

Integrative genomics of human and bovine tuberculosis

Kate E. Killick^{1, 2}, Michael P. Mullen³, Thomas Hall¹, Nicolas C. Nalpas⁴, Ian W. Richardson⁵, David A. Magee¹, Kirsten E. McLoughlin¹, Carol N. Correia¹, Lucy Burkitt-Gray², Heather L. Wiencko⁶, John A Browne¹, Bernardo Villarreal-Ramos⁷, H. Martin Vordermeier⁷, Dan G. Bradley⁸, Vivek Naranbhai⁹, Adrian Hill⁹, Eamonn Gormley¹, Stephen V. Gordon^{1, 2} and David E. MacHugh^{1, 2}

¹ UCD College of Health and Agricultural Sciences, University College Dublin, Ireland; ² UCD Conway Institute of Biomolecular and Biomedical Research, University College Dublin, Ireland; ³ Athlone Institute of Technology, Athlone, Ireland; ⁴ Proteome Center, University of Tubingen, Germany; ⁵ Identigen Ltd., Blackrock Business Park, Dublin, Ireland; ⁶ Plusvital Ltd., Dun Laoghaire Business Park, Dublin, Ireland; ⁷ Animal and Plant Health Agency, Weybridge, United Kingdom; ⁸ Department of Genetics, Trinity College Dublin, Ireland. ⁹ The Wellcome Trust Centre for Human Genetics, University of Oxford, United Kingdom.

Email: kate.killick@ucd.ie

Project Overview

Human tuberculosis (TB), caused by *Mycobacterium tuberculosis*, is second only to HIV/AIDS as the greatest killer worldwide due to a single infectious agent. Bovine TB, caused by the closely related *Mycobacterium bovis* (99.95% genome sequence identity), is a major endemic disease affecting global cattle production, particularly in many developing countries. In the current study we used a network-based approach to integrate host gene expression data from human monocyte-derived macrophages (hMDM) and bovine alveolar macrophages (bAM) infected with *M. tuberculosis* and *M. bovis*, respectively. Single-nucleotide polymorphism (SNP) genome-wide association (GWA) data for bovine TB resistance/susceptibility was then prioritised using module information from the integrative network analyses to enhance detection of genomic variants for susceptibility/resistance to *M. bovis* infection.

Materials and Methods

Human monocyte-derived macrophage gene expression data

- RNA-seq data from human monocyte-derived macrophages (hMDM) challenged *in vitro* with *M. tuberculosis* (strain: H37Rv) and harvested 18 h post-infection [$n = 10$ and MOI = 2:1] (Blishak *et al.* 2015; DOI: [10.1038/srep16882](https://doi.org/10.1038/srep16882)).
- 4,937 DE genes were obtained for the contrast between hMDM challenged with *M. tuberculosis* and non-infected control hMDM at 18 h post-infection (Benjamini-Hochberg [B-H] FDR adjusted P -value ≤ 0.05).

Bovine alveolar macrophage gene expression data

- Post-mortem lungs were used to obtain bovine alveolar macrophage (bAM) samples from age-matched male Holstein-Friesian calves [$n = 10$] (Nalpas *et al.* 2015; DOI: [10.1038/srep13629](https://doi.org/10.1038/srep13629)).
- RNA-seq data from bAM challenged *in vitro* with *M. bovis* (strain AF2122/97) and harvested 24 h post-infection (MOI $\approx 10:1$).
- RNA-seq data from parallel non-infected control bAM obtained from the male calves described above and cultured for 24 h ($n = 10$).
- 5,515 differentially expressed (DE) genes were obtained for the contrast between bovine AM challenged with *M. bovis* and non-infected control bAM at 24 h post-infection (B-H FDR adjusted P -value ≤ 0.05).

Gene interaction network integration

- A base tuberculosis network was created using 982 genes identified in a GeneCards (www.genecards.org) search of human genes for the search terms: tuberculosis OR mycobacterium OR mycobacteria OR mycobacterial.
- A GeneCards Relevance Score cut-off of 2.0 was imposed: 213 genes retained.
- A large undirected interaction network for 213 key genes was generated using InnateDB (www.innatedb.com): 6,412 nodes (6,139 unique genes) and 14,679 edges (interactions).
- Networks were visualised in Cytoscape (cytoscape.org) and JActiveModules (apps.cytoscape.org/apps/jactivemodules) was used to extract functional modules with DE gene sets from hMDM and bAM transcriptomics experiments (stringent B-H FDR adjusted P -value ≤ 0.001).

Bovine genome-wide association study for bovine TB resistance/susceptibility

- High-density (Illumina® BovineHD) genome-wide association (GWA) SNP data (597,114 SNPs) from 841 Holstein-Friesian sires with estimated breeding values (EBVs) for susceptibility to *M. bovis* infection (Richardson *et al.* 2016; DOI: [10.1186/s12711-016-0197-x](https://doi.org/10.1186/s12711-016-0197-x)).
- SNPs from top five bAM modules (5 kb up- and downstream of each gene) were prioritised in the GWA single-SNP regression analysis.

Results

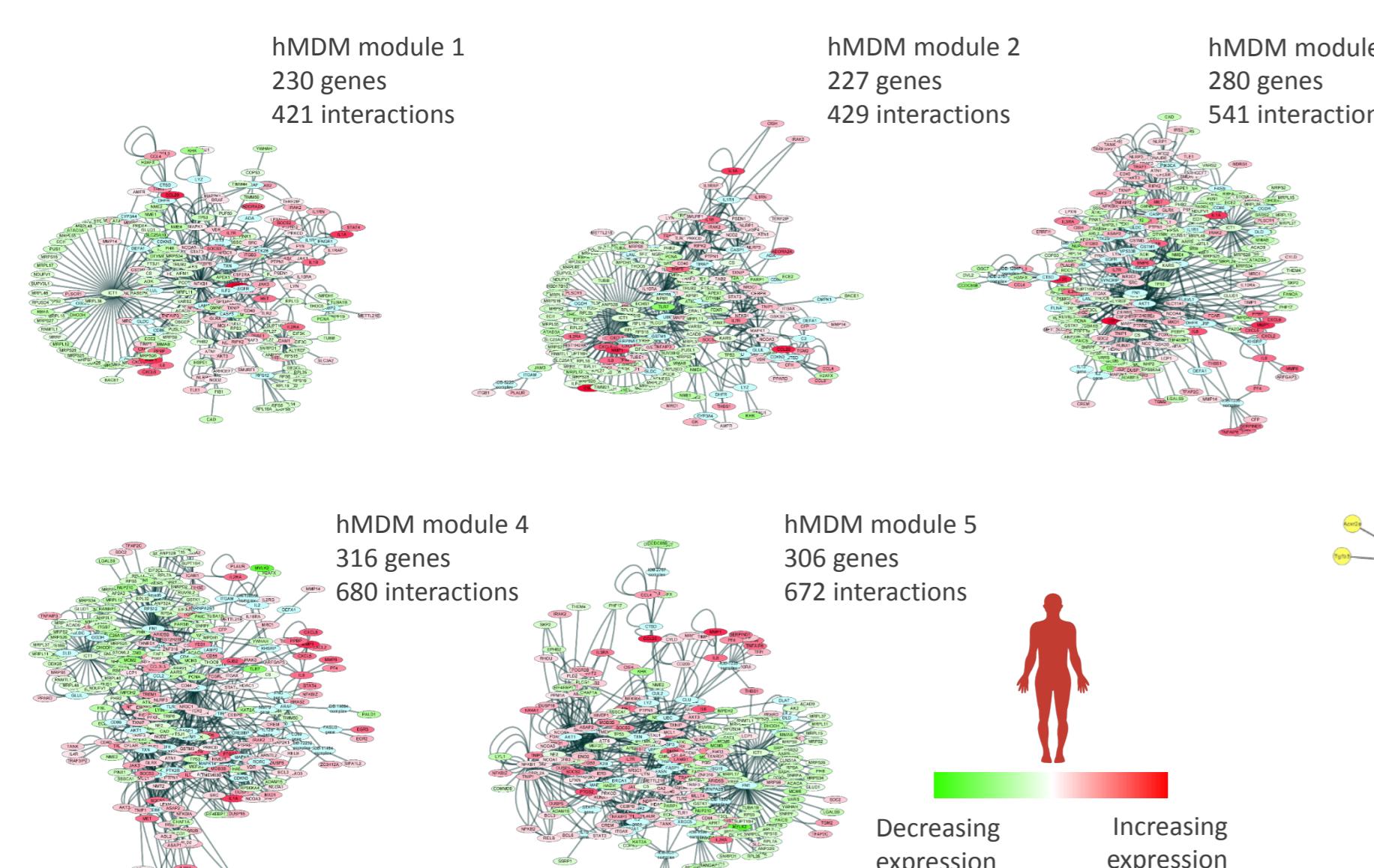


Fig. 1: Five top functional modules (18 h human MDM *M. tuberculosis* infection time point), corresponding to a union set of 484 unique genes.

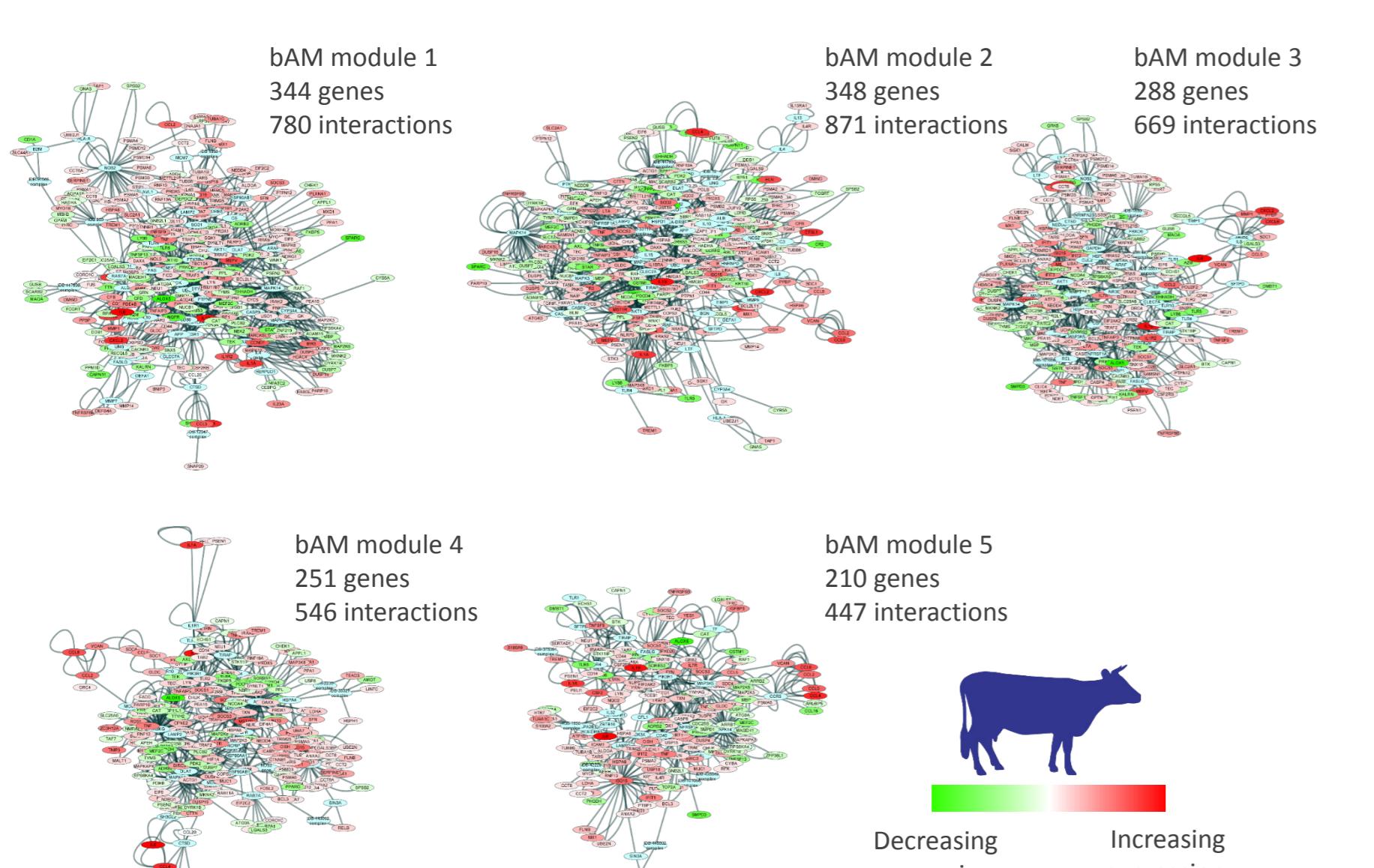


Fig. 2: Five top functional modules (24 h bovine AM *M. bovis* infection time point), corresponding to a union set of 516 unique genes.

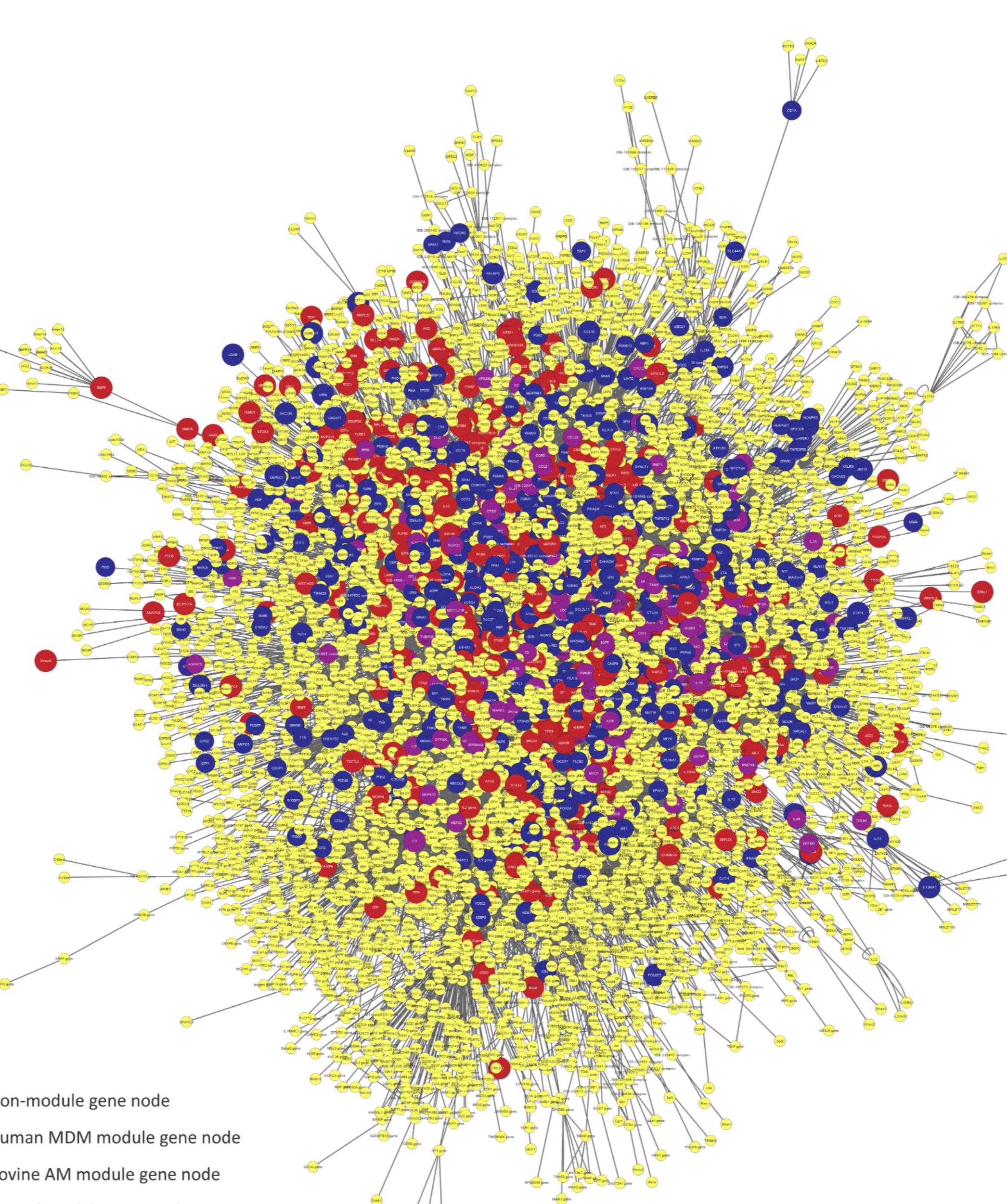


Fig. 3: Base tuberculosis network overlaid with functional module gene nodes enlarged and highlighted as shown in the legend. There are 364 gene nodes unique to hMDM modules, 396 unique to bAM modules and 120 shared across both human and bovine modules.

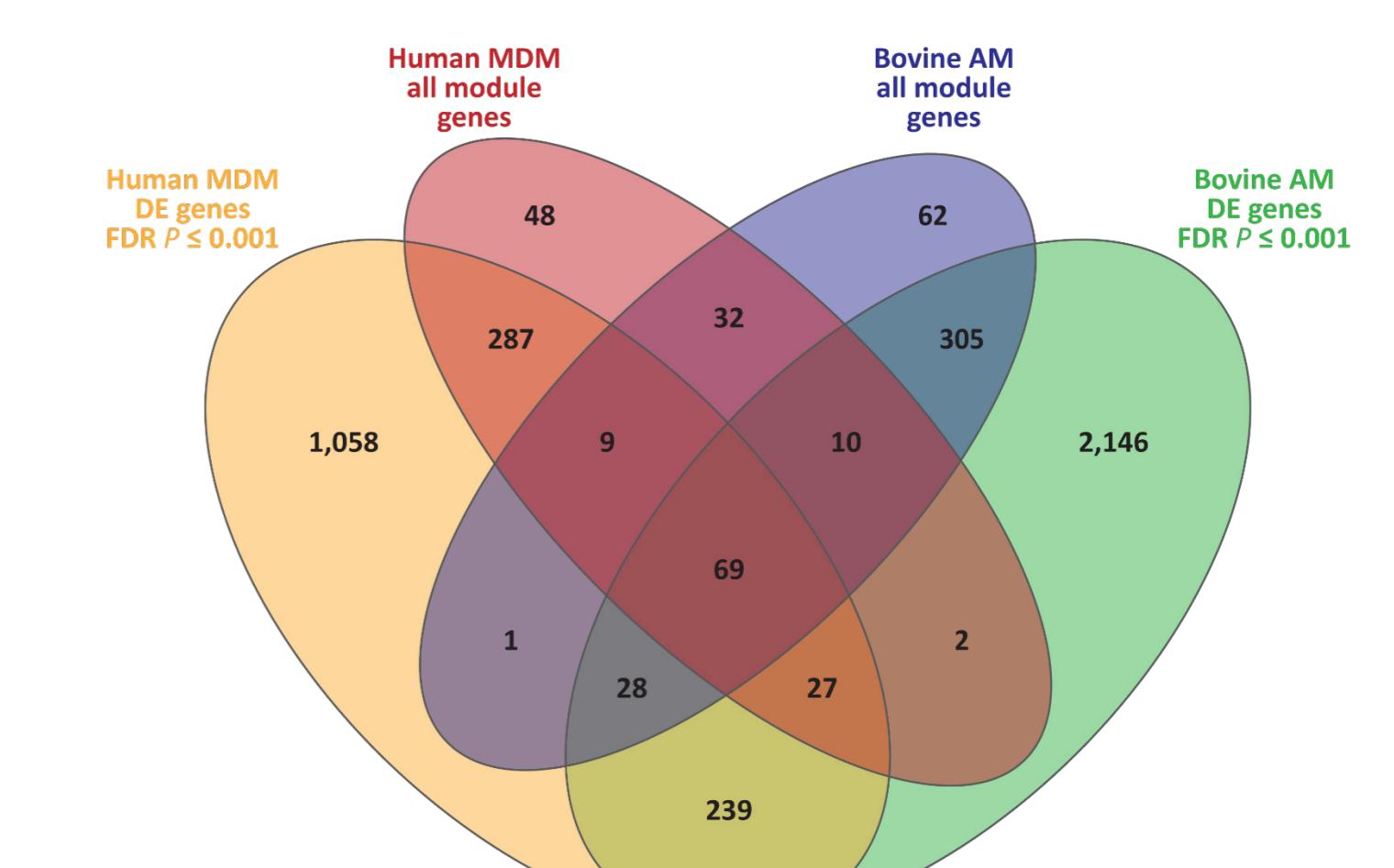


Fig. 4: Overlap of hMDM DE genes (1,718; FDR $P \leq 0.001$), bAM DE genes (2,826 FDR $P \leq 0.001$), all hMDM functional module genes (484) and all bAM functional module genes (516).

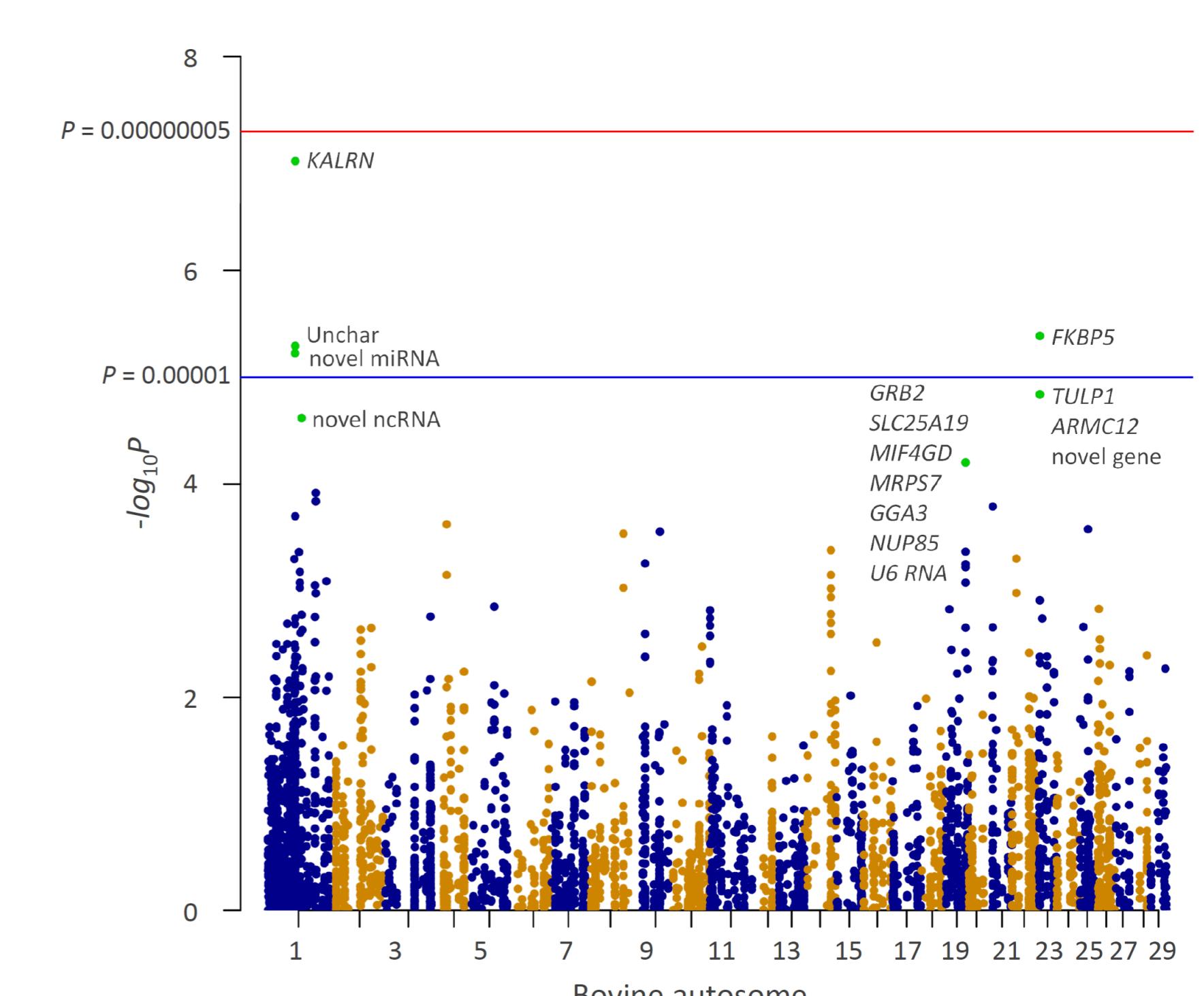


Fig. 5: Multi-module-wide association study (MMWAS) for 6,882 SNPs (516 genes) within five top modules (24 h bAM *M. bovis* infection time point). FDR q value < 0.05 .

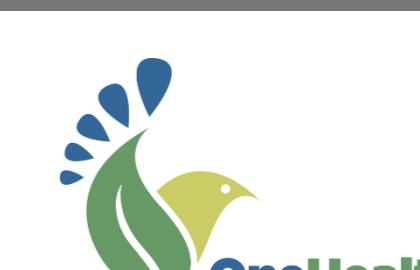
Future work and concluding remarks

- In vitro* macrophage infection challenge experiments using comparative transcriptomics is a powerful approach for dissecting host-pathogen interactions for human and bovine TB caused by infection with *M. tuberculosis* and *M. bovis*, respectively.
- Integrative gene network analyses can leverage information from large complementary molecular data sets to gain new biological knowledge.
- Integration of bovine TB transcriptomics and GWA data sets can provide functional information for genome-enabled breeding programmes in cattle.
- Gene network analysis outputs will be used to prioritise GWA SNP data for human TB susceptibility/resistance, including data from the Wellcome Trust Case Control Consortium (WTCCC – www.wtccc.org.uk) resource.
- GWA work will be extended using a Bayesian mixture model approach (BayesRC – MacLeod *et al.* 2016; DOI: [10.1186/s12864-016-2443-6](https://doi.org/10.1186/s12864-016-2443-6)) to assign biological priors to SNPs located in genes within functional modules identified using JActiveModules.
- Additional functional genomics data sets will be incorporated – e.g. peripheral blood transcriptomes from humans and cattle infected with *M. tuberculosis* and *M. bovis*, respectively.

Funding and support



Supported by
wellcome trust



Supported by
CAPES

