Author’s response to reviews

Title: Human newborn Bacille Calmette-Guerin vaccination and risk of tuberculosis disease: a case-control study

Authors:

Helen Fletcher (Helen.Fletcher@lshtm.ac.uk)
Ali Filali-Mouhim (axf282@case.edu)
Elisa Nemes (elisa.nemes@uct.ac.za)
Anthony Hawkridge (anthony.hawkridge@westerncape.gov.za)
Alana Keyser (alana.keyser@uct.ac.za)
Samuel Njikan (snjikan@gmail.com)
Mark Hatherill (mark.hatherill@uct.ac.za)
Thomas Scriba (thomas.scriba@uct.ac.za)
Brian Abel (gabrioli01@gmail.com)
Benjamin Kagina (bm.kagina@uct.ac.za)
Ashley Veldsman (ashley.veldsman@uct.ac.za)
Nancy Agudelo (nadoma1@yahoo.es)
Gilla Kaplan (Gilla.Kaplan@gatesfoundation.org)
Gregory Hussey (gregory.hussey@uct.ac.za)
Rafick-Pierre Sekaly (rafick.sekaly@case.edu)
WILLEM HANEKOM (willem.hanekom@gatesfoundation.org)

Version: 1 Date: 18 Mar 2016

Author’s response to reviews:
Dear Editor,

RE-SUMMISSION: Human newborn BCG vaccination induces distinct patterns of immune activation, suggesting diverse mechanisms of protection

We would like to thank you and the reviewers for the useful comments and the opportunity to ameliorate and re-submit our manuscript.

Please find below responses to individual reviewer comments, including the list of changes made (and tracked) to the manuscript.

We hope that the revised manuscript could still be considered for the World TB Day article collection.

Yours sincerely

Willem Hanekom

Reviewer reports and our responses:

Reviewer #1: This is an elegant and comprehensive study that arises from work within a large neonatal BCG trial in non-HIV infected infants in South Africa. The authors' bottom line is that there are two distinct patterns of immune activation in blood drawn at age 10 weeks, and that these may translate into distinct mechanisms of protection. Whilst the immune responses do appear to be reasonably distinct when investigated at transcriptomic and proteomic levels, I am not sure that it can be concluded that both (or possibly either) are protective, as the data presented do not seem to suggest that active TB over the following two years is any more or less likely with either immune response. Hence, to my mind the authors demonstrate that the measured BCG-related response is broadly dichotomous, rather than necessarily relevant to protection against active TB. The authors, therefore, need to be more cautious in their conclusions.
The reviewer raises a valid point. The sample size was not adequate to independently interrogate protection in each of the two groups. We wish to make the point that host responses to a candidate TB vaccine may be different in distinct groups of recipients, and that this may have implications for vaccine efficacy. We have revised the abstract:

“Conclusions: Distinct patterns of host responses to Mycobacterium bovis BCG suggest that novel TB vaccines may also elicit distinct patterns of host responses. This diversity should be considered in future TB vaccine development.”

The first sentence of the conclusions paragraph in the discussion has been modified as follows:

“Our findings suggest that distinct patterns of host responses to Mycobacterium bovis BCG are present in infants vaccinated with BCG.”

Other points are discussed below as they arise in the text:

1. Methods - BCG was used to stimulate whole blood and PBMCs. Have the authors any experience of using M. tuberculosis? Clearly this is much more demanding from a lab point of view, but if so I would be interested to know if this gave similar results between cases and controls as they found with BCG.

We elected to use BCG in assays to enable comparison with previously published work from this cohort. Our aim was to assess BCG vaccine specific immune responses that might associate with risk of TB disease. We agree that incubation with M. tuberculosis would have been ideal, but logistic constraints and limited numbers of PBMC available precluded this option.

2. Methods - there is no mention of how they selected the number of study participants. Was this pragmatic, or a powered approach (not shown if it were) or just the samples that were available for assessment? This is important as in the Discussion the authors speculate that the small size of the study population may have affected their results; and Figure S1 indicates that there were more infants who developed TB or could have acted as controls within the whole population.
The number of study participants chosen on pragmatic grounds, based on experience from BCG studies in this infant population, where we have shown that group sizes of 30 individuals result in sufficient power to describe host responses on a population level. To identify correlates of risk, we initially adopted a hypothesis-driven approach, and measured T cell functions thought to be important to control TB in n=29 definite TB cases and 2 groups of controls (n=55 each), as shown in supplementary figures S3 to S5. As we did not demonstrate immune function associated with risk of TB, we then adopted an unbiased approach, and measured gene expression in the same participants. We included all definite cases and household controls from the primary cohort for whom PBMC were still available. Similar numbers of participants were randomly selected for the validation cohort.

The legend of figure 1 and the “participants” section of the results (page 13) have been amended to clarify participant selection:

“For functional assays, up to 29 definite cases and 110 controls (household controls, n=55, and community controls, n=55) were included in different analyses.

Primary analysis of transcriptional profiling was restricted to those cases and controls included in functional assay analysis for whom PBMC were available.”

We recognize that this small sample size (further divided into 2 groups, which we couldn't foresee) has limited the interpretation of our findings. This study taught us the importance of individual case-control matching and increased control-to-case ratio, which we have applied to other prospective biomarker studies (Zak et al, Lancet; Fletcher et al, Nature Communication, in press).

3. Results - Generally I felt that the selection of figures for the main manuscript was excellent, and each was clear. I would, though, suggest that Figure S1 is moved into the main text as it is a useful diagram that indicates to the reader the various populations and their relevant timelines.

We agree that this would help the reader to navigate the manuscript and have now moved Figure S1 into the main manuscript. The first reference to Figure 1 is on page 4.
4. Results - only infants who developed pulmonary TB were selected. What happened to those who had extrapulmonary TB? They may have been even harder to diagnose - and if BCG were more effective against this, in total might have been a very small number of cases. However, it can be argued that for all these reasons they are the population who really need to be studied.


5. Results - no indication is provided of when the infants developed TB. Did the authors investigate the relationship between this (ie age or time from BCG) and immune response?

There was no significant difference in median time to diagnosis between cluster 1 and cluster 2 case infants (8.43 versus 11.11 months). The following sentence on page 15 has been amended to read:

“We found no clinical variable that differed between the 2 clusters, including vaccination route, weeks of gestation, gender, ethnicity, birth weight (not shown); there were also no experimental variables that differed, including RNA quantity and quality (not shown); there was no significant difference between the time to TB diagnosis in cluster 1 and cluster 2 case infants (data not shown).”

6. Results - Figure 1: suggest order panel E same as panel D

The order of the cytokines for panel E agrees with that of panel D, where the cytokines overlap – the two panels show similar and different cytokines.
7. Results - Figure 3: panel B, I can't seem to identify the red open circles which indicate TB cases from cluster 2.

Apologies, this has now been corrected.

8. Results - the data presented on M1 and M2 monocytes is, to me, less convincing than the author's conclusions, which I feel need to tone down.

We agree that the final statement in this results section should be modified, and we have chosen to do this as follows (page 20):

“These findings suggest that monocyte phenotype may additionally contribute to differences in immunogenicity in the two infant clusters.”

Reviewer #3: Major comments:

1) This manuscript reports the results of a comprehensive investigation for immune surrogates of risk of TB in 10-week old infants utilizing a systems biology approach. The finding that no surrogate of risk in 10-week-old infants could be identified is of great importance to the field of TB vaccine development. Specifically, this was an extraordinarily well-done and comprehensive study performed in a region of the world hyper-endemic for TB, allowing for study feasibility. Moreover, the immunology was very well done and comprehensive. It is therefore unfortunate that all of these the data are presented in supplementary figures.

It is more challenging to present a negative finding than a positive one. The negative data are included to highlight how comprehensive the analyses were, in exploring potential differences between case and control infants. We tried to find the balance between presenting the key findings, while also sharing the negative data, and attempting to not overwhelm the reader, and thought it best to include the negative data as supplementary figures. (Other reviewers were concerned that we show too much data.) We thank the reviewer, but would like to propose that the negative data remain in the supplementary files.
2) The second main theme of the manuscript relates to intriguing evidence supporting that infants respond to BCG vaccination with two distinct responses. Additional data analysis suggests that one type of response, characterized by strong TH1 responses or high frequencies of monocytes, is associated with a higher risk of TB disease. While these findings have the potential to inform our understanding of BCG-induced immunity and inform future vaccine trials, there are two weaknesses that limit the significance as currently presented. First, while immunologic assay results are presented that confirm the gene expression data within the primary cohorts, no supportive analysis of the validation cohorts is presented.

It is true that we were not able to support these transcriptomic findings of risk of TB disease in an independent cohort. However, the validation cohort we used consisted of infants with probable TB (the most common manifestation of TB in childhood) and not definite (culture confirmed) TB. An inaccurate diagnosis of TB (always a challenge in infants) would confound attempts to validate the original observation. This may be why we observe partial validation of our original transcriptomic findings in a second cohort. However, we were able to validate our observation of higher frequencies of monocytes associating with risk of culture confirmed TB disease using data available from 3 independent cohorts in Durban, South Africa, using full differential blood count data – these studies are referenced in the discussion.

Secondly, there are problems with the figures and references to the figures, which make interpretation of the data presented very difficult to follow. These are:

a. Figure 2 on Page 15 - discussion of GSEA for unstimulated samples refers to Figure 2 - which doesn't show these results.

We apologize, during the submission process some of the formatting in the Figures was altered and we have now corrected these errors. We have also modified the figure 2 (now figure 3) legend as follows for clarity:

“Figure 3. Pathway analysis, using gene expression data from unstimulated PBMC, in the two clusters of infants.”

b. Page 31 - no figure referred to in line 26.
The pathways names are provided in a tabular form (Table S6); there is no Figure for this data.

c. Figure 3A: I believe the colors in the legends are switched around? (cluster 1 data is shown in blue and cluster 2 data is shown in green?)

Apologies, this has now been corrected.

d. Figure 3B & C: These figures cannot be easily interpreted. Where are the open red circles depicted? What are second and third columns in each figure? Because what is being depicted is unclear the conclusion stated in the manuscript (p 17, line 43; "Cases in cluster 1 displayed higher monocyte to T cell ratios compared with the pool of controls and with cluster 2 cases" is not substantiated by what is depicted in Figure 3b. Similarly, the conclusion on page 17, line 51, "cluster 1 cases had higher frequencies of BCG-specific CD4+ T cells producing type 1 cytokines" is not substantiated by what is depicted in Figure 3c.

Apologies, the formatting of some symbols (open red circles) was altered during the submission process. We have now corrected this and have checked that our conclusions are supported by the data shown in this Figure.

e. For Figure 4 A, 4: It would be helpful to keep the color for clusters 1 & 2 consistent between Figures 1 & 2 versus Figures 4 A (Cluster 1 is blue in Figures 1& 2 and is green in Figures 4A).

Apologies, this has now been corrected.

For Figure 4A, it is disappointing that the heat map for all the cases and controls (all four groups) is not shown rather than just the cases. This is a main point of the paper - yet data supporting this point is not depicted. In addition, it is unclear why the following text is buried within the figure legend, rather than included in the manuscript text. "M1/M2 pathways included those defined but not yet included in the MSigDB [33], and pathways associated with the development of an M2 phenotype, such as activated MAP kinase signaling and platelets [18, 34]. The latter
two functions have been shown to drive monocyte differentiation into giant cells, which can suppress the mycobacteria-specific immune response [18, 34]."

From current Figure 4 (which includes all available data from case and control samples) it can be seen that the strongest differences in myeloid to T cell ratio are between the case infants, which is why Figure 5 focuses on case infants only. We appreciate the reviewer may have missed this point as the Figure 4 legend had been corrupted in the previous version of the paper.

The discussion text in the Figure legend has now been moved from the Figure legend and into the main manuscript file to the end of the third paragraph of the discussion on page 22 as follows:

“The M1/M2 pathways used in this paper included those defined but not yet included in the MSigDB database, and pathways associated with the development of an M2 phenotype, such as activated MAP kinase signaling and platelets. The latter two functions have been shown to drive monocyte differentiation into giant cells, which can suppress the mycobacteria-specific immune response.”

In addition, we took the opportunity to further clarify how the heatmap was built, by modifying the legend of figure 5A as follows:

“To build the heatmap the infants were first ranked by increasing expression intensity for each gene. Then, the mean-rank, across the set of genes, for each infant was used to order infants from the lowest to the highest mean-rank. A Spearman correlation was used to assess the significance (p value<0.0008) of the association between the ordering of the infants and the monocyte to T cell ratio.”

f. For Figure 4B, the legend does not describe adequately what is shown in the table and describes two rows for M1 and M2 macrophages while M1 and M2 are depicted as separate columns.

This has now been corrected to read as follows:
“(B) P value table from an over-representation test performed using Fishers Exact Test to identify M1 (right column) and M2 (left column) gene signature enrichment amongst genes up-regulated in each of the four groups of infants.”

**g.** Figure 4C cannot be interpreted for the same reasons as described for Figures 3B & C.

Apologies, this has now been corrected.

For further clarity, the results section (page 19) has been modified to include reference to the specific panels (A, B, C) discussed in this figure.

**h.** Figure S7: The reference to this supplementary figure comes first in the manuscript.

This has been corrected to Figure S1 and the numbering of other supplementary Figures adjusted accordingly.

**i.** Figure S10: The manuscript text and figure legend for S10 insufficiently explain this data. The data shown appears to show that for both clusters inflammatory genes go down and cytokine genes go up with BCG stimulation. In addition this is more dramatic for Cluster 1 than Cluster 2. However, showing the comparison between clusters for each condition as the primary comparison makes this point hard to discern and the there are no p values shown for the cytokine comparison between UNS cluster 1 and BCG cluster 1 or between UNS cluster 2 and BCG cluster 2 - which presumably are significant (at least for cluster 1) if the conclusion stated in the manuscript that inflammatory gene expression goes down and cytokine expression goes up is supported by this data.

We agree that this Figure is difficult to interpret in its current format and have replaced this Figure with a version showing comparison between stimulation condition as well as between clusters (including p values).
j. Figure S11. Analogous to my comments on Figure 3 B & C.

This has been corrected.

3) Finally, in general this manuscript is dense and presents too much data for a single manuscript. There are 4 figures containing 13 parts, within which there are 32 graphs/illustrations. In addition, there are 11 supplementary figures containing 22 parts within which there are 32 graphs/illustrations. In this regard, the highly significant information presented in Figures 3 & 4 are given insufficient space and emphasis.

We thank the reviewer for this comment. It was challenging to find a balance between positive and negative findings. As stated above, we propose keeping the balance as is – please also refer to the response to comment 1 above.

Reviewer #5: The authors present an interesting analysis where they find no distinct differences between case and control infants vaccinated with BCG in terms of

1. Why are 33 infants used on page 6 of the clustering algorithm? How is this number justified?

Please refer to our comments about the sample size to reviewer 1.

Figure 1 and the “participants” result section have been amended to better clarify participant selection, which now reads (page 13):

“For functional assays, up to 29 definite cases and 110 controls (household controls, n=55, and community controls, n=55) were included in different analyses.

Primary analysis of transcriptional profiling was restricted to those cases and controls included in functional assay analysis for whom PBMC were available.”
2. Infants were randomly selected from a RCT, presumably blinded to what arm the infants were in. Is there any concern that infants might respond differently to the vaccine depending on the treatment arm of the RCT they were randomized to?

This is an important point and we did indeed check that there was no impact of BCG vaccine route on clustering. The following paragraph on page 15 briefly addresses this:

“We found no clinical variable that differed between the 2 clusters, including vaccination route, weeks of gestation, gender, ethnicity, birth weight (not shown); there were also no experimental variables that differed, including RNA quantity and quality (not shown); there was no significant difference between the time to TB diagnosis between cluster 1 and cluster 2 infants (data not shown).”

Additional statistical adviser:

The microarray analysis approach appears reasonable, they have used R/Bioconductor to normalise and log transform the raw expression data. Further analysis includes clustering, differential gene expression and pathway analysis. I would recommend that software citations and if possible version numbers are included.

These have now been included (see below)

Specific comments:

Page 6, line 12: what does "screened for quality" actually mean? Were there criteria by which data was excluded? If so please specify.

This has been specified in the method section, page 6:
“Scanned array images were visually inspected for artifacts or mishandling. Diagnostic plots including density plots, box plots and a heatmaps of between-array distances were used to assess hybridization quality within the whole data set. One array failed quality screening.”

Page 6, line 14: The citation for the R statistical language is missing. Also for the purposes for reproducibility, version numbers should be included.

www.r-project.com, version 3.1.2

Page 6, line 44: "Using 4 different pre-clustering.....", could this sentence be clarified, "the four different sets of input genes resulting from the filtering process", is referring to the results of the genefilter package? Also the four numbers "3077, 5306, 7902, 9878" are these the size of these gene sets? Are these lists of genes available in the supplementary? The sentence is not clear to me.

For clarity we have replaced this paragraph with the following (method section, page 6-7):

“Using the genefilter package with 4 different filtering criteria, we selected different numbers of probes to assess the stability of clustering; probes retained with expression values greater than 100 in at least 2 samples and interquartile (IQR) range in expression across samples >0.2 (9878 probes), probes retained with expression values greater than 200 in at least 2 samples and IQR range in expression across samples >0.2 (7902 probes), probes retained with expression values greater than 200 in at least 2 samples and IQR range in expression across samples >0.3 (5306 probes), probes retained with expression values greater than 200 in at least 2 samples and IQR range in expression across samples >0.4 (3077 probes). The selected probes were used as input to test the stability and robustness of the clustering process.”

Page 7, line 22: a citation for limma should be included. It is listed here


This reference has now been added into the manuscript text