained by a given SNP; *q*-value, *p*-values corrected for multiple testing using a false discovery rate approach; rs, identifier code of the SNP according to the RefSNP database; β±SE, allelic substitution effect ± standard error, the β-values in bold correspond to rank-based transformed phenotypes.

TABLE 3 Genome-wide significant associations between SNP markers and red blood cell parameters recorded in 882 Murciano-Granadina goats (multivariate analysis).

Group	Chr	rs	Pos	A1	MAF	β_RBC	β_MCV	β_HGB	β_MCH	β_HCT	β_MCHC	β_RDW	*p*-value	*q*-value
RC	21	rs268273004	19 565 629	G	0.204	0.238	−0.511	1.422	−0.135	−0.460	2.811	−2.578	2.10×10^{-10}	6.70×10^{-06}

Abbreviations: A1, minor allele; Chr, chromosome; HCT, hematocrit; HGB, hemoglobin; MAF, minor allele frequency; MCH, mean corpuscular hemoglobin; MCHC, mean corpuscular hemoglobin concentration; MCV, mean corpuscular volume; Pos, position in base pairs; *p*-value, raw *p*-values; *q*-value, *p*-values corrected for multiple testing using a false discovery rate approach; RBC, red blood cell count; RDW, red cell distribution width; RC, red cell phenotypes; rs, identifier code of the SNP according to the RefSNP database; β±SE, allelic substitution effect ± standard error, the β-values in bold correspond to rank-based transformed phenotypes.

Multiple genomic regions are associated at the chromosome-wide level of significance with red blood cell traits

At the chromosome-wide level of significance, the MV-GWAS did not yield any significant association while the u-GWAS made possible to detect several associations between genomic regions and red blood cell traits (Table 4). Interestingly, some of these genomic regions harbored candidate genes involved in red blood cell homeostasis. For example, the rs26824749 marker on CHI19 (43 Mb) associated with MCHC was located close to two genes related to red blood cell function: the solute carrier family 4 member 1 (Diego Blood Group) locus (*SLC4A1* CHI19: 43.7 Mb) and the dual specificity protein phosphatase 3 (*DUSP3* CHI19: 43.3 Mb) The main function of the SLC4A1 protein is to mediate the chloride/bicarbonate anion exchange across the plasma membrane of the red blood cell, a process necessary for efficient respiration (Reithmeier et al., 2016). In humans, mutations in the *SLC4A1* gene produce pleiotropic effects associated with distal renal tubular acidosis (More & Kedar, 2021), along with red blood cell membrane protein disorders such as hereditary spherocytosis, Southeast Asian ovalocytosis, and hereditary stomatocytosis (Kager et al., 2017; Yang et al., 2018). by contrast, there is growing evidence that the *DUSP3* gene regulates the immune response, thrombosis, hemostasis, angiogenesis, and genomic stability (Russo et al., 2018). In the mice, $DUSP3^{-/-}$ mutants had a 30% decrease in the hemoglobin content of tumors compared to wild type mice. This suggests that DUSP3 deficiency leads to tumor-induced defective angiogenesis (Vandereyken et al., 2017).

Chromosome-wide significant associations with leukocyte traits

No genome-wide significant association was detected for leukocyte traits neither in the MV-GWAS nor in the u-GWAS, but several chromosome-wide significant associations and positional candidate genes implicated in immunity were identified in the u-GWAS (Table 5). For instance, the rs268270592 SNP on CHI19: 41.50 Mb is associated with EOSI count and resides ~0.5 Mb apart from the signal transducer and activator of transcription 5B (*STAT5B*, CHI19: 41.95 Mb) and the signal transducer and activator of transcription 3 (*STAT3*, CHI19: 42.05 Mb) genes. The STAT5B transcription factor regulates the functional properties of immune cells activated by STAT5-dependent cytokines (Majri et al., 2018). The abolishment of STAT5 function causes growth retardation and immunodeficiency characterized by T and NK cell lymphopenia, impaired regulatory T cell frequency and function and hypergammaglobulinemia (Majri et al., 2018). However, the *STAT3* gene is a key regulator of inflammation and immunity, and its malfunction has been linked to immunodeficiency, autoimmunity and cancer (Hillmer et al., 2016). Interestingly, mutations in the *STAT3* gene cause an autosomal dominant hyper-immunoglobulin E syndrome characterized by immunodeficiency and

TABLE 4 Chromosome-wide significant associations between SNP markers and red blood cell parameters recorded in 882 Murciano-Granadina goats (univariate analysis).

Trait	Chr	rs	Pos	A1	MAF	β±SE	*p*-value	*q*-value	PVE
RBC	10	rs268241981	19 118 502	C	0.421	0.198±0.045	1.878×10^{-05}	0.030	0.021
	21	rs268273004	19 565 629	G	0.204	0.247±0.054	5.561×10^{-06}	0.006	0.023
HCT	13	rs268239059	9 615 190	G	0.214	1.074±0.215	1.035×10^{-06}	1.099×10^{-03}	0.028
	21	rs268272996	19 225 290	G	0.163	1.060±0.247	2.421×10^{-05}	0.014	0.020
		rs268273004	19 565 629	G	0.204	1.548±0.212	2.788×10^{-012}	8.860×10^{-10}	0.057
HGB	1	rs268261815	149 849 592	G	0.300	−0.255±0.057	1.519×10^{-05}	0.032	0.022
	21	rs268243835	36 736 182	A	0.266	0.245±0.056	1.723×10^{-05}	0.019	0.021
MCV	21	rs268273004	19 565 629	G	0.204	**22.109±3.957**	4.828×10^{-08}	3.948×10^{-05}	0.034
		rs268239456	24 671 090	G	0.086	**−23.518±5.664**	4.210×10^{-05}	0.024	0.019
MCH	5	rs268281605	51 636 882	G	0.200	**24.335±5.228**	6.978×10^{-06}	0.009	0.024
	10	rs268234586	27 862 169	A	0.492	**18.793±4.396**	3.216×10^{-05}	0.046	0.020
	21	rs268272996	19 225 290	G	0.163	**−30.125±5.873**	5.110×10^{-07}	2.365×10^{-04}	0.029
	21	rs268273004	19 565 629	G	0.204	**−28.520±5.226**	9.492×10^{-08}	8.134×10^{-05}	0.033
MCHC	19	rs268247491	43 604 975	A	0.107	**−23.630±5.516**	2.542×10^{-05}	0.023	0.020
	21	rs268272996	19 225 290	G	0.163	**−26.549±4.603**	8.068×10^{-07}	7.275×10^{-06}	0.036
	21	rs268273004	19 565 629	G	0.204	**−30.629±4.033**	3.811×10^{-14}	9.513×10^{-11}	0.061
RDW	26	rs268287192	23 996 545	C	0.428	**−18.407±4.380**	4.234×10^{-05}	0.031	0.020
	28	rs268288378	25 429 471	A	0.162	**21.989±5.615**	1.065×10^{-04}	0.041	0.017

Abbreviations: A1, minor allele; Chr, chromosome; HCT, hematocrit; HGB, hemoglobin; MAF, minor allele frequency; MCH, mean corpuscular hemoglobin; MCHC, mean corpuscular hemoglobin concentration; MCV, mean corpuscular volume; Pos, position in base pairs; *p*-value, raw *p*-values; PVE, proportion of variance in phenotype explained by a given SNP; *q*-value, *p*-values corrected for multiple testing using a false discovery rate approach; RBC, red blood cell count; RDW, red cell distribution width; rs, identifier code of the SNP according to the RefSNP database; β±SE, allelic substitution effect ± standard error, the β-values in bold correspond to rank-based transformed phenotypes.

TABLE 5 Chromosome-wide significant associations between SNP markers and white blood cell parameters recorded in 882 Murciano-Granadina goats (univariate analysis).

Trait	Chr	rs	Pos	A1	MAF	β±SE	*p*-value	*q*-value	PVE
WBC	27	rs268283565	3 323 775	C	0.449	**−16.989±3.942**	2.134×10^{-05}	0.015	0.021
NE	14	rs268282961	56 866 131	A	0.237	**20.754±4.718**	1.431×10^{-05}	0.021	0.021
	19	rs268288222	37 122 148	G	0.327	**−18.691±4.101**	7.206×10^{-06}	0.006	0.023
	29	rs268262584	15 882 374	A	0.295	**−19.212±4.595**	5.505×10^{-05}	0.035	0.019
EO	2	rs268276883	87 591 464	G	0.299	**−21.030±4.802**	1.573×10^{-05}	0.034	0.021
	19	rs268270592	41 496 242	A	0.346	**20.622±4.522**	7.648×10^{-06}	0.006	0.023
MO	29	rs268270265	19 590 673	A	0.241	**18.583±4.407**	3.252×10^{-05}	0.024	0.020
LYM	28	rs268271853	14 868 625	C	0.282	**18.722±4.576**	4.674×10^{-05}	0.040	0.019

Abbreviations: A1, minor allele; Chr, chromosome; EO, eosinophils; LYM, lymphocytes; MAF, minor allele frequency; MO, monocytes; NE, neutrophils; Pos, position in base pairs; *p*-value, raw *p*-values; PVE, proportion of variance in phenotype explained by a given SNP; *q*-value, *p*-values corrected for multiple testing using a false discovery rate approach; rs, identifier code of the SNP according to the RefSNP database; WBC, leukocytes; β±SE, allelic substitution effect ± standard error, the β-values in bold correspond to rank-based transformed phenotypes.

chronic eosinophilia (Holland et al., 2007). Moreover, STAT3 activation in tumoral T-cells has been associated with the in situ presence of eosinophils (Fredholm et al., 2014), providing additional support for the potential involvement of *STAT3* genotype on the determination of EOSI counts in goats.

Another interesting association is between the rs268288222 (CHI19: 37.12 Mb) genetic marker and NE count. This SNP maps near to the calcium binding and coiled-coil domain 2 (*CALCOCO2*) gene, which is a known xenophagy receptor mediating intracellular bacterial degradation (Cui et al., 2021).

In summary, we have detected several genomic regions associated with red blood cell and leukocyte count parameters in goats. The identification of these regions would be a first step towards detect mutations with causal effects on a set of phenotypes which are tightly related to immunity, susceptibility to disease and animal welfare. Indeed, leukocyte counts can be considered as biomarkers of the immune response against pathogens

as well as of inflammation. In consequence, investigating the genetic factors that regulate the abundance of leukocytes in blood might yield valuable knowledge that, in the midterm future, could be applied to the selection of goats resistant (or tolerant) to specific infectious diseases. Similarly, the determination of the genetic factors governing red blood cell and hemoglobin parameters in goats might have future applications in the selection of individuals resistant (or tolerant) to blood-borne parasites, such as trypanosoma, theileria, babesia and anaplasma, which cause great economic losses mostly in tropical and subtropical areas.

CONCLUSIONS

Our study provides new insights about the genomic architecture of hematological traits in goats. We have identified the existence of multiple genomic regions associated, either at the genome or chromosome-wide level of significance, with several hematological parameters, thus demonstrating that these phenotypes are genetically determined in goats. The most promising result is the genome-wide significant association between multiple red blood cell phenotypes and one genetic marker on CHI21 (rs268273004, 19.5 Mb), which was consistently detected in the MV-GWAS and the u-GWAS. This marker maps close to the *FANCI* gene, which encodes a DNA repair protein implicated in Fanconi anemia. One pleiotropic causal mutation or several causal mutations tightly linked to rs268273004 might explain the associations with erythrocyte phenotypes found in our study.

ACKNOWLEDGMENTS

This research was funded by project PID2019-105805RB-I00 funded by MCIN/AEI/10.13039/501100011033. We also acknowledge the support of the CERCA Programme/Generalitat de Catalunya and of the Spanish Ministry of Economy and Competitivity for the Center of Excellence Severo Ochoa 2020-2023 (CEX2019-000902-S) grant awarded to the Centre for Research in Agricultural Genomics (CRAG, Bellaterra, Spain). Maria Luigi-Sierra was funded with a PhD fellowship 'Formación de Personal Investigador' (BES-C-2017-079709) awarded by the Spanish Ministry of Economy and Competitivity and by 'ESF Investing in your future'. Dailu Guan was funded by a PhD fellowship from the China Scholarship Council (CSC).

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no competing interests.

DATA AVAILABILITY STATEMENT

Phenotypes and genotypes of Murciano-Granadina goats are available from the Figshare https://figshare.com/articles/dataset/A_genome_wide_association_study_for_hematological_traits_in_Murciano-Granadina_goats/14459598.

ORCID

Martina Macri https://orcid.org/0000-0003-0114-384X
Maria Gracia Luigi-Sierra https://orcid.org/0000-0002-6414-0550
Dailu Guan https://orcid.org/0000-0001-8800-3158
Juan Vicente Delgado https://orcid.org/0000-0003-1657-8838
Marcel Amills https://orcid.org/0000-0002-8999-0770
Amparo Martínez Martínez https://orcid.org/0000-0002-6944-0501

REFERENCES

Astle, W.J., Elding, H., Jiang, T., Allen, D., Ruklisa, D., Mann, A.L. et al. (2016) The allelic landscape of human blood cell trait variation andchilinks to common complex disease. *Cell*, 167, 1415–1429. Available from: https://doi.org/10.1016/j.cell.2016.10.042

Aulchenko, Y.S., Ripke, S., Isaacs, A. & Van Duijn, C.M. (2007) GenABEL: an R library for genome-wide association analysis. *Bioinformatics*, 23, 1294–1296. Available from: https://doi.org/10.1093/bioinformatics/btm108

Benjamini, Y. & Hochberg, Y. (1995) Controlling the false discovery rate-a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society B*, 57, 289–300. Available from: https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Buttari, B., Profumo, E. & Riganò, R. (2015) Crosstalk between red blood cells and the immune system and its impact on atherosclerosis. *BioMed Research International*, 2015, 616834. Available from: https://doi.org/10.1155/2015/616834

Chang, C.C., Chow, C.C., Tellier, L.C.A.M., Vattikuti, S., Purcell, S.M. & Lee, J.J. (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*, 4, 7. Available from: https://doi.org/10.1186/s13742-015-0047-8

Chinchilla-Vargas, J., Kramer, L.M., Tucker, J.D., Hubbell, D.S., III, Powell, J.G., Lester, T.D. et al. (2020) Genetic basis of blood-based traits and their relationship with performance and environment in beef cattle at weaning. *Frontiers in Genetics*, 11, 717. Available from: https://doi.org/10.3389/fgene.2020.00717

Cui, F., Wang, S., Tan, J., Tang, H., Fan, Y. & Hu, J. (2021) Calcium-binding and coiled-coil domain 2 promotes the proliferation and suppresses apoptosis of prostate cancer cells. *Experimental and Therapeutic Medici.E*, 21, 1–9. Available from: https://doi.org/10.3892/etm.2021.9836

Dubois, E.L., Guitton-Sert, L., Béliveau, M., Parmar, K., Chagraoui, J., Vignard, J. et al. (2019) A Fanci knockout mouse model reveals common and distinct functions for FANCI and FANCD2. *Nucleic Acids Research*, 47, 7532–7547. Available from: https://doi.org/10.1093/nar/gkz514

Fredholm, S., Gjerdrum, L.M.R., Willerslev-Olsen, A., Petersen, D.L., Nielsen, I.Ø., Kauczok, C.S. et al. (2014) STAT3 activation and infiltration of eosinophil granulocytes in mycosis fungoides. *Anticancer Research*, 34, 5277–5286.

Galesloot, T.E., Van Steen, K., Kiemeney, L.A., Janss, L.L. & Vermeulen, S.H. (2014) A comparison of multivariate genome-wide association methods. *PLoS One*, 9, e95923. Available from: https://doi.org/10.1371/journal.pone.0095923

Guan, D., Landi, V., Luigi-Sierra, M.G., Delgado, J.V., Such, X., Castelló, A. et al. (2020) Analyzing the genomic and transcriptomic architecture of milk traits in Murciano-Granadina goats. *Journal of Animal Science and Biotechnology*, 11, 35. Available from: https://doi.org/10.1186/s40104-020-00435-4

Hillmer, E.J., Zhang, H., Li, H.S. & Watowich, S.S. (2016) STAT3 signaling in immunity. *Cytokine & Growth Factor Reviews*, 31, 1–15. Available from: https://doi.org/10.1016/j.cytogfr.2016.05.001

Holland, S.M., DeLeo, F.R., Elloumi, H.Z., Hsu, A.P., Uzel, G., Brodsky, N. et al. (2007) STAT3 mutations in the hyper-IgE syndrome. *New England Journal of Medicine*, 357, 1608–1619. Available from: https://doi.org/10.1056/NEJMoa073687

Huang, W., Sherman, B.T. & Lempicki, R.A. (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*, 4, 44–57. Available from: https://doi.org/10.1038/nprot.2008.211

Janeway, C.A., Jr., Travers, P., Walport, M. & Shlomchik, M.J. (2001) The components of the immune system. In: *Immunobiology: the immune system in health and disease*, 5th edition. New York, USA: Garland Science.

Joo, W., Xu, G., Persky, N.S., Smogorzewska, A., Rudge, D.G., Buzovetsky, O. et al. (2011) Structure of the FANCI-FANCD2 complex: insights into the Fanconi anemia DNA repair pathway. *Science*, 333, 312–316. Available from: https://doi.org/10.1126/science.1205805

Kager, L., Bruce, L.J., Zeitlhofer, P., Flatt, J.F., Maia, T.M., Ribeiro, M.L. et al. (2017) Band 3 null VIENNA, a novel homozygous *SLC4A1* p.Ser477X variant causing severe hemolytic anemia, dyserythropoiesis and complete distal renal tubular acidosis. *Pediatric blood. Cancer*, 64, e26227. Available from: https://doi.org/10.1002/pbc.26227

Kinsella, R.J., Kähäri, A., Haider, S., Zamora, J., Proctor, G., Spudich, G. et al. (2011) Ensembl BioMarts: a hub for data retrieval across taxonomic space. *Database*, 2011, bar030. Available from: https://doi.org/10.1093/database/bar030

Luigi-Sierra, M.G., Landi, V., Guan, D., Delgado, J.V., Castelló, A., Cabrera, B. et al. (2020) A genome-wide association analysis for body, udder, and leg conformation traits recorded in Murciano-Granadina goats. *Journal of Dairy Science*, 103, 1605–1617. Available from: https://doi.org/10.3168/jds.2020-18461

Ma, X., Jia, C., Fu, D., Chu, M., Ding, X., Wu, X. et al. (2019) Analysis of hematological traits in polled yak by genome-wide association studies using individual SNPs and haplotypes. *Genes*, 10, 463. Available from: https://doi.org/10.3390/genes10060463

Majri, S.S., Fritz, J.M., Villarino, A.V., Zheng, L., Kanellopoulou, C., Chaigne-Delalande, B. et al. (2018) STAT5B: a differential regulator of the life and death of $CD4^{+}$ effector memory. *Journal of Immunology*, 200, 110–118. Available from: https://doi.org/10.4049/jimmunol.1701133

Malorni, W., Straface, E., Pagano, G., Monti, D., Zatterale, A., Del Principe, D. et al. (2000) Cytoskeleton alterations of erythrocytes from patients with Fanconi's anemia. *FEBS Letters*, 468, 125–128. Available from: https://doi.org/10.1016/S0014-5793(00)01187-X

More, T.A. & Kedar, P.S. (2021) Genotypic analysis of *SLC4A1* A858D mutation in Indian population associated with distal renal tubular acidosis (dRTA) coupled with hemolytic anemia. *Gene*, 769, 145241. Available from: https://doi.org/10.1016/j.gene.2020.145241

Newcomer, B.W., Cebra, C., Chamorro, M.F., Reppert, E., Cebra, M. & Edmondson, M.A. (2021) Diseases of the hematologic, immunologic, and lymphatic systems (multisystem diseases). In: Pugh, D.G., Baird, N., Edmondson, M. & Passler, T. (Eds.) *Sheep, Goat, and Cervid Medicine*, 3rd edition. Missouri: Elsevier, pp. 405–438. Available from: https://doi.org/10.1016/B978-0-323-62463-3.00025-6

R Core Team. (2017) *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Available from: https://www.r-project.org/

Reithmeier, R.A., Casey, J.R., Kalli, A.C., Sansom, M.S., Alguel, Y. & Iwata, S. (2016) Band 3, the human red cell chloride/bicarbonate anion exchanger (AE1, SLC4A1), in a structural context. *Biochimica et Biophysica Acta (BBA) – Biomembranes*, 1858, 1507–1532. Available from: https://doi.org/10.1016/j.bbamem.2016.03.030

Russo, L.C., Farias, J.O. & Forti, F.L. (2018) Revisiting the roles of VHR/DUSP3 phosphatase in human diseases. *Clinics*, 73, e466s. Available from: https://doi.org/10.6061/clinics/2018/e466s

Savage, S.A., Sumislawski, J.J., Zarzaur, B.L., Dutton, W.P., Croce, M.A. & Fabian, T.C. (2015) The new metric to define large-volume hemorrhage: results of a prospective study of the critical administration threshold. *Journal of Trauma and Acute Care Surgery*, 78, 224–230. Available from: https://doi.org/10.1097/TA.0000000000000502

Shapiro, S.S. & Wilk, M.B. (1965) An analysis of variance test for normality (complete samples). *Biometrika*, 52, 591–611.

Shim, H., Chasman, D.I., Smith, J.D., Mora, S., Ridker, P.M., Nickerson, D.A. et al. (2015) A multivariate genome-wide association analysis of 10 LDL subfractions, and their response to statin treatment, in 1868 caucasians. *PLoS One*, 10, e0120758. Available from: https://doi.org/10.1371/journal.pone.0120758

Sondalle, S.B., Longerich, S., Ogawa, L.M., Sung, P. & Baserga, S.J. (2019) Fanconi anemia protein FANCI functions in ribosome biogenesis. *Proceedings of the National Academy of Sciences of the USA*, 116, 2561–2570. Available from: https://doi.org/10.1073/pnas.1811557116

Tomaszowski, K.H., Roy, S., Keshvani, C., Ott, M., DiNardo, C., Schindler, D. et al. (2020) Polygenic mutations model the pleiotropic disease of Fanconi anemia. *BioRxiv*. Available from: https://doi.org/10.1101/2020.09.01.277038

Tosser-Klopp, G., Bardou, P., Bouchez, O., Cabau, C., Crooijmans, R., Dong, Y. et al. (2014) Design and characterization of a 52K SNP chip for goats. *PLoS One*, 9, e86227. Available from: https://doi.org/10.1371/journal.pone.0152632

UniProt Consortium. (2019) UniProt: a worldwide hub of protein knowledge. *Nucleic. Nucleic Acids Research*, 47, D506–D515. Available from: https://doi.org/10.1093/nar/gky1049

Van der Most, P.J., Küpers, L.K., Snieder, H. & Nolte, I. (2017) QCEWAS: automated quality control of results of epigenome-wide association studies. *Bioinformatics*, 33, 1243–1245. Available from: https://doi.org/10.1093/bioinformatics/btw766

Vandereyken, M., Jacques, S., Van Overmeire, E., Amand, M., Rocks, N., Delierneux, C. et al. (2017) Dusp3 deletion in mice promotes experimental lung tumour metastasis in a macrophage dependent manner. *PLoS One*, 12, e0185786. Available from: https://doi.org/10.1371/journal.pone.0185786

Yang, E., Seo-Mayer, P., Lezon-Geyda, K., Badior, K.E., Li, J., Casey, J.R. et al. (2018) A Ser725Arg mutation in band 3 abolishes transport function and leads to anemia and renal tubular acidosis. *Blood*, 131, 1759–1763. Available from: https://doi.org/10.1182/blood-2018-01-827725

Zhang, F., Zhang, Z., Yan, X., Chen, H., Zhang, W. & Hong, Y. (2014) Genome-wide association studies for hematological traits in Chinese Sutai pigs. *BMC Genetics*, 15, 41. Available from: https://doi.org/10.1186/1471-2156-15-41

Zhou, X. & Stephens, M. (2012) Genome-wide efficient mixed-model analysis for association studies. *Nature Genetics*, 44, 821–824. Available from: https://doi.org/10.1038/ng.2310

SUPPORTING INFORMATION

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

How to cite this article: Macri, M., Luigi-Sierra, M.G., Guan, D., Delgado, J.V., Alvarez, J.F., Amills, M. et al. (2023) Univariate and multivariate genome-wide association studies for hematological traits in Murciano-Granadina goats. *Animal Genetics*, 54, 491–499. Available from: https://doi.org/10.1111/age.13328