Inflammation and immune activation are associated with risk of Mycobacterium tuberculosis infection in BCG-vaccinated infants

Iman Satti
   The Jenner Institute, University of Oxford

Rachel Wittenberg
   University of Oxford  https://orcid.org/0000-0002-0823-5635

Shuailin Li
   Oxford University

Stephanie Harris
   Oxford University

Rachel Tanner
   University of Oxford

Deniz Cizmeci
   Oxford University  https://orcid.org/0000-0003-3231-7726

Ashley Jacobs
   Oxford University

Nicola Williams

Humphrey Mulenga
   University of Cape Town

Helen Fletcher
   London School of Tropical Hygiene and Medicine  https://orcid.org/0000-0002-0435-1006

Thomas Scriba
   University of Cape Town  https://orcid.org/0000-0002-0641-1359

Michele Tameris
   South African Tuberculosis Vaccine Initiative, University of Cape Town  https://orcid.org/0000-0002-7983-7203

Mark Hatherill
   South African Tuberculosis Vaccine Initiative, University of Cape Town

Helen McShane (✉ helen.mcshane@ndm.ox.ac.uk)
   Oxford University  https://orcid.org/0000-0002-2126-5142
Abstract

Tuberculosis vaccine development is hindered by the lack of validated immune correlates of protection. Exploring immune correlates of risk of disease and/or infection in prospective samples can inform this field. We investigated whether previously identified immune correlates of risk of TB disease also associate with increased risk of *M. tb* infection in BCG-vaccinated South African infants, who became infected with *M. tb* during 2–3 years of follow-up. *M. tb* infection was defined by conversion to positive reactivity in the QuantiFERON test. We demonstrate that inflammation and immune activation are associated with risk of *M. tb* infection. Ag85A-specific IgG was elevated in infants that were subsequently infected with *M. tb*, and this was coupled with upregulated gene expression of immunoglobulin-associated genes and type-I interferon. Plasma levels of IFN-α2, TNF-α, CXCL10 (IP-10) and complement C2 were also higher in infants that were subsequently infected with *M. tb*.

Introduction:

Tuberculosis (TB), caused by infection with *Mycobacterium tuberculosis* (*M. tb*), is a global health threat that, until the COVID-19 pandemic, killed more people annually than any other infectious pathogen. In 2020, 9.9 million new cases and 1.5 million deaths were due to TB [1]. In most *M. tb* infected individuals, infection is either eliminated or contained by the host immune response resulting in a status of asymptomatic infection [2, 3]. Bacille Calmette-Guérin (BCG), the only licenced vaccine against TB, does not adequately protect against pulmonary disease, with especially low efficacy in TB endemic areas; hence, a new and improved vaccine is urgently needed [4].

Development of an effective and efficacious TB vaccine would be facilitated by an improved understanding of protective immunity. Although many aspects of host immunity are known to play a role in protection against TB, defined and validated correlates of protection are lacking. The identification of immune correlates of protection in human studies would allow for the rational design and development of TB vaccine candidates, and antigen delivery systems could be optimised to induce these immune correlates. Validation of immune correlates in field efficacy trials would ultimately enable the development and licensure of vaccines without the need for expensive, time consuming field efficacy studies.

In the absence of sample sets from trials of effective TB vaccines, where correlates of protection can be identified, exploring immune correlates of risk in prospective samples can inform this field. Although MVA85A did not confer improved efficacy over BCG vaccination alone, samples collected from BCG-vaccinated South African infants who participated in the MVA85A efficacy trial provide a good opportunity for studying correlates of risk of TB disease and *M. tb* infection in this important target population [5]. We have previously shown, in a case-control study investigating immune correlates of risk of TB disease in the same cohort of South African infants, that lower BCG-specific ex-vivo IFN-γ responses, lower Ag85A-specific IgG antibodies and higher non-antigen specific CD4+ T-cell activation were all associated with an increased risk of TB disease [6]. We have identified transcriptomic signatures
that correlated with progression from *M. tb* infection to TB disease in South African adolescents [7]. Measurement of these signatures has been translated to microfluidic RT-qPCR, allowing more efficient application to further studies [7]. Persistent viral or bacterial infections can result in T-cell activation and dysfunction of antigen-specific T-cells [8–10]. We have previously studied the association between human cytomegalovirus (CMV) infection and T-cell activation using samples from the same cohort of South African infants. We found that a CMV-specific IFN-γ response was associated with CD8 + T-cell activation and with increased risk of subsequent TB disease [11]. Similarly, CMV antibody responses were associated with increased risk of TB in Ugandan adults [12]. CMV infection acquired in the first year of life is associated with increased risk of TB disease in South African infants [13]. Respiratory viral infections have also been shown to be associated with increased risk of TB progression and to drive conversion to a positive TB disease gene signature [14].

Given the time and expense required to evaluate candidate TB vaccines in Prevention of Disease (PoD) efficacy trials, there has been a recent shift of focus to Prevention of *M. tb* Infection (PoI) studies, as a way of evaluating the efficacy of candidate TB vaccines in humans prior to PoD efficacy trials. As only ~5–10% of infected individuals progress to active disease during their lifetime, the incidence of *M. tb* infection is higher than the incidence of TB disease. PoI studies therefore have more endpoints and are thus shorter and more cost-effective than PoD studies [10]. A possible assumption implicit in this approach is that the same immunological mechanisms confer protection against both TB disease and *M. tb* infection. It is important to determine the validity of this assumption.

In our current study, we examined samples collected from South African infants who participated in the MVA85A efficacy trial [5] to determine whether the correlates of risk of TB disease we identified in this population [6] are also correlates of risk of *M. tb* infection. We characterised baseline immune responses, prior to acquisition of *M. tb* infection, in two groups of BCG-vaccinated 4–6 months old infants: infants who went on to become infected with *M. tb* during a follow-up period of 2–3 years, but who did not develop disease, and infants who remained uninfected throughout the study, referred to as *M. tb*-infected and *M. tb*-uninfected, respectively. *M. tb* infection was defined using the QuantiFERON (QFT) TB Gold In-tube test, which has been used extensively in TB endemic and non-endemic countries to define *M. tb* infection [15].

In this study we show that factors associated with increased risk of TB disease in South African infants [6] are not associated with increased risk of *M. tb* infection in this cohort. However, we identified inflammation and activation immune markers as well as gene transcripts that are associated with susceptibility to *M. tb* infection in South African infants.

**Materials And Methods:**

**Study design and volunteers:**

Samples for this study were collected from 3-6 month old, HIV-negative South African infants who received BCG within 7 days of birth and participated in the MVA85A efficacy trial (ClinicalTrials.gov
number NCT00953927, [5]). Infants without known TB exposure and who tested negative with QFT test were randomized to receive a single intradermal dose of MVA85A or placebo. The trial was approved by the University of Cape Town Faculty of Health Sciences Human Research Ethics Committee, Oxford University Tropical Research Ethics Committee and the Medicines Control Council of South Africa.

Informed consent was obtained from the mothers of all infants [15]. *M. tb* infection was defined as QFT test conversion from negative to positive at the manufacturer's threshold of 0.35 IU/mL IFN-g. Control infants were those who remained QFT negative and were included in the study at a ratio of 3 *M. tb*-uninfected to 1 *M. tb*-infected. Controls were matched based on sex, ethnic group, Centre for Disease Control (CDC) weight-for-age percentile and time on study. Infants who lived in a household with a smear-positive TB case or received Isoniazid Preventive Therapy (IPT) were excluded from this study, and none of the infants included in this study had active TB disease during the study follow-up period.

Blood samples, collected at baseline, from 43 infants that subsequently met the criteria for *M. tb* infection, according to the study protocol, were used in this study [5]; 17/43 infants were from the group that received MVA85A (Figure 1). Samples for this study were mainly from Day -7 of the study (7 days before BCG-vaccinated infants were randomized to receive vaccination with either MVA85A or placebo). Fourteen infants did not have Day -7 PBMC samples and in these infants Day 28 post-vaccination samples were used; six of these infants were in the MVA85A group and eight were in the placebo group. Seventeen infants did not have Day -7 plasma samples, hence Day 28 post-vaccination samples from these infants were used instead. Eight of these infants were in the MVA85A group and nine were in the placebo group. Blood samples from all 127 matched controls were from Day -7 of the study. 69/127 were from infants that received MVA85A.

**Sample processing:**

Immune assays were performed on frozen PBMC that were thawed and rested for 2 hours in a 5% CO₂ incubator at 37°C as previously described [6].

Samples were processed in 6 batches and the assays performed included, in priority order: flow cytometry surface staining for characterisation of T-cells, Enzyme-Linked ImmunoSpot (ELISpot) assay, flow cytometry surface staining for characterisation of B-cells, Mycobacteria Growth Inhibition Assay (MGIA), gene expression and flow cytometry surface staining for characterisation of mucosal-associated invariant T (MAIT) cells. The laboratory researchers performing the assays were blinded to the group allocation. Multiplex Assay and ELISA (IgG against Ag85A and whole cell lysate) were performed on all available plasma samples. Supplementary Table 1 summarises the assays performed and the percentage of samples that had all assays and conditions performed in each batch. Immune assays were performed in line with those used in the previous correlates of risk of TB disease study [6,16].
Gene expression:

Cell stimulation:

Cells were stimulated in RMPI with 10% Foetal calf serum (FCS) at 5 $10^6$/ml with 1 $10^6$ Colony Forming Units (CFU)/ml BCG (SII) or left unstimulated in a total volume of 200 µl/well in 96-well plates. Cells were incubated for 12 hours (37˚C, 5% CO$_2$). Plates were centrifuged and 200 µl supernatant was removed without disrupting the pellet and transferred into clean U-bottom plates. The plates were centrifuged again, the supernatant flicked-out and the cells resuspended in 200 µl RLT buffer with -mercaptoethanol (10 µl/ml). The samples were mixed with the buffer to lyse the cells and the plates were sealed and stored at -20˚C.

At the time of RNA extraction, plates were thawed at 37˚C for 10 minutes. RNA extraction was performed using the Qiagen RNeasy kit following the manufacturer's instructions (including optional DNase digestion) and RNA was eluted twice, 30 µl per elution. RNA was quantified by nanodrop and samples were stored at -80˚C.

RNA-Seq:

Total RNA quantity and integrity were assessed using Quant-IT RiboGreen RNA Assay Kit (Invitrogen, Carlsbad, CA, USA) and Agilent Tapestation4200. Samples were normalised to 70ng input. Purification of mRNA, generation of double stranded cDNA and library construction were performed using NEBNext Poly(A) mRNA Magnetic Isolation Module (E7490) and NEBNext Ultra II Directional RNA Library Prep Kit for Illumina (E7760L) with adapters and barcode tags (dual indexing) [17]. The concentrations used to generate the multiplex pool were determined by Picogreen. The final size distribution of the pool was determined using a Tapestation system (Agilent), and quantification was determined by Qubit (Thermofisher) before sequencing on an Illumina HiSeq 4000 as 75bp paired end.

Gene expression data analysis:

Quality control checks on raw sequence data were performed using fastqc. Reads were aligned to the genome using HISAT2. Sequences were quantified as gene counts using the featureCounts tool. A linear model was fitted using DESeq2 (v1.22.2) to determine differentially expressed genes (DEGs) adjusted for matching group and stimulus. Genes with an adjusted $p$ value of less than 0.2 and a fold change of more than 5% were considered DEG.

Pathway analysis was performed using the Coincident Extreme Ranks in Numerical Observations (CERNO) algorithm from tmod package (v0.40) in R. Modules used for enrichment analysis were adapted
from Li et al, 2014 [18]. Modules with an adjusted $p$ value of less than 0.001 were considered enriched in the analysis.

GO enrichment analysis was performed using the clusterProfile package (v3.10.1) in R [19]. DEGs were grouped into upregulated genes and downregulated genes, and then enriched separately. We used the get_child_nodes function in GOfuncR package (v1.2.0) in R to obtain all child GO terms of immune system process (GO:0002376) and cytokine production (GO:0001816). Then, we removed all enriched GO terms that are not child GO terms of immune system process (GO:0002376) and cytokine production (GO:0001816). If the number of remaining GO terms of a group was larger than 50, we only kept the most significant terms among terms with similar semantics (cut-off = 0.7) using the simplify function in the clusterProfile package, which is measured by a similarity measure [20].

**Enzyme-Linked ImmunoSpot (ELISpot) assay:**

IFN-γ ELISpot was performed as previously described [6]. Briefly, 3 $10^5$ PBMC/well were stimulated in duplicate with each of 2 $10^5$ CFU/ml BCG SII, 2 g/ml Cytomegalovirus (CMV) peptides and 2 g/ml Epstein-Barr Virus (EBV) peptides (both from AnaSpec; CMV=5 peptides (4 from pp65, 1 from IE1) and EBV=15 peptides), 20 g/ml purified protein derivative (PPD) (AJ Vaccines), 10 g/ml phytohaemagglutinin (PHA) (SIGMA) or unstimulated negative control. Antigen specific responses are presented as Spot Forming Cells (SFC)/1 $10^6$ PBMC, calculated by subtracting the mean of the unstimulated responses from the mean of the antigen responses and correcting for the number of PBMC per well. A response was considered positive if the mean number of spots in the antigen well was at least twice the mean of the unstimulated wells and at least 5 spots or greater.

**Enzyme-Linked Immunosorbent (ELISA) assay:**

Ag85A-specific IgG was measured as previously described [6]. ELISA plates were coated with 2 g/ml rAg85A (BEI resources) and each sample was assayed performed in triplicate. Results are presented as mean OD values with the mean OD values of the blank wells (without plasma) subtracted.

**CMV reactivity:**

Evaluation of CMV infection was performed by detection of a cellular immune response to CMV peptides by ELISpot and by detection of CMV-specific IgG and IgM responses at day -7 and day 28, using commercial kits (Human Anti-Cytomegalovirus IgG ELISA Kit, Abcam) and following the manufacturer’s instructions. The absorbance of samples was transformed to standard units according to the manufacturer’s recommendation. Samples with standard units greater than 11 were considered positive,
whereas those with standard units of less than 9 were considered negative. The assay was repeated for samples with standard units of between 9 and 11, which were considered inconclusive. If the results of the second test were less than 11, then the samples were considered negative. Otherwise, they were considered positive. CMV-specific IgG responses were only evaluated in the infants from whom both day -7 and day 28 samples were available. For CMV-specific IgM responses, we initially tested day 28 samples, and tested day -7 samples where day 28 samples were not available.

Flow cytometry surface staining:

Surface staining of PBMC was performed as previously described [6,16]. Cells were stained with antibodies to characterize T, B and MAIT-cells (Supplementary Table 2). Cells were acquired on an LSRll (BD Biosciences) and data analysed using Flowjo v8.8 (BD Biosciences). Dead and doublet cells were excluded from the analysis. Gating strategies for the 3 panels are illustrated in Supplementary Figure 1.

Mycobacterial growth inhibition assay (MGIA):

PBMC (3 $10^6$) were incubated with 500 CFU BCG Pasteur and 120 μl autologous serum in a total volume of 480 μl RPMI/well in a 48-well tissue culture plate. Cultures were incubated at 37°C for 96 hours. Cells were then transferred to 2 ml screw-cap tubes and centrifuged at 12000 rpm for 10 minutes followed by removal of the supernatant. 500 μl sterile water was added to the wells to lyse adherent monocytes and release intracellular BCG, and was then added to the corresponding pellet in the 2ml tubes. The 2ml tubes were pulse-vortexed and the lysate was transferred to BACTEC MGIT tubes supplemented with BBL MGIT OADC and PANTA (Becton Dickinson, UK). MGIT tubes were placed on the BACTEC 960 machine (Becton Dickinson, UK) and incubated at 37°C until the detection of positivity by fluorescence. The time to positivity (TTP) read-out was converted to log$_{10}$ CFU using stock standard curves of TTP against inoculum volume and CFU. Control tubes were set up at day 0 by inoculating supplemented BACTEC MGIT tubes with the same volume of mycobacteria as the samples. Results are presented as normalised mycobacterial growth which is equal to (log$_{10}$ CFU of sample − log$_{10}$ CFU of growth control).

Multiplex:

A bead-based multiplex assay to analyse a range of 38 Th1 and Th2 cytokines, chemokines and complement factors was performed on plasma samples. The MILLIPLEX MAP Human Cytokine/Chemokine Magnetic Bead Panel was used following the manufacturer's recommendations. Details of the measured cytokines and chemokines are included in Supplementary Table 1A.
Statistical analysis:

A statistical analysis plan was prepared prior to data analysis in which the methods of analyses were specified and the primary and secondary analyses were defined (Supplementary document 1). The primary objective of this study was to identify correlates of risk of *M. tb* infection in BCG-vaccinated, HIV-uninfected infants and in particular, to determine whether the three markers that were associated with risk of *M. tb* disease were also associated with risk of *M. tb* infection [6]. Accordingly, the primary objective was to assess the ability of a set of three selected variables, activated CD4+ T-cells, BCG-specific IFN-γ and Ag85A-specific IgG to predict subsequent *M. tb* infection during study follow-up in a set of 43 *M. tb*-infected infants with 3 matched *M. tb*-uninfected controls per *M. tb*-infected infant. An exploratory objective was to identify a secondary set of variables measuring the immune response at Study Day -7 or Day 28 that best classified participants in terms of their risk of *M. tb* infection; these variables are shown in Supplementary Table 3. A series of univariate conditional logistic regression analyses were performed for each assay to assess the association between Day -7 or Day 28 immune response and risk of *M. tb* infection. In these analyses, the occurrence of *M. tb* infection was the outcome variable with the matched case-control set number as the strata variable and the Day -7 or Day 28 immune response as the independent variable. Results were summarized as a list of all immune response variables together with their estimated odds ratios, 95% confidence intervals, and two-sided p-value. Benjamini-Hochberg adjusted p-values (False Discovery Rate, FDR) are presented and an FDR of less than 0.20 was considered evidence of a significant association. Furthermore, the area under the receiver operating characteristic (AUROC) curve was calculated for each immune response. Immune responses with an AUROC of between 0.80 and 0.90 were considered to be 'good' discriminators of infants who were at risk of going on to become *M. tb* infected, while responses with an AUROC of greater than 0.90 were to be considered 'excellent' discriminators. All adjusted p-values shown in this paper are adjusted by Benjamini-Hochberg multiple testing correction.

Results:

QuantiFERON-TB Gold In-Tube conversion:

*M. tb* infection was assessed by measuring mycobacteria-specific IFN-γ production using the QFT test, performed prior to enrolment, at day 336 and at the end of study. All infants had a negative QFT response at enrolment. 16/43 had a positive QFT ≥ 0.35 IU/ml on day 336 but only two of these remained QFT positive at end of study; they had QFT values of 2.54 IU/ml and 0.82 IU/ml on day 336 and 5.05 IU/ml and 1.33 IU/ml at end of study, respectively. 22/43 infants who were QFT negative at day 336 had QFT values ≥ 0.35 IU/ml at the end of study. 5/43 infants converted to positive QFT at time points other than day 336 and end of study (Figure 2).

Ag85A-specific IgG response on Day -7 was associated with increased risk of *M. tb* infection
Baseline characteristics used to investigate possible associations between the studied immune parameters and *M. tb* infection are presented for *M. tb*-infected and uninfected infants (Supplementary Table 4). The number analysed was dependent on the variables used in the model and how many samples were missing for each of the included variables. The primary aim of this study was to investigate if the three immune correlates of risk of TB disease identified in our previous study in the same population were similarly associated with risk of *M. tb* infection [6]. We evaluated HLA-DR expression on CD4+ T-cells, frequencies of BCG-specific IFN-γ producing cells in PBMC and plasma Ag85A-specific IgG responses. HLA-DR expression on CD4+ T-cells (OR 1, 95% CI 0.95 – 1.06, p = 0.921, FDR = 0.921, adjusted only for the three immune correlates of TB disease, same for the remaining two parameters) and BCG-specific IFN-γ producing cells (OR 0.99, 95% CI 0.97 – 1.01, p = 0.238, FDR = 0.357) were not associated with risk of *M. tb* infection, whereas Ag85A-specific IgG response on Day -7 was associated with increased risk of *M. tb* infection (OR 8.85, 95% CI 2.05 – 38.09, p = 0.003, FDR = 0.009).

**Systematic upregulation of cytokines, chemokines and complement component factors in infants with subsequent *M. tb* infection:**

As an exploratory analysis, we quantified a number of cytokines, chemokines, and complement factors in plasma samples collected from the study subjects in a multiplex assay. The statistical significance of these immune parameters was much higher than those measured by other assays including ELISpot, flow cytometry and MGIA (Figure 3A, Supplementary Table 3). Almost all of the immune parameters measured by the multiplex assay had an odds ratio >1, indicating that the increased values of these immune parameters were associated with an increased risk of *M. tb* infection (Figure 3A, Supplementary Table 3). Given that the immune parameters measured by the multiplex assay were closely correlated with each other (Figure 3B), it is unsurprising that almost all of them have a consistent odds ratio. In addition, the AUROC values of the immune parameters measured by the multiplex assay were also higher than those measured by other assays (Figure 3C, Supplementary Table 3). Therefore, it is likely that multiple testing correction with other immune parameters caused the loss of statistical power of the immune parameters measured by the multiplex assay. On the basis of these observations, we performed multiple testing correction within the immune parameters measured by the multiplex assay. The results showed that 26 out of the 45 available immune parameters measured by the multiplex assay had an FDR of no greater than 0.2, indicating that systematic upregulation of chemokines, cytokines and complement components in the plasma was associated with the increased risk of *M. tb* infection (Supplementary Table 5).

Approximately one third of the blood samples from infants who went on to be infected with *M. tb* were collected on Day 28, while all samples from *M. tb*-uninfected infants were collected on Day -7. To investigate the possible confounding effect of the sample collection time points on the univariate conditional logistic regression, we compared data from Day -7 samples only with combined Day -7 and
Day 28 data from infants who went on to be infected with *M. tb*. We selected parameters that are different between Day 28 and Day -7 samples from infants who would be infected with *M. tb* (unadjusted p-values less than 0.1, Supplementary Table 6A). The odds ratios, calculated by univariate conditional logistic regression, of some of these parameters using Day -7 samples only changed qualitatively (namely from less than 1 to larger than 1 or vice versa) compared to those calculated using all samples (combined Day -7 and Day 28) (Supplementary Figure 2A and Supplementary Table 6B). Of the immune parameters which were significantly different between *M. tb*-infected and *M. tb*-uninfected infants (Supplementary Table 5), only Complement Factor I and Complement C5 differed when we looked at data from Day -7 only compared to Day 28 and Day -7 data combined (Day 28 vs Day -7, unadjusted p-value less than 0.1) (Supplementary Figure 2B and Supplementary Table 6B). Taken together, the results suggested that the significant change of 24/26 cytokines, chemokines and complements were not confounded by the time point of sample collection.

**CMV infection is not associated with increased risk of *M. tb* infection:**

Cellular and humoral immune responses to CMV were measured by IFN-ELISpot and IgG and IgM ELISA, respectively. We used the manufacturer’s recommendation to determine the cut-off for a positive CMV-specific IgM response. For the CMV-specific IgG response, considering that IgG antibodies can be vertically transmitted from mothers to babies [21], we introduced a more stringent cut-off criteria: only infants who were IgG positive (according to the manufacturer’s cut-off) on day 28 and maintained ≥ 90% of their day -7 IgG responses on day 28 were considered CMV-specific IgG positive (Figure 4A). Infants who had a CMV-specific IFN- response of >17 SFC/10^6 PBMC were considered positive, based on the cut-off used in our previous study on samples collected from the same population [11] (Figure 4B). All infants who had a positive CMV-specific IgM response and all with positive CMV-specific IFN- responses, except one infant who maintained 85% rather than 90% of their day -7 CMV-IgG response on day 28, had a positive CMV-specific IgG response, indicating that measuring IgG is more sensitive in detection of CMV infection (Figure 4C). Therefore, we defined the CMV infection status of infants as follows: infants were defined as CMV-uninfected if they had a negative CMV-specific IgG response; infants were defined as CMV-infected if they had either a positive IgG response, or IgM response or IFN- response; if the infant had negative IgM and IFN- responses with an unknown IgG response because of sample unavailability, they were considered ‘unknown’ (Figure 4C). Using this definition, *M. tb* infection incidence (based on QFT conversion) was 33.9% (21/62) in the CMV-infected group, compared to 16.7% (6/36) in the CMV-uninfected group (p = 0.0699, Conditional logistic regression (OR 2.607, 95% CI 0.9252-7.344), not corrected for multiple testing) (Figure 4D).

Although we did not detect a direct association between the 3 immune correlates of risk of TB disease and risk of *M. tb* infection, we detected increased frequencies of HLA-DR+ CD4+ T-cells and HLA-DR+...
CD8+ T-cells and decreased levels of BCG-specific IFN-γ-producing cells in CMV-infected infants (Supplementary Figure 3 and Supplementary Table 7).

We evaluated the association between immune parameters measured by the multiplex assay and CMV infection and found that IP-10, TNF-β and complement C2 were all significantly (FDR  0.2) upregulated in CMV-infected infants, compared with CMV-uninfected infants (Supplementary Figure 3; Supplementary Table 7). The plasma levels of these three proteins were also significantly upregulated in infants who became infected with *M. tb* (FDR  0.2, Supplementary Table 5).

**Differential gene expression between infants with subsequent *M. tb* infection and uninfected infants (controls):**

Gene expression analysis was performed in samples from 21 *M. tb*-infected who had QFT values ≥ 0.35 IU and 52 matched controls. Day 28 samples were used for 2 out of the 21 *M. tb*-infected because of lack of samples. When cell availability allowed, each infant was represented by a set of 2 samples: BCG-stimulated and unstimulated. A linear model was fitted to determine differential gene expression using the matched case-control set number as the strata variable and adjusted for the stimulus (BCG stimulated or unstimulated), given that BCG stimulation had a strong influence on gene expression. After differential gene expression analysis (Supplementary Figure 4A, Supplementary Table 8A), we performed gene set enrichment analysis using two methods and gene set annotation. The first method was the CERNO algorithm using gene sets adapted from Li *et al*, 2014 [18]. The second method was overrepresentation analysis using gene sets defined in Gene Ontology (GO).

The number of infants who became infected with *M. tb* was higher in the CMV-infected group, though this did not reach statistical significance. We therefore performed differential gene expression analysis in CMV-infected and CMV-uninfected infants separately (Supplementary Figure 4B-C, Supplementary Table 8B-C). We also identified transcripts differentially expressed between CMV-infected and CMV-uninfected infants (Supplementary Figure 4D, Supplementary Table 8D).

When we analysed combined data from CMV-infected and CMV-uninfected infants, gene sets associated with B-cells, immunoglobulins, complement activation, type I interferon response and anti-viral response were upregulated in infants who were subsequently infected with *M. tb*. These included genes such as IGKV4-1, TNFRSF17, IFIH1 and IFIT2 (Figure 5A, Supplementary Figure 5A, Supplementary Table 8A, 9-10).
In CMV-infected infants, gene sets associated with anti-viral response, type-I interferon response, B-cells, immunoglobulins, complement activation, neutrophil activation and cytokine production and secretion were upregulated in infants who were subsequently infected with \textit{M.tb} compared with those who remained \textit{M.tb}-uninfected, these included genes such as DDX58, IFIH1, IFI35, CD19, IGKV4-1, C5, C3AR1, FUCA2, NCF4 and IL-10 (Figure 5A, Supplementary Figure 5B, Supplementary Table 8B, 9-10).

In CMV-uninfected infants, gene sets related to complement activation, B-cells and immunoglobulins were upregulated in infants that were subsequently infected with \textit{M.tb}, these included genes such as C1QBP, CR2, C4BPB, CD19 and IGKV4-1 (Figure 5A, Supplementary Figure 5C, Supplementary Table 8-10). Gene sets related to neutrophil activation, natural killer cell activation, complement activation were downregulated in infants who were subsequently infected with \textit{M.tb}, these included genes such as C1QA, C1QB, C1QC, C3, C3AR1, C5AR1, C5AR2, FUCA2, KLRB1, KLRD1, KLRC3 and NKG7 (Figure 5A, Supplementary Figure 5D, Supplementary Table 8C, 9-10).

Many genes that were differentially expressed between \textit{M.tb}-infected infants and \textit{M.tb}-uninfected infants in CMV-infected infants had the opposite patterns of fold change between \textit{M.tb}-infected and \textit{M.tb}-uninfected infants in CMV-uninfected infants (Figure 5B). This led to a lower number of differentially expressed genes, when we analysed combined data from CMV-infected and CMV-uninfected infants, compared to when we analysed them separately by CMV infection status. Gene sets associated with \textit{M.tb} infection were strikingly different among CMV-infected and CMV-uninfected infants (Figure 5A and Supplementary Figure 5B-5D).

To further validate our RNA-seq results, we analysed a dataset from the training set of \textit{M.tb}-infected adolescents in the South African Adolescent Cohort Study (ACS), who either progressed to active TB or remained disease-free [7]. Because the gene expression abundance in the ACS was quantified at the level of splice junction, we used the sum of all splice junction counts of each gene as the read count of the gene and performed differential gene expression analysis using DESeq2. Considering the high prevalence of CMV infection in South Africa [22], we compared the results in the ACS and our CMV-infected infants. We found that gene sets related to complement activation, neutrophils, anti-viral response and type-I interferon response were also upregulated in TB disease progressors in the ACS, which included genes such as CR1, C1QA, C1QB, C1QC, C2, C5AR1, NCF4, IFIH1, DDX58. However, gene sets related to B-cells and immunoglobulins were downregulated in progressors in ACS, though most of the differentially expressed genes in these sets were not immunoglobulin genes, which differs from those observed in our study (Figure 5C and Supplementary Table 11).
To investigate the association between CMV infection and susceptibility to *M. tb* infection we looked at the overlap between gene sets that were enriched in the differential expression analysis between *M. tb*-infected and *M. tb*-uninfected infants and also between CMV-infected and CMV-uninfected infants. The results showed overlap of differential expression of genes that were associated with B-cells and immunoglobulins (Figure 5A and Supplementary Figure 5A, 5E-5F, Supplementary Table 8D, 9-10).

**Ag85A-specific IgG response correlates negatively with BCG-specific IFN-γ response:**

We evaluated the possible associations between the BCG induced antibody response and T-cell response in our infection cohort and found a significant negative correlation between the ELISpot BCG response and Ag85A-specific IgG response on Day 28. This was the case both when the analysis included the ELISpot BCG response data from Day -7 samples only and also when Day -7 and Day 28 data were combined (Figure 6A, *p<0.05*). We then classified the infants into BCG-specific IFN-γ responders (SFC ≥ 60) and non-responders, selecting a cut-off (SFC = 60) which can separate the distribution of the BCG-specific IFN-γ response into a high-density region and a low-density region (Figure 6B). The Ag85A IgG response was significantly higher in IFN-γ non-responders (Figure 6C).

**Discussion:**

In this study we investigated whether immune correlates of risk of TB disease were also associated with susceptibility to *M. tb* infection in BCG-vaccinated South African infants, using the same protocols to ensure comparability of the results obtained from both studies [6]. Findings of the present study showed that correlates of risk of TB disease were not associated with an increased risk of *M. tb* infection in infants who became infected with *M. tb*, but did not progress to active TB during the course of the study. Susceptibility to TB disease and *M. tb* infection may be mediated in part by different immunological mechanisms, which might have implications in designing vaccines targeting prevention of *M. tb* infection and in extrapolating findings from Pol studies to PoD studies.

We evaluated a range of exploratory immune parameters in the *M. tb*-infected infants and control study groups, none of which were associated with increased risk of subsequent *M. tb* infection after correcting for multiple comparisons. However, immune parameters measured by the multiplex assay had much higher statistical significance and higher AUROC values than those measured by other methods. We found upregulation of cytokines, chemokines and complement components such as TNF- and IP-10, in the plasma when we analysed these parameters separately, indicating that they may mediate, at least in part, the increased risk of *M. tb* infection. Some of these proteins, such as IFN-2, IP-10 and VEGF, have previously been shown to be correlated with risk of TB disease [23]. Complement Factor I and Complement C5, which were both associated with increased risk of *M. tb* infection, had significantly different profiles at Day -7 compared with Day 28 samples. It is not possible to determine whether the
significant results observed for these two factors when the two time-points were analysed together were confounded by the MVA85A vaccination, late BCG-induced response or just an effect of increased sample size.

CMV is a common persistent herpesvirus infection that modulates immune functions. CMV infection is highly prevalent in low-and middle-income countries, is known to cause immune activation, and is associated with risk of TB disease [8,11,24,25]. The samples used in this study were from 3-6 month old infants, an age considered to be the peak age of primary CMV infection [26]. We evaluated CMV infection status by measuring CMV-specific IFN- responses using a pool of CMV peptides and also by measuring CMV-specific IgG and IgM responses. The goal of measuring CMV-specific IgG responses at two time points and only considering those infants that maintained at least 90% of their initial response as CMV-infected was to differentiate exposure to CMV from passive maternal CMV antibody transmission. Measuring CMV-specific IgG was more sensitive than measuring CMV cellular responses to peptides, which was probably limited by the number of peptides used and HLA-restriction.

We observed a tendency towards an increased prevalence of subsequent *M. tb* infection in CMV-infected infants compared to those with negative CMV reactivity. CMV-infected and CMV-uninfected infants have distinct *M. tb* infection classifier signatures; hence, CMV infection status should be taken into consideration when characterising *M. tb* infection gene signatures, in line with our previous observation from the disease cohort [11]. CMV infection induces inflammatory responses that may modulate the immune system, such as the upregulation of TNF- , IP-10 and complement C2 in the plasma and HLA-DR expression on CD4+ T-cells and CD8+ T-cells. CMV-induced immune activation might have perturbed the immune system and led to an impairment of BCG-induced immunity, which could explain the trend towards increased numbers of *M. tb* infected infants in the CMV reactive pool. In addition, immune activation may influence early immune defence mechanisms in the airways, which might contribute to higher susceptibility to *M. tb* infection.

As samples from infants who went on to develop TB disease were used in our previous correlates of risk of disease analysis, in this study we have evaluated immune correlates of risk of *M. tb* infection for a subset of infants who did not go on to develop TB disease. Thus these infants could be considered to be protected from TB disease, in parallel with being susceptible to *M. tb* infection. This is particularly true as the majority of infants infected with *M. tb* at Day 336 of the study (14/16) reverted to negative QFT reactivity during the study follow-up period. Unfortunately, there was not an adequate number of sustained QFT converters to perform any direct comparison with QFT reverters. Correlates of risk of *M. tb* infection might be different in infants who progress to active TB disease, study subjects were not followed up beyond end of study date to identify TB disease progressors. It is still unclear whether TB
disease incidence is lower in QFT reverters than in those with sustained QFT reactivity. In South African adolescents, TB incidence among QFT reverters was no different from whose with sustained QFT reactivity [15].

Total RNA-seq analysis revealed a number of genes that were differentially expressed between *M. tb*-infected and *M. tb*-uninfected infants. Unexpectedly, immunoglobulin-related genes were among those upregulated in *M. tb*-infected infants. The upregulation of immunoglobulin genes in *M. tb*-infected infants was found in both CMV-infected and CMV-uninfected infants when we analysed the data separately, suggesting that the upregulation of immunoglobulin genes was not simply a response to the elevated inflammation driven by CMV infection. Levels of Day -7 Ag85A-specific plasma IgG were also higher in infants who became infected with *M. tb* during the follow-up period. We have previously shown that Ag85A-specific IgG was associated with a reduced risk of TB disease [27], whereas our present study illustrates an association between antibodies and increased risk of *M. tb* infection. The role of antibodies in protection might be limited to development of TB disease rather than acquisition of *M. tb* infection. Further investigations are needed to interrogate the role of antibodies in more detail. Although these immunoglobulin-related transcripts were upregulated in infants who became infected with *M. tb*, none of the infants who converted to positive QFT at Day 336 progressed to active TB disease in the study and the majority of infants reverted to a negative QFT, suggesting these infants cleared their *M. tb* infection. The negative correlation between Ag85A-specific IgG response and BCG-specific T-cell response in this cohort, may suggest a potential skewing of BCG response towards antibody response and away from T-cell response.

The overlap between genes that were differentially expressed between the *M. tb*-infected vs. control groups and the CMV-infected vs. CMV-uninfected groups was dominated by immunoglobulin-related genes. Within each of CMV-infected and CMV-uninfected infants, these signatures were still higher in *M. tb*-infected infants compared to control groups, indicating that this effect is not entirely attributable to CMV infection and the antibodies translated from these immunoglobulin genes might not all be CMV-specific.

Transcripts encoding components of the complement system and pathways related to complement components were upregulated in infants who were subsequently infected with *M. tb* in this study. This was coupled with increased plasma levels of complement components such as complement C4b and C2. The complement system is an important part of the innate immune system and functions as a proteolytic cascade. The role of the complement system in TB has been investigated previously, and the complement
cascade was found to be elevated prior to TB diagnosis in the ACS cohort previously described [28]. At a molecular level, pathogenic mycobacteria invasion mechanisms require the association of the complement cleavage product C2a with mycobacteria resulting in the formation of a C3 convertase, which leads to C3b opsonization of the mycobacteria and recognition by macrophages [29].

IFN-related transcripts, together with plasma levels of IFN-2 and IP-10, were upregulated in infants who became infected with *M. tb*. This is consistent with previous observations of the association between increased IFN response and risk of TB disease and the presence of increased IFN transcripts in *M. tb*-infected adolescents who progressed to active TB [7,11]. One of the upregulated IFN-related genes is the negative regulator of IFN response (NRIR), which is one of the negative regulators that can be exploited by pathogens to escape type-I IFN-mediated antimicrobial activity, a feature that can be utilised to support *M. tb* survival [30]. Modular analysis revealed upregulation of type-I IFN and innate antiviral pathways in *M. tb*-infected. In addition, plasma levels of many inflammatory cytokines, such as TNF-, IFN-, IL-12 and GM-CSF, were also upregulated in the *M. tb*-infected infants. A recent nested case-control study using serum samples from TB patients in Uganda prior to a diagnosis of TB showed that the IP-10 level was positively correlated with increased CMV IgG level and the risk of TB disease [12]. In a murine model of chronic lymphocytic choriomeningitis virus (LCMV) and *M. tb* coinfection, chronic infection-induced TNF- was found to inhibit the induction of *M. tb*-specific T-cell immunity, contributing to the increased burden of *M. tb* in the LCMV and *M. tb* coinfected mice [31].

Our study identifies that inflammation and immune activation are associated with risk of *M. tb* infection in a subset of BCG-vaccinated South African infants who converted to positive QFT but did not progress to active TB disease. We have also shown that our previously identified correlates of risk of TB disease are not associated with susceptibility to *M. tb* infection in this cohort. Further work on these and other clinical trial cohorts is needed to understand the underlying immunological and molecular pathways in order to inform vaccine design and development.

**Declarations**

**Acknowledgements:**

This work was funded by TBVAC2020 and HMcS Wellcome Trust Senior Clinical Fellowship. Bioinformatics support was provided by the VALIDATE network.

**References**

1. WHO. Global tuberculosis report. 2021;


**Figures**
Figure 1

Flow chart illustrates the number of samples from QFT positive and QFT negative infants (IFN-g > 0.35 IU/ml). All *M. tb*-infected infants had 3 matched *M. tb*-uninfected controls based on sex, ethnic group, CDC weight-for-age percentile (±10 points), and time on study (±9 months). HH: Household.
Figure 2

QuantiFERON-TB Gold In-Tube conversion of infants

A: QFT over time for all *M.tbc*-infected infants with quantitative QFT, QFT IFN-g is presented in IU/m at days 0, 336 and end of study. Each line represents an infant.

B: Conversion and reversion data for all *M.tbc*-infected shown as percentages of total *M.tbc* infected infants.
Figure 3

Univariate conditional logistic regression results

A: Unadjusted p-values of immune parameters measured by different methods. Each point represents an immune parameter. Different columns include immune parameters measured by different assays. Immune parameters included in each column are listed in Supplementary Table 3. The red and blue points represent immune parameters upregulated and downregulated in \( M.tb \)-infected, respectively. The grey bar represents the median value of p-values of immune parameters measured by the corresponding assay. The dotted line represents the p-value threshold of 0.05. Mann-Whitney test without multiple testing correction was done to compare p-values of immune parameters measured by different methods.

B: Spearman correlation between plasma levels of cytokines, chemokines and complements. The colour intensity indicates the correlation coefficient. Only significant correlations were shown in the figure. P-values were adjusted using Benjamini-Hochberg multiple testing correction.

C: AUROC values of immune parameters measured by different assays. The higher the AUROC value, the better the immune parameter can classify \( M.tb \)-infected and \( M.tb \)-uninfected infants. The grey bar represents the median value of AUROC values of immune parameters measured by the corresponding
assay. Mann-Whitney test without multiple testing correction was done to compare p-values of immune parameters measured by different methods.

Figure 4

Cytomegalovirus (CMV) reactivity and rate of M.tb infection in CMV-reactive (CMV-infected) and non-reactive (CMV-uninfected) BCG-vaccinated South African infants.

A: CMV-specific IgG response measured on days -7 and 28 of the study. The horizontal line represents the kit manufacturer's recommended CMV-specific IgG positivity threshold; all infants with CMV-specific IgG below this value on day 28 were considered to have negative IgG reactivity to CMV (IgG-). Above this line, only infants who maintained 90% of their day -7 CMV-IgG levels on day 28 were considered to be IgG+.
B: CMV-specific IgM and IFN-g response: The vertical line represents the kit manufacturer’s recommended CMV-specific IgM positivity threshold. The horizontal line represents the threshold of CMV-specific IFN-g response, which is 17 SFC/million PBMC.

C: CMV antigen-specific IgG, IgM and IFN-g responses are shown in the first 3 rows. The last row is the CMV infection status of infants defined as detailed in the Results section.

Red: positive, blue: negative and grey: unknown (data unavailable due to limited sample availability).

D: Rate of *M.tb* infection in CMV-infected and CMV-uninfected BCG-vaccinated South African infants.
Figure 5

Gene set enrichment results using the CERNO algorithm and gene sets defined by Li et al. [18].

A: The first, third and last columns show enriched gene sets for differential expression between *M. tb*-infected and *M. tb*-uninfected infants among all, CMV-infected and CMV-uninfected infants respectively. The second column shows enriched gene sets for differential expression between CMV-infected and
CMV-uninfected infants. Each row represents a gene set. The size of the rug corresponds to the AUROC value in the CERNO test and the colour intensity of the rug corresponds to the enrichment q value. The proportion of red, blue and grey indicates the proportion of upregulated, downregulated and not differentially expressed genes within the gene set, respectively. Only gene sets with AUROC values larger than 0.75 are shown here. Other gene sets are listed in Supplementary Table 9.

B: The horizontal axis shows log$_2$ fold change of gene expression between *M. tb*-infected and *M. tb*-uninfected infants among CMV-infected infants, while the vertical axis shows that among CMV-uninfected infants. Each point represents a gene. The red and green points indicate differentially expressed genes (DEGs) in CMV-uninfected infants and CMV-infected infants, respectively, while the blue points indicate DEGs in both CMV-infected and CMV-uninfected infants. Only part of the DEGs were shown in the figure.

C: The first column shows enriched gene sets for differential expression between *M. tb*-infected and *M. tb*-uninfected infants among CMV-infected infants in the current study, while the second shows enriched gene sets for differential expression between TB progressor and non-progressor in adolescents in the South African Adolescent Cohort Study ACS [6]. Only gene sets with AUROC value larger than 0.75 are shown here. Other gene sets are listed in Supplementary Table 11.
Figure 6

Correlation between Ag85A-specific IgG response and BCG-specific IFN-γ response

A: Correlation between Ag85A-specific IgG response on Day -7 (Left) or Day 28 (Right) and BCG-specific IFN-γ response. Day -7 and Day 28 samples are included in the upper panel and only Day -7 samples are included in the bottom panel (Only unadjusted p-values are shown).

B: The distribution of BCG specific IFN-γ response for all samples (both Day -7 and Day 28 samples are included).

C: Ag85A-specific IgG response on Day -7 (Left) or Day 28 (Right) in BCG T-cell responders (ELISpot BCG response > 60 SFC) and non-responders (ELISpot BCG response ≤ 60 SFC) (only unadjusted p-values are shown).
Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTable2.xlsx
- SupplementaryTable1.xlsx
- SupplementaryTable3.xlsx
- SupplementaryTable4.xlsx
- SupplementaryTable6.xlsx
- SupplementaryTable9.xlsx
- SupplementaryTable10.xlsx
- SupplementaryTable11.xlsx
- SupplementaryTable8.xlsx
- SupplementaryTable7.xlsx
- SupplementaryDocument1.pdf
- SupplementaryFigure1A.pdf
- SupplementaryFigure1B.pdf
- SupplementaryFigure1C.pdf
- SupplementaryFigure2.pdf
- SupplementaryFigure3.pdf
- SupplementaryFigure4.pdf
- SupplementaryFigure5.pdf