Short Communication

Genome Wide Association Study for Antibody Response to *Mycobacterium avium ssp. paratubeculosis* in Goats

Giulietta Minozzi[1*], Maria Grazia De Iorio[1,] Fiorentina Palazzo[2], Gustavo Gandini[1], Stefano Biffani[3], Gianluigi Paolillo[1], Elena Ciani[4], Vincenzo Di Marco Lo Presti[5], Alessandra Stella[3], John L. Williams[6,7]

1 Department of Veterinary Medicine DIMEVET, Università degli Studi di Milano, Lodi, 26900, Italy;

2 Faculty of Bioscience and Technology for Food, Agriculture and Environment, Università degli Studi di Teramo, Teramo, 64100, Italy;

3 IBBA-CNR, Milano, 20133, Italy;

4 Department of Biosciences, Biotechnologies and Biopharmaceutics, Università degli Studi di Bari "Aldo Moro", Bari, 70121, Italy;

5 Istituto Zooprofilattico Sperimentale della Sicilia "A. Mirri", Area Territoriale Barcellona Pozzo di Gotto, Messina, 90129, Italy; University of Adelaide, Roseworthy, 5371, Australia;

6 Davies Research Centre, School of Animal and Veterinary Sciences, University of Adelaide, Roseworthy, 5371, Australia;

7 Present address: Dipartimento di Scienze Animali, della Nutrizione e degli Alimenti, Università Cattolica del Sacro Cuore, 29122 Piacenza, Italy.

*Correspondence: Giulietta.minozzi@unimi.it; (G.M);
telephone number 0039.025033473, fax 0039.02.503334500

Summary: *Mycobacterium avium ssp. paratuberculosis* (MAP), causes Johne's disease (JD), or paratuberculosis, a chronic enteritis of ruminants, which in goats is characterized by ileal lesions. The work described here is a case-control association study using the Illumina Caprine SNP50 BeadChip to unravel the genes involved in susceptibility of goats to JD. Goats in herds with high occurrence of Johne's disease were classified as healthy or infected based on the level of serum antibodies against MAP, and 352 animals were selected for the association study. Whole genome association analysis (GWAS) identified four chromosomal regions that were significantly associated with antibody response to MAP ($P < 5x10^{-5}$) in the Jonica breed, on chromosomes 3, 9 and 11, while 3 regions, on chromosomes 1, 10 and 23, were identified in the Siriana breed. The regions on chromosomes 10 and 11 overlapped in the combined GWAS of both breeds. These results provide evidence for genetic loci involved in the antibody response to MAP in goats and identified breed specific and non-specific loci.
Keywords: caprine; SNPs; paratuberculosis; ELISA; genetic association.

Paratuberculosis or Johne's disease (JD) is a chronic granulomatous gastroenteritis caused by infection with *Mycobacterium avium subspecies paratuberculosis* (MAP) in ruminants. JD is not treatable, and is present worldwide [Nielsen & Toft (2009); Fernández-Silva *et al.* 2014]. In goats JD is clinically less evident than in cattle, and is characterized by chronic inflammatory lesions of the intestine and lymphoid organs that reduce nutrient uptake [Windsor 2015]. An Italian study found a high incidence of goat JD with farm level incidence between 29% and 39% [Galiero *et al.* 2017]. The estimated heritability of JD in goats ranges from 0.12 and 0.07 [van Hulzen *et al.* 2012].

A genome wide association study (GWAS) in Garfagnina goats identified several regions associated with serological response to MAP [Cecchi *et al.* 2017]. In addition to the GWAS, RNAseq and MicroRNA sequencing studies in cattle have identified expression patterns associated with JD progression and have suggested their potential as diagnostic tools [Malvisi *et al.* 2020; Malvisi *et al.* 2016; Marino *et al.* 2017; Casey *et al.* 2015; Fiorentina et al. 2020]. The extension of these studies to goats is of interest considering the differences in progression of the disease among species.

The present study used a case-control design to identify genetic loci associated with antibody response against MAP in goats.

Samples were collected in goat herds with high occurrence of JD, based on the serum antibodies produced in response to MAP infection detected using the ELISAID Screen® Paratubercolosis Indirect Screening test (Id.Vet Montpellier, France). All animals were at least two years of age, females and selected from as many herds as possible. The goats belonged to two breeds, the Siriana breed (174 samples, 87 cases and 87 controls) selected from 14 herds, and the Jonica breed (157 samples, 77 cases and 80 controls) from 10 herds. The minimum number of positive animals belonging to one herd was 2. Negative controls were chosen from the same herd sampled on the same day. Cases were defined as animals serologically positive for MAP by ELISA with a sample to positive ratio (S/P) higher than 0.7 and MAP negative animals had a S/P lower than 0.6. Positive animals were tested twice with the ID Screen Paratuberculosis confirmation test (ID. Vet Montpellier, France).

Blood samples were collected in EDTA vacutainers (BD Vacutainer Systems, Plymouth, UK), refrigerated at 4°C and then stored at -20°C. DNA was extracted from whole blood samples using the NucleoSpinVRBlood kit (Macherey-Nagel, Duren, Germany). The 331 samples were genotyped using the Illumina GoatSNP50 BeadChip (Illumina Inc., San Diego). Genotype quality control was performed in the R statistical environment using the "check.marker" function of the GenABEL package [Aulchenko *et al.* 2007a]. SNP missing 5% of data, or with MAF of less than 1% were removed. Genotyping efficiency for samples was also verified, and samples with more than 5% missing data were removed.

Genome-wide association analysis was performed using the GenABEL package [Aulchenko *et al.* 2007a] in R using a three step GRAMMAR-CG approach, (Genome Wide Association using Mixed Model and Regression - Genomic Control), using the genomic kinship matrix estimated from genomic marker data with the ''ibs'' (option weight= ''freq'') function of GenABEL [Amin *et al.*

2007; Aulchenko *et al*. 2007b]. Analysis of the combined cohort included breed as fixed effect in the model. Uncorrected p-values of $\leq 5 \times e^{-05}$ were considered as significant associations . SNP location and gene names were based on the *Capra hircus* ARS1 assembly. All analyses were carried out within the R statistical environment (R: The R project for statistical computing 2020).

For the Jonica breed, the final data set that passed the quality controls consisted of 50864 SNPs and 148 samples. Genome wide analysis identified 4 significant SNPs on chromosomes 3, 9 and 11 (Table 1). All four highly significant SNPs were located in intronic regions of genes: respectively *REV1*, *IMPG1*, *NCK2* and *SETDB1*. A further SNP on chromosome 14 was located in an intergenic region showed a p-value close to significance. The ten most significant SNPs are given in Table 1.

For the Siriana breed the final data set contained 51215 SNPs and 151 samples. In total, 3 significant SNPs were identified on chromosomes 1, 23 and 10, all located in intergenic regions. Three further SNPs on chromosomes 4 showed p-values close to significance, these were in an intron of *GSDME*, on chromosome 29, in an intron of *PACS1* and on chromosome 8 in an intron of *MFSD14B*.

A combined analysis of data for both breeds was conducted on a final dataset set of 51151 genome wide SNPs and 299 samples, which identified significant SNPs on chromosomes 11, 10 and 7 (Figure 1). The SNP on chromosome 11 was significant in the Jonica breed analysis, while the SNP on chromosome 10 was significant in the Siriana cohort. Two SNPs that had p-values close to significance on chromosomes 1 and 8 which were among the 10 SNPs identified in the Siriana breed. Five additional SNPs with p-values close to significance were identified on chromosomes 3, 11, 15, 20 and 21 (Table 1).

In summary, the analyses identified 26 SNPs located on 17 different chromosomes, 16 fall within genes, of these 2 are both functional and positional candidates (*DOK5, PACS1*), 2 are functional candidates (*NCK2, GSDME*) and 3 positional candidates (*DDX51, SETDB1, IMPG1*).

The SNP rs268277403, identified in the Jonica cohort, is located on chromosome 13 in an intronic region of the gene *DOK5*. *DOK5* is a member of the Docking Protein gene family, expressed in activated T cells that probably plays a role in immune regulation [Favre *et al.* 2003]. In both sheep

and goat *DOK5* is located on chromosome 13 and is ~5 megabases from a significant SNPs identified in a GWAS for paratuberculosis in sheep, at positions 75853910 and 75853910 on OAEv3.1 [Moioli *et al.* 2016].

The SNP rs268285112 which was close to significance in the Siriana cohort is located on chromosome 29 and lies in the intronic region of the gene Phosphofurin Acidic Cluster Sorting Protein 1 (*PACS1*), involved in the interaction of MAP with the immune system and may result in the persistence of infection and the long incubation period that occurs before the clinical stages of disease. *PACS1* is located on chromosome 2 goat and cattle and is ~7Mb from *APLP2* which was suggested as a candidate gene in two GWAS of JD in cattle [Kirkpatrick *et al.* 2011; Del Corvo *et al.* 2017].

The SNP rs268256774 was found to be significant in both Jonica cohort and Jonica and Siriana combined analysis. It is located on chromosome 11 in an intronic region of NCK adaptor protein 2 (*NCK2*). *NCK2* is involved in the stimulation of the expression of the T cell antigen receptors [Ngoenkam *et al.* 2014], suggesting a potential role in disease progression [Koo *et al.* 2004].

The only other Johne's disease GWAS of goats identified significant loci on 9 chromosomes [Cecchi *et al.* 2017], of which 5 are in common with present study. Among these, Cecchi et al found a significant SNP on chromosome 17 located less than 1Mb from the SNP rs268293054 identified in the present study.

The significant SNP rs268264217 found in Jonica cohort located on Chromosome 9 is within the Interphotoreceptor Matrix Proteoglycan 1 gene (*IMPG1*), which also has no obvious connection with JD. However, *IMPG1* is located on chromosome 8 in sheep, and is 2-3 mega bases away from *CD109* and *LOC10111282*, which have been identified as positional candidates for JD resistance [Moioli *et al.* 2016]. In addition, in cattle, *IMPG1* is located on chromosome 9, close to a region that has been associated with JD in several studies [Kirkpatrick *et al.* 2011; Alpay *et al.* 2014; Del Corvo *et al.* 2017]. Therefore, it is likely that this region of chromosomes 9 in goat has a functional variant with effect on response to JD.

In conclusion this study identified several chromosomal regions significantly associated with JD. However, identifying the functional variations will require much larger sample sets.

Acknowledgments

This work was carried out with the support of grant PON01_01841 PON EPISUD from the Italian Ministry of Education University and Research (www.istruzione.it) awarded to JLW. JLW is currently supported by the project LEO Livestock Environment Open-data, 16.2 – PSRN 2014-2020 funded through Fondo Europeo Agricolo per lo Sviluppo Rurale (FEASR). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Funding

This research was funded by The Italian Ministry of Education and Research (www.istruzione.it), grant number PON01_01841 PON EPISUD. G.P and M.G.D.I were funded by the BEENOMIX 2.0 project funded by the Lombardy Region (FEASR program) PSR (grant number 201801057971 – G44I19001910002).

Conflicts of Interest

The authors declare they do not have any conflict of interest.

Data Availability Statement

The data presented in this study are available at the following link https://cloud.cnr.it/owncloud/index.php/s/u8yfoXKCaIPxelQ.

Institutional Review Board Statement

The ethical approval was given by the OPBA (Organismo Preposto al Benessere Animale) of the University of Milan.

Author Contributions: Conceptualization, J.L.W and G.M.; methodology, G.M, F.P and A.S.; formal analysis, G.M., G.P..; resources, V.D.M.L.P.; data curation, S.B.; writing—original draft

preparation, G.M, M.G.D.I, and G.P..; writing—review and editing, J.L.W, A.S., G.G, E.C. and G.G.; project administration, J.L.W; funding acquisition, J.L.W, G.M. All authors have read and agree to the published version of the manuscript.

References

Alpay F., Zare Y., Kamalludin M.H. *et al*. (2014) Genome-Wide Association Study of Susceptibility to Infection by Mycobacterium avium Subspecies paratuberculosis in Holstein Cattle. Neilan B, ed. *PLoS ONE* 2014 **9(12)**, e111704.

Amin N., van Duijn C.M. & Aulchenko Y.S. (2007) A genomic background based method for association analysis in related individuals. *PLoS ONE* **2(12)**, e1274.

Aulchenko Y.S., de Koning D.J. & Haley, C. (2007) Genomewide rapid association using mixed model and regression: a fast and simple method for genome-wide pedigree-based quantitative trait loci association analysis. *Genetics* **177(1)**, 577-85.

Aulchenko Y.S., Ripke S., Isaacs A. & van Duijn C.M. (2007) GenABEL: an R package for genome-wide association analysis. *Bioinformatics*, **23**, 1294-6.

Casey M.E., Meade K.G., Nalpas N.C., Taraktsoglou M., Browne J.A., Killick K.E., Park S.D., Gormley E., Hokamp K., Magee D.A. & MacHugh D.E. (2015) Analysis of the Bovine Monocyte-Derived Macrophage Response to Mycobacterium avium Subspecies Paratuberculosis Infection Using RNA-seq. *Frontiers in Immunology* **6**, 23.

Cecchi F., Russo C., Iamartino D., Galiero A., Turchi B., Fratini F., Degl'Innocenti S., Mazza R., Biffani S., Preziuso G. & Cantile C. (2017) Identification of candidate genes for paratuberculosis resistance in the native Italian Garfagnina goat breed. *Tropical Animal Health and Production* **49(6)**, 1135-1142.

Del Corvo M., Luini M., Stella A., Pagnacco G., Ajmone-Marsan P., Williams J.L. & Minozzi, G. (2017) Identification of additional loci associated with antibody response to Mycobacterium

avium ssp. Paratuberculosis in cattle by GSEA-SNP analysis. *Mammalian genome: official journal of the International Mammalian Genome* **28(11-12)**, 520-527.

Favre C., Gérard A., Clauzier E., Pontarotti P., Olive D. & Nunès J.A. (2003) DOK4 and DOK5: New Dok-related Genes Expressed in Human T Cells. *Genes and Immunity* **4(1)**, 40-45.

Fernández-Silva J.A., Correa-Valencia N.M. & Ramírez N.F. (2014) Systematic review of the prevalence of paratuberculosis in cattle, sheep, and goats in Latin America and the Caribbean. *Tropical Animal Health and Production* **46(8),** 1321-1340.

Fiorentina P., Martino C., Mancini Y., De Iorio M.G., Williams J.L., Minozzi G. (2021) Using Omics Approaches in the Discovery of Biomarkers for Early Diagnosis of Johne's Disease in Sheep and Goats. *Animals (Basel)* **11(7),** 1912.

Galiero A., Turchi B., Pedonese F., Nuvoloni R., Cantile C., Colombani G., Forzan M., Cerri D., Bandecchi P. & Fratini F. (2017) Serological, culture and molecular survey of Mycobacterium avium paratuberculosis in a goat flock in Tuscany. *Folia Microbiolgica (Praha)* **62(6),** 471-477.

Kirkpatrick B.W., Shi X., Shook G.E. & Collins M.T. (2011) Whole-Genome association analysis of susceptibility to paratuberculosis in Holstein cattle. *Animal Genetics* **42(2),** 149-160.

Koo H.C., Park Y.H., Hamilton M.J., Barrington G.M., Davies C.J., Kim J.B., Dahl J.L., Waters W.R. & Davis W.C. (2004) Analysis of the immune response to Mycobacterium avium subsp. paratuberculosis in experimentally infected calves. *Infection and Immunity* **72(12),** 6870-6883.

Malvisi M., Curti N., Remondini D., De Iorio M.G., Palazzo F., Gandini G., Vitali S., Polli M., Williams J.L. & Minozzi G. (2020) Combinatorial Discriminant Analysis Applied to RNAseq Data Reveals a Set of 10 Transcripts as Signatures of Exposure of Cattle to Mycobacterium avium subsp. paratuberculosis. *Animals (Basel).* **10(2),** 253.

Malvisi M., Palazzo F., Morandi N., Lazzari B., Williams J.L., Pagnacco G. & Minozzi G. (2016) Responses of Bovine Innate Immunity to Mycobacterium avium subsp.paratuberculosis Infection Revealed by Changes in Gene Expression and Levels of MicroRNA. *PLoS One* **11(10),** e0164461.

Marino R., Capoferri R., Panelli S., Minozzi G., Strozzi F., Trevisi E., Snel G.G.M., Ajmone-Marsan P. & Williams. J.L. (2017) Johne's disease in cattle: an in vitro model to study early response to infection of Mycobacterium avium subsp. paratuberculosis using RNA-seq. *Molecular Immunology* **91**, 259-271.

Moioli B., D'Andrea S., De Grossi L., Sezzi E., De Sanctis B., Catillo G., Steri, R., Valentini A. & Pilla F. (2016) Genomic scan for identifying candidate genes for paratuberculosis resistance in sheep. *Animal Production Science* **56,** 1046-1055.

Ngoenkam J., Paensuwan P., Preechanukul K., Khamsri B., Yiemwattana I., Beck-García E., Minguet S., Schamel W.W. & Pongcharoen S. (2014) Non-overlapping functions of Nck1 and Nck2 adaptor proteins in T cell activation. *Cell communication and signaling: CCS* **12**, 21.

Nielsen S.S. & Toft, N. (2009) A review of prevalences of paratuberculosis in farmed animals in Europe. *Preventive Veterinary Medicine* **88(1),** 1-14.

R: The R project for statistical computing. (2020) http://www.r-project.org

van Hulzen K.J., Koets A.P., Nielen M., Hoeboer J., van Arendonk J.A. & Heuven H.C. (2012) Genetic variation for infection status as determined by a specific antibody response against Mycobacterium avium subspecies paratuberculosis in milk of Dutch dairy goats. *Journal of Dairy Science* **95(10),** 6145-6151.

Windsor P.A. (2015) Paratuberculosis in sheep and goats. *Veterinary Microbioliogy* **181(1-2),** 161-169.

Table 1. List of TOP 10 SNP identified by GWAs to be associated with antibody response to MAP.

Joniche	rs	chr	Position	A1	n	EFF.B	p-value	Variant type	Gene name	Ensembl ID
1	rs268236287	11	4499463	G	148	-0.26	9.38e-	intron variant	REV1	ENSCHIG00000012748
2	rs268264217	9	2833936	A	148	0.29	1.28e-	intron variant	IMPG1	ENSCHIG00000008626
3	rs268256774	11	45510753	G	148	0.21	4.93e-	intron variant	NCK2	ENSCHIG00000010383
4	rs268250033	3	100339710	G	148	0.22	4.94e-	intron variant	SETDB1	ENSCHIG00000010698
5	rs268282873	14	52902030	G	148	-0.21	8.18e-05	intergenic variant	-	-
6	rs268283652	8	25285942	C	148	-0.23	1.22e-04	intron variant	ADAMTSL1	ENSCHIG00000025981
7	rs268264223	9	3115401	G	148	0.25	1.29e-04	intergenic variant	-	-
8	rs268280287	1	89136936	A	148	0.28	1.32e-04	intergenic variant	-	-
9	rs268277403	13	81560549	G	148	-0.24	1.61e-04	intron variant	DOK5	ENSCHIG00000019478
10	rs268254513	15	9755045	G	148	0.20	2.41e-04	intron variant	-	ENSCHIG00000001728
Siriana	**rs**	**chr**	**Position**	**A1**	**n**	**EFF.B**	**p-value**	**Variant Type**	**Gene name**	**Ensembl ID**
1	rs268272525	1	79233377	A	151	0.27	3.92e-	intergenic variant	-	-
2	rs268281657	23	38870024	C	151	-0.27	1.49e-	intergenic variant	-	-
3	rs268289803	10	67331443	G	151	-0.25	4.69e-	intergenic variant	-	-
4	rs268259981	4	49132543	A	151	-0.23	6.48e-05	intron variant	GSDME	ENSCHIG00000026791
5	rs268264182	9	12275894	G	151	-0.23	6.75e-05	intergenic variant	-	-
6	rs268285112	29	44576267	A	151	0.22	8.31e-05	intron variant	PACS1	ENSCHIG00000022748
7	rs268235970	22	7214691	A	151	-0.21	9.25e-05	intergenic variant	-	-
8	rs268293054	17	26179499	A	151	0,25	1.14e-04	Downstream	DDX51	ENSCHIG00000014268
9	rs268293054	17	26179499	A	151	0.25	1.14e-04	Downstream		ENSCHIG00000021392
10	rs268275405	3	88668156	A	151	-0.21	1.59e-04	5' UTR variant	PIFO	ENSCHIG00000008928
11	rs268286138	8	251428	A	151	-0.24	1.63e-04	intron variant	MFSD14B	ENSCHIG00000015957
Jonica&Siriana	**rs**	**chr**	**Position**	**A1**	**n**	**EFF.B**	**p-value**	**Variant Type**	**Gene name**	**Ensembl ID**
1	rs268256774	11	45510753	G	299	0.17	1.22e-	intron variant	NCK2	ENSCHIG00000010383
2	rs268289803	10	67331443	G	299	-0.18	1.68e-	intergenic variant	-	-
3	rs268247914	7	66849525	A	299	0.29	3.81e-	intergenic variant	-	-
4	rs268272525	1	79233377	A	299	0.15	8.66e-05	intergenic variant	-	-
5	rs268286138	S8	251428	A	299	-0.16	1.05e-04	intron variant	MSFD14B	ENSCHIG00000015957
6	rs268290018	20	60967107	G	299	-0.15	1.16e-04	intergenic variant	-	-
7	rs268275901	3	27660329	G	299	0.16	1.33e-04	intron variant	SLC1A7	ENSCHIG00000019509
8	rs268247750	21	8424938	G	299	-0.14	1.71e-04	intergenic variant	-	-
9	rs268272606	11	28473481	A	299	0.15	1.74e-04	intergenic variant	-	-
10	rs268274835	15	20174073	G	299	0.16	2.17e-04	intergenic variant	-	-

rs: SNP rs; chr: goat chromosome; Position: position on the goat genome in base pairs; A1: coding of the minor allele A1; n: number of animals tested; EFF.B: effect of the minor allele; p-value: p-values adjusted for Genomic Control;

Figure 1. Manhattan plot displaying the results of the Genome-wide scan with respect to their genomic position in the combined cohort of the Jonica and Siriana breeds. The threshold for genome-wide significance (P < 5 x 10-5) is indicated by the red line.