CORA Cork Open Research Archive
Cartlann Taighde Oscailte Chorcaí

Title	Candidate genes associated with the heritable humoral response to Mycobacterium avium subspecies paratuberculosis in dairy cows have factors in common with gastrointestinal diseases in humans
Authors	McGovern, S. P.;Purfield, D. C.;Ring, S. C.;Carthy, T. R.;Graham, D. A.;Berry, Donagh P.
Publication date	2019-03-07
Original Citation	McGovern, S. P., Purfield, D. C., Ring, S. C., Carthy, T. R., Graham, D. A. and Berry, D. P. (2019) 'Candidate genes associated with the heritable humoral response to Mycobacterium avium subspecies paratuberculosis in dairy cows have factors in common with gastrointestinal diseases in humans', Journal of Dairy Science. doi: 10.3168/jds.2018-15906
Type of publication	Article (peer-reviewed)
Link to publisher's version	10.3168/jds.2018-15906
Rights	© 2019, American Dairy Science Association. Published by Elsevier B.V. All rights reserved. - https://creativecommons.org/licenses/by-nc-nd/4.0/
Download date	2026-11-02 22:16:03
Item downloaded from	https://hdl.handle.net/10468/7634

UCC
University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

Candidate genes associated with the heritable humoral response to Mycobacterium avium subspecies paratuberculosis in dairy cows have factors in common with gastrointestinal diseases in humansBy McGovern et al. XXXX. Paratuberculosis in cattle, also commonly known as Johne's disease causes serious performance, and by extension, economic losses on farms. Its tentative links with Crohn's disease in humans is also of concern. The present study revealed that breeding for differences in humoral response to paratuberculosis in cattle is possible should routine access to test results be available and also that the underlying genetic variants contributing to this genetic variation have indeed factors in common with gastrointestinal diseases in humans

RUNNING HEADING: GENETICS AND GENOMICS OF PARATUBERCULOSIS IN CATTLE

Candidate genes associated with the heritable humoral response to *Mycobacterium avium subspecies paratuberculosis* in dairy cows have factors in common with gastrointestinal diseases in humans

S. P. McGovern[*], D. C. Purfield[‡], S. C. Ring[†], T. R. Carthy[‡], D. A. Graham[§], and D.P. Berry[‡1]

[*]University College Cork, Coláiste na hOllscoile Corcaigh, College Road, Cork City, Co. Cork, Ireland.

[†]Irish Cattle Breeding Federation, Highfield House, Shinagh, Bandon, Co. Cork, Ireland, P72 X050.

[§]Animal Health Ireland, 4-5 The Archways, Carrick-on-Shannon, Co. Leitrim, Ireland, N41WN27.

[‡]Teagasc, Animal & Grassland Research and Innovation Centre, Moorepark, Fermoy, Co. Cork, Ireland, P61C996.

[1]**Corresponding author:** Donagh Berry. Donagh.Berry@teagasc.ie

ABSTRACT

Infection of cattle with bovine paratuberculosis (i.e., Johne's disease) is caused by *Mycobacterium avium* subspecies *paratuberculosis* (MAP) and results in a chronic incurable gastroenteritis. This disease, which has economic ramifications for the cattle industry, is increasing in detected prevalence globally; subclinically infected animals can silently shed the bacterium into the environment for years, exposing contemporaries and hampering disease control programs. The objective of the present study was to firstly quantify the genetic parameters for humoral response to MAP in dairy cattle. This was followed by a genome-based association analysis and subsequent downstream bioinformatic analyses from imputed whole genome sequence single nucleotide polymorphism (SNP) data. After edits, ELISA test records were available on 136,767 cows; analyses were also undertaken on a subset of 33,818 of these animals from herds with at least 5 MAP ELISA positive cows, with at least 1 of those positive cows being homebred. Variance components were estimated using univariate animal and sire linear mixed models. The heritability calculated from the animal model for humoral response to MAP using alternative phenotype definitions varied from 0.02 (SE=0.003) to 0.05 (SE=0.008). The genome-based associations were undertaken within a mixed model framework using weighted deregressed estimated breeding values as a dependent variable on 1,883 phenotyped animals that were ≥87.5% Holstein-Friesian. Putative susceptibility quantitative trait loci (QTLs) were identified on BTA 1, 3, 5, 6, 8, 9, 10, 11, 13, 14, 18, 21, 23, 25, 26, 27 and 29; mapping the most significant SNPs to genes within and overlapping these QTLs revealed that the most significant associations were with the 10 functional candidate genes *KALRN*, *ZBTB20*, *LPP*, *SLA2*, *FI3A1*, *LRCH3*, *DNAJC6*, *ZDHHC14*, *SNX1* and *HAS2*. Pathway analysis failed to reveal significantly enriched biological pathways, when both bovine-specific pathway data and human ortholog data were taken in to account. The existence of genetic variation for MAP susceptibility in a large dataset of dairy cows signifies the potential of breeding programs for reducing MAP susceptibility. Furthermore, the identification of susceptible QTLs facilitates greater biological understanding of bovine paratuberculosis and potential therapeutic targets for future investigation. The novel molecular similarities identified between bovine paratuberculosis and human inflammatory bowel disease suggest potential for human therapeutic interventions to be translated to veterinary medicine, and vice versa.

Keywords: Johne's disease, resistance, QTL, GWAS, sequence

INTRODUCTION

Paratuberculosis, also known as Johne's disease, caused by the gram-positive aerobic bacterium *Mycobacterium avium* subspecies *paratuberculosis* (MAP), is a contagious disease primarily affecting ruminants. Paratuberculosis results in chronic, progressive gastroenteritis for which there is no cure (Harris and Barletta, 2001). First reported in Europe by Johne and Frothingham (1895) as a "peculiar case of tuberculosis in cattle", MAP is primarily spread via the fecal-oral route; younger animals are most susceptible to clinical MAP infection upon exposure (Windsor and Whittington, 2009). Clinical signs of MAP infection in cattle, primarily observed in older cattle (MAP has an incubation period of up to 10 years; Collins, 2003) include weight loss due to characteristic pipestream diarrhea, hypoproteinemia (Sweeney et al., 2012), reduced milk yield (Richardson and More, 2009) and reduced cull cow value (Richardson and More, 2009), all of which adversely impact both animal well-being and farm profitability. Infected animals silently shed MAP in their environment via contaminated faeces, spreading paratuberculosis to uninfected animals (Koets et al., 2015).

Testing for MAP infection, concurrent with management (including measures to address both biocontainment and bioexclusion at herd level) and vaccination strategies, are the advocated control measures used to curtail the spread of bovine paratuberculosis. Unfortunately, all of these have limited efficacy and will not guarantee eradication, even at herd level, despite continued effort over an extended period. Correctly classifying animals as infected with paratuberculosis is challenging due to the available tests being suboptimal in both sensitivity and specificity; moreover, vaccination does not result in a reduction in the number of infected animals or offer long-term immunity (Park and Yoo, 2016, Geraghty *et al.*, 2014). In addition, vaccination against MAP is prohibited in some countries (e.g., the Republic of Ireland). Previous studies on the contribution of genetic variability to paratuberculosis in dairy cattle suggest heritability estimates of susceptibility to the disease ranging from <0.01 (Koets et al., 1999) to 0.283 (Küpper *et al.*, 2012). Differences in parameter estimates among studies could be due to a multitude of reasons, including the extent of variability (residual and genetic) in the populations sampled, sample sizes, trait definition, and models used (i.e., linear, threshold, animal, sire). Nonetheless, the non-zero heritability estimates of MAP susceptibility, coupled with the existence of considerable genetic variability, suggest that genetic selection could improve resistance to MAP infection in the bovine population; this could prove useful in the control and eradication of the disease concomitant with other control measures already in place.

Prior studies based on genome-wide associations (Settles et al., 2009, Pant et al., 2010, Alpay et al., 2014) have identified multiple different loci putatively associated with bovine paratuberculosis. The degree of concordance in reported quantitative trait loci (QTL) across studies is, however, poor. Only one study exists that used imputed whole genome SNP data to detect loci associated with MAP infection in cattle, although the study was based on relatively small cohorts of up to 459 dairy cattle (Kiser et al., 2017). Kiser et al. (2017) identified loci on chromosomes 3, 8, 10, 12, 14, 16, 21 and 22 associated with paratuberculosis susceptibility. Three studies exist that have used a gene-set enrichment analysis approach following a genome-based association analysis to investigate modest-effect SNPs and genes, enriched pathways and gene ontologies associated with MAP infection in cattle (Neibergs et al., 2010; Del Corvo et al., 2017; Kiser et al., 2017).

The objective of the present study was to quantify the genetic parameters of the humoral response to MAP infection in a large cohort of Irish dairy cows; the derived parameters were subsequently used to estimate breeding values as an input variable for an association analyses using imputed whole genome single nucleotide polymorphism (SNP) data to detect regions of the bovine genome putatively associated with humoral response to MAP. The biology underpinning the detected regions was further investigated by conducting bioinformatics analyses to identify the underlying gene functions and related biochemical pathways.

MATERIALS AND METHODS

The data used in the present study were obtained from a pre-existing database managed by the Irish Cattle Breeding Federation (ICBF). Therefore, it was not necessary to obtain animal care and use committee approval in advance of conducting this study.

Data

A total of 663,719 enzyme-linked immunosorbent assay (ELISA) test records on humoral (blood and milk antibody) response to MAP were available from 9 Irish laboratories between June 2012 and November 2017 inclusive on 282,396 cows in 2,704 dairy and beef herds. The health status of these herds for other diseases was unknown. Only cows aged between 2 and 12 years at the time of testing were considered. Records classified as “inconclusive” (n = 3,639), “suspect” (n = 1,406) or “low positive” (n = 4,210) were not considered further; subsequently, all individual cow test records were classified as either “positive” (13,685 records) or

"negative" (640,779 records) to MAP infection based on the respective test manufacturer's guidelines (Bovine ELISA Paratuberculosis Antibody Screening Kit (Institut Pourquier, France), ID Screen Paratuberculosis Indirect Screening Test (ID Vet, Montpellier, France, *Mycobacterium paratuberculosis* Antibody Test Kit PARACHEK (Prionics, Zurich, Switzerland) and Paratuberculosis Antibody Screening Test (Idexx Laboratories, Westbrook, ME, USA)).

Of the 10,137 cows with at least 1 positive ELISA record, 2,651 also had a negative MAP result following a positive MAP result; these cows were discarded from the dataset. Only data from dairy herds were considered further. A herd was classified as beef or dairy based on the average breed composition of the cows, as per Twomey et al. (2016). Cow breed composition was determined from the recorded breed composition of ancestors. A dairy herd was defined as a herd whose average dairy breed proportion of the cows was ≥75%. Only herd-years with ≥25 cows tested were retained. Following these edits, 612,375 test records from 260,740 cows in 4,543 herd-years from 2,170 herds remained.

Herd-years defined as MAP naïve were those which only had MAP negative cattle residing in them; these herd-years were not considered further, leaving a total of 378,701 records from 186,174 cows in 2,485 herd-years (1,487 herds). The remaining herd-years were considered exposed to MAP infection, as they had at least one cow that yielded an ELISA positive result. To be retained, all cows must have calved at least once in the herd prior to being tested in the positive herd-year. All cow inter-location movement data were available from the ICBF's database as it is a legal requirement to record these movements in Ireland. A cow was considered exposed to MAP if she resided in an exposed herd-year for at least one year prior to an ELISA test. This was to allow MAP-negative cows adequate time to become exposed to MAP via infected contemporary(ies).

Cow parity at test was categorized as 1, 2, 3, 4, or ≥5. Stage of lactation (i.e., days in milk) at ELISA test was categorized into 9 categories: 10 – 49, 50 – 99, ….. , 400 – 450; cows <10 or >450 days in milk at test were not considered further. Cows with no sire information were discarded and only the most recent test result per cow was considered further. Following these edits, 155,072 cows remained. General heterosis and recombination loss coefficients for each animal were calculated as $1 - \sum_{i=1}^{n} \text{sire}_i \times \text{dam}_i$ and $1 - \sum_{i=1}^{n} \frac{\text{sire}_i^2 \times \text{dam}_i^2}{2}$, respectively,

where $sire_i$ and dam_i are the proportion of breed *i* in the sire and dam, respectively (VanRaden and Sanders, 2003).

Contemporary groups were defined as herd-year-season of test using an algorithm described in detail by Berry and Evans (2014). The algorithm clusters herd-contemporaries that were tested around the same period of the year together but within no more than 90 days of each other. Only contemporaries groups with at least 5 cows and which had at least one positive and one negative MAP test result were retained. After edits, the final dataset comprised of 136,767 cows from 2,463 herd-years with test records reported from 8 different laboratories.

Genetic parameter estimation and genetic evaluation

Variance components for humoral response to MAP were estimated using animal and sire linear mixed models in ASReml (Gilmour et al., 2009). The fitted linear mixed model was:

$$Y = CG + heterosis + recombination + parity \times stage\ of\ lactation + a + e$$

where *Y* is the binary dependent variable of the MAP phenotype, *CG* is the fixed effect of contemporary group, *heterosis* is the fixed effect of a general heterosis coefficient (0.0%, >0.0 to <0.1%, ≥0.1 to <0.2%, …, ≥0.9 to <100%, 100%), *recombination* is the fixed effect of a general recombination loss coefficient (0.00%, >0.00 to <0.05%, ≥0.05 to <0.10%, …, ≥0.45 to <0.50%, 0.50%, >0.50%), *parity* is the fixed effect of the parity of the cow, *stage of lactation* is the fixed effect of stage of lactation, *a* is the random additive genetic effect of the animal where $a \sim N(0, \mathbf{A}\ \sigma_a^2)$ with σ_a^2 representing the additive genetic variance of the animal and **A** as the additive genetic relationship matrix among animals, and *e* is the random residual effect where $e \sim N(0, \mathbf{I}\ \sigma_e^2)$ with σ_e^2 representing the residual variance, and **I** representing the identity matrix. Each cow's pedigree was traced back to the founder population, which was assigned to one of 10 genetic groups. Sire models used were as described above except that the direct animal genetic effect was replaced by a sire genetic effect.

Genetic parameters were calculated for the overall dataset of 136,767 cows but also within each of the 8 laboratories separately. Genetic parameters were also estimated for humoral response to MAP where the data was restricted to 33,818 cows that resided in herd-years that contained at least 5 MAP ELISA positive cows, with at least 1 of those positive cows being homebred. The observed binary-scale heritability estimates were transformed to the

underlying liability scale via the formula of Robertson and Lerner (1949), using the average MAP prevalence of the respective cohort:

$$h_L^2 = h_O^2 \left[\frac{p(1-p)}{z^2}\right]$$

where h_L^2 = heritability on the liability scale (i.e., threshold model), h_O^2 = heritability on the observed scale, p = the trait prevalence/100, and z^2 = the height of the ordinate of normal distribution corresponding to a truncation point applied to p.

Two genetic evaluations were conducted for the present study, the first for validating estimated breeding values (EBVs) and the second for subsequent use in the genome-wide association analyses. Estimated breeding values and their corresponding reliabilities for MAP susceptibility were calculated using the MiX99 software suite (Stranden and Lidauer, 1999) with the fitted animal model being the same as previously described. EBVs were estimated for the set of 33,818 phenotyped cows and their relatives that resided in herd-years that contained at least 5 MAP positive cows, with at least 1 of those positive cows being homebred.

Validation of EBVs

For the purposes of validation of the genetic evaluations, the MAP phenotype of a sub-cohort of the 33,818 cows was masked and their breeding values estimated via their pedigree links with related phenotyped animals. The validation dataset was chosen based on herd-year characteristics in that only herd-years with at least 20 phenotyped animals and a mean herd incidence of between 10 and 25% were retained. In total, 6,600 animals from 94 herds were considered in the validation; the mean prevalence of ELISA positivity in this cohort was 15%. Once the MAP phenotypes were masked, a genetic evaluation was undertaken using the phenotypes of all remaining 27,218 animals. The EBVs of the validation animals with an estimated reliability of >0.05 were retained; 5,477 animals remained. Within herd-year, the validation animals were stratified equally (where the modular of 3 per herd was 0, otherwise as close to equal as possible) into high EBV (poor), average EBV and low EBV (good) groups. Logistic regression was used to model the association between EBV stratum (n=3) and the logit of the probability of a positive MAP outcome; a binomial distribution of errors was assumed. The area under the receiver operating curve (ROC) was estimated and the odds of a positive outcome calculated from the model solutions; the low EBV group was used as the referent category. A further logistic regression analysis was undertaken where the fixed effects model

terms of herd-year, heterosis, recombination and a two-way interaction between parity and stage of lactation were also fitted.

Whole-genome sequence imputation and association analysis

Of the 433,989 animals with an EBV from the genetic evaluation (with no masked phenotypes), only genotyped animals ≥87.5% Holstein-Friesian were considered for the genome-wide analysis; principal component analysis of the genotypes (along with animals from other breeds) was used to ensure all animals were Holstein-Friesian. A total of 8,780 animals remained. The EBVs estimated for the Holstein-Friesian animals were deregressed using the Secant method in MiX99 (Strandén and Mäntysaari, 2010), and subsequently, effective record contributions (ERCs) were calculated for each animal. Only the subset of 1,883 genotyped animals (662 male, 1,221 female) that had an ERC value ≥1 were considered further. These animals were initially genotyped using one of 7 Illumina arrays, namely 3k (2,909 SNPs, n = 36), LD (7,931 SNPs, n = 230), Bovine SNP50 (54,001 SNPs, n = 300), International Dairy and Beef version 1 (IDBv1) (17,137 SNPs, n = 121), IDBv2 (18,004 SNPs, n = 677), IDBv3 (53,450 SNPs, n = 236) and HD (777,962, n = 283). All animals had a call rate ≥ 90% and only autosomal SNPs, SNPs with a reported position as on UMDv3.1, and SNPs with a call rate ≥ 90% were retained within each panel.

All animals were imputed to HD using a two-step approach with FImpute2 software (Sargolzaei et al., 2014); this involved imputing the IDB, LD and 3k genotyped animals to the Bovine50 beadchip density (i.e., 54,001 SNPs) and consequently imputing all resulting genotypes (including the Bovine50 Beadchip genotypes) to HD using a multi-breed reference population of 5,504 HD-genotyped animals. The genotypes of all animals were imputed to whole-genome sequence level using a reference population of 2,333 *Bos taurus* animals (using multiple breeds) from Run6.0 of the 1,000 Bulls Genomes Project (Daetwyler et al., 2014). A consensus SNP density across all animals was achieved using SAMtools version 1.3.1 (Li, 2011), followed by Beagle software version 4.1 imputation (Browning and Browning, 2016) to call variants in the reference population and improve genotype calls. Details surrounding the alignment to UMDv3.1, variant calling and quality controls conducted on the reference population are described by (Daetwyler et al., 2014). A total of 41.39 million SNP variants were called with an average coverage of 12.85X. The imputation procedure was completed by

initially using Eagle v2.3.2 (Loh et al., 2016) to phase imputed HD genotypes, followed by imputation to whole-genome sequence level using minimac3 (Das et al., 2016).

Regions of poor WGS imputation accuracy perhaps due to local mis-assemblies or mis-orientated contigs, were identified using an additional dataset of 147,309 verified parent-progeny relationships. Mendelian errors, defined as the proportion of opposing homozygotes in a parent-progeny pair, were estimated for each relationship and the subsequent Mendelian error rate per SNP was determined. To accurately identify genomic regions of poor imputation, the R package GenWin (Beissinger et al., 2015) which fits a β-spline to the data to find likely inflection points, was used to determine genomic region breakpoints of high Mendelian errors. Windows were analysed using an initial window size of 5 kb and Genwin pooled windows for which the SNP Mendelian error rate were similar. The average SNP Mendelian error rate per window was estimated and all variants within windows where the mean SNP Mendelian error rate was >0.02 were removed (687,137 SNPs were removed).

A genomic relationship matrix, using just the autosomal high density SNP genotypes, was constructed among animals using the VanRaden method 1 (VanRaden, 2008). Association analyses were undertaken for each SNP separately using linear mixed models in WOMBAT (Meyer, 2007), to calculate SNP effects for all 1,883 animals. The model fitted for each SNP analysis was:

$$\textit{Deregressed EBV} = \mu + \textit{SNP} + a + e$$

where *deregressed EBV* is the dependent variable, μ is the fixed effect of the population mean, *SNP* is the fixed effect of allele dosage for each SNP (coded as 0, 1, or 2), a is the random effect of the animal, where $a \sim N(0, \mathbf{G}\,\sigma_a^2)$ with σ_a^2 representing the additive genetic variance of the animal and **G** is the genomic relationship matrix among animals; e represents the residual, where $e \sim N(0, \mathbf{I}\,\sigma_e^2)$ with σ_e^2 representing the residual variance and **I** represents the identity matrix. The dependent variable was weighted using the formula by Garrick et al. (2009):

$$w_i = \frac{1 - h^2}{\left[c + \frac{1 - r_i^2}{r_i^2}\right] h^2}$$

where w_i = is the weighting factor of the deregressed EBV of the *ith* animal, h^2 is the heritability estimate (i.e., $h^2 = 0.05$ as estimated in the present study), r_i^2 is the reliability of the deregressed EBV for the *ith* animal, and *c* is the genetic variance not accounted for by the SNPs (i.e., $c = 0.90$). Test statistics for all SNPs were obtained and SNPs with a p-value of $\leq 5 \times 10^{-8}$ were considered to be genome-wide significant.

Defining QTLs

Genome-wide significant SNPs (p-value $\leq 5 \times 10^{-8}$) informed the initial positions for QTL regions associated with humoral response to MAP. The QTL start and end positions were defined based on SNPs that were in strong linkage disequilibrium (LD) with these significantly associated SNPs. An r^2 threshold value of ≥ 0.7 was utilized to define whether or not SNPs were in strong LD with the significantly associated SNPs. The QTL boundaries were defined as being the SNPs within a 5Mb window from the significantly associated SNPs that passed the LD threshold. In cases where QTL boundaries were overlapping, these QTL were merged and considered as a single (larger) QTL.

Downstream Bioinformatic Analyses

Once defined, the QTLs were subsequently mined for the presence of annotated candidate genes using Ensembl (https://www.ensembl.org/) based on the UMDv3.1 genome build. Only non-intergenic SNPs were considered for further analysis, i.e. SNPs with annotation information being one of the following: intron variant, splice donor variant, stop gained variant, missense variant, synonymous variant, downstream gene variant or upstream gene variant. Pathway analyses were then conducted based on these identified candidate genes using InnateDB (http://www.innatedb.com/) (Breuer et al., 2013), which utilized integrated pathway data from the Reactome Pathway Knowledgebase (https://reactome.org/) (Fabregat et al., 2018), and the Pathway Interaction Database (Schaefer et al., 2009). The hypergeometric algorithm and Benjamini-Hochberg correction were used for querying InnateDB.

RESULTS

Variance components

The prevalence of positive humoral response to MAP in the overall edited national cow data which included 136,767 cows was 4.0%; in the restricted cohort of 33,818 cows it was 7.7% and it ranged from 3.1% to 5.5% among the 8 different laboratories (Table 1). For the animal

312 models, heritability estimates ranged from 0.006 to 0.084 (Table 1); heritability estimates for
313 the sire models ranged from 0.022 to 0.046. The heritability estimate for the overall population
314 of 136,767 animals was 0.020 (SE = 0.003). The heritability estimate for the cohort of 33,818
315 animals that were subsequently used for the genomic analysis was 0.050 (SE = 0.008). The
316 additive genetic standard deviation for the prevalence of positive humoral response to MAP in
317 the 10 (i.e., data from 8 laboratories and either the full or reduced dataset) different datasets
318 ranged from 0.014 to 0.058 (Table 1). The additive genetic standard deviation for humoral
319 response to MAP in the overall population was 0.027; the additive genetic standard deviation
320 for the reduced cohort used for genomic analysis was 0.058.

321 ***Estimated Breeding Values and their validation***

322 The mean EBV reliability of the 433,989 animals (i.e. the 33,818 with MAP humoral
323 phenotypes and their 400,171 non-phenotyped relatives) was 0.089 (SD = 0.087). The EBV of
324 individual animals ranged from -0.190 to 0.159; the maximum EBV reliability was 0.92.
325 Considering only the cows that had their own MAP phenotype available (i.e. the restricted
326 cohort of 33,818 cows), the mean EBV reliability was 0.192 (SD = 0.058).

327 Summary statistics relating to the validation of EBVs are in Table 2. The raw mean
328 prevalence of MAP in the poor, average and good EBV strata was 0.16, 0.14 and 0.12,
329 respectively. The mean reliability of the EBVs was similar for each stratum (poor = 0.149,
330 average = 0.144, good = 0.143). Stratum for MAP EBV was associated (p-value $\leq$0.001) with
331 the logit of the probability of a positive MAP outcome; the area under the receiver operating
332 curve when just MAP EBV stratum was included in the logistic regression was 0.539.
333 Irrespective of whether only MAP stratum alone was included in the model or whether MAP
334 stratum was also included simultaneously with other fixed effects, the poor EBV stratum had
335 a greater (p-value <0.05) odds (1.37 to 1.43) of having a positive MAP outcome, relative to the
336 good EBV stratum; although numerically worse, the odds of a positive MAP outcome in the
337 animals within the average EBV stratum did not differ significantly from the good EBV
338 stratum. The predicted probability of a positive MAP outcome in the poor, average and good
339 EBV stratum for a third parity cow, 100 to 149 DIM, with no heterosis or recombination in the
340 average herd-year was 0.134, 0.115 and 0.101, respectively.

341 ***Genome-based associations and downstream bioinformatic analyses***

A Manhattan plot depicting the association between each SNP and the deregressed EBV phenotype for MAP humoral response is presented in Figure 1. The present study identified 223 SNPs as being associated (p-value $\leq 5 \times 10^{-8}$) with MAP humoral response. A total of 17,960 SNPs were identified as being in strong LD with these 223 SNPs, resulting in a total of 18,181 SNPs considered as being associated with MAP humoral response. These SNPs resided within 47 QTL regions, which were distributed across 17 chromosomes namely BTA 1, 3, 5, 6, 8, 9, 10, 11, 13, 14, 18, 21, 23, 25, 26, 27 and 29 (Table 3). The chromosome which harbored the most QTLs was BTA1 (10 QTLs), the largest of the QTLs being 6.87Mb in length. Mining these 47 QTL regions for SNP annotation data and candidate genes resulted in 623 positional candidate genes being identified for further investigation; further information on these findings can be found in Supplementary Table S1. Of the 623 genes, the majority (301) were on BTA 18 (Table 3). The single QTL which harbored the most annotated positional candidate genes was on BTA 18 (from 53905719 – 61291660 bp; 243 genes), which contains genes such as *PKD2*, *HIF3A* and *KLK1*.

The QTL that harbored the most genome-wide significantly associated SNPs was on BTA 1 (from 70108608 – 71811570 bp), with 65 SNPs identified and a further 14,543 in strong LD, and was 1.7 Mb long. The 5 most significantly associated SNPs with MAP resistance resided in a QTL upstream of the QTL which harboured the most significantly associated SNPs on BTA 1 (SNP genome-wide threshold p-value = 2.137×10^{-13}) and all within a range of 16.7 kb of each other. These SNPs were all mapped to the *Kalirin* (*KALRN*) gene; 4 were intronic variants (rs378864226, rs719379694, rs37959091, and rs384286217) and one was identified as a splice donor variant (rs378147396). Indeed, a variety of DNA variant annotations such as downstream variants, upstream variants, intronic variants, and splice region variants within the *KALRN* gene were significantly associated with MAP in the present study.

In total, there were 22 positional candidate genes identified within the 47 QTL associated with MAP (Table 3), of which 10 were identified as potential functional candidate genes. Upon investigating the literature for potential biological functions of each, the most likely functional candidate genes identified along with *KALRN*, are *Zinc Finger And BTB Domain Containing 20* (*ZBTB20*), *lipoma-preferred partner* (*LPP*), *src-like-adapter 2* (*SLA2*), *Coagulation factor XIII A chain* (*F13A1*), *leucine-rich repeats and calponin homology (CH) domain containing 3 (LRCH3), DnaJ Heat Shock Protein Family (Hsp40) Member C6*

(DNAJC6), Zinc Finger DHHC-Type Containing 14 (ZDHHC14), sorting nexin-1 (SNX1) and *hyaluronan synthase 2 (HAS2).*

Seventy-six distinct bovine biological pathways were initially identified using the list of 623 positional candidate genes. After applying the hypergeometric algorithm, 29 pathways had an unadjusted p-value <0.05 (Supplementary Table S2). After applying the Benjamini-Hochberg correction, none of the pathways were identified as being enriched (minimum corrected p-value = 0.108). When known data on human biological pathways were included in the pathway analyses, 482 orthologous biological pathways were identified. Of these, 64 pathways had an unadjusted p-value <0.05 (Supplementary Table S3). Once the Benjamini-Hochberg adjustment was applied, none of the pathways were identified as being enriched (minimum corrected p-value = 0.239).

DISCUSSION

The challenges in addressing many cattle diseases, including bovine paratuberculosis, necessitates consideration of other disease mitigation strategies, one of which could be animal breeding. One of the advantages of breeding as a strategy to improve animal health is that it is cumulative and permanent with the genetic merit of a given animal being a function of all selection decisions made throughout its ancestral generations. Since in dairy cattle, the female must become pregnant to initiate a subsequent lactation, the marginal cost incurred to use superior germplasm semen is usually minimal. Given the detected existence of substantial genetic variation in humoral response to MAP in the present study, the justification for considering genetic merit for humoral response to MAP in a breeding program is therefore strong. This was substantiated by the exercise in the present study that validated MAP EBVs where a 1.43 greater odds of yielding a positive MAP result was detected in cows ranked poorly on genetic merit for MAP relative to their herd contemporaries ranked best on genetic merit for MAP. Twomey et al. (2016) undertook a similar EBV validation exercise based on cows divergent in parental average EBV for liver fluke. The area reported under the ROC reported by Twomey et al (2016) was 0.522, with cows in the top (i.e., worst) 10% being 1.28 (95% CI 1.05 – 1.36) times more likely to have livers damaged by liver fluke, compared to contemporaries in the bottom (i.e. best) 10%. This difference in odds translated to a 6% unit probability difference between the top and bottom 10% of cows in the study cohort. Similarly,

using bovine tuberculosis (bTB) infection data in cattle, Ring et al. (2018 under review) reported a mean prevalence of 9.3% in cows in the worst 20% on parental EBV for bTB versus a mean prevalence of 6.9% in cows in the best 20% on parental EBV for bTB; this equated to an odds ratio of 1.44. The area under the ROC reported was 0.529 (95% CI 0.5236 – 0.5342) for the present study. These two validation studies, together with the results from the validation exercise in present study, clearly demonstrate the potential health gains achievable with genetic selection. While the mean EBV reliability of the validation cows in the present study was low (i.e., on average 0.145), further differentiative ability could have been achieved if the reliability of the EBV was greater. Such an increase in reliability could be achieved through access to more phenotypic data or the inclusion of genomic information in the prediction process (Meuwissen et al., 2001). Using the approach proposed by Wray et al. (2001), assuming a heritability of humoral response to MAP of 0.05 and a prevalence of 7.7%, the area under the ROC curve to predict the outcome for humoral response to MAP would increase from 0.56, if a only quarter of the genetic variance in the phenotype could be explained to 0.59 and eventually 0.63 if half and all of the genetic variance could be explained, respectively.

A key question to be addressed in future studies is whether selecting for or against positive antibody response to MAP is beneficial for cattle populations. It could be surmised that selecting against positive antibody response is, in reality, selecting for the animals that can maintain a degree of internal cellular homeostasis despite having been infected by MAP. Such animals may in fact be shedding MAP, yet may not progress to the advanced clinical stages of bovine paratuberculosis. Irrespective, single trait selection for any trait is never recommended and thus any consideration of breeding for changes in the antibody response to MAP should be undertaken within the framework of a multi-trait breeding objective.

Estimated variance components for MAP in comparison to previous studies

Previous studies that estimated genetic parameters for paratuberculosis in cattle suggest heritability estimates of susceptibility to the disease range from <0.01 (Koets et al., 1999) to 0.283 (Küpper *et al.,* 2012); all studies to-date (which reported the actual breed of cattle used) have been undertaken in dairy cattle. A meta-analysis of the two available studies (Hinger et al., 2008, Berry et al., 2010) on dichotomized humoral response to MAP using blood ELISA and linear animal models, similar to the strategy adopted in the present study, resulted in a

pooled heritability estimate of 0.072 (pooled SE = 0.014); this was not very different to the 0.05 (SE = 0.008) estimated in the present study. Once the prevalence of the previous estimates of heritability for dichotomized humoral response to MAP using ELISA were adjusted to a prevalence of 7.7%, as observed in the present study, the heritability on the underlying scale (Robertson and Lerner, 1949) varied from 0.19 (Hinger et al., 2008) to 0.34 (Berry et al., 2010), with an overall mean pooled heritability estimate of 0.234 for the two studies; the corresponding value for the present study was 0.17. Neither the additive genetic standard deviation nor the residual standard deviation for the dichotomised humoral response to MAP were reported by Hinger et al. (2008) or Berry et al. (2010). As such, the additive genetic standard deviation and residual standard deviation in dichotomised humoral response to MAP infection reported in the present study cannot be compared.

The heritability estimate for the humoral response to MAP infection using a sire model and the 33,818 cows in herds with at least 5 positive cows and at least one of those cows yielding a positive ELISA to MAP in the present study is of a similar magnitude of that reported by Kirkpatrick and Lett (2018). The range reported by Kirkpatrick and Lett (2018) using sire models and dichotomised humoral (or milk) antibody response to MAP was 0.041 (SE = 0.004) – 0.062 (SE = 0.007). The upper and lower limits of this range were based on cohorts of herds with at least one positive test (n = 999 sires represented by 222,872 daughter ELISA records) and ≥5% positive tests (n = 475 sires represented by 65,289 daughter ELISA records), respectively. Perhaps, as suggested by Kirkpatrick and Lett (2018), a sire threshold model analysis is a better analysis to conduct to gain an accurate estimate of the true heritability value for antibody responses to MAP infection. This approach may ameliorate the effects of the prevalence of MAP on a per-cohort basis on the estimates of heritability for dichotomized humoral response to MAP infection. When the heritability estimates for dichotomized antibody response to MAP using blood or milk ELISAs were adjusted to the prevalence reported in the present study (7.7%), the heritability on the underlying scale (Robertson and Lerner, 1949) varied from 0.14 – 0.21; the corresponding value for the present study was 0.16.

The additive genetic standard deviation for humoral response to MAP was reported only by Gao et al. (2018), and was 0.125 which is considerably higher than the genetic standard deviation of 0.058 in the present study. The residual standard deviation (again, only reported by Gao et al. (2018)) was 0.55, compared to the lower estimate of 0.25 reported in the present study. Hence, the lower heritability reported in the present study could be attributed to a

relatively larger residual variance and lower genetic variance compared to other studies. Compounding this could be the fact that heritability estimate ranges reported in the present study on a per lab basis demonstrate site-specific levels of noise (i.e. residual variance). Noise could have been introduced in part due to environmental differences that existed between the laboratories, e.g. laboratory technicians at each laboratory could have possessed various levels of experience in conducting ELISAs. Furthermore, these assays may have been a routine procedure in some laboratories which specialize in such assays, whilst being an infrequent assay conducted in others. These factors, among others may have contributed to varying levels of residual variance being introduced on a per laboratory basis. Moreover, the somewhat low heritability estimate could also be attributed to the imperfect nature of ELISA tests in diagnosing cases and controls. Since ELISAs rely on the humoral response of individuals, they may not detect animals that are in the early stages of MAP infection, as they may not have mounted a humoral response to MAP infection at the point of testing (and may not for months, or even years); this results in infected animals being reported as ELISA negative to MAP, despite shedding MAP into the environment (Milner et al., 1987). False positive humoral results can also arise due to bTB infected animals, and bTB testing in herds prior to MAP testing without sufficient time elapsing between tests (i.e., 90 days) (Lilenbaum et al., 2007; Varges et al., 2009). Environmental mycobacteria (closely related to, but excluding MAP) could also potentially result in false positive ELISA results in the present study (Osterstock et al., 2007). As such, this misclassification introduces noise, thus biasing down heritability estimates (Milner et al., 1987, Bishop and Woolliams, 2010).

Despite the relatively low heritability for humoral response to MAP in the present study, an accuracy of selection of 0.34 could be achieved with phenotypic information on just 10 progeny; similarly, based on solely progeny phenotypes in a univariate genetic evaluation, MAP phenotypes on just 76 progeny would be required to achieve an accuracy of selection for humoral response to MAP of 0.70. High accuracy of selection is thus achievable in the presence of a national recording system for humoral response to MAP in cattle.

Candidate genes associated with humoral response to MAP

Of the 22 positional candidate genes identified in the 47 identified QTLs, ten were identified as potential functional candidate genes, based upon their gene ontology results and previous reports on their function in the literature (Bannantine and Bermudez, 2013, Liu et al., 2013, Marton et al., 2015). Some of these genes are novel in the sense that they have not previously

been associated with bovine paratuberculosis. The functional candidate genes of interest identified were *KALRN*, *ZBTB20*, *LPP*, *SLA2*, *F13A1*, *LRCH3*, *DNAJC6*, *ZGHHC14*, *SNX1* and *HAS2*; the (potential) biological functions of these are outlined below. All candidate genes have been shown to exhibit an expression profile in the bovine colon, duodenum, ileum, spleen and lymph nodes (Harhay et al., 2010). As stated by Brito et al. (2018), resistance or susceptibility to MAP infection appears to be a highly polygenic trait, reflecting the nature of bovine paratuberculosis, which is indeed, a highly complex trait. As such, the environmental influence on the manifestation of the trait must not be igrnoed. This influence may, in part, give rise to the fact that previous studies have implicated all *bos Taurus* autosomes as having an association with bovine paratuberculosis, yet there is little consistency observed among the reported genomic locations of these studies. Other factors which must be considered as contributors to this lack of coherence include the low heritability of the trait, inconsistent methods to classify MAP positive / negative animals, and divergent statistical methodologies employed on varying sample sizes of different cattle populations (Brito et al., 2018). Specific to the present study, another factor contributing to spurious results is the error in genotype imputation from sequence data. Despite the methods employed in the present study yielding a high imputation accuracy of 98%, the 2% error may have impacted upon the allele frequencies of the population, and thus the significance of some of the SNP associations obtained. Nevertheless, the results of the present study corroborate some previous findings in cattle, whilst some results are agree with those identified in human populations.

KALRN. Gene ontology (GO) results for *KALRN*, found in a QTL on BTA1 (between positions 64369225 and 69754155), support it as a strong candidate gene influencing humoral response to MAP infection because of its role in encoding for the RhoGEF kinase protein. Human tissues such as the duodenum, colon, lymph nodes, small intestine and spleen have been reported to have expressed *KALRN* (Fagerberg et al., 2014), all of which are important in the pathology of Johne's disease. Biological processes in which the RhoGEF kinase protein is involved include the positive regulation of Rho protein signal transduction as well as positive regulation of Rho GTPase activity. This is particularly interesting considering that the RhoA GTPase protein is located in intestinal epithelial cells and, upon stimulation by MAP, the bacteria infiltrates the intestinal submucosa (Bannantine and Bermudez, 2013). Infiltration of, and transcytosis through, the intestinal submucosa is crucial for MAP growth and replication, as it facilitates the bacterium crossing from the intestinal lumen to its target cells (i.e. macrophage and dendritic cells). From within macrophage and dendritic cells, MAP

successfully modulates the host's immune responses, subverting immune attack. The intracellular subversion enables the bacterium to establish a niche within the host, providing it sufficient time to grow, replicate and spread through the surrounding intestines and lymph nodes (Bannantine and Bermudez, 2013, Arsenault et al., 2014, Koets et al., 2015). Furthermore, the rho kinase signal pathway has previously been implicated in inflammatory bowel disease (IBD) in humans, playing roles in intestinal barrier damage, abnormal immune response and intestinal fibrosis (Huang et al., 2015). Brito et al. (2018) also identified *KALRN* as a candidate gene in an independent population of ELISA-tested Canadian Holstein cattle. Given this, it is not unrealistic to hypothesize that *KALRN* plays an important role in the aforementioned processes in cattle susceptible to Johne's disease.

ZBTB20. Gene ontology results implicate the gene's product in metal ion binding processes. The fact that MAP cannot endogenously produce mycobactin (necessary for iron transport), makes it an obligate intracellular parasite dependent upon the host for iron uptake and metabolism (Clark et al., 2008, McNees et al., 2015, Rathnaiah et al., 2017). Perhaps whilst subverting the host immune system inside target cells, MAP utilizes *ZTBT20* to provide it with a means of exogenous iron to grow and replicate. The *ZTBT20* gene has also been implicated in promoting toll-like receptor (TLR)-triggered innate immune responses in hosts via repressing the transcription of *IκBα*, which in turn promotes the activation of *NFκB* (Liu et al., 2013). This corroborates the results of the gene-set enrichment analysis conducted by Kiser et al. (2017) in cattle, where enrichment analyses identified pathways involved with *NFκB*. TLRs have also been previously implicated in bovine paratuberculosis through candidate gene studies (Mucha et al., 2009; Koets et al., 2010). Expression of *ZBTB20* has been observed in human tissues relevant to Johne's disease such as the colon, duodenum, lymph nodes, small intestine and spleen (Fagerberg et al., 2014). Moreover, *ZBTB20* was also identified as a candidate gene in the GWAS conducted by Brito et al. (2018) in dairy cows, further substantiating the evidence that *ZBTB20* could indeed be associated with the resistance status of cattle to MAP.

LPP. The potential role for *LPP* in the etiology of Johne's disease could lie in the fact that it aids in signal transduction, cell migration, cytoskeletal remodeling and cell-cell adhesion (Jin et al., 2009). Although its exact function has yet to be established, *LPP* has been shown to enhance cell migration in response to cellular injury and has been linked to diseases such as atherosclerosis, a disorder which is characterized by an excessive chronic inflammatory immune response. Initially, this acute inflammatory response is beneficial in protecting the host

during immune challenges, but it becomes detrimental to the host when it turns chronic in nature (Ross and Agius, 1992). The chronicity of the inflammatory response causes the disease due to the fibroproliferative properties of the response resulting in granulomatous tissue (Ross and Agius, 1992; Kunkel et al., 1998). Perhaps what is an initial beneficial host immune response to MAP infection, progresses to become a chronic malignant over-activation of the host immune response. Coordinated in part by *LPP*, this response could result in the chronic granulomatous gastroenteritis characteristic of Johne's disease. It could also be surmised that MAP uses *LPP* to intracellularly modulate macrophage lipoprotein metabolism for the benefit of the bacteria, whilst the recruitment of macrophages to the intestinal epithelium would maximize the bacteria's opportunity to infiltrate host cells that enable it to grow and replicate. Human tissue expression studies report that *LPP* is expressed in the colon, duodenum, lymph nodes, small intestine and spleen (Fagerberg et al., 2014).

SLA2. Gene ontology results for *SLA2* includes T cell activation, which is an integral part of the host immune response. Src-like adaptor protein 2 (SLAP-2), encoded by *SLA2*, has been observed to be expressed in the small intestine, the colon, the spleen and the lymph nodes in humans (Fagerberg et al., 2014). As reviewed by Marton et al. (2015), the SLAP-2 protein negatively regulates the antigen receptor signaling action of both B and T lymphocytes, thus suppressing host immune responses. It has also been suggested that SLAP-2 is necessary for the maturation and activation of monocyte and dendritic cells (Liontos et al., 2011). Perhaps MAP modulates the expression of *SLA2* within monocytes, thus subverting the host immune responses long enough for sufficient growth and replication to thrive.

F13A1. Encoding for coagulation factor XIIIA, *F13A1* is a strong candidate gene in the etiology of Johne's disease as it promotes wound healing in damaged tissues, providing structural support and repaired vascularization. Coagulation factor XIIIA has also been implicated in IBD, as similar to atherogenesis, these initial beneficial mechanisms can in fact become detrimental to the host. The chronic inflammation observed in humans suffering from IBD is characterized by a hypercoagulable state accompanied by aberrations in coagulation processes (Danese et al., 2007). Mechanisms coordinated by coagulation factor XIII can cause increased neovascularization and impaired mucosal healing, which can cause additional inflammation in patients suffering from IBD (Gemmati et al., 2016). Genetic variation in *F13A1* could result in susceptibility to Johne's disease through abnormal coagulation

processes. The human colon, duodenum, lymph node, small intestine and spleen express *F13A1* (Fagerberg et al., 2014).

LRCH3. Although there is a relative paucity of information in the literature regarding this gene, members of the leucine-rich repeat (LRR) family are thought to be critical in initiating inflammatory innate immune responses (Bell et al., 2003). Specifically, *LRCH3*, which is a member of this family, is a potential candidate gene involved in the pathology of Johne's disease as it has previously been implicated in NFκB activation mechanisms in the NFκB immunoregulatory pathways (Gewurz et al., 2012). This evidence corroborates the aforementioned evidence of the gene-set enrichment analysis conducted by Kiser et al. (2017). Moreover, evidence in the literature has shown that *LRCH3* is significantly differentially expressed between healthy colorectal tissue sample controls and samples presenting with colorectal cancer; the cancerous tissue samples exhibit a higher level of *LRCH3* expression (Piepoli et al., 2012).

DNAJC6. The DnaJ Heat Shock Protein Family (Hsp40) Member C6 has been shown to be a useful biomarker in the design of predictive algorithms which classify the severity of IBD in patients to better diagnose and improve patient quality of life (Montero-Meléndez et al., 2013). Also known as *PARK19*, *DNAJC6* has also been implicated in Parkinson's disease (Ran & Belin, 2014). Following a review of the available literature, Dow (2014) hypothesised that there could indeed be a link between MAP and Parkinson's disease in humans, a hypothesis which is yet to be confirmed.

ZDHHC14. Zinc finger DHHC domain-containing 14 protein has been shown by Oo et al. (2014) to promote migration and invasion of scirrhous type gastric cancer, promoting tumour progression when overexpressed. The overexpression of this gene was significantly associated with scirrhous-patterning and the degree of the depth of cancer tumour invasion of gastric mucosa tissue collected from patients suffering from gastric cancer. As such, it is not unreasonable to hypothesise that *ZDHHC14* plays a role in Johne's disease, considering it is also a chronic, progressive disease of the gastrointestinal tract.

SNX1. Downregulation of *SNX1* has been linked to drug resistance in colorectal cancer, whilst also predicting poor prognosis for the disease (Bian et al., 2016). Specifically, the downregulation of *SNX1* has been strongly associated with poor overall survival rate of patients suffering from colorectal cancer. Bian et al. (2016) hypothesised that upregulation of the gene

could provide a novel therapeutic intervention for this disease, an approach which could potentially be translated to veterinary medicine for the treatment of Johne's disease in the future. Interestingly, there may be a synergy between the susceptible alleles for the aforementioned *KALRN* gene and the susceptible alleles for the *SNX1* gene, as Prosser et al. (2010) have shown that *SNX1* interacts with the Rac1 and RhoG guanine nucleotide exchange factor Kalirin-7; this process recruits the inactive Rho GTPase to its exchange factor, influencing cell membrane remodeling – perhaps a similar interaction occurs to initiate the process outlined above.

HAS2. Upregulation of this gene, which produces hyaluronan synthesis, has been shown to be associated with IBD (among other inflammatory diseases), and has been shown to be expressed in human intestinal cells under the influence of the proinflammatory cytokine IL-1beta (Ducale et al., 2005). In fact, Vigetti et al. (2010) demonstrated that in endothelial cells, monocyte adhesion depends strongly on hyaluronan. Throughout the inflammatory response, cells produce hyaluronan matrices, which in turn promote the adhesion of monocytes and macrophages to cell surfaces. This aids in the processes governing the transformation of macrophages in to foam cells, which contributes toward atherosclerotic plaque formations, very similar to *F13A1* outlined above. Vigetti et al. (2010) also demonstrated that IL1-beta, and tumour necrosis factors alpha and beta strongly induce hyaluronan synthesis via the *NFκB* pathway, which again, was previously implicated in the pathology of Johne's disease by Kiser et al., (2017).

CONCLUSIONS

Exploitable genetic variation in humoral response to MAP exists in the Holstein-Friesian population sampled in the present study. Thus, ample opportunity exists to capitalize on this variation through a breeding program for reduced MAP susceptibility. Moreover, the biological insights achieved from the identified QTLs and candidate genes provide novel evidence to support the notion that bovine paratuberculosis is molecularly similar to inflammatory bowel disease in humans; validation of the results reported herein are, nonetheless, crucial especially for the low MAF SNPs. Candidate genes and QTL regions reported within could also be used a priors in any further analyses, either association studies or genomic/phenotypic predictions that exploit Bayesian techniques. Furthermore, the results from the present study potentially enables researchers to gain a greater understanding of bovine paratuberculosis, and potential

therapeutic targets for future investigation, as breakthroughs in the treatment for inflammatory bowel disease may also be transferable to veterinary medicine.

ACKNOWLEDGEMENTS

Funding from the Irish Department of Agriculture, Food and the Marine STIMULUS research grant HealthyGenes (Dublin) and Science Foundation Ireland (SFI) principal investigator award grant number 14/IA/2576 is greatly appreciated as is the support from a research grant from both Science Foundation Ireland and the Department of Agriculture, Food and Marine on behalf of the Government of Ireland under the Grant 16/RC/3835 (VistaMilk). We gratefully acknowledge the 1000 Bull Genomes Consortium for providing accessibility to whole-genome sequence data which was used in this study.

REFERENCES

Alpay, F., Y. Zare, M.H. Kamalludin, X. Huang, X. Shi, G.E. Shook, M.T. Collins, and B.W. Kirkpatrick. 2014 Genome-wide association study of susceptibility to infection by Mycobacterium avium subspecies paratuberculosis in Holstein cattle. PLoS One 9:e111704.

Arsenault, R.J., P. Maattanen, J. Daigle, A. Potter, P. Griebel, and S. Napper 2014 From mouth to macrophage: mechanisms of innate immune subversion by Mycobacterium avium subsp. paratuberculosis. Vet Res 45:54.

Bannantine, J.P., and J.E. Bermudez 2013 No holes barred: invasion of the intestinal mucosa by Mycobacterium avium subsp. paratuberculosis. Infect Immun 81:3960-3965.

Beissinger, T.M., G.J. Rosa, S.M. Kaeppler, D. Gianola, and N. De Leon 2015 Defining window-boundaries for genomic analyses using smoothing spline techniques. Genetics Selection Evolution 47:30.

Bell, J.K., G.E. Mullen, C.A. Leifer, A. Mazzoni, D.R. Davies, and D.M. Segal 2003. Leucine-rich repeats and pathogen recognition in Toll-like receptors. Trends Immunol 2003;24:528– 33.

Berry, D.P., and R.D. Evans 2014 Genetics of reproductive performance in seasonal calving beef cows and its association with performance traits. Journal of Animal Science 92(4):1412-1422.

Berry, D.P., M. Good, P. Mullowney, A.R. Cromie, and S.J. More. 2010. Genetic variation in serological response to Mycobacterium avium subspecies paratuberculosis and its

association with performance in Irish Holstein-Frieisian dairy cows. Livestock Science 131 : 102-107

Bian, Z., Y. Feng, Y. Xue, Y. Hu, Q. Wang, L. Zhou, Z. Liu, J. Zhang, Y. Yin, B. Gu, and Z. Huang 2016. Down-regulation of SNX1 predicts poor prognosis and contributes to drug resistance in colorectal cancer. Tumour Biol 37(5):):6619-25. doi: 10.1007/s13277-015-3814-3.

Bishop, S.C., and J.A. Woolliams 2010 On the genetic interpretation of disease data. PLoS One 5:e8940.

Breuer, K., A.K. Foroushani, M.R. Laird, C. Chen, A. Sribnaia, R. Lo, G.L. Winsor, R.E. Hancock, F.S. Brinkman, and D.J. Lynn 2013 InnateDB: systems biology of innate immunity and beyond--recent updates and continuing curation. Nucleic Acids Res 41:D1228-1233.

Brito, L.F., S. Mallikarjunappa, M. Sargolzaei, A. Koeck, J. Chesnais, F.S. Schenkel, K.G. Meade, F. Miglior, and N.A. Karrow 2018 The genetic architecture of milk ELISA scores as an indicator of Johne's disease (paratuberculosis) in dairy cattle. J. Dairy Sci 101:10062-10075.

Browning, B.L., and S.R. Browning 2016 Genotype Imputation with Millions of Reference Samples. Am J Hum Genet 98:116-126.

Clark, D.L., Jr., J.J. Koziczkowski, R.P. Radcliff, R.A. Carlson, and J.L. Ellingson 2008 Detection of Mycobacterium avium subspecies paratuberculosis: comparing fecal culture versus serum enzyme-linked immunosorbent assay and direct fecal polymerase chain reaction. J Dairy Sci 91:2620-2627.

Collins, M.T., S.J. Wells, K.R. Petrini, J.E. Collins, R.D. Schultz, and R.H. Whitlock 2005 Evaluation of Five Antibody Detection Tests for Diagnosis of Bovine Paratuberculosis. Clin Diagn Lab Immunol 12(6):685-692.

Collins, M.T. 2003 Paratuberculosis: review of present knowledge. Acta Vet Scand 44:217-221.

Daetwyler HD, Capitan A, Pausch H, Stothard P, Van Binsbergen R, Brondum RF, Liao XP, Djari A, Rodriguez SC, Grohs C, Esquerre D, Bouchez O, Rossignol MN, Klopp C, Rocha D, Fritz S, Eggen A, Bowman PJ, Coote D, Chamberlain AJ, Anderson C, VanTassell CP, Hulsegge I, Goddard ME, Guldbrandtsen B, Lund MS, Veerkamp RF, Boichard DA, Fries R, Hayes BJ 2014 Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. Nat Genet 46:858-865.

Danese, S., A. Papa, S. Saibeni, A. Repici, A. Malesci, and M. Vecchi 2007 Inflammation and coagulation in inflammatory bowel disease: The clot thickens. Am J Gastroenterol 102:174-186.

Das S, Forer L, Schonherr S, Sidore C, Locke AE, Kwong A, Vrieze SI, Chew EY, Levy S, McGue M, Schlessinger D, Stambolian D, Loh PR, Iacono WG, Swaroop A, Scott LJ, Cucca F, Kronenberg F, Boehnke M, Abecasis GR, Fuchsberger C (2016) Next-generation genotype imputation service and methods. Nat Genet 48:1284-1287.

Del Corvo M, Luini M, Stella A, Pagnacco G, Ajmone-Marsan P, Williams JL, Minozzi G (2017) Identification of additional loci associated with antibody response to Mycobacterium avium ssp Paratuberculosis in cattle by GSEA-SNP analysis. Mamm Genome 28:520-527.

Dow CT (2014). M. paratuberculosis and Parkinson's disease--is this a trigger. Med Hypotheses 83(6):709-12. doi: 10.1016/j.mehy.2014.09.025.

Ducale AE, Ward SI, Dechert T, Yager DR (2005). Regulation of hyaluronan synthase-2 expression in human intestinal mesenchymal cells: mechanisms of interleukin-1beta-mediated induction. Am J Physiol Gastrointest Liver Physiol 289(3):G462-70.

Fabregat A, Jupe S, Matthews L, Sidiropoulos K, Gillespie M, Garapati P, Haw R, Jassal B, Korninger F, May B, Milacic M, Roca CD, Rothfels K, Sevilla C, Shamovsky V, Shorser S, Varusai T, Viteri G, Weiser J, Wu G, Stein L, Hermjakob H, D'Eustachio P (2018) The Reactome Pathway Knowledgebase. Nucleic Acids Res 46:D649-D655.

Fagerberg L, Hallstrom BM, Oksvold P, Kampf C, Djureinovic D, Odeberg J, Habuka M, Tahmasebpoor S, Danielsson A, Edlund K, Asplund A, Sjostedt E, Lundberg E, Szigyarto CA, Skogs M, Takanen JO, Berling H, Tegel H, Mulder J, Nilsson P, Schwenk JM, Lindskog C, Danielsson F, Mardinoglu A, Sivertsson A, von Feilitzen K, Forsberg M, Zwahlen M, Olsson I, Navani S, Huss M, Nielsen J, Ponten F, Uhlen M (2014) Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. Mol Cell Proteomics 13:397-406.

Gao Y, Cao J, Zhang S, Zhang Q, Sun D (2018) Short communication: Heritability estimates for susceptibility to Mycobacterium avium ssp paratuberculosis infection in Chinese Holstein cattle. Journal of Dairy Science 101:7274-7279.

Garrick, D. J., J. F. Taylor, and R. L. Fernando. 2009. Deregressing estimated breeding values and weighting information for genomic regression analyses. Genet. Sel. Evol. 41:55.doi: 10.1186/1297-9686-41-55.

Geraghty T., Graham D. A., Mullowney P., More S. J. 2014. A review of bovine Johne's disease control activities in 6 endemically infected countries. Prev Vet Med 116:1–11. doi: 10.1016/j.prevetmed.2014.06.003.

Gewurz BE, Towfic F, Mar JC, Shinners NP, Takasaki K, Zhao B, Cahir-McFarland ED, Quackenbush J, Xavier RJ and Kieff E (2012). Genome-wide siRNA screen for mediators of NF-κB activation. PNAS 109 (7) 2467-2472; doi: 10.1073/pnas.1120542109

763 Gilmour, A. R., B. J. Gogel, B. R. Cullis, and R. Thompson (2009). ASReml User Guide
764 Release 3.0. VSN International Ltd, Hemel Hempstead, HP1 1ES, UK. www.vsni.co.uk

765 Harhay GP, Smith TP, Alexander LJ, Haudenschild CD, Keele JW, Matukumalli LK,
766 Schroeder SG, Van Tassell CP, Gresham CR, Bridges SM, Burgess SC, Sonstegard TS
767 (2010) An atlas of bovine gene expression reveals novel distinctive tissue characteristics
768 and evidence for improving genome annotation. Genome Biol 11:R102.

769 Harris, N. B. and Barletta, R. G. 2001. Mycobacterium avium subsp. paratuberculosis in
770 Veterinary Medicine. Clinical Microbiology Reviews Jul; 14(3): 489–512. doi:
771 10.1128/CMR.14.3.489-512.2001.

772 Hinger M, Brandt H, Erhardt G (2008) Heritability estimates for antibody response to
773 Mycobacterium avium subspecies paratuberculosis in German Holstein cattle. J Dairy
774 Sci 91:3237-3244.

775 Oo, H. Z., Sentani, K., Sakamoto, N., Anami, K., Naito, Y., Uraoka, N., Oshima, T.,
776 Yanagihara, K., Oue, N., Yasui, W (2014). Overexpression of ZDHHC14 promotes
777 migration and invasion of scirrhous type gastric cancer. Oncology Reports 32(1): 403-
778 410. doi: 10.3892/or.2014.3166.

779 Huang Y, Xiao SY, Jiang QH (2015) Role of Rho kinase signal pathway in inflammatory bowel
780 disease. Int J Clin Exp Med 8:3089-3097.

781 Jin L, Hastings NE, Blackman BR, Somlyo AV (2009) Mechanical properties of the
782 extracellular matrix alter expression of smooth muscle protein LPP and its partner
783 palladin; relationship to early atherosclerosis and vascular injury. J Muscle Res Cell M
784 30:41-55.

785 Johne, H. A, and Frothingham, L. 1895. Ein eigenthuemlicher fall von tuberkulose beim rind.
786 Dtsch Z. Tiermed. Pathol. 21:438–454.

787 Kirkpatrick B. W., Lett B. M. 2018. Short communication: Heritability of susceptibility to
788 infection by Mycobacterium avium ssp. paratuberculosis in Holstein cattle.J Dairy Sci.
789 101:1-5. Doi: https://doi.org/10.3168/jds.2018-15021.

790 Kirkpatrick B. W., Shi X., Shook G. E. and Collins M. T. 2011. Whole-genome association
791 analysis of susceptibility to paratuberculosis in Holstein cattle. Anim. Genet. 42(2):149–
792 160. doi: 10.1111/j.1365-2052.2010.02097.x.

793 Kiser, J. N., White, S. N, Johnson, K. A., Hoff, J. L., Taylor, J. F and Neibergs, H. L. 2017.
794 Identification of loci associated with susceptibility to Mycobacterium avium subspecies
795 paratuberculosis (Map) tissue infection in cattle. J Anim Sci 95:1080–1091. doi:
796 10.2527/jas.2016.1152

797 Koets AP, Adugna G, Janss LG, van Weering HJ, Kalis CHJ, Harbers A, Rutten VPMG,
798 Schukken YH (1999) Genetic variation in susceptibility to M. a. paratuberculosis

799 infection in cattle. Proceedings of the Sixth International Colloquium on Paratuberculosis
800 169-175.

801 Koets, Ad. P., Eda, Shigetoshi and Sreevatsan, Srinand. 2015. The within host dynamics of
802 Mycobacterium avium ssp. paratuberculosis infection in cattle: where time and place
803 matter. Vet Res 46(1): 61. doi: 10.1186%2Fs13567-015-0185-0.

804 Kunkel SL, Lukacs NW, Strieter RM, Chensue SW (1998) Animal models of granulomatous
805 inflammation. Semin Respir Infect 13:221-228.

806 Küpper, J., Brandt, H., Donat, K. and Erhardt. 2012. G. Heritability estimates for
807 Mycobacterium avium subspecies paratuberculosis status of German Holstein cows
808 tested by fecal culture. Journal of Dairy Science. 95:2734–2739. doi: 10.3168/jds.2011-
809 4994.

810 Li H (2011) A statistical framework for SNP calling, mutation discovery, association mapping
811 and population genetical parameter estimation from sequencing data. Bioinformatics
812 27:2987-2993.

813 Lilenbaum W, Ferreira R, Dray Marassi C, Ristow P, Martins Roland Oelemann W, de Souza
814 Fonseca L (2007). Interference of Tuberculosis on the performance of ELISAs used in
815 the Diagnosis of Paratuberculosis in Cattle. Brazilian Journal of Microbiology 38:472-
816 477.

817 Liontos LM, Dissanayake D, Ohashi PS, Weiss A, Dragone LL, McGlade CJ (2011) The Src-
818 Like Adaptor Protein Regulates GM-CSFR Signaling and Monocytic Dendritic Cell
819 Maturation. J Immunol 186:1923-1933.

820 Liu XG, Zhang P, Bao Y, Han YM, Wang Y, Zhang Q, Zhan ZZ, Meng J, Li YK, Li N, Zhang
821 WPJ, Cao XT (2013) Zinc finger protein ZBTB20 promotes toll-like receptor-triggered
822 innate immune responses by repressing I kappa B alpha gene transcription. P Natl Acad
823 Sci USA 110:11097-11102.

824 Loh PR, Danecek P, Palamara PF, Fuchsberger C, Reshef YA, Finucane HK, Schoenherr S,
825 Forer L, McCarthy S, Abecasis GR, Durbin R, Price AL (2016) Reference-based phasing
826 using the Haplotype Reference Consortium panel. Nat Genet 48:1443-1448.

827 Marton N, Baricza E, Ersek B, Buzas EI, Nagy G (2015) The Emerging and Diverse Roles of
828 Src-Like Adaptor Proteins in Health and Disease. Mediat Inflamm.

829 McNees AL, Markesich D, Zayyani NR, Graham DY (2015) Mycobacterium paratuberculosis
830 as a cause of Crohn's disease. Expert Rev Gastroenterol Hepatol 9:1523-1534.

831 Meuwissen TH, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-
832 wide dense marker maps. Genetics 157:1819-1829.

833 Meyer K (2007) WOMBAT: a tool for mixed model analyses in quantitative genetics by
834 restricted maximum likelihood (REML). J Zhejiang Univ Sci B 8:815-821.

Milner AR, Lepper AWD, Symonds WN, Gruner E (1987) Analysis by Elisa and Western Blotting of Antibody Reactivities in Cattle Infected with Mycobacterium-Paratuberculosis after Absorption of Serum with M-Phlei. Res Vet Sci 42:140-144.

Minozzi G., Buggiotti L., Stella A., Strozzi F., Luini M. and Williams J. L. 2010. Genetic loci involved in antibody response to Mycobacterium avium ssp. paratuberculosis in cattle. PLoS One 5(6):e11117. doi: 10.1371/journal.pone.0011117.

Montero-Meléndez T, Llor X, García-Planella E, Perretti M, Suárez A (2013). Identification of novel predictor classifiers for inflammatory bowel disease by gene expression profiling. PLoS One 14;8(10):e76235. doi: 10.1371/journal.pone.0076235.

Neibergs HL, Settles ML, Whitlock RH, Taylor JF (2010) GSEA-SNP identifies genes associated with Johne's disease in cattle. Mamm Genome 21:419-425.

Osterstock JB, Fosgate GT, Norby B, Manning EJ, Collins MT, Roussel AJ (2007). Contribution of environmental mycobacteria to false-positive serum ELISA results for paratuberculosis. J Am Vet Med Assoc 230(6): 896-901.

Pant S. D., Schenkel F. S., Verschoor C. P., You Q. Kelton D. F., Moore S. S. and Karrow N. A. 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(3):176–182. doi: 10.1016/j.ygeno.2010.01.001.

Park, H. T. and Yoo, H. S. 2016. Development of vaccines to Mycobacterium avium subsp. paratuberculosis infection. Clinical and Experimental Vaccine Research. 5: 108-116. doi: 10.7774/cevr.2016.5.2.108.

Piepoli A, Palmieri O, Maglietta R, Panza A, Cattaneo E, Latiano A, Laczko E, Gentile A, Carella M, Mazzoccoli G, Ancona N, Marra G, Andriulli A (2012). The expression of leucine-rich repeat gene family members in colorectal cancer. Experimental Biology and Medicine 2012; 237: 1123 –1128. doi: 10.1258/ebm.2012.012042.

Pierce, C. F., Kiser, J. N., Hoff, J. L., Neupane, M., White, S. N., Taylor, J. F and Neibergs, H. L. 0203 Identification of loci on chromosome 3 associated with susceptibility to bovine paratuberculosis using genotypes imputed to whole genome sequence in Holstein cows. 94 issue supplement 5: 96-97. doi: 10.2527/jam2016-0203.

Prosser DC, Tran D, Schooley A, Wendland B, Nqsee JK. A novel, retromer-independent role for sorting nexins 1 and 2 in RhoG-dependent membrane remodeling. Traffic 11(10):1347-62. doi: 10.1111/j.1600-0854.2010.01100.x.

Ran C, Belin A (2014). The genetics of Parkinson’s disease: review of current and emerging candidates. Journal of Pankinsonism and Restless Legs Syndrome 2014(4):63-75. doi: 10.2147/JPRLS.S38954.

870 Rathnaiah G, Zinniel DK, Bannantine JP, Stabel JR, Grohn YT, Collins MT, Barletta RG
871 (2017) Pathogenesis, Molecular Genetics, and Genomics of Mycobacterium avium
872 subsp. paratuberculosis, the Etiologic Agent of Johne's Disease. Front Vet Sci 4:187.

873 Richardson, E. K. B. and More, S. J. 2009. Direct and indirect effects of Johne's disease on
874 farm and animal productivity in an Irish dairy herd. Irish Veterinary Journal. 62(8):526-
875 532. doi:10.1186/2046-0481-62-8-526.

876 Robertson A, Lerner IM (1949) The Heritability of All-or-None Traits: Viability of Poultry.
877 Genetics 34:395-411.

878 Ross R, Agius L (1992) The Process of Atherogenesis - Cellular and Molecular Interaction -
879 from Experimental Animal-Models to Humans. Diabetologia 35:S34-S40.

880 Sargolzaei M, Chesnais JP, Schenkel FS (2014) A new approach for efficient genotype
881 imputation using information from relatives. BMC Genomics 15:478.

882 Schaefer CF, Anthony K, Krupa S, Buchoff J, Day M, Hannay T, Buetow KH (2009) PID: the
883 Pathway Interaction Database. Nucleic Acids Res 37:D674-679.

884 Settles M, Zanella R, McKay SD, Schnabel RD, Taylor JF, Whitlock R, Schukken Y, Van
885 Kessel JS, Smith JM, Neibergs H (2009) A whole genome association analysis identifies
886 loci associated with Mycobacterium avium subsp. paratuberculosis infection status in US
887 holstein cattle. Anim Genet 40:655-662.

888 Sheng YH, Hasnain SZ, Florin THJ, McGuckin MA (2012) Mucins in inflammatory bowel
889 diseases and colorectal cancer. J Gastroen Hepatol 27:28-38.

890 Strandén, I., and E. A. Mäntysaari. 2010. A recipe for multiple trait deregression. Interbull
891 Bulletin. 42:21-24.

892 Stranden I, Lidauer M (1999) Solving large mixed linear models using preconditioned
893 conjugate gradient iteration. J Dairy Sci 82:2779-2787.

894 Sweeney, R. W., Collins, M. T., Koets, A. P., McGuirk, S. M. and Roussel, A. J. 2012.
895 Paratuberculosis (Johne's Disease) in Cattle and Other Susceptible Species. Journal of
896 Veterinary Medicine 26(6). doi: 10.1111/j.1939-1676.2012.01019.x.

897 Twomey AJ, Sayers RG, Carroll RI, Byrne N, Brien EO, Doherty ML, McClure JC, Graham
898 DA, Berry DP (2016) Genetic parameters for both a liver damage phenotype caused by
899 Fasciola hepatica and antibody response to Fasciola hepatica phenotype in dairy and beef
900 cattle. Journal of Animal Science 94:4109-4119.

901 VanRaden PM (2008) Efficient Methods to Compute Genomic Predictions. Journal of Dairy
902 Science 91:4414-4423.

903 VanRaden PM, Sanders AH (2003) Economic merit of crossbred and purebred US dairy cattle.
904 J Dairy Sci 86:1036-1044.

van Hulzen, K. J, Nielen, M., Koets, A. P., de Jong, G., van Arendonk, J. A. and Heuven, H. C. 2011. Effect of herd prevalence on heritability estimates of antibody response to Mycobacterium avium subspecies paratuberculosis. Journal of Dairy Science. 94(2):992-7. doi: 10.3168/jds.2010-3472.

van Hulzen, K. J. E., Schopen, G. C. B., van Arendonk, J. A. M., Nielen M., Koets A. P., Schrooten C. and Heuven H. C. M. 2012. Genome-wide association study to identify chromosomal regions associated with antibody response to Mycobacterium avium subspecies paratuberculosis in milk of Dutch Holstein-Friesians. J. Dairy Sci. 95(5):2740–2748. doi: 10.3168/jds.2011-5005.

Varges R, Dray Marassi C, Oelemann W, Lilenbaum W (2009). Interference of intradermal tuberculin tests on the serodiagnosis of paratuberculosis in cattle. Research in Veterinary Science 86(3):371-372.

Vigetti D, Genaseetti A, Karousou E, Viola M, Moretto P, Clerici M, Deleonibus S, De Luca G, Hascall VC, Passi A (2010). Proinflammatory cytokines induce hyaluronan synthesis and monocyte adhesion in human endothelial cells through hyaluronan synthase 2 (HAS2) and the nuclear factor-kappaB (NF-kappaB) pathway. J Biol Chem 285(32):24639-45. doi: 10.1074/jbc.M110.134536.

Whittington, R. J., Marshall, D. J., Nicholls, P. J., Marsh, I. B., and Reddacliff, L. A. 2004. Survival and dormancy of Mycobacterium avium subsp. paratuberculosis in the environment. Appl Environ Microbiol 70(5):2989-3004.

Whittington, R. J., Marsh, I. B. and Reddacliff, L. A. 2005. Survival of Mycobacterium avium subsp. paratuberculosis in dam water and sediment. Apply Environ Microbiol 71(9):5302-8.

Windsor, P. A and Whittington, R. J. 2010. Evidence for age susceptibility of cattle to Johne's disease. The Veterinary Journal. 184: 37-44.

Wopereis S, Lefeber DJ, Morava E, Wevers RA (2006) Mechanisms in protein O-glycan biosynthesis and clinical and molecular aspects of protein O-glycan biosynthesis defects: a review. Clin Chem 52:574-600.

Wray NR, Yang J, Goddard ME, Visscher PM (2010) The genetic interpretation of area under the ROC curve in genomic profiling. Plos Genet 6:e1000864.

Zare, Y., Shook, G. E., Collins, M. T. and Kirkpatrick, B. W. 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.

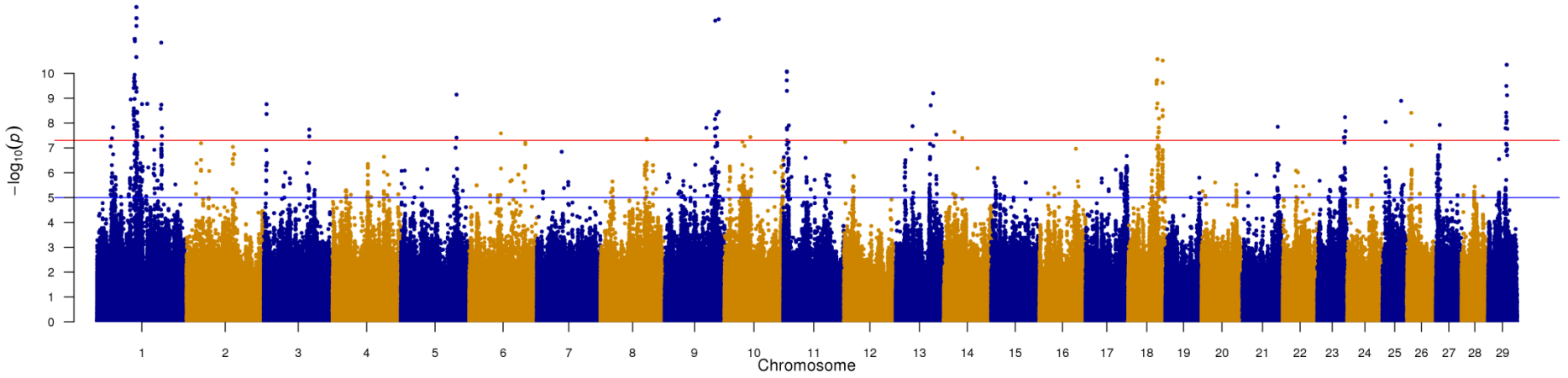


Figure 1. Manhattan plot of the sequence GWAS using the deregressed estimated breeding values for humoral response to *Mycobacterium avium* subspecies *paratuberculosis*. These animals had an effective record contribution ≥1 (n = 1,883). Red threshold line = genome-wide significance (p-value ≤ 5×10^{-8}), blue threshold line = suggestive SNPs (p-value ≤ 1×10^{-5}).

Table 1. Summary statistics for *Mycobacterium avium* subspecies *paratuberculosis* infection for each cohort investigated. SE = standard error, σ_a = additive genetic standard deviation, (h^2_L) = transformed heritability on the liability scale.

Cohort	No. animals	Prevalence (%)	σ_a	h^2 (SE)	h^2_L
Entire dataset[a]	136,767	0.04	0.027	0.02 (0.003)	0.1014
Entire dataset[s]	136,767 (7,105 sires)	0.04	0.114	0.02	0.1132
5 positive, 1 homebred[a]	33,818	0.077	0.058	0.05 (0.008)	0.1721
5 positive, 1 homebred[s]	33,818 (2,692 sires)	0.077	0.057	0.046	0.1572
Lab 1[a]	33,727	0.045	0.025	0.016 (0.005)	0.0770
Lab 2[a]	24,377	0.055	0.033	0.02 (0.008)	0.0937
Lab 3[a]	22,546	0.031	0.023	0.018 (0.006)	0.1141
Lab 4[a]	19,838	0.038	0.054	0.084 (0.013)	0.4536
Lab 5[a]	17,972	0.031	0.014	0.0063 (0.006)	0.0388
Lab 6[a]	5,433	0.055	0.021	0.009 (0.015)	0.0381
Lab 7[a]	10,197	0.029	0.018	0.012 (0.0095)	0.0731
Lab 8[a]	2,260	0.055	0.036	0.025 (0.032)	0.1039

[a] = animal model used, [s] = sire model used.

Table 2 Summary statistics and odds ratios (95% confidence interval in parenthesis) for animals stratified on estimated breeding values (EBV) for *Mycobacterium avium* subspecies *paratuberculosis*.

EBV Stratum	No. Animals	Prevalence (%)	EBV (SD)	Reliability (SD)	Simple Logistic Regression odds (95% CI)	Multiple Logistic Regression odds (95% CI)
Poor EBV	1812	0.16	0.019 (0.015)	0.15 (0.059)	1.43 (1.18,1.72)	1.37 (1.13,1.67)
Average EBV	1816	0.14	0.002 (0.009)	0.14 (0.062)	1.19 (0.98,1.44)	0 (1.44,0)
Good EBV	1849	0.12	-0.011 (0.009)	0.14 (0.063)	1	1

Table 3 Quantitative trait loci (QTL) identified as being associated with humoral response to *Mycobacterium avium* subspecies *paratuberculosis*. Note, for cases where >1 single nucleotide polymorphism (SNP) possessed the same p-value but was associated with the same positional candidate gene, the first (position-wise) SNP was reported. MAF = minor allele frequency of the favourable allele.

BTA	QTL start (bp)	QTL end (bp)	p-value for most sig SNP	Favourable Allele	Favourable allele frequency	Allele substitution effect	Significant SNP position	Annotation	Candidate Gene	No. Genes in QTL	No. significant SNPs in QTL
1	26167312	26446316	$4.129x10^{-8}$	C	0.005	0.007	26279218	intergenic		1	1
1	28407185	28494094	$1.483 x10^{-8}$	C	0.007	0.006	28407185	upstream	*ENSBTAG00000005101*	1	1
1	59604462	59612006	$1.124x10^{-9}$	G	0.006	0.008	59612006	intron	*ZBTB20*	1	1
1	64369225	69754155	$2.137x10^{-13}$	C	0.006	0.009	69624778	Intron	*KALRN*	60	51
1	69764180	69764180	$3.433x10^{-9}$	T	0.012	0.005	69764180	Intron	*UMPS*	1	1
1	70108608	71811570	$3.611x10^{-9}$	G	0.007	0.006	70854219	intron	*LRCH3*	27	65
1	79588937	79588937	$1.723x10^{-9}$	A	0.006	0.011	79588937	Intron	*LPP*	1	1
1	80447980	80447980	$3.651x10^{-8}$	C	0.007	0.005	80447980	intergenic	-	-	1
1	88842592	88847970	$1.679x10^{-9}$	G	0.006	0.010	88842592	intergenic	-	-	3
1	112715335	119581742	$5.777x10^{-12}$	A	0.006	0.012	113546452	intergenic	-	40	10
3	4670952	4670966	$1.735x10^{-9}$	T	0.006	0.010	4670952	intergenic	-	-	3
3	79626954	80275067	$1.814x10^{-8}$	A	0.003	0.997	80275067	intron	*DNAJC6*	5	2
5	96254052	99194836	$7.170x10^{-10}$	A	0.008	0.006	98598926	intergenic		28	2
6	55845383	55845383	$2.589X10^{-8}$	A	0.007	0.006	55845383	intergenic	-	-	1
8	81381959	81857714	$4.294x10^{-8}$	G	0.005	0.010	81782508	intergenic		2	2
9	73680357	73821094	$1.557x10^{-8}$	T	0.005	0.012	73726695	intergenic	-	-	1
9	87476619	90634355	$7.548x10^{-13}$	C	0.005	0.021	89425331	intergenic		27	2
9	91561782	91571541	$4.495x10^{-9}$	T	0.006	0.012	91571541	intergenic	-	-	1
9	92215914	92215914	$1.531x10^{-8}$	G	0.007	0.005	92215914	intergenic		-	1
9	95503522	95897049	$6.589x10^{-13}$	C	0.010	0.005	95619684	intron	*ZDHHC14*	3	2
10	44616944	48347572	$3.683x10^{-8}$	G	0.006	0.010	45898484	downstream	*SNX1*	35	1
11	5908958	5908958	$1.915x10^{-10}$	A	0.007	0.006	5908958	intron	*NPAS2*	1	1
11	6005215	6005215	$8.296 x10^{-11}$	C	0.008	0.005	6005215	downstream	*NPAS2*	1	1
11	6150881	6150890	$8.777 x10^{-11}$	T	0.008	0.005	6150881	intergenic		-	2
11	6187500	6241603	$1.568 x10^{-11}$	A	0.004	0.017	6227418	intron	*RNF149*	2	1
11	6327860	6327860	$4.972 x10^{-8}$	G	0.006	0.007	6327860	intron	*RFX8*	1	1

11	6333224	6333267	1.807×10^{-8}	G	0.003	0.030	6333255	intron	*RFX8*	1	2
11	9606899	9606899	1.246×10^{-8}	A	0.003	0.030	9606899	intron	*TACR1*	1	1
13	28234519	30909480	1.338×10^{-8}	A	-0.007	0.005	29951397	intergenic		19	1
13	62037755	62037755	1.942×10^{-9}	C	0.007	0.006	62037755	intron	*XKR7*	1	1
13	66373805	66373805	6.292×10^{-10}	T	0.006	0.008	66373805	intron	*SLA2*	1	1
13	70488028	72329336	2.910×10^{-8}	A	0.003	0.027	72023756	intergenic		12	1
14	16944837	21128294	2.276×10^{-8}	A	0.006	0.008	19711487	intron	*HAS2*	32	1
14	33563774	33601252	3.990×10^{-8}	T	-0.001	0.297	33588051	intron	*CPA6*	1	1
18	50390456	52265553	2.677×10^{-11}	G	0.007	0.009	51899478	intron	*TEX101*	58	28
18	53905719	61291660	3.048×10^{-11}	A	0.009	0.005	61212502	intron	*ENSBTAG00000040392*	243	7
21	62644820	62644820	1.414×10^{-8}	G	0.005	0.013	62644820	intergenic	-	-	1
23	46741649	46741649	3.792×10^{-8}	A	0.003	0.040	46741649	intergenic	-	-	1
23	48411393	48482001	3.669×10^{-8}	T	0.006	0.009	48460057	intergenic		1	1
23	48741897	48741897	5.799×10^{-9}	T	0.007	0.006	48741897	intron	*F13A1*	1	1
23	49577226	49659683	2.135×10^{-8}	T	0.006	0.009	49577226	intron	*CDYL*	1	1
25	5454814	5454814	9.015×10^{-9}	T	0.006	0.012	5454814	intergenic		-	1
25	32829869	32829869	1.277×10^{-9}	T	0.004	0.024	32829869	intergenic		-	1
26	7959610	7959610	3.889×10^{-9}	C	0.005	0.012	7959610	intron	*PRKG1*	1	1
27	6246011	6544550	1.185×10^{-8}	A	0.004	0.018	6544550	intergenic		2	1
29	31234645	31234645	1.608×10^{-8}	T	0.006	0.008	31234645	intergenic			1
29	32448786	34312188	4.508×10^{-11}	G	0.006	0.009	33280610	intergenic		11	10

Supplementary Table S1. All positional candidate genes residing within the 47 quantitative trait loci (QTL) that harbor the SNPs that passed the genome-wide significance threshold of 5 x 10^{-8}. Cases where there was no candidate genes identified are denoted with a dash (-).

BTA	QTL Start (bp)	QTL End (bp)	No. SNPs (in LD)	Genes
1	26167312	26446316	1(3)	*ROBO1*
1	28407185	28494094	1(1)	*ENSBTAG00000005101*
1	59604462	59612006	1(1)	*ZBTB20*
1	64369225	69754155	51(1038)	*IGSF11,ENSBTAG00000014157,UPK1B,B4GALT4,ARHGAP31,TMEM39A,POGLUT1,TIMMDC1,CD80,ADPRH,PLA1A,POPDC2, COX17,MAATS1,SNORA31,NR1I2,ENSBTAG00000040597,ENSBTAG00000048057,GPR156,LRRC58,FSTL1, ENSBTAG00000015892,HGD,RABL3,GTF2E1,STXBP5L,POLQ,ARGFX,FBXO40,HCLS1,GOLGB1,IQCB1,EAF2,SLC15A2,ENSBTAG0 0000001442,ILDR1,CD86,ENSBTAG00000003865,7SK,ENSBTAG00000014133,CSTA,CCDC58,KPNA1, PARP9, DTX3L, PARP14,HSPBAP1,DIRC2,SEMA5B,PDIA5,SEC22A,ADCY5,5S_rRNA,HACD2,ENSBTAG00000023608,MYLK,CCDC14,ENSBTAG000 00006947,KALRN,UMPS*
1	69764180	69764180	1(0)	*UMPS*
1	70108608	71811570	65(14543)	*HEG1,SLC12A8,ZNF148,SNX4,OSBPL11,ENSBTAG00000003383,LMLN,ENSBTAG00000014208,IQCG,LRCH3,FYTTD1,RUBCN,MUC 20,ENSBTAG00000014721,TNK2,TFRC,ZDHHC19,SLC51A,PCYT1A,ENSBTAG00000032684,ENSBTAG00000040161,UBXN7,ENSBTA G00000013317,SMCO1,WDR53,FBXO45,NRROS*
1	79588937	79588937	1(0)	*LPP*
1	80447980	80447980	1(0)	-
1	88842592	88847970	3(6)	-
1	112715335	119581742	10(719)	*SLC33A1,C3orf33,PLCH1,U6atac,SNORA70,MME,ENSBTAG00000024623,GPR149,DHX36,ARHGEF26,Metazoa_SRP,YWHAH,ENSB TAG00000047341,RAP2B,U6,P2RY1,ENSBTAG00000033740,MBNL1,U2,SUCNR1,AADAC,ENSBTAG00000013378,Y_RNA,ENSBTAG0 0000010385,ENSBTAG00000023447,IGSF10,MED12L,ENSBTAG00000011267,ENSBTAG00000038069,SIAH2,ERICH6,SELENOT,EIF 2A,SERP1,TSC22D2,PFN2,RNF13,ANKUB1,COMMD2,WWTR1*
3	4670952	4670966	3(6)	-
3	79626954	80275067	2(6)	*PDE4B,ENSBTAG00000030852,LEPR,LEPROT,DNAJC6*
5	96254052	99194836	2(70)	*GRIN2B,EMP1,GSG1,FAM234B,ENSBTAG00000005868,GPRC5D,GPRC5A,DDX47,APOLD1,ENSBTAG00000048294,CDKN1B,GPR1 9,CREBL2,DUSP16,BORCS5,ENSBTAG00000039496,MANSC1,LRP6,ENSBTAG00000030431,BCL2L14,ETV6,ENSBTAG00000030476, ENSBTAG00000030472,ENSBTAG00000030471,ENSBTAG00000001336,ENSBTAG00000023258,ENSBTAG00000030468,ENSBTAG00 000030466*

6	55845383	55845383	1(0)	-
8	81381959	81857714	2(8)	*GAS1, SNORA70*
9	73680357	73821094	1(0)	-
9	87476619	90674280	2(131)	*TAB2,ZC3H12D,PPIL4,GINM1,KATNA1,LATS1,NUP43,PCMT1,LRP11,RAET1G,ENSBTAG00000038891,ENSBTAG00000024751,ENSBTAG00000036061,ENSBTAG00000039875,PPP1R14C,IYD,PLEKHG1,MTHFD1L,AKAP12,bta-mir2285e2, ZBTB2, RMND1, ARMT1,TOMM7,CCDC170,ESR1,SYNE1*
9	91561782	91571541	1(0)	-
9	92215914	92215914	1(0)	-
9	95503522	95897049	2(1)	*ZDHHC14, 7SK, SNX9*
10	44616944	48347572	1(0)	*FRMD6,GNG2,C14orf166,NID2,ENSBTAG00000001423,PTGDR,PLEKHO2,PIF1,RBPMS2,OAZ2,ZNF609,TRIP4,ENSBTAG00000039462,CSNK1G1,PPIB,SNX22,SNX1,FAM96A,DAPK2,HERC1,FBXL22,USP3,CA12,APH1B,RAB8B,ENSBTAG00000012898,LACTB,TPM1, TLN2,bta-mir-190a,ENSBTAG00000020814,SNORA2,U6,ENSBTAG00000010952,VPS13C*
11	5908958	5908958	1(0)	*NPAS2*
11	6005215	6005215	1(0)	*NPAS2*
11	6150881	6150890	2(2)	-
11	6187500	6241603	1(0)	*CNOT11, RNF149*
11	6327860	6327860	1(0)	*RFX8*
11	6333224	6333267	2(6)	*RFX8*
11	9606899	9606899	1(0)	*TACR1*
13	28234519	30909480	1(0)	*ENSBTAG00000007700,SEPHS1,BEND7,ENSBTAG00000017244,FRMD4A,ENSBTAG00000010023,CDNF,HSPA14,SUV39H2,DCLRE1C,MEIG1,OLAH,ACBD7,RPP38,NMT2,FAM171A1,U6,ITGA8,FAM188A*
13	62037755	62037755	1(0)	*XKR7*
13	66373805	66373805	1(0)	*SLA2*
13	70488028	72329336	1(0)	*U6,ENSBTAG00000046785,PLCG1,ZHX3,ENSBTAG00000045671,LPIN3,EMILIN3,CHD6,bta-mir-544b-1, ENSBTAG00000048055, ENSBTAG00000046673,ENSBTAG00000000309*
14	16944837	21128294	1(0)	*MTSS1,NDUFB9,TATDN1,RNF139,TRMT12,ENSBTAG00000010813,TMEM65,FER1L6,FAM91A1,ENSBTAG00000047212,ANXA13,ENSBTAG00000045178,KLHL38,7SK,FBXO32,WDYHV1,ATAD2,ZHX1,ENSBTAG00000015319,FAM83A,TBC1D31,DERL1,ZHX2,ENSB*

				TAG00000043938,U6,HAS2,ENSBTAG00000009474,ENSBTAG00000040351,ENSBTAG00000046307,SPIDR,ENSBTAG00000017016,PRKDC
14	33563774	33601252	1(0)	*CPA6*
18	50390456	52265553	28(168)	*ENSBTAG00000047815,ENSBTAG00000003272,CYP2B6,ENSBTAG00000047196,CYP2S1,AXL,HNRNPUL1,CCDC97,TGFB1,B9D2,TMEM91,EXOSC5,BCKDHA,ATP5SL,ERICH4,CEACAM1,LIPE,CNFN,MEGF8,TMEM145,PRR19,ENSBTAG00000019785,ERF,ENSBTAG00000020756,ZNF526,DEDD2,SNORD112,POU2F2,ZNF574,GRIK5,ATP1A3,RABAC1,ENSBTAG00000048100,ARHGEF1,CD79A,ENSBTAG00000011963,DMRTC2,LYPD4,ENSBTAG00000006859,U6atac,CXCL17,CD177,TEX101,ENSBTAG00000003886,ENSBTAG00000018882,ENSBTAG00000023434,LYPD3,PHLDB3,ETHE1,ZNF575,XRCC1,PINLYP,IRGQ,ZNF428,CADM4,PLAUR,ENSBTAG00000023439,IRGC*
18	53905719	61291660	7(865)	*CCDC61,PGLYRP1,ENSBTAG00000023367,IGFL1,HIF3A,PPP5C,PNMAL1,ENSBTAG00000046451,CCDC8,ENSBTAG00000048021,ENSBTAG00000014583,PTGIR,GNG8,DACT3,PRKD2,STRN4,FKRP,SLC1A5,AP2S1,ARHGAP35,NPAS1,TMEM160,ZC3H4,SAE1,CCDC9,C5AR1,C5AR2,DHX34,MEIS3,SLC8A2,KPTN,NAPA,ZNF541,GLTSCR1,EHD2,GLTSCR2,SELENOW,CRX,ENSBTAG00000023419,ENSBTAG00000039971,ENSBTAG00000040054,ELSPBP1,CABP5,LIG1,C19orf68,ENSBTAG00000038322,ENSBTAG00000038986,ENSBTAG00000011119,ZNF114,CCDC114,EMP3,TMEM143,SYNGR4,KDELR1,GRIN2D,GRWD1,KCNJ14,CYTH2,LMTK3,SULT2B1,SPACA4,RPL18,SPHK2,DBP,CA11,NTN5,ENSBTAG00000014514,ENSBTAG00000045968,FUT2,MAMSTR,RASIP1,IZUMO1,FGF21,BCAT2,HSD17B14,PLEKHA4,PPP1R15A,TULP2,NUCB1,DHDH,BAX,ENSBTAG00000013343,GYS1,RUVBL2,ENSBTAG00000038735,NTF4,KCNA7,SNRNP70,LIN7B,PPFIA3,TRPM4,ENSBTAG00000030505,ENSBTAG00000030502,SLC6A16,ENSBTAG00000030544,CD37,TEAD2,DKKL1,CCDC155,PTH2,SLC17A7,PIH1D1,ALDH16A1,FLT3LG,ENSBTAG00000005296,RPS11,FCGRT,RCN3,NOSIP,PRRG2,PRR12,RRAS,SCAF1,BCL2L12,ENSBTAG00000006646,CPT1C,TSKS,AP2A1,FUZ,MED25,PTOV1,PNKP,AKT1S1,TBC1D17,IL4I1,ATF5,ENSBTAG00000016997,VRK3,ZNF473,IZUMO2,MYH14,KCNC3,NAPSA,ENSBTAG00000048283,NR1H2,POLD1,ENSBTAG00000043953,MYBPC2,FAM71E1,EMC10,JOSD2,ASPDH,LRRC4B,SYT3,C19orf81,SHANK1,CLEC11A,ACPT,ENSBTAG00000011079,KLK1,KLK15,KLK4,KLK5,KLK6,KLK7,KLK8,ENSBTAG00000040177,KLK10,KLK11,KLK12,KLK13,KLK14,ENSBTAG00000039212,CEACAM18,ENSBTAG00000004608,ENSBTAG00000037537,ENSBTAG00000030440,ENSBTAG00000047301,ENSBTAG00000037710,ENSBTAG00000037699,SIGLECL1,ENSBTAG00000038526,IGLON5,VSIG10L,ETFB,NKG7,LIM2,SIGLEC10,ENSBTAG00000008851,ENSBTAG00000047675,U6,ZNF175,ENSBTAG00000023365,ENSBTAG00000045880,ENSBTAG00000035868,ENSBTAG00000019227,bta-mir-99b,bta-let-7e,bta-mir-125a, SPACA6, HAS1, ENSBTAG00000039491,ENSBTAG00000014593,ENSBTAG00000039111,ZNF613,ZNF432,ZNF614,ENSBTAG00000037906,ENSBTAG00000010046,PPP2R1A,ENSBTAG00000046864,VN1R4,ENSBTAG00000046472,ENSBTAG00000047617,ENSBTAG00000046510,ENSBTAG00000047791,ENSBTAG00000047761,ENSBTAG00000038903,ENSBTAG00000030470,ENSBTAG00000038125,ZNF665,ENSBTAG00000011926,ENSBTAG00000033604,ENSBTAG00000045571,ENSBTAG00000037440,ENSBTAG00000047712,ENSBTAG00000030454,ENSBTAG00000011052,bta-mir-2887-1,ENSBTAG00000040411,ENSBTAG00000021433,ENSBTAG00000015899,ENSBTAG00000011844,ENSBTAG00000047053,ENSBTAG00000045985,ENSBTAG00000004925,ZNF677,ENSBTAG00000039969,ENSBTAG00000017993,ENSBTAG00000030444,ZNF23,ENSBTAG00000034090,ENSBTAG00000045581,ENSBTAG00000033642,ENSBTAG00000038755,ENSBTAG00000046403,ZNF331,ENSBTAG00000015061,NLRP12,bta-mir-371,ENSBTAG00000040392,ENSBTAG00000000336,ENSBTAG00000046961*
21	62644820	62644820	1(0)	-
23	46741649	46741649	1(0)	-
23	48411393	48482001	1(0)	*LY86*

23	48741897	48741897	1(0)	*F13A1*
23	49577226	49659683	1(0)	*CDYL*
25	5454814	5454814	1(0)	-
25	32829869	32829869	1(0)	-
26	7959610	7959610	1(0)	*PRKG1*
27	6246011	6544550	1(0)	*GPM6A, ENSBTAG00000046071*
29	31234645	31234645	1(0)	-
29	32448786	34312188	10(55)	*FLI1,KCNJ1,KCNJ5,ENSBTAG00000024731,ARHGAP32,BARX2,SNORD112,JAM3,ENSBTAG00000045650,IGSF9B,SPATA19*

967

968

969

970

971 **Supplementary Table S2** Bovine biological pathways with an unadjusted p-value <0.05 identified (n = 29). Pathway Id relates to the
972 corresponding Reactome pathway ID.

Pathway Name	Pathway Id	Gene Count	Pathway p-value	Corrected p-value	Gene Symbols
Antigen activates B Cell Receptor (BCR) leading to generation of second messengers	19637	2	0.004222589	0.320916736	*CALM; CD79A;*
ADP signalling through P2Y purinoceptor 1	18736	2	0.006249223	0.158313652	*GNG2; P2RY1;*
Activation of Kainate Receptors upon glutamate binding	16714	2	0.006249223	0.158313652	*CALM; GNG2;*
Signal amplification	18462	2	0.011356063	0.215765188	*GNG2; P2RY1;*
Eukaryotic Translation Termination	19690	5	0.013724741	0.208616057	*RPL13A; RPL18; RPL35A; RPS11; RPS19;*
Peptide chain elongation	17599	5	0.014553831	0.184348521	*RPL13A; RPL18; RPL35A; RPS11; RPS19;*
Nonsense Mediated Decay (NMD) independent of the Exon Junction Complex (EJC)	17994	5	0.015416329	0.167377286	*RPL13A; RPL18; RPL35A; RPS11; RPS19;*
Eukaryotic Translation Elongation	19898	5	0.016312806	0.154971658	*RPL13A; RPL18; RPL35A; RPS11; RPS19;*
Formation of a pool of free 40S subunits	18106	5	0.020249388	0.170994831	*RPL13A; RPL18; RPL35A; RPS11; RPS19;*
Nonsense Mediated Decay (NMD) enhanced by the Exon Junction Complex (EJC)	17605	5	0.021323786	0.147327973	*RPL13A; RPL18; RPL35A; RPS11; RPS19;*
Nonsense-Mediated Decay (NMD)	19399	5	0.021323786	0.147327973	*RPL13A; RPL18; RPL35A; RPS11; RPS19;*
Metabolism of nitric oxide	17490	2	0.021428562	0.125274669	*CALM; NOSIP;*
eNOS activation and regulation	17377	2	0.021428562	0.125274669	*CALM; NOSIP;*
Transmission across Chemical Synapses	17968	4	0.025563686	0.138774297	*AP2S1; CALM; GNG2; SLC17A7;*
SRP-dependent cotranslational protein targeting to membrane	16985	5	0.025996652	0.13171637	*RPL13A; RPL18; RPL35A; RPS11; RPS19;*
Metabolism of proteins	18621	10	0.028395646	0.13487932	*LHB; PDIA5; PLAUR; RPL13A; RPL18; RPL35A; RPS11; RPS19; SAE1; SERP1;*
GTP hydrolysis and joining of the 60S ribosomal subunit	18494	5	0.028564244	0.127698973	*RPL13A; RPL18; RPL35A; RPS11; RPS19;*
Platelet homeostasis	18659	2	0.029589522	0.124933536	*GNG2; PRKG1;*
L13a-mediated translational silencing of Ceruloplasmin expression	17580	5	0.029907203	0.119628811	*RPL13A; RPL18; RPL35A; RPS11; RPS19;*
Ca2+ pathway	18159	2	0.034064589	0.117677671	*CALM; GNG2;*

IRE1alpha activates chaperones	18497	2	0.034064589	0.117677671	*PDIA5; SERP1;*
XBP1(S) activates chaperone genes	18959	2	0.034064589	0.117677671	*PDIA5; SERP1;*
Cap-dependent Translation Initiation	17876	5	0.034176885	0.108226802	*RPL13A; RPL18; RPL35A; RPS11; RPS19;*
Eukaryotic Translation Initiation	17853	5	0.034176885	0.108226802	*RPL13A; RPL18; RPL35A; RPS11; RPS19;*
Hemostasis	17834	7	0.03588147	0.109079669	*CALM; F13A1; GNG2; LOC512741; P2RY1; PLAUR; PRKG1;*
Inactivation, recovery and regulation of the phototransduction cascade	19685	2	0.038786223	0.113375113	*CALM; NMT2;*
Neuronal System	18958	4	0.042031414	0.118310647	*AP2S1; CALM; GNG2; SLC17A7;*
The phototransduction cascade	17305	2	0.043742531	0.118729728	*CALM; NMT2;*
Neurotransmitter Receptor Binding And Downstream Transmission In The Postsynaptic Cell	18514	3	0.04473712	0.117242108	*AP2S1; CALM; GNG2;*

973

Supplementary Table S3 Biological pathways using human ortholog data with an unadjusted p-value <0.05 identified (n = 64). Pathway Id relates to the corresponding Reactome pathway ID.

Pathway Name	Pathway Id	Gene Count	Pathway p-value	Corrected p-value	Gene Symbols
Inwardly rectifying K+ channels	18309	5	6.39E-04	0.308173259	*BT.62324; GNG2; GNG8; KCNJ1; KCNJ5;*
Neuronal System	17312	14	0.002257084	0.543957324	*ADCY5; AP2A1; BT.62324; GNG2; GNG8; GRIK5; GRIN2D; KCNA7; KCNC3; KCNJ1; KCNJ5; PPFIA3; RRAS; SLC17A7;*
AKT phosphorylates targets in the cytosol	12988	3	0.002500313	0.301287698	*BT.52156; CDKN1B; GSK3A;*
DNA-PK pathway in nonhomologous end joining	15829	3	0.002500313	0.301287698	*DCLRE1C; PNKP; PRKDC;*
Thromboxane A2 receptor signaling	14941	4	0.00266897	0.257288669	*ARHGEF1; BT.45031; GNG2; PTGDR;*
Hyaluronan biosynthesis and export	13620	2	0.003284616	0.263864138	*BT.63577; HAS1;*
Golgi Associated Vesicle Biogenesis	13839	5	0.004965629	0.341918993	*BT.65042; FTL; NAPA; SNX9; TFRC;*
G alpha (s) signalling events	13224	8	0.005154194	0.310540179	*ADCY5; BT.45031; GNG2; GNG8; LHB; PDE4B; PTGDR; PTH2;*
Nef Mediated CD8 Down-regulation	13682	2	0.00538863	0.288591092	*AP2A1; AP2S1;*
Arf1 pathway	15337	3	0.005933294	0.285984751	*AP2A1; CYTH2; KDELR1;*
Potassium Channels	18590	7	0.00737954	0.323358022	*BT.62324; GNG2; GNG8; KCNA7; KCNC3; KCNJ1; KCNJ5;*
Neurotransmitter Receptor Binding And Downstream Transmission In The Postsynaptic Cell	19389	8	0.00738831	0.296763794	*ADCY5; AP2A1; GNG2; GNG8; GRIK5; GRIN2D; KCNJ5; RRAS;*
Transmission across Chemical Synapses	18918	10	0.007610675	0.282180407	*ADCY5; AP2A1; GNG2; GNG8; GRIK5; GRIN2D; KCNJ5; PPFIA3; RRAS; SLC17A7;*
FOXM1 transcription factor network	15700	4	0.008284666	0.285229223	*BT.32558; GAS1; GSK3A; XRCC1;*
Base Excision Repair	18332	3	0.008348716	0.251505076	*LIG1; POLD1; XRCC1;*
Resolution of Abasic Sites (AP sites)	17264	3	0.008348716	0.251505076	*LIG1; POLD1; XRCC1;*
Clathrin derived vesicle budding	13840	5	0.008958367	0.239885156	*BT.65042; FTL; NAPA; SNX9; TFRC;*
trans-Golgi Network Vesicle Budding	16988	5	0.008958367	0.239885156	*BT.65042; FTL; NAPA; SNX9; TFRC;*
Prostacyclin signalling through prostacyclin receptor	13104	3	0.009742664	0.247156012	*BT.45031; GNG2; GNG8;*
Sema4D mediated inhibition of cell attachment and migration	13907	2	0.010965481	0.264268088	*ARHGAP35; RRAS;*
Activation of GABAB receptors	17880	4	0.012251089	0.268410216	*ADCY5; GNG2; GNG8; KCNJ5;*
GABA B receptor activation	13345	4	0.012251089	0.268410216	*ADCY5; GNG2; GNG8; KCNJ5;*
Glycosphingolipid biosynthesis	552	3	0.012914624	0.270645607	*B4GALT4; BT.21655; FUT1;*

Influenza Infection	16853	7	0.013740425	0.275953543	*KPNA1; NUP43; RPL13A; RPL35A; RPS11; RPS19; TGFB1;*
Nef Mediated CD4 Down-regulation	13684	2	0.014393004	0.277497116	*AP2A1; AP2S1;*
Ctcf: first multivalent nuclear factor	4040	3	0.01469612	0.272443457	*CD79A; CDKN1B; TGFB1;*
Class I PI3K signaling events mediated by Akt	15022	3	0.016609692	0.285923987	*CDKN1B; KPNA1; PRKDC;*
PAR1-mediated thrombin signaling events	15229	3	0.016609692	0.285923987	*ARHGEF1; GNG2; SNX1;*
Prostanoid ligand receptors	13236	2	0.018217734	0.302791305	*BT.45031; PTGDR;*
ADP signalling through P2Y purinoceptor 1	13119	3	0.018656228	0.272493999	*GNG2; GNG8; P2RY1;*
Activation of G protein gated Potassium channels	13351	3	0.018656228	0.272493999	*GNG2; GNG8; KCNJ5;*
G protein gated Potassium channels	17648	3	0.018656228	0.272493999	*GNG2; GNG8; KCNJ5;*
Inhibition of voltage gated Ca2+ channels via Gbeta/gamma subunits	13344	3	0.018656228	0.272493999	*GNG2; GNG8; KCNJ5;*
Sema4D in semaphorin signaling	13909	3	0.020836309	0.29538532	*ARHGAP35; MYH14; RRAS;*
Retrograde neurotrophin signalling	13194	2	0.022419003	0.308741692	*AP2A1; AP2S1;*
Adrenaline,noradrenaline inhibits insulin secretion	13553	3	0.023150227	0.309955819	*ADCY5; GNG2; GNG8;*
Activation of Kainate Receptors upon glutamate binding	18335	3	0.025598005	0.324690486	*GNG2; GNG8; GRIK5;*
Extension of Telomeres	19607	3	0.025598005	0.324690486	*LIG1; POLD1; RUVBL2;*
Processive synthesis on the C-strand of the telomere	12917	2	0.02697689	0.333406698	*LIG1; POLD1;*
Opioid Signalling	13260	5	0.029702303	0.35791275	*ADCY5; GNG2; GNG8; PDE4B; PPP2R1A;*
GABA receptor activation	17604	4	0.030698604	0.360895788	*ADCY5; GNG2; GNG8; KCNJ5;*
Platelet homeostasis	13108	5	0.031261582	0.358763874	*BT.45031; GNG2; GNG8; PPP2R1A; SLC8A2;*
Carm1 and regulation of the estrogen receptor	4134	2	0.031872206	0.349145529	*BT.32558; GTF2E1;*
Non-homologous end-joining	2800	2	0.031872206	0.349145529	*DCLRE1C; PRKDC;*
Glucagon signaling in metabolic regulation	13556	3	0.036719604	0.384757592	*ADCY5; GNG2; GNG8;*
Signal amplification	19554	3	0.036719604	0.384757592	*GNG2; GNG8; P2RY1;*
Syndecan-1-mediated signaling events	15809	2	0.037086461	0.372409876	*PPIB; TGFB1;*
WNT5A-dependent internalization of FZD4	16977	2	0.037086461	0.372409876	*AP2A1; AP2S1;*

Influenza Life Cycle	19285	6	0.037542603	0.369296625	*KPNA1; NUP43; RPL13A; RPL35A; RPS11; RPS19;*
Base excision repair	2819	3	0.039828675	0.383948427	*LIG1; POLD1; XRCC1;*
Canonical Wnt signaling pathway	15524	2	0.042601847	0.354036042	*CSNK1G1; LRP6;*
Eicosanoid ligand-binding receptors	13238	2	0.042601847	0.354036042	*BT.45031; PTGDR;*
Glycosphingolipid biosynthesis	565	2	0.042601847	0.354036042	*BT.21655; FUT1;*
Hyaluronan metabolism	19552	2	0.042601847	0.354036042	*BT.63577; HAS1;*
Mismatch repair (MMR) directed by MSH2:MSH3 (MutSbeta)	19229	2	0.042601847	0.354036042	*LIG1; POLD1;*
Mismatch repair (MMR) directed by MSH2:MSH6 (MutSalpha)	18261	2	0.042601847	0.354036042	*LIG1; POLD1;*
Processive synthesis on the lagging strand	12950	2	0.042601847	0.354036042	*LIG1; POLD1;*
Termination of O-glycan biosynthesis	13367	2	0.042601847	0.354036042	*MUC20; MUC4;*
Telomere Maintenance	17962	3	0.043066945	0.35183504	*LIG1; POLD1; RUVBL2;*
Nonsense Mediated Decay (NMD) enhanced by the Exon Junction Complex (EJC)	13341	5	0.045581053	0.36016504	*PPP2R1A; RPL13A; RPL35A; RPS11; RPS19;*
Nonsense-Mediated Decay (NMD)	16937	5	0.045581053	0.36016504	*PPP2R1A; RPL13A; RPL35A; RPS11; RPS19;*
NGF signalling via TRKA from the plasma membrane	18753	8	0.045937984	0.35713078	*ADCY5; AP2A1; AP2S1; BT.52156; CDKN1B; GSK3A; PLCG1; PPP2R1A;*
Mismatch Repair	19342	2	0.048401218	0.370307735	*LIG1; POLD1;*
Signaling by EGFR	13216	7	0.049612884	0.373647035	*ADCY5; AP2A1; AP2S1; BT.52156; CDKN1B; GSK3A; PLCG1;*

976

977