CSIRO PUBLISHING

Animal Production Science, 2016, **56**, 1046–1055
http://dx.doi.org/10.1071/AN14826

Genomic scan for identifying candidate genes for paratuberculosis resistance in sheep

Bianca Moioli[A,E], *Silvia D'Andrea*[B], *Luigi De Grossi*[C], *Erminia Sezzi*[C], *Bruno De Sanctis*[C], *Gennaro Catillo*[A], *Roberto Steri*[A], *Alessio Valentini*[D] *and Fabio Pilla*[B]

[A]Consiglio per la Ricerca e la sperimentazione in Agricoltura, via Salaria 31, 00015 Monterotondo, Italy.
[B]Dipartimento Agricoltura, Ambiente e Alimenti, Università degli Studi del Molise, 86100 Campobasso, Italy.
[C]Istituto Zooprofilattico Sperimentale, Strada Terme, 01100 Viterbo, Italy.
[D]DIBAF, Università degli Studi della Tuscia, Via S. Camillo de Lellis, 01100, Viterbo, Italy.
[E]Corresponding author. Email: bianca.moioli@entecra.it

Abstract. Breeding objectives relating to health, functional traits and welfare need to receive priority in the research programs and selection schemes, but very few reports are available on natural resistant genotypes in livestock, where some important diseases cause severe economic losses and pose serious zoonotic threats. In this study, diagnosis of paratuberculosis was performed on 759 adult sheep, from a single flock, with the serum antibody enzyme-linked immunosorbent assay; 100 sheep were selected among the extreme divergent animals for the S/P ratio obtained from the test, and were genotyped on the Illumina Ovine SNP50K BeadChip. A genome-wide scan was then performed on the individual marker genotypes, in the attempt to identify genomic regions associated with disease resistance in sheep. For each marker, the allelic substitution effect was calculated by regressing the S/P value on the number of copies of the reference allele. The position on the OARv3.1 Genome Assembly was searched for 32 markers, which showed a statistically significant allelic substitution effect (*Raw P* < 0.0006 and *FDR P* < 0.09). All markers were located within, or close to, annotated genes. Five of these genes, SEMA3, CD109, PCP4, PRDM2 and ITFG2 are referred in literature to play a role in either disease resistance or cell-mediated immune response.

Additional keywords: disease susceptibility, GWAS, Ovine SNP50K BeadChip.

Received 26 June 2014, accepted 3 November 2014, published online 12 February 2015

Introduction

Paratuberculosis, or Johne's, disease is a chronic granulomatous enteritis, which affects ruminants, caused by *Mycobacterium avium* subsp. *Paratuberculosis* (MAP). It impacts not only on agriculture economy but also on public health. In fact, although the relationship between MAP and Crohn's disease in humans is still subject of debate (Balfour Sartor 2005) migrant studies showed that the incidence of the disease, in people moving from a low Crohn's disease and Johne's disease incidence area to a high incidence area, subsequently rises to that of the host population (Hermon-Taylor 2009). Moreover Olsen *et al*. (2009) suggested a possible role of MAP in the inflammation seen in Crohn's disease. There are no effective treatments against MAP; control programs are very complex and expensive for farmers, and also the vaccine did not show to be sufficiently protective (Reddacliff *et al*. 2006). However, although MAP is widespread in the environment, the low rate of infected individuals, even in the same flock, suggested the presence of genetic factors influencing disease resistance.

Some functional candidate genes have been associated with MAP in cattle, and with increasing risk of Crohn's disease in humans. Toll-like receptors (TLR) showed polymorphisms associated with susceptibility to MAP infection (Mucha *et al*. 2009; Pinedo *et al*. 2009). Their role in the pathogenesis of Johne's disease has been identified based on gene expression in sheep (Nalubamba *et al*. 2008; Taylor *et al*. 2008). TLR, being members of mammalian pattern-recognition receptors, play a primary role in the recognition of pathogen-associated molecular patterns in bacteria, viruses, protozoa and fungi (Vasselon and Detmers 2002; Kaisho and Akira 2006). Several studies have shown that mutations in TLR genes may reduce the ability to recognise pathogen-associated molecular patterns and hence interfere with innate immune activation. TLR6 may be a potential marker of exposure to MAP and could be used to identify sheep resistant to MAP infection (Plain *et al*. 2010). Mutations found in TLR1 and 2 may increase the susceptibility to ovine mycobacterial infection (Bhide *et al*. 2009). NOD2 and CARD15 genes were associated with increasing risk of Crohn's disease in humans (Hugot *et al*. 2001; Hugot 2006); however, Sezzi *et al*. (2007) could not find any significant associations between the polymorphisms detected in the NOD2/CARD15 gene and the disease.

In order to identify genomic regions associated with Johne's disease paratuberculosis, whole-genome association studies

(GWAS) using the Illumina Bovine BeadChip, have been performed in cattle (Settles *et al*. 2009; Minozzi *et al*. 2010; Pant *et al*. 2010; Kirkpatrick *et al*. 2011; Zanella *et al*. 2011), but, in sheep, the studies are still scarce.

The recent availability of the Ovine SNP50k BeadChip (Kijas *et al*. 2012) containing more than 50 000 single nucleotide polymorphism (SNP), distributed along the whole genome, with knowledge of the position on each specific chromosome, recently started to give impulse to the research also in sheep breeding and sheep genomics, with a small number of examples of detected selection sweeps (Moradi *et al*. 2012; Moioli *et al*. 2013). With the aim to identify putative candidate genes that play a role in the immune responses, and potentially affect paratuberculosis disease resistance, we genotyped with the Illumina SNP50K BeadChip two groups of dairy sheep of the Sarda breed, selected among the most divergent after diagnosis assessment using enzyme-linked immunosorbent assay (ELISA) test. The genomic regions either surrounding or encoding the markers, that showed a significant effect on the results of the ELISA test, were then checked for the presence of annotated genes either in sheep or other mammals.

Materials and methods

Animal samples and diagnostic assessment

Because the complicated nature of the long, slowly developing MAP infections and the lack of good reference tests might likely introduce selection bias in traditional test evaluations (Nielsen and Toft 2008), the trial was performed in a flock of 1000 dairy sheep of the Sarda breed, under the monitoring of the National Veterinary Services Laboratories for several years, that was known to have sheep infected with MAP. Seven-hundred and fifty-nine adult sheep were screened on serum through ELISA. A commercial kit (IDVet, Montpellier, France) was used to diagnose paratuberculosis with the serum antibody ELISA test. Sera were treated according to the manufacturer's protocol. Animals were considered either serologically positive or negative based on the optical density ratio between the sample and the positive control, as follows:

$$\mathrm{S/P} = [\text{Optical density (OD) sample} - \text{Avg. OD normal control}]/(\text{Avg. OD positive control} - \text{Avg. OD normal control}).$$

With the purpose of characterising selective sweep in sheep for resistance to MAP infection, we applied the selective genotyping strategy, which consists of genotyping the extreme divergent animals for a trait, in this case the potential disease resistance. Within the screened 759 sheep, 50 positive and 50 negative sheep were selected, the kit considering positivity S/P values >70. A second ELISA assessment was performed on serum of positive individuals so to confirm the previous results. Average S/P value of the positive sheep was 145.02 ± 35.34; min. = 76.00; max. = 224.00; the corresponding value of the negative sheep was 0.44 ± 16.67; min. = –20.89; max. = 50.78. To further support resistance to MAP, all selected negative sheep were over 6 years of age, under the assumption that older animals were to be considered resistant for not having showed the infection for several consecutive years. The trial was performed in a commercial flock, which keeps no pedigree of the sheep, except the registration number of the ram; breeding groups are composed of ~30 sheep for each ram, and rams are maintained for breeding for about three consecutive years. Therefore, in order to prevent that the between-individuals relatedness affected the results, the sheep of each of the two groups were selected among the daughters of the same rams. The number of the positive-negative daughters of each ram was reported in Table 1.

By selecting highly divergent sheep in the same flock and of the same breed, the linkage disequilibrium (LD) between the SNP and the alleles of the true responsible genes affecting the trait is likely to have been maintained (Goddard and Hayes 2009; Wientjes *et al*. 2013), this being the required premise for a GWAS aiming to detect the genes influencing the target trait.

Genotyping

DNA of the 100 sheep of the two groups, extracted from blood sample by using Qiagen QIAamp DNA blood mini/midi kit (Qiagen, San Diego, CA, USA), was genotyped using the Ovine SNP50 BeadChip manufactured by Illumina (San Diego, CA, USA). Genotyping was performed by LGS (Cremona, Italy). Raw data were analysed using the GenomeStudio Genotyping Module version 1.7 (Illumina) by applying a no-call threshold of 0.15. Moreover, the markers not satisfying the following filtering parameters were excluded: SNP call rate $\leq$ 90%; SNP minor allele frequency $\leq$ 5%; out of Hardy–Weinberg equilibrium at $P < 0.01$.

Previous works on the genetic diversity of sheep breeds (Kijas *et al*. 2012; Ciani *et al*. 2014) made evident that the Ovine SNP50k BeadChip clearly separates the breeds and the geographical breeding areas on the basis of the markers' allele frequencies. Because the sheep of the present trial belonged to the same breed, the Sarda, and the same farm, a high frequency of markers with similar allele frequencies between the positive and the negative group had to be expected. For this reason, the markers where the frequency of the reference allele was <0.10 different between positive and negative sheep were considered non-informative and were excluded from the statistical analysis.

Statistical analyses and search for the annotated genes in the candidate regions

For each marker, the allelic substitution effect was calculated by regressing the S/P value on the number of copies of the reference allele, using the GLM procedure of SAS software (SAS Institute 2007). To control the chance of any false positive among those markers, the False Discovery Rate (FDR) correction was applied using the Multtest procedure of SAS software (SAS

Table 1. Distribution of the rams in the two groups of sheep

Ram	No. of daughters	
	Positive	Negative
IT VTA 0228200	8	10
IT VTA 0296664	9	8
IT VTA 0292797	8	7
IT VTA 0296898	8	11
IT VTA 0296622	8	8
IT VTA 0296375	9	6
Total	50	50

Institute 2007). It was arbitrarily decided that the allelic substitution effect of the potential markers for disease resistance should have *FDR P* < 0.09. For these markers, the position on the OARv3.1 genome assembly was assessed (http://www.livestockgenomics.csiro.au/sheep/). The presence of an annotated gene close to the marker was then checked in the National Center for Biotechnology Information (NCBI) Ovis aries genome database (http://www.ncbi.nlm.nih.gov/genome/?term=Ovis+aries, verified 23 November 2014), by exploring ~500 kb upstream and downstream the marker of the OARv3.1 region, as suggested by Zare *et al*. (2014). The explored region was larger than the haplotype block of LD estimated in humans, on one side because the size of these regions are still subject of debate (Reich *et al*. 2001); on the other side because modern cattle breeds are characterised by large LD blocks, likely caused by evolutionary forces such as genetic drift, admixture, selection and small effective population size, which are common in livestock (Odani *et al*. 2006; Khatkar *et al*. 2007).

The position of the marker either in the gene, or in the flanking regions, was assessed by performing a standard nucleotide Basic Local Alignment Search Tool (Altschul *et al*. 1990) of the OARv3.1 region encompassing the annotated gene with the deposited annotated gene sequence in the NCBI database.

There is increasing evidence that the onset of many human diseases is due to mutations in the intronic regions of genes (Alshatwi *et al*. 2012). Because such mutations may cause alterations in regulatory regions and splicing process (Doss and Sethumadhavan 2009), to verify whether any of these mutations occurred within the putative binding sites for transcription factors, *in silico* analysis was performed using the Match software (http://www.gene-regulation.com, verified 23 November 2014).

Results

Genotyping and data mining

After data editing, 5091 markers were available for the analysis; their distribution on the ovine chromosomes was reported in Table 2. Thirty-two markers satisfied the significance threshold, having *Raw P* ≤ 0.0006 and *FDR P* < 0.9. Fig. 1 represents the Manhattan plot of the *F*-test values obtained, for each marker, from the GLM procedure of SAS software (SAS Institute 2007). The horizontal line separates the 32 significant markers. For these

Table 2. Distribution on the ovine chromosomes of the markers used in the present study

Chromosome	Total markers	% markers
1	567	11.12
2	491	9.64
3	474	9.31
4	258	5.07
5	193	3.79
6	249	4.88
7	222	4.36
8	192	3.88
9	182	3.57
10	162	3.18
11	97	1.90
12	213	4.18
13	181	3.55
14	123	2.41
15	175	3.43
16	153	3.00
17	135	2.63
18	156	3.06
19	84	1.65
20	113	2.22
21	90	1.76
22	100	1.96
23	74	1.45
24	113	2.22
25	58	1.14
26	58	1.14
X	178	3.50
Total	5091	100.00

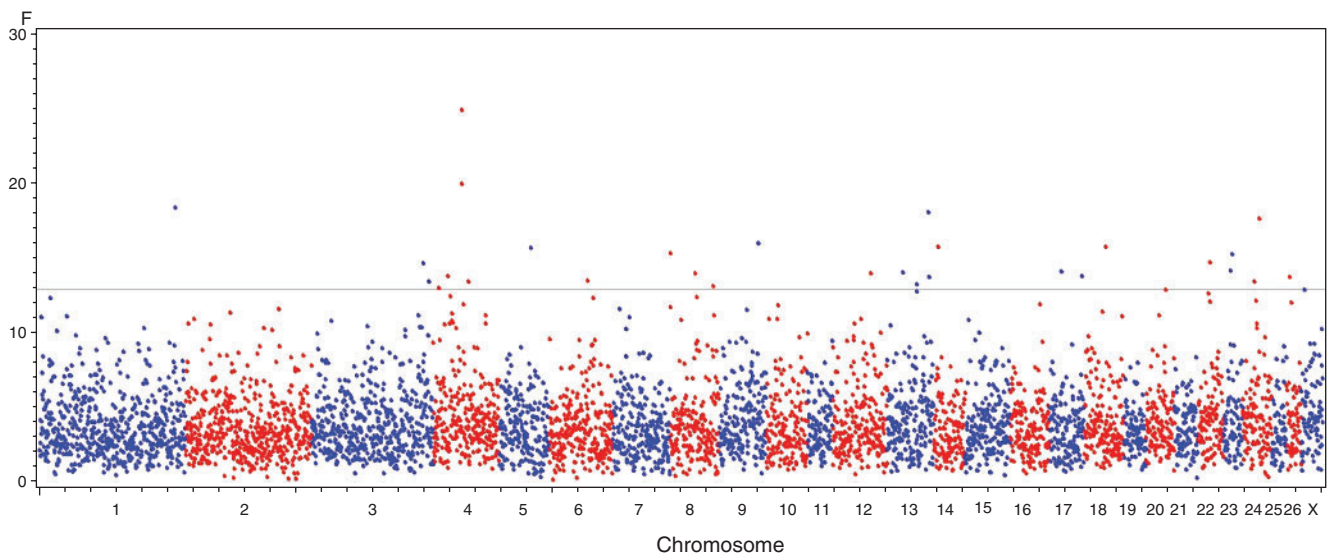

Fig. 1. Manhattan plot of the *F*-test values obtained, for each marker, from the GLM procedure. The horizontal line separates the 32 significant markers.

markers, the allelic substitution effect of the reference allele was reported in Table 3; in this Table, markers were sorted according to the statistical significance of the effect. The 32 markers fell either within the sequence, or in the 5′-3′ UTR, of a described gene. The *in silico* analysis (http://gene-regulation.com, verified 23 November 2014) indicated that two markers were located in putative binding site of transcription factors: OAR8_270360.1 was in the binding site of the myeloid zinc finger 1 (MZF1), reported as one regulator of transcriptional events during hemopoietic development (Hromas *et al*. 1991); OAR24_23721386.1 was in the binding site of the CCAAT Enhancer Binding Protein β (C/EBPbeta), which is required for the gene transcription of the nuclear factor of activated T-cells (Yang and Chow 2003).

Study of the annotated genes in the candidate regions

The search of the annotated genes was performed for each of the 32 markers of Table 3; the genes and the position in the gene of these markers as well the accession number of an annotated mRNA of the gene, when available, were reported in Table 4. In this Table the markers were sorted according to the chromosome and the position on the chromosome, to allow for a simplified evaluation of a potential LD of some of them, which might corroborate the effect of a putative candidate gene on the analysed performance. The Table reports also the gene function, as in the databases http://ghr.nlm.nih.gov/gene (verified 23 November 2014) and http://www.uniprot.org/uniprot (verified 23 November 2014); according to their function, the genes were classified in a pie graph based on the gene ontology (Fig. 2).

Discussion

With the purpose to detect genes that might play a role in disease resistance, this study was performed in a dairy flock that was known to have sheep infected with MAP; sheep were screened on serum through ELISA. Serology provides a cost-effective alternative to organism detection-based diagnostic methods for bovine paratuberculosis, although assay specificity is less than that for faecal culture (Collins 2002). Sensitivity and specificity estimates from studies of serum antibody ELISA for detection of affected, infectious and infected sheep were reported by Nielsen and Toft (2008) and Gumber *et al*. (2006). Both authors refer values for specificity ranging from 95% to 100% and suggest that appropriate animal sampling for flock diagnoses can be

Table 3. Allelic substitution effect of the reference allele, *F*-value and probability values for the 32 markers with *FDR P* < 0.09, with the position of the marker on OARv3.1

OAR	Marker	Position on OARv3.1	Ref. allele	Substitution effect of the ref. allele	*F*	*Raw_P*	*FDR_P*
4	OAR4_37455919.1	35463534	A	–69.9	24.87	.000003	.013
4	OAR4_37402301.1	35413050	G	–63.2	19.96	.000021	.054
1	OAR1_278980576.1	258148544	A	48.1	18.35	.000043	.064
13	OAR13_81680798.1	75853910	A	–49.36	18.02	.000050	.064
24	OAR24_23721386.1	21966370	A	–41.33	17.59	.000063	.080
9	OAR9_87184530.1	82402017	A	46.58	15.97	.000124	.080
18	s53361.1	31059794	G	74.83	15.73	.000139	.080
14	s16439.1	9866960	A	42.31	15.73	.000139	.080
5	s75037.1	67865996	A	–52.34	15.67	.000144	.080
8	OAR8_270360.1	274469	G	–45.73	15.27	.000173	.080
23	OAR23_37479093.1	35491925	G	64.77	15.23	.000174	.080
22	s42486.1	24605942	G	39.79	14.71	.000223	.081
3	OAR3_228560068.1	210915761	G	41.63	14.63	.000231	.081
23	s74857.1	30775199	G	56.37	14.12	.000292	.081
17	OAR17_29162292.1	26567600	G	–49.95	14.09	.000295	.081
13	OAR13_39835484.1	36523324	A	61.63	14.02	.000307	.081
8	OAR8_44440146_X.1	41273738	A	38.87	13.98	.000313	.081
12	s75315.1	51132322	A	–42.95	13.77	.000316	.081
17	OAR17_68309863.1	62688843	G	–47.52	13.75	.000345	.081
4	OAR4_21356609.1	20526954	G	48.53	13.71	.000348	.081
13	s40860.1	76542311	A	–50.10	13.70	.000355	.081
26	OAR26_9766087.1	7605814	G	45.91	13.47	.000356	.081
6	OAR6_80988051.1	74188622	A	–43.37	13.42	.000397	.081
3	s33966.1	218139413	A	–42.96	13.38	.000404	.081
24	s64989.1	15589004	G	–40.83	13.37	.000412	.081
4	OAR4_48113653.1	45393335	A	–37.89	13.24	.000414	.081
13	OAR13_68959441.1	63881639	G	–53.84	13.06	.000441	.083
8	s54499.1	84399113	A	–38.57	12.99	.000480	.085
4	s38487.1	13410025	A	38.37	12.89	.000499	.085
20	s67035.1	32151561	A	40.36	12.83	.000528	.085
27	OARX_11161852.1	4561008	G	36.57	12.70	.000533	.085
13	s21696.1	63867006	G	–46.37	12.62	.000569	.085

Table 4. Gene symbol, name and function, accession number and position in the gene of the markers of Table 3, ordered by chromosomes
n.a., no deposited mRNA sequence

OAR	Marker	Gene symbol	Gene name and function	Accession	Position in the gene
1	OAR1_278980576.1	PCP4	Purkinje cell protein 4 is involved in calcium deposition of and modulation of Ca modulation-dependent protein kinases	XM_004003899.1	29K bp downstream
3	OAR3_228560068.1	NRIP2	Nuclear receptor interacting protein 2, transcript variant 3, downregulates transcriptional activation by nuclear receptors	XM_004006958.1	Intron 1
3	s33966.1	ITFG2	Integrin α FG-GAP repeat containing 2 encodes a cytoskeletal protein, member of integrin adhesion family	XM_004006959.1	6 K bp upstream
4	s38487.1	SLC25A13	Solute carrier family 25, member 13, provides instructions for making the citrin protein, involved in transporting molecules into and out of mitochondria; mutations have been shown to cause disruption in the urea cycle	XM_004022710.1	Intron 11
4	OAR4_21356609.1	LOC101117414	Messenger RNA60S ribosomal protein L14-like	XR_173156.1	866 bp upstream
4	OAR4_37455919.1	SEMA3D	Semaphorin 3D is involved in cell differentiation and encodes a protein belonging to the immunoglobulin domain	XM_004007803.1	5′ UTR
4	OAR4_37402301.1				547 bp upstream start codon
4	OAR4_48113653.1	ORC5	Origin recognition complex, subunit 5, is a highly conserved protein complex essential for the initiation of the DNA replication in eukaryotic cells	XM_004007832.1	Intron 13
5	s75037	EBF1	Transcription factor COE1 is a transcriptional activator, which recognises variations of specific palindromic sequences	XM_004009024.1	Intron 10
6	OAR6_80988051	LOC101118954	F-box/WD repeat-containing protein 7-like	n.a.	Within locus
8	OAR8_270360.1	CD109	CD109 antigen precursor represents a novel and independent branch of the α2-macroglobulin/complement gene family	XM_004011463.1	Intron 2
8	OAR8_44440146_X	LOC101112824	Small ubiquitin-related modifier 1-like	n.a.	Within locus
8	s54499	PARK2	Parkinson protein 2, E3 ubiquitin protein ligase, plays a role in the cell machinery that breaks down unneeded proteins by tagging damaged and excess proteins with ubiquitin	XM_004011554.1	Intron 5
9	OAR9_87184530.1	KIAA1429	KIAA1429 orthologue gene encodes a nuclear membrane protein involved in mRNA transport; the gene belonging to the vir family, may be involved in mRNA splicing regulation	XM_004011834.1	Intron 1
12	s75315	PRDM2	PR domain containing 2, with ZNF domain: may function as a DNA-binding transcription factor, and specifically as a transcriptional activator of the heme oxygenase 1 gene	XM_004014075.1	698 K upstream
13	OAR13_39835484	PC2	Prohormone convertase 2 precursor: is part of a family of convertases involved in proprotein processing	XM_004014283.1	165 K upstream
13	OAR13_68959441.1	TRPC4	Transient receptor potential cation channel subfamily C member 4 associated protein: forms a receptor-activated non-selective calcium permeant cation channel, and has been shown to be calcium-selective, and be activated by intracellular calcium store depletion	XM_004014518.1	Intron 4

(*continued next page*)

Table 4. (*continued*)

OAR	Marker	Gene symbol	Gene name and function	Accession	Position in the gene
13	s21696.1				Intron 9
13	s40860	ARFGEF2	ADP-ribosylation factor guanine nucleotide-exchange factor 2 is involved in the regulation of Golgi vesicular transport	XM_004014875.1	145 K upstream
13	OAR13_81680798.1	SULF2	Sulfatase 2: exhibits highly specific endoglucosamine-6-sulfatase activity	XM_004014874.1	66 k downstream
14	s16439	MBTPS1	Membrane-bound transcription factor peptidase catalyses the first step in the proteolytic activation of the sterol regulatory element-binding proteins	XM_004014937.1	Intron 7
17	OAR17_29162292	LOC101108425	Transforming acidic coiled-coil-containing protein 3-like: plays a role in transcriptional and chromatin remodelling events	n.a.	92 K bp upstream
17	OAR17_68309863	SPPL3	Signal peptide peptidase-like 3 encodes an intramembrane-cleaving aspartic protease	XM_004017424.1	4 K bp upstream start codon
18	s53361	RCN2	Reticulocalbin 2 binds calcium ions	XM_004023557.1	Intron 3
20	s67035	ALDH5A1	Aldehyde dehydrogenase 5 family, member A1, catalyses one step in the degradation of the inhibitory neurotransmitter gamma-aminobutyric acid	XM_004019334.1	1.2 K downstream
22	s42486	SORCS3	VPS10 domain-containing receptor SorCS3 functions as a sorting receptor in the Golgi compartment and as a clearance receptor on the cell surface	XM_004020175.1	Intron 2
23	OAR23_37479093	THOC1	THO complex 1 regulates transcriptional elongation of a subset of genes	XM_004020628.1	186 bp upstream
23	s74857	KCTD1	Potassium channel tetramerisation domain containing 1 may repress the transcriptional activity of AP-2 family members	XM_004020460.1	304 K downstream
24	s64989	LOC101111512	Xylosyltransferase 1-like enzyme catalyses the first step in biosynthesis of glycosaminoglycan	n.a.	100.6 K bp upstream
24	OAR24_23721386.1	SPTSSA	Serine palmitoyltransferase small subunit A-like catalyses the first committed step in sphingolipid biosynthesis	XM_004020843.1	2K bp upstream
26	OAR26_9766087	AGA	Aspartylglucosaminidase plays a role in the catabolism of N-linked oligosaccharides of glycoproteins	XM_004021838.1	173 K downstream
X	OARX_11161852	STS	Steroid sulfatase, microsomal, isozyme S is a membrane-bound microsomal enzyme that hydrolyses various alkyl and aryl steroid sulfates	n.a.	29 K upstream

performed to meet specific levels of confidence. To achieve the best possible conditions, the selective genotyping strategy was here applied, by selecting two groups of 50 positive and 50 negative sheep, out of 759 screened animals, representing the highest divergent animals for the target trait. On the sheep of the two divergent groups, we performed a GWAS with the Illumina SNP50K BeadChip with the purpose to detect genes that might have a role in disease resistance.

In order to discern whether the genes in proximity of the 32 markers were either directly or indirectly involved in Johne's disease, we examined the literature that included experimental trials performed with non-infected versus *Mycobacterium bovis*-infected cattle. These works are based on two different approaches, the first is a gene expression screenings with specific DNA microarrays containing immunity genes; the second is the identification of loci associated with tolerance to the disease through a GWAS. Because single genetic markers of susceptibility genes with small effects have limited predictive value, a genetic risk profile requires extensive knowledge of gene–gene interactions (Janssens *et al.* 2008); therefore, to expedite the search for genes associated with polygenic diseases, the complementary approaches of whole-genome scans and microarray gene profiling is used to identify and validate clusters of relevant genes (Gibbons *et al.* 2004). Coussens *et al.* (2005) followed the first approach and evaluated gene-expression profiles of peripheral blood mononuclear cells from two groups of Holstein cows, consisting of four animals naturally infected with MAP and four healthy uninfected controls from a commercial herd. These authors performed microarray hybridisations using a commercial microarray hybridisation station (BOTL-3 bovine cDNA microarrays, CAFG – Michigan State University)

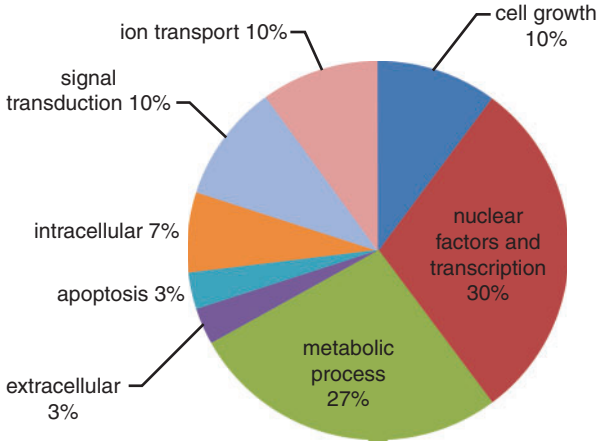

Fig. 2. Major ontological and functional classification of the genes in proximity of the markers significantly affecting paratuberculosis positivity (*FDR P* < 0.09).

containing expressed sequence tag clones representing genes known to function in immune response, and evaluated function and ontology of the differentially expressed genes; they showed that 40% of the genes could be classified into the following ontological and functional categories: cytokines and receptors, signal transduction, nuclear and transcription factors, cell cycle control, signal transduction, cytoskeletal; a further 30% of the differentially expressed genes were involved in apoptotic processes, cell growth and migration, adhesion, inhibitors and metabolism. Similarly to what was reported in the work by Coussens *et al*. (2005), in Fig. 2 the here detected genes were classified in a pie graph based on the gene ontology. The majority of the genes, also in the present study, fell in the category of the activator of transcription, the growth factor activators and the signal transduction (50%; Fig. 2); another important category (37%) included the genes involved in metabolic processes and ion transport.

Also Meade *et al*. (2007), Magee *et al*. (2012) and Nalpas *et al*. (2013) have performed gene expression screenings, with DNA microarrays containing immunity genes, on peripheral blood mononuclear cells from non-infected versus *M. bovis*-infected cattle. Meade *et al*. (2007) reported the lower relative expression of key innate immune genes in the infected animals. Magee *et al*. (2012) revealed a differential expression for genes involved in: (1) the inflammatory response; (2) cell signalling pathways, including TLR and intracellular pathogen recognition receptors; and (3) apoptosis. Nalpas *et al*. (2013) demonstrated the differential expression of immune, apoptotic and cell signalling genes in the non-infected versus infected animals. The results of the reported studies actually showed that *M. bovis* infection is associated with the repression of host immunity gene expression.

The second approach to identify loci associated with tolerance to the disease is the GWAS. Similarly to the present study, Zanella *et al*. (2011) performed a GWAS in Holstein cattle with the Illumina Bovine SNP50 BeadChip, and found that only one gene, the GNA12 gene (guanine nucleotide binding protein α 12) could be considered a positional candidate associated to tolerance to Johne's disease, so that the previous findings, emphasising the important role of CARD15/NOD2 and TLR4 genes, were not corroborated. Zare *et al*. (2014) performed a case-control GWAS to identify genomic regions underlying susceptibility to MAP infection in Jersey cattle, but using a less dense SNP panel (7K SNP). They reported that the significant markers were in proximity of the following candidate genes: major histocompatibility complex, TCF19, FAT10, HIVEP1, CCDC17, ZNF684, UBE2 L3, UBE2K, FAM5C, FAM109A. These genes included transcription regulators and ubiquitins. Kirkpatrick *et al*. (2011) performed a GWAS with the SNP50 BeadChip on MAP-infected Holstein cows and compared SNP allele frequencies with specific reference populations, their aim being the identification of a set of SNP that could be used as a predictor for Holstein cattle susceptibility to MAP infection. Of the genes detected though GWAS in humans and mice, inferred as potential candidates of Crohn's disease and colitis, Kirkpatrick *et al*. (2011) found that only the PTGER4 gene was in proximity of one of the SNP of the proposed set.

In the present study, 5 of the 30 detected genes had been previously reported to play a direct role in the immune system: SEMA3D, CD109, PCP4, PRDM2 and ITFG2. SEMA3D is a member of the semaphorins that have been demonstrated to be important for immune response by reducing the activation of T-cells through its cell-surface receptors, including members of the neurophilin and plexin families (Moretti *et al*. 2006). Emerging evidence also indicated that additional semaphorins and related molecules seem to function in the reciprocal stimulation of T-cells and antigen-presenting cells.

The CD109 gene, belonging to the complement gene family, has been studied in depth in humans, where the differential expression of the gene was reported for different human carcinoma, compared with the control (Hashimoto *et al*. 2004; Zhang *et al*. 2005). A direct effect of the marker in proximity of this gene, on the differential expression of the CD109 gene in resistant versus non-resistant sheep, might furthermore be inferred, falling marker OAR8_270360.1 in the binding site of the MZF1 transcription factor.

The PCP4 gene is involved in the activation of B-cells: in the murine B-cell co-receptor complex, Jacobson *et al*. (2009) demonstrated an increased expression of PCP4, in wild type mice, compared with complement-deficient animals; the same authors correlated the increased expression of PCP4 with B-cell maturation into end stage phenotypes.

The PRDM2 gene belongs to the positive regulatory domain I containing genes, with zinc fingers (PRDM) family, which has recently invoked considerable interest as it has been implicated in fundamental aspects of cellular differentiation and exhibits expanding ties to human diseases (Fog *et al*. 2012). It is interesting to note that the PRMD1 gene was considered by Minozzi *et al*. (2010) a strong functional candidate of MAP resistance; these authors have, in fact, performed a GWAS in Holstein cows, using the Illumina Bovine SNP50 BeadChip, and have identified SNP on chromosomes 12, 8, 9, 11, and 27 with significant association with MAP antibody response, but reported that the only gene that could be considered a strong functional candidate was PRMD1, a transcription repressor that acts on the β interferon gene expression and affects the maturation of B-lymphocytes into antibody secreting cells.

The ITFG2 gene, belonging to the integrin α family, was reported to be involved in cell-mediated responses by Zhuang *et al*. (2008).

With the exception of the PRDM members, proposed as potential candidate of MAP resistance in sheep, in the present study, and in cattle, by Minozzi *et al*. (2010), each GWAS has identified different genes. It is therefore evident that the complexity of the immune system cannot be summarised in few genes with major effect.

On the contrary, in this study it was evident that the genes responsible for basic cellular processes play a role in MAP resistance, corroborating the findings of Coussens *et al*. (2005), Meade *et al*. (2007) and Nalpas *et al*. (2013).

Conclusion

Most of the existing studies on immune response to MAP infection, both in humans and in livestock, used the candidate gene approach, which is based on a pre-selection of the genes that are hypothesised to influence natural resistance, and following analysis of these genes, either of the polymorphisms or the gene expression, in the positive and control groups. This approach precludes the detection of new genes, potentially involved in the immune response, simply because their role was not previously observed or hypothesised. For this reason, in the present study we aspired to identify regions associated with resistance to paratuberculosis in sheep. Although the present study should be considered preliminary, because the GWAS was performed on a limited number of individuals, it describes the first genome-wide characterisation of selective sweeps in sheep addressing disease resistance. Although animal health has been identified as a priority, breeding for disease resistance is still a scarcely explored area and there is a clear need of research, particularly for sheep.

In this study, 32 genomic regions were made evident for encoding polymorphic markers with significant effect on the S/P value indicating paratuberculosis positivity. These regions harboured already described ovine genes; for five of these genes, a direct involvement in the immune response was reported in the literature.

The explorative value of this study might be corroborated by using the obtained results, on one side, to deeply analyse each gene, so to detect the polymorphism that alter gene expression, likely being in LD with the marker. On the other side, specific quantitative real-time PCR trials of gene expression for these genes could be designed on experimental resistant versus non-resistant populations, so to find markers to be used in the selection for disease resistance.

Acknowledgement

This study is part of the GENZOOT research program, funded by the Italian Ministry of Agriculture (Rome, Italy).

References

Alshatwi AA, Hasan TN, Syed NA, Shafi G, Grace L (2012) Identification of functional SNPs in *BARD1* gene and *In Silico* analysis of damaging SNPs: based on data procured from dbSNP database. *PLoS ONE* **7**, e43939. doi:10.1371/journal.pone.0043939

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *Journal of Molecular Biology* **215**, 403–410. doi:10.1016/S0022-2836(05)80360-2

Balfour Sartor R (2005) Does *Mycobacterium avium* subspecies *paratuberculosis* cause Crohn's disease? *Gut* **54**, 896–898. doi:10.1136/gut.2004.055889

Bhide MR, Cepkova M, Mikula I Jr, Kisova L, Skrabana R, Novak M, Mikula I Sr (2009) Novel mutations in TLR genes cause hyporesponsiveness to *Mycobacterium avium* subsp. *Paratuberculosis* infection. *BMC Genetics* **10**, 21. doi:10.1186/1471-2156-10-21

Ciani E, Crepaldi P, Nicoloso L, Lasagna E, Sarti FM, Moioli B, Napolitano F, Carta A, Usai G, D'Andrea M, Marletta D, Ciampolini R, Riggio V, Occidente M, Matassino D, Kompan D, Modesto P, Macciotta N, Ajmone-Marsan P, Pilla F (2014) Genome-wide analysis of Italian sheep diversity reveals a strong geographic pattern and cryptic relationships between breeds. *Animal Genetics* **45**, 256–266. doi:10.1111/age.12106

Collins MT (2002) Interpretation of a commercial bovine paratuberculosis enzyme-linked immunosorbent assay by using likelihood ratios. *Clinical and Diagnostic Laboratory Immunology* **9**, 1367–1371.

Coussens PM, Pudrith CB, Skovgaard K, Ren X, Suchyta SP, Stabel JR, Heegaard PM (2005) Johne's disease in cattle is associated with enhanced expression of genes encoding IL-5, GATA-3, tissue inhibitors of matrix metalloproteinases 1 and 2, and factors promoting apoptosis in peripheral blood mononuclear cells. *Veterinary Immunology and Immunopathology* **105**, 221–234. doi:10.1016/j.vetimm.2005.02.009

Doss CG, Sethumadhavan R (2009) Investigation on the role of nsSNPs in HNPCC genes – a bioinformatics approach. *Journal of Biomedical Science* **16**, 42. doi:10.1186/1423-0127-16-42

Fog CK, Galli GG, Lund AH (2012) PRDM proteins: important players in differentiation and disease. *BioEssays* **34**, 50–60. doi:10.1002/bies.201100107

Gibbons GH, Liew CC, Goodarzi MO, Rotter JI, Hsueh WA, Siragy HM, Pratt R, Dzau VJ (2004) Genetic markers, progress and potential for cardiovascular disease. *Circulation* **109**, IV-47–IV-58. doi:10.1161/01.CIR.0000133440.86427.26

Goddard ME, Hayes BJ (2009) Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nature Reviews. Genetics* **10**, 381–391. doi:10.1038/nrg2575

Gumber S, Eamens G, Whittington RJ (2006) Evaluation of a Pourquier ELISA kit in relation to agar gel immunodiffusion (AGID) test for assessment of the humoral immune response in sheep and goats with and without *Mycobacterium paratuberculosis* infection. *Veterinary Microbiology* **115**, 91–101. doi:10.1016/j.vetmic.2006.01.003

Hashimoto M, Ichihara M, Watanabe T, Kawai K, Koshikawa K, Yuasa N, Takahashi T, Yatabe Y, Murakumo Y, Zhang JM, Nimura Y, Takahashi M (2004) Expression of CD109 in human cancer. *Oncogene* **23**, 3716–3720. doi:10.1038/sj.onc.1207418

Hermon-Taylor J (2009) *Mycobacterium avium* subspecies *paratuberculosis*, Crohn's disease and the Doomsday Scenario. *Gut Pathogens* **1**, 15. doi:10.1186/1757-4749-1-15

Hromas R, Collins SJ, Hickstein D, Raskind W, Deaven LL, O'Hara P, Hagen FS, Kaushansky K (1991) A retinoic acid-responsive human zinc finger gene, MZF-1, preferentially expressed in myeloid cells. *The Journal of Biological Chemistry* **278**, 14183–14187.

Hugot JP (2006) CARD15 mutations in Crohn's disease. *Annals of the New York Academy of Sciences* **1072**, 9–18. doi:10.1196/annals.1326.011

Hugot JP, Chamaillard M, Zouali H, Lesage S, Cézard JP, Belaiche J, Almer S, Tysk C, O'Morain CA, Gassull M, Binder V, Finkel Y, Cortot A, Modigliani R, Laurent-Puig P, Gower-Rousseau C, Macry J, Colombel JF, Sahbatou M, Thomas G (2001) Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature* **411**, 599–603. doi:10.1038/35079107

Jacobson A, Weis JJ, Weis JH (2009) CD21 signaling via C3 regulates purkinje cell protein 4 expression. *Molecular Immunology* **46**, 1488–1493. doi:10.1016/j.molimm.2008.12.030

Janssens AC, Gwinn M, Bradley LA, Oostra BA, van Duijn CM, Khoury MJ (2008) A critical appraisal of the scientific basis of commercial genomic profiles used to assess health risks and personalize health interventions. *American Journal of Human Genetics* **82**, 593–599. doi:10.1016/j.ajhg.2007.12.020

Kaisho T, Akira S (2006) Toll-like receptor function and signaling. *The Journal of Allergy and Clinical Immunology* **117**, 979–987. doi:10.1016/j.jaci.2006.02.023

Khatkar M, Zenger KR, Hobbas M, Hawken R, Cavanagh JA, Barris W, McClintock A, McClintock S, Thomson P, Tier B, Nicholas F, Raadsma HW (2007) A primary assembly of a bovine haplotype block map based on a 15,036-single-nucleotide polymorphism panel genotyped in Holstein–Friesian cattle. *Genetics* **176**, 763–772. doi:10.1534/genetics.106.069369

Kijas JW, Lenstra JA, Hayes B, Boitard S, Porto Neto LR, San Cristobal M, Servin B, McCulloch R, Whan V, Gietzen K, Paiva S, Barendse W, Ciani E, Raadsma H, McEwan J, Dalrymple B (2012) Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. *PLoS Biology* **10**, e1001258. doi:10.1371/journal.pbio.1001258

Kirkpatrick BW, Shi X, Shook GE, Collins MT (2011) Whole-genome association analysis of susceptibility to paratuberculosis in Holstein cattle. *Animal Genetics* **42**, 149–160. doi:10.1111/j.1365-2052.2010.02097.x

Magee DA, Taraktsoglou M, Killick KE, Nalpas NC, Browne JA, Park SDE, Conlon K, Lynn DJ, Hokamp K, Gordon SV, Gormley E, MacHugh DE (2012) Global gene expression and systems biology analysis of bovine monocyte-derived macrophages in response to in vitro challenge with *Mycobacterium bovis*. *PLoS ONE* **7**, e32034. doi:10.1371/journal.pone.0032034

Meade KG, Gormley E, Doyle MB, Fitzsimons T, O'Farrelly C, Costello E, Keane J, Zhao Y, MacHugh DE (2007) Innate gene repression associated with *Mycobacterium bovis* infection in cattle: toward a gene signature of disease. *BMC Genomics* **8**, 400. doi:10.1186/1471-2164-8-400

Minozzi GL, Buggiotti L, Stella A, Strozzi F, Luini M, Williams JL (2010) Genetic loci involved in antibody response to *Mycobacterium avium* ssp. *Paratuberculosis* in cattle. *PLoS ONE* **5**, e11117. doi:10.1371/journal.pone.0011117

Moioli B, Scatà MC, Steri R, Napolitano F, Catillo G (2013) Signatures of selection identify loci associated with milk yield in sheep. *BMC Genetics* **14**, 76. doi:10.1186/1471-2156-14-76

Moradi MH, Nejati-Javaremi A, Moradi-Shahrbabak M, Dodds K, McEwan JC (2012) Genomic scan of selective sweeps in thin and fat tail sheep beeds for identifying of candidate regions associated with fat deposition. *BMC Genetics* **13**, 10. doi:10.1186/1471-2156-13-10

Moretti S, Procopio A, Boemi M, Catalano A (2006) Neuronal semaphorins regulate a primary immune response. *Current Neurovascular Research* **3**, 295–305. doi:10.2174/156720206778792939

Mucha R, Bhide MR, Chakurkar EB, Novak M, Mikula I Sr (2009) Toll-like receptors TLR1,TLR2 and TLR4 gene mutations and natural resistance to *Mycobacterium avium* subsp. *paratuberculosis* infection in cattle. *Veterinary Immunology and Immunopathology* **128**, 381–388. doi:10.1016/j.vetimm.2008.12.007

Nalpas NC, Park SD, Magee DA, Taraktsoglou M, Browne JA, Conlon K, Killick K, Hokamp K, Lohan AJ, Loftus BJ, Gormley E, Gordon SV, MacHugh DE (2013) Whole-transcriptome, high-throughput RNA sequence analysis of the bovine macrophage response to *Mycobacterium bovis* infection in vitro. *BMC Genomics* **14**, 230. doi:10.1186/1471-2164-14-230

Nalubamba K, Smeed J, Gossner A, Watkins C, Dalziel R, Hopkins J (2008) Differential expression of pattern recognition receptors in the three pathological forms of sheep paratuberculosis. *Microbes and Infection* **10**, 598–604. doi:10.1016/j.micinf.2008.02.005

Nielsen SS, Toft N (2008) Ante mortem diagnosis of paratuberculosis: a review of accuracies of ELISA, interferon-gamma assay and faecal culture techniques. *Veterinary Microbiology* **129**, 217–235. doi:10.1016/j.vetmic.2007.12.011

Odani M, Narita A, Watanabe T, Yokouchi K, Sugimoto Y, Fujita T, Oguni T, Matsumoto M, Sasaki Y (2006) Genome-wide linkage disequilibrium in two Japanese beef cattle breeds. *Animal Genetics* **37**, 139–144. doi:10.1111/j.1365-2052.2005.01400.x

Olsen I, Tollefsen S, Aagaard C, Reitan LJ, Bannantine JP, Andersen P, Sollid LM, Lundin KE (2009) Isolation of *Mycobacterium avium* subspecies *paratuberculosis* reactive CD4 T cells from intestinal biopsies of Crohn's disease patients. *PLoS ONE* **4**, e5641. doi:10.1371/journal.pone.0005641

Pant SD, Schenkel FS, Verschoor CP, You Q, Kelton DF, Moore SS, Karrow NA (2010) A principal component regression based genome wide analysis approach reveals the presence of a novel QTL on BTA7 for MAP resistance in Holstein cattle. *Genomics* **95**, 176–182. doi:10.1016/j.ygeno.2010.01.001

Pinedo PJ, Buergelt CD, Donovan GA, Melendez P, Morel L, Wu R, Langaee TY, Rae DO (2009) Association between CARD15/NOD2 gene polymorphisms and paratuberculosis infection in cattle. *Veterinary Microbiology* **134**, 346–352. doi:10.1016/j.vetmic.2008.09.052

Plain KM, Purdie AC, Begg DJ, de Silva K, Whittington RJ (2010) Toll-like receptor (TLR)6 and TLR1 differentiation in gene expression studies of Johne's disease. *Veterinary Immunology and Immunopathology* **137**, 142–148. doi:10.1016/j.vetimm.2010.04.002

Reddacliff L, Eppleston J, Windsor P, Whittington R, Jones S (2006) Efficacy of a killed vaccine for the control of paratuberculosis in Australian sheep flocks. *Veterinary Microbiology* **115**, 77–90. doi:10.1016/j.vetmic.2005.12.021

Reich DE, Cargill M, Bolk S, Ireland J, Sabeti PC, Richter DJ, Lavery T, Kouyoumjian R, Farhadian SF, Ward R, Lander ES (2001) Linkage disequilibrium in the human genome. *Nature* **411**, 199–204.

SAS Institute (2007) 'SAS/STAT user's guide. Version 9.1.' (SAS Institute Inc.: Cary, NC)

Settles M, Zanella R, McKay SD, Schnabel RD, Taylor JF, Whitlock R, Schukken Y, VanKessel JS, Smith JM, Neibergs H (2009) A whole genome association analysis identifies loci associated with *Mycobacterium avium* subsp. *Paratuberculosis* infection status in US Holstein cattle. *Animal Genetics* **40**, 655–662. doi:10.1111/j.1365-2052.2009.01896.x

Sezzi E, Lillini E, De Sanctis B, Scorsino G, Valentini A, Pariset L, De Grossi L (2007) In Proceedings of the 9th International Colloquium on Paratuberculosis. 29 October – 2 November 2007, Tsukuba, Japan. (Ed. SS Nielsen) pp. 94–97. (International Association for Paratuberculosis: Derio, Bizkaia, Spain)

Taylor DL, Zhong L, Begg DJ, de Silva K, Whittington RJ (2008) Toll-like receptor genes are differentially expressed at the sites of infection during the progression of Johne's disease in outbred sheep. *Veterinary Immunology and Immunopathology* **124**, 132–151. doi:10.1016/j.vetimm.2008.02.021

Vasselon T, Detmers PA (2002) Toll receptors: a central element in innate immune responses. *Infection and Immunity* **70**, 1033–1041. doi:10.1128/IAI.70.3.1033-1041.2002

Wientjes Y, Veerkamp RF, Calus MP (2013) The effect of linkage disequilibrium and family relationships on the reliability of genomic prediction. *Genetics* **193**, 621–631. doi:10.1534/genetics.112.146290

Yang TT, Chow CW (2003) Transcription cooperation by NFAT.C/EBP composite enhancer complex. *The Journal of Biological Chemistry* **278**, 15 874–15 885. doi:10.1074/jbc.M211560200

Zanella R, Settles ML, McKay SD, Schnabel R, Taylor J, Whitlock RH, Schukken Y, Van Kessel JS, Smith JM, Neibergs HL (2011) Identification

of loci associated with tolerance to Johne's disease in Holstein cattle. *Animal Genetics* **42**, 28–38. doi:10.1111/j.1365-2052.2010.02076.x

Zare Y, Shook GE, Collins MT, Kirkpatrick BW (2014) Genome-wide association analysis and genomic prediction of *Mycobacterium avium* subspecies *paratuberculosis* infection in US Jersey cattle. *PLoS ONE* **9**, e88380. doi:10.1371/journal.pone.0088380

Zhang JM, Hashimoto M, Kawai K, Murakumo Y, Sato T, Ichihara M, Nakamura S, Takahashi M (2005) CD109 expression in squamous cell carcinoma of the uterine cervix. *Pathology International* **55**, 165–169. doi:10.1111/j.1440-1827.2005.01807.x

Zhuang S, Kelo L, Nardi JB, Kanost MR (2008) Multiple alpha subunits of integrin are involved in cell-mediated responses of the Manduca immune system. *Developmental and Comparative Immunology* **32**, 365–379. doi:10.1016/j.dci.2007.07.007