Integrative genomics sheds light on the immunobiology of tuberculosis in cattle

John F. O'Grady[1], Gillian P. McHugo[1], James A. Ward[1], Thomas J. Hall[1], Sarah L. Faherty O'Donnell[1,9], Carolina N. Correia[1,10], John A. Browne[1], Michael McDonald[1], Eamonn Gormley[2], Valentina Riggio[3,4], James G. D. Prendergast[3,4], Emily L. Clark[3,4], Hubert Pausch[5], Kieran G. Meade[1,6], Isobel C. Gormley[7], Stephen V. Gordon[2,6,8*], David E. MacHugh[1,6,8*]

[1] UCD School of Agriculture and Food Science, University College Dublin, Belfield, Dublin, D04 V1W8, Ireland.

[2] UCD School of Veterinary Medicine, University College Dublin, Belfield, Dublin, D04 V1W8, Ireland.

[3] The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Midlothian, EH25 9RG, UK.

[4] Centre for Tropical Livestock Genetics and Health (CTLGH), Roslin Institute, University of Edinburgh, Easter Bush Campus, EH25 9RG, UK.

[5] Animal Genomics, ETH Zurich, Universitaetstrasse 2, 8006, Zurich, Switzerland.

[6] UCD Conway Institute of Biomolecular and Biomedical Research, University College Dublin, Belfield, Dublin, D04 V1W8, Ireland.

[7] UCD School of Mathematics and Statistics, University College Dublin, Belfield, Dublin, D04 V1W8, Ireland.

[8] UCD One Health Centre, University College Dublin, Belfield, Dublin, D04 V1W8, Ireland.

[9] Present address: Irish Blood Transfusion Service, National Blood Centre, James's Street, Dublin, D08 NH5R, Ireland.

[10] Present address: Children's Health Ireland, 32 James's Walk, Rialto, Dublin, D08 HP97, Ireland.

[*] Corresponding authors

E-mail addresses:

JFOG: john.ogrady@ucdconnect.ie

GPM: gillian.mc-hugo@ucdconnect.ie

JAW: james.ward1@ucdconnect.ie

TJH: thomas.hall@ucd.ie

SFOD: sarah.fahertyo'donnell@ibts.ie

CNC: carolin.correia@gmail.com

JAB: john.a.browne@ucd.ie

MM: mcdonalm@ucd.ie

EG: egormley@ucd.ie

VR: valentina.riggio@roslin.ed.ac.uk

JDGP: james.prendergast@ed.ac.uk

ELC: emily.clark@roslin.ed.ac.uk

HP: hubert.pausch@usys.ethz.ch

KGM: kieran.meade@ucd.ie

ICG: claire.gormley@ucd.ie

SVG: stephen.gordon@ucd.ie

DEM: david.machugh@ucd.ie

Abstract

Mycobacterium bovis causes bovine tuberculosis (bTB), an infectious disease of cattle that poses a zoonotic threat to humans. Research has shown that bTB susceptibility is a heritable trait, and that the peripheral blood (PB) transcriptome is perturbed during bTB disease. Hitherto, no study has integrated PB transcriptomic, genomic and GWAS data to study bTB disease, and little is known about the genomic architecture underpinning the PB transcriptional response to *M. bovis* infection. Here, we perform transcriptome profiling of PB from 63 control and 60 confirmed *M. bovis* infected animals and detect 2,592 differently expressed genes that perturb multiple immune response pathways. Leveraging imputed genome-wide SNP data, we characterise thousands of *cis*- and *trans*-expression quantitative trait loci (eQTLs) and show that the PB transcriptome is substantially impacted by intrapopulation genomic variation. We integrate our gene expression data with summary statistics from multiple GWAS data sets for bTB susceptibility and perform the first transcriptome-wide association study (TWAS) in the context of tuberculosis disease. From this TWAS, we identify 136 functionally relevant genes (including *RGS10*, *GBP4*, *TREML2*, and *RELT*) and provide important new omics data for understanding the host response to mycobacterial infections that cause tuberculosis in mammals.

Introduction

Tuberculosis (TB) is a chronic infectious disease and a major source of ill health globally with over one billion people having died as a consequence of human TB (hTB) during the past two centuries[1] and with a further 1.3 million deaths reported in 2022[2], illustrating both the historical and persistent threat of the disease. The primary causative agent of hTB, *Mycobacterium tuberculosis*, forms part of the *Mycobacterium tuberculosis* complex (MTBC), a group of phylogenetically closely related bacteria exhibiting extreme genomic homogeneity that cause TB disease in mammals[3-6]. Another member of the MTBC, *Mycobacterium bovis*, is the chief causative agent of bovine tuberculosis (bTB), an endemic disease principally associated with cattle that imposes a significant economic impact on individual farmers and national economies[7,8]. As a zoonotic pathogen, *M. bovis* can transmit from animals to humans causing zoonotic TB (zTB), which disproportionally affects the Global South[9,10]. The most recent estimates, available for 2019, attributed more than 140,000 of new hTB cases and more than 11,000 deaths to zTB[11].

Previous research has shown that there are many shared characteristics between the pathogenesis of hTB and bTB, such that cattle can serve as a valuable large animal model to study TB disease in humans[12-15]. The primary route of infection for both *M. tuberculosis* and *M. bovis* is via the inhalation of aerosolised bacilli expelled by an infected individual or animal that are then phagocytosed by host alveolar macrophages (AM), establishing the primary site of infection in the lung. Normally, efficient pathogen killing is achieved by AMs through a range of innate immune response mechanisms including encasement of the bacilli within a phagolysosome, autophagy and apoptosis of infected cells, and by the production of antimicrobial peptides[16,17]. However, mycobacteria have evolved a range of strategies to manipulate innate immune responses, thereby facilitating colonisation, persistence, and replication within AMs[18-20]. Given the marked genomic similarities between *M. tuberculosis* and *M. bovis*, the close parallels between host-pathogen interactions and disease progression for hTB and bTB, and the zoonotic threat of *M. bovis*, a One Health approach to understanding the molecular mechanisms that underpin host immune responses and pathology in bTB can also provide important new information for tackling both hTB and zTB.

The genetic basis of susceptibility to *M. bovis* infection and bTB disease traits has been examined in cattle using focused candidate gene approaches[21-24]. Previous work has also highlighted the existence of substantial genetic variation for susceptibility to *M. bovis* infection in cattle populations[25,26]. In addition, genome-wide association studies (GWAS) have suggested susceptibility to *M. bovis* infection and bTB disease resilience traits are highly polygenic and influenced by interbreed genetic variation, which is reflected in modest replication of GWAS signals across multiple experiments[27-33]. Ultimately, identifying, cataloguing, and measuring the functional effects of these

polymorphisms will expand and enhance genomic prediction models for economically important traits such as resistance to *M. bovis* infection[34].

Expression quantitative trait loci (eQTLs) are genomic sequence variations—primarily single-nucleotide polymorphisms (SNPs)—that modulate gene expression and mRNA transcript abundance[35-39]. In this regard, SNPs that are significantly associated with a trait of interest often exert an eQTL regulatory effect[40-42]. This is observed for hTB, where infection response eQTLs detected in dendritic cells challenged with *M. tuberculosis* were enriched for SNPs associated with susceptibility to hTB[43]. In cattle, eQTLs and other regulatory polymorphisms have been shown to contribute a substantial proportion of the genetic variation associated with multiple complex traits[44,45]. A transcriptome wide association study (TWAS) is a multi-omics integrative strategy that combines gene expression data and independently generated GWAS summary data to discern explanatory links between genotypic variation, molecular phenotype variation, and phenotypic variation for a particular complex trait[46-50]. Notwithstanding recent methodological concerns[51], the TWAS approach can provide meaningful insights into the molecular basis of quantitative trait loci and an integrated knowledgebase of tissue-specific human TWAS associations, the TWAS Atlas, has recently been developed[52]. TWAS approaches have also been leveraged to identify genes with expression patterns that modulate phenotypic variability for economically important traits in cattle[53,54]. Various TWAS methods have been developed to study the effects of proximal genetic variants (*cis*-eQTLs) on transcriptional regulation[46,47,55,56] that do not consider distal/interchromosomal regulatory polymorphisms (*trans*-eQTLs), which are a major component of the omnigenic model of complex trait inheritance[57]. To address this, the Multi-Omic Strategies for TWAS (MOSTWAS) suite of tools has been developed, which extend traditional TWAS approaches to include *trans*-acting variants around regulatory biomarkers (e.g., transcription factor and microRNA genes) to increase the power to detect significant gene-trait associations[58].

It has previously been reported that the peripheral blood (PB) immune responses reflect those at the site of infection for bTB disease[59]. In this regard, our group and others have detected and characterised PB transcriptional biosignatures of *M. bovis* infection and bTB[60-70]. However, functional integration of PB transcriptomes, host genomic variation, and GWAS data sets for bTB susceptibility has not been performed previously. Additionally, to-date there have been no published studies that use the TWAS approach to understand the regulatory genome in the context of the host response to mycobacterial infections that cause TB in mammals. Therefore, using PB RNA-seq data from *M. bovis*-infected and control non-infected cattle, and imputed genome-wide SNP data, we combine an eQTL analysis with multiple bTB GWAS data sets[33] and conduct a summary TWAS

131 incorporating *trans*-acting genomic variants[58], which identifies important new genes underpinning
132 the mammalian host response to mycobacterial infections that cause TB.

133

Results

Animal disease phenotyping

Fig. 1 provides an overview of the experimental workflow and computational pipeline used for this study. A large panel of bTB reactor (bTB+; $n = 60$) and control (bTB−; $n = 63$) cattle were recruited that had a positive (reactor) and negative reaction, respectively, to the single intradermal comparative tuberculin test (SICTT). All animals were male, and the mean age of the animals was 21.9 ± 8.3 months. **Supplementary Table 1** provides detailed information about these animals, including the last four digits of the ear tag ID, date of sampling, and breed ancestry based on comprehensive pedigree information.

For the purposes of this study, and as a confirmatory test, the interferon gamma (IFN-γ) diagnostic assay was used to evaluate *M. bovis* infection status in all 123 recruited animals. The criterion for IFN-γ test positivity was a test result difference greater than 80 ELISA units for the purified protein derivative (PPD)-bovine (PPDb) IFN-γ value minus the PPD-avian (PPDa) IFN-γ value (ΔPPD)[71]. The mean ΔPPD (± SE) for the bTB+ animal group was 1170.35 ± 84.48 compared to −360.46 ± 55.17 for the bTB− group and this group difference was highly significant (two tailed Wilcoxon rank-sum test; $P < 3.258 \times 10^{-21}$) (**Supplementary Fig. 1**, **Supplementary Table 2**). One designated bTB− control animal produced a positive result for the IFN-γ test (C050, ΔPPD = 263.1) and two designated bTB+ animals elicited a negative result (T007, ΔPPD = 36.0; T062, ΔPPD = −52.3). These results yielded test sensitivity and specificity rates of 96.67% and 98.41%, respectively, which is in line with IFN-γ test performance under Irish conditions[71]. These animals were still designated as bTB− and bTB+, respectively, and included in subsequent analyses.

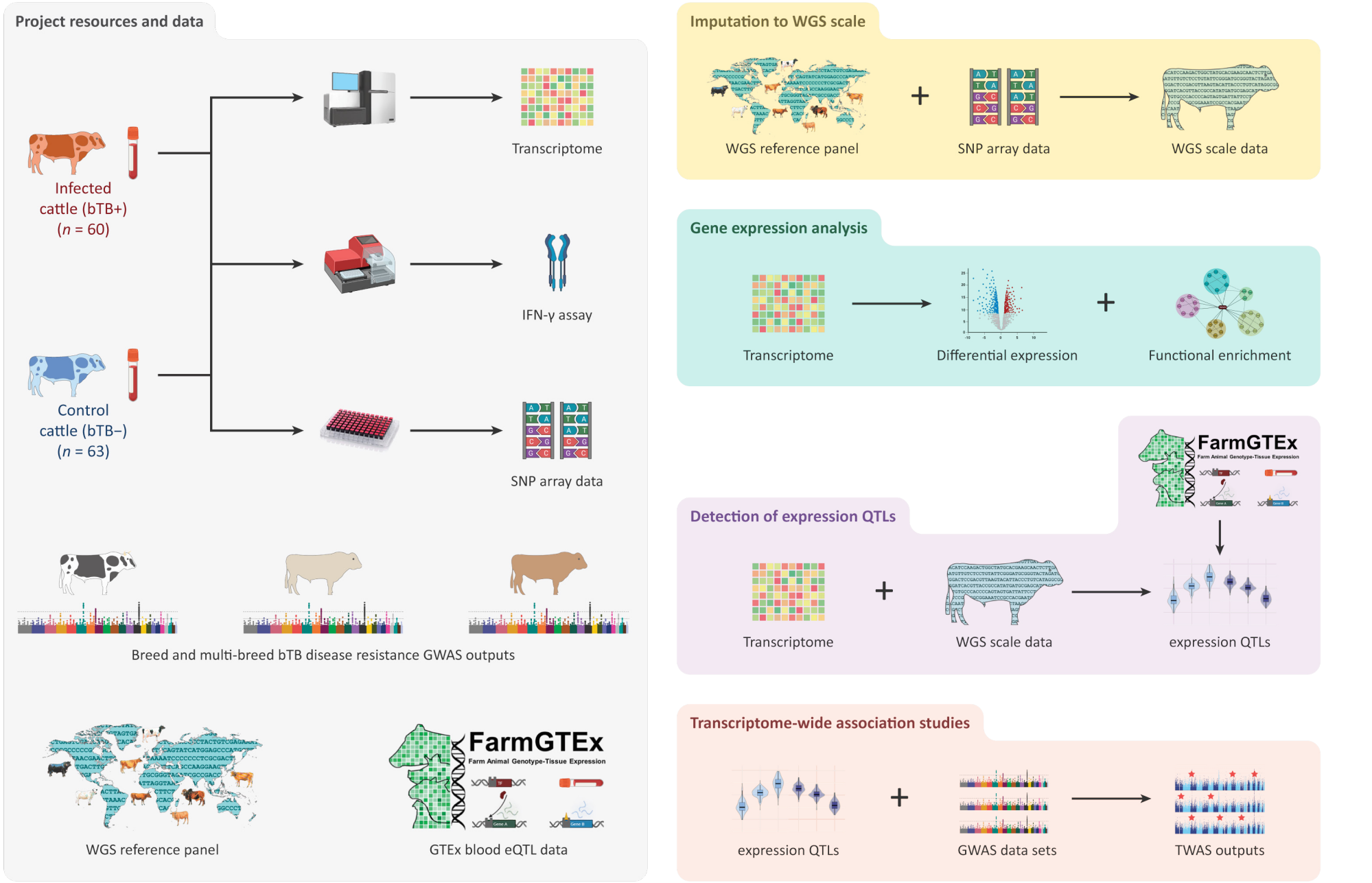
Project resources and data
Infected cattle (bTB+) (n = 60)
Control cattle (bTB–) (n = 63)
Transcriptome
IFN-γ assay
SNP array data
Breed and multi-breed bTB disease resistance GWAS outputs
WGS reference panel
FarmGTEx
Farm Animal Genotype-Tissue Expression
GTEx blood eQTL data
Imputation to WGS scale
WGS reference panel
SNP array data
WGS scale data
Gene expression analysis
Transcriptome
Differential expression
Functional enrichment
FarmGTEx
Farm Animal Genotype-Tissue Expression
Detection of expression QTLs
Transcriptome
WGS scale data
expression QTLs
Transcriptome-wide association studies
expression QTLs
GWAS data sets
TWAS outputs

Fig. 1: Experimental and computational workflow. Data resources for the project included; 1) newly generated high-resolution SNP-array data, peripheral blood RNA-seq data and interferon gamma (IFN-γ) release assay (IGRA) measurements from a reference panel of *n* = 60 bovine tuberculosis (bTB) reactor (bTB+) and *n* = 63 control (bTB−) cattle; 2) single and multi-breed GWAS summary statistics for bTB susceptibility from Ring *et al*., (2019)[33]; 3) whole genome sequence (WGS) data from Dutta *et al*., (2020)[72] and; 4) whole blood eQTL summary statistics from the Cattle GTEx consortium[53]. For the reference panel, SNP array genotype data was remapped to the ARS-UCD1.2 bovine genome build and imputed using the WGS cohort as a reference panel. RNA-seq data was aligned to ARS-UCD1.2 with the resulting count matrices normalised using various methodologies (See Methods) for inclusion in the differential expression, functional enrichment, and expression quantitative trait loci (eQTL) analyses. The normalised expression matrix was integrated with the imputed SNP-array data for the eQTL analysis. To assess the replication of eQTLs, we leveraged whole blood eQTL summary statistics from the Cattle GTEx consortium[53] and separately performed various permutation tests on identified *trans*-eQTLs. Finally, the GWAS summary statistics were remapped to ARS-UCD1.2 before being integrated with the reference panel eQTL results to conduct three single- and one multi-breed transcriptome wide association study (TWAS) for bTB susceptibility using the MOSTWAS software[58] (some figure components created with a Biorender.com license).

RNA-seq mapping statistics and genome-wide SNP imputation

Peripheral blood RNA sequencing yielded a mean of 35,129,315 ± 3,430,729 reads per individual sample library (*n* = 123 libraries and ± standard deviation). Reads were aligned to the ARS-UCD1.2 *B. taurus* genome build with a mean of 33,352,903 ± 3,206,593 (95.06% ± 0.76%) reads mapping uniquely, 779,866 ± 110,837 (2.22% ± 0.17%) mapping to multiple loci, 14,168 ± 2,639 (0.04 ± 0.008%) mapping to an excessive number of loci, 97,358 ± 274,880 (2.74 ± 0.68%) that were too short, and 15,020 ± 2,867 (0.04% ± 0.008%) that could not be assigned to any genomic locus. The mean mapped length was 297.8 ± 0.3 bp (**Supplementary Table 3**). None of the libraries exhibited an abnormal distribution of gene counts (**Supplementary Fig. 2**).

A total of 591,947 array-genotyped SNPs were available for analysis. To determine if any animal samples were inadvertently duplicated, we first LD-pruned the array genotype data following filtering of variants which were rare (MAF < 0.1) and which deviated from HWE ($P < 1 \times 10^{-6}$) to yield 34,272 SNPs. We then calculated the identity by state (IBS) among all animals using PLINK. We set a cut-off of 0.85 for deeming two samples as duplicates. All pairs of animals returned an IBS distance value < 0.8 (**Supplementary Fig. 3a**). Following this, we remapped the raw SNP array-data from the UMD 3.1 genome build to the ARS-UCD 1.2 reference genome and imputed the remapped variants up to WGS scale using a Global Reference Panel as a reference[72] (**Supplementary Note 4**). Imputation performance increased as MAF increased with poor performance observed at variants with a MAF ≤ 1% (**Supplementary Fig. 5)**. Following the removal of imputed variants that displayed poor imputation performance (R^2 < 0.6), possessed a low MAF (< 0.05), and that deviated

significantly from HWE ($P < 1 \times 10^{-6}$), a total of 3,866,506 imputed autosomal SNPs were retained for the eQTL analysis. Lastly, comparison of the imputed SNP profiles with RNA-seq reads using QTLtools showed that there were no sample mismatches and that the imputed WGS data correctly matched the transcriptomics data for all animals (**Supplementary Fig. 3b**).

Population genomics, differential gene expression, and functional enrichment analyses

The results of the genetic structure analysis using the ADMIXTURE program with 34,272 pruned genome-wide array SNPs and an inferred number of ancestral populations $K = 2$ are shown in **Fig. 2a**. A principal component analysis (PCA) plot of principal components (PC) 1 and 2 generated from the same set of pruned SNPs is shown in **Fig. 2b** with percentage Holstein ancestry and component 1 from the ADMIXTURE structure analysis also shown for each animal sample (see also **Supplementary Table 4**). The results of these two analyses were mutually compatible; component 1 from the ADMIXTURE structure plot was in concordance with PC1 (10.8% of the variance derived from the top 20 PCs from the PCA and likely corresponded, at least in part, to Holstein ancestry for the animals that had pedigree-derived breed composition data (113 out of 123 animals) (**Fig. 2b**, **Supplementary Table 1**). There was also a highly significant positive correlation (Spearman correlation (ρ) = 0.829, $P < 2.2 \times 10^{-16}$) between the pedigree-derived percentage Holstein ancestry values and component 1 from the ADMIXTURE structure plot (**Supplementary Fig. 6a**). PC2 (8.0% of the total variance of the top 20 PCs) likely accounts for population structure within the Holstein-Friesian populations, which has been documented previously in an independent cohort[27]. We observed that the genetic structure of the study population (bTB− and bTB+) was a confounder in the transcriptomics data set because there was sample clustering caused by breed ancestry observed in the PCA of the top 1,500 most variable genes determined from the variance stabilised transformed (VST) count matrix in DESeq2 (**Supplementary Fig. 6b**).

We performed a differential expression analysis (DEA) to identify differentially expressed genes (DEGs) between the reactor (bTB+) and control (bTB−) animal groups, which incorporated PC1 and PC2 (**Fig. 2b**), age in months, and sequencing batch (1 or 2) as covariates in the generalised linear model. With this approach, we identified 2,592 DEGs (FDR $P_{adj.} < 0.05$) for the bTB+ versus bTB− contrast (**Fig. 2c**, **Supplementary Table 5**). Within the bTB+ group, increased expression was observed for 1,638 DEGs and 954 DEGs exhibited decreased expression. We then selected a subset of 1,091 highly significant DEGs (FDR $P_{adj.} < 0.01$) for gene set overrepresentation and functional enrichment analyses using g:Profiler and IPA®. In this subset of DEGs, 602 and 489 genes exhibited increased and decreased expression, respectively in the bTB+ cohort.

Using the g:Profiler tool (FDR $P_{adj.} < 0.05$) we observed a clear enrichment for innate immune response, pathogen internalisation, and host-pathogen interaction GO terms and biological pathways

(**Fig. 2d**). The top significantly enriched functional entity was the *Defense response to virus* (FDR $P_{adj.} = 9.02 \times 10^{-13}$) GO:BP term. Other significantly enriched functional entities included: *Cytosolic pattern recognition receptor signalling pathway* (FDR $P_{adj.} = 1.24 \times 10^{-4}$) GO:BP term; *RIG-I-like receptor signalling pathway* (FDR $P_{adj.} = 3.88 \times 10^{-5}$) from KEGG; and *Antiviral mechanism by IFN-stimulated genes* (FDR $P_{adj.} = 7.67 \times 10^{-10}$) from the Reactome database. All significant results obtained from g:Profiler, in addition to the intersection of DE genes with the respective functional entities, are provided in **Supplementary Table 6**. For IPA®, a total of 996 DE genes and 14,228 background genes were successfully mapped. The significantly enriched (FDR $P_{adj.} < 0.05$) pathways identified from IPA® included *Interferon alpha/beta signalling* (FDR $P_{adj.} = 4.09 \times 10^{-8}$), *Oxidative phosphorylation* (FDR $P_{adj.} = 2.91 \times 10^{-6}$) and *Activation of IRF by cytosolic pattern recognition receptors* (FDR $P_{adj.} = 9.12 \times 10^{-3}$) (**Supplementary Fig. 7a**, **Supplementary Table 7**). Upstream transcriptional regulator analysis using IPA® revealed that the transcriptional regulator, ETV3 was the most significant upstream biological regulator of the inputted DE genes (FDR $P_{adj.} = 4.89 \times 10^{-19}$) (**Supplementary Fig. 7b**) Other important statistically significant (FDR $P_{adj.} < 0.05$) upstream regulators include TLR3, STING1, IRF5, and STAT1 (for complete results see **Supplementary Table 8**).

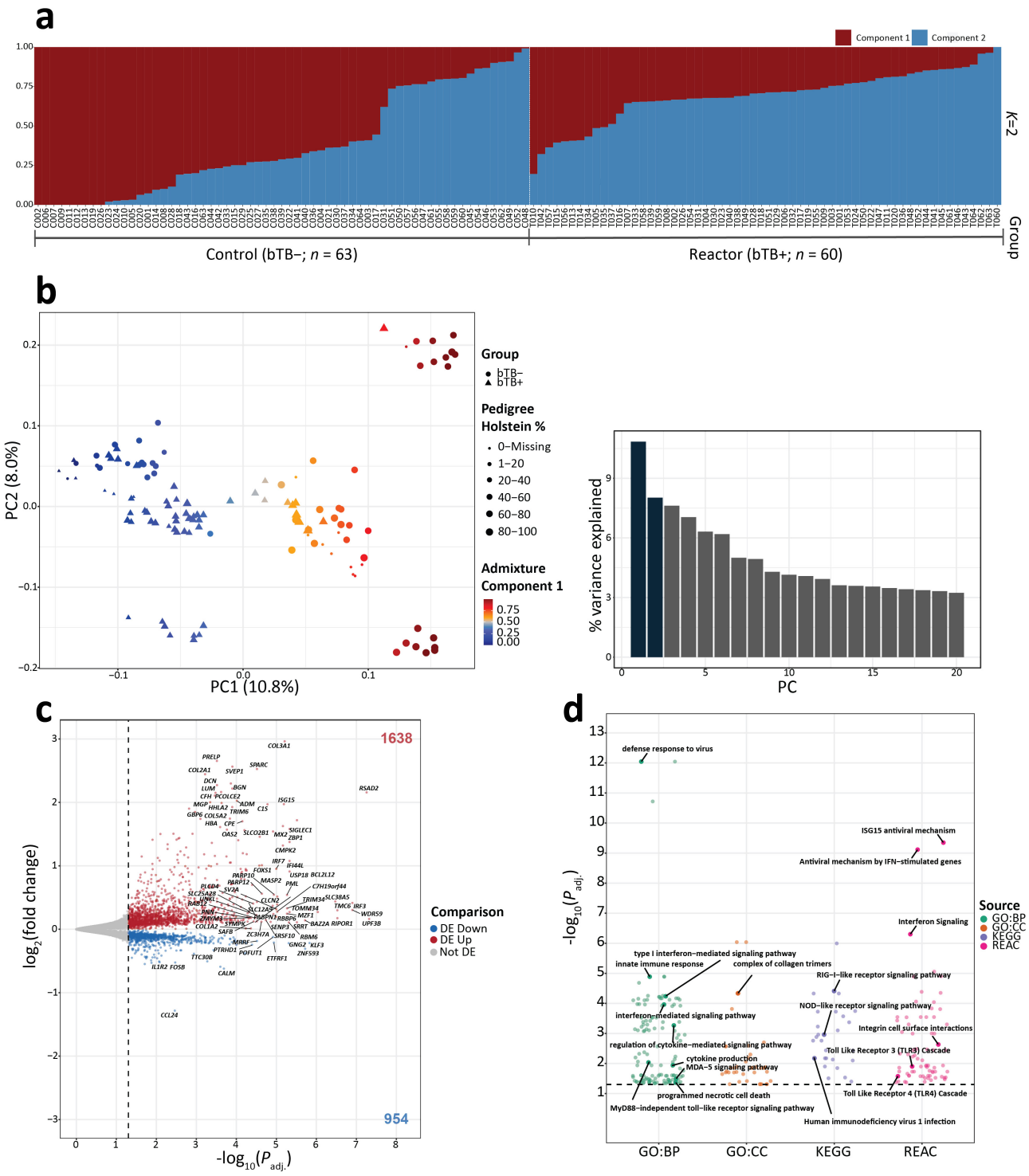

Fig. 2: Population genomics, differential expression, and functional enrichment analyses. a Structure plot showing the proportion of ancestry components 1 and 2 from the ADMIXTURE analysis for 123 animals (reactor bTB+ and control bTB−). **b** Principal component analysis (PCA) plot of PC1 and PC2 derived from 34,272 pruned SNPs genotyped in 123 animals. The data points are shaped based on their experimental designation, coloured based on the inferred ancestry component 1 from the ADMIXTURE analysis, and sized based on their reported pedigree Holstein percentage. A histogram plot of the relative variance contributions for the first 20 PCs is also shown with PC1 and PC2 accounting for 10.8% and 8.0% of the variation in the top 20 PCs, respectively. **c** Horizontal volcano plot of differentially expressed genes (DEGs) for the bTB+ (n = 60) versus bTB− (n = 63) contrast with thresholds determined by FDR $P_{adj.} < 0.05$ and an absolute $\log_2$ fold-change (LFC) > 0. The x-axis represents the $-\log_{10} P_{adj.}$ and the y-axis represents the $\log_2$ fold change. **d** Jitter plot of significantly impacted pathways/GO terms identified across the Gene Ontology (GO) Biological Processes (GO:BP), Cellular Compartment (GO:CC), Reactome (REAC) and Kyoto Encyclopaedia of Genes and Genomes (KEGG) databases using g:Profiler. The data points are coloured according to the corresponding database.

Identification of *cis*-expression quantitative trait loci

We used a linear regression model in TensorQTL to test associations between expressed genes and SNPs that passed filtering thresholds to identify local (± 1 Mb) *cis*-eQTLs in the reactor (bTB+) group (n = 60), the control (bTB−) group (n = 63), and a combined all animals group (AAG, n = 123). As covariates, we also included 1) the top five SNP genetic variation PCs (PC1-5) inferred for each group separately to account for interbreed differences between the animals; 2) age in months; 3) sequencing batch; 4) disease status (where applicable); and 5) transcriptomic PCs with PC1-8, PC1-9, and PC1-14 for the bTB+, bTB−, and AAG cohorts, respectively. The number of transcriptomic PCs to use was determined using the elbow method (**Supplementary Fig. 8**). We also removed known covariates (genotype PCs, age, batch, disease status) that were well captured by the inferred covariates (unadjusted $R^2 \geq 0.9$) and the final set of covariates for each cohort are detailed in **Supplementary Tables 9-11**. In total, we tested 14,701, 14,598, and 14,612 genes in the bTB+, bTB−, and AAG cohorts, respectively for *cis* SNP variants associated with their expression levels (**Supplementary Fig. 9a**).

Table 1 summarises the number of significant (FDR $P_{adj.} < 0.05$) *cis*-eQTLs, *cis*-eVariants, and *cis*-eGenes identified in all three groups. We identified 2,235, 3,419, and 6,676 *cis*-eGenes in the bTB+, bTB−, and AAG cohorts, respectively, with the largest proportion captured by the AAG group (**Fig. 3a, Supplementary Tables 12-14**). For each *cis*-eGene in each group, variants with a nominal P-value below the gene-level threshold (**Supplementary Fig. 9b**) were considered significant *cis*-eVariants. Overall, we identified 168,251, 415,861 and 1,103,004 significant *cis*-eVariant:gene associations in the bTB+, bTB−, and AAG cohorts. Of these *cis*-eVariants, 21.0%, 23.1% and 35.7% were associated with >1 *cis*-eGene. For all three groups, we identified hundreds to thousands of *cis*-eGenes with multiple independent acting *cis*-eQTLs (**Fig. 3b**). The conditional analysis detected

13.2%, 27.1% and 80.1% additional independent *cis*-eQTLs in the bTB+, bTB−, and AAG cohorts, respectively. We observed that the top *cis*-eQTL identified by the permutation analysis tended to cluster close to the transcriptional start site (TSS) of the associated gene, whereas conditionally independent *cis*-eQTLs were located at variable distances to the TSS (Wilcoxon rank-sum test; $P < 2.2 \times 10^{-16}$) (**Fig. 3c**). The permuted and conditional *cis*-eQTL associations were symmetrical around the TSS with no enrichment in the 5' or 3' directions (**Supplementary Fig. 10**). We noted a moderately negative but highly significant Spearman correlation in the effect size estimates of *cis*-eQTLs and their respective distances to the TSS of the associated gene in all three groups (**Fig. 3d**).

Table 1: Total number of and unique number of significant (FDR $P_{adj.} < 0.05$) *cis*-eQTLs, *cis*-eVariants and their corresponding *cis*-eGenes identified across the reactor (bTB+), control (bTB−) and combined all animals (AAG) cohorts, respectively.

Class of eQTL/eGene	Distance between associated pair	bTB+ (*n* = 60)	bTB− (*n* = 63)	AAG (*n* = 123)
cis-eQTLs (permutation)	± 1 Mb	2,235	3,419	6,676
Conditionally independent *cis*-eQTLs	± 1 Mb	295	925	5,385
Total number of independent *cis*-eQTLs	± 1 Mb	2,530	4,344	12,061
cis-eGenes	± 1 Mb	2,235	3,419	6,676
cis-eVariant associations	± 1 Mb	168,251	415,861	1,103,004
Unique *cis*-eVariants	± 1 Mb	139,616	319,734	709,337

To assess replication of the *cis*-eQTLs identified in this study, we used three metrics: allelic concordance (AC), the π_1 statistic to measure the proportion of true positive associations, and the Spearman correlation coefficient of effect size estimates in an external set of whole-blood *cis*-eQTL summary statistics obtained from the Cattle GTEx Consortium[53]. We observed high AC between top and significant *cis*-eQTLs identified in this study and significant *cis*-eQTLs identified in the Cattle GTEx ($AC_{bTB+} = 99.27\%$, $AC_{bTB-} = 99.10\%$, and $AC_{AAG} = 98.87\%$). We observed moderate to high π_1 statistics across all groups indicating good replication ($\pi_{1bTB+} = 0.791 \pm 0.0009$, $\pi_{1bTB-} = 0.685 \pm 0.0007$, and $\pi_{1AAG} = 0.605 \pm 0.0004$) (**Fig. 3e**). We also noted a positive and significant Spearman correlation in effect size estimates for the top significant eQTLs identified in our study and the matched variants from the Cattle GTEx across all three groups ($\rho_{bTB+} = 0.797$, $\rho_{bTB-} = 0.783$, and $\rho_{AAG} = 0.761$) (**Fig. 3f**).

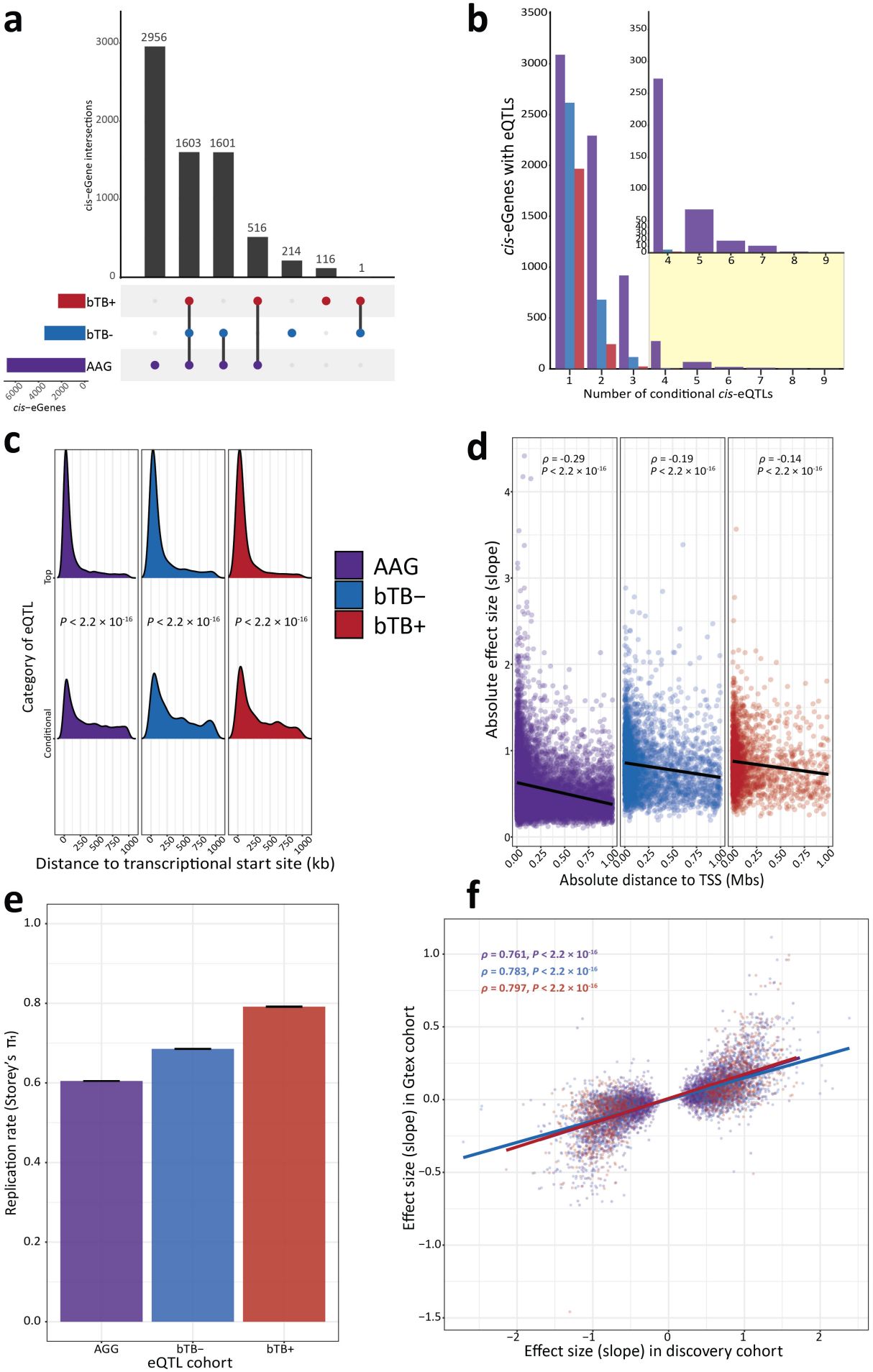

Fig. 3: *Cis*-expression quantitative trait loci (eQTL) mapping and external replication results. a Upset plot showing the intersection of shared *cis*-eGenes identified in the reactor (bTB+), control (bTB−), and combined all animal (AAG) cohorts, respectively. **b** Barplot illustrating the number of genes with a significant primary or conditional *cis*-eQTL for degrees 1–9 across all three groups. Inset shows the number of genes for *cis*-eQTL degrees 4–9. **c** Ridgeline plot showing the distribution of the absolute distance from the transcriptional start site (TSS) of top and conditional *cis*-eQTLs identified in all three groups. *P*-values are inferred from the Wilcoxon rank-sum test between top and conditional *cis*-eQTLs within each group. **d** Scatter plot illustrating the relationship between absolute *cis*-eQTL effect size and distance to the TSS for all significant *cis*-eQTLs (top and conditional) identified in each group separately. Spearman correlation values are also reported in addition to the corresponding *P*-value representing the significance level of each respective correlation. The black line indicates line of best fit. **e** Replication rate as measured by Storey's π_1 in the current study and whole blood *cis*-eQTLs identified in the Cattle GTEx[53]. The error bars indicate the standard error from 100 bootstrap samplings. **f** Scatterplot illustrating the effect sizes of significant *cis*-eQTLs identified in this study and matched variant-gene pairs identified in the Cattle GTEx. Spearman correlation values are also reported in addition to the corresponding *P*-value representing the significance level of each respective correlation. The coloured lines indicate lines of best fit within each group, respectively. The colour scheme for each group is consistent throughout the figure.

Cis-eQTLs regulate the peripheral blood transcriptional response to *M. bovis* infection

We compared the *cis*-eGenes identified in the bTB+ and the bTB− groups and that were also replicated in the AAG group (**Fig. 3a**) to assess if there were genomic variants influencing the PB transcriptomes for each of these biological states. This approach facilitated identification of; 509 bTB+ only (bTB+) *cis*-eGenes, identified in the bTB+ and AAG cohorts but not the bTB− group; 1603 *cis*-eGenes that were identified across all three groups; and 1593 bTB− only (bTB−) *cis*-eGenes identified in the bTB− and AAG cohorts but not the bTB+ group (**Fig. 4a**). We then performed a gene set overrepresentation analysis of *cis*-eGenes for four groups (bTB−, bTB− and AAG, bTB+, bTB+ and AAG) using g:Profiler (**Fig. 4b**). Significantly overrepresented (FDR $P_{adj.} < 0.05$) functional entities identified using *cis*-eGenes identified in the bTB− only and the AAG cohorts included the *MHC class II protein complex* GO:CC term; The *ER-Phagosome pathway* Reactome term and the *Leishmaniasis* KEGG term. Considering *cis*-eGenes identified only in the bTB− group, significantly impacted pathways included the *Succinyl-CoA metabolic process* and the *antigen processing and presentation of peptide antigen* GO:BP terms. *Cis*-eGenes identified in the bTB+ and AAG group were significantly overrepresented in pathways that included the *Th1 and Th2* and *Th17 cell differentiation* KEGG terms and *Phsophorylation of CD3 and TCR Zeta chains* Reactome term. In the bTB+ group, we observed a number of GO:BP terms significantly overrepresented by *cis*-eGenes including; *Negative regulation of chemokine (C-C motif) ligand 4* and *5 production* and *Negative regulation of macrophage inflammatory protein 1 alpha production*. All overrepresentation

results obtained using g:Profiler for the analysis of these four gene sets are detailed in **Supplementary Table 15-18**.

To identify DE *cis*-eGenes, we focused on the *cis*-Genes identified in the AAG group that overlapped with the DEG results (**Fig. 2c**). Of the 2,592 DE genes, 2,388 (92.12%) were tested in the *cis*-eQTL analysis. A total of 1,059 DEGs were characterised as *cis*-eGenes and 1,329 DEGs were not (**Fig. 4c**). We did not identify a significant association between DEGs and genes characterised as being *cis*-eGenes (chi-square test; $\chi^2 = 2.0068$, $P = 0.1566$). For the 1,059 DE *cis*-eGenes, we conducted a g:Profiler overrepresentation analysis using the set of genes that overlapped between the DEG and the AAG *cis*-eQTL analyses as the background set. Significantly impacted pathways and GO terms perturbed by these DE *cis*-eGenes included the *Defense response to virus* GO:BP term (FDR $P_{adj.} = 1.12 \times 10^{-3}$), the *Viral life cycle – HIV-1* (FDR $P_{adj.} = 1.76 \times 10^{-2}$) KEGG pathway and the *Methylosome* (FDR $P_{adj.} = 2.57 \times 10^{-2}$) GO:CC term (**Fig. 4d**, **Supplementary Table 19**).

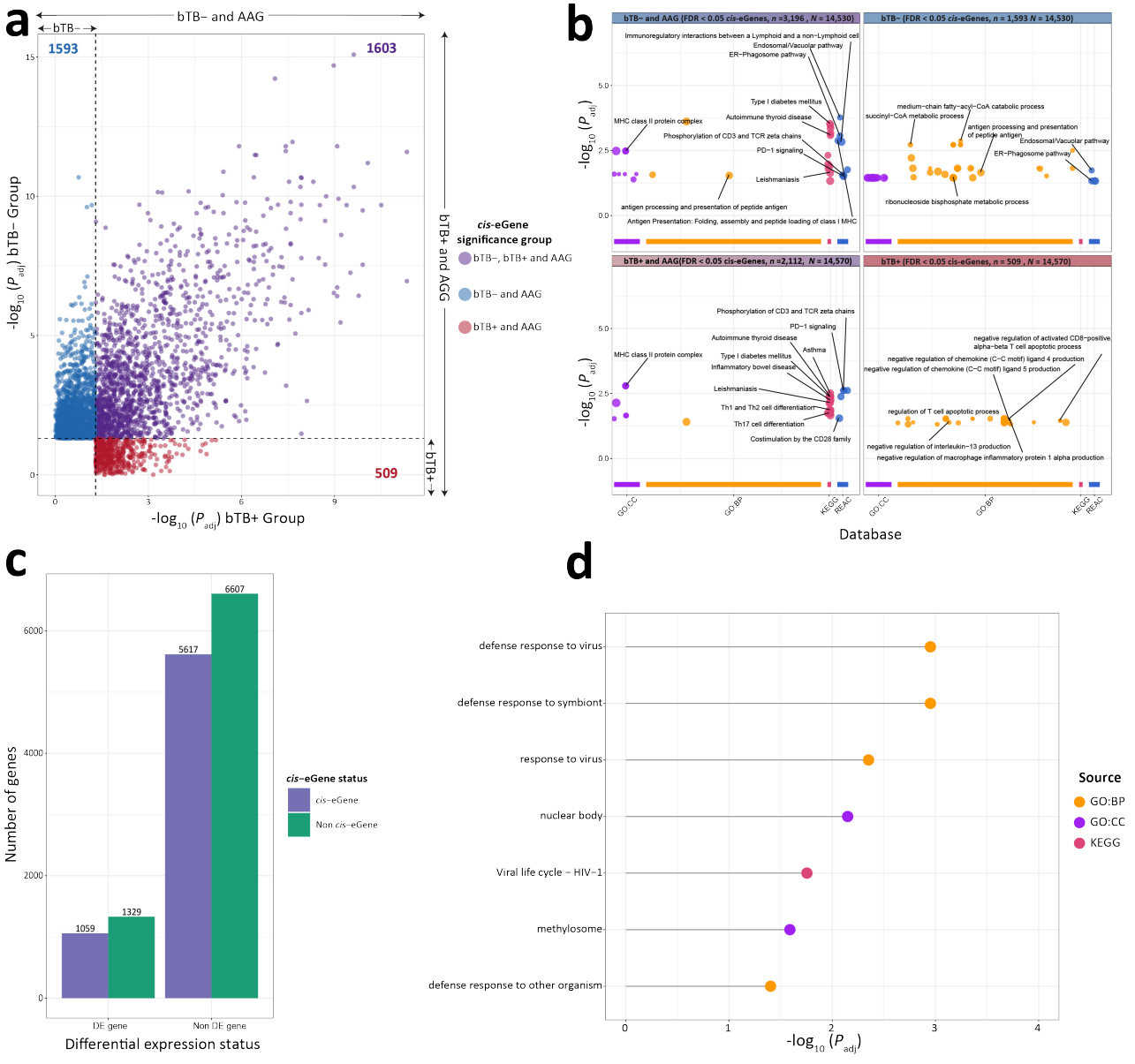

Fig. 4: Integrative *cis*-eGene analysis. a Scatter plot of significant *cis*-eGenes identified in the control (bTB−) and combined all animals (AAG) cohorts but not the reactor (bTB+); those identified across all three groups (bTB−, bTB+, and AAG); and those identified in the bTB+ and AAG cohorts but not the bTB− group (bTB+). The *y*-axis corresponds to the most significant FDR $P_{adj.}$ variant-gene pair identified in the bTB− group and the *x*-axis corresponds to the most significant $P_{adj.}$ variant-gene pair identified in the bTB+ group. Dashed lines indicate and FDR $P_{adj.} < 0.05$. **b** Significantly impacted pathways/GO terms by *cis*-eGenes from the bTB−, bTB−/AAG, bTB+, and bTB+/AAG cohorts across Gene Ontology (GO) Biological Processes (GO:BP), Cellular Compartment (GO:CC), Reactome (REAC), and the Kyoto Encyclopaedia of Genes and Genomes (KEGG) databases using g:Profiler. The number of input genes for each set (*n*) and number of background genes (*N*) for each set is also detailed. Data points are coloured based on their corresponding database **c** Barplot showing the classification of genes tested in both the *cis*-eQTL and differential expression analysis that were classified as differentially expressed (DE) or not DE *cis*-eGenes. **d** Lollipop chart showing significantly impacted pathways and GO terms (FDR $P_{adj.} < 0.05$) for the 1,059 DE-*cis*-eGenes across the GO:BP, GO:CC, and KEGG databases. The pathways are ordered based on adjusted *P*-value and are coloured based on their corresponding database.

Mapping of *trans*-expression quantitative trait loci is confounded by bovine population genetic structure.

We employed a linear regression model in QTLtools that included the same inputs as the *cis*-eQTL mapping procedure to characterise distal intrachromosomal (> 5 Mb) and interchromosomal *trans*-eVariants. **Table 2** summarizes the numbers of intra- and interchromosomal *trans*-eVariants and *trans*-eGenes detected in all three groups (bTB+, bTB−, and AAG) and **Fig. 5a** shows the overlap of *trans*-eGenes across these groups. In total, we identified 497, 916, and 5,314 *trans*-eVariants (FDR $P_{adj.} < 0.05$) in the bTB+, bTB−, and AAG cohorts, which were associated with 13, 17 and 107 *trans*-eGenes, respectively (**Fig. 5a, Supplementary Table 20-22**). Because of the relatively small numbers of *trans*-eGenes identified in the bTB+ and bTB− groups, we focused on the AAG set of *trans*-eGenes for a more detailed analysis.

Identification, biological interpretation, and replication of peripheral blood *trans*-eQTLs is challenging owing to the heterogenous nature of the tissue and the small effect sizes associated with putative distal variants[73]; however, notwithstanding these limitations, we observed an inflated number of *trans*-eQTLs in this study compared to previous reports in humans[74]. We first focused on the 588 intrachromosomal *trans*-eVariants associated with 26 *trans*-eGenes (**Supplementary Fig. 11a**). Among the intrachromosomal *trans*-eVariants associated with the same intrachromosomal *trans*-eGene, we observed a high LD genotype correlation between these variants ($r = 0.889 \pm 0.243$SD) (**Supplementary Fig. 11b**). We therefore selected the most significant intrachromosomal *trans*-eVariant for each gene and computed the LD between this variant and the top *cis*-eQTL of the same gene. In total, 24 genes had a significant *cis*-eQTL and intrachromosomal *trans*-eVariant associated

with its expression levels. We observed high LD amongst top intrachromosomal *trans*-eVariants and top *cis*-eQTLs of the same gene (r = 0.562 ± 0.203SD). To determine whether our observed LD pattern was significantly greater than what would be expected by chance, we randomly sampled 10,000 sets of 24 variant pairs which were no less than 5 Mb and no greater than 14,065,301 bp apart (the latter cutoff was two standard deviations of the distribution of distances between top *trans*-eVariants and top *cis*-eQTLs for the same gene). We calculated the medians and means of these 10,000 sets to generate two null distributions. We then calculated a permuted *P*-value ($P_{perm.}$) defined as the proportion of permutations with a median and mean intrachromosomal LD relationship at least as large or greater than the observed set. After this procedure, we obtained a permuted *P*-value of < 0.0001 indicating that our observed set of intrachromosomal *trans*-eVariants was significantly inflated by LD (**Supplementary Fig. 11c, Supplementary Table 23**).

We next focused on the 4,726 interchromosomal *trans*-eVariants associated with 81 *trans*-eGenes. We again selected the most significant SNP associated with each *trans*-eGene and calculated the LD between these interchromosomal *trans*-eVariants and the top *cis*-eQTL of the same gene. In total, 23 genes had a significant interchromosomal *trans*-eVariant and *cis*-eQTL associated with its expression levels. We observed a complex interchromosomal LD pattern between *cis*-eVariants and *trans*-eVariants of the same gene (r = 0.280 ± 0.199SD) (**Fig. 5b**). To assess whether our observed LD pattern was significantly greater than what would be expected by chance, we first sampled for each *trans*-eVariant with replacement, 1000 null *trans*-eVariants from the same chromosome and same allele frequency as putative *trans*-eVariants. We then computed the LD relationship between these null *trans*-eVariants and the *cis*-eQTLs of interest and then randomly generated 10,000 sets of 23 null interchromosomal *trans*-eVariants and the corresponding top *cis*-eQTL pairs. We performed the same procedure used for the intrachromosomal analysis to generate two null distributions with two $P_{perm.}$ values < 0.0001, which indicated that our top interchomosomal *trans*-eVariants were in high LD with top *cis*-eQTLs of the same gene (**Fig. 5c**, **Supplementary Table 24**).

Table 2: Total number of and unique number of significant (FDR $P_{adj.} < 0.05$) intrachromosomal and interchromosomal *trans*-eVariants and *trans*-eGenes identified across the reactor (bTB+), control (bTB−) and combined all animals (AAG) cohorts, respectively.

Class of eQTL/eGene	Distance between associated pair	bTB+ (*n* = 60)	bTB− (*n* = 63)	AAG (*n* = 123)
Trans-eVariant associations	> 5 Mb	497	916	5,314
Trans-eGenes	> 5 Mb	13	17	107
Unique *trans*-eVariants	> 5 Mb	497	704	4,976
Intrachromosomal *trans*-eVariants	> 5 Mb and on same chromosome	0	215	588
Interchromosomal *trans*-eVariants	Different chromosome	497	701	4,726
Intrachromosomal *trans*-eGenes	> 5Mb and on same chromosome	0	2	26
Interchromosomal *trans*-eGenes	Different chromosome	13	15	81

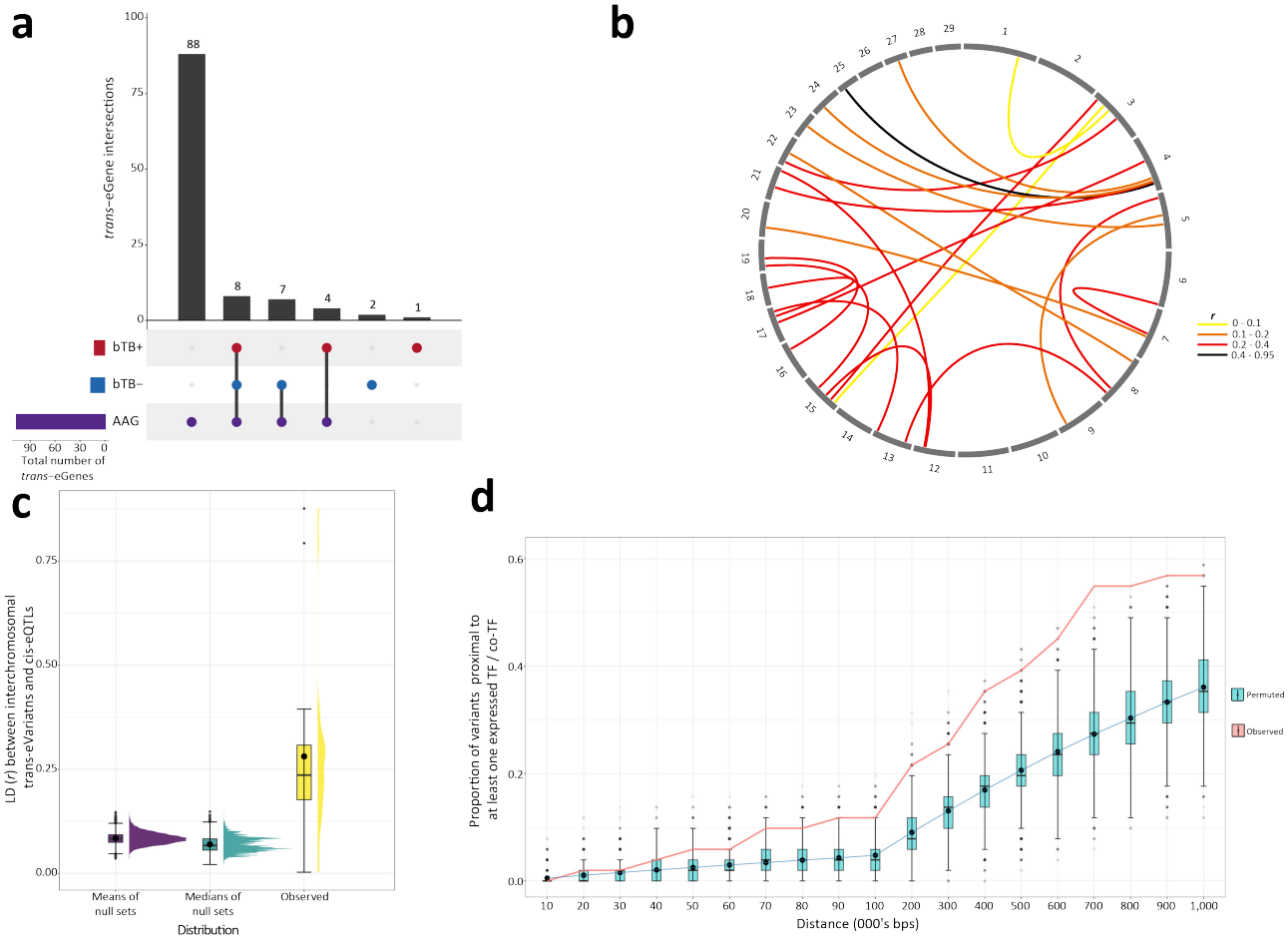

Fig. 5: *Trans*-expression quantitative loci mapping results and downstream analysis of inter-chromosomal *trans*-eVariants. a Upset plot showing the intersection of shared *trans*-eGenes identified in the reactor (bTB+), the control (bTB−), and the combined all animal group (AAG) cohorts, respectively. **b** Circos plot showing the linkage disequilibrium (LD) (*r*) relationship between the top interchromosomal *trans*-eVariants and top *cis*-eQTLs for the same gene. **c** Comparison of observed interchromosomal LD relationship (*r*) between 23 top interchromosomal *trans*-eVariants and *cis*-eQTLs of the same gene versus the mean and median distributions of 10,000 sets of 23 interchromosomal null *trans*-eVariant and top *cis*-eQTL pairs. **d** The proportion of observed (blue) highly significant top *trans*-eVariants (FDR $P_{adj.} < 0.01$) not in LD with top *cis*-eQTLs of the same gene ($r^2 < 0.01$) residing close to at least one expressed transcription factor (TF) or TF-cofactor at various (±) intervals versus 10,000 null sets of 51 SNPs (red). Horizontal lines inside the boxplots show the medians, solid circles indicate the means. Box bounds show the lower quartile (Q1, the 25th percentile) and the upper quartile (Q3, the 75th percentile). Whiskers are minima (Q1 – 1.5 × IQR) and maxima (Q3 + 1.5 × IQR) where IQR is the interquartile range (Q3-Q1).

Trans-eVariants cluster close to expressed transcription factors and co-transcription factors.

We next filtered putative *trans*-eVariants to retain variants with a highly significant (FDR $P_{adj.} < 0.01$) *trans* association and that were not in LD (genotype squared correlation (r^2) > 0.01) with a *cis*-eQTL of the same gene. This reduced the number of *trans*-eVariants and *trans*-eGenes available for analysis to 3,934 and 51, respectively. We hypothesised that these *trans*-acting variants resided close to expressed transcription factors (e-TFs) or transcription factor co-factors (e-coTFs) and would regulate *trans*-eGenes through influencing expression of the e-TFs/e-coTFs in *cis*. To investigate this, we first selected the top *trans*-eVariant associated with each *trans*-eGene and downloaded the genomic locations of 2,384 annotated TFs/coTFs from the AnimalTFDB v.4.0 database[75]. Of these, 973 (40.81%) passed expression filtering thresholds for inclusion in the AAG eQTL analysis. We next calculated the proportion of the 51 most significant *trans*-eVariants that resided close to at least one of the 973 TFs/co-TFs at various distances ranging from ±10 kb to ±1 Mb versus a random set of 51 SNPs computed 10,000 times to generate a null distribution. We calculated a permuted *P*-value ($P_{perm.}$) defined as the number of sets with a proportion of null *trans*-eVariants proximal to at least one expressed TF/co-TF equal to or greater than the observed proportion divided by 10,000. Across distance windows from ± 70kb – 1Mb, we noted that our observed proportion was significantly higher ($P_{perm.} < 0.05$) than that expected by chance. (**Fig. 5d**, **Supplementary Table 25**).

Transcriptome wide association analyses highlight genes associated with bTB susceptibility

To assess if expression patterns in the three groups of animals (bTB+, bTB−, and AAG) were correlated to bTB susceptibility, we used MOSTWAS[58] to generate predictive models of expression and combined these, using a TWAS approach, with SNP summary statistics from multiple GWAS data sets for bTB susceptibility in four breed cohorts (Holstein-Friesian – HF, Charolais – CH, Limousin – LM, and a multi-breed panel – MB)[33]. The SNPs in these GWAS data sets were originally mapped to the UMD3.1 genome assembly and were therefore remapped to the ARS-UCD1.2 assembly for this TWAS. We first computed 29,905, 91,822, and 1,046,632 significant (FDR $P_{adj.} < 0.01$) correlations between expressed cis-eTFs/coTFs and *cis*-eGenes in the bTB+, bTB−, and AAG cohorts, respectively. We then used the *MeTWAS* function in MOSTWAS to build predictive models of expression for *cis*-eGenes within each group. In total, we trained 1,604, 2,502, and 3,957 expression models in the bTB+, bTB−, and AAG cohorts, respectively (**Table 3**). The expression patterns of these genes were significantly heritable ($P < 0.05$) and achieved a McNemar's five-fold cross-validated predicted R^2 value ≥ 0.01 within the *MeTWAS* function. For each reference group and each GWAS cohort, we conducted a weighted burden test using the MOSTWAS *BurdenTest* function to identify genes with expression patterns correlated to bTB susceptibility. For genes that were significant at a Bonferroni-adjusted P-value < 0.05, we conducted a permutation test conditioning on the GWAS effect size and genes with a permuted P-value < 0.05 were considered significantly associated with bTB susceptibility.

The number of genes that were significant after Bonferroni correction, and that remained significant after the permutation procedure in each of the 12 TWAS groups are shown in **Table 3**. In total, across all four GWAS cohorts (HF, CH, LM, and MB) we identified 31, 33 and 72 TWAS genes significantly associated with bTB susceptibility in the bTB+, bTB−, and AAG cohorts, respectively (**Fig. 6**). Among the cohorts, there was little overlap between TWAS genes, with many genes emerging as breed- and expression model-specific (**Supplementary Fig. 12**). Overall, we identified 136 genes dispersed across the genome with expression patterns correlated with bTB susceptibility (**Fig. 6**). Our TWAS analysis highlighted immunobiologically relevant genes such as *RGS10, GBP4, TREML2*, and *RELT* and the full results of all TWAS associations for each reference panel are provided in **Supplementary Table 26-28**.

Table 3: Total number of significantly heritable ($P < 0.05$) predictive expression models ($R^2 > 0.01$) generated for the reactor (bTB+), control (bTB−) and combined all animals (AAG) cohorts with the corresponding Bonferroni adjusted *P*-value cut-off for association and number of significant genes identified across all four GWAS data sets. Numbers in brackets indicate the number of TWAS genes significant after permutation testing.

Group	Expression models	*P*-value cut-off	Limousin (LM) TWAS genes	Holstein-Friesian (HF) TWAS genes	Charolais (CH) TWAS genes	Multi-breed (MB) TWAS genes
bTB+	1,604	3.12×10^{-5}	11 (3)	14 (7)	14 (9)	27 (12)
bTB−	2,502	2.00×10^{-5}	10 (3)	30 (15)	11 (6)	18 (9)
AAG	3,957	1.26×10^{-5}	22 (12)	46 (31)	24 (10)	53 (19)

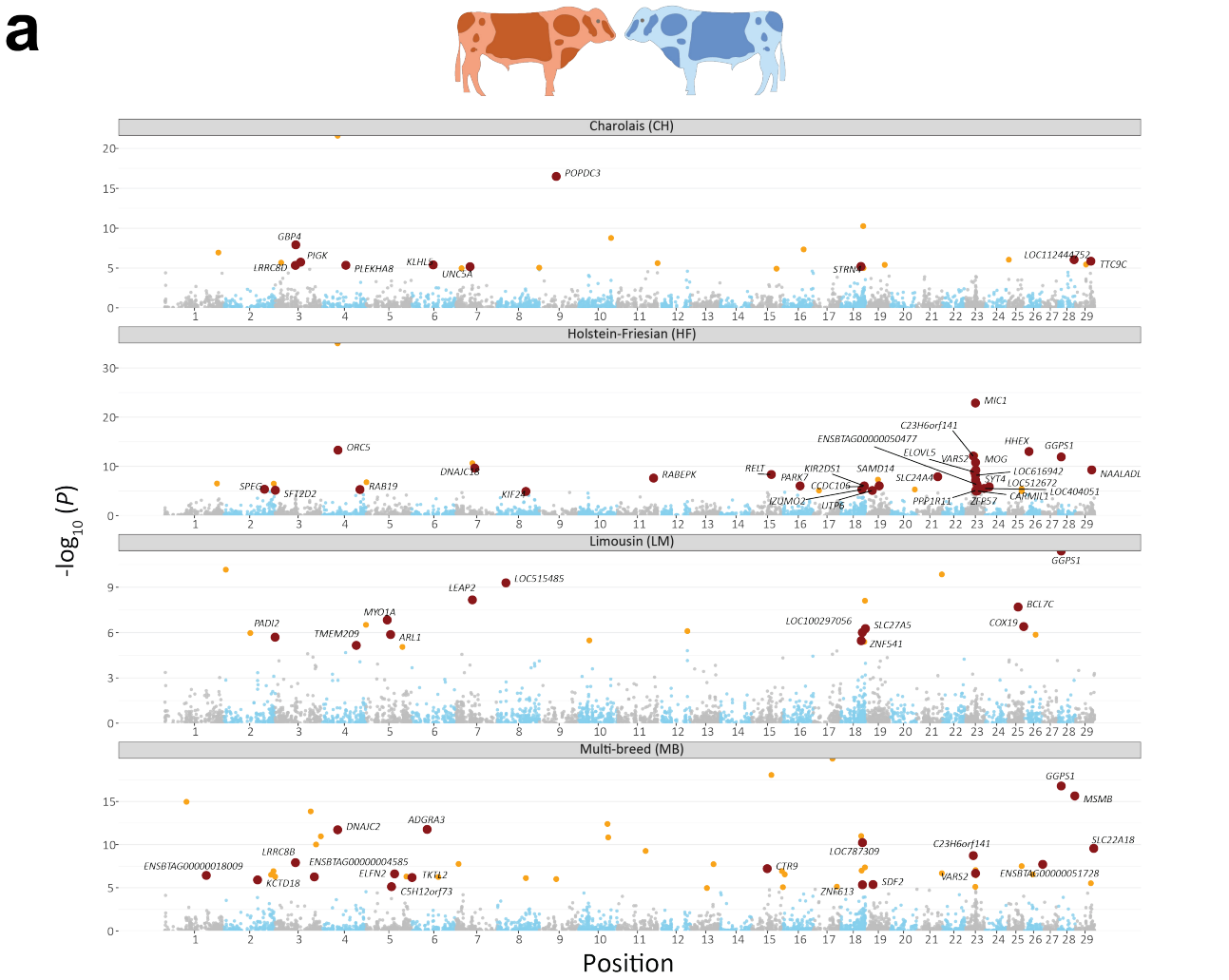

Fig. 6: Transcriptome wide association analysis (TWAS) results. a Manhattan plots showing all TWAS associations for expression models generated in the analysis of all animals combined and imputed into four GWAS data sets (Charolais (CH), Holstein-Friesian (HF), Limousin (LM), and Multi-Breed (MB)). Yellow data points have a Bonferroni FDR $P_{adj.} < 0.05$, and red points correspond to genes that have a Bonferroni $P_{adj.} < 0.05$, and $P_{perm.} < 0.05$. Labelled genes correspond to red data points in the plot. **b** Volcano plot highlighting significant TWAS associations for expression models generated in the reactor group (bTB+). The *x*-axis indicates the TWAS *Z*-score, and the *y*-axis shows the nominal ($-\log_{10}$ scale) *P*-value of association. Associations are coloured based on the GWAS data set for which the expression model was imputed into. Associations are shaped according to whether they had a Bonferroni $P_{adj.} > 0.05$ (circle), $P_{adj.} < 0.05$ (triangle), or $P_{adj.} \leq 0.05$ and $P_{perm.} < 0.05$ (square). The dashed line corresponds to a Bonferroni $P_{adj.}$ cut-off ($P < 3.12 \times 10^{-5}$). **c** Volcano plot highlighting significant TWAS associations for expression models generated in the control group. The *x*-axis indicates the TWAS *Z*-score, and the *y*-axis shows the nominal *P*-value of association. Associations are coloured based on the GWAS data set for which the expression model was imputed into. Associations are shaped according to whether they had an FDR $P_{adj.} > 0.05$ (circle), $P_{adj.} < 0.05$ (triangle), or $P_{adj.} \leq 0.05$ and $P_{perm.} < 0.05$ (square). The dashed line corresponds to a Bonferroni $P_{adj.}$ cut-off ($P < 2.00 \times 10^{-5}$). The figure legend for panel **b** and panel **c** is common to both.

Discussion

We present a comprehensive multi-omics analysis, which integrates genomics, bovine PB transcriptomics and GWAS data sets for bTB susceptibility to improve our understanding of how genetic factors contribute to the interindividual variability in response to *M. bovis* infection and mycobacterial infections more broadly in a One Health context. Moreover, this study is the first application of the TWAS approach to dissecting the genomic architecture of a susceptibility trait for a mycobacterial infection that causes TB in mammals.

Bovine TB disease susceptibility is a moderately heritable quantitative trait (estimated h^2 ranges between 0.08 and 0.14) with a highly polygenic and breed-specific genetic architecture that poses significant challenges for functional assignment of QTLs identified from GWAS experiments[33,76-78]. However, understanding the biology of these QTLs will be important in bridging the genome to phenome gap for bTB disease resilience because regulatory QTLs, especially *cis*- and *trans*-eQTLs, contribute a large proportion of the variance in complex trait heritability[44,45]. Additionally, it has been estimated that up to 50% of GWAS signals are shared with at least one molecular phenotype in humans[79], with a particular enrichment observed for regulatory QTLs associated with proximal and distal gene expression regulation in PB[80].

Analysis of differential gene expression using RNA-seq showed that the bovine PB transcriptome is substantially perturbed by *M. bovis* infection with 2,592 genes significantly (FDR $P_{adj.} < 0.05$) DE (**Fig. 2c**, **Supplementary Table 5**). We detected fewer DEGs in comparison to those reported by McLoughlin, et al. [64] who analysed PB leukocytes from cattle infected with *M. bovis*. However, we identified more DEGs than McLoughlin, et al. [68] who analysed whole blood RNA-seq data from calves infected with *M. bovis* across an experimental time-course. The variability observed in this study among animals, characterised by differences in breed composition, age, duration since *M. bovis* infection, and the varied biological tissue analysed, along with the diverse cell composition associated with PB may explain the heterogenous nature of the bovine transcriptomic profile. Contrary to this, the experiments conducted by McLoughlin, et al. [64] and McLoughlin, et al. [68] featured a more controlled setting, involving Holstein-Friesian calves matched for age and breed. These observations are supported by other studies showing that population genetic structure impacts gene expression due to allele frequency differences at *cis*-eQTL sites[81], and that ancestry effects impact the human response to viral infection in a cell type-specific manner[82]. Many of the DEGs detected here (38%) were also observed to be DE at 48 hours post infection (hpi) in bovine alveolar macrophages (bAM) challenged with *M. bovis* and were components of gene modules key to the innate immune response[83]. These shared genes included, but were not limited to, *MX2, MX1, OAS2, ISG15*, and *IRF7* that collectively constitute interferon-stimulated genes[84]. This is also reflected in

our gene set enrichment analysis of highly significant DEGs where many of the top significant overrepresented functional entities were biological pathways and GO terms related to interferon signalling and induction of interferon genes (**Fig. 2d**, **Supplementary Table 6**).

Our *cis*-eQTL analysis highlighted hundreds of thousands *cis*-eVariants that were associated with thousands of *cis*-eGenes (**Table 1**), and our power of detection was dependent on sample size, which has been previously reported[53]. We also showed that there are multiple independent *cis*-eQTLs acting on thousands of genes (**Table 1**, **Fig. 3b**). Although PB is cellularly heterogeneous, we obtained good replication of *cis*-eQTLs in an external cohort from the Cattle GTEx Consortium[53] using AC, Storey's π_1 statistic, and Spearman correlation of effect size estimates (**Fig. 3e**, **Fig. 3f**). While nearly all expressed genes appear to have a *cis*-eQTL in a relevant context/tissue[85], we demonstrated that PB DEGs, which differentiate bTB+ and bTB− cattle have genomic variants associated with transcript abundance and that perturbation of these genes significantly impacts host immunobiology, most notably functions associated with defence response to virus and HIV-1 viral life cycle (**Fig. 4d**). Peripheral blood DEGs have recently been characterised as reflecting disease-induced expression perturbations rather than mechanistic disease causing changes[86]; however, the DE *cis*-eGenes identified in our study should be prioritised for further downstream functional analysis and the eVariants associated with these genes may be incorporated as prior information in future genome-enabled breeding programmes for bTB disease susceptibility traits[87,88].

Cis-eQTLs explain a small proportion of expression heritability whereas *trans*-eQTLs have been estimated to contribute up to 70% of the interindividual variance in gene expression[57,89] and tag important genomic regulatory elements and transcriptional regulators (e.g., TFs/coTFs), which will be important for bridging the genome to phenome gap in livestock species[90]. We mapped *trans*-eVariants located more than 5 Mb from the associated gene and observed an inflated number of *trans*-eGenes despite a limited sample size (n = 123). In humans, with a sample size of approximately 120 subjects, we would expect to detect less than five *trans*-eGenes[74]. Conversely, in the present study using 123 bovine PB transcriptomes, we detected 107 *trans*-eGenes that were associated with 5,314 *trans*-eVariants. An inflated number of *trans*-eGenes was also previously reported by the Cattle GTEx Consortium[53] and we posit that many of these *trans*-eVariants are false positives owing to a complex intra- and interchromosomal LD relationship existing between *trans*-eVariants and top *cis*-eQTLs of the same gene. We empirically showed via permutation analysis that the LD relationship between intra- and interchromosomal *trans*-eVariants and top *cis*-eQTLs of the same gene was significantly higher than that expected by chance (**Fig. 5b**, **Fig. 5c**, **Supplementary Fig. 11c**). This LD pattern is likely a consequence of intense human-mediated selection for production traits (e.g., milk yield)[91] and the relatively small and decreasing effective population size (N_e) of European *B.*

taurus breeds[92], particularly the Holstein-Friesian breed[93]. Diminishing N_e accentuates the contribution of random genetic drift to allelic frequency changes, which can lead to random associations between loci on different chromosomes that arise at a rate inversely proportional to the N_e[94]. Another evolutionary force, admixture, particularly admixture between genetically isolated populations can contribute to LD arising between unlinked sites, however such LD is likely to degrade rapidly[95-97].

Our TWAS analyses revealed a total of 136 genes associated with susceptibility to bTB disease in cattle, many of which were breed- and expression model cohort-specific, an observation that reflects the polygenicity of this phenotypic trait[33] (**Supplementary Fig. 12**). However, several of the genes we identified have documented roles in the host response to mycobacterial infection and the immunopathology of TB disease. For example, in our AAG-CH TWAS cohort, we identified the guanylate binding protein 4 gene (*GBP4*; $P = 2.5 \times 10^{-6}$; $Z = -4.7$) as being significantly associated with bTB disease susceptibility. A negative *Z*-score can be interpreted as decreased expression of this gene being associated with the trait of interest[47]. *GBP4* is an interferon-inducible gene that is upregulated and contributes to the Type 1 immune response during *M. tuberculosis* infection[98]. Additionally, expression of *GBP4* was shown to be significantly upregulated at +1 week, + 2 weeks and +10 weeks in blood samples from cattle experimentally infected with *M. bovis* and stimulated with PPD-b compared to control non-stimulated blood samples[70] and in PB leukocytes of *M. bovis*-infected cattle compared to non-infected control animals[64]. The most significant gene associated with bTB disease susceptibility in our bTB− group was the regulator of G-protein signalling 10 gene (*RGS10*; $P = 7.34 \times 10^{-18}$; $Z = 8.6$), which was identified in the Holstein-Friesian GWAS panel. (**Fig. 6b**). *RGS10* encodes an important anti-inflammatory protein and has previously been implicated *in vitro* in regulating macrophage activity, specifically limiting activation of the NF-κB pathway, reducing expression of tumour necrosis factor (TNF), and regulating macrophage M1 to M2 repolarisation[99]. In murine models challenged with a lethal dose of influenza A virus, loss of *RGS10* resulted in increased cytokine and chemokine activity, and a more pronounced recruitment of neutrophils and monocytes to the site of infection[100].

Members of the MTBC, including *M. tuberculosis* and *M. bovis*, have evolved a diverse range of strategies to modulate, subvert, and evade the host innate immune response and an important facet of this is manipulation of granuloma formation and function[101]. Recent multi-modal profiling of the granuloma in cynomolgus macaques (*Macaca fascicularis)* experimentally infected with *M. tuberculosis* has shown that high-burden granulomas are characterised by Type 2 immunity and tissue-protective responses that maintain essential tissue functionality while paradoxically creating a niche for bacterial persistence, whereas low *M. tuberculosis* burden granulomas are governed by an

adaptive Type 1–Type 17 (Th1-Th17) and cytotoxic T cell responses that kills and destroys invading bacilli[102]. In this regard, we also identified the RELT TNF receptor gene (*RELT*; $P = 4.4 \times 10^{-9}$; $Z = 5.9$) in our AAG-HF TWAS cohort as being significantly associated with bTB disease susceptibility. *RELT* is a member of the TNF superfamily and evidence suggests that RELT may promote an immunosuppressive environment through suppression of IFN-γ, TNF, and IL-5 production in $CD4^+$ and $CD8^+$ T cells[103]. The triggering receptor expressed on myeloid like cells 2 gene (*TREML2*) was also significantly associated with bTB susceptibility in our bTB−/HF TWAS cohort ($P = 1.1 \times 10^{-8}$; $Z = 5.7$). In monocytes infected with *M. tuberculosis*, overexpression of *TREML2* was shown to promote *IL6* transcription through activation of STAT3 and to supress the Th1 response[104]. Expression of IL-6 induced by *M. tuberculosis* infection was also shown to inhibit the macrophage response to IFN-γ[105] and impaired intracellular killing of mycobacteria[106].

Taken together, based on our TWAS results, we can therefore hypothesise that the combined increased and decreased expression of several immunoregulatory genes dampens the proinflammatory immune response during *M. bovis* infection, supresses the Th1 T-cell response and contributes to macrophage M2 polarisation and Type 2 immunity characteristics that lead to bTB disease susceptibility, bacterial persistence, pathogenesis, and disease.

Although the TWAS approach is being increasingly applied to complex traits in plant and animal species, the statistical framework underpinning the methodology has come under criticism due to an inflated type 1 error rate as a consequence of failing to account for predictive expression model uncertainty in the gene-trait association test[51]. Stringent gene filtering through application of a two-step statistical significance process with a Bonferroni $P_{adj.} < 0.05$ threshold followed by a post-hoc permutation test ($P < 0.05$) appears to control for this inflated false positive rate[51]. However, the permutation scheme itself is highly conservative and as such, true causal genes associated with the trait of interest may be filtered out owing to insufficient power[47]. Immunologically relevant genes that did not achieve a $P_{perm.} < 0.05$, but that may be associated with bTB disease susceptibility, included the polymeric immunoglobulin receptor gene (*PIGR*) in the AAG-MB TWAS cohort ($P = 2.83 \times 10^{-7}$; $Z = -5.1$). *PIGR* encodes an important transmembrane receptor involved in the transport of dimeric immunoglobulin A (IgA) from the lamina propria across the epithelial/mucosal barrier enabling the production of secretory immunoglobulins that mediate innate host protection through specific and non-specific pathogen interactions[107]. Previous work has shown that $PIGR^{-/-}$ mice are more susceptible to *M. tuberculosis* infection and have reduced IFN-γ and TNF expression and a delayed induction of mycobacteria-induced immune responses[108]. PTEN induced kinase 1 gene (*PINK1*) was also identified as being associated with bTB susceptibility in the bTB−/HF TWAS cohort ($P = 7.0 \times 10^{-8}$; $Z = 5.4$). In bovine monocyte-derived macrophages (MDM) challenged with

654 *M. bovis*, expression of *PINK1* was shown to benefit the pathogen in the host cell through induction
655 of mitophagy, promoting its intracellular survival by inhibiting xenophagy[109].

656 While the TWAS reference panels (bTB+, bTB−, and AAG) used in this study are primarily
657 composed of crossbred cattle, with Holstein representing the bulk of animal ancestry (**Fig. 2a**, **Fig.**
658 **2b)**, three of the GWAS data sets that were used to impute the expression models were generated
659 from single breed population samples (Charolais, Holstein-Friesian, and Limousin). The genetic
660 heterogeneity across the bTB+/bTB−/AAG reference panels and GWAS cohorts, therefore, makes it
661 challenging to impute reference expression models. A reference panel that better matches the GWAS
662 cohort would result in more power to detect genes associated with *M. bovis* infection
663 susceptibility/resistance traits. This issue may account for the detection of more significant TWAS
664 genes in the Holstein-Friesian GWAS cohort versus the other breed cohorts across the bTB− and
665 AAG reference panels as Holstein was the predominant ancestry (**Table 3**). This observation aligns
666 with results from a previous integrative genomics study. Compared to the Charolais and Limousin
667 GWAS data sets, substantially more significant bTB susceptibility associated SNPs were detected in
668 the Holstein-Friesian GWAS data set following integration of functional genomics outputs from
669 Holstein-Friesian bovine AMs challenged *in vitro* with *M. bovis*[83].

670 The animals in the bTB+ reference panel have a confirmed bTB diagnosis and are maintained
671 for bTB diagnostics potency testing; therefore, it is possible that some of the results from the TWAS
672 may be confounded by horizontal pleiotropy owing to the same causal variant having independent
673 effects on both expression and the trait[47]. We would therefore prioritise significant TWAS
674 associations identified in the analysis of the bTB− group for further downstream analyses. Lastly, it
675 is challenging to evaluate causality from our TWAS results due to issues associated with sharing of
676 GWAS variants between expression models, coregulation of a putatively causal and non-causal
677 gene(s), and correlation of predicted expression models between tested genes[48]. A combination of a
678 larger tissue/cell specific reference panel in conjunction with other integrative and functional genomic
679 techniques such as colocalization[110] and Mendelian randomisation[111] would facilitate this approach.

680

Methods

Animal recruitment, sampling, and data acquisition.

A total of n = 60 cattle infected with *M. bovis* and n = 63 non-infected control animals were recruited for the purpose of this study. All animals were male and were born between 2014 and 2019. The *M. bovis*-infected cattle (bTB+) were selected from a panel of naturally infected animals maintained for on-going tuberculosis surveillance at the Department of Agriculture, Food, and the Marine (DAFM) Backweston Laboratory Campus farm (Celbridge, Co. Kildare, Ireland). These animals were skin tested by an experienced veterinary practitioner and had positive single intradermal comparative tuberculin test (SICTT) results where the skin-fold thickness response to purified protein derivative (PPD)-bovine (PPDb) exceeded that of PPD-avian (PPDa) by at least 12 mm. As an ancillary diagnostic test carried out in series, all animals were tested for *M. bovis* infection using the whole blood IFN-γ release assay (IGRA) (BoviGAM® – Prionics AG, Switzerland)[71]. During *post-mortem* examination, all the animals disclosed multiple lesions consistent with bovine tuberculosis. The non-infected control animals (bTB−) were selected from bTB-free cattle herds (all SICTT negative) and with no recent history of *M. bovis* infection.

Peripheral blood was sampled from each animal using blood collection tubes with blood harvested from the tail vein. All tubes were inverted 4-6 times immediately after sampling and transported to the laboratory in a refrigerated cool box within three hours of collection. Whole blood was collected in Tempus™ blood RNA stabilisation tubes (Thermo Fisher Scientific) for isolation of total RNA. All tubes intended for RNA extraction were stored at −80°C prior to isolation and purification. A single heparin-coated tube was also collected from each animal (bTB+ and bTB−) for same-day IGRA testing to confirm infection status at the UCD Tuberculosis Diagnostics and Immunology Research Centre. Further details on genomic and transcriptomic data acquisition are detailed in **Supplementary Note 1**.

Basic population genomics analysis and imputation

All animals in the study were genotyped using the Affymetrix Axiom™ Genome-Wide BOS-1 Array (Thermo Fisher Scientific) with SNP positions originally mapped to the UMD3.1 bovine reference genome assembly[112]. The CEL files were imported into Axiom Analysis Suite software tool v.5.1.1.1 following the Axiom Best Practices Genotyping Analysis Workflow with the required sample attributes[113]. The SNPolisher Recommended Probesets were exported in PLINK format annotated with the Axiom_GW_Bos_SNP_1.na35.annot.db annotation database using genome version UMD 3.1 and NCBI version 6. This analysis yielded a total of 591,947 SNPs (91.23%) for downstream analyses. Prior to remapping of SNPs to the current ARS-UCD1.2 genome build[114], the

genetic structure and diversity of the study population was evaluated as follows. PLINK v1.90b6.25[115] was used to filter out SNPs with a minor allele frequency (MAF) < 10%; that deviated from Hardy-Weinberg equilibrium (HWE; $P < 1 \times 10^{-6}$); and with a call rate < 0.95. The PLINK --*indep-pairwise* command was then used to prune variants in linkage disequilibrium (LD) with the following parameters: window size = 1000 kb; step size = 5 variants; and $r^2 > 0.2$. Following these steps there were 34,272 SNPs available for examination of genetic structure using ADMIXTURE v.1.3[116] and principal component analysis (PCA) using PLINK.

The ADMIXTURE analysis was performed with the --*cv* option such that setting the number of ancestral populations to $K = 2$ produced the lowest cross-validation error. We then used pophelper v.2.3.1[117] to generate a structure plot. For the PCA analysis, we used the --*pca* function in PLINK and used a custom R v.4.3.2[118] script to plot the PCA results using ggplot2 v.3.4.1[119].

For the imputation up to whole-genome sequence (WGS) scale data, raw genotyped variants were first remapped from UMD3.1 to the ARS-UCD1.2 bovine genome assembly (**Supplementary Note 2**). Following this, a global cattle reference panel from Dutta, et al. [72] was used as the imputation reference panel, which comprised a total of 10,282,187 SNPs derived from $n = 287$ distinct animals spanning a diverse range of breeds and geographic locations (55 populations: 13 European, 12 African, 28 Asian, and two Middle Eastern). Imputation was performed using Minimac4 v.1.03[120] with default parameters to impute the target genotype data set up to WGS, which resulted in a master imputed data set consisting of all $n = 123$ animals with genotypes for 10,282,037 SNPs (**Supplementary Note 2**).

Transcriptomics data quality control, read alignment, and read mapping

The paired-end RNA-seq FASTQ files ($n = 123$; 60 bTB+ and 63 bTB−) was assessed using FastQC v.0.11.5[121], which showed that the RNA-seq data set was of sufficiently high quality to negate the requirement for hard or soft trimming. Following this, RNA-seq reads were aligned to the ARS-UCD1.2 bovine reference genome using STAR v.2.7.1a[122]. Read counts for each gene were then quantified using featureCounts v.2.0.6[123] and the ARS-UCD1.2 ensemble annotation file (https://ftp.ensembl.org/pub/release-110/gtf/bos_taurus/Bos_taurus.ARS-UCD1.2.110.gtf.gz) excluding chimeric fragments, aligning reads in a reversely stranded manner, and considering only fragments with both ends successfully aligned for quantification.

Missing data imputation and sample mismatch assessment

Control sample C028 did not have any date of birth information available (**Supplementary Table 1**). Therefore, we inferred the age of C028 as the mean of all other animals that were sampled on the same date (02/05/2017). Control samples C039 and C041 were assigned the same animal

identification number; therefore, to ensure that these animals were not duplicates, we estimated the identity-by-state (IBS) distance values between all samples by using the pruned SNPs prior to imputation to identify and remove duplicate animals using PLINK. The IBS distance values were calculated as:

$$IBS\ distance = (IBS2 + 0.5 \times IBS1)/(IBS0 + IBS1 + IBS2)$$

where *IBS0* is the number of IBS 0 non-missing variants, *IBS1* is the number of IBS 1 non-missing variants and *IBS2* is the number of IBS 2 non-missing variants. Sample pairs with IBS distance values > 0.85 were considered duplicates and only one sample was retained for subsequent analyses[53].

To ensure that the transcriptomics data and genome-wide SNP data for all 123 animals (bTB− and bTB+) were matched, we assessed the genotype consistency using the match BAM to VCF (MBV) function[124] that is part of the QTLtools (v 1.3.1) package[125]. Briefly, MBV reports the proportion of heterozygous and homozygous genotypes (for each sample in a VCF file) for which both alleles are captured by the sequencing reads in all BAM files. Correct sample matches can then be verified, as they should have a high proportion of concordant heterozygous and homozygous sites between the genotype data and the mapped sequencing reads.

Differential expression analysis

A differential expression analysis (DEA) was conducted between the control (bTB−) and reactor (bTB+) animal groups using DESeq2 v.1.40.2[126] and a design matrix, which included the following covariates: age in months, RNA-seq sequencing batch, and genetic structure in the form of PC1 and PC2 from the PCA of the pruned SNP data set prior to imputation with reactor status as the variable of interest. The PC1 and PC2 covariates were included because the crossbred/multibreed nature of the animals in our study population should be incorporated in the DEA contrast for the bTB− and bTB+ animal groups. Genes with raw expression counts ≥ 6 in at least 20% of samples were retained prior to the DEA. For the DEA, the null hypothesis was that the logarithmic fold change (LFC) between the control and the reactor group, for the expression of a particular gene is exactly 0. To account for potential heteroscedasticity of LFCs, we implemented the approximate posterior estimation for generalised linear model coefficients (APEGLM) method[127] using the *lfcShrink* function. Genes with a Benjamini-Hochberg (BH) false discovery rate (FDR) adjusted *P*-value[128] ($P_{adj.}$) < 0.05 and a LFC > 0 or < 0 were considered significantly differently expressed (DE).

Cis-eQTL mapping

For the mapping of *cis*-eQTLs, we used the human GTEx Consortium[74] pipeline with some minor modifications. We conducted the *cis*-eQTL analysis on the control group (bTB−), the reactor group (bTB+), and a combined group of all 123 animals (AAG). Raw RNA-seq read counts were

normalised using the trimmed mean of the M values (TMM) method[129] and the expression values for each gene were then inverse normally transformed across samples to ensure the molecular phenotypes followed a normal distribution. Genes with raw expression counts ≥ 6 and a transcript per million (TPM)[130,131] normalised expression count ≥ 0.1 in at least 20% of samples were retained for the eQTL analysis. For each group, we used the PCAForQTL R package v.0.1.0[132] to identify hidden confounders in the normalised and filtered expression matrices. The number of latent variables selected was determined using the elbow method via the *runElbow* function in PCAForQTL. We then merged these inferred covariates with known covariates (the top five genotype PCs of the imputed data set, age in months, sequencing batch, and infection status, where applicable) and removed highly correlated known covariates captured well by the inferred covariates (unadjusted $R^2 \geq 0.9$) using the PCAForQTL *filterKnownCovariates* function.

For the *cis*-eQTL mapping procedure, we used TensorQTL v.1.0.8[133]. We defined the *cis* window as +/− 1 Mb from the transcriptional start site (TSS) of a gene. To identify significant *cis*-eQTLs, we invoked the permutation strategy in TensorQTL[134] to estimate variant-phenotype associated empirical *P*-values with the parameter --mode *cis* to account for multiple variants being tested per molecular phenotype. We then used the Storey and Tibshirani FDR procedure[135] to correct the *beta* distribution-extrapolated empirical *P*-values to account for multiple phenotypes being tested genome-wide. A gene with at least one significant associated *cis*-eQTL was considered a *cis*-eGene.

To identify significant *cis*-eVariants associated with detected *cis*-eGenes, we followed the procedure implemented by the Pig GTEx Consortium[136]. Briefly, we first obtained nominal *P*-values of association for each variant-gene pair using the parameter --mode *cis_nominal*. We then defined the empirical *P*-value of a gene which was closest to an FDR of 0.05 as the genome wide empirical *P*-value threshold (*pt*). Next, we calculated the gene-level threshold for each gene from the beta distribution by using the *qbeta(pt*, *beta_shape1*, *beta_shape2*) command in R with beta_shape1 and beta_shape2 being derived from TensorQTL. Variants with a nominal *P*-value of association below the gene-level threshold were included in the final list of variant-gene pairs and were considered as significant *cis*-eVariants.

Following the Pig GTEx Consortium[136], to identify genes with multiple independent-acting *cis*-eQTLs, we performed a conditional stepwise regression analysis using the parameter --mode *cis_independent*. Briefly, the most significant variant was considered a putative *cis*-eQTL if it had a nominal *P*-value below the genome-wide FDR threshold inferred above. Next, using a forward stepwise regression procedure, the genotypes of this variant were residualized out from the phenotype quantifications and the process of regression, selection, and residualization was repeated until no more variants were below the *P*-value threshold resulting in *n* independent signals per gene. Finally,

using backward stepwise regression, nearby significant variants were assigned to inferred independent signals.

Trans-eVariant mapping and permutation analysis

We conducted *trans*-eQTL mapping on all three groups of animals using QTLtools v.1.3.1[125] using the --*trans* option. We first tested all variant phenotype pairs using the --*nominal* and --*normal* parameters together, including the same covariates described for the *cis*-eQTL mapping procedure and reported those with a nominal association below a threshold of $P < 1 \times 10^{-5}$ and that were not proximal (< 5 Mb) to the tested phenotype. We then characterised the null distribution of associations by employing the --*permute* option and used the QTLtools *runFDR_ftrans.R* script to estimate the FDR. Briefly, the nominal and permuted *P*-values are ranked in descending order and the FDR for a particular variant-phenotype pair is calculated by counting the number of permutation hits with smaller *P*-values than the nominal *P*-value for a variant-phenotype pair and, finally, dividing this number by the rank of the pair. Variants with an FDR < 0.05 were considered significant *trans*-eVariants. Given the small number of *trans*-eGenes identified in the control (bTB−) and reactor (bTB+) cohorts, we decided to focus on the larger cohort (bTB− and bTB+) for analysis of *trans*-eVariants.

We hypothesized that top intra and interchromosomal *trans*-eQTLs were in high LD with top *cis*-eQTLs of the same gene. To test this hypothesis, we performed a permutation analysis where we randomly sampled 10,000 sets of null intrachromosomal variant pairs and interchromsoomal *trans*-eVariants respectively. For the intrachromosomal set, we computed the LD (*r*) between each set and compared the distribution of the means and medians to our observed distribution. For the interchromsomal set, we computed the LD (*r*) between null intrachromsomal *trans*-eVariants and top *cis*-eQTLs of the same gene and compared the means and medians of the 10,000 sets to our observed distribution. For both the inter and intrachromosomal LD analyses, we calculated a permuted *P*-value ($P_{\text{perm.}}$) defined as the number of sets with a mean or median LD (*r*) value respectively greater than or equal to our observed LD values divided by 10,000. A more detailed description of this analysis is outlined in **Supplementary Note 3**.

Lastly, we hypothesised that the remaining top *trans*-eVariants were proximal to expressed transcription factors (TFs) or transcription factor co-factors (co-TFs). To empirically test this, we first removed *trans*-eVariant gene pairs if the top *trans*-eVariant was in LD ($r^2 > 0.01$) with the top *cis*-eVariant associated to the same gene and only considered remaining *trans*-eVariants that were highly significant (FDR < 0.01). We then downloaded genomic coordinates for annotated TFs/co-TFs from the AnimalTFDB: v.4.0 database[75]. We calculated the proportion of the filtered top *trans*-eVariants that were proximal to at least one expressed TF/co-TF at genomic intervals ranging from ±10 kb to

±1 Mb versus 10,000 random sets of SNPs to generate a null distribution. For each distance window, we obtained a $P_{perm.}$ value defined as the number of sets with a proportion of null *trans*-eVariants proximal to at least one expressed TF/co-TF equal to or greater than the observed proportion divided by 10,000.

Replication of *cis*-eQTLs

To assess the replicability of *cis*-eQTLs identified for each group in an independent cohort, we first downloaded blood *cis*-eQTL summary statistics (both permuted and nominal associations) from the Cattle GTEx Consortium (https://cgtex.roslin.ed.ac.uk/wp-content/plugins/cgtex/static/rawdata/Full_summary_statisitcs_cis_eQTLs_FarmGTEx_cattle_V0.tar.gz). We used three different measurements of agreement of eQTL effects when comparing eQTLs across the two studies: allele concordance (AC), π_1 and Spearman correlation (ρ). AC provides an indication of the proportion of effects that have a consistent direction of effect (slope) within the set of eQTLs that is significant in both the discovery (here, denoted as the control bTB−, reactor bTB+, and combined (AAG; bTB− and bTB+ cohorts) and the replication cohort (the Cattle GTEx) and is expected to be 50% for random eQTL effects[137]. The parameter π_1[135] represents the proportion of true positive eQTL *P*-values in the replication cohort and is calculated as $1 - \pi_0$ (the proportion of true null eQTL *P*-values). The Spearman ρ statistic estimates the correlation between the effect sizes (slope) of significant eQTLs in the discovery cohort and matched associations in the replication cohort, regardless of significance in the latter.

To calculate AC, we matched significant eQTLs in the discovery cohort to significant eQTLs in the replication cohort. We then calculated the proportion of these eQTLs that showed the same direction of effect. To calculate π_1, we obtained the *P*-values in the replication cohort of significant associations identified in the discovery cohort and used the *qvalue* function in R to estimate π_0. We then calculated π_1 as $1 - \pi_0$. Uncertainty estimates of π_1 were obtained using 100 bootstraps where SNPs were sampled with replacement and π_1 was recomputed each time[138]. To obtain the Spearman ρ statistics, we calculated the Spearman correlation between significant eQTLs identified in this study to matched variant:gene pairs in the replication cohort, regardless of significance.

GWAS data pre-processing

GWAS summary statistics for the present study were obtained from a single and multi-breed GWAS experiment that leveraged WGS data from Run 6 of the 1000 Bull Genomes Project[139] as an imputation reference panel. The GWAS used estimated breeding values (EBVs) derived from an *M. bovis* infection phenotype as the trait of interest for n = 2,039 Charolais, n = 1,964 Limousin, and n = 1,502 Holstein-Friesian cattle[33]. Variants were remapped from UMD 3.1 to ARS-UCD1.2 using a

custom R script that was developed for a previous study that integrated the GWAS summary statistics with functional genomics data obtained from *M. bovis*-infected bovine alveolar macrophages (bAM)[83]. To check for instances of strand flips, the reference and alternative allele pairs derived from Run 6 of the 1000 Bull Genomes Project were compared to reference and alternative allele pairs in the ARS-UCD1.2 reference genome (https://sites.ualberta.ca/~stothard/1000_bull_genomes/ARS1.2PlusY_BQSR.vcf.gz). If a strand flip occurred, the beta values for each SNP were also inverted. A Wald-statistic Z score for each GWAS SNP was calculated by dividing the effect size (β) of a SNP with the standard error of the effect size.

Transcriptome-wide association study (TWAS) analysis

Imputed genotype data for the three groups were converted to binary (.bed) format using PLINK with the *--keep-allele-order* parameter. The resulting files were then loaded into R using the bigsnpr v.1.10.8 and bigstatsr v.1.5.6 R packages[140]. Predictive models of expression for each gene were generated using the Mediator-enriched TWAS (MeTWAS) function within the MOSTWAS package v.0.1.0[58]. Briefly, MeTWAS first identifies an association between a mediating biomarker (e.g., a TF) and a gene of interest. It then builds a predictive model of expression for the mediating biomarker considering SNPs local to the biomarker. The predicted expression pattern of the biomarker (determined via five-fold cross-validation) is then included as a fixed effect with the effect sizes of putative mediators on the expression levels of the gene of interest estimated by ordinary least squares regression. Lastly, for the final predictive model of the gene of interest, the *cis*-eVariants are fitted as random effects using either elastic net regression or linear mixed modelling, whichever produces the highest five-fold McNemar's cross-validated adjusted R^2 value.

The mediating biomarkers used in MeTWAS included expressed regulatory proteins (TFs and co-TFs) curated from the AnimalTFDB database[75]. We first computed associations between mediating biomarkers and genes through correlation analysis with significant associations (BH-FDR < 0.01) being retained. We then retained mediating biomarker:gene associations in instances where the mediating biomarker was considered a *cis*-eGene. Genes that had significant non-zero heritabilities (nominal $P < 0.05$) for their expression levels, as computed by the likelihood ratio test (LRT) from the genome-wide complex trait analysis (GCTA) software tool v.1.94.1[141] and for which MOSTWAS-derived predictive models achieved a five-fold McNemar's cross-validated adjusted R^2 value ≥ 0.01 were retained for the gene–trait association test. The maximum number of mediating biomarkers to include in the expression model for a gene was set to ten.

Within the MOSTWAS framework, expression models were imputed into the GWAS summary statistics using the ImpG-Summary algorithm[47,142] and a weighted burden Z-test was employed in the gene–trait association test[47,142]. Genes with a Bonferroni-adjusted P-value < 0.05 were considered

915 candidate genes associated with bTB susceptibility. To assess whether the same distribution of
916 GWAS SNP effect sizes could yield a significant association by chance, we implemented a
917 permutation scheme on significant (Bonferroni-adjusted *P*-value < 0.05) TWAS genes where we
918 sampled, without replacement, the SNP effect sizes 1000 times and recomputed the weighted burden
919 test statistic to generate a permuted null distribution[47]. Genes with a permuted *P*-value < 0.05 were
920 considered significantly associated with bTB disease status.

921 **Gene set overrepresentation and functional enrichment analyses**

922 Gene set overrepresentation and functional enrichment analyses was conducted using a
923 combination of the g:GOSt tool within g:Profiler v.0.2.2[143] and Ingenuity® Pathway Analysis – IPA®
924 (Summer 2023 release; Qiagen). For IPA®, the target species selected included *Homo sapiens*, *Mus*
925 *musculus*, and *Rattus rattus* with all cell types selected in addition to the Experimentally Observed
926 and High Predicted confidence settings. We followed best practice recommendations to account for
927 tissue-specific sampling biases in gene set overrepresentation and functional enrichment analyses[144].
928 Consequently, for analysis of differentially expressed genes (DEGs), the background set consisted of
929 all expressed genes that were tested for differential expression. For analyses of genes between the
930 eQTL and DEGs, the background set consisted of the intersection between the genes tested in both
931 analyses. For g:Profiler, the organism selected was *B. taurus* and an ordered query list (based on the
932 adjusted *P*-value from the differential expression analysis) was inputted. For analyses of our query
933 gene sets, we selected the gene ontology biological processes (GO:BP) and the cellular component
934 (GO:CC)[145] databases in addition to the Kyoto encyclopaedia of genes and genomes (KEGG)[146] and
935 Reactome[147] repositories. To identify significantly enriched/overrepresented pathways, a BH-FDR
936 multiple testing correction was applied ($P_{adj.} < 0.05$).

937 **Computational infrastructure and reproducibility**

938 All data-intensive computational procedures were performed on a 36-core/72-thread compute
939 server (2× Intel® Xeon® CPU E5–2697 v4 processors, 2.30 GHz with 18 cores each), with 512 GB of
940 RAM, 96 TB SAS storage (12 × 8 TB at 7200 rpm), 480 GB SSD storage, and with Ubuntu Linux OS
941 (version 18.04 LTS).

942

References

1 Paulson, T. Epidemiology: A mortal foe. *Nature* **502**, S2-3, doi:10.1038/502S2a (2013).

2 World Health Organization. *Global Tuberculosis Report 2023*. (World Health Organization, 2023).

3 Brosch, R. *et al.* Comparative genomics of the mycobacteria. *Int. J. Med. Microbiol.* **290**, 143-152, doi:10.1016/S1438-4221(00)80083-1 (2000).

4 Garnier, T. *et al.* The complete genome sequence of *Mycobacterium bovis*. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 7877-7882, doi:10.1073/pnas.1130426100 (2003).

5 Gutierrez, M. C. *et al.* Ancient origin and gene mosaicism of the progenitor of *Mycobacterium tuberculosis*. *PLoS Pathog.* **1**, e5, doi:10.1371/journal.ppat.0010005 (2005).

6 Gagneux, S. Ecology and evolution of *Mycobacterium tuberculosis*. *Nat. Rev. Microbiol.* **16**, 202-213, doi:10.1038/nrmicro.2018.8 (2018).

7 Steele, J. in *Mycobacterium bovis Infection in Animals and Humans* (ed C. O. & Steele Thoen, J.H.) 169-172 (Iowa State University Press, 1995).

8 Waters, W. R., Palmer, M. V., Buddle, B. M. & Vordermeier, H. M. Bovine tuberculosis vaccine research: historical perspectives and recent advances. *Vaccine* **30**, 2611-2622, doi:10.1016/j.vaccine.2012.02.018 (2012).

9 Olea-Popelka, F. *et al.* Zoonotic tuberculosis in human beings caused by *Mycobacterium bovis*–a call for action. *Lancet Infect. Dis.* **17**, e21-e25, doi:10.1016/S1473-3099(16)30139-6 (2017).

10 Kock, R. *et al.* Zoonotic tuberculosis - the changing landscape. *Int. J. Infect. Dis.* **113 Suppl 1**, S68-S72, doi:10.1016/j.ijid.2021.02.091 (2021).

11 World Health Organization. *Global Tuberculosis Report 2020*. (World Health Organization, 2020).

12 Waters, W. R., Maggioli, M. F., McGill, J. L., Lyashchenko, K. P. & Palmer, M. V. Relevance of bovine tuberculosis research to the understanding of human disease: historical perspectives, approaches, and immunologic mechanisms. *Vet. Immunol. Immunopathol.* **159**, 113-132, doi:10.1016/j.vetimm.2014.02.009 (2014).

13 Waters, W. R. & Palmer, M. V. *Mycobacterium bovis* infection of cattle and white-tailed deer: translational research of relevance to human tuberculosis. *ILAR J.* **56**, 26-43, doi:10.1093/ilar/ilv001 (2015).

14 Waters, W. R. *et al.* Tuberculosis immunity: opportunities from studies with cattle. *Clin. Dev. Immunol.* **2011**, 768542, doi:10.1155/2011/768542 (2011).

15 Van Rhijn, I., Godfroid, J., Michel, A. & Rutten, V. Bovine tuberculosis as a model for human tuberculosis: advantages over small animal models. *Microbes Infect.* **10**, 711-715, doi:10.1016/j.micinf.2008.04.005 (2008).

16 Kaufmann, S. H. E. & Dorhoi, A. Molecular determinants in phagocyte-bacteria interactions. *Immunity* **44**, 476-491, doi:10.1016/j.immuni.2016.02.014 (2016).

17 Weiss, G. & Schaible, U. E. Macrophage defense mechanisms against intracellular bacteria. *Immunol. Rev.* **264**, 182-203, doi:10.1111/imr.12266 (2015).

18 Ehrt, S. & Schnappinger, D. Mycobacterial survival strategies in the phagosome: defence against host stresses. *Cell. Microbiol.* **11**, 1170-1178, doi:https://doi.org/10.1111/j.1462-5822.2009.01335.x (2009).

19 Chai, Q., Wang, L., Liu, C. H. & Ge, B. New insights into the evasion of host innate immunity by Mycobacterium tuberculosis. *Cell. Mol. Immunol.* **17**, 901-913, doi:10.1038/s41423-020-0502-z (2020).

20 Chandra, P., Grigsby, S. J. & Philips, J. A. Immune evasion and provocation by Mycobacterium tuberculosis. *Nat. Rev. Microbiol.* **20**, 750-766, doi:10.1038/s41579-022-00763-4 (2022).

21 Driscoll, E. E., Hoffman, J. I., Green, L. E., Medley, G. F. & Amos, W. A preliminary study of genetic factors that influence susceptibility to bovine tuberculosis in the British cattle herd. *PLoS ONE* **6**, e18806, doi:10.1371/journal.pone.0018806 (2011).

22 Allen, A. R. *et al.* Bovine tuberculosis: the genetic basis of host susceptibility. *Proc. Biol. Sci.* **277**, 2737-2745, doi:10.1098/rspb.2010.0830 (2010).

23 Kadarmideen, H. N., Ali, A. A., Thomson, P. C., Muller, B. & Zinsstag, J. Polymorphisms of the *SLC11A1* gene and resistance to bovine tuberculosis in African zebu cattle. *Anim. Genet.* **42**, 656-658, doi:10.1111/j.1365-2052.2011.02203.x (2011).

24 Amos, W. *et al.* Genetic predisposition to pass the standard SICCT test for bovine tuberculosis in British cattle. *PLoS ONE* **8**, e58245, doi:10.1371/journal.pone.0058245 (2013).

25 Bermingham, M. L. *et al.* Genetics of tuberculosis in Irish Holstein-Friesian dairy herds. *J. Dairy Sci.* **92**, 3447-3456, doi:10.3168/jds.2008-1848 (2009).

26 Brotherstone, S. *et al.* Evidence of genetic resistance of cattle to infection with *Mycobacterium bovis*. *J. Dairy Sci.* **93**, 1234-1242, doi:10.3168/jds.2009-2609 (2010).

27 Finlay, E. K., Berry, D. P., Wickham, B., Gormley, E. P. & Bradley, D. G. A genome wide association scan of bovine tuberculosis susceptibility in Holstein-Friesian dairy cattle. *PLoS ONE* **7**, e30545, doi:10.1371/journal.pone.0030545 (2012).

28 Bermingham, M. L. *et al.* Genome-wide association study identifies novel loci associated with resistance to bovine tuberculosis. *Heredity (Edinb)* **112**, 543-551, doi:10.1038/hdy.2013.137 (2014).

29 Richardson, I. W. *et al.* A genome-wide association study for genetic susceptibility to *Mycobacterium bovis* infection in dairy cattle identifies a susceptibility QTL on chromosome 23. *Genet. Sel. Evol.* **48**, 19, doi:10.1186/s12711-016-0197-x (2016).

30 Raphaka, K. *et al.* Genomic regions underlying susceptibility to bovine tuberculosis in Holstein-Friesian cattle. *BMC Genet.* **18**, 27, doi:10.1186/s12863-017-0493-7 (2017).

31 Wilkinson, S. *et al.* Fine-mapping host genetic variation underlying outcomes to *Mycobacterium bovis* infection in dairy cows. *BMC Genomics* **18**, 477, doi:10.1186/s12864-017-3836-x (2017).

32 Tsairidou, S. *et al.* An analysis of effects of heterozygosity in dairy cattle for bovine tuberculosis resistance. *Anim. Genet.* **49**, 103-109, doi:10.1111/age.12637 (2018).

33 Ring, S. C. *et al.* Variance components for bovine tuberculosis infection and multi-breed genome-wide association analysis using imputed whole genome sequence data. *PLoS ONE* **14**, e0212067, doi:10.1371/journal.pone.0212067 (2019).

34 Clark, E. L. *et al.* From FAANG to fork: application of highly annotated genomes to improve farmed animal production. *Genome Biol.* **21**, 285, doi:10.1186/s13059-020-02197-8 (2020).

35 Wittkopp, P. J. Genomic sources of regulatory variation in cis and in trans. *Cell. Mol. Life Sci.* **62**, 1779-1783, doi:10.1007/s00018-005-5064-9 (2005).

36 Gilad, Y., Rifkin, S. A. & Pritchard, J. K. Revealing the architecture of gene regulation: the promise of eQTL studies. *Trends Genet.* **24**, 408-415, doi:10.1016/j.tig.2008.06.001 (2008).

37 Nica, A. C. & Dermitzakis, E. T. Expression quantitative trait loci: present and future. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **368**, 20120362, doi:10.1098/rstb.2012.0362 (2013).

38 Hill, M. S., Vande Zande, P. & Wittkopp, P. J. Molecular and evolutionary processes generating variation in gene expression. *Nat. Rev. Genet.* **22**, 203-215, doi:10.1038/s41576-020-00304-w (2021).

39 Aguet, F. *et al.* Molecular quantitative trait loci. *Nat. Rev. Methods Primers* **3**, 4, doi:10.1038/s43586-022-00188-6 (2023).

40 Nicolae, D. L. *et al.* Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. *PLoS Genet.* **6**, e1000888, doi:10.1371/journal.pgen.1000888 (2010).

41 Nica, A. C. *et al.* Candidate causal regulatory effects by integration of expression QTLs with complex trait genetic associations. *PLoS Genet.* **6**, e1000895, doi:10.1371/journal.pgen.1000895 (2010).

42 Fehrmann, R. S. *et al.* Trans-eQTLs reveal that independent genetic variants associated with a complex phenotype converge on intermediate genes, with a major role for the HLA. *PLoS Genet.* **7**, e1002197, doi:10.1371/journal.pgen.1002197 (2011).

43 Barreiro, L. B. *et al.* Deciphering the genetic architecture of variation in the immune response to *Mycobacterium tuberculosis* infection. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 1204-1209, doi:10.1073/pnas.1115761109 (2012).

44 Xiang, R. *et al.* Quantifying the contribution of sequence variants with regulatory and evolutionary significance to 34 bovine complex traits. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 19398-19408, doi:10.1073/pnas.1904159116 (2019).

45 Xiang, R. *et al.* Gene expression and RNA splicing explain large proportions of the heritability for complex traits in cattle. *Cell Genom* **3**, 100385, doi:10.1016/j.xgen.2023.100385 (2023).

46 Gamazon, E. R. *et al.* A gene-based association method for mapping traits using reference transcriptome data. *Nat. Genet.* **47**, 1091-1098, doi:10.1038/ng.3367 (2015).

47 Gusev, A. *et al.* Integrative approaches for large-scale transcriptome-wide association studies. *Nat. Genet.* **48**, 245-252, doi:10.1038/ng.3506 (2016).

48 Wainberg, M. *et al.* Opportunities and challenges for transcriptome-wide association studies. *Nat. Genet.* **51**, 592-599, doi:10.1038/s41588-019-0385-z (2019).

49 Li, B. & Ritchie, M. D. From GWAS to gene: transcriptome-wide association studies and other methods to functionally understand GWAS discoveries. *Front. Genet.* **12**, 713230, doi:10.3389/fgene.2021.713230 (2021).

50 Mai, J., Lu, M., Gao, Q., Zeng, J. & Xiao, J. Transcriptome-wide association studies: recent advances in methods, applications and available databases. *Commun. Biol.* **6**, 899, doi:10.1038/s42003-023-05279-y (2023).

51 de Leeuw, C., Werme, J., Savage, J. E., Peyrot, W. J. & Posthuma, D. On the interpretation of transcriptome-wide association studies. *PLoS Genet.* **19**, e1010921, doi:10.1371/journal.pgen.1010921 (2023).

52 Lu, M. *et al.* TWAS Atlas: a curated knowledgebase of transcriptome-wide association studies. *Nucleic Acids Res.* **51**, D1179-D1187, doi:10.1093/nar/gkac821 (2023).

53 Liu, S. *et al.* A multi-tissue atlas of regulatory variants in cattle. *Nat. Genet.* **54**, 1438-1447, doi:10.1038/s41588-022-01153-5 (2022).

54 Cai, W. *et al.* The eQTL colocalization and transcriptome-wide association study identify potentially causal genes responsible for economic traits in Simmental beef cattle. *J Anim Sci Biotechnol* **14**, 78, doi:10.1186/s40104-023-00876-7 (2023).

55 Li, Z. *et al.* METRO: Multi-ancestry transcriptome-wide association studies for powerful gene-trait association detection. *Am. J. Hum. Genet.* **109**, 783-801, doi:10.1016/j.ajhg.2022.03.003 (2022).

56 Chen, F. *et al.* Multi-ancestry transcriptome-wide association analyses yield insights into tobacco use biology and drug repurposing. *Nat. Genet.* **55**, 291-300, doi:10.1038/s41588-022-01282-x (2023).

57 Boyle, E. A., Li, Y. I. & Pritchard, J. K. An expanded view of complex traits: from polygenic to omnigenic. *Cell* **169**, 1177-1186, doi:10.1016/j.cell.2017.05.038 (2017).

58 Bhattacharya, A., Li, Y. & Love, M. I. MOSTWAS: Multi-Omic Strategies for Transcriptome-Wide Association Studies. *PLoS Genet.* **17**, e1009398, doi:10.1371/journal.pgen.1009398 (2021).

59 Rhodes, S. G., Buddle, B. M., Hewinson, R. G. & Vordermeier, H. M. Bovine tuberculosis: immune responses in the peripheral blood and at the site of active disease. *Immunology* **99**, 195-202, doi:10.1046/j.1365-2567.2000.00944.x (2000).

60 Killick, K. E. *et al.* Genome-wide transcriptional profiling of peripheral blood leukocytes from cattle infected with *Mycobacterium bovis* reveals suppression of host immune genes. *BMC Genomics* **12**, 611, doi:10.1186/1471-2164-12-611 (2011).

61 Blanco, F. C., Soria, M., Bianco, M. V. & Bigi, F. Transcriptional response of peripheral blood mononuclear cells from cattle infected with *Mycobacterium bovis*. *PLoS ONE* **7**, e41066, doi:10.1371/journal.pone.0041066 (2012).

62 Aranday-Cortes, E. *et al.* Transcriptional profiling of disease-induced host responses in bovine tuberculosis and the identification of potential diagnostic biomarkers. *PLoS ONE* **7**, e30626, doi:10.1371/journal.pone.0030626 (2012).

63 Bhuju, S. *et al.* Global gene transcriptome analysis in vaccinated cattle revealed a dominant role of IL-22 for protection against bovine tuberculosis. *PLoS Pathog.* **8**, e1003077, doi:10.1371/journal.ppat.1003077 (2012).

64 McLoughlin, K. E. *et al.* RNA-seq transcriptional profiling of peripheral blood leukocytes from cattle infected with *Mycobacterium bovis*. *Front. Immunol.* **5**, 396, doi:10.3389/fimmu.2014.00396 (2014).

65 Klepp, L. I. *et al.* Identification of bovine tuberculosis biomarkers to detect tuberculin skin test and IFNgamma release assay false negative cattle. *Res. Vet. Sci.* **122**, 7-14, doi:10.1016/j.rvsc.2018.10.016 (2019).

66 Fang, L. *et al.* Potential diagnostic value of the peripheral blood mononuclear cell transcriptome from cattle with bovine tuberculosis. *Front. Vet. Sci.* **7**, 295, doi:10.3389/fvets.2020.00295 (2020).

67 Wiarda, J. E. *et al.* Severity of bovine tuberculosis is associated with innate immune-biased transcriptional signatures of whole blood in early weeks after experimental *Mycobacterium bovis* infection. *PLoS ONE* **15**, e0239938, doi:10.1371/journal.pone.0239938 (2020).

68 McLoughlin, K. E. *et al.* RNA-seq transcriptome analysis of peripheral blood from cattle infected with *Mycobacterium bovis* across an experimental time course. *Front. Vet. Sci.* **8**, 662002, doi:10.3389/fvets.2021.662002 (2021).

69 Abdelaal, H. F. M., Thacker, T. C., Wadie, B., Palmer, M. V. & Talaat, A. M. Transcriptional profiling of early and late phases of bovine tuberculosis. *Infect. Immun.* **90**, e0031321, doi:10.1128/iai.00313-21 (2022).

70 Correia, C. N. *et al.* High-resolution transcriptomics of bovine purified protein derivative-stimulated peripheral blood from cattle infected with *Mycobacterium bovis* across an experimental time course. *Tuberculosis (Edinb)* **136**, 102235, doi:10.1016/j.tube.2022.102235 (2022).

71 Clegg, T. A., Doyle, M., Ryan, E., More, S. J. & Gormley, E. Characteristics of *Mycobacterium bovis* infected herds tested with the interferon-gamma assay. *Prev. Vet. Med.* **168**, 52-59, doi:10.1016/j.prevetmed.2019.04.004 (2019).

72 Dutta, P. *et al.* Whole genome analysis of water buffalo and global cattle breeds highlights convergent signatures of domestication. *Nat. Commun.* **11**, 4739, doi:10.1038/s41467-020-18550-1 (2020).

73 Vosa, U. *et al.* Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. *Nat. Genet.* **53**, 1300-1310, doi:10.1038/s41588-021-00913-z (2021).

74 GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318-1330, doi:10.1126/science.aaz1776 (2020).

75 Shen, W. K. *et al.* AnimalTFDB 4.0: a comprehensive animal transcription factor database updated with variation and expression annotations. *Nucleic Acids Res.* **51**, D39-D45, doi:10.1093/nar/gkac907 (2023).

76 Richardson, I. W. *et al.* Variance components for susceptibility to *Mycobacterium bovis* infection in dairy and beef cattle. *Genet. Sel. Evol.* **46**, 77, doi:10.1186/s12711-014-0077-1 (2014).

77 Banos, G. *et al.* Genetic evaluation for bovine tuberculosis resistance in dairy cattle. *J. Dairy Sci.* **100**, 1272-1281, doi:10.3168/jds.2016-11897 (2017).

78 Raphaka, K. *et al.* Impact of genetic selection for increased cattle resistance to bovine tuberculosis on disease transmission dynamics. *Front. Vet. Sci.* **5**, 237, doi:10.3389/fvets.2018.00237 (2018).

79 Wu, Y. *et al.* Joint analysis of GWAS and multi-omics QTL summary statistics reveals a large fraction of GWAS signals shared with molecular phenotypes. *Cell Genom* **3**, 100344, doi:10.1016/j.xgen.2023.100344 (2023).

80 Brown, A. A. *et al.* Genetic analysis of blood molecular phenotypes reveals common properties in the regulatory networks affecting complex traits. *Nat. Commun.* **14**, 5062, doi:10.1038/s41467-023-40569-3 (2023).

81 Brown, B. C., Bray, N. L. & Pachter, L. Expression reflects population structure. *PLoS Genet.* **14**, e1007841, doi:10.1371/journal.pgen.1007841 (2018).

82 Randolph, H. E. *et al.* Genetic ancestry effects on the response to viral infection are pervasive but cell type specific. *Science* **374**, 1127-1133, doi:10.1126/science.abg0928 (2021).

83 Hall, T. J. *et al.* Integrative genomics of the mammalian alveolar macrophage response to intracellular mycobacteria. *BMC Genomics* **22**, 343, doi:10.1186/s12864-021-07643-w (2021).

84 Schneider, W. M., Chevillotte, M. D. & Rice, C. M. Interferon-stimulated genes: a complex web of host defenses. *Annu. Rev. Immunol.* **32**, 513-545, doi:10.1146/annurev-immunol-032713-120231 (2014).

85 Aguet, F. *et al.* Genetic effects on gene expression across human tissues. *Nature* **550**, 204-213, doi:10.1038/nature24277 (2017).

86 Porcu, E. *et al.* Differentially expressed genes reflect disease-induced rather than disease-causing changes in the transcriptome. *Nat. Commun.* **12**, 5647, doi:10.1038/s41467-021-25805-y (2021).

87 MacLeod, I. M. *et al.* Exploiting biological priors and sequence variants enhances QTL discovery and genomic prediction of complex traits. *BMC Genomics* **17**, 144, doi:10.1186/s12864-016-2443-6 (2016).

88 Breen, E. J. *et al.* BayesR3 enables fast MCMC blocked processing for largescale multi-trait genomic prediction and QTN mapping analysis. *Commun. Biol.* **5**, 661, doi:10.1038/s42003-022-03624-1 (2022).

89 Liu, X., Li, Y. I. & Pritchard, J. K. Trans effects on gene expression can drive omnigenic inheritance. *Cell* **177**, 1022-1034 e1026, doi:10.1016/j.cell.2019.04.014 (2019).

90 Rexroad, C. *et al.* Genome to phenome: improving animal health, production, and well-being - a new USDA blueprint for animal genome research 2018-2027. *Front. Genet.* **10**, 327, doi:10.3389/fgene.2019.00327 (2019).

91 Sargolzaei, M., Schenkel, F. S., Jansen, G. B. & Schaeffer, L. R. Extent of linkage disequilibrium in Holstein cattle in North America. *J. Dairy Sci.* **91**, 2106-2117, doi:10.3168/jds.2007-0553 (2008).

92 Gibbs, R. A. *et al.* Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. *Science* **324**, 528-532, doi:10.1126/science.1167936 (2009).

93 de Roos, A. P., Hayes, B. J., Spelman, R. J. & Goddard, M. E. Linkage disequilibrium and persistence of phase in Holstein-Friesian, Jersey and Angus cattle. *Genetics* **179**, 1503-1512, doi:10.1534/genetics.107.084301 (2008).

94 Waples, R. K., Larson, W. A. & Waples, R. S. Estimating contemporary effective population size in non-model species using linkage disequilibrium across thousands of loci. *Heredity (Edinb)* **117**, 233-240, doi:10.1038/hdy.2016.60 (2016).

95 Pritchard, J. K. & Przeworski, M. Linkage disequilibrium in humans: models and data. *Am. J. Hum. Genet.* **69**, 1-14, doi:10.1086/321275 (2001).

96 Arcos-Burgos, M. & Muenke, M. Genetics of population isolates. *Clin. Genet.* **61**, 233-247, doi:10.1034/j.1399-0004.2002.610401.x (2002).

97 Qanbari, S. On the extent of linkage disequilibrium in the genome of farm animals. *Front. Genet.* **10**, 1304, doi:10.3389/fgene.2019.01304 (2019).

98 Akter, S. *et al. Mycobacterium tuberculosis* infection drives a type I IFN signature in lung lymphocytes. *Cell Rep.* **39**, 110983, doi:10.1016/j.celrep.2022.110983 (2022).

99 Lee, J. K., Chung, J., Kannarkat, G. T. & Tansey, M. G. Critical role of regulator G-protein signaling 10 (RGS10) in modulating macrophage M1/M2 activation. *PLoS ONE* **8**, e81785, doi:10.1371/journal.pone.0081785 (2013).

100 Almutairi, F. *et al.* RGS10 reduces lethal influenza infection and associated lung inflammation in mice. *Front. Immunol.* **12**, 772288, doi:10.3389/fimmu.2021.772288 (2021).

101 Zhai, W., Wu, F., Zhang, Y., Fu, Y. & Liu, Z. The immune escape mechanisms of *Mycobacterium tuberculosis*. *Int. J. Mol. Sci.* **20**, doi:10.3390/ijms20020340 (2019).

102 Gideon, H. P. *et al.* Multimodal profiling of lung granulomas in macaques reveals cellular correlates of tuberculosis control. *Immunity* **55**, 827-846 e810, doi:10.1016/j.immuni.2022.04.004 (2022).

103 Cusick, J. K., Alcaide, J. & Shi, Y. The RELT family of proteins: an increasing awareness of their importance for cancer, the immune system, and development. *Biomedicines* **11**, doi:10.3390/biomedicines11102695 (2023).

104 Li, J. *et al.* TLT2 suppresses Th1 response by promoting IL-6 production in monocyte through JAK/STAT3 signal pathway in tuberculosis. *Front. Immunol.* **11**, 2031, doi:10.3389/fimmu.2020.02031 (2020).

105 Nagabhushanam, V. *et al.* Innate inhibition of adaptive immunity: *Mycobacterium tuberculosis*-induced IL-6 inhibits macrophage responses to IFN-gamma. *J. Immunol.* **171**, 4750-4757, doi:10.4049/jimmunol.171.9.4750 (2003).

106 Péan, C. B. *et al.* Regulation of phagocyte triglyceride by a STAT-ATG2 pathway controls mycobacterial infection. *Nat. Commun.* **8**, 14642, doi:10.1038/ncomms14642 (2017).

107 Turula, H. & Wobus, C. E. The role of the polymeric immunoglobulin receptor and secretory immunoglobulins during mucosal infection and immunity. *Viruses* **10**, doi:10.3390/v10050237 (2018).

108 Tjärnlund, A. *et al.* Polymeric IgR knockout mice are more susceptible to mycobacterial infections in the respiratory tract than wild-type mice. *Int. Immunol.* **18**, 807-816, doi:10.1093/intimm/dxl017 (2006).

109 Song, Y. *et al.* Mycobacterium bovis induces mitophagy to suppress host xenophagy for its intracellular survival. *Autophagy* **18**, 1401-1415, doi:10.1080/15548627.2021.1987671 (2022).

110 Giambartolomei, C. *et al.* Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.* **10**, e1004383, doi:10.1371/journal.pgen.1004383 (2014).

111 Davey Smith, G. & Hemani, G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum. Mol. Genet.* **23**, R89-98, doi:10.1093/hmg/ddu328 (2014).

112 Zimin, A. V. *et al.* A whole-genome assembly of the domestic cow, *Bos taurus*. *Genome Biol.* **10**, R42, doi:10.1186/gb-2009-10-4-r42 (2009).

113 Applied Biosystems. Axiom™ Genotyping Solution Data Analysis: User Guide. (2023). <https://assets.thermofisher.com/TFS-Assets/LSG/manuals/axiom_genotyping_solution_analysis_guide.pdf>.

114 Rosen, B. D. *et al.* De novo assembly of the cattle reference genome with single-molecule sequencing. *Gigascience* **9**, doi:10.1093/gigascience/giaa021 (2020).

115 Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559-575, doi:10.1086/519795 (2007).

116 Alexander, D. H. & Lange, K. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics* **12**, 246, doi:10.1186/1471-2105-12-246 (2011).

117 Francis, R. M. pophelper: an R package and web app to analyse and visualize population structure. *Mol. Ecol. Resour.* **17**, 27-32, doi:10.1111/1755-0998.12509 (2017).

118 R: A language and environment for statistical computing (R Foundation for Statistical Computing, Vienna, Austria, 2023).
119 Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. 2nd ed. edn, (Springer-Verlag, 2016).
120 Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat. Genet.* **48**, 1284-1287, doi:10.1038/ng.3656 (2016).
121 Andrews, S. *FastQC: a quality control tool for high throughput sequence data.*, <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (2010).
122 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21, doi:10.1093/bioinformatics/bts635 (2013).
123 Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-930, doi:10.1093/bioinformatics/btt656 (2014).
124 Fort, A. *et al.* MBV: a method to solve sample mislabeling and detect technical bias in large combined genotype and sequencing assay datasets. *Bioinformatics* **33**, 1895-1897, doi:10.1093/bioinformatics/btx074 (2017).
125 Delaneau, O. *et al.* A complete tool set for molecular QTL discovery and analysis. *Nat. Commun.* **8**, 15452, doi:10.1038/ncomms15452 (2017).
126 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550, doi:10.1186/s13059-014-0550-8 (2014).
127 Zhu, A., Ibrahim, J. G. & Love, M. I. Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. *Bioinformatics* **35**, 2084-2092, doi:10.1093/bioinformatics/bty895 (2019).
128 Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate - a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Method.* **57**, 289-300, doi:DOI 10.1111/j.2517-6161.1995.tb02031.x (1995).
129 Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140, doi:10.1093/bioinformatics/btp616 (2010).
130 Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323, doi:10.1186/1471-2105-12-323 (2011).
131 Wagner, G. P., Kin, K. & Lynch, V. J. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory Biosci* **131**, 281-285, doi:10.1007/s12064-012-0162-3 (2012).
132 Zhou, H. J., Li, L., Li, Y., Li, W. & Li, J. J. PCA outperforms popular hidden variable inference methods for molecular QTL mapping. *Genome Biol.* **23**, 210, doi:10.1186/s13059-022-02761-4 (2022).
133 Taylor-Weiner, A. *et al.* Scaling computational genomics to millions of individuals with GPUs. *Genome Biol.* **20**, 228, doi:10.1186/s13059-019-1836-7 (2019).
134 Ongen, H., Buil, A., Brown, A. A., Dermitzakis, E. T. & Delaneau, O. Fast and efficient QTL mapper for thousands of molecular phenotypes. *Bioinformatics* **32**, 1479-1485, doi:10.1093/bioinformatics/btv722 (2016).
135 Storey, J. D. & Tibshirani, R. Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 9440-9445, doi:10.1073/pnas.1530509100 (2003).
136 Teng, J. *et al.* A compendium of genetic regulatory effects across pig tissues. *Nat. Genet.* **56**, 112-123, doi:10.1038/s41588-023-01585-7 (2024).
137 de Klein, N. *et al.* Brain expression quantitative trait locus and network analyses reveal downstream effects and putative drivers for brain-related diseases. *Nat. Genet.* **55**, 377-388, doi:10.1038/s41588-023-01300-6 (2023).

138 Zeng, B. *et al.* Multi-ancestry eQTL meta-analysis of human brain identifies candidate causal variants for brain-related traits. *Nat. Genet.* **54**, 161-169, doi:10.1038/s41588-021-00987-9 (2022).

139 Daetwyler, H. D. *et al.* Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. *Nat. Genet.* **46**, 858-865, doi:10.1038/ng.3034 (2014).

140 Prive, F., Aschard, H., Ziyatdinov, A. & Blum, M. G. B. Efficient analysis of large-scale genome-wide data with two R packages: bigstatsr and bigsnpr. *Bioinformatics* **34**, 2781-2787, doi:10.1093/bioinformatics/bty185 (2018).

141 Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76-82, doi:10.1016/j.ajhg.2010.11.011 (2011).

142 Pasaniuc, B. *et al.* Fast and accurate imputation of summary statistics enhances evidence of functional enrichment. *Bioinformatics* **30**, 2906-2914, doi:10.1093/bioinformatics/btu416 (2014).

143 Raudvere, U. *et al.* g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res.* **47**, W191-W198, doi:10.1093/nar/gkz369 (2019).

144 Timmons, J. A., Szkop, K. J. & Gallagher, I. J. Multiple sources of bias confound functional enrichment analysis of global -omics data. *Genome Biol.* **16**, 186, doi:10.1186/s13059-015-0761-7 (2015).

145 Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25**, 25-29, doi:10.1038/75556 (2000).

146 Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27-30, doi:10.1093/nar/28.1.27 (2000).

147 Gillespie, M. *et al.* The reactome pathway knowledgebase 2022. *Nucleic Acids Res.* **50**, D687-D692, doi:10.1093/nar/gkab1028 (2022).

Declarations

Ethics approval and consent to participate

All experimental procedures involving animals were conducted under ethical approval from the University College Dublin (UCD) Animal Research Ethics Committee (AREC-19-09-MacHugh) and experimental license AE18982/P141 from the Irish Health Products Regulatory Authority (HPRA) in accordance with the Cruelty to Animals Act 1876 and in agreement with the European Union (Protection of Animals Used for Scientific Purposes) regulations 2012 (S.I. No.543 of 2012).

Consent for publication

Not applicable.

Availability of data and materials

The RNA-seq data from the 60 *M. bovis*-infected (bTB+) and 63 control (bTB−) cattle is available at the Gene Expression Omnibus (GEO) with the BioProject Accession GSE255724. The raw UMD 3.1 high density genotype data and ARS UCD1.2 imputed and filtered data is available at 10.5281/zenodo.10658453. GWAS summary statistics data were obtained from the Irish Cattle Breeding Federation (ICBF) and additional information about sequence and genotype data availability is provided by Ring et al.[33]. The computer code and scripts used in this study are available at the following GitHub link: https://github.com/jfogrady1/BovineQTL.

Funding

J.F.O.G was supported by Science Foundation Ireland (SFI) through the SFI Centre for Research Training in Genomics Data Science (grant no. 18/CRT/6214). This study was also supported by SFI Investigator Programme Awards to D.E.M. and S.V.G. (grant nos. SFI/08/IN.1/B2038 and SFI/15/IA/3154), and the University College Dublin – University of Edinburgh Strategic Partnership in One Health awarded to D.E.M., S.V.G., J.G.D.P., and E.L.C.

Competing interests

The authors declare no competing interests.

Author contributions

J.F.O.G, K.G.M, E.G, I.C.G, S.V.G, and D.E.M conceived and designed the study. D.E.M, I.C.G, S.V.G, J.G.D.P, and E.L.C acquired funding for the study. E.G, K.G.M and M.M facilitated access to animals for the study and C.N.C, G.P.M, S.L.F.O.D, J.A.W and J.A.B performed experimental work. J.F.O.G performed all bioinformatics analyses in collaboration with J.G.D.P, V.R, H.P, J.A.W, G.P.M., and T.J.H. and V.R. supplied the WGS Global Reference Panel and

provided advice for variant remapping, strand flipping, and genomic imputation. H.P contributed to biological interpretation of results and discussions around *trans*-eQTL mapping. J.F.O.G wrote the first draft of the manuscript, and prepared all figures, tables, and supplementary information with input from D.E.M. All authors read and approved the final manuscript.

Acknowledgements

The authors would like to thank Mairead Doyle for their help in interpreting the interferon-γ release assay test results for the animals used in the study. The authors would also like to express their gratitude to Bojan Stojkovic for assistance with animal ethics and Di Wang for assistance with experimental work.