

RESEARCH

Open Access



Summary-data based Mendelian randomization identifies gene expression regulatory polymorphisms associated with bovine paratuberculosis by modulation of the *nuclear factor Kappa β* (*NF- κ B*)-mediated inflammatory response

Gerard Badia-Bringué^{1,2}, Maria Canive¹, Nora Fernandez-Jimenez³, José Luis Lavín⁴, Rosa Casais⁵, Cristina Blanco-Vázquez⁵, Patricia Vázquez¹, Almudena Fernández⁶, Jose Ramón Bilbao³, Joseba M. Garrido¹, Ramón A. Juste¹, Oscar González-Recio⁶ and Marta Alonso-Hearn^{1*}

Abstract

Genome-wide association studies (GWAS) have identified host genetic variants associated with paratuberculosis (PTB) susceptibility. Most of the GWAS-identified SNPs are in non-coding regions. Connecting these non-coding variants and downstream affected genes is a challenge and, up to date, only a few functional mutations or expression quantitative loci (cis-eQTLs) associated with PTB susceptibility have been identified. In the current study, the associations between imputed whole-genome sequence genotypes and whole RNA-Sequencing data from peripheral blood (PB) and ileocecal valve (ICV) samples of Spanish Holstein cows ($N = 16$) were analyzed with *TensorQTL*. This approach allowed the identification of 88 and 37 cis-eQTLs regulating the expression levels of 90 and 37 genes in PB and ICV samples, respectively (False discovery rate, $FDR \leq 0.05$). Next, we applied summary-based data Mendelian randomization (SMR) to integrate the cis-eQTL dataset with GWAS data obtained from a cohort of 813 culled cattle that were classified according to the presence or absence of PTB-associated histopathological lesions in gut tissues. After multiple testing corrections ($FDR \leq 0.05$), we identified two novel cis-eQTLs affecting the expression of the *early growth response factor 4* (*EGR4*) and the bovine *neuroblastoma breakpoint family member 6-like protein isoform 2* (*MGC134040*) that showed pleiotropic associations with the presence of multifocal and diffuse lesions in gut tissues; $P = 0.002$ and $P = 0.017$, respectively. While *EGR4* acts as a brake on T-cell proliferation and cytokine production through interaction with the *nuclear factor Kappa β* (*NF- κ B*), *MGC134040* is a target gene of *NF- κ B*. Our findings provide a better understanding of the genetic factors influencing PTB outcomes, confirm that the multifocal lesions are localized/confined lesions that have different underlying host genetics than the diffuse lesions, and highlight regulatory SNPs and regulated-gene targets to design future functional studies.

*Correspondence:

Marta Alonso-Hearn
malonso@neiker.eus

Full list of author information is available at the end of the article



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Keywords Summary-data based Mendelian Randomization, GWAS, eQTL, Pleiotropy, Paratuberculosis, Holstein cattle, NF- κ B

Background

Bovine paratuberculosis (PTB) or Johne's disease is caused by *Mycobacterium avium* subsp. *paratuberculosis* (MAP) [1, 2]. PTB is a granulomatous enteritis that affects ruminants worldwide and must be notified to the World Organization for Animal Health. The economic burden of PTB on the dairy industry relates mainly to decreased milk production and premature killing of infected animals [3]. Each year, an estimated US\$198 million is lost due to PTB in the United States, US\$364 million in the European Union, US\$75 million in Germany, US\$56 million in France, and US\$12 million in Spain [4]. Infection usually occurs through the fecal–oral route at an early stage of life and can remain subclinical for years. In the jejunal-ileal Peyer's patches, MAP gains entry to the intestinal mucosa by interacting with M and epithelial cells [5–7]. MAP can survive within infected macrophages by inhibiting apoptosis and phagosome acidification, and by preventing the presentation of antigens to the immune system [8]. As the infection progresses, the lesions in the intestine and lymph nodes become more severe and the cellular infiltrate becomes diffuse, disrupting the mucosal structure and affecting the jejunum and ileum [9, 10]. Moreover, there is evidence suggesting that MAP infection is associated with human inflammatory bowel diseases (IBD), autoimmune diseases, as well as colorectal cancer and Alzheimer's disease [11–13]. Colorectal cancer is a complication of two forms of idiopathic IBD; colonic Crohn's disease (CD) and ulcerative colitis. Interestingly, MAP bacilli have been detected in the intestines of patients with CD, ulcerative colitis, and IBD-associated colorectal cancer [12–14].

Currently, there is not any effective treatment to cure bovine PTB, and the parenteral vaccination with heat-killed inactivated vaccines is not widely accepted by animal health authorities on the grounds of a slight interference with the diagnosis of bovine tuberculosis [15]. The identification and selection of animals naturally less susceptible to MAP infection or with increased resistance to PTB is important for disease control and breeding purposes. Genome-wide association studies (GWAS) have allowed the identification of single-nucleotide polymorphisms (SNPs) associated with PTB susceptibility, resistance, and tolerance using whole-genome sequence data (WGS) [16–25]. Although previous GWAS identified loci associated with MAP tissue infection assessed by PCR

and culture [26–28], we provided the first comparison of the genetic variants associated with three distinct PTB-associated lesions in gut tissues (focal, multifocal, and diffuse) in Spanish Holstein cattle ($N=813$) [19]. Most of the GWAS-identified SNPs are located in non-coding regions of the genome, including intergenic and intronic regions that are enriched in regulatory elements indicating that those variants exert their effects through the modulation of gene expression [29]. Connecting non-coding variants and downstream affected genes is a challenge and, up to date, only a few functional mutations or expression quantitative loci (eQTLs) with causative effects over PTB susceptibility or resistance have been characterized [30–32]. A common practice following GWAS is to map genes near the identified SNPs based on haplotype and linkage disequilibrium. However, this approach does not take into account that cis-eQTLs can be located within 1 Mb upstream of a gene transcription start site (TSS) and modulate the transcription of distant target genes [33]. Another approach for mapping cis-eQTLs-affected genes consists in finding statistically significant associations between gene expression levels and genetic variants within 1 Mb upstream of a TSS, commonly performed by linear regression [34].

In our previous study, *Matrix eQTL* [35] was used to integrate gene expression data that were mapped to the *Bos Taurus* reference genome UMD3.1 v87 with TopHat 2.1.1 [36], and genotypes obtained with the EuroG Medium Density (MD) Bead Chip of Illumina (54,609 SNPs). This approach allowed the identification of 192 and 48 cis-eQTLs associated with the expression of 145 and 43 genes in bovine peripheral blood (PB) and ileocecal valve (ICV) samples, respectively [30]. The association between these cis-eQTLs and PTB susceptibility was addressed by performing a case–control association analysis using the genotypes of the identified cis-eQTLs and phenotypical data from 986 culled cows. This approach allowed the identification of three cis-eQTLs associated with PTB susceptibility and with the up-regulation of the *MDS1* and *EVII complex (MECOM)*, *eukaryotic elongation factor 1- α 2 (eEF1A2)*, and *U1 spliceosomal* mRNA expression. Although this study was the first in providing insights into the role of cis-eQTLs in gene transcription regulation and PTB susceptibility, it was not performed at the WGS level.

Cis-eQTLs can alter mRNA expression leading to changes in the level, timing, and or localization of gene

expression which can significantly cause variations in individual phenotypes. If the expression of a gene governed by a cis-eQTL influences a complex trait, such as a disease outcome, differences in gene expression levels between individuals will translate into differences in the disease outcome. Combined genetic-transcriptomic approaches like Mendelian Randomization (MR) allow the identification of cis-eQTLs that lead to manifestations of complex diseases or disease outcomes due to genetically regulated transcriptional activity [37]. In MR, a genetic variant is used to test the causative effect of an exposure (gene expression) on an outcome (disease outcome). However, phenotypes, WGS-derived genotypes, and gene expression data measured in a large sample size are rarely available. For this reason, Zhu et al. proposed a method called summary-data-based MR (SMR), which integrates summary-level data from an independent GWAS with data from an eQTL study to identify genes whose expression levels are associated with a complex trait due to pleiotropy, defined as “association between a trait and gene expression due to either pleiotropy (both gene expression and the trait are affected by the same causal variant) or causality (the effect of a causal variant on the trait

is mediated by gene expression)” [38]. Identifying genes and genetic markers with causative effects over disease outcomes or resilience will help improve breeding progress in livestock.

The main goal of this study was to identify candidate genes with pleiotropic association with cis-eQTLs and PTB-associated lesions using an SMR-based approach. First, we mapped cis-eQTLs significantly associated with gene expression changes in PB (N=15) and ICV (N=15) samples from 16 Holstein cows using *Tensor QTL*. For cis-eQTLs mapping, WGS-derived genotypes and RNA-Sequencing (RNA-Seq) data from ICV and PB were used [39]. Next, to test whether the pleiotropic association between these cis-eQTLs and PTB outcomes existed, an SMR analysis integrating expression data from PB and ICV datasets and three previously generated GWAS datasets from a cohort of 813 culled cows was performed. The three GWAS consisted in case–control studies in which the animals with a specific PTB-associated lesion (focal, multifocal, or diffuse) were compared with control cows without lesions in gut tissues [19]. The workflow of the study is presented in Fig. 1.

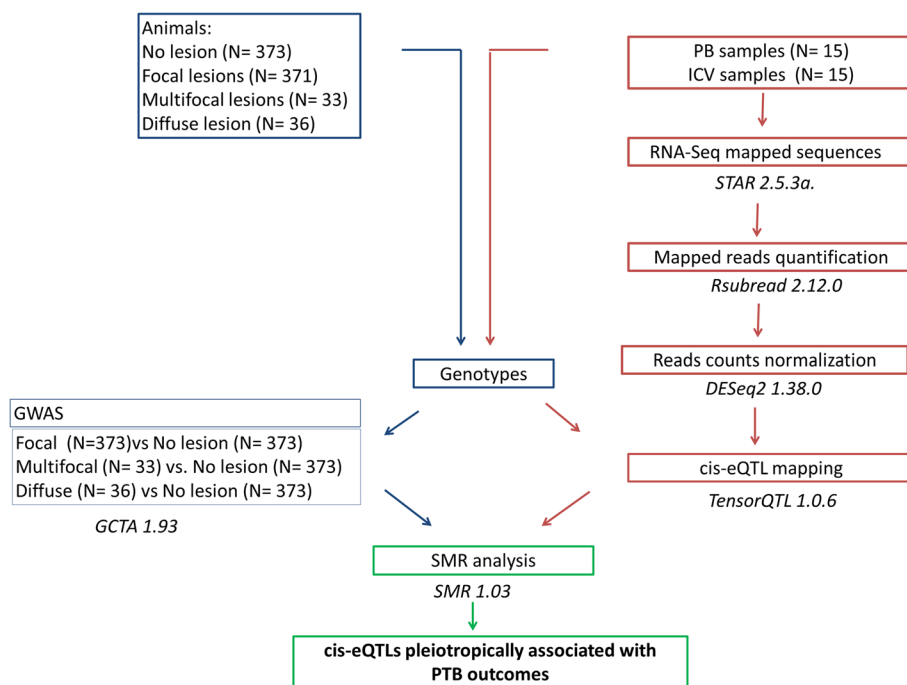


Fig. 1 Study design. Summary data from GWAS and cis-eQTL mapping were used to perform the SMR analysis. The approach starts with the individual normalized RNA-counts, then identifies genetic variants controlling differences in gene expression, and finally checks whether the identified cis-eQTLs are indeed associated with disease outcomes using summary GWAS data from a bigger population classified according to the presence or absence of PTB-associated lesions. Two eQTL datasets from peripheral blood (PB) and ileocecal valve (ICV) were applied together with three GWAS datasets from the comparisons: focal, multifocal or diffuse lesions versus (vs) controls. Using these two different omics data, the SMR method discover biological mechanisms behind PTB outcomes

Materials and methods

Animals and PTB diagnosis

PB ($N=15$) and ICV ($N=15$) samples used in the RNA-Seq analysis were previously collected from 16 Holstein cows from a single farm in Asturias (Spain) at the time of slaughter as previously described (Table 1) [39]. Animals were 18 months old or older (mean: 4.6 years old). PB from cow ID13 and the ICV from cow ID16 were not available and were not included in the study. The infectious status of these animals was determined by histopathological analysis of gut tissues, ELISA for the detection of anti-MAP antibodies, and fecal and gut tissues PCR and bacteriological culture as previously described [40]. On the other hand, 813 culled Holstein cattle from eight different regions of Spain were used in a previous GWAS analysis [19]. Histopathological analysis of gut tissues collected from these animals was previously performed [41]. For the GWAS, cases were animals with focal ($N=371$), multifocal ($N=33$), or diffuse lesions ($N=36$), while controls ($N=373$) did not have lesions in gut tissues and had a negative ELISA, PCR, and bacteriological culture of gut tissues at the time of slaughter. The average age of the animals without lesions and with multifocal and diffuse lesions was 5.45, 5.09, and 4.38 years old, respectively.

Genotyping and imputation

PB samples from the animals included in the cis-eQTL mapping and GWAS analyses were previously collected

into 10 ml Vacutainer EDTA tubes (BD, Franklin Lakes, USA) for genotyping [19, 30]. Briefly, total DNA was extracted from PB samples using the QIAamp DNA Blood Mini Kit according to the manufacturer's instructions (Qiagen, Hilden, Germany). Purified DNA was then quantified by spectrophotometry and genotyped using the EuroG medium density (MD) Bead Chip of Illumina at the molecular genetic laboratory service of the Spanish Federation of Holstein Cattle (CONAFE) using the *InfiniumTM iScan* software for allele assignment (Illumina, San Diego, USA). Individual genotypes were imputed to WGS as previously described [18]. Briefly, genotypes were phased using *Eagle 2.4* [42] and imputed with *minimac4* [43] to the Bovine High-Density Bead Chip using a reference panel of 1,278 *Bos taurus* from Run7.0 of the 1,000 Bull Genomes project and 581,712 SNPs (ARS-UCD1.2). In this step, the Run 7 population was restricted to Holstein animals. Imputation to the WGS level was then undertaken using a reference population of 2,333 *Bos taurus* from Run7.0 of the 1,000 Bull Genomes project [44]. In total, 33.77 million variants per animal, including small insertions and deletions, were obtained. SNPs with a minimum allele frequency (MAF) < 0.01 and an imputation score (R^2) < 0.7 were excluded. After applying these two filters, all the filtered SNPs had a call rate > 95%. After applying these quality parameters, 12.38 million and 13.88 million SNPs passed the filters for the cis-eQTL mapping and GWAS populations, respectively.

Table 1 Histopathological analysis, ELISA, PCR, and bacteriological culture results from all the animals included in the current study

Animal ID	Histopathological analysis			ELISA	Fecal PCR	Fecal Culture	Tissue PCR	Tissue Culture
	Microscopic	Macroscopic	ZN					
1	Negative	No	Negative	Negative	Negative	Negative	Negative	Negative
2	Negative	No	Negative	Negative	Negative	Negative	Negative	Negative
3	Negative	No	Negative	Negative	Negative	Negative	Negative	Negative
15	Negative	No	Negative	Negative	Negative	Negative	Negative	Negative
4	Focal	No	Negative	Negative	Negative	Negative	Positive	Negative
5	Focal	No	Negative	Negative	Negative	Negative	Positive	Low
6	Focal	No	Negative	Negative	Negative	Negative	Positive	Negative
7	Focal	No	Negative	Negative	Negative	Negative	Positive	Negative
8	Focal	No	Negative	Positive	Positive	Negative	Positive	Medium
9	Multifocal	Yes	Positive	Negative	Negative	Negative	Negative	Negative
16	Multifocal	No	Positive	Positive	Positive	Heavy	Positive	Heavy
10	Diffuse	Yes	Positive	Positive	Positive	Negative	Positive	Heavy
11	Diffuse	Yes	Positive	Positive	Positive	Negative	Positive	Heavy
12	Diffuse	Yes	Positive	Positive	Positive	Heavy	Positive	Heavy
13	Diffuse	Yes	Positive	Positive	Positive	Negative	Positive	Low
14	Diffuse	Yes	Positive	Positive	Positive	Negative	Positive	Heavy

Bacterial load was classified as low (< 10 CFU), medium (between 10 to 50 CFU) or heavy (> 50 CFU)

ZN Ziehl–Neelsen, CFU Colony Forming Units

Gene expression data

Total RNA was previously isolated from PB ($N=15$) and ICV ($N=15$) samples of 16 Holstein cows from a single farm in Asturias (Spain) at the time of slaughter [39]. PB from cow ID13 and the ICV from cow ID16 were not available and were not included. RNA-Seq libraries were generated and then single-end sequenced in a 1×75 bp format using the Illumina NextSeq 500 sequencer at the Genomic Unit of the Scientific Park of Madrid, Spain. Raw reads were filtered by length (minimum size 75 bp long) and percentage of ambiguous bases less than 10% using *Prinseq-lite* 0.20.4 [45]. In the current study, reads from our previous RNA-Seq study [39] were mapped to the most recent *Bos taurus* reference genome (ARS-UCD1.2.105) with *STAR (Spliced Transcripts Alignment to a Reference)* 2.5.3a, an ultrafast RNA-Seq aligner capable of mapping full-length RNA sequences [46]. The reads mapped 27,607 genes. Alignment.sam files were converted to bam files using *Samtools* 1.13 [47]. The number of reads for each gene was counted using the function “feature counts” of *Rsubread* 2.12.0 [48] and normalized with the mean-of-ratios method included in the *DESeq2* 1.38.0 software [49].

Cis-eQTL mapping

For cis-eQTL mapping, the gene expression data (normalized counts of the 27,607 mapped genes) from the PB and ICV samples of Holstein cows from a single farm in Asturias (Spain) and corresponding WGS-derived genotypes (12,377,073 SNPs) were used to run *Tensor QTL*, which uses a fast permutation scheme that relies on the β -distribution to compute P_β -values [50, 51]. Significant associations between cis-eQTLs located within 1 Mb upstream of a transcription start site (TSS) and normalized gene counts were detected. P_β -values were corrected for multiple testing corrections with the Benjamini–Hochberg (BH) method [52] using the *R* p.adjust (pvalues, method=“fdr”) function [53]. Age was included as a covariate in the analysis.

Genome-wide association studies (GWAS)

Associations between the imputed genotypes (13,881,067 SNPs) and the absence ($N=373$) or presence of focal ($N=371$), multifocal ($N=33$), or diffuse ($N=36$) lesions were previously analyzed in a case–control study ($N=813$) using the mixed linear model association (mlma) analysis of the *GCTA* 1.93.2 software [19]. The model is defined as $y = Xb + g + e$ where \mathbf{g} is the vector of polygenic effects and the relationship matrix is a genomic relationship matrix, with the usual notation (G) [54]. More specifically, \mathbf{y} is a vector of phenotypes of length equal to n , which is

the number of animals, \mathbf{X} is a $n \times p$ matrix of the variables with fixed effects coded as either 0, 1 or 2, with p being the number of SNPs, \mathbf{b} is a vector of the additive fixed effects of each SNP with a length of n , \mathbf{g} is a vector of length n with a distribution $\sim N(0, \mathbf{A} \sigma_g^2)$, where \mathbf{A} is an $N \times N$ genetic relationship matrix between individuals, and σ_g^2 is the variance explained by all the SNPs included in the model, and \mathbf{e} is a vector of residual effects with length n with $\mathbf{e} \sim N(0, \mathbf{I} \sigma_e^2)$, where \mathbf{I} is an $n \times n$ identity matrix and σ_e^2 is any variance not explained by the SNPs included in the model [55]. Age was included as a covariate in the analysis.

SMR analysis

Both the GWAS (.mlma file) and the cis-eQTL mapping summary data (.txt file) were used to perform the SMR analysis using the *SMR* 1.03 software [38]. Cis-eQTLs with nominal P -values of more than 5×10^{-8} , and/or with differences in allele frequency between the populations of the GWAS and eQTL summary data larger than 0.2 were excluded according to Zhu et al.³⁸. When using SMR, specifying a threshold to remove SNPs with discrepant allele frequencies between data sets is required. That is, the SNPs with allele frequency differences between any pairwise data sets (the cis-eQTL and the GWAS summary data) large than the specified threshold (default value = 0.2) will be excluded to filter possible false positives due to allele frequency differences between the two studied populations. MR uses a cis-eQTL as a variable to estimate and test for the causative effect of an exposure variable (gene expression levels) on an outcome (presence of a specific type of PTB-associated lesion). The effect of gene expression on a specific disease outcome would then be explained by the effect of the cis-eQTL on both disease outcome and gene expression: $\hat{b}_{xy} = b_{zy} / \hat{b}_{zx}$ where z is the SNP, x a gene’s expression level, y the phenotype, \hat{b}_{zy} and \hat{b}_{zx} are the least-square estimates of y and x on z , respectively, and \hat{b}_{xy} is interpreted as the effect size of x on y free of confounding from non-genetic factors [56]. The statistic that tests for pleiotropic association (T_{SMR}) would be then calculated as follows:

$$T_{SMR} = b_{xy}^2 / \text{var}(b_{xy}) = \frac{z_{zy}^2 z_{zx}^2}{z_{zy}^2 + z_{zx}^2}$$

where z_{zy} and z_{zx} being the statistics of the GWAS and eQTL analyses, respectively. P -values were corrected with the BH method and filtered by $FDR \leq 0.05$ using *R* [53]. Finally, to correct for linkage disequilibrium (LD), an r^2 threshold (default value $r^2 > 0.9$) was used to remove SNPs in very strong LD with the top associated cis-eQTLs.

Results

RNA-Seq databases and WGS-derived genotypes

Holstein cattle ($N=16$) included in the RNA-Seq study were tested by ELISA for the detection of MAP-specific antibodies, real-time quantitative PCR (qPCR) for the detection of MAP DNA in feces and gut issues, bacteriological culture of feces and gut tissues, and histopathological analysis, as previously described (Table 1) [39]. PB from cow ID13 and the ICV from cow ID16 were not available and were not included in the study. All control animals ($N=4$) were negative for all the tests. Total RNA was isolated from 15 PB and 15 ICV samples, RNA-Seq libraries were generated and sequenced. In the current study, the RNA-Seq reads were mapped to the most recent *Bos taurus* reference genome (ARS-UCD1.2.105) with STAR 2.5.3a [46]. The number of total reads and mapped reads per individual RNA-Seq library is provided in Supplementary Table 1. Alignments of the RNA-Seq reads to the *Bos taurus* reference genome yielded mean values per library of 21.14 million reads. From the mapped reads, an average of 5% of the reads was mapped to multiple locations in the genome and was excluded. The number of mapped reads for each sample was counted and normalized using Rsubread 2.12.0 [48] and DESeq2 [49], respectively. On the other hand, DNA from PB samples of the 16 animals was genotyped using the Illumina EuroG MD Bead Chip (54,609 SNPs), imputed to the Bovine HD Bead Chip, and then to WGS as previously

described [18]. After filtering for $MAF < 0.01$, a total of 12,377,073 remained.

Cis-eQTL mapping

Significant associations ($FDR \leq 0.05$) between cis-eQTLs located within 1 Mb upstream of a TSS and normalized gene counts were detected using Tensor QTL [51]. The cis-eQTLs identified within 1 Mb upstream of a TSS that are associated with changes in gene expression levels are presented in Fig. 2. We identified 88 cis-eQTLs associated with the expression of 90 genes in the PB samples (Fig. 2a) and 37 cis-eQTLs associated with expression changes of 37 genes in the ICV samples (Fig. 2b) ($FDR \leq 0.05$). Most of the identified cis-eQTLs we located in intronic regions in the PB (53%) and ICV (71%) samples (Figs. 2c and d). These cis-eQTLs found in the TensorQTL analysis of PB and ICV samples and their targeted genes are presented in Supplementary Tables 2 and 3, respectively. The cis-eQTL with the lowest P_β -values in the PB dataset (1.93×10^{-34}) was associated with changes in the expression of the *Protein Phosphatase 1 Regulatory Subunit 3A (PPP1R3A)*, a regulator of glucose homeostasis and lipid metabolism. Several cis-eQTLs identified in the PB samples regulate the expression of miRNAs such as *bta-mir-148b*, *bta-mir-2285CK*, *bta-mir-2291*, *bta-mir-2481*, and *bta-mir-10172*. In the ICV samples, the two cis-eQTLs with the lowest P_β -value (3.60×10^{-11} and 4.47×10^{-11}) regulate the expression of *LOC112449666* (small nucleolar RNAs *SNORA70*) and *bta-mir-2462*

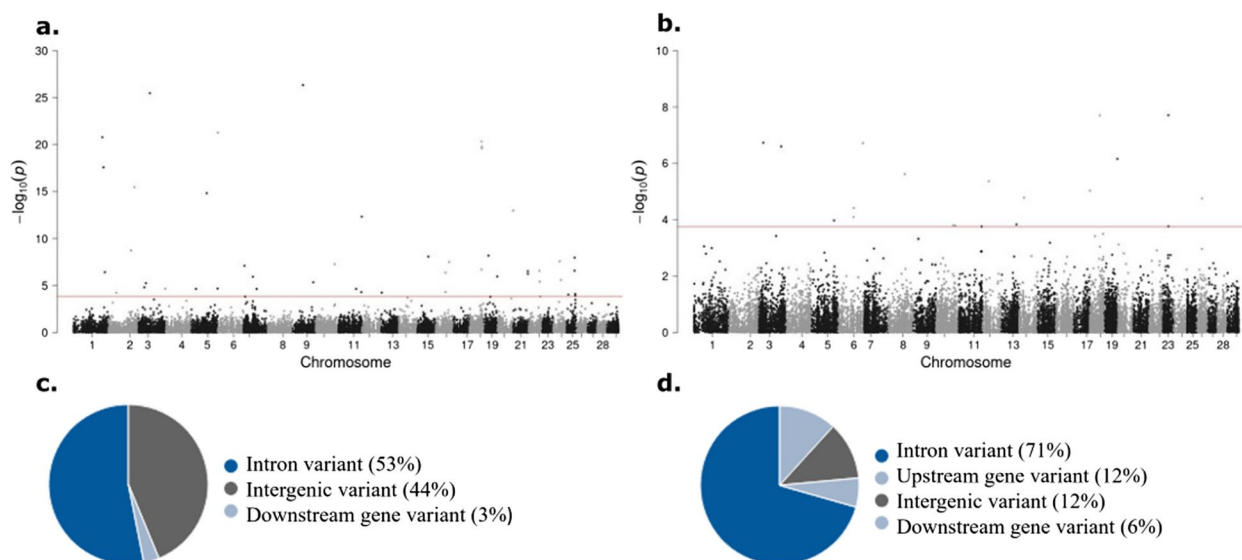


Fig. 2 Manhattan plots of cis-eQTL mapping results. **a** cis-eQTLs identified in PB samples within 1 Mb upstream of a TSS. **b** cis-eQTLs identified in ICV samples within 1 Mb upstream of a TSS. The plot shows in the Y axis the $-\log_{10}(P_\beta\text{-values})$ of each SNP and in the X axis the chromosome where each cis-eQTL is located. Each dot represents a SNP along the *Bos taurus* genome. The dotted lines represent the P_β -values that correspond to a FDR equal to 0.05. **c, d** The chart shows the genomic distributions of the PB (**c**) and ICV (**d**) cis-eQTLs identified within 1 Mb upstream of a TSS according to the Ensembl Variant Effector Predictor (VEP)

expression. Several *cis*-eQTLs identified in the ICV database were found to modify the expression of several small nucleolar RNAs (*SNORA63*, *SNORA47*, *SNORA116*), miRNAs (*bta-mir-126*, *bta-mir-2462*) and genes implicated in splicing (*U1*, *U5*, *U6*).

Summary-based data Mendelian randomization (SMR) analysis

To perform the SMR analysis, *cis*-eQTLs identified in PB and ICV samples (P -value $\leq 5 \times 10^{-8}$) and GWAS data from a previous study¹⁹ were used. In this GWAS, animals were categorized according to the presence or absence of focal, multifocal, and diffuse lesions. Therefore, a total of six SMR analyses were performed using the two sets of *cis*-eQTLs information in PB and ICV samples and the three GWAS summary statistics for the comparisons (animals with focal lesions *vs* controls, animals with multifocal lesions *vs* controls, and animals with diffuse lesions *vs* controls). For each SMR analysis performed, the FDR was determined. No significant *cis*-eQTL-gene expression-focal lesions relationship was found likely due to the lack of SNPs associated with the presence of focal lesions in the GWAS [19]. The top 10 SNPs identified in the SMR analysis using the *cis*-eQTL identified in PB and ICV samples (P -value $\leq 5 \times 10^{-8}$) and

GWAS data of the comparisons of cows with multifocal lesions *vs* controls and cows with diffuse lesions *vs* control can be found in Tables 2 and 3, respectively. Although several *cis*-eQTLs passed the genome-wide significant threshold in the SMR test ($P \leq 0.05$), only two *cis*-eQTLs were significant after correction for multiple testing (False discovery rate, $FDR \leq 0.05$). Several genes tagged by the top-associated *cis*-eQTLs ($P \leq 0.05$, $FDR > 0.05$) were involved in splicing (*U6*), transcriptional regulation (*TCEA3*, *MECOM*, *EHF*), innate immune response (*SERPINB12*, *BPIFB6*), apoptosis (*CIDEA*), blood coagulation (*Coagulation Factor F3*, *anticoagulant protein S*), and regulation of epithelial cells adhesion (*CLDN14*). A previously identified *cis*-eQTL-rs43744169, associated with the up-regulation of the *MECOM* expression and with increased risk for developing diffuse lesions in gut tissues after MAP infection³⁰, was the top eQTL in the PB-SMR database; $P = 0.0005$ and $FDR = 0.2$ (Table 3).

After multiple corrections, the *cis*-eQTL-rs383097118 (intron variant) involved in the upregulation of the *early growth response factor 4* (*EGR4*) showed a pleiotropic association with the presence of multifocal lesions (P_{β} [Tensor] = 6.47×10^{-43} , P [GWAS] = 1.22×10^{-6} , FDR [SMR] = 0.002, SNP effect [SMR] = 0.222) (Fig. 3a). Using the *cis*-eQTLs ($P \leq 5 \times 10^{-8}$) identified in the ICV

Table 2 Top 10 most significant *cis*-eQTL found in the SMR analyses (P -value ≤ 0.05) using GWAS data from a case-control study where cows with multifocal lesions ($N = 33$) and without lesions ($N = 373$) were compared

Sample	Chr	Position	Gene ID	Gene symbol	SNP ID	SNP effect	P-value	FDR
ICV	23	27,146,264	ENSBTAG00000049699	U6	rs382341668	0.27394	0.00052194	0.134
	24	60,894,426	ENSBTAG00000035171	SERPINB12	rs477516898	0.13948	0.00171221	0.22
	18	63,798,265	ENSBTAG00000053810		rs432670715	0.220168	0.00404204	0.346
	5	106,389,826	ENSBTAG00000042924	SNORA70	rs464151990	-0.0624398	0.01562814	0.963
	29	1,195,578	ENSBTAG00000054019		rs379620385	0.0697581	0.02785573	0.963
	3	17,894,282	ENSBTAG00000053103		rs109701681	0.00163663	0.03640438	0.963
	3	20,098,666	ENSBTAG00000022741	MGC134040	rs478694916	0.055018	0.03735433	0.963
	14	63,608,763	ENSBTAG00000042171	SNORA70	rs441304008	-0.0235147	0.03771466	0.963
	6	94,896,092	ENSBTAG00000035776	CFAP299	rs459014612	0.0827706	0.04253706	0.963
	22	55,021,464	ENSBTAG00000043073	U6	rs443419569	0.0815626	0.04585341	0.963
PB	11	10,810,645	ENSBTAG00000024058	EGR4	rs383097118	0.222932	4.75E-06	0.002
	3	47,966,652	ENSBTAG00000007101	F3	rs380850681	0.00536243	0.00193635	0.207
	23	16,744,416	ENSBTAG00000015131	SLC29A1	rs137606813	0.00021436	0.00189969	0.207
	1	38,894,238	ENSBTAG00000023652	PROS1	rs385387727	0.00120501	0.00103787	0.207
	19	18,988,823	ENSBTAG00000008788	SLC13A2	rs478589506	0.24342	0.00511518	0.365
	22	6,288,529	ENSBTAG00000022744	TRIM71	rs136689901	0.0169698	0.00434811	0.365
	2	128,473,027	ENSBTAG00000038865	TCEA3	rs482027819	0.0143887	0.00868258	0.532
	3	94,262,264	ENSBTAG00000009674	C3H1orf185	rs466365210	0.216626	0.01357773	0.677
	1	72,577,888	ENSBTAG00000048826		rs109233921	-0.0774901	0.0142077	0.677
	5	106,654,304	ENSBTAG00000008553	B4GALNT3		-0.0420428	0.02157095	0.74

SNP effect- the effect size of a SNP on the phenotype that is mediated by gene expression.

ICV Ileocecal valve, PB Peripheral blood, Chr Chromosome, SNP single nucleotide polymorphism, FDR false discovery rate, ID identification code

Table 3 Top 10 most significant cis-eQTL found in the SMR analyses (P -value ≤ 0.05) using GWAS data from a case-control study where cows with diffuse lesions ($N=36$) and without lesions ($N=373$) were compared

Sample	Chr	Position	Gene ID	Gene symbol	SNP ID	SNP effect	P-value	FDR
ICV	3	20,098,666	ENSBTAG00000022741	MGC134040	rs478694916	0.103	7.53E-05	0.017
	10	25,670,772	ENSBTAG00000039241	OR6S1		0.14	1.85E-03	0.212
	3	114,640,987	ENSBTAG00000044472	5S_rRNA	rs469342481	0.065	9.70E-03	0.741
	3	17,181,908	ENSBTAG00000053743		rs471339029	0.108	0.016	0.754
	22	15,880,322	ENSBTAG00000007969	CIDEC	rs209078215	-0.0008	0.016	0.754
	13	73,824,843	ENSBTAG00000045630		rs801310416	0.086	0.027	0.798
	19	20,354,910	ENSBTAG00000042268	LOC112442820	rs481186610	0.07	0.029	0.798
	7	58,851,099	ENSBTAG00000033759	LOC777593	rs383285289	-0.087	0.03	0.798
	13	63,197,058	ENSBTAG00000010112	BPIFB6	rs110136678	0.001	0.031	0.798
	5	93,655,946	ENSBTAG00000052966	bta-mir-2285be	rs453650622	0.154	0.038	0.8
PB	1	97,501,420	ENSBTAG00000005871	MECOM	rs382448539	0.022	5.22E-04	0.201
	15	21,656,148	ENSBTAG00000015810	C15H11orf34	rs110448688	0.112	1.39E-03	0.267
	1	148,065,848	ENSBTAG00000030585	CLDN14		0.356	3.44E-03	0.441
	16	48,321,792	ENSBTAG00000045373	bta-mir-2320		0.108	7.63E-03	0.734
	15	64,738,340	ENSBTAG00000017150	EHF	rs876546467	0.053	9.83E-03	0.757
	13	74,882,914	ENSBTAG00000002942	SLC2A10	rs41711514	0.002	0.014	0.874
	3	47,966,652	ENSBTAG00000007101	F3	rs380850681	0.004	0.017	0.916
	9	9,801,563	ENSBTAG00000035054	COL9A1	rs435010900	0.04	0.021	0.986
	5	114,503,489	ENSBTAG00000006973	PHF21B	rs444663763	0.027	0.023	0.986
	15	21,822,607	ENSBTAG00000053780		rs109404405	0.029	0.026	0.986

SNP effect- the effect size of a SNP on the phenotype that is mediated by gene expression

ICV Ileocec valve, PB Peripheral blood, Chr Chromosome, SNP Single nucleotide polymorphism, FDR False discovery rate, ID Identification code

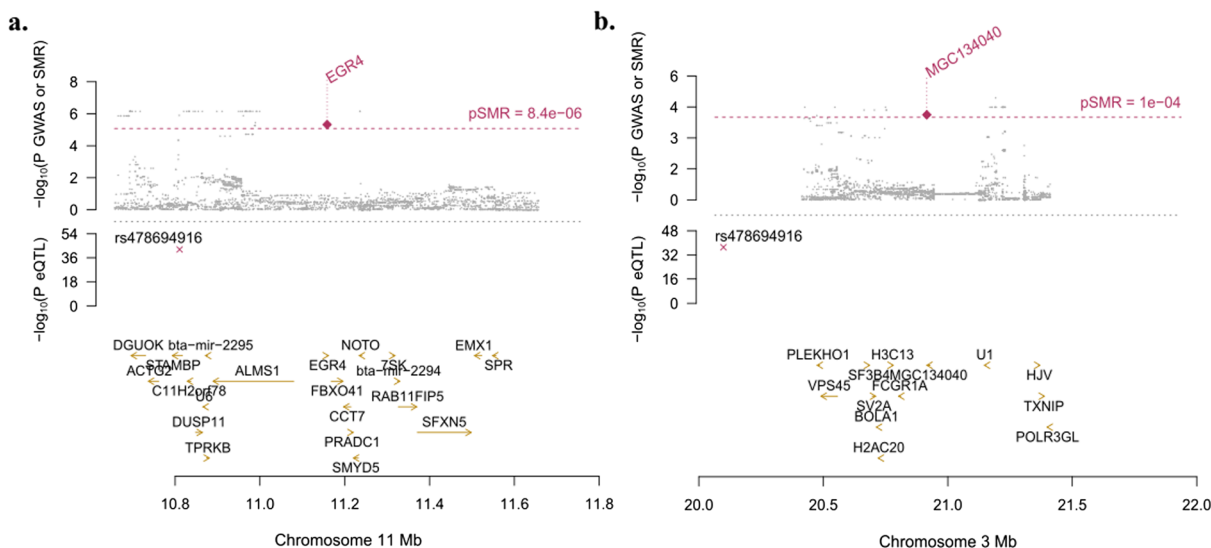


Fig. 3 Locus plots showing the significant genes *EGR4* and *MGC134040*, their locations within chromosomes 11 and 3, respectively. Y axis represents the negative log of the significant P -values and X axis represents the negative log of the P -values. Dots in the top plots represent the P -values for SNPs from the GWAS analysis. Rhombuses represent the P -values from the SMR analyses, highlighted in red if significant. In the middle plots, crosses represent the P_{β} -values of the cis-eQTLs located in the plotted region. In the bottom plots are represented all genes in each plotted region. a. SMR results using cis-eQTLs identified in PB samples and GWAs results of a case-control study where cases were animals with multifocal lesions and controls were animals without lesions. The dotted red line represents $FDR \leq 0.05$ corresponding to $P [SMR] = 10^{-5}$. b. SMR results using ci-eQTLs identified in ICV samples and GWAs results of a case-control study where cases were animals with diffuse lesions and controls were animals without lesions. The dotted red line represents $FDR \leq 0.05$ corresponding to $P [SMR] = 10^{-4}$

samples, we found that the *cis*-eQTL-rs478694916 (upstream gene variant) upregulated the bovine *neuroblastoma breakpoint family member 6-like protein isoform 2* (*MGC134040*) expression and showed pleiotropic association with the presence of diffuse lesions (P_{β} [Tensor] = 8.85×10^{-38} , P [GWAS] = 3.16×10^{-5} , FDR [SMR] = 0.017, SNP effect [SMR] = 0.103) (Fig. 3b).

Discussion

Combined genetic-transcriptomic approaches allow the identification of *cis*-eQTLs that affect the expression levels of genes with pleiotropic effects on complex traits. In the current study, we integrated WGS-derived genotypes and RNA-Seq reads that were mapped to the most recent *Bos taurus* reference genome (ARS-UCD1.2.105) with *STAR* 2.5.3a, an ultrafast RNA-Seq aligner that exhibits better alignment precision and sensitivity than other RNA-seq aligners [46]. Using *TensorQTL* for *cis*-eQTL mapping, we identified 88 *cis*-eQTLs in PB and 37 *cis*-eQTLs in ICV associated with the expression of 90 and 37 genes, respectively. We recognize that the number of individual samples used in the first part of the study was small ($N=30$). Although the number of animals used in the GWAS could be considered moderate ($N=813$), a total of 192 and 92 SNPs defining 13 and 9 distinct QTLs were highly associated ($FDR \leq 0.05$, $P \leq 5 \times 10^{-7}$) with the multifocal (heritability = 0.075) and the diffuse (heritability = 0.189) lesions, respectively [19]. *cis*-eQTL data was combined with GWAS data to successfully detect two pleiotropic associations between the *EGR4* expression in PB and susceptibility to develop multifocal lesions and between the *MGC134040* expression in ICV and susceptibility to develop diffuse lesions. The list of *cis*-eQTLs associated with disease outcomes will be larger in the future when larger *cis*-eQTLs and GWAS datasets will become available. We used *cis*-eQTLs data as *trans*-eQTLs have a weaker effect size and less direct effect.

To our knowledge, this is the first study that used an SMR approach to identify *cis*-eQTLs regulating genes associated with PTB outcomes by pleiotropy [57]. Using this multi-omics approach to combine various sources of data, no significant *cis*-eQTL-gene expression-focal lesions relationship was found, probably due to the lack of discrimination between focal lesions and controls using GWAS [19]. In contrast, we found two *cis*-eQTLs significantly associated with the presence of multifocal and diffuse PTB-associated lesions in gut tissues. More specifically, the heterozygous (T/C) genotype in the *cis*-eQTL-rs383097118 was associated with upregulation of the *EGR4* expression and with the presence of multifocal lesions (SNP effect = 0.222). This suggests that either the variant (T/C) in the *cis*-eQTL-rs383097118 results in upregulation of *EGR4* expression and increases the

susceptibility to develop multifocal lesions or that the (T/C) variant results in upregulation of *EGR4* expression, which in turn increases the susceptibility to present multifocal lesions. In other words, the presence of the heterozygous (T/C) genotype in the *cis*-eQTL-rs383097118 and the increase in *EGR4* expression were associated with the presence of multifocal lesions. *EGR4* is a key regulator of T-cell activation, differentiation, and function. MAP-antigens stimulation of T cells is induced by T-cell receptor ligation and results in the activation and the *novo* transcription of DNA-binding transcription factors including *EGR4*, *NF- κ B*, *NT-AT*, *AP-1* ad *STAT* proteins. *EGR4* is a transcription factor that contains two zinc finger domains that interact with *NF- κ B* in T-cells leading to the transcriptional control of genes encoding pro-inflammatory cytokines [58]. The absence of *EGR4* markedly enhanced the proliferation of both CD4+ and CD8 T+ cells and increased the levels of *IFN γ* , *IL-9*, *IL-10*, and *IL-21* in *EGR4* -/- mice even in the absence of stimulation [59]. These results suggest that *EGR4* induction acts as a brake on T cell proliferation and cytokine production by limiting the strength and duration of *NF- κ B* activation, making T cells poised to respond efficiently to further stimulation. In line with these findings, we hypothesize that the presence of the heterozygous genotype in the *cis*-eQTL-rs383097118 would increase *EGR4* expression and limit the *NF- κ B*-induced pro-inflammatory immune response in response to MAP infection controlling the inflammation, resulting in the induction of an anergy stage, and allowing a long-term association of MAP with the host. This will be in line with our recent findings suggesting that the multifocal lesions are localized/confined lesions that have different underlying host genetics than the diffuse lesions [19]. Similarly, it was previously demonstrated that the presence of granulomas with multifocal distribution in tuberculoid leprosy leads to the control of *Mycobacterium leprae* replication and the containment of its spread [60]. Therefore, the T/C variant in the *cis*-eQTL-rs383097118 and high *EGR4* expression could be considered markers of PTB resilience, reflecting a T-cell intrinsic property. Since it is well recognized that not all the infected animals will progress into clinical forms during their productive life, the use of genetic markers for the identification of resilient cattle might help farmers or animal health managers to select resilient and long-time asymptomatic cattle able to tolerate the disease without having their health and milk production compromised.

The heterozygous variant in the *cis*-eQTL-rs478694916 was associated with the upregulation of the bovine *MGC134040* in ICV samples and with the presence of diffuse lesions. These findings suggest that high levels of *MGC134040* are associated with PTB susceptibility,

as diffuse lesions are usually a sign of clinical infection. Although there is a lack of information about the *MGC134040* gene function in cattle, *MGC134040* belongs to the *neuroblastoma breakpoint family (NBPF)*, whose overexpression has been associated with cell proliferation, and with several cancers including sarcoma in humans [61]. *NBPF* members are DNA-binding transcription factors that are directly regulated by *NF- κ B* [62].

Although no other cis-eQTLs were significant after correction for multiple testing ($FDR \leq 0.05$), genes tagged by significant cis-eQTLs ($P [SMR] \leq 0.05$) were involved in splicing (*U6*), transcriptional regulation (*TCEA3*, *MECOM*, *EHF*), innate immune response (*SERPINB12*, *BPIFB6*), apoptosis (*CIDEA*), blood coagulation (*Coagulation Factor F3*, *anticoagulant protein S*), and regulation of epithelial cells adhesion (*CLDN14*). The cis-eQTL-rs43744169 (T/C), associated with the upregulation of the *MECOM* and increased risk of clinical PTB in a previous study [30], was also significantly associated with an increased risk of progression to diffuse lesions by SMR ($P[SMR]=0.0005$, $FDR=0.2$). Recent findings indicated that the *MECOM* is upregulated by inflammatory stimuli, including bacteria, and that mutations in the *MECOM* make mice more susceptible to bacterial infections [63]. Allelic variants affecting the human *MECOM* have been also associated with human IBD [64]. *MECOM* is a transcriptional regulator of the *NF- κ B*-mediated inflammatory response and an oncogene whose upregulation has been associated with many types of solid cancers in humans, including colorectal cancer and acute myeloid leukemia [65, 66].

Since the *NF- κ B* is a critical factor in the gut immune response against pathogens and in promoting inflammation-associated carcinomas in the gastrointestinal tract, it has been proposed that the *NF- κ B* might provide a common and critical mechanistic link between inflammation and cancer [67]. In addition, recent epidemiological and experimental studies have revealed that mycobacterium-related inflammation may be a possible mechanism of cancer pathogenesis [68]. For instance, it has been documented that *Mycobacterium tuberculosis* and non-tuberculous *Mycobacterium avium* complex infections increase the risk and mortality of pulmonary cancer, whereas *Mycobacterium ulcerans* has been correlated with skin carcinogenesis and MAP with IIBD-associated and sporadic colorectal cancer [12, 68, 69]. Our findings suggest that heterozygous variants in a specific cis-eQTL might upregulate the expression of the *MGC134040*, a transcriptional regulator of the *NF- κ B* mediated inflammatory response in macrophages and epithelial cells, causing an uncontrolled and aberrant inflammatory response which might exacerbate tissue

injury in PTB-infected cattle. It has been suggested that the activation of the inflammatory response mediated by *NF- κ B* is accompanied by an increase in the proliferation rate of cells with damaged DNA which might lead to the initiation of tumorigenesis [70]. Contrarily, our study suggests that cis-eQTL-mediated upregulation of the transcriptional regulator *EGR4* might limit the *NF- κ B*-induced proinflammatory immune response to MAP infection leading to a more restricted type of lesions.

SMR allowed us to include independently collected GWAS and cis-eQTL data to enlarge the number of individuals and thus increase the statistical power. The two cis-eQTLs, rs383097118 and rs478694916, with effects on the expression of the *EGR4* and *MGC134040* genes and the resilience/susceptibility to PTB were statistically significant in the previous GWAS, but they were not among the most significant SNPs, i.e. at the top of their peaks. This is because the *Tensor QTL* algorithm only identifies cis-eQTLs within 1 Mb of a gene, so SNPs that fall out of that range will not be identified as cis-eQTLs and, therefore, the most significant variants in the GWAS might be filtered out. Further functional studies are needed to evaluate the critical significance of the identified genetic variants for diagnosis and breeding purposes and to confirm the roles of the *MGC134040* and *EGR4* in PTB susceptibility and resilience, respectively.

Conclusions

Our results provide a better understanding of the genetic factors influencing PTB resilience and susceptibility at the whole-genome level. Using SMR we identified two novel cis-eQTLs regulating the expression levels of two transcriptional factors (*EGR4* and *MGC134040*) functionally involved in the *NF- κ B* inflammatory response. These cis-eQTLs affect the occurrence of multifocal or diffuse lesions by upregulating the expression of *EGR4* and *MGC134040*, respectively. The introduction of the cis-eQTLs identified in the current study into marker-assisted breeding programs is expected to have a relevant effect on increasing PTB natural resilience and reducing susceptibility through selection, which in turn would reduce economic losses and improve animal health. Applications of animal genetics in breeding programs are currently one of the important motors for efficient livestock production, not only to increase performance and productivity but also to ensure the resilience and health of livestock while improving longevity of animals.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-023-09710-w>.

Additional file 1: Supplementary Table I. RNA-Seq mapped reads.

Additional file 2: Supplementary Table 2. Significant cis-eQTLs found in the Tensor QTL analysis of PB samples and their targeted genes.

Additional file 3: Supplementary Table 3. Significant cis-eQTLs found in the Tensor QTL analysis of ICV samples and their controlled genes.

Acknowledgements

Special thanks are due to the dairy producers and veterinarians for their collaboration. We would like to thank Mertxe Bascones and Ainara Badiola for their help in preparing blood samples for genotyping. We would like to thank the i2basque Research and Academic Network for computational support. We gratefully acknowledge the 1000 Bull Genomes Consortium for providing accessibility to WGS data that was used in this study. We are grateful to Kyle P. Hearn for the careful editing of the manuscript.

Authors' contributions

M.A.H. coordinated the study. N.F., J.R.B., M.C., G.B.B., and M.A.H. participated in the study design. R.C. and C.B.V. participated in sample collection and phenotypic characterization of the animals used in the RNA-Seq. P.V., J.G., and R.A.J. characterized the disease status of the animals included in the GWAS analysis. M.C. performed the GWAS analysis. J.L.L. performed the RNA-Seq analysis. A.F. and O.G.R. supervised the imputation to WGS. G.B.B. performed the *Tensor* and *SMR* analysis and prepared all tables and figures. G.B.B., M.C., R.A.J., and M.A.H. collaborated on the interpretation and discussion of the results. G.B.B. drafted the manuscript and M.A.H. reviewed and edited the final draft. All the authors carefully read, reviewed, and approved the final manuscript.

Funding

Financial support for this study was provided by grants (RTA2014-00009; RTI2018-094192, PID2021-122197OR-C21) funded by the National Institute for Agricultural Research (INIA), MCIN/AEI/10.13039/501100011033, and by FEDER, "Una manera de hacer Europa". Funds were also provided by the Gobierno del Principado de Asturias Regional funds PCTI 2021–2023 (GRUPIN: IDI2021-000102) co-funded by FEDER. MC, CBV, and GBB have been awarded fellowships from INIA and MCIN/AEI/10.13039/501100011033 and "FSE Invierte en tu futuro"; grants FPI2016-00041, CPD2016-0142, and PRE2019-090562, respectively.

Availability of data and materials

The dataset supporting the conclusions of this article are included within the article. RNA-Seq raw data are deposited in the NCBI Gene Expression Omnibus (GEO) database (<https://www.ncbi.nlm.nih.gov/geo/>) under the accession number GSE137395. Sequence data used in this study for WGS imputation are owned by the 1000 Bull Genomes Project Consortium. The sequence variants for 1800 animals from Run8 are public at the European Variation Archive under project number PRJEB42783 (<https://www.ebi.ac.uk/ena/browser/view/PRJEB42783>). Individual genotype data used in this study is managed by a third party, the Spanish Friesian Cattle National Federation (CONAFE). Requests for genotype data can be made to CONAFE, Ctra. de Andalucía, km. 23.600—28,340 Valdemoro, Madrid, Spain; email: conafe@conafe.com; phone: +34 (91) 8,952,412; website: www.conafe.com. This article was submitted to an online preprint archive; <https://doi.org/10.21203/rs.3.rs-2471714/v1>.

Declarations

Ethics approval and consent to participate

All methods were carried out in accordance with relevant guidelines and regulations. PB and ICV collection from the 16 animals included in the RNA-Seq analysis were approved by the Animal Ethics Committee of the Servicio Regional de Investigación y Desarrollo Agroalimentario (SERIDA) and authorized by the Regional Consejería de Agroganadería y Recursos Autóctonos of the Principality of Asturias (authorization code PROAE29/2015). All the procedures were conducted following the European Guidelines for welfare (Basque Government Decree 454/1994, Spanish Government Law 32/2007, Royal decree 731/ the Care and Use of Animals for Research Purposes (2012/63/EU). Blood, fecal, and tissue samples were collected by trained personnel and in accordance with good veterinary practices. Slaughtered animals ($N=813$) used in the GWAS study were sampled under pertinent legislation for safeguarding animal 2007, and European council regulation number 1099/2009).

All methods are reported in accordance with ARRIVE guidelines (<https://arriveguidelines.org>) for the reporting of animal experiments.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Department of Animal Health, NEIKER- Basque Institute for Agricultural Research and Development, Basque Research and Technology Alliance (BRTA), Derio, Bizkaia, Spain. ²Doctoral Program in Molecular Biology and Biomedicine, Universidad del País Vasco/Euskal Herriko Unibertsitatea (UPV/EHU), Leioa, Bizkaia, Spain. ³Department of Genetics, Physical Anthropology and Animal Physiology, University of the Basque Country (UPV/EHU), Biocruces-Bizkaia HRI, Leioa, Bizkaia, Spain. ⁴Department of Applied Mathematics, NEIKER-Basque Institute for Agricultural Research and Development, Basque Research and Technology Alliance (BRTA), Derio, Bizkaia, Spain. ⁵Center of Animal Biotechnology, SERIDA, Servicio Regional de Investigación y Desarrollo Agroalimentario, Deva, Asturias, Spain. ⁶Departamento de Mejora Genética Animal, Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria, CSIC, Madrid, Spain.

Received: 12 January 2023 Accepted: 2 October 2023

Published online: 11 October 2023

References

- Saxegaard F, Baess I. Relationship between *Mycobacterium avium*, *Mycobacterium paratuberculosis* and "wood pigeon mycobacteria." *APMIS*. 1988;96(1–6):37–42. <https://doi.org/10.1111/j.1699-0463.1988.tb05265.x>.
- Thorel M, Blom-Potar M-C, Rastogi N. Characterization of *Mycobacterium paratuberculosis* and "wood-pigeon" mycobacteria by isoenzyme profile and selective staining of immunoprecipitates. *Res Microbiol*. 1990;141(5):551–61. [https://doi.org/10.1016/0923-2508\(90\)90019-M](https://doi.org/10.1016/0923-2508(90)90019-M).
- Garvey M. *Mycobacterium Avium Paratuberculosis*: A Disease Burden on the Dairy Industry. *Animals*. 2020;10(10):1773. <https://doi.org/10.3390/ani10101773>.
- Rasmussen P, Barkema HW, Mason S, Beaulieu E, Hall DC. Economic losses due to Johne's disease (paratuberculosis) in dairy cattle. *J Dairy Sci*. 2021;104(3):3123–43. <https://doi.org/10.3168/jds.2020-19381>.
- Hines ME, Stiver S, Giri D, Whittington L, Watson C, Johnson J, et al. Efficacy of spheroplastic and cell-wall competent vaccines for *Mycobacterium avium* subsp. *paratuberculosis* in experimentally-challenged baby goats. *Vet Microbiol*. 2007;120(3–4):261–283. <https://doi.org/10.1016/j.vetmic.2006.10.030>
- Bermudez LE, Petrofsky M, Sommer S, Barletta RG. Peyer's Patch-Deficient Mice Demonstrate that *Mycobacterium avium* subsp. *paratuberculosis* Translocates across the Mucosal Barrier via both M Cells and Enterocytes but Has Inefficient Dissemination. *Infect Immun*. 2010;78(8):3570–3577. <https://doi.org/10.1128/IAI.01411-09>
- Rees WD, Lorenzo-Leal AC, Steiner TS, Bach H. *Mycobacterium avium* Subspecies *paratuberculosis* Infects and Replicates within Human Monocyte-Derived Dendritic Cells. *Microorganisms*. 2020;8(7):994. <https://doi.org/10.3390/microorganisms8070994>.
- Khare S, Lawhon SD, Drake KL, Nunes JES, Figueiredo JF, Rossetti CA, et al. Systems Biology Analysis of Gene Expression during In Vivo *Mycobacterium avium* paratuberculosis Enteric Colonization Reveals Role for Immune Tolerance. *PLoS ONE*. 2012;7(8):e42127. <https://doi.org/10.1371/journal.pone.0042127>.
- González J, Geijo MV, García-Pariente C, Verna A, Corpa JM, Reyes LE, et al. Histopathological Classification of Lesions associated with Natural Paratuberculosis Infection in Cattle. *J Comp Pathol*. 2005;133(2–3):184–96. <https://doi.org/10.1016/j.jcpa.2005.04.007>.
- Balseiro A, Perez V, Juste RA. Chronic regional intestinal inflammatory disease: A trans-species slow infection? *Comp Immunol Microbiol Infect Dis*. 2018;2019(62):88–100. <https://doi.org/10.1016/j.cimid.2018.12.001>.

11. Juste RA, Elguezal N, Pavón A, Garrido JM, Geijo M, Sevilla I, et al. Association between *Mycobacterium avium* subsp. *paratuberculosis* DNA in blood and cellular and humoral immune response in inflammatory bowel disease patients and controls. *Int J Infect Dis.* 2009;13(2):247–254. <https://doi.org/10.1016/j.ijid.2008.06.034>
12. Pierce ES. Could *Mycobacterium avium* subspecies *paratuberculosis* cause Crohn's disease, ulcerative colitis... and colorectal cancer? *Infect Agent Cancer.* 2018;13(1):1–6. <https://doi.org/10.1186/s13027-017-0172-3>
13. Dow CT. Warm, Sweetened Milk at the Twilight of Immunity - Alzheimer's Disease - Inflammaging, Insulin Resistance M paratuberculosis and Immunosenescence. *Front Immunol.* 2021;12:1–11. <https://doi.org/10.3389/fimmu.2021.714179>
14. Jeyanathan M, Boutros-Tadros O, Radhi J, Semret M, Bitton A, Behr MA. Visualization of *Mycobacterium avium* in Crohn's tissue by oil-immersion microscopy. *Microbes Infect.* 2007;9(14–15):1567–73. <https://doi.org/10.1016/j.micinf.2007.09.001>
15. Garrido JM, Vazquez P, Molina E, Plazaola JM, Sevilla IA, Geijo MV, et al. Paratuberculosis vaccination causes only limited cross-reactivity in the skin test for diagnosis of bovine tuberculosis. *PLoS One.* 2013;8(11):2–8. <https://doi.org/10.1371/journal.pone.0080985>
16. McGovern SP, Purfield DC, Ring SC, Carthy TR, Graham DA, Berry DP. Candidate genes associated with the heritable humoral response to *Mycobacterium avium* ssp. *paratuberculosis* in dairy cows have factors in common with gastrointestinal diseases in humans. *J Dairy Sci.* 2019;102:4249–4263. <https://doi.org/10.3168/jds.2018-15906>
17. Sanchez MP, Guatteo R, Davergne A, Saout J, Grohs C, Deloche MC, et al. Identification of the ABC4, IER3, and CBFA2T2 candidate genes for resistance to paratuberculosis from sequence-based GWAS in Holstein and Normande dairy cattle. *Genet Sel Evol.* 2020;52:1–17. <https://doi.org/10.1186/s12711-020-00535-9>
18. Canive M, González-Recio O, Fernández A, Vázquez P, Badia-Bringué G, Lavín JL, et al. Identification of loci associated with susceptibility to *Mycobacterium avium* subsp. *paratuberculosis* infection in Holstein cattle using combinations of diagnostic tests and imputed whole-genome sequence data. *PLoS One.* 2021;16(8):e0256091. <https://doi.org/10.1371/journal.pone.0256091>
19. Canive M, Badia-Bringué G, Vázquez P, González-Recio O, Fernández A, Garrido JM, et al. Identification of loci associated with pathological outcomes in Holstein cattle infected with *Mycobacterium avium* subsp. *paratuberculosis* using whole-genome sequence data. *Sci Rep.* 2021;11(1):20177. <https://doi.org/10.1038/s41598-021-99672-4>
20. Kirkpatrick BW, Cooke ME, Frie M, Sporer KRB, Lett B, Wells SJ, et al. Genome-wide association analysis for susceptibility to infection by *Mycobacterium avium* ssp. *paratuberculosis* in US Holsteins. *J Dairy Sci.* 2022;105:4301–13. <https://doi.org/10.3168/jds.2021-21276>
21. Sanchez MP, Tribout T, Fritz S, Guatteo R, Fourichon C, Schibler L, et al. New insights into the genetic resistance to paratuberculosis in Holstein cattle via single-step genomic evaluation. *Genet Sel Evol.* 2022;054(1):67. <https://doi.org/10.1186/s12711-022-00757-z>
22. Alonso-Hearn M, Badia-Bringué G, Canive M. Genome-wide association studies for the identification of cattle susceptible and resilient to paratuberculosis. *Front Vet Sci.* 2022;9. <https://doi.org/10.3389/fvets.2022.935133>
23. Canive M, Badia-Bringué G, Vázquez P, Garrido JM, Juste RA, Fernández A, et al. A Genome-Wide Association Study for Tolerance to Paratuberculosis Identifies Candidate Genes Involved in DNA Packaging, DNA Damage Repair, Innate Immunity, and Pathogen Persistence. *Front Immunol.* 2022;13. <https://doi.org/10.3389/fimmu.2022.820965>
24. Badia-Bringué G, Canive M, Alonso-Hearn M. Control of *Mycobacterium avium* subsp. *paratuberculosis* load within infected bovine monocyte-derived macrophages is associated with host genetics. *Front Immunol.* 2023; 14:1042638. <https://doi.org/10.3389/fimmu.2023.1042638>
25. Badia-Bringué G, Canive M, Vázquez P, Garrido JM, Fernández A, Juste RA, et al. Association between High Interferon-Gamma Production in Avian Tuberculin-Stimulated Blood from *Mycobacterium avium* subsp. *paratuberculosis*-Infected Cattle and Candidate Genes Implicated in Necroptosis. *Microorganisms.* 2023;11(7):1817. <https://doi.org/10.3390/microorganisms11071817>
26. Minozzi G, Buggiotti L, Stella A, Strozzi F, Luini M, Williams JL. Genetic loci involved in antibody response to *Mycobacterium avium* ssp. *paratuberculosis* in cattle. *PLoS One.* 2010;5(6). <https://doi.org/10.1371/journal.pone.0011117>
27. Kiser JN, White SN, Johnson KA, Hoff JL, Taylor JF, Neiberghs HL. Identification of loci associated with susceptibility to *Mycobacterium avium* subspecies *paratuberculosis* (Map) tissue infection in cattle. *J Anim Sci.* 2017;95(3):1080–91. <https://doi.org/10.2527/jas2016.1152>
28. Koets AP, Aduagna G, Janss LLG, van Weering HJ, Kalis CHJ, Wentink GH, et al. Genetic variation of susceptibility to *Mycobacterium avium* subsp. *paratuberculosis* infection in dairy cattle. *J Dairy Sci.* 2000;83(11):2702–2708. [https://doi.org/10.3168/jds.S0022-0302\(00\)75164-2](https://doi.org/10.3168/jds.S0022-0302(00)75164-2)
29. Farh KKH, Marson A, Zhu J, Kleinewietfeld M, Housley WJ, Beik S, et al. Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature.* 2015;518(7539):337–43. <https://doi.org/10.1038/nature13835>
30. Canive M, Fernandez-Jimenez N, Casais R, Vázquez P, Lavín JL, Bilbao JR, et al. Identification of loci associated with susceptibility to bovine paratuberculosis and with the dysregulation of the MECOM, eEF1A2, and U1 spliceosomal RNA expression. *Sci Rep.* 2021;11(1):313. <https://doi.org/10.1038/s41598-020-79619-x>
31. Neiberghs HL, Settles ML, Whitlock RH, Taylor JF. GSEA-SNP identifies genes associated with Johne's disease in cattle. *Mamm Genome.* 2010;21(7–8):419–25. <https://doi.org/10.1007/s00335-010-9278-2>
32. Canive M, Badia-Bringué G, Alonso-Hearn M. The Upregulation of Cathepsin G Is Associated with Resistance to Bovine Paratuberculosis. *Animals.* 2022;12(21):1–12. <https://doi.org/10.3390/ani12213038>
33. Li B, Ritchie MD. From GWAS to Gene: Transcriptome-Wide Association Studies and Other Methods to Functionally Understand GWAS Discoveries. *Front Genet.* 2021;12(September). <https://doi.org/10.3389/fgene.2021.713230>
34. Servin B, Stephens M. Imputation-based analysis of association studies: Candidate regions and quantitative traits. *PLoS Genet.* 2007;3(7):1296–308. <https://doi.org/10.1371/journal.pgen.0030114>
35. Shabalin AA. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics.* 2012;28(10):1353–8. <https://doi.org/10.1093/bioinformatics/bts163>
36. Trapnell C, Pachter L, Salzberg SL. TopHat: Discovering splice junctions with RNA-Seq. *Bioinformatics.* 2009;25(9):1105–11. <https://doi.org/10.1093/bioinformatics/btp120>
37. Gusev A, Ko A, Shi H, Bhatia G, Chung W, Penninx BWJH, et al. Integrative approaches for large-scale transcriptome-wide association studies. *Nat Genet.* 2016;48(3):245–52. <https://doi.org/10.1038/ng.3506>
38. Zhu Z, Zhang F, Hu H, Bakshi A, Robinson MR, Powell JE, et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet.* 2016;48(5):481–7. <https://doi.org/10.1038/ng.3538>
39. Alonso-Hearn M, Canive M, Blanco-Vázquez C, Torremocha R, Balseiro A, Amado J, et al. RNA-Seq analysis of ileocecal valve and peripheral blood from Holstein cattle infected with *Mycobacterium avium* subsp. *paratuberculosis* revealed dysregulation of the CXCL8/IL8 signaling pathway. *Sci Rep.* 2019;9(1):14845. <https://doi.org/10.1038/s41598-019-51328-0>
40. Blanco-Vázquez C, Alonso-Hearn M, Juste RA, Canive M, Iglesias T, Iglesias N, et al. Detection of latent forms of *Mycobacterium avium* subsp. *paratuberculosis* infection using host biomarker-based ELISAs greatly improves paratuberculosis diagnostic sensitivity. *PLoS One.* 2020;15(9):e0236336. <https://doi.org/10.1371/journal.pone.0236336>
41. Vázquez P, Ruiz-Larrañaga O, Garrido JM, Iriondo M, Manzano C, Agirre M, et al. Genetic Association Analysis of Paratuberculosis Forms in Holstein-Friesian Cattle. *Vet Med Int.* 2014;2014:1–8. <https://doi.org/10.1155/2014/321327>
42. Loh P-R, Palamara PF, Price AL. Fast and accurate long-range phasing in a UK Biobank cohort. *Nat Genet.* 2016;48(7):811–6. <https://doi.org/10.1038/ng.3571>
43. Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, et al. Next-generation genotype imputation service and methods. *Nat Genet.* 2016;48(10):1284–7. <https://doi.org/10.1038/ng.3656>
44. Hayes BJ, Daetwyler HD. 1000 Bull Genomes Project to Map Simple and Complex Genetic Traits in Cattle: Applications and Outcomes. *Annu Rev Anim Biosci.* 2019;7(1):89–102. <https://doi.org/10.1146/annurev-animal-020518-115024>

45. Schmieder R, Edwards R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics*. 2011;27(6):863–4. <https://doi.org/10.1093/bioinformatics/btr026>.
46. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29(1):15–21. <https://doi.org/10.1093/bioinformatics/bts635>.
47. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. *Gigascience*. 2021;10(2):1–4. <https://doi.org/10.1093/gigascience/giab008>.
48. Liao Y, Smyth GK, Shi W. The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. *Nucleic Acids Res*. 2019;47(8):e47. <https://doi.org/10.1093/nar/gkz114>.
49. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15(12):550. <https://doi.org/10.1186/s13059-014-0550-8>.
50. Ongen H, Buil A, Brown AA, Dermitzakis ET, Delaneau O. Fast and efficient QTL mapper for thousands of molecular phenotypes. *Bioinformatics*. 2016;32(10):1479–85. <https://doi.org/10.1093/bioinformatics/btv722>.
51. Taylor-Weiner A, Aguet F, Haradhvala NJ, Gosai S, Anand S, Kim J, et al. Scaling computational genomics to millions of individuals with GPUs. *Genome Biol*. 2019;20(1):228. <https://doi.org/10.1186/s13059-019-1836-7>.
52. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J R Stat Soc Ser B*. 1995;57(1):289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>.
53. R Core team. R: A language and environment for statistical computing. Published online 2022. <https://www.r-project.org/>
54. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: A Tool for Genome-wide Complex Trait Analysis. *Am J Hum Genet*. 2011;88(1):76–82. <https://doi.org/10.1016/j.ajhg.2010.11.011>.
55. Yang J, Zaitlen NA, Goddard ME, Visscher PM, Price AL. Advantages and pitfalls in the application of mixed-model association methods. *Nat Genet*. 2014;46(2):100–6. <https://doi.org/10.1038/ng.2876>.
56. Smith GD, Ebrahim S. “Mendelian randomization”: Can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol*. 2003;32(1):1–22. <https://doi.org/10.1093/ije/dyg070>.
57. Porcu E, Rueger S, Lepik K, eQTLGen Consortium, BIOS Consortium, Santoni FA, et al. Mendelian randomization integrating GWAS and eQTL data reveals genetic determinants of complex and clinical traits. *Nat Commun*. 2019;10(1):1–12. <https://doi.org/10.1038/s41467-019-10936-0>
58. Wieland GD, Nehmann N, Müller D, Eibel H, Sieben U, Sühnel J, et al. Early growth response proteins EGR-4 and EGR-3 interact with immune inflammatory mediators NF- κ B p50 and p65. *J Cell Sci*. 2005;118(14):3203–12. <https://doi.org/10.1242/jcs.02445>.
59. Mookerjee-Basu J, Hooper R, Gross S, Schultz B, Go CK, Samakai E, et al. Suppression of Ca²⁺ signals by EGR 4 controls Th1 differentiation and anti-cancer immunity *in vivo*. *EMBO Rep*. 2020;21(5):1–15. <https://doi.org/10.15252/embr.201948904>
60. Leal-Calvo T, Martins BL, Bertoluci DF, Rosa PS, de Camargo RM, Germano GV, et al. Large-Scale Gene Expression Signatures Reveal a Microbicidal Pattern of Activation in *Mycobacterium leprae*-Infected Monocyte-Derived Macrophages With Low Multiplicity of Infection. *Front Immunol*. 2021;12(April):1–12. <https://doi.org/10.3389/fimmu.2021.647832>.
61. Vandepoele K, Van Roy N, Staes K, Speleman F, van Roy F. A Novel Gene Family NBPF: Intricate Structure Generated by Gene Duplications During Primate Evolution. *Mol Biol Evol*. 2005;22(11):2265–74. <https://doi.org/10.1093/molbev/msi222>.
62. Zhou F, Xing Y, Xu X, Yang Y, Zhang J, Ma Zhengliang et al. NBPF is a potential DNA-binding transcription factor that is directly regulated by NF- κ B. *Int J Biochem Cell Biol*. 2013;45(11):2479–2490. <https://doi.org/10.1016/j.biocel.2013.07.022>
63. Hood D, Moxon R, Purnell T, Richter C, Williams D, Azar A, et al. A new model for non-typeable *Haemophilus influenzae* middle ear infection in the Junbo mutant mouse. *DMM Dis Model Mech*. 2016;9(1):69–79. <https://doi.org/10.1242/dmm.021659>.
64. Naito T, Yokoyama N, Kakuta Y, Ueno K, Kawai Y, Onodera M, et al. Clinical and genetic risk factors for decreased bone mineral density in Japanese patients with inflammatory bowel disease. *J Gastroenterol Hepatol*. 2018;33(11):1873–81. <https://doi.org/10.1111/jgh.14149>.
65. Fehringer G, Kraft P, Pharoah PD, Eeles RA, Chatterjee N, Schumacher F, et al. Cross-Cancer Genome-Wide Analysis of Lung, Ovary, Breast, Prostate, and Colorectal Cancer Reveals Novel Pleiotropic Associations. *Cancer Res*. 2016;76(17):5103–14. <https://doi.org/10.1158/0008-5472.CAN-15-2980>.
66. Glass C, Wuertzer C, Cui X, Bi Y, Davuluri R, Xiao YY, et al. Global Identification of EVI1 Target Genes in Acute Myeloid Leukemia. *PLoS One*. 2013;8(6). <https://doi.org/10.1371/journal.pone.0067134>
67. DiDonato JA, Mercurio F, Karin M. NF- κ B and the link between inflammation and cancer. *Immunol Rev*. 2012;246(1):379–400. <https://doi.org/10.1111/j.1600-065X.2012.01099.x>.
68. Fol M, Kozirski P, Kulesza J, Bialecki P, Druszczńska M. Dual nature of relationship between mycobacteria and cancer. *Int J Mol Sci*. 2021;22(15). <https://doi.org/10.3390/ijms22158332>
69. Cope RB, Stang B, Valentine BA, Bermudez LE. Topical exposure to exogenous ultraviolet-irradiated urocanic acid enhances *Mycobacterium ulcerans* infection in a Crl:IAF(HA)-hrBR hairless guinea-pig model of Buruli ulcer disease. *Photodermatol Photoimmunol Photomed*. 2004;20(1):14–20. <https://doi.org/10.1111/j.1600-0781.2004.00073.x>.
70. Ardies CM. Inflammation as cause for scar cancers of the lung. *Integr Cancer Ther*. 2003;2(3):238–46. <https://doi.org/10.1177/1534735403256332>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

