BMC Genomics

Integrative genomics of the mammalian alveolar macrophage response to intracellular mycobacteria

Thomas J. Hall[1], Michael P. Mullen[2], Gillian P. McHugo[1], Kate E. Killick[1,3], Siobhán C. Ring[4], Donagh P. Berry[5], Carolina N. Correia[1], John A. Browne[1], Stephen V. Gordon[6,7] and David E. MacHugh[1,7]*

Abstract

Background: Bovine TB (bTB), caused by infection with *Mycobacterium bovis*, is a major endemic disease affecting global cattle production. The key innate immune cell that first encounters the pathogen is the alveolar macrophage, previously shown to be substantially reprogrammed during intracellular infection by the pathogen. Here we use differential expression, and correlation- and interaction-based network approaches to analyse the host response to infection with *M. bovis* at the transcriptome level to identify core infection response pathways and gene modules. These outputs were then integrated with genome-wide association study (GWAS) data sets to enhance detection of genomic variants for susceptibility/resistance to *M. bovis* infection.

Results: The host gene expression data consisted of RNA-seq data from bovine alveolar macrophages (bAM) infected with *M. bovis* at 24 and 48 h post-infection (hpi) compared to non-infected control bAM. These RNA-seq data were analysed using three distinct computational pipelines to produce six separate gene sets: 1) DE genes filtered using stringent fold-change and *P*-value thresholds (DEG-24: 378 genes, DEG-48: 390 genes); 2) genes obtained from expression correlation networks (CON-24: 460 genes, CON-48: 416 genes); and 3) genes obtained from differential expression networks (DEN-24: 339 genes, DEN-48: 495 genes). These six gene sets were integrated with three bTB breed GWAS data sets by employing a new genomics data integration tool—*gwinteR*. Using GWAS summary statistics, this methodology enabled detection of 36, 102 and 921 prioritised SNPs for Charolais, Limousin and Holstein-Friesian, respectively.

(Continued on next page)

* Correspondence: david.machugh@ucd.ie
[1]Animal Genomics Laboratory, UCD School of Agriculture and Food Science, UCD College of Health and Agricultural Sciences, University College Dublin, Belfield, Dublin D04 V1W8, Ireland
[7]UCD Conway Institute of Biomolecular and Biomedical Research, University College Dublin, Belfield, Dublin D04 V1W8, Ireland
Full list of author information is available at the end of the article

(Continued from previous page)

Conclusions: The results from the three parallel analyses showed that the three computational approaches could identify genes significantly enriched for SNPs associated with susceptibility/resistance to *M. bovis* infection. Results indicate distinct and significant overlap in SNP discovery, demonstrating that network-based integration of biologically relevant transcriptomics data can leverage substantial additional information from GWAS data sets. These analyses also demonstrated significant differences among breeds, with the Holstein-Friesian breed GWAS proving most useful for prioritising SNPS through data integration. Because the functional genomics data were generated using bAM from this population, this suggests that the genomic architecture of bTB resilience traits may be more breed-specific than previously assumed.

Keywords: Alveolar macrophage, GWAS, Integrative genomics, Mycobacterium bovis, Network, RNA-seq, Tuberculosis

Background

Bovine tuberculosis (bTB) is a chronic disease of livestock, particularly among domestic dairy and beef cattle, which has been conservatively estimated to cause more than $3 billion annual losses to global agriculture [1, 2]. The disease can also establish across a large variety of wildlife species including, for example, American bison (*Bison bison*), African buffalo (*Syncerus caffer*), the brushtail possum (*Trichosurus vulpecula*), red deer (*Cervus elaphus*), wild boar (*Sus scrofa*), and the European badger (*Meles meles*) [3–6]. The aetiological agent of bTB, *Mycobacterium bovis*, is a member of the *Mycobacterium tuberculosis* complex (MTBC) and has a genome sequence 99.95% identical to *M. tuberculosis*, the primary cause of human tuberculosis (TB) [7] and the leading cause of human deaths from a single infectious agent—approximately 1.25 million in 2018 [8]. In addition, for several low- and middle-income countries, the human TB disease burden is increased by zoonotic TB (zTB) caused by infection with *M. bovis* [9–12].

Scientific understanding of bTB and human TB has been synergistically intertwined since the nineteenth century and the foundational research work of Theobald Smith and others [5, 13, 14]. The pathogenesis of bTB disease in cattle is comparable with human TB disease and many aspects of *M. bovis* infection are also characteristic of *M. tuberculosis* infection [15–19]. Consequently, it is now widely recognised that *M. bovis* infection of cattle and bTB disease represent an important comparative system for understanding human TB caused by *M. tuberculosis* [20–24].

Inhalation of aerosolized bacteria is the main route of transmission for *M. bovis* in cattle and the primary site of infection is normally the lungs [17, 25, 26]. Here the bacilli are phagocytosed by alveolar macrophages (AM)—key effector cells of the innate immune system, which provide surveillance of pulmonary surfaces and can normally destroy or restrict inhaled intracellular bacilli [27, 28]. *M. bovis* and other facultative intracellular MTBC pathogens have evolved a complex range of mechanisms to evade, subvert, and exploit innate immune responses, thereby facilitating colonisation, persistence and replication within host macrophages [29–32]. These mechanisms include: recruitment of cell surface receptors on the host macrophage through molecular mimicry; restricting phagosome maturation and autophagy; detoxification of reactive oxygen species and reactive nitrogen intermediates (ROSs and RNIs); modulation of type I interferon (IFN) signalling; suppression of antigen presentation; rewiring and short-circuiting of macrophage signal transduction pathways; manipulation of host macrophage metabolism; egress of bacilli into the macrophage cytosol; and inhibition of apoptosis with concomitant induction of necrosis leading to immunopathology and shedding by the host to complete the pathogenic life cycle [33–39]. Hence after infection, a two-way response is triggered between the pathogen and macrophage, the outcome of which ultimately leads to establishment of infection or clearance of the pathogen. The latter outcome of clearance may, or may not, require engagement of the adaptive immune system. As the detection of *M. bovis* infection in cattle generally relies on detecting an adaptive immune response to the pathogen, the outcome of which is slaughter of positive animals ('reactors'), identifying genes that underpin efficacious innate responses promises to reveal favourable genomic variants for incorporation into breeding programmes.

Since 2005, substantial efforts have been made to better understand host-pathogen interaction for bTB using transcriptomics technologies such as gene expression microarrays and RNA sequencing (RNA-seq) at the host cellular level—specifically the bovine alveolar macrophage (bAM) and initial innate immune responses to infection by *M. bovis* [40–46]. These studies have helped to define a "pathogenic signature" [31, 47] of *M. bovis* infection in bAM, which reflects the tension between macrophage responses to contain and kill intracellular pathogens and evasion and avoidance mechanisms evolved by these mycobacteria. Using functional genomics data mining of transcriptomics data, it has also been shown that bAM responses to *M. bovis*

infection can be clearly differentiated from infection with *M. tuberculosis*, the primary cause of human TB [45]. In addition, these studies have been expanded to encompass surveys of the bAM epigenome using methylome sequencing and chromatin immunoprecipitation sequencing (ChIP-seq). This work has demonstrated that the transcriptional reprogramming of bAM caused by *M. bovis* infection is profoundly shaped by chromatin re-modelling at gene loci associated with critical components of host-pathogen interaction [46, 48].

In parallel to functional genomics studies of bTB, genome-wide association studies (GWAS) have been performed in Irish and UK cattle populations using estimated breeding values (EBVs; estimate of genetic merit of an animal derived from a statistical model) for several *M. bovis* infection resistance traits with heritabilities ranging from 0.04 to 0.37, depending on the phenotype used [49–55]. These GWAS have used medium- and high-density single-nucleotide polymorphism (SNP) arrays and, more recently, imputed whole-genome sequence (WGS) data sets for a large multi-breed GWAS on 7346 bulls, which identified 64 quantitative trait loci (QTLs) associated with resistance to *M. bovis* infection [55]; the association study was based on phenotypic data from 781,270 individual animals.

We have recently shown that integration of bAM functional genomics data sets—RNA-seq, microRNA-seq and ChIP-seq—with a GWAS data set for resistance to *M. bovis* infection can be used to enhance detection of genomic regions associated with reduced incidence of bTB disease [46]. For the present study, we substantially expand this work by leveraging gene-focused network- and pathway-based methods under a statistical framework based on a new software tool, *gwinteR*, to integrate transcriptomics data from *M. bovis*-challenged bAM with WGS-based GWAS results for resistance to *M. bovis* infection [55]. Functional genomics data and downstream data mining (e.g., to generate lists of differentially expressed genes and outputs from network and pathway analyses) can be used to obtain prioritised subsets of genes that are likely to be important for a specific biological process or phenomenon [56, 57]. The *gwinteR* tool can leverage these prioritised gene subsets and combine them with summary statistics from biologically relevant GWAS data sets. Biologically meaningful SNP-phenotype associations can therefore be identified and enriched that would otherwise be filtered out because of stringent multiple test correction for the very large numbers of observations in a typical GWAS. The primary aim of this work was to evaluate whether this approach can systematically enhance detection of genomic sequence variants and genes underpinning bTB disease resistance in cattle populations.

Results

Differential gene expression and pathway analyses of *M. bovis*-infected bovine AM

Quality filtering of RNA-seq read pairs yielded a mean of $22,681,828 \pm 3,508,710$ reads per individual library ($n = 78$ libraries). A mean of $19,582,959 \pm 3,021,333$ read pairs (86.17%) were uniquely mapped to locations in the ARS-UCD1.2 bovine genome assembly. Detailed filtering and mapping statistics are provided in a data file available from the Dryad Digital Repository (https://doi.org/10.5061/dryad.83bk3j9q6) and multivariate PCA analysis of the individual animal sample expression data using DESeq2 revealed separation of the control and *M. bovis*-infected bAM groups at the 24 and 48 h post infection (hpi) time points, but not at the 2 and 6 hpi time points (S1 Fig).

Using default criteria for differential expression (FDR $P_{\text{adj.}} < 0.05$; $|\log_2 FC| > 1$), and considering the *M. bovis*-infected bAM relative to the control non-infected macrophages, three DE genes were detected at 2 hpi (all three exhibited increased expression in the *M. bovis*-infected group); 97 DE genes were detected at 6 hpi (40 increased and 57 decreased); 1345 were detected at 24 hpi (764 increased and 581 decreased); and 2915 were detected at 48 hpi (1528 increased and 1387 decreased) (Fig. 2a; Additional file 1). Figure 2b shows that 2982 genes were differentially expressed across the 24 and 48 hpi time points. Table 1 shows a breakdown of DE genes across the infection time course for a range of statistical thresholds and fold-change cut-offs, including the default criteria (FDR $P_{\text{adj.}} < 0.05$; $|\log_2 FC| > 1$). To ensure manageable computational loads, the DE gene sets that were used for GWAS integration with *gwinteR* were filtered with $|\log_2 FC| > 2$, and $P_{\text{adj.}} < 0.01$ and $P_{\text{adj.}} < 0.000001$ for 24 and 48 hpi, respectively. With these criteria, there were 378 input genes for GWAS integration identified at 24 hpi and 390 input genes at 48 hpi. (24 hpi and 48 hpi DEG gene sets). In addition, 210 genes overlapped between the two time points. The two DEG *gwinteR* input gene sets (DEG-24 and DEG-48 – see Fig. 1) are also detailed in Additional file 1.

To produce gene sets for the IPA Core Analysis within the recommended range for the number of input entities [58, 59] and to include DE genes with small fold-change values, gene sets were filtered using only $P_{\text{adj.}}$ thresholds of 0.05 and 0.01 at 24 hpi and 48 hpi, respectively. This resulted in 1957 input genes (1071 upregulated and 886 downregulated) from a background detectable set of 16,084 at 24 hpi and 2492 input genes (1401 upregulated and 1091 downregulated) from a background detectable set of 17,492 genes at 48 hpi. The IPA analysis was focused on the 24 and 48 hpi time points because a relatively small numbers of DE genes were detected at 2 and 6 hpi (Fig. 2 and Table 1).
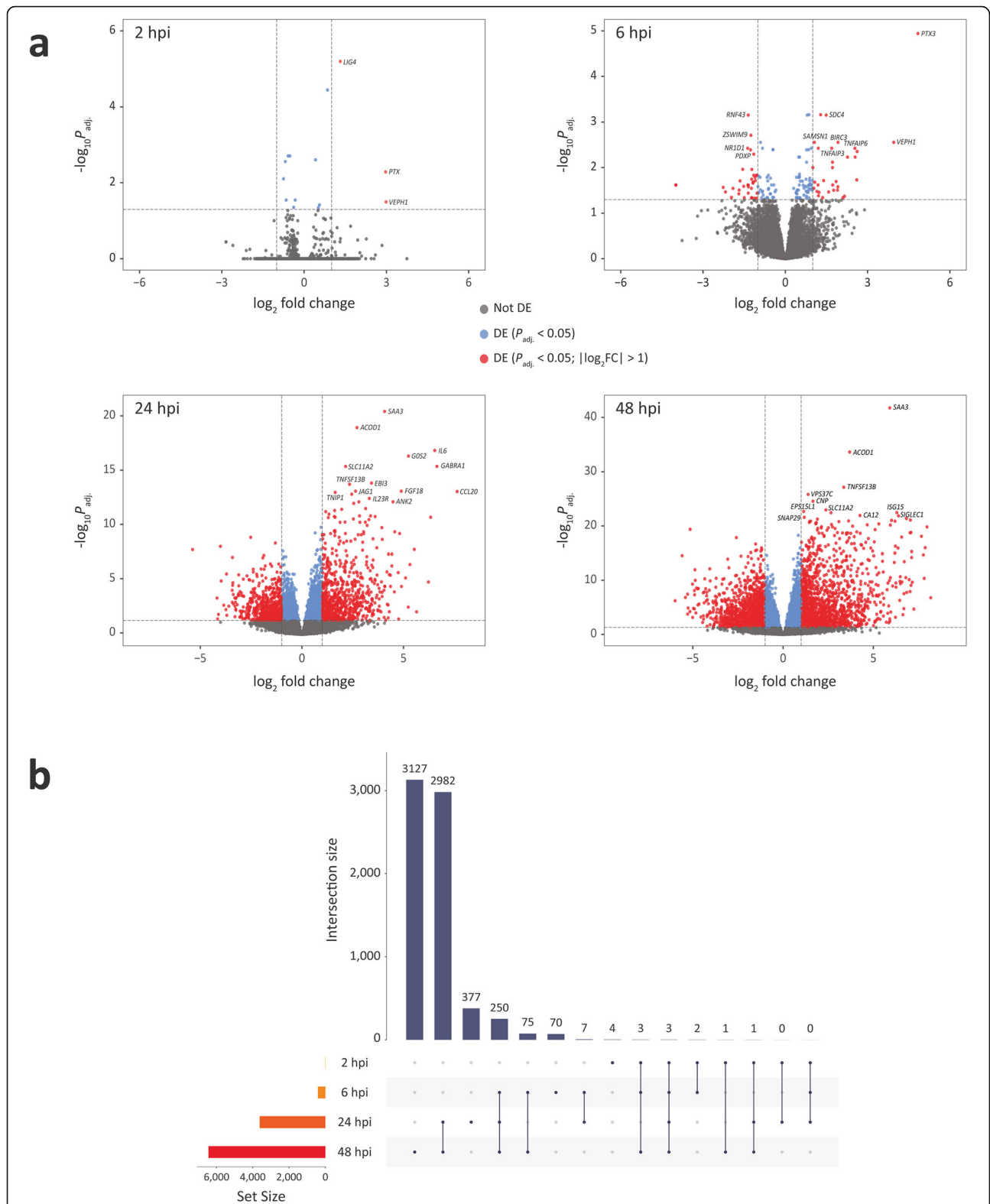
Table 1 Differentially expressed genes detected in *M. bovis*-infected bovine AM relevant to control non-infected bAM

Post-infection time point	$P_{adj.} < 0.05$; $|log_2FC| > 0$ (increased/decreased)	$P_{adj.} < 0.05$; $|log_2FC| > 1$ (increased/decreased)	$P_{adj.} < 0.01$; $|log_2FC| > 0$ (increased/decreased)	$P_{adj.} < 0.01$; $|log_2FC| > 1$ (increased/decreased)
2 hpi	14 (7/7)	3 (3/0)	8 (4/4)	2 (2/0)
6 hpi	410 (203/207)	97 (40/57)	119 (63/56)	32 (14/18)
24 hpi	3620 (1898/1722)	1345 (764/581)	2059 (1168/891)	933 (577/356)
48 hpi	6442 (3295/3147)	2915 (1528/1387)	4737 (2516/2221)	2386 (1294/1092)

Using the B-H method for multiple test correction in IPA ($P_{adj.} < 0.05$), there were 68 and 48 statistically significant enriched IPA canonical pathways at 24 hpi and 48 hpi, respectively (Additional file 2). Enriched pathways at 24 hpi included *Role of Pattern Recognition Receptors in Recognition of Bacteria and Viruses, IL-6 Signalling, TNFR2 Signalling, Role of RIG1-like Receptors in Antiviral Innate Immunity, Role of Cytokines in Mediating Communication between Immune Cells, Communication between Innate and Adaptive Immune Cells, IL-*

12 Signalling and Production in Macrophages, IL-10 Signalling, Protein Ubiquitination Pathway, Toll-like Receptor Signalling, NF-κB Signalling, PI3K/AKT Signalling, and *TNFR1 Signalling.* The most highly activated pathway at 24 hpi was *PI3K/AKT Signalling.* Enriched pathways at 48 hpi included *Protein Ubiquitination Pathway, Role of Cytokines in Mediating Communication between Immune Cells, IL-12 Signalling and Production in Macrophages, Role of RIG1-like Receptors in Antiviral Innate Immunity, Role of Pattern Recognition Receptors in*
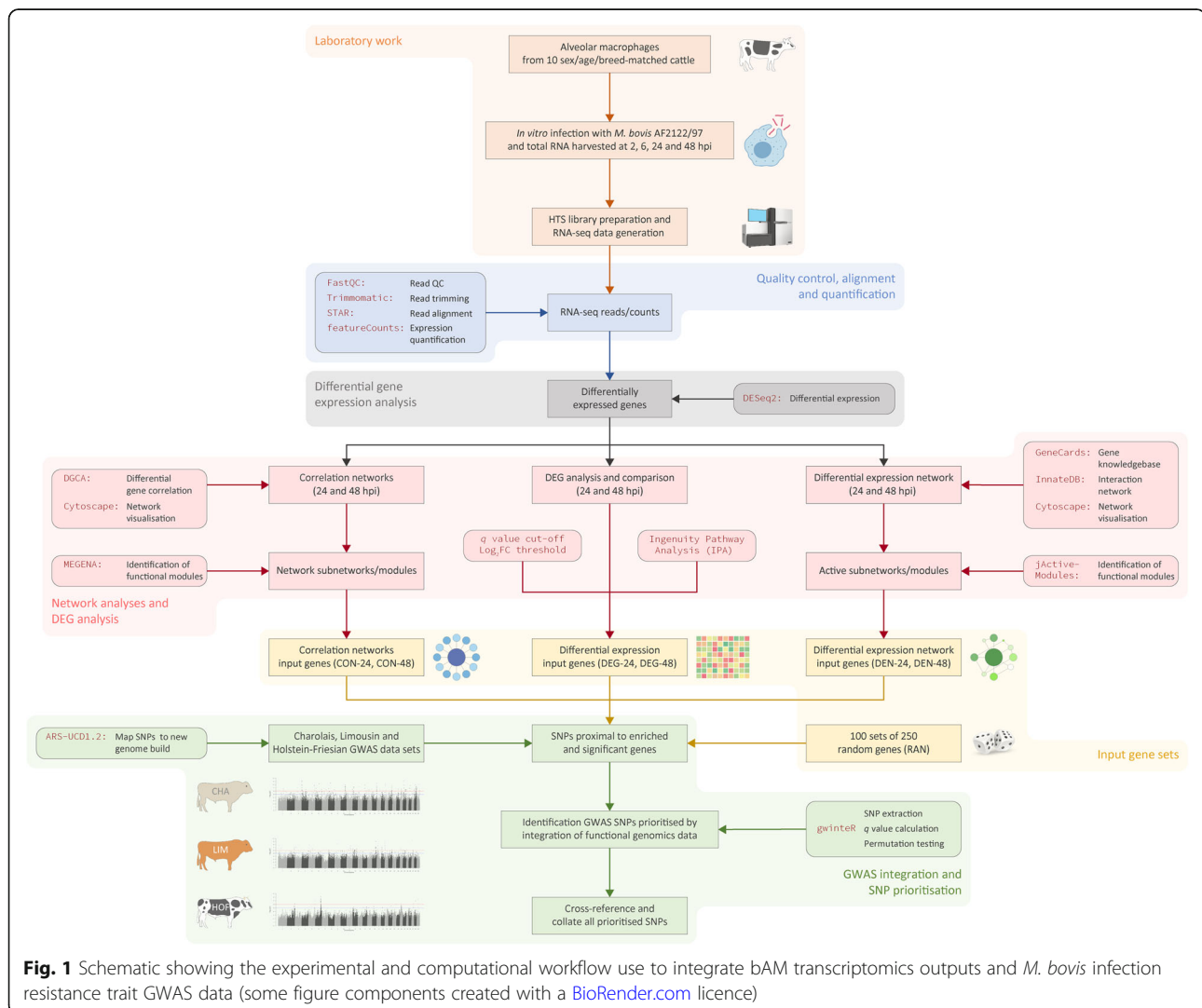
Fig. 1 Schematic showing the experimental and computational workflow use to integrate bAM transcriptomics outputs and *M. bovis* infection resistance trait GWAS data (some figure components created with a BioRender.com licence)

Fig. 2 Differentially expressed genes in *M. bovis*-infected bAM at 2, 6, 24, and 48 hpi. **a** Volcano plots of differentially expressed genes with FDR $P_{adj.}$ value thresholds of 0.05 and absolute $\log_2$ fold-change > 1. **b** UpSet plot showing the intersection of shared differentially expressed genes across the four post-infection time points

Recognition of Bacteria and Viruses, Communication between Innate and Adaptive Immune Cells, TNFR2 Signalling, Role of PI3K/AKT Signalling in the Pathogenesis of Influenza, IL-10 Signalling, and *Toll-like Receptor Signalling*. The SIGORA software tool [60] has been previously used to identify biological pathways associated with a robust 'core' bAM response to infection with both *M. bovis* and *M. tuberculosis* [45]. It is therefore reassuring that many of these pathways—including *PI3K-Akt Signalling Pathway, RIG-I-like Receptor Signalling Pathway, Toll-like Receptor Signalling* and *Protein Ubiquitination Pathway*—were also enriched using the IPA methodology at 24 and 48 hpi.

Differential co-expression correlation networks and identification of functional gene modules

For the generation of bAM differential co-expression correlation networks, filtering of genes with low measure of central tendency, which reduces the number of potential spurious correlations [61], resulted in 11,354 and 11,170 genes at 24 and 48 hpi, respectively. Following this step, differential correlation analysis using DGCA with an empirical $P_{adj.}$ value threshold of 0.10 resulted in 3507 differentially correlated gene pairs out of 128,913,316 total pairwise correlations at 24 hpi; and 1135 from a total of 124,768,900 at 48 hpi (Additional file 3). The correlation networks generated at 24 hpi and 48 hpi (Fig. 3a) yielded a total of 22 and 14 functional gene modules, respectively (Fig. 3b and c, and Additional file 3). After removal of duplicates, consolidated totals of 460 genes and 416 genes were contained in the functional modules at 24 hpi and 48 hpi, respectively. There were also 26 genes that overlapped between the functional modules for the two time points. The two correlation network (CON) *gwinteR* input gene sets (CON-24 and CON-48 – see Fig. 1) are also detailed in Additional file 3.

GO term enrichment was also performed for each functional module at 24 hpi and 48 hpi, with the top three GO terms retained for each functional module (S2 Fig and S3 Fig). The top five GO terms at 24 hpi (ranked by $P_{adj.}$) were *translation* (GO:0006412), *peptide biosynthetic process* (GO:0043043), *amide biosynthetic process* (GO:0043604), *structural constituent of ribosome* (GO:0003735), and *cellular amide metabolic process* (GO:0043603) (S2 Fig). The top five enriched GO terms at 48 hpi (ranked by $P_{adj.}$) were *signalling receptor activity* (GO:0038023) and *molecular transducer activity* (GO:0060089), *transforming growth factor beta activation* (GO:0036363), *chemokine activity* (GO:0008009), and *signalling receptor binding* (GO:0005102) (S3 Fig).

Differential expression network analysis and identification of activated modular subnetworks

To provide a computationally manageable number of genes for an InnateDB input data set [62], a GeneCards *Relevance Score* (GCRS) threshold was used (GCRS > 2.5). This GCRS cut-off produced an input list of 258 functionally prioritised genes for generation of an InnateDB gene interaction network (GIN) and the top ten genes from this list ranked by GCRS were: interferon gamma receptor 1 (*IFNGR1*), interleukin 12 receptor subunit beta 1 (*IL12RB1*), toll like receptor 2 (*TLR2*), solute carrier family 11 member 1 (*SLC11A1*), signal transducer and activator of transcription 1 (*STAT1*), interleukin 12B (*IL12B*), cytochrome b-245 beta chain (*CYBB*), tumour necrosis factor (*TNF*), interferon gamma receptor 2 (*IFNGR2*), and interferon gamma (*IFNG*).

The large GIN produced by InnateDB starting with the input list of 258 functionally prioritised genes was visualised using Cytoscape and consisted of 7001 nodes (individual genes) and 19,713 edges (gene interactions) (Fig. 4a). Additional file 4 provides information for all gene interactions represented in Fig. 4a. Following visualisation of the large GIN in Cytoscape, the jActivesModules Cytoscape plugin was used to detect statistically significant differentially activated subnetworks (modules) at the 24 hpi and 48 hpi time points. The top five subnetworks at each time point were retained for downstream analyses and consisted of 198 genes in module 1 at 24 hpi (M1−24), 287 genes in M2−24, 272 genes in M3−24, 53 genes in M4−24, 171 genes in M5−24, 381 genes in M1−48, 330 genes in M2−48, 403 genes in M3−48, 371 genes in M4−48, and 399 genes in M5−48 (Additional file 4). As an example, Fig. 4b shows the subnetwork of genes and gene interactions representing module 5 at 24 hpi (M5−24).

The genes contained in the top five modules at 24 hpi and 48 hpi were filtered to remove duplicates and consolidated into two separate gene sets for GWAS integration with *gwinteR*. The consolidated gene sets for the top five modules at 24 hpi and 48 hpi contained 339 and 495 unique genes, respectively. There were 245 genes that overlapped between the two subnetwork gene sets for the two post-infection time points. The two differential expression network (DEN) *gwinteR* input gene sets (DEN-24 and DEN-48 – see Fig. 1) are also detailed in Additional file 4.

GWAS integration and identification of additional SNP–trait associations

The six gene sets generated from the three separate analyses of DE genes in bAM challenged with *M. bovis* at 24 hpi and 48 hpi are summarised in Table 2 and further detailed in Additional files 1, 2 and 3. Also, S4 Fig and S5 Fig show Venn diagrams with the overlaps for the DEG, CON and DEN input gene sets at 24 hpi and

Hall *et al. BMC Genomics*   (2021) 22:343

Page 7 of 20
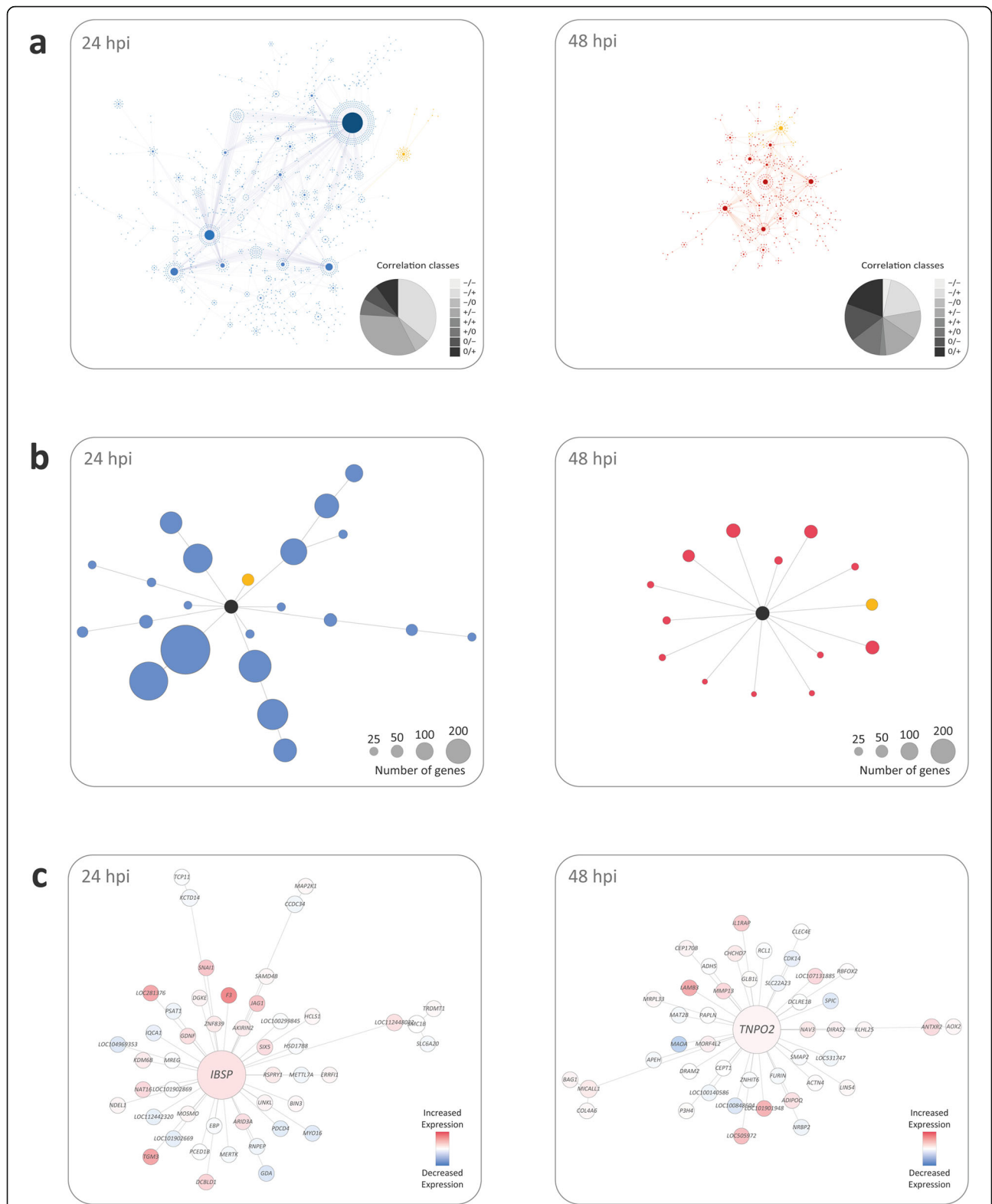
Fig. 3 Differential co-expression correlation networks and submodules at 24 and 48 hpi. **a** The complete correlation networks for *M. bovis*-infected bAM at 24 and 48 hpi. **b** The 22 subnetwork modules detected at 24 hpi and the 14 subnetwork modules detected at 48 hpi with individual example modules highlighted in yellow. **c** Individual example subnetwork modules at 24 and 48 hpi

Hall *et al. BMC Genomics*   (2021) 22:343

Page 8 of 20
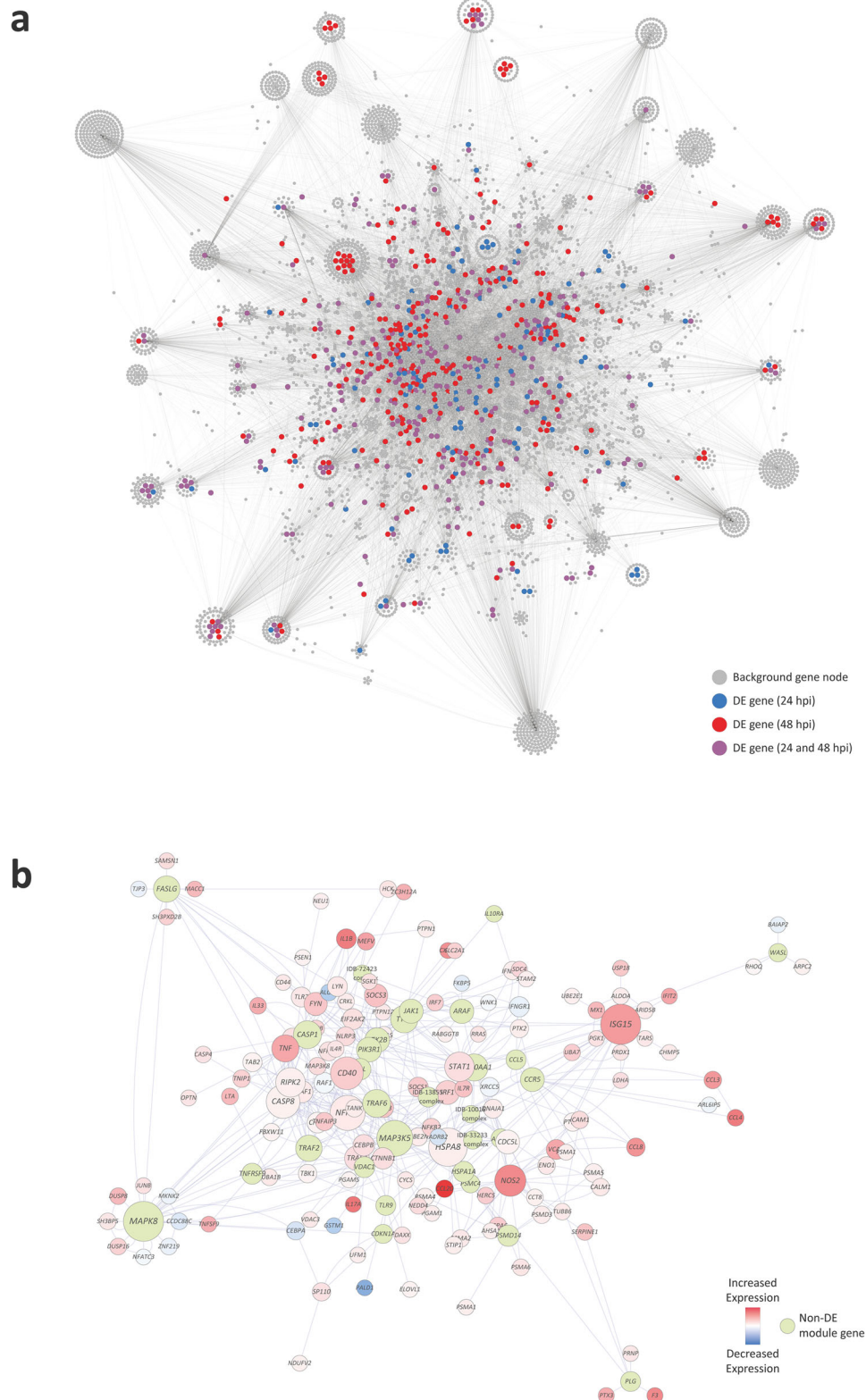
Fig. 4 A large gene interaction network (GIN) with superimposed differentially expressed genes. **a** Cytoscape tuberculosis gene interaction network with superimposed differentially expressed genes at 24 and 48 hpi. **b** Example subnetwork showing module number 5 detected with the 24 hpi gene expression data

Table 2 Six different input gene sets used for GWAS integration

Post-infection time point	Filtered differentially expressed genes (DEG)		Correlation networks analysis (CON)		Differential expression network analysis (DEN)	
	Input gene set code	No. of genes	Input gene set code	No. of genes	Input gene set code	No. of genes
24 hpi	DEG-24	378	CON-24	460	DEN-24	338
48 hpi	DEG-48	390	CON-48	416	DEN-48	493

48 hpi, respectively. In addition to these six putative functionally relevant gene sets, one hundred sets of 250 genes randomly sampled from the bovine genome were used for statistical context and comparison. These random gene sets (RAN) are detailed in Additional file 5 (see also Fig. 1). The results from the integrative analyses using *gwinteR* with the DEG-24, DEG-48, CON-24, CON-48, DEN-24, DEN-48, and RAN gene sets are summarised graphically in Fig. 5 and detailed in Additional files 6 and 7. We also used DE gene sets at 2 and 6 hpi with the default statistical threshold and fold-change cut-offs (FDR $P_{adj.} < 0.05$; $|\log_2FC| > 1$) for GWAS integration using gwinteR (i.e., DEG-2 and DEG-6). However, no significant SNP enrichment were observed for these input gene sets, most likely because of the low numbers of input genes (3 and 97, respectively).
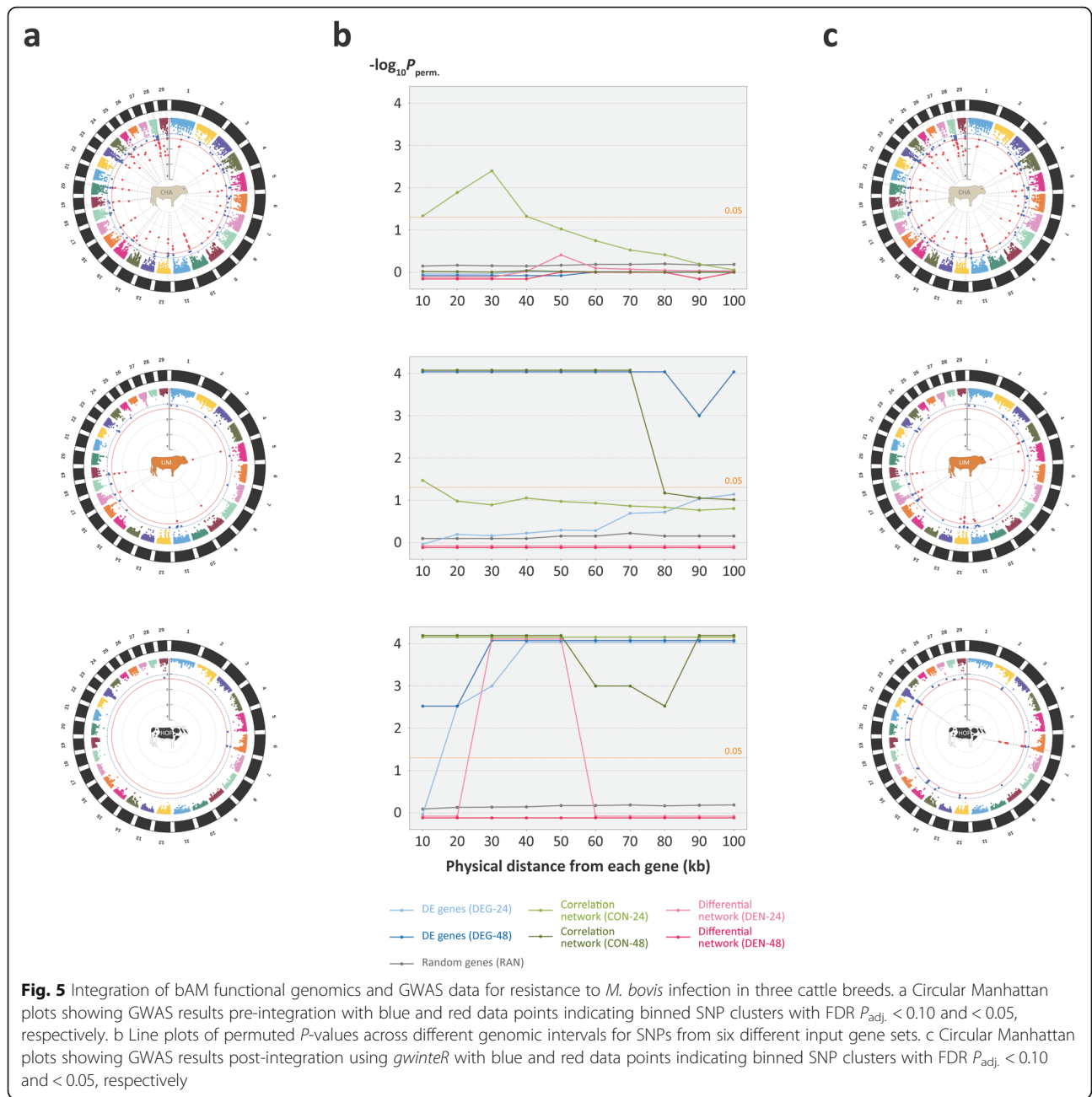
Figure 5a shows circular Manhattan plots with GWAS results ($P_{adj.}$ values) for each of the three breeds prior to data integration using *gwinteR*. Figure 5b shows the *gwinteR* permuted *P*-values ($P_{perm.}$) for each of the 10 genomic intervals used and for each of the six input gene sets plus the RAN gene set with $P_{adj.} < 0.10$. Figure 5c shows circular Manhattan plots with GWAS results post data integration using *gwinteR*. Inspection of Fig. 5b shows that, in terms of SNP enrichment ($P_{perm.} < 0.05$), the integrative analyses using *gwinteR* were most effective for the HOF breed group where the CON-24, CON-48, and DEG-48 input gene sets produced enriched SNPs across all 10 genomic ranges. In addition, the DEG-24 and DEN-24 input gene sets were effective for the HOF breed across the ±20 to 100 kb and ±30 to 50 kb genomic ranges, respectively. In the case of the LIM breed, the DEG-48 input gene set produced enriched SNPs across all 10 genomic ranges, the CON-48 between ±10 to 70 kb and the CON-24 at ±10 kb. For the CHA breed, SNP enrichment using *gwinteR* was only observed for the CON-24 input gene set for the genomic interval between ±10 to 40 kb.

Figure 6 summarises the numbers of statistically significant SNPs pre- and post-data integration, again with SNP enrichment being most evident for the HOF breed with a 24-fold post-*gwinteR* SNP enrichment at $P_{adj.} < 0.10$ from 40 to 961 SNPs. The SNP enrichments for the CHA and LIM breed were more modest, although there

was a 2.3-fold enrichment at $P_{adj.} < 0.10$ from 80 to 182 SNPs for the LIM breed. Inspection of Additional file 7 reveals notable gene loci associated with enriched GWAS SNPs for the HOF breed including the allograft inflammatory factor 1 gene (*AIF1*), which encodes a protein that promotes macrophage activation and proinflammatory activity [63]; the ciliogenesis associated kinase 1 gene (*CILK1* aka *ICK*); *IL17A* and *IL7F*, which encode proinflammatory cytokines and contain polymorphisms in the human orthologs that have been associated with lower human TB disease incidence [64, 65]; the integrin subunit beta 3 gene (*ITGB3*), which encodes a protein that has been shown to regulate matrix metalloproteinase secretion in pulmonary human TB [66]; the neuraminidase 1 gene (*NEU1*), which encodes a protein that regulates phagocytosis in macrophages [67]; and the *TNF* gene that encodes TNF (aka TNF-α), a key cytokine for generation and maintenance of the granuloma, and where gene polymorphisms have been linked to resistance to *M. bovis* infection [68]. Gene loci associated with enriched GWAS SNPs for the CHA breed group also included the *CILK1* gene (aka *ICK*) and the Kelch repeat and BTB domain containing 3 gene (*KCNJ15*), which has been detected as an expression biomarker for human TB [69]; the T cell immune regulator 1, ATPase H+ transporting V0 subunit a3 gene (*TCIRG1*), a known antimycobacterial host defence gene that has been shown to be a key hub gene associated with IFN-γ stimulation of human macrophages [70]; and the Von Willebrand factor gene (*VWF*). Notable gene loci associated with enriched GWAS SNPs for the Limousin breed group included the cellular repressor of E1A stimulated genes 1 gene (*CREG1*), which encodes a regulator of core macrophage differentiation genes [71]; the desmoplakin gene (*DSP*), which increases in expression during *M. tuberculosis*-derived ESAT6-regulated transition of bone marrow-derived macrophages (BMDMs) into epithelioid macrophages [72]; and the SP110 nuclear body protein gene (*SP110*), which encodes a protein that modulates growth of MTBC pathogens in macrophages and has been successfully exploited for genome editing of cattle to enhance resistance to *M. bovis* infection [73].
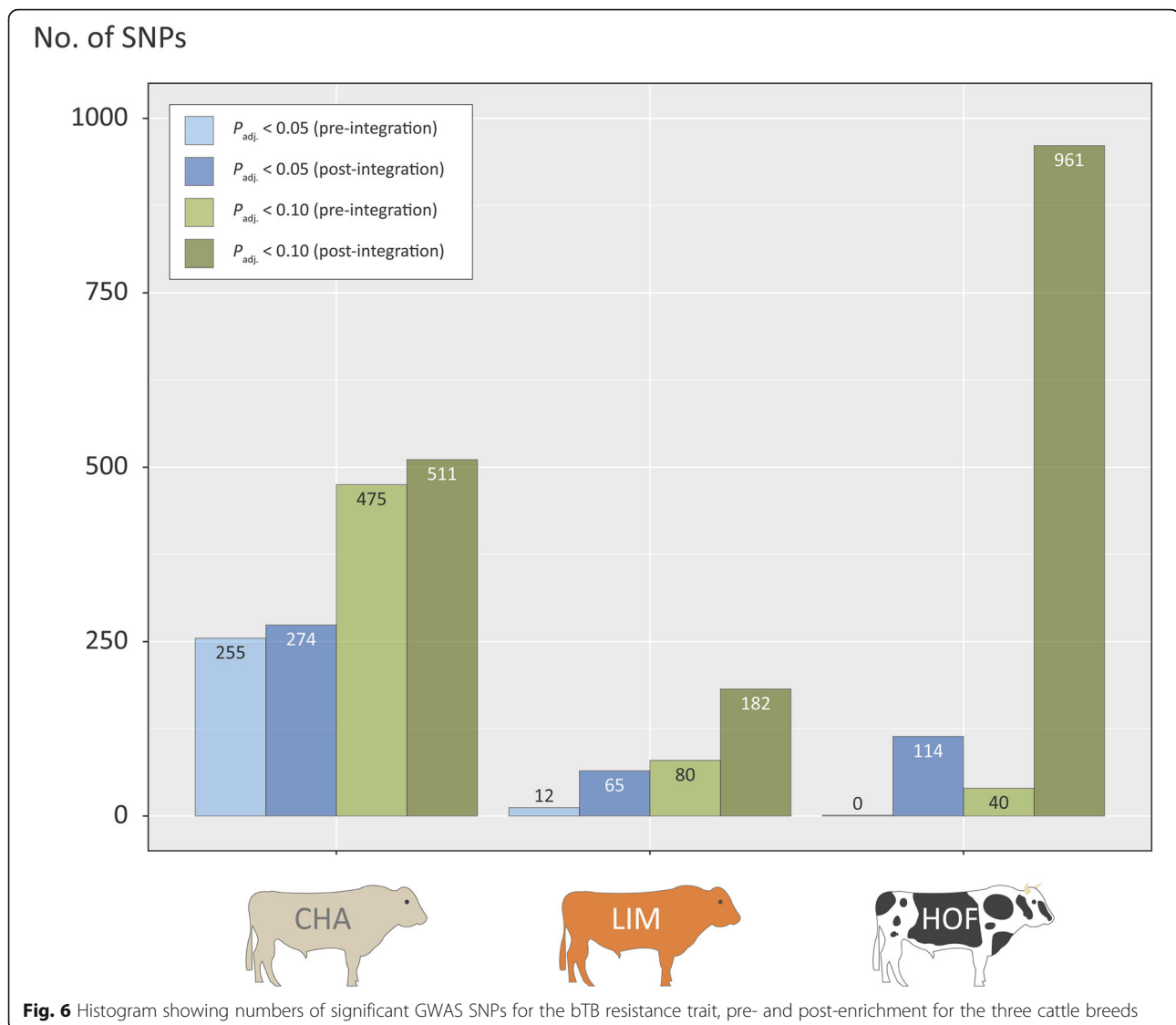
Discussion

During the last decade, integrative genomics, multi-omics analyses and network biology have come to the fore as powerful strategies for exploring, dissecting

Fig. 5 Integration of bAM functional genomics and GWAS data for resistance to *M. bovis* infection in three cattle breeds. a Circular Manhattan plots showing GWAS results pre-integration with blue and red data points indicating binned SNP clusters with FDR $P_{adj.} < 0.10$ and < 0.05, respectively. b Line plots of permuted *P*-values across different genomic intervals for SNPs from six different input gene sets. c Circular Manhattan plots showing GWAS results post-integration using *gwinteR* with blue and red data points indicating binned SNP clusters with FDR $P_{adj.} < 0.10$ and < 0.05, respectively

and unpicking the complexities of the vertebrate immune system and immune responses to specific microbial pathogens [74–77]. In the present study we have used these approaches to integrate transcriptomics data from a pivotal immune effector cell in *M. bovis* infection with high-resolution GWAS data for a bTB resistance trait. We developed a new computational tool, *gwinteR*, to enhance detection of QTLs by leveraging nominal SNP *P*-values from large GWAS data sets for resistance to infection by *M. bovis* in cattle. Three different integration strategies were employed with transcriptomics data from bAM

infected with *M. bovis* across a 48-h time course. The first and most straightforward method was based on DE gene sets at 24 and 48 hpi with stringent fold-change and *P*-value thresholds as filtering criteria (DEG-24 and DEG-48). For the second method, a correlation network approach was used to identify subnetworks (modules) of co-expressed functional gene clusters from the bAM transcriptomics data at the two post-infection time points (CON-24 and CON-48). The third method was also network-based but took advantage of the extensive scientific literature and curated biomolecular data for mycobacterial

Fig. 6 Histogram showing numbers of significant GWAS SNPs for the bTB resistance trait, pre- and post-enrichment for the three cattle breeds

infections and tuberculosis disease. For this approach, a base GIN was constructed and functional modules containing overlaid differentially expressed genes were identified to provide two post-infection input gene sets (DEN-24 and DEN-48) for downstream data integration.

Integration of these six input gene sets with bTB GWAS data from three different cattle breeds (CHA, LIM and HOF) revealed substantial differences among the three methods in their capacity to detect additional QTLs for a *M. bovis* infection resistance trait in these particular GWAS data sets. For example, the correlation network approach was the only method that enriched SNPs for the CHA breed and that worked for at least one bAM post-infection time point across all three breeds (see Fig. 5). Surprisingly, perhaps, the functional

modules obtained using the GIN differential network (DEN) approach—the most complex method to implement—produced the least effective input gene sets for prioritising additional SNPs from the GWAS data set. This method proved effective only for the DEN-24 input gene set with the HOF breed and then only for the ±30, ±40, and ± 50 kb genomic intervals. Conversely, the simplest method based on DE genes at 24 and 48 hpi enriched SNPs for both the LIM and HOF breed groups, with the DEG-48 input gene set being the most effective (Fig. 5). The relatively poor performance of the DEN approach for integration of functional genomics and GWAS data may be a consequence of the human-focused GeneCards and InnateDB resources we used to generate the base GIN [62, 78]. Therefore, it would be instructive to conduct similar integrative analyses for

human TB using appropriate functional genomics and GWAS data sets.

In summary, across the two different bAM infection time points and the three breeds, the CON method enriched 970 SNPs, the DEG method enriched 163 SNPs and the DEN method enriched only 11 SNPs (Additional file 7). Although other factors such as linkage disequilibrium need to be considered in interpreting these differences, it is reasonable to hypothesise that GWAS integration using the correlation network approach is more sensitive to regulatory genomic variants that alter expression of co-ordinately regulated protein components of the alveolar macrophage pathways and processes underpinning host-pathogen interaction for the early stages of intracellular MTBC infection [27, 28, 30]. Figure 7 illustrates this using the example of SNPs within and proximal to gene loci associated with PI3K/AKT signalling, which based on IPA data mining for bAM DE genes was a highly activated pathway, particularly at 24 hpi. In this regard, previous work has shown that macrophage PI3K/AKT signalling is key to a range of cellular processes associated with host-pathogen interaction in MTBC infections, including modulation of cell death pathways, manipulation of signalling downstream of TLRs, and initiation of granuloma formation [79–84]. We have also recently demonstrated that genes encoding protein products embedded in the PI3K/AKT pathway are primary targets for chromatin modifications that substantially alter bAM gene expression in response to *M. bovis* infection [46].

There were also notable breed differences in the effectiveness of the three methods used for multi-omics data integration and enrichment of GWAS SNPs. The original GWAS data set for the HOF breed was the least powered of the three breeds in terms of both sample size ($n = 1502$ sires) and genetic markers (12,740,315 genome-wide SNPs). This is reflected in the relatively small number of GWAS SNPs that were detected pre-integration for the HOF breed: 40 SNPs at $P_{\text{adj.}} < 0.10$ compared to 475 for the CHA breed and 80 for the LIM breed (see Fig. 6). However, the SNP enrichment post-integration was markedly more effective for the HOF breed with a 24-fold increase from 40 to 961 SNPs ($P_{\text{adj.}} < 0.10$) compared to 1.08-fold (475 to 511 SNPs; $P_{\text{adj.}} < 0.10$) and 2.28-fold (80 to 182 SNPs; $P_{\text{adj.}} < 0.10$) for the CHA and LIM breeds, respectively (Fig. 6). In addition, a total of 48 genes were captured by the enriched GWAS SNPs for the HOF breed compared to 20 and 16 genes for the CHA and LIM breeds, respectively (Additional file 7).

Regarding the enhanced SNP detection observed for the HOF breed GWAS after multi-omic integration, it is noteworthy that the *M. bovis* infection transcriptomics data set was generated using bAM from this population [43]. Most of the SNPs ($n = 829$, 88%) and genes ($n = 36$, 67%) obtained post-integration in the HOF breed were detected using the correlation co-expression network method, which, again, is likely caused by enrichment of genomic regulatory variants that modulate expression of genes associated with alveolar macrophage pathways and processes critical to early host-pathogen interaction. This pattern of variation may also reflect the polygenic architecture of *M. bovis* infection resistance in domestic cattle [52, 54, 55, 85].

It is well established that intracellular mycobacterial pathogens use a host of evolved mechanisms to survive within, and ultimately disseminate from host macrophage cells. Several of these mechanisms are ultimately linked to regulators of TNF and cognate downstream signalling pathways [29, 30, 39]. It is noteworthy, therefore, that our analysis prioritised SNPs at the *TNF* gene locus, a key nexus for host-pathogen interaction in mycobacterial infections. In addition, another important component of host-pathogen interaction and target for immunoevasion by intracellular mycobacterial infections is the PI3K/AKT pathway [33, 79, 83] and our integrative genomics approach was able to prioritise genomic variation at genes encoding ligands and receptors within this signalling network, specifically *TNXB*, *GNG11*, *ITGB3*, and *VWF* (see Fig. 7).

Conclusions

Elucidation of the mechanisms used by *M. bovis* to establish infection in cattle, and ultimately cause disease, requires an intimate knowledge of host-pathogen interactions, especially at the interface between the bAM and the invading bacilli. Using a systems biology approach, we identified and catalogued patterns of gene expression and gene-gene interactions that occur in bAM during the initial stages of *M. bovis* infection, highlighting key response pathways and hub genes. Additionally, a new R package, *gwinteR*, facilitated integration of these functional genomics outputs with three large breed based GWAS data sets for resistance/susceptibility to *M. bovis* infection. This revealed genomic variants associated with resistance in key innate immune genes and supported the hypothesis that the response to *M. bovis* infection at the level of the alveolar macrophage may be more breed-specific than previously assumed.

Integrative multi-omics approaches to data integration are now widely used to explore and dissect the genomic architecture and physiological basis of complex traits in domestic livestock, including network-based methods to integrate functional genomics and GWAS data [86–91]. However, to the best of our knowledge, this study is the first that uses network biology to systematically combine transcriptomics data from *M. bovis*-infected macrophages with GWAS data for *M. bovis* infection resistance in
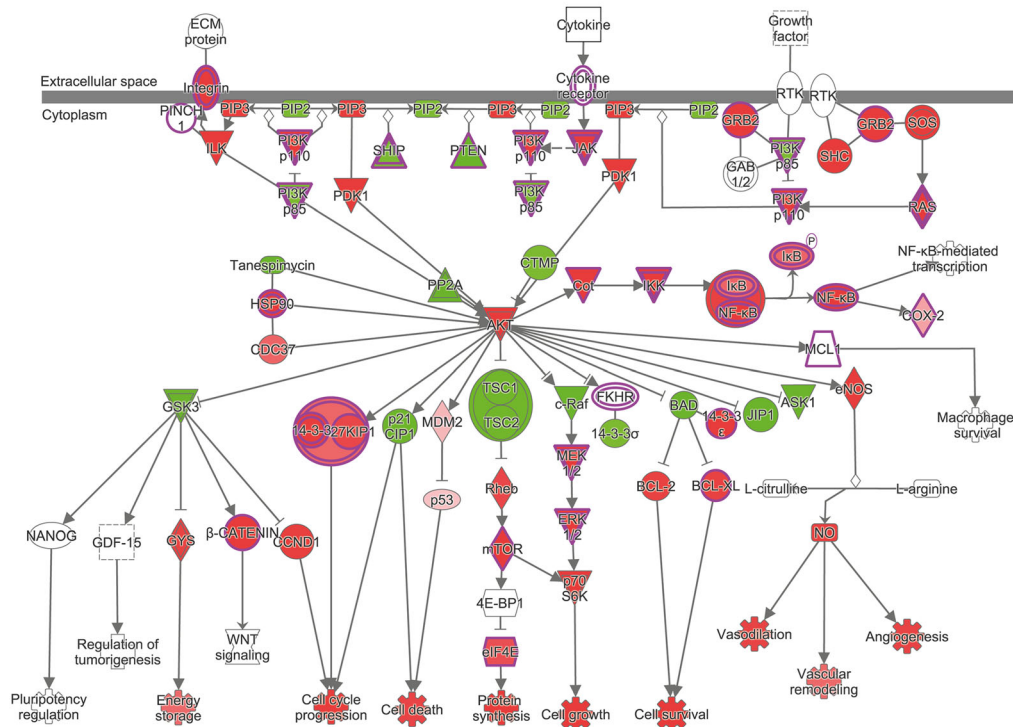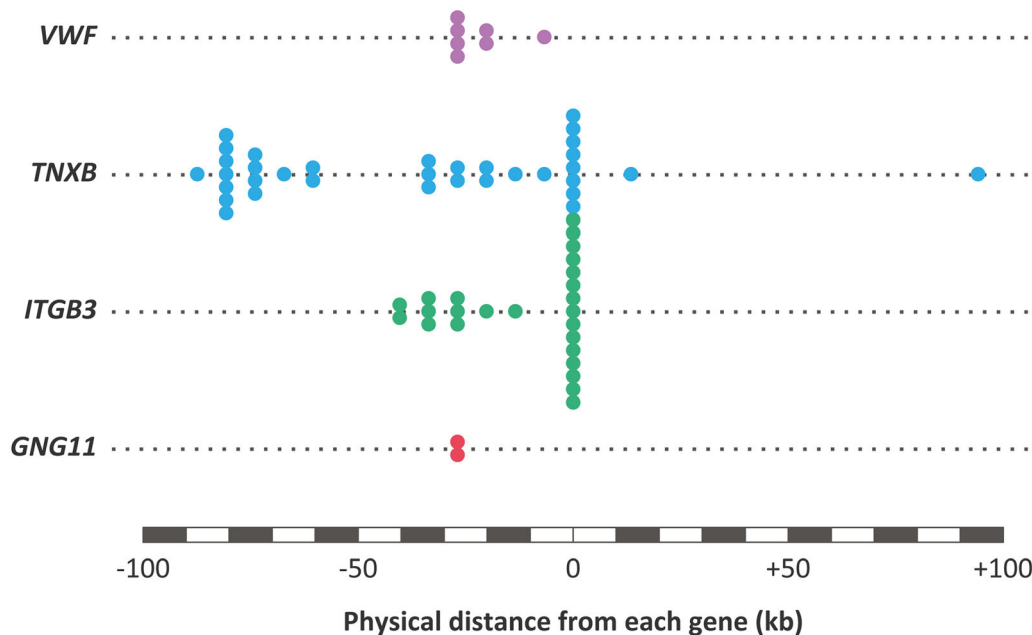
Fig. 7 The PI3K/AKT signalling pathway and genomic variants within associated genes. **a** The IPA-predicted activation state of the PI3K/AKT signalling pathway with red and blue colours indicating increased and decreased activity of pathway components, respectively. **b** Schematic of enriched bTB disease resistance GWAS SNPs in four genes associated with the PI3K/AKT signalling pathway

cattle. Therefore, it provides a novel framework for integrative genomics studies of complex infectious disease resistance traits in livestock, particularly those involving other intracellular bacterial pathogens such as *Brucella abortus*, *Mycobacterium avium subsp. paratuberculosis* and *Salmonella enterica*. The work is also relevant to development of methods for integrative analyses of outputs from the Functional Annotation of Animal Genomics (FAANG) initiative [92] and for identification and prioritization of targets for genome editing to enhance resistance to infection in domestic livestock species [93]. The results from this study may also inform genome-enabled breeding programmes for resistance to *M. bovis* infection in production cattle populations [85, 94]. Finally, this integrative multi-omics approach could also be used to combine relevant functional genomics and GWAS data sets to improve knowledge of innate immune responses and establishment of infection in human TB caused by *M. tuberculosis*.

Methods

Genomics data acquisition and computational and bioinformatics workflow

Genome-wide RNA-seq transcriptomics data from a 48-h bAM time course challenge experiment using the sequenced *M. bovis* AF2122/97 strain have been previously generated by our group (GEO accession: GSE62506). The complete laboratory methods used to isolate, culture and infect bAM with *M. bovis* AF2122/9 and generate strand-specific RNA-seq libraries using RNA harvested from these cells are described in detail elsewhere [42, 43, 45]. Briefly, these RNA-seq data were generated using bAM obtained by lung lavage of ten unrelated age-matched 7 to 12-week-old male Holstein-Friesian calves. Bovine AM were either infected in vitro with *M. bovis* AF2122/97 or incubated with media only. Following total RNA extraction from *M. bovis*-infected and control non-infected alveolar macrophages, 78 strand-specific RNA-seq libraries were prepared (paired-end 2 × 90 nucleotide reads). These comprised *M. bovis*- and non-infected samples from each post-infection time point (2, 6, 24 and 48 hpi) across 10 animals with the exception of one animal that did not yield sufficient alveolar macrophages for in vitro infection at 48 hpi.

GWAS data sets for the present study were obtained from intra-breed imputed WGS-based GWAS analyses that used estimated breeding values (EBVs) derived from a bTB infection phenotype, which were generated for 2039 Charolais, 1964 Limousin and 1502 Holstein-Friesian sires [55]. The bTB phenotype, the WGS-based imputed SNP data, and the quantitative genetics methods are described in detail elsewhere [55]; however,

the following provides a brief summary. The bTB infection phenotype was defined for every animal present during each herd-level bTB breakdown when a bTB reactor or a slaughterhouse case was identified. Cattle that yielded a positive single intradermal comparative tuberculin test (SICTT), and/or post-mortem lymph node lesion, or laboratory culture result/s were coded as bTB = 1 and all other cattle present in the herd during the bTB-breakdown were coded as bTB = 0; potential exposure of cattle within the bTB breakdown was also considered in this study [55]. After phenotype data edits, bTB resistance EBVs were generated for 781,270 phenotyped cattle (plus their recorded ancestors). After within-breed SNP filtering using thresholds for minor allele frequency (MAF ≤ 0.002) and deviation from Hardy-Weinberg equilibrium (HWE; $P < 1 \times 10^{-6}$), there were 17,250,600, 17,267,260 and 15,017,692 autosomal SNPs for the 2039 Charolais (CHA), 1964 Limousin (LIM) and 1502 Holstein-Friesian (HOF) sire analyses. A single-SNP regression analyses was performed for each breed separately using weighted (i.e., by an effective record contribution) sire EBVs for *M. bovis* infection resistance/susceptibility and the nominal *P*-values were used for downstream integrative genomics analyses.

All data-intensive computational procedures were performed on a 36-core/72-thread compute server (2× Intel® Xeon® CPU E5–2697 v4 processors, 2.30 GHz with 18 cores each), with 512 GB of RAM, 96 TB SAS storage (12 × 8 TB at 7200 rpm), 480 GB SSD storage, and with Ubuntu Linux OS (version 18.04 LTS). The complete computational and bioinformatics workflow is available with additional information as a public GitHub repository (github.com/ThomasHall1688/Bovine_multi-omic_integration). The individual components of the experimental and computational workflows are shown in Fig. 1 and described in more detail below.

Differential gene expression analysis of RNA-seq data

A custom Perl script was used to deconvolute barcoded RNA-seq reads into individual libraries, filter out adapter sequence reads, and remove poor quality reads [43]. At each stage of the process, a quality check was performed on the FASTQ files with FastQC (version 0.11.8) [95]. Paired-end sequence reads were then aligned to the *Bos taurus* reference genome (ARS-UCD1.2, GenBank assembly accession: GCA_002263795.2) [96] using the STAR aligner (version 2.7) [97]. Read counts for each gene were calculated using featureCounts (version 1.6.4) [98], set to unambiguously assign uniquely aligned paired-end reads in a stranded manner to gene exon annotation. Using the R statistical programming language (version 3.5.3) [99], gene annotation was derived from the NCBI database via a GFF annotation file GCF_

002263805.1 with additional descriptions and chromosomal locations annotated using GO.db (version 3.8.2) [100] and biomaRt packages (version 2.40.0) [101]. Differential gene expression analysis was performed using the DESeq2 package (version 1.24.0) [102] with a longitudinal time series design that accounted for time (hpi) and experimental treatment (*M. bovis*-infected versus control). Lowly expressed reads were removed using the mean of normalized counts as a filter statistic; individual genes with very low read counts would typically not exhibit significant differential expression due to high dispersion [102]. In addition, extreme count outliers were removed within DESeq2 using the Cook's distance [103] as previously described [102]. Multiple testing correction was performed on each time point using the Benjamini-Hochberg (B-H) false discovery rate (FDR) method [104]. The default criteria for differentially expressed (DE) genes were an FDR-adjusted *P*-value less than 0.05 ($P_{adj.} < 0.05$).

Ingenuity® pathway analysis (IPA) of differentially expressed genes

Ingenuity® Pathway Analysis—IPA® (version 1.1, summer 2020 release; Qiagen, Redwood City, CA, USA) was used to perform a statistical enrichment analysis of DE gene sets and expression data [58]. This enabled identification of canonical pathways and functional processes of biological importance in alveolar macrophages challenged with *M. bovis* across the longitudinal infection time course. The target species selected was *Homo sapiens* and the cell type used was *Macrophage* with the *Experimentally Observed* and *High Predicted* confidence settings. Following best practice, the default background gene set for pathway and functional process enrichment testing was the set of detectable genes across all RNA-seq libraries for each time point contrast and not the complete bovine transcriptome [105].

Identification of functional gene modules using differential co-expression network analysis

An integrated computational pipeline for differential co-expression network analysis was implemented using: 1) the Differential Gene Correlation Analysis (DGCA) R package (version 1.0.2) [61]; 2) the Cytoscape open source Java platform for visualisation and integration of biomolecular interaction networks (version 3.7.0) [106]; and 3) the Multiscale Embedded Gene Co-expression Network Analysis (MEGENA) R package (version 1.3.7) [107].

Due to the low variance in gene expression observed among samples at 2 and 6 hpi (Fig. 2a and b), only data from the 24 and 48 hpi time points were used for the differential correlation analysis. The DGCA R package was used to filter normalised gene

counts such that genes in the lower 30th percentile of median expression values were removed. Pearson correlation coefficients were then calculated for each gene pair between the control non-infected and *M. bovis*-infected samples at 24 and 48 hpi. Following this, for each time point, the infected and control samples were randomly shuffled, and the analysis was repeated for a total of ten iterations. Additionally, using the permutation testing and reference pool distribution approaches implemented in DGCA, an empirical *P*-value was calculated for each observed correlation coefficient and *q*-values were calculated based on empirical *P*-values and the estimated proportion of null hypotheses; gene pairs were then considered to be differentially correlated with a *q*-value threshold of 0.10 [61].

The correlation networks and network parameters generated by DGCA were initially visualised, examined and evaluated using the Cytoscape platform. In correlation networks, based on gene co-expression, each gene acts as a node and each correlation acts as a weighted edge, depending on the strength of the correlation coefficient [108, 109]. The DGCA network data for the 24 and 48 hpi time points were then imported into MEGENA for identification of functional subnetworks (modules) using differentially correlated ($q < 0.10$) gene pairs ranked by *q*-value to a maximum of 3500 gene pairs; this gave 3500 and 1085 gene pairs for the 24 and 48 hpi correlation networks, respectively. Each gene pair was assigned to a class that described the change in correlation depending on infection status, and only those genes that exhibited a change in expression pattern were included in the MEGENA analysis to identify functional modules.

Functional modules in the 24 and 48 hpi correlation networks were detected as locally coherent subnetwork clusters with a minimum of 20 unique genes that MEGENA classified as statistically significant (*P*-value < 0.05) based on analyses of shortest path indices, local path index, weight of the correlation and overall modularity. The resulting MEGENA functional modules were visualised using the Cytoscape platform and genes embedded in functional modules at 24 and 48 hpi were combined and annotated for downstream GWAS integration as described below. The genes contained in these functional modules were also subject to gene ontology (GO) term enrichment analyses within the MEGENA package [107].

Detection of active gene subnetworks using a tuberculosis and mycobacterial infection gene interaction network

The GeneCards® gene compendium and knowledgebase (www.genecards.org; version 4.9), which integrates

multiple sources of biological information on all annotated and predicted human genes [78], was used to identify a set of genes that are functionally associated with the host response to TB and other diseases caused by infection with mycobacteria. The GeneCards search query generated a total of 2291 gene hits (Additional file 4) using the search terms: `tuberculosis OR mycobacterium OR mycobacteria OR mycobacterial`. Genes were ranked by a GeneCards statistic—the *Relevance Score*—based on the *Elasticsearch* algorithm [110], which determines the strength of the relationships between genes and keyword terms. Gene IDs were converted to human Ensembl gene IDs [111] and retained for downstream analysis using the InnateDB knowledgebase and analysis platform for systems level analysis of the innate immune response (www.innatedb.com; version 5.4) [62].

A gene interaction network (GIN) was generated with the gene list output from GeneCards using InnateDB with default settings and this network was visualised using Cytoscape. The jActivesModules Cytoscape plugin (version 3.12.1) [112] was then used to superimpose the bAM RNA-seq gene expression data and detect, through a greedy search algorithm, differentially active subnetworks (modules) of genes at the 24 and 48 hpi time points. Locally coherent clusters that contain genes that are differentially expressed were identified using the $\log_2$FC and $P_{adj.}$ values of each differentially expressed gene; the overall connectivity of those genes with their immediate module co-members; and the comparison of that connectivity with a background comprised of randomly drawn networks using the same genes, but independent of the base network. Genes embedded in active modules that were detected as statistically significant at 24 and 48 hpi were combined and annotated for downstream GWAS integration as described below.

Integration of *M. bovis*-infected bovine AM gene expression data with bTB GWAS data

To facilitate integration of GWAS data with gene sets generated from functional genomics data analyses, an R software package was developed—*gwinteR* (github.com/ThomasHall1688/gwinteR), which can be used to test the hypothesis that a specific set of genes is enriched for signal in a GWAS data set relative to the genomic background. This gene set, for example, could be an output from an active gene module network analysis of transcriptomics data from a cell type or tissue relevant to the GWAS phenotype. To formally test the primary hypothesis, the *gwinteR* tool was designed to determine if genomic regions containing GWAS SNPs that are proximal to genes within a gene set are enriched for statistical associations with the trait/s analysed in the GWAS.

The *gwinteR* tool works as follows: 1) a set of significant and non-significant SNPs (named the target SNP set) is collated across all genes in a specific gene set at increasing genomic intervals upstream and downstream from each gene inclusive of the coding sequence (e.g., ± 10 kb, ±20 kb, ±30 kb … … ±100 kb); 2) for each genomic region, a null distribution of 1000 SNP sets, each of which contains the same number of total significant and non-significant combined SNPs as the target SNP set, is generated by resampling with replacement from the search space of the total population of SNPs in the GWAS data set; 3) the nominal (uncorrected) GWAS *P*-values for the target SNP set and the null distribution SNP sets are converted to local FDR-adjusted *P*-values ($P_{adj.}$) using the fdrtool R package (current version 1.2.15) [113]; 4) a permuted *P*-value ($P_{perm.}$) to the test the primary hypothesis for each observed genomic interval target SNP set is generated based on the proportion of permuted random SNP sets where the same or a larger number of SNPs exhibiting significant *q*-values (e.g. $q < 0.05$ or $q < 0.10$) are observed; 5) *gwinteR* generates data to plot $P_{perm.}$ results by genomic interval class and obtain a graphical representation of the GWAS signal surrounding genes within the target gene set; and 6) a summary output file of all SNPs in the observed target SNP set with genomic locations and *q*-values is generated for subsequent investigation.

In the original bTB GWAS data set used for the present study [55], the WGS-imputed SNPs were mapped to the UMD3.1 bovine genome assembly [114]. Consequently, prior to GWAS data integration, the imputed and previously filtered SNPs for each of the three breed groups were mapped, using a custom R pipeline (github.com/ThomasHall1688/Bovine_multi-omic_integration), to the most recent ARS-UCD1.2 cattle reference genome assembly [96]. After this step, there were 14,583,567, 14,586,972 and 12,740,315 autosomal SNPs with nominal GWAS *P*-values that could be used for integrative genomics analyses of the CHA, LIM and HOF breeds, respectively.

For the integrative analyses of bAM functional genomics outputs with the bTB GWAS data, three different subsets of genes were used: 1) basic DE gene sets that were filtered to ensure manageable computational loads using stringent expression threshold criteria of $|\log_2$FC$| > 2$ and $P_{adj.} < 0.01$ and $P_{adj.} < 0.000001$ for 24 and 48 hpi, respectively; 2) genes embedded in functional modules at 24 and 48 hpi that were detected using the MEGENA package in the differential co-expression network analyses; and 3) genes embedded in active modules at 24 and 48 hpi that were identified using jActiveModules within the tuberculosis and mycobacterial infection interaction network.

Abbreviations
AM: Alveolar macrophage; bAM: Bovine alveolar macrophage; B-H: Benjamini-Hochberg; bTB: Bovine tuberculosis; CHA: Charolais; ChIP-seq: Chromatin immunoprecipitation sequencing; CON: Correlation network; DEG: Differentially expressed gene; DEN: Differential expression network; DGCA: Differential Gene Correlation Analysis; EBV: Estimated breeding value; GCRS: GeneCards Relevance Score; FDR: False discovery rate; GIN: Gene interaction network; GO: Gene ontology; GWAS: Genome-wide association study; HOF: Holstein-Friesian; HWE: Hardy-Weinberg equilibrium; IFN: Interferon; IPA: Ingenuity Pathway Analysis; LIM: Limousin; MEGE NA: Multiscale Embedded Gene Co-expression Network Analysis; MAF: Minor allele frequency; MTBC: *Mycobacterium tuberculosis* complex; QTL: Quantitative trait locus; RNA-seq: RNA sequencing; RNI: Reactive nitrogen intermediate; ROS: Reactive oxygen species; SICTT: Single intradermal comparative tuberculin test; SNP: Single nucleotide polymorphism; TB: Tuberculosis; WGS: Whole-genome sequence; zTB: Zoonotic tuberculosis

Supplementary Information
The online version contains supplementary material available at https://doi.org/10.1186/s12864-021-07643-w.

Additional file 1. Differentially expressed genes across the infection time course and DEG-24 and DEG-48 input gene sets for GWAS integration.

Additional file 2. Enriched IPA canonical pathways at 24 and 48 hpi.

Additional file 3. Outputs from correlation network analyses using DGCA with MEGENA and CON-24 and CON-48 input gene sets for GWAS integration.

Additional file 4. Outputs from differential network analyses using GeneCards®, InnateDB with Cytoscape/jActiveModules and DEN-24 and DEN-48 input gene sets for GWAS integration.

Additional file 5. 100 random (RAN) input gene sets used for GWAS integration.

Additional file 6. Functional genomics and GWAS integration results for Charolais (CHA), Limousin (LIM) and Holstein-Friesian (HOF).

Additional file 7. GWAS prioritised SNP results for Charolais (CHA), Limousin (LIM) and Holstein-Friesian (HOF).

Additional file 8: S1 Fig. Principal component analysis (PCA) plots for individual animal bAM gene expression data at **a** 2 hpi, **b** 6 hpi, **c** 24 hpi, and **d** 48 hpi. **S2 Fig.** Gene ontology (GO) enrichment for functional modules identified from the differential co-expression correlation network generated from *M. bovis*-infected bAM gene expression at 24 hpi. **S3 Fig.** Gene ontology (GO) enrichment for functional modules identified from the differential co-expression correlation network generated from *M. bovis*-infected bAM gene expression at 48 hpi. **S4 Fig.** Venn diagram illustrating the overlaps for the three 24 hpi input gene sets used for integration with cattle GWAS data sets. **S5 Fig.** Venn diagram illustrating the overlaps for the three 48 hpi input gene sets used for integration with cattle GWAS data sets. RNA-seq statistics and results for this study are available at the Dryad Digital Repository: https://doi.org/10.5061/dryad.83bk3j9q6.

Authors' contributions
TJH, MPM, KEK, DPB, SVG and DEM conceived and designed the study. JAB performed experimental work. SCR and DPB provided cattle genotype and other genomic data. TJH, MPM, GPM, KEK, SCR and CNC performed bioinformatics and computational analyses. TJH, CNC and DEM wrote and prepared the manuscript and figures. All authors read and approved the final manuscript.

Availability of data and materials
The RNA-seq data set was generated by the authors and can be obtained from the NCBI Gene Expression Omnibus (GEO): accession number GSE62506. GWAS summary statistics data were obtained from a published study that provides additional information about sequence and genotype data availability [55]. The source code to reproduce the results of the analyses is available from a GitHub repository (https://github.com/ThomasHall1688/Bovine_multi-omic_integration). The source code used to integrate the outputs from functional genomics and GWAS data sets is available from a GitHub repository (https://github.com/ThomasHall1688/gwinteR) Detailed RNA-seq filtering and mapping statistics are provided in a data file available from the Dryad Digital Repository: https://doi.org/10.5061/dryad.83bk3j9q6. The computational pipeline and software used for this study is detailed below.
Complete pipeline: https://github.com/ThomasHall1688/Bovine_multi-omic_integration
gwinteR: https://github.com/ThomasHall1688/gwinteR
DESeq2: https://github.com/mikelove/DESeq2
DGCA: https://github.com/andymckenzie/DGCA
MEGENA: https://github.com/cran/MEGENA
Cattle genome assembly and annotation: www.ncbi.nlm.nih.gov/datasets/genomes/?txid=9913

Declarations

Ethics approval and consent to participate
All animal procedures were performed in accordance with EU Directive 2010/63/EU and Irish Statutory Instrument 543/2012, with ethical approval from the University College Dublin (UCD) Animal Research Ethics Committee (AREC-13–14-Gordon).

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Animal Genomics Laboratory, UCD School of Agriculture and Food Science, UCD College of Health and Agricultural Sciences, University College Dublin, Belfield, Dublin D04 V1W8, Ireland. [2]Bioscience Research Institute, Athlone Institute of Technology, Dublin Road, Athlone, Westmeath N37 HD68, Ireland. [3]Present address: Genuity Science, Cherrywood Business Park. Loughlinstown, Dublin D18 K7W4, Ireland. [4]Irish Cattle Breeding Federation, Highfield House, Shinagh, Bandon, Cork P72 X050, Ireland. [5]Teagasc, Animal and Grassland Research and Innovation Centre, Moorepark, Fermoy, Cork P61 C996, Ireland. [6]UCD School of Veterinary Medicine, UCD College of Health and Agricultural Sciences, University College Dublin, Belfield, Dublin D04 V1W8, Ireland. [7]UCD Conway Institute of Biomolecular and Biomedical Research, University College Dublin, Belfield, Dublin D04 V1W8, Ireland.

References
1. Steele JH. Introduction (Part 2 Regional and Country Status Reports). In: Thoen CO, Steele JH, editors. Mycobacterium bovis infection in animals and humans. Ames: Iowa State University Press; 1995. p. 169–72.
2. Waters WR, Palmer MV, Buddle BM, Vordermeier HM. Bovine tuberculosis vaccine research: historical perspectives and recent advances. Vaccine. 2012; 30(16):2611–22. https://doi.org/10.1016/j.vaccine.2012.02.018.

Hall *et al. BMC Genomics*   (2021) 22:343

Page 18 of 20

3.   Fitzgerald SD, Kaneene JB. Wildlife reservoirs of bovine tuberculosis worldwide: hosts, pathology, surveillance, and control. Vet Pathol. 2013; 50(3):488–99. https://doi.org/10.1177/0300985812467472.

4.   Gormley E, Corner LAL. Wild animal tuberculosis: stakeholder value systems and management of disease. Front Vet Sci. 2018;5:327. https://doi.org/10.3389/fvets.2018.00327.

5.   Malone KM, Gordon SV. *Mycobacterium tuberculosis* complex members adapted to wild and domestic animals. Adv Exp Med Biol. 2017;1019:135–54. https://doi.org/10.1007/978-3-319-64371-7_7.

6.   Palmer MV. *Mycobacterium bovis*: characteristics of wildlife reservoir hosts. Transbound Emerg Dis. 2013;60(Suppl 1):1–13. https://doi.org/10.1111/tbed.12115.

7.   Garnier T, Eiglmeier K, Camus JC, Medina N, Mansoor H, Pryor M, et al. The complete genome sequence of *Mycobacterium bovis*. Proc Natl Acad Sci U S A. 2003;100(13):7877–82. https://doi.org/10.1073/pnas.1130426100.

8.   World Health Organization. Global Tuberculosis Report 2019. Geneva: World Health Organization; 2019.

9.   Luciano SA, Roess A. Human zoonotic tuberculosis and livestock exposure in low- and middle-income countries: a systematic review identifying challenges in laboratory diagnosis. Zoonoses Public Health. 2020;67(2):97–111. https://doi.org/10.1111/zph.12684.

10.   Olea-Popelka F, Muwonge A, Perera A, Dean AS, Mumford E, Erlacher-Vindel E, et al. Zoonotic tuberculosis in human beings caused by *Mycobacterium bovis* – a call for action. Lancet Infect Dis. 2017;17(1):e21–5. https://doi.org/10.1016/S1473-3099(16)30139-6.

11.   Thoen CO, Kaplan B, Thoen TC, Gilsdorf MJ, Shere JA. Zoonotic tuberculosis. A comprehensive ONE HEALTH approach. Medicina (B Aires). 2016;76(3):159–65.

12.   Vayr F, Martin-Blondel G, Savall F, Soulat JM, Deffontaines G, Herin F. Occupational exposure to human *Mycobacterium bovis* infection: a systematic review. PLoS Negl Trop Dis. 2018;12(1):e0006208. https://doi.org/10.1371/journal.pntd.0006208.

13.   Daniel TM. The history of tuberculosis. Respir Med. 2006;100(11):1862–70. https://doi.org/10.1016/j.rmed.2006.08.006.

14.   Cambau E, Drancourt M. Steps towards the discovery of *Mycobacterium tuberculosis* by Robert Koch, 1882. Clin Microbiol Infect. 2014;20(3):196–201. https://doi.org/10.1111/1469-0691.12555.

15.   Neill SD, Bryson DG, Pollock JM. Pathogenesis of tuberculosis in cattle. Tuberculosis (Edinb). 2001;81(1–2):79–86. https://doi.org/10.1054/tube.2000.0279.

16.   Russell DG. Highlighting the parallels between human and bovine tuberculosis. J Vet Med Educ. 2003;30(2):140–2. https://doi.org/10.3138/jvme.30.2.140.

17.   Cassidy JP. The pathogenesis and pathology of bovine tuberculosis with insights from studies of tuberculosis in humans and laboratory animal models. Vet Microbiol. 2006;112(2–4):151–61. https://doi.org/10.1016/j.vetmic.2005.11.031.

18.   Pollock JM, Rodgers JD, Welsh MD, McNair J. Pathogenesis of bovine tuberculosis: the role of experimental models of infection. Vet Microbiol. 2006;112(2–4):141–50. https://doi.org/10.1016/j.vetmic.2005.11.032.

19.   Waters WR, Maggioli MF, McGill JL, Lyashchenko KP, Palmer MV. Relevance of bovine tuberculosis research to the understanding of human disease: historical perspectives, approaches, and immunologic mechanisms. Vet Immunol Immunopathol. 2014;159(3–4):113–32. https://doi.org/10.1016/j.vetimm.2014.02.009.

20.   Hein WR, Griebel PJ. A road less travelled: large animal models in immunological research. Nat Rev Immunol. 2003;3(1):79–84. https://doi.org/10.1038/nri977.

21.   Van Rhijn I, Godfroid J, Michel A, Rutten V. Bovine tuberculosis as a model for human tuberculosis: advantages over small animal models. Microbes Infect. 2008;10(7):711–5. https://doi.org/10.1016/j.micinf.2008.04.005.

22.   Waters WR, Palmer MV, Thacker TC, Davis WC, Sreevatsan S, Coussens P, et al. Tuberculosis immunity: opportunities from studies with cattle. Clin Dev Immunol. 2011;2011:768542.

23.   Williams A, Orme IM. Animal models of tuberculosis: an overview. Microbiol Spectr. 2016;4(4). https://doi.org/10.1128/microbiolspec.TBTB2-0004-2015.

24.   Gong W, Liang Y, Wu X. Animal models of tuberculosis vaccine research: an important component in the fight against tuberculosis. Biomed Res Int. 2020;2020:4263079. https://doi.org/10.1155/2020/4263079.

25.   Palmer MV, Waters WR, Whipple DL. Aerosol delivery of virulent *Mycobacterium bovis* to cattle. Tuberculosis (Edinb). 2002;82(6):275–82. https://doi.org/10.1054/tube.2002.0341.

26.   Palmer MV, Wiarda J, Kanipe C, Thacker TC. Early pulmonary lesions in cattle infected via aerosolized *Mycobacterium bovis*. Vet Pathol. 2019;56(4):544–54. https://doi.org/10.1177/0300985819833454.

27.   Kaufmann SHE, Dorhoi A. Molecular determinants in phagocyte-bacteria interactions. Immunity. 2016;44(3):476–91. https://doi.org/10.1016/j.immuni.2016.02.014.

28.   Weiss G, Schaible UE. Macrophage defense mechanisms against intracellular bacteria. Immunol Rev. 2015;264(1):182–203. https://doi.org/10.1111/imr.12266.

29.   Awuh JA, Flo TH. Molecular basis of mycobacterial survival in macrophages. Cell Mol Life Sci. 2017;74(9):1625–48. https://doi.org/10.1007/s00018-016-2422-8.

30.   Schorey JS, Schlesinger LS. Innate immune responses to tuberculosis. Microbiol Spectr. 2016;4(6). https://doi.org/10.1128/microbiolspec.TBTB2-0010-2016.

31.   Cambier CJ, Falkow S, Ramakrishnan L. Host evasion and exploitation schemes of *Mycobacterium tuberculosis*. Cell. 2014;159(7):1497–509. https://doi.org/10.1016/j.cell.2014.11.024.

32.   de Chastellier C. The many niches and strategies used by pathogenic mycobacteria for survival within host macrophages. Immunobiology. 2009; 214(7):526–42. https://doi.org/10.1016/j.imbio.2008.12.005.

33.   BoseDasgupta S, Pieters J. Macrophage-microbe interaction: lessons learned from the pathogen *Mycobacterium tuberculosis*. Semin Immunopathol. 2018; 40(6):577–91. https://doi.org/10.1007/s00281-018-0710-0.

34.   Chaurasiya SK. Tuberculosis: smart manipulation of a lethal host. Microbiol Immunol. 2018;62(6):361–79. https://doi.org/10.1111/1348-0421.12593.

35.   Hussain Bhat K, Mukhopadhyay S. Macrophage takeover and the host-bacilli interplay during tuberculosis. Future Microbiol. 2015;10(5):853–72. https://doi.org/10.2217/fmb.15.11.

36.   Leopold Wager CM, Arnett E, Schlesinger LS. *Mycobacterium tuberculosis* and macrophage nuclear receptors: What we do and don't know. Tuberculosis (Edinb). 2019;116s:S98–s106. https://doi.org/10.1016/j.tube.2019.04.016.

37.   Queval CJ, Brosch R, Simeone R. The macrophage: a disputed fortress in the battle against *Mycobacterium tuberculosis*. Front Microbiol. 2017;8:2284. https://doi.org/10.3389/fmicb.2017.02284.

38.   Russell DG, Huang L, VanderVen BC. Immunometabolism at the interface between macrophages and pathogens. Nat Rev Immunol. 2019;19(5):291–304. https://doi.org/10.1038/s41577-019-0124-9.

39.   Stutz MD, Clark MP, Doerflinger M, Pellegrini M. *Mycobacterium tuberculosis*: rewiring host cell signaling to promote infection. J Leukoc Biol. 2018;103(2):259–68. https://doi.org/10.1002/JLB.4MR0717-277R.

40.   Widdison S, Watson M, Piercy J, Howard C, Coffey TJ. Granulocyte chemotactic properties of *M. tuberculosis* versus *M. bovis*-infected bovine alveolar macrophages. Mol Immunol. 2008;45(3):740–9. https://doi.org/10.1016/j.molimm.2007.06.357.

41.   Widdison S, Watson M, Coffey TJ. Early response of bovine alveolar macrophages to infection with live and heat-killed *Mycobacterium bovis*. Dev Comp Immunol. 2011;35(5):580–91. https://doi.org/10.1016/j.dci.2011.01.001.

42.   Magee DA, Conlon KM, Nalpas NC, Browne JA, Pirson C, Healy C, et al. Innate cytokine profiling of bovine alveolar macrophages reveals commonalities and divergence in the response to *Mycobacterium bovis* and *Mycobacterium tuberculosis* infection. Tuberculosis (Edinb). 2014;94(4):441–50. https://doi.org/10.1016/j.tube.2014.04.004.

43.   Nalpas NC, Magee DA, Conlon KM, Browne JA, Healy C, McLoughlin KE, et al. RNA sequencing provides exquisite insight into the manipulation of the alveolar macrophage by tubercle bacilli. Sci Rep. 2015;5(1):13629. https://doi.org/10.1038/srep13629.

44.   Vegh P, Magee DA, Nalpas NC, Bryan K, McCabe MS, Browne JA, et al. MicroRNA profiling of the bovine alveolar macrophage response to *Mycobacterium bovis* infection suggests pathogen survival is enhanced by microRNA regulation of endocytosis and lysosome trafficking. Tuberculosis. 2015;95(1):60–7. https://doi.org/10.1016/j.tube.2014.10.011.

45.   Malone KM, Rue-Albrecht K, Magee DA, Conlon K, Schubert OT, Nalpas NC, et al. Comparative 'omics analyses differentiate *Mycobacterium tuberculosis* and *Mycobacterium bovis* and reveal distinct macrophage responses to infection with the human and bovine tubercle bacilli. Microb Genom. 2018; 4(3):e000163. https://doi.org/10.1099/mgen.0.000163.

46.   Hall TJ, Vernimmen D, Browne JA, Mullen MP, Gordon SV, MacHugh DE, et al. Alveolar macrophage chromatin is modified to orchestrate host response to *Mycobacterium bovis* infection. Front Genet. 2020;10:1386. https://doi.org/10.3389/fgene.2019.01386.

47.   Falkow S. I never met a microbe I didn't like. Nat Med. 2008;14(10):1053–7. https://doi.org/10.1038/nm1008-1053.

48. O'Doherty AM, Rue-Albrecht KC, Magee DA, Ahting S, Irwin RE, Hall TJ, et al. The bovine alveolar macrophage DNA methylome is resilient to infection with *Mycobacterium bovis*. Sci Rep. 2019;9(1):1510. https://doi.org/10.1038/s41598-018-37618-z.

49. Finlay EK, Berry DP, Wickham B, Gormley EP, Bradley DG. A genome wide association scan of bovine tuberculosis susceptibility in Holstein-Friesian dairy cattle. PLoS One. 2012;7(2):e30545. https://doi.org/10.1371/journal.pone.0030545.

50. Bermingham ML, Bishop SC, Woolliams JA, Pong-Wong R, Allen AR, McBride SH, et al. Genome-wide association study identifies novel loci associated with resistance to bovine tuberculosis. Heredity (Edinb). 2014;112(5):543–51. https://doi.org/10.1038/hdy.2013.137.

51. Richardson IW, Berry DP, Wiencko HL, Higgins IM, More SJ, McClure J, 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. 2016;48(1):19. https://doi.org/10.1186/s12711-016-0197-x.

52. Raphaka K, Matika O, Sanchez-Molano E, Mrode R, Coffey MP, Riggio V, et al. Genomic regions underlying susceptibility to bovine tuberculosis in Holstein-Friesian cattle. BMC Genet. 2017;18(1):27. https://doi.org/10.1186/SP1863-017-0493-7.

53. Wilkinson S, Bishop SC, Allen AR, McBride SH, Skuce RA, Bermingham M, et al. Fine-mapping host genetic variation underlying outcomes to *Mycobacterium bovis* infection in dairy cows. BMC Genomics. 2017;18(1):477. https://doi.org/10.1186/s12864-017-3836-x.

54. Tsairidou S, Allen AR, Pong-Wong R, McBride SH, Wright DM, Matika O, et al. An analysis of effects of heterozygosity in dairy cattle for bovine tuberculosis resistance. Anim Genet. 2018;49(2):103–9. https://doi.org/10.1111/age.12637.

55. Ring SC, Purfield DC, Good M, Breslin P, Ryan E, Blom A, et al. Variance components for bovine tuberculosis infection and multi-breed genome-wide association analysis using imputed whole genome sequence data. PLoS One. 2019;14(2):e0212067. https://doi.org/10.1371/journal.pone.0212067.

56. Karczewski KJ, Snyder MP. Integrative omics for health and disease. Nat Rev Genet. 2018;19(5):299–310. https://doi.org/10.1038/nrg.2018.4.

57. Cano-Gamez E, Trynka G. From GWAS to function: using functional genomics to identify the mechanisms underlying complex diseases. Front Genet. 2020;11:424. https://doi.org/10.3389/fgene.2020.00424.

58. Kramer A, Green J, Pollard J Jr, Tugendreich S. Causal analysis approaches in ingenuity pathway analysis. Bioinformatics. 2014;30(4):523–30. https://doi.org/10.1093/bioinformatics/btt703.

59. Qiagen: Qiagen Ingenuity Pathway Analysis Online Manual. 2019.

60. Foroushani AB, Brinkman FS, Lynn DJ. Pathway-GPS and SIGORA: identifying relevant pathways based on the over-representation of their gene-pair signatures. PeerJ. 2013;1:e229. https://doi.org/10.7717/peerj.229.

61. McKenzie AT, Katsyv I, Song WM, Wang M, Zhang B. DGCA: a comprehensive R package for differential gene correlation analysis. BMC Syst Biol. 2016;10(1):106. https://doi.org/10.1186/s12918-016-0349-1.

62. Breuer K, Foroushani AK, Laird MR, Chen C, Sribnaia A, Lo R, et al. InnateDB: systems biology of innate immunity and beyond—recent updates and continuing curation. Nucleic Acids Res. 2013;41(Database issue):D1228–33. https://doi.org/10.1093/nar/gks1147.

63. Liu G, Ma H, Jiang L, Zhao Y. Allograft inflammatory factor-1 and its immune regulation. Autoimmunity. 2007;40(2):95–102. https://doi.org/10.1080/08916930601083946.

64. Eskandari-Nasab E, Moghadampour M, Tahmasebi A, Asadi-Saghandi A. Interleukin-17 a and F gene polymorphisms affect the risk of tuberculosis: an updated meta-analysis. Indian J Tuberc. 2018;65(3):200–7. https://doi.org/10.1016/j.ijtb.2017.08.027.

65. Wang M, Xu G, Lü L, Xu K, Chen Y, Pan H, et al. Genetic polymorphisms of IL-17A, IL-17F, TLR4 and miR-146a in association with the risk of pulmonary tuberculosis. Sci Rep. 2016;6(1):28586. https://doi.org/10.1038/srep28586.

66. Brilha S, Wysoczanski R, Whittington AM, Friedland JS, Porter JC. Monocyte adhesion, migration, and extracellular matrix breakdown are regulated by integrin αVβ3 in *Mycobacterium tuberculosis* infection. J Immunol. 2017; 199(3):982–91. https://doi.org/10.4049/jimmunol.1700128.

67. Seyrantepe V, Iannello A, Liang F, Kanshin E, Jayanth P, Samarani S, et al. Regulation of phagocytosis in macrophages by neuraminidase 1. J Biol Chem. 2010;285(1):206–15. https://doi.org/10.1074/jbc.M109.055475.

68. Cheng Y, Huang C, Tsai HJ. Relationship of bovine TNF-α gene polymorphisms with the risk of bovine tuberculosis in Holstein cattle. J Vet Med Sci. 2016;78(5):727–32. https://doi.org/10.1292/jvms.15-0506.

69. Satproedprai N, Wichukchinda N, Suphankong S, Inunchot W, Kuntima T, Kumpeerasart S, et al. Diagnostic value of blood gene expression signatures in active tuberculosis in Thais: a pilot study. Genes Immun. 2015;16(4):253–60. https://doi.org/10.1038/gene.2015.4.

70. Steiger J, Stephan A, Inkeles MS, Realegeno S, Bruns H, Kröll P, et al. Imatinib triggers phagolysosome acidification and antimicrobial activity against *Mycobacterium bovis* Bacille Calmette-Guérin in glucocorticoid-treated human macrophages. J Immunol. 2016;197(1):222–32. https://doi.org/10.4049/jimmunol.1502407.

71. Gautier EL, Shay T, Miller J, Greter M, Jakubzick C, Ivanov S, et al. Gene-expression profiles and transcriptional regulatory pathways that underlie the identity and diversity of mouse tissue macrophages. Nat Immunol. 2012; 13(11):1118–28. https://doi.org/10.1038/ni.2419.

72. Lin J, Jiang Y, Liu D, Dai X, Wang M, Dai Y. Early secreted antigenic target of 6-kDa of *Mycobacterium tuberculosis* induces transition of macrophages into epithelioid macrophages by downregulating iNOS / NO-mediated H3K27 trimethylation in macrophages. Mol Immunol. 2020;117:189–200. https://doi.org/10.1016/j.molimm.2019.11.013.

73. Wu HB, Wang YS, Zhang Y, Yang MQ, Lv JX, Liu J. TALE nickase-mediated *SP110* knockin endows cattle with increased resistance to tuberculosis. Proc Natl Acad Sci U S A. 2015;112:E1530–9. https://doi.org/10.1073/pnas.1421587112.

74. Schubert C. Systems immunology: complexity captured. Nature. 2011; 473(7345):113–4. https://doi.org/10.1038/nj7345-113a.

75. Kidd BA, Peters LA, Schadt EE, Dudley JT. Unifying immunology with informatics and multiscale biology. Nat Immunol. 2014;15(2):118–27. https://doi.org/10.1038/ni.2787.

76. Vodovotz Y, Xia A, Read EL, Bassaganya-Riera J, Hafler DA, Sontag E, et al. Solving immunology? Trends Immunol. 2017;38(2):116–27. https://doi.org/10.1016/j.it.2016.11.006.

77. Hao S, Yan KK, Ding L, Qian C, Chi H, Yu J. Network approaches for dissecting the immune system. iScience. 2020;23(8):101354. https://doi.org/10.1016/j.isci.2020.101354.

78. Stelzer G, Rosen N, Plaschkes I, Zimmerman S, Twik M, Fishilevich S, et al. The GeneCards suite: from gene data mining to disease genome sequence analyses. Curr Protoc Bioinformatics. 2016;54:1.30.31–31.30.33. https://doi.org/10.1002/cpbi.5.

79. Brace PT, Tezera LB, Bielecka MK, Mellows T, Garay D, Tian S, et al. *Mycobacterium tuberculosis* subverts negative regulatory pathways in human macrophages to drive immunopathology. PLoS Pathog. 2017;13(6): e1006367. https://doi.org/10.1371/journal.ppat.1006367.

80. Cho JE, Park S, Lee H, Cho SN, Kim YS. *Mycobacterium tuberculosis*-induced expression of granulocyte-macrophage colony stimulating factor is mediated by PI3-K/MEK1/p38 MAPK signaling pathway. BMB Rep. 2013; 46(4):213–8. https://doi.org/10.5483/BMBRep.2013.46.4.200.

81. Maiti D, Bhattacharyya A, Basu J. Lipoarabinomannan from *Mycobacterium tuberculosis* promotes macrophage survival by phosphorylating bad through a phosphatidylinositol 3-kinase/Akt pathway. J Biol Chem. 2001; 276(1):329–33. https://doi.org/10.1074/jbc.M002650200.

82. Senthil Kumar A, Bansal K, Holla S, Verma-Kumar S, Sharma P, Balaji KN. ESAT-6 induced COX-2 expression involves coordinated interplay between PI3K and MAPK signaling. Mol Immunol. 2012;49(4):655–63. https://doi.org/10.1016/j.molimm.2011.11.011.

83. Yang Y, Sun Y, Xu J, Bao K, Luo M, Liu X, et al. Epithelial cells attenuate toll-like receptor-mediated inflammatory responses in monocyte-derived macrophage-like cells to *Mycobacterium tuberculosis* by modulating the PI3K/Akt/mTOR signaling pathway. Mediat Inflamm. 2018;2018:3685948. https://doi.org/10.1155/2018/3685948.

84. Liu Y, Li JY, Chen ST, Huang HR, Cai H. The rLrp of *Mycobacterium tuberculosis* inhibits proinflammatory cytokine production and downregulates APC function in mouse macrophages via a TLR2-mediated PI3K/Akt pathway activation-dependent mechanism. Cell Mol Immunol. 2016;13(6):729–46. https://doi.org/10.1038/cmi.2015.58.

85. Tsairidou S, Allen A, Banos G, Coffey M, Anacleto O, Byrne AW, et al. Can we breed cattle for lower bovine TB infectivity? Front Vet Sci. 2018;5:310. https://doi.org/10.3389/fvets.2018.00310.

86. Cai Z, Guldbrandtsen B, Lund MS, Sahana G. Prioritizing candidate genes post-GWAS using multiple sources of data for mastitis resistance in dairy cattle. BMC Genomics. 2018;19(1):656. https://doi.org/10.1186/s12864-018-5050-x.

87. Canovas A, Reverter A, DeAtley KL, Ashley RL, Colgrave ML, Fortes MR, et al. Multi-tissue omics analyses reveal molecular regulatory networks for

Hall *et al. BMC Genomics*   (2021) 22:343

Page 20 of 20

puberty in composite beef cattle. PLoS One. 2014;9(7):e102551. https://doi.org/10.1371/journal.pone.0102551.

88. Deng T, Liang A, Liang S, Ma X, Lu X, Duan A, et al. Integrative analysis of transcriptome and GWAS data to identify the hub genes associated with milk yield trait in buffalo. Front Genet. 2019;10:36. https://doi.org/10.3389/fgene.2019.00036.

89. Fang L, Sørensen P, Sahana G, Panitz F, Su G, Zhang S, et al. MicroRNA-guided prioritization of genome-wide association signals reveals the importance of microRNA-target gene networks for complex traits in cattle. Sci Rep. 2018;8(1):9345. https://doi.org/10.1038/s41598-018-27729-y.

90. Yan Z, Huang H, Freebern E, Santos DJA, Dai D, Si J, et al. Integrating RNA-Seq with GWAS reveals novel insights into the molecular mechanism underpinning ketosis in cattle. BMC Genomics. 2020;21(1):489. https://doi.org/10.1186/s12864-020-06909-z.

91. Fang L, Sahana G, Su G, Yu Y, Zhang S, Lund MS, et al. Integrating sequence-based GWAS and RNA-seq provides novel insights into the genetic basis of mastitis and milk production in dairy cattle. Sci Rep. 2017; 7(1):45560. https://doi.org/10.1038/srep45560.

92. Giuffra E, Tuggle CK. The FAANG consortium: functional annotation of animal genomes (FAANG): current achievements and roadmap. Annu Rev Anim Biosci. 2019;7(1):65–88. https://doi.org/10.1146/annurev-animal-020518-114913.

93. Bishop TF, Van Eenennaam AL. Genome editing approaches to augment livestock breeding programs. J Exp Biol. 2020;223(Pt Suppl 1):jeb207159. https://doi.org/10.1242/jeb.207159.

94. Banos G, Winters M, Mrode R, Mitchell AP, Bishop SC, Woolliams JA, et al. Genetic evaluation for bovine tuberculosis resistance in dairy cattle. J Dairy Sci. 2017;100(2):1272–81. https://doi.org/10.3168/jds.2016-11897.

95. Andrews S. FastQC: a quality control tool for high throughput sequence data. Cambridge: Babraham Research Campus: Bioinformatics Group, Babraham Institute; 2019.

96. Rosen BD, Bickhart DM, Schnabel RD, Koren S, Elsik CG, Tseng E, et al. *De novo* assembly of the cattle reference genome with single-molecule sequencing. Gigascience. 2020;9(3):giaa021. https://doi.org/10.1093/gigascience/giaa021.

97. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29(1):15–21. https://doi.org/10.1093/bioinformatics/bts635.

98. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014;30(7):923–30. https://doi.org/10.1093/bioinformatics/btt656.

99. R Core Team. R: a language and environment for statistical computing. Vienna: R Foundation for statistical Computing; 2019.

100. Carlson M. GO.db: A set of annotation maps describing the entire Gene Ontology. R package version 3.8.2; 2019.

101. Durinck S, Spellman PT, Birney E, Huber W. Mapping identifiers for the integration of genomic datasets with the R/bioconductor package biomaRt. Nat Protoc. 2009;4(8):1184–91. https://doi.org/10.1038/nprot.2009.97.

102. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550. https://doi.org/10.1186/s13059-014-0550-8.

103. Cook RD. Detection of influential observation in linear-regression. Technometrics. 1977;19(1):15–8. https://doi.org/10.2307/1268249.

104. Benjamini Y, Hochberg Y. Controlling the false discovery rate - a practical and powerful approach to multiple testing. J R Stat Soc Ser B Method. 1995; 57(1):289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.

105. Timmons JA, Szkop KJ, Gallagher IJ. Multiple sources of bias confound functional enrichment analysis of global -omics data. Genome Biol. 2015; 16(1):186. https://doi.org/10.1186/s13059-015-0761-7.

106. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13(11):2498–504. https://doi.org/10.1101/gr.1239303.

107. Song WM, Zhang B. Multiscale embedded gene co-expression network analysis. PLoS Comput Biol. 2015;11(11):e1004574. https://doi.org/10.1371/journal.pcbi.1004574.

108. van Dam S, Vosa U, van der Graaf A, Franke L, de Magalhaes JP. Gene co-expression analysis for functional classification and gene-disease predictions. Brief Bioinform. 2018;19(4):575–92. https://doi.org/10.1093/bib/bbw139.

109. Schaefer RJ, Michno JM, Myers CL. Unraveling gene function in agricultural species using gene co-expression networks. Biochimica et biophysica acta

Gene regulatory mechanisms. 2017;1860(1):53–63. https://doi.org/10.1016/j.bbagrm.2016.07.016.

110. Gormley C, Tong Z. Elasticsearch: the definitive guide. 1st ed. Sebastopol: O'Reilly Media, Inc; 2015.

111. Yates AD, Achuthan P, Akanni W, Allen J, Allen J, Alvarez-Jarreta J, et al. Ensembl 2020. Nucleic Acids Res. 2020;48(D1):D682–d688. https://doi.org/10.1093/nar/gkz966.

112. Ideker T, Ozier O, Schwikowski B, Siegel AF. Discovering regulatory and signalling circuits in molecular interaction networks. Bioinformatics. 2002; 18(Suppl 1):S233–40. https://doi.org/10.1093/bioinformatics/18.suppl_1.S233.

113. Strimmer K. Fdrtool: a versatile R package for estimating local and tail area-based false discovery rates. Bioinformatics. 2008;24(12):1461–2. https://doi.org/10.1093/bioinformatics/btn209.

114. Zimin AV, Delcher AL, Florea L, Kelley DR, Schatz MC, Puiu D, et al. A whole-genome assembly of the domestic cow, *Bos taurus*. Genome Biol. 2009; 10(4):R42. https://doi.org/10.1186/gb-2009-10-4-r42.

Publisher's Note