nature portfolio

Peer Review File

The frequency of CD38^{+} alveolar macrophages correlates with early control of *M. tuberculosis* in the murine lung

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

The study is largely based on multimodal scRNA-seq analysis of Mtb-infected murine lung tissue at various timepoints post infection. The authors focused on macrophage populations and identified early permissive and restrictive populations. Usage of reporter bacteria, cell sorting and scRNA-seq allowed identification of various subsets of lung macrophages, their trajectories and DGE, particularly in subsets which cause stress responses in Mtb. CD38 seemed to be expressed by AM which restrict Mtb and were epigenetically characterized by a chromatin structure enabling pro-inflammatory and antimicrobial programs (e.g. iNOS). Whereas in naïve animals the AM were largely CD38-, later in infection CD38+ AM accumulated and differentiated primarily from recruited monocytes. Intranasal delivery of BCG enriched for CD38+ expressing tissue-resident macrophages with capacity to limit Mtb growth. The study underlines the dynamic roles of macrophage populations during infection. The authors propose that CD38+ AM mediate the early control of Mtb and that CD38 may be used as a biomarker for immune protection in TB.

Overall, these observations are timely, as the universe of macrophage subsets is expanding, and relevant for immunology of TB. The studies are largely descriptive and correlative, relying primarily on scRNA-seq data with minimal validation. It would be relevant to validate the findings and discuss the data in the context of iNOS-independent immune responses to TB, as described in humans.

Several issues need to be addressed and these are presented below:

-Does ablation of CD38+ AM (or modulation of its expression) interfere with the outcome of TB? Does adaptive immunity regulate expression of this marker, and AM functions, during TB? Is CD38 expression directly linked to and controls iNOS (and NO production) and lysosomal function/autophagy? Are these biological functions (e.g. release NO, autophagy) enhanced in CD38+ macrophages?

-Figure 1 depicts subsets of AM, including CD38-expressing subsets, but a reasoning for clustering based on this marker is provided only later in the manuscript.
What is the explanation for detecting Ly6c expression on T and NK cells, and on B cells (1F)? Are doublets or the short tissue digestion an issue?

-It is unclear how many cells were analyzed in bystander vs infected moAM (2D). What is the reason for detecting expression of iNOS in T and B cells (2E)?

-Does the differential abundance testing take in the consideration potential sample errors inherent to scRNA-seq samples , e.g. variability due to loading, etc (4A)? Does it compute credible changes?

An alternative could be a compositional analysis with scCODA, a tool designed for scRNA-seq to identify credible changes in cell type composition (doi: 10.1038/s41467-021-27150-6).
In the same vein, are calculations in 4C, 4D, Fig 5 reflecting credible changes/representations?

-The BCG effect is very interesting. It remains unclear what cells are infected in this case because depicted % of TR-AM are way below 100%. How many cells, from how many mice, were profiled for the BCG group (6A)? Quantification (6E), and an exhaustive list of genes related to lysosomal function, phagocytosis and autophagy list (and biological processes), should be provided (6E).

-scATAC-seq studies remain unclear. What populations are shown in 7A? What was the rationale for the selection of the clusters in 7C, 7E? Validation would be helpful.

-The authors may reconsider whether all data is required to be presented in the main figures. Some figures cannot be properly read, panels in Fig 6, Fig 7.

-Certain statistical analyses may be re-evaluated, e.g. t-test for CFU with 4 data points per group. Was data normality checked? And considered for the choice of the statistical tests?

-Details about some data presentation (mean, median; SD, SEM, IQR etc) are missing in legends, what is shown?

-Statements in the discussion part should be toned down in absence of solid proof. For instance: “A significant shift occurs … BCG-treated-mice …enhancing intracellular trafficking and lysosomal/autophagic functions”. There is no data for causality, there is no data that showing enhanced trafficking etc. Solely transcripts do not allow concluding on a biological process such as cell trafficking, phagocytosis etc.

Reviewer #2 (Remarks to the Author):

In this manuscript, Pisu et al applied a number of cutting edge mycobacteriological and immunological tools, largely utilizing single cell sequencing applications, to study tissue resident (TR) and recruited, monocyte derived alveolar macrophages (AMs), their activation patters and the ability of these immune cells to control M. tuberculosis in the murine model.
They identified CD38-positive AMs, which comprise both TM-AMs and, later during infection, recruited/infiltrating moAMs, as the cells that best controlled Mtb counts and induced in vivo “immune-related” bacterial stress. These effector responses by activated, CD38-positive AMs were associated with gene expression signatures consistent with pro-inflammatory responses. They also showed that CD38+ TR-AMs pre-exist, at low numbers, in the lungs uninfected mice and, through

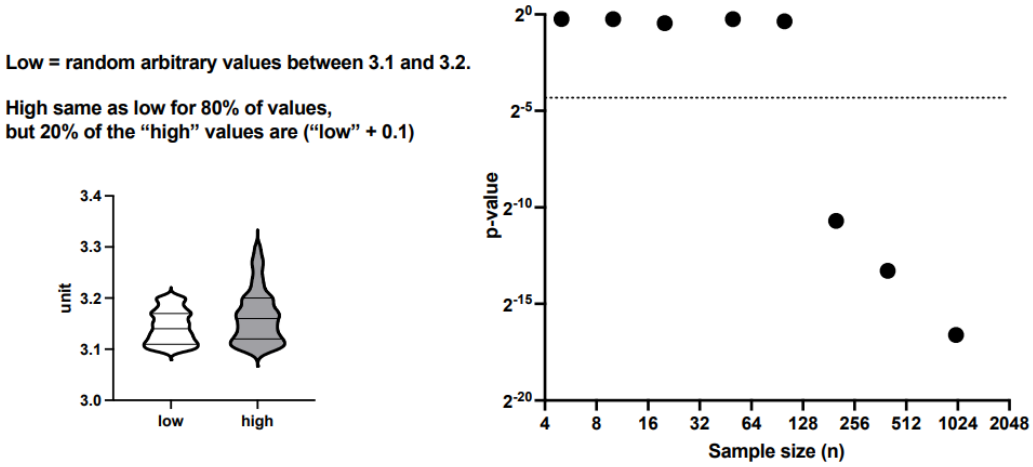
scATAC-seq analysis, that these cells have an epigenetic predisposition to such a proinflammatory response.
Finally, they showed that intranasal BCG administration led to increases in CD38+
AMs in the lungs and significant control of Mtb growth.
Overall the paper is quite well written, although very long. The experiments are elegant and very comprehensive and the analyses are complex and at times difficult to follow.
I have two major points that should be addressed to ensure that the interpretation is robust, while the discussion lacks broader perspective within the context of settings beyond the SPF mouse facility.

1. It is not shown how many mice are included in each experiment for virtually all analyses. Based on the methods section, it seems the number per group/timepoints ranges from 3 to 5 mice. Fig 6I, Suppl Fig 3G, H and I, seem to support this. First of all, it should be clear to the reader what the actual sample size (in number of individual mice, not only total number of single cells) is for each experiment. Secondly, a very large number of analyses are presented where protein marker or gene expression, or proportion of cells, are compared between hundreds to thousands of cells that are pooled from 3-5 mice, from 2 or more groups/conditions/ timepoints, applying statistical tests such as T-tests, Kruskal-Wallis with Dunns' or Wilcoxon signed-rank test. Unsurprisingly, in light of the massive sample size (# cells) per group, these analyses yield p-values that are very small or in some cases = zero. This statistical approach is just not appropriate for such single cell data (even though it is commonly used for ss-seq analyses in the literature).
I did a simple simulation which illustrates that this type of statistical analysis will yield very small p-values even if only 1 mouse out of 5 shows a small difference, as soon as the sample size (number of cells) exceeds a few hundred (see figure in the file attached). If we compare random values between 3.1 and 3.2 between 2 groups (called low and high in the figure below), where 20% (representing cells from 1 out of 5 mice) in the high group have values in a range between 3.2 and 3.3, then a Mass-Whitney U test will return very small p-values at sample sizes beyond ~200 cells per group. It is therefore likely that some of the comparisons that the authors interpret as showing highly significant differences may be driven by a single animal.

The authors should consult a statistician to develop statistical models that account for which individual mouse the cells come from (random effect) to ensure that the observed effects are consistently seen in multiple animals and not driven by an outlier mouse.

2. The authors should address if BCG persists in vaccinated mice up to the 2 month timepoint. This is important because persistence of BCG would be somewhat analogous to a chronic infection, which reflects a very different scenario ITO macrophage activation and protective effects, than a scenario where BCG has been cleared and trained innate "memory" is observed. For a vaccine to be realistically feasible as a public health intervention it should have long-term effects that last years beyond vaccine clearance. Even if BCG does get effectively cleared, the authors should discuss the likely duration of the effects they observe and how this may have implications for vaccination.

3. The discussion makes claims about the results informing rational vaccine design and host directed therapies. It would be useful for the authors to comment on how likely it is that the effects they see in SPF mice may translate to humans. In countries where interventions against TB are most urgently needed infants receive BCG vaccination at birth (i.e. most people are not "naïve") and people are continuously exposed to a variety of organisms, while many environmental exposures are experienced. Among older individuals infection with Mtb is also common. How would these factors affect lung AM activation, trained immunity and the phenotypes reported in the paper? What does this mean for the interpretation of the results?



Reviewer #3 (Remarks to the Author):

In this work the authors take a systematic approach to examining the time course of lung macrophage response to an intracellular lung pathogen. The data sets are intriguing and exciting for those interested in the kinetics and responsiveness of macrophages in the lung. The data sets provide new knowledge about the behaviour of different macrophage populations and the paper informs us how the different populations interact and respond - and how this response relates to the viability of the intracellular pathogen.
I think the authors were wise to focus on the lung and macrophage biology, their link through to eventual markers of immune status to be cautious but meaningful.

Reviewer 1:

The study is largely based on multimodal scRNA-seq analysis of Mtb-infected murine lung tissue at various timepoints post infection. The authors focused on macrophage populations and identified early permissive and restrictive populations. Usage of reporter bacteria, cell sorting and scRNA-seq allowed identification of various subsets of lung macrophages, their trajectories and DGE, particularly in subsets which cause stress responses in Mtb. CD38 seemed to be expressed by AM which restrict Mtb and were epigenetically characterized by a chromatin structure enabling pro-inflammatory and antimicrobial programs (e.g. iNOS). Whereas in naïve animals the AM were largely CD38-, later in infection CD38+ AM accumulated and differentiated primarily from recruited monocytes. Intranasal delivery of BCG enriched for CD38+ expressing tissue-resident macrophages with capacity to limit Mtb growth. The study underlines the dynamic roles of macrophage populations during infection. The authors propose that CD38+ AM mediate the early control of Mtb and that CD38 may be used as a biomarker for immune protection in TB.

Overall, these observations are timely, as the universe of macrophage subsets is expanding, and relevant for immunology of TB. The studies are largely descriptive and correlative, relying primarily on scRNA-seq data with minimal validation. It would be relevant to validate the findings and discuss the data in the context of iNOS-independent immune responses to TB, as described in humans.

Several issues need to be addressed and these are presented below:

Does ablation of CD38$^+$ AM (or modulation of its expression) interfere with the outcome of TB? Does adaptive immunity regulate expression of this marker, and AM functions, during TB? Is CD38 expression directly linked to and controls iNOS (and NO production) and lysosomal function/autophagy? Are these biological functions (e.g. release NO, autophagy) enhanced in CD38$^+$ macrophages?

We appreciate the reviewer's point. However, we want to emphasize that we are presenting CD38 as a marker that correlates with protective macrophage responses in tuberculosis infection, not as a mechanism directly linked to antimicrobial functions. CD38 is a well-established surface marker which is expressed by a large variety of immune cells in different conditions, including macrophages, T-cells, B-cells and NK cells (https://10.3390/cells8121527, https://10.3389/fonc.2022.775649) and is specifically linked to M1 macrophage inflammatory activation and polarization (https://10.1016/j.cellsig.2017.10.014, https://doi.org/10.1128/iai.00340-13, https://10.1007/s10059-012-0263-3, https://doi.org/10.1371/journal.pone.0126007, https://10.1155/2019/7026067). In our study we found it to specifically associate with protective phenotypes, allowing us to study how these macrophage subpopulations change overtime and in response to BCG vaccination.

In response to the specific questions:

- Ablation of CD38 AMs: We have previously performed ablation of AMs and IMs (Reference 6). At 2 weeks post-infection where 88% of the infected AMs are CD38$^-$ (Figure 5G), ablation of the AM population leads to a log decrease in bacterial load, whereas ablation of IMs (which are CD38$^+$) results in an increase.
- Regulation by Adaptive Immunity: We observe that the adaptive immune response is critical for activating CD38$^+$ AMs from a quiescent to a protective state, as highlighted through the manuscript. In addition, we have conducted flow cytometry analyses at 2wpi (when most infected macrophages are AMs – Figure 4C -) in both Rag1 and IFN-Y KO mice, which show severe lung pathology, observing no induction of the hspx'::GFP (Suppl Figure 2A), confirming the role of adaptive immunity in regulating macrophage activation and NO release.
- Link to antimicrobial function: We identified increased autophagy as a specific signature of BCG-infected cells, in alignment with previous studies (https://doi.org/10.1038/s41541-019-0122-8 ,10.1038/nm.1928). Hspx has been very well characterized to be a specific Mtb readout for intracellular NO release (https://doi.org/10.1371/journal.ppat.1004394), and hspx$^+$ Mtb are only present in CD38$^+$ macrophages (Figure 2E). Finally, we also have evidence that antimicrobial Mtb

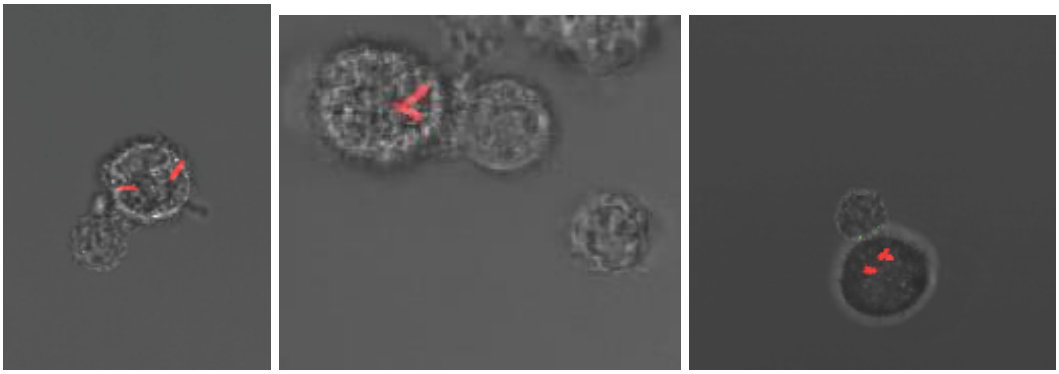
activity is specifically enhanced in CD38^{+} macrophages: our scRNA-seq analysis enabled the detection of intracellular Mtb mRNA within CD38^{+} macrophages. Mechanical lysis of the bacteria is required to release Mtb mRNA (see Reference 8 for more details), indicating that active degradation of Mtb within these cells is occurring. Our data reveals low levels of recovered Mtb reads at 2 weeks post-infection, with a significant increase from 3 to 6 weeks, coinciding with the adaptive immune response's onset and the heightened activity of CD38^{+} AMs and IMs (Supp Figure 2E). This temporal pattern and unique increased recovery of Mtb RNA in CD38^{+} populations provide further evidence of their role in controlling Mtb infection.

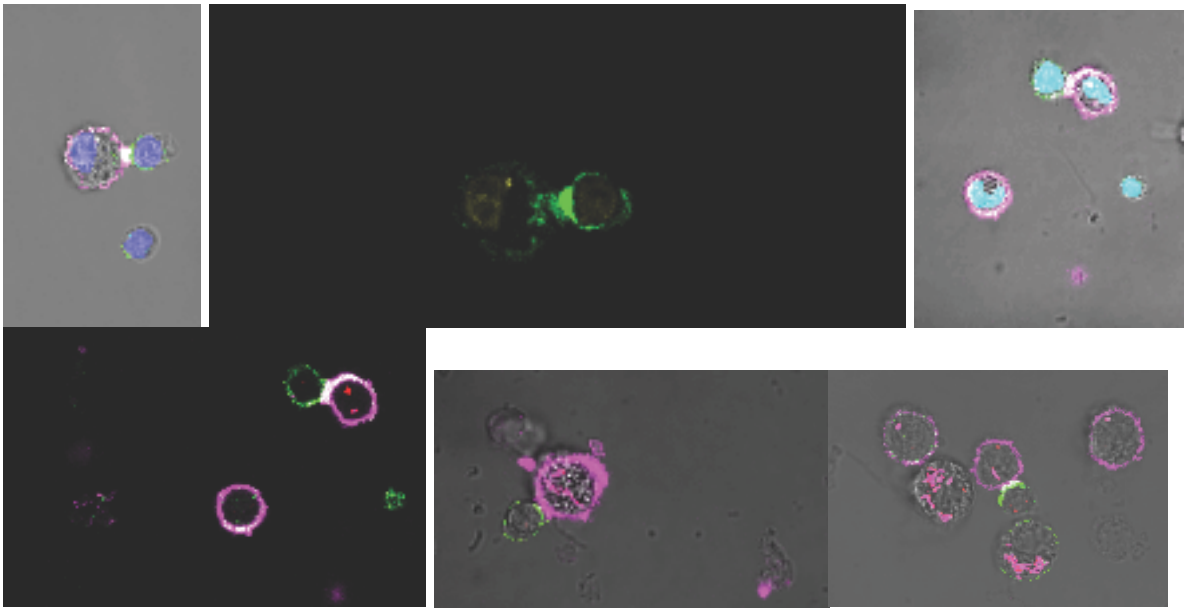
We have modified the manuscript to clarify and include these points.

Figure 1 depicts subsets of AM, including CD38-expressing subsets, but a reasoning for clustering based on this marker is provided only later in the manuscript.
What is the explanation for detecting Ly6c expression on T and NK cells, and on B cells (1F)? Are doublets or the short tissue digestion an issue?

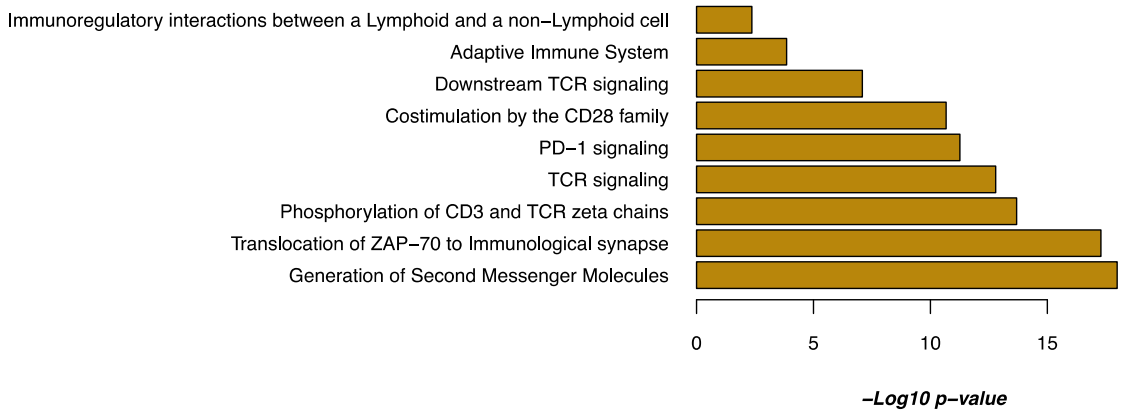
Cells in Figure 1A are clustered based on their transcriptional profiles in an unbiased manner, not by CD38 surface marker expression. We have modified the text detailing Figure 1 to make clear that this clustering is the result of unbiased clustering based on transcriptional profiling and not a direct consequence of classifying cells based on the expression of surface marker CD38. As detailed later in the manuscript, there is a perfect overlap between CD38 expression and the clusters identified through transcriptional analysis.

Regarding the second question. This is a very good point and something we have been exploring in some detail, but is still a work in progress. We had a similar observation in our published scRNA-seq study (Pisu et al, 2021). While the presence of autofluorescence and artificial doublets due to sample processing were initial concerns, subsequent analysis, including confocal microscopy of the sorted mCherry^{+} cells, revealed the presence of immune synapses between T-cells and infected macrophages, which last for some time after sorting. These synapses, visualized by localized CD3 expression at the contact points of the fusing membranes, indicate functional TCR activation. We and others have also observed similar “complexes” in ongoing NHP studies. We detect co-expression of genes specifically linked to TcR activation and immune synapsis formation in this population. The presence of monocyte/macrophage reads in this subset of T-cells is due to the presence of a small amount of these immune synapses among the sorted cell population. Our ongoing studies are focused on deepening our understanding of these immune synapses and their functional implications in tuberculosis infection. See below some pictures from ongoing work:





Red= mCherry, Green= CD3, Purple=CD64, Blue=DAPI



Gene Sets enriched in T-cells from immune synapsis.

It is unclear how many cells were analyzed in bystander vs infected moAM (2D). What is the reason for detecting expression of iNOS in T and B cells (2E)?

The number of cells analyzed in the different groups are n = 1048 cells for moAMs Bystander and n = 1537 cells for moAMs infected. We have added the numbers to the figure legend. Regarding the reason why the vast majority of “infected” T and B cells are associated with GFP high bacteria and there is *Nos2* detection, this is due to the recovery of a small amount of immune synapsis, discussed in the previous response.

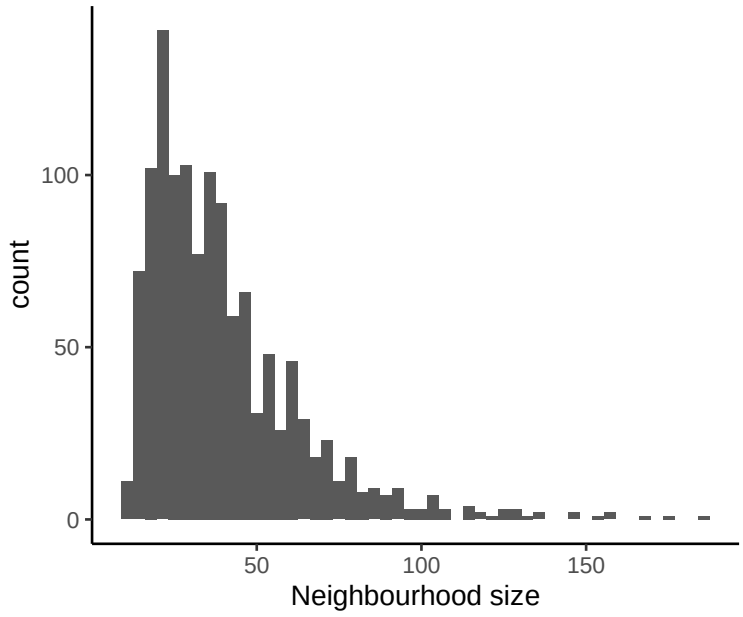
Does the differential abundance testing take in the consideration potential sample errors inherent to scRNA-seq samples, e.g. variability due to loading, etc (4A)? Does it compute credible changes? An alternative could be a compositional analysis with scCODA, a tool designed for scRNA-seq to identify

credible changes in cell type composition (doi: 10.1038/s41467-021-27150-6).
In the same vein, are calculations in 4C, 4D, Fig 5 reflecting credible changes/representations?

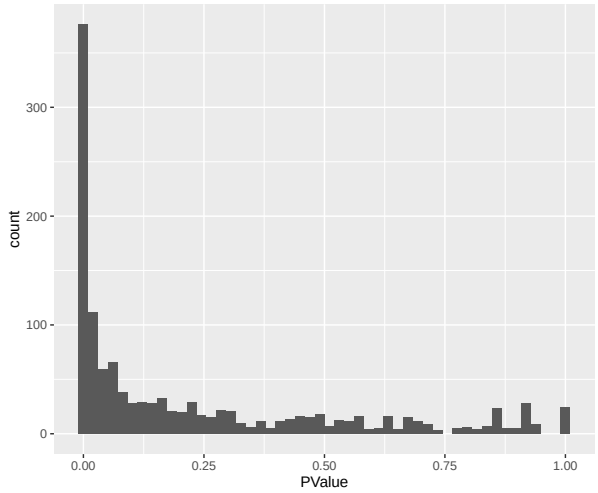
We believe that Milo (https://doi.org/10.1038/s41587-021-01033-z) and scCODA are the two most widely used compositional analysis tools on scRNA-seq datasets comparable to ours.
The reason we used Milo over scCODA is that while scCODA is effective for compositional analysis in datasets where clear cell-type definitions are applicable, it is less suited for studies like ours that involve transitional states between cell types and where only subpopulations within a cell type may be affected by the condition.
To name a few advantages of using Milo vs scCODA:
1. Milo uses k-nearest neighbor (KNN) graphs to define neighborhoods of cells in low-dimensional space, which is ideal for detecting compositional changes within smaller subpopulations. This approach is well-suited for our data's dynamic nature, allowing for the identification of subtle intra-cluster shifts.
2. The KNN construction in Milo incorporates steps to minimize the influence of technical factors on cell-cell similarities, which is crucial for maintaining the integrity of data interpretation across timepoints.
3. Milo employs a generalized linear model (GLM) to handle continuous covariates like time, providing a way to directly test for temporal changes. This is preferable over methods like scCODA, which focus on categorical differences. Despite potential compositional biases not explicitly modeled by the GLM, the widespread testing across thousands of neighborhoods helps mitigate this issue. Milo also uses TMM normalization to robustly adjust for compositional differences across samples.
4. The integration of negative binomial testing with edgeR for differential abundance analysis helps accurately estimate dispersion and manage variability inherent in scRNA-seq data. This approach minimizes the likelihood of false discoveries, which is vital in time-course studies with high sample variability.

We would also like to point out that in preparation for our data analysis, we conducted several 'sanity checks' as recommended by the developers of the package to ensure the robustness and appropriateness of our analysis parameters. These included:

1. Neighborhood Cell Distribution: We analyzed the distribution of the number of cells forming each neighborhood (nHood). The developers suggest an ideal empirical peak around 20 cells per neighborhood, indicating that the chosen 'k' and 'prop' parameters are well-suited for our dataset. This check confirmed that our settings were optimal for detecting meaningful biological variations without being skewed by over- or under-sampling of cell populations.



2) Distribution of Uncorrected P-values: We assessed the distribution of uncorrected P-values to ensure the statistical tests were balanced and unbiased. Under the null hypothesis, we anticipate these P-values to be uniformly distributed, which supports the appropriateness of the test conditions. A peak close to zero indicates significant findings, while a uniform distribution without skew suggests no artificial inflation of significance. A bimodal distribution, particularly with a second peak near one, could suggest the presence of residual batch effects or insufficient variance between replicates within one condition. Our analysis did not exhibit such bimodal patterns, confirming the absence of underlying batch effects and validating the robustness of our experimental setup.



As suggested by the reviewer, we expanded our analysis to include scCODA, which allowed us to compare results across different statistical methods. We categorized our scRNA-seq samples into two conditions based on the time of collection: 'Early Timepoint' for samples from 2 and 3 weeks post-infection, and 'Late Timepoint' for samples from 4 to 6 weeks post-infection.

Following the guidelines provided in the scCODA vignette (https://sccoda.readthedocs.io/en/latest/getting_started.html) we executed the analysis using both conservative and relaxed settings. Our findings from scCODA largely mirrored those obtained from the Milo analysis. Specifically, under conservative settings, scCODA identified an increase in $Nos2^{+}$ IMs at the late timepoint. The relaxed settings allowed to highlight a decrease in $CD38^{-}$ TR-AMs at the late timepoint, while $CD38^{+}$ TR-AMs and moAMs remained stable, consistent with the trends we observed in the Milo analysis (Supp. File 1).

Regarding the bulk populations changes in Figure 4C and 4D, and specifically the decrease in AMs and increase in IMs, the data aligns with what we see by flow cytometry (Suppl Figure 3) as well as what have been reported previously in literature by Cohen et al (Reference 2) in Figure 1A of their manuscript. We have added both the scCODA results and independent flow cytometry analysis in the main text. We have included text explaining the steps for the scCODA analysis and “sanity checks” in “Data Analysis” part of the Materials and Methods.

The BCG effect is very interesting. It remains unclear what cells are infected in this case because depicted % of TR-AM are way below 100%. How many cells, from how many mice, were profiled for the BCG group (6A)? Quantification (6E), and an exhaustive list of genes related to lysosomal function, phagocytosis and autophagy list (and biological processes), should be provided (6E).

In our analysis on the effect of BCG vaccination on macrophage populations phenotypes following subsequent Mtb challenge, we specifically analyzed Mtb-infected cells from the lungs of BCG-vaccinated mice.

We pooled cells from five individually-sorted mice to maximize recovery of infected cells for scRNA-seq analysis.
All cells included in the scRNA-seq analysis for the BCG-vaccination paragraph were Mtb-infected, confirmed by sorting based on mCherry positivity (Figure 6B). Regarding the comment "% of TR-AM are way below 100%", we believe the reviewer is referring to plot 6D. That plot compares the proportion of each TR-AM subtype (for the two groups – BCG/Unvaccinated) among the total infected macrophage population (including resident and monocyte-derived macs) in each group (ie: in Unvaccinated mice $CD38^-$ TR-AM_1 are 24.11% of the total infected macrophages, while they only represent ~7.27% of the total infected macrophages in BCG). We used that plot to clearly convey, in a visually impactful way, the shift in the composition of infected TR-AMs caused by BCG vaccination, which leads to a decrease in the percentage of infected non-responsive TR-AM populations.
We profiled these cells at the 2-week post-infection timepoint, recovering a total of 3,931 cells after QC. For comparative analysis, these Mtb-infected BCG-vaccinated cells were analyzed against 2,278 Mtb-infected unvaccinated control cells at the same timepoint.
This approach enabled us to focus on how intranasal BCG vaccination impacts macrophage responses upon subsequent Mtb re-infection, specifically examining the infected cell population due to their critical role in Mtb control and pro-inflammatory responses, as highlighted earlier in the manuscript.

We have included the table with all the 463 genes from Module 2 and respective membership values and p-values from the scWGCNA analysis and highlighted the genes related to lysosomal function/phagocytosis and autophagy, as requested. The table is provided as a Supplementary File (Supplementary Table 3). We have also included the pathway analysis table for both GO biological processes and Kegg analysis, with pathways related to autophagy/programmed cell death highlighted, as a Supplementary file as well. (Supplementary Table 4).

We have updated the main text in the paragraph to provide the readers with details about cell numbers and references to the Supplementary Tables. We also specified the numbers of mice used for BCG experiment in the Material and Methods section, paragraph "Sorting of the murine lung suspensions for scRNA-seq analysis". Finally, we also updated the figure legends to provide both the actual cell numbers and more details about the respective plots.

scATAC-seq studies remain unclear. What populations are shown in 7A? What was the rationale for the selection of the clusters in 7C, 7E? Validation would be helpful.

In the scATAC-seq part of the manuscript, we analyzed immune cells from naïve mice lungs (which are mainly constituted of tissue-resident macrophage populations, since there is almost no recruitment of monocyte-derived cells in pathogen-free mice) by scATAC-seq to study the epigenomic organization of AMs before infection.
In an effort to try to keep the section as concise as possible, we tried to summarize different scATAC-seq analysis at once, as well as the results of the integration of the scATAC-seq data with the scRNA-seq, and we agree with the reviewer that the way the section was written some parts were not too clear.

Figure 7A highlights the scATAC-seq unbiased clustering of chromatin profiles from naïve mice, using the ArchR package (reference 96). This sets the stage to explore the epigenomic landscape before infection.

Subsequently, we performed an unconstrained integration of scATAC-seq data with scRNA-seq data. This is a completely agnostic approach that takes all of the cells in the scATAC-seq experiment and attempt to align them to any of the cells in the scRNA-seq dataset, facilitating a comparison that is not biased by pre-defined conditions. For this purpose, we generated 'gene scores' for each cell in the scATAC-seq data. These scores, calculated as detailed in reference 96 and on the ArchR website, predict gene expression by assessing the accessibility of regulatory elements, aiding in understanding how chromatin structure might influence gene expression in various cell types.

Unexpectedly, this integration showed that scATAC-seq cluster C7 aligned with gene expression profiles typically absent in naïve mice (as seen in scRNA-seq figure 2A). This unexpected alignment is illustrated in Figures 7B and 7C, where we show a clear correspondence between the inferred gene expression of

scATAC-seq clusters prior to infection and their actual expression post-infection in scRNA-seq. For example, from Figure 7C is easy to see that scATAC-seq cluster C5 inferred gene expression (based on their chromatin organization) matches the gene expression profile of $Mki67^+$ replicative AMs in scRNA-seq, while cluster C6 and C8 match $CD38^-$ anti-inflammatory AMs and Cluster C7 matches $CD38^+$ pro-inflammatory AMs.

To further validate the findings, we performed marker peak and motif enrichment (Figure 7D), which showed significant enrichment of transcription factor binding sites associated with pro-inflammatory gene activation in cluster C7. These findings were corroborated by examining the expression of these transcription factors in our scRNA-seq data, revealing their specific transcriptional expression within pro-inflammatory populations during tuberculosis infection (Figure 7E). Open chromatin peak analysis within promoter regions of key pro-inflammatory genes (detailed in Supplementary Figure 6C) confirmed higher activity within cluster C7, solidifying the link between chromatin accessibility before infection and gene expression patterns after infection.

We therefore extensively modified the scATAC-seq paragraph of the results section, to provide more details about the analysis and the significance of the results to our findings. We also increased the amount of details provided in the associated Figure Legends, to make it easier for the readers to understand the visualizations and the relative data that it is presented. We also provide extensive technical details and explanation of the analysis in the Materials and Methods section under the “Data Analysis” part, so that readers can easily replicate and/or perform similar analysis following our manuscript.

The authors may reconsider whether all data is required to be presented in the main figures. Some figures cannot be properly read, panels in Fig 6, Fig 7.

We reorganized Figure 6 and 7 and moved some images in the Supplementary Figure Panels.

Certain statistical analyses may be re-evaluated, e.g. t-test for CFU with 4 data points per group. Was data normality checked? And considered for the choice of the statistical tests?

We fully understand the reviewer’s concern. In the first draft of the manuscript we had a full explanation of the statistical tests used in the Figure legends, which have been subsequently removed. We are aware that with 4 data points per group, the central limit theorem cannot be relied upon to assume normality of sample means. While it is usually assumed that CFU data is often log-normally distributed, for both CFU analyses (Figures 3D and 6C) we performed normality tests, equality of variance tests, as well as visual inspection of the underlying data to determine the most appropriate statistical method to use.

We performed Shapiro-Wilk tests for normality, Levene’s test to assess the homogeneity of variances between the two groups and analyzed the underlying data using Q-Q plots.

1) For Figure 3D:

Shapiro-Wilk Test Results

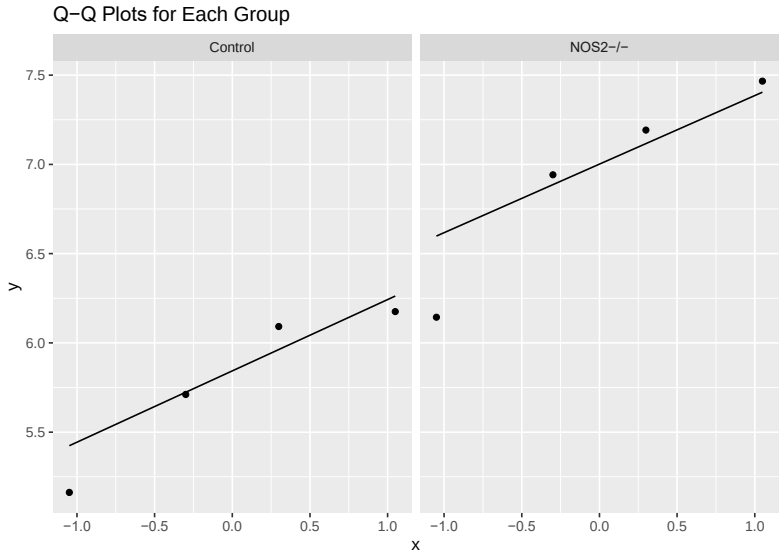
Control Group: $p = 0.440$
NOS2-/- Group: $p = 0.579$

These p-values indicate that we do not have evidence to reject the null hypothesis that the data in each group comes from a normally distributed population. Since the p-values are well above the typical threshold of 0.05, it suggests that the data distribution in both groups does not significantly deviate from normality.

Levene’s Test Results
F value: 0.0382

p-value: 0.8516

Levene’s test results, with a p-value of 0.8516, indicate that there is no significant difference in the variances between the two groups. This supports the assumption of homogeneity of variances, which is another key assumption for conducting a t-test.



As can be seen from the Q-Q plots the quantiles of the data approximately match the quantiles of a normal distribution.

2) For Figure 6C:

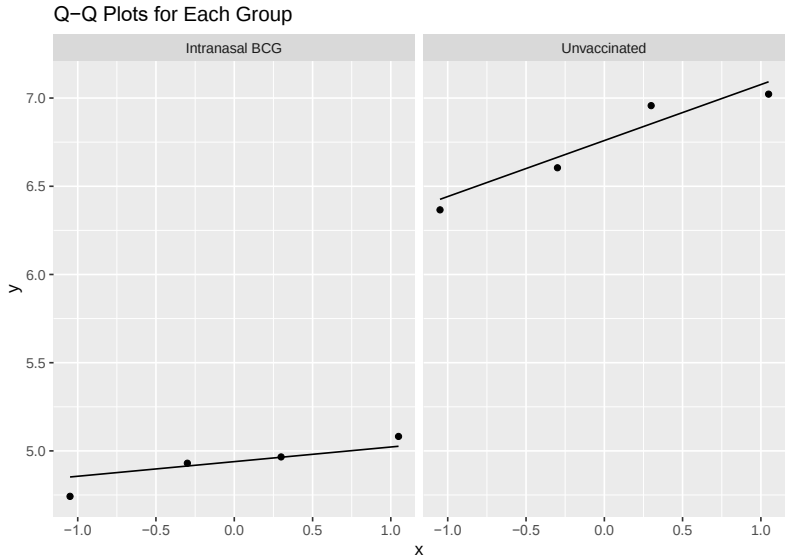
Shapiro-Wilk Test Results

Intranasal BCG: Shapiro-Wilk test p-value = 0.786
Unvaccinated: Shapiro-Wilk test p-value = 0.490

These p-values are significantly greater than 0.05, suggesting that the data for both groups does not deviate significantly from a normal distribution. This supports the assumption of normality needed for the t-test.

Levene’s Test Results
Levene's Test: F value = 4.7135, p-value = 0.07295

The p-value is greater than 0.05 but close to the threshold, which indicates a trend towards unequal variances but not strong enough to definitively reject the assumption of equal variances for the purposes of a t-test.



As can be seen from the Q-Q plots the quantiles of the data approximately match the quantiles of a normal distribution.

Overall, both groups passed the normality tests with high p-values, supporting the use of parametric tests which assume normal data distribution.

We modified the “CFU data” paragraph, in the “Statistical Analysis” section of the material and methods as follows to provide more context:

“*In this study, CFUs were quantified and analyzed to assess differences in bacterial burden between different conditions/treatments. Log10 transformations were applied to each CFU count prior to further analysis. To ensure the appropriateness of parametric tests, the data were first checked for normality using the Shapiro-Wilk test, a method suited for small sample sizes. This test evaluates the hypothesis that a sample comes from a normally distributed population, which is a critical assumption for the application of a two-sample t-test. Further, to assess the equality of variances between the treatment groups, Levene’s test was performed. Upon confirming that the assumptions of normality and homogeneity of variances were satisfied, a two-sample t-test was conducted to determine if there were statistically significant differences in the mean log-transformed CFU counts between the groups. The results of the t-test, including the p-values, were used to infer the statistical significance of the differences observed. The CFU plots presented in the results section are annotated with summary statistics including the mean and standard deviation (SD) of each group, along with the sample size (n), and the p-value from the t-test.”*

Details about some data presentation (mean, median; SD, SEM, IQR etc) are missing in legends, what is shown?

We updated the figure legends for the plots 3D and 6C to include a description of the summary statistics included in each plot.

Statements in the discussion part should be toned down in absence of solid proof. For instance: “A significant shift occurs … BCG-treated-mice …enhancing intracellular trafficking and lysosomal/autophagic functions”. There is no data for causality, there is no data that showing enhanced trafficking etc. Solely transcripts do not allow concluding on a biological process such as cell trafficking, phagocytosis etc.

We have revised the language in the discussion section to more accurately reflect the transcriptional nature of our data.

Reviewer 2:

In this manuscript, Pisu et al applied a number of cutting edge mycobacteriological and immunological tools, largely utilizing single cell sequencing applications, to study tissue resident (TR) and recruited, monocyte derived alveolar macrophages (AMs), their activation patters and the ability of these immune cells to control M. tuberculosis in the murine model.

They identified CD38-positive AMs, which comprise both TM-AMs and, later during infection, recruited/infiltrating moAMs, as the cells that best controlled Mtb counts and induced in vivo "immune-related" bacterial stress. These effector responses by activated, CD38-positive AMs were associated with gene expression signatures consistent with pro-inflammatory responses. They also showed that CD38+ TR-AMs pre-exist, at low numbers, in the lungs uninfected mice and, through scATAC-seq analysis, that these cells have an epigenetic predisposition to such a proinflammatory response.

Finally, they showed that intranasal BCG administration led to increases in CD38+ AMs in the lungs and significant control of Mtb growth.

Overall the paper is quite well written, although very long. The experiments are elegant and very comprehensive and the analyses are complex and at times difficult to follow.

I have two major points that should be addressed to ensure that the interpretation is robust, while the discussion lacks broader perspective within the context of settings beyond the SPF mouse facility.

1. It is not shown how many mice are included in each experiment for virtually all analyses. Based on the methods section, it seems the number per group/timepoints ranges from 3 to 5 mice. Fig 6I, Suppl Fig 3G, H and I, seem to support this. First of all, it should be clear to the reader what the actual sample size (in number of individual mice, not only total number of single cells) is for each experiment. Secondly, a very large number of analyses are presented where protein marker or gene expression, or proportion of cells, are compared between hundreds to thousands of cells that are pooled from 3-5 mice, from 2 or more groups/conditions/ timepoints, applying statistical tests such as T-tests, Kruskal-Wallis with Dunns' or Wilcoxon signed-rank test. Unsurprisingly, in light of the massive sample size (# cells) per group, these analyses yield p-values that are very small or in some cases = zero. This statistical approach is just not appropriate for such single cell data (even though it is commonly used for ss-seq analyses in the literature).

I did a simple simulation which illustrates that this type of statistical analysis will yield very small p-values even if only 1 mouse out of 5 shows a small difference, as soon as the sample size (number of cells) exceeds a few hundred (see figure in the file attached). If we compare random values between 3.1 and 3.2 between 2 groups (called low and high in the figure below), where 20% (representing cells from 1 out of 5 mice) in the high group have values in a range between 3.2 and 3.3, then a Mass-Whitney U test will return very small p-values at sample sizes beyond ~200 cells per group. It is therefore likely that some of the comparisons that the authors interpret as showing highly significant differences may be driven by a single animal.

The authors should consult a statistician to develop statistical models that account for which individual mouse the cells come from (random effect) to ensure that the observed effects are consistently seen in multiple animals and not driven by an outlier mouse.

Following comments from both reviewers we added the information about mice numbers and number of cells analyzed in both Figure Legends (when appropriate) and into the Materials and Method sections where this information was missing. In general, we used 5 mice for sorting of infected cells, and 3 mice for sorting of Bystander and Naïve populations.

Regarding the use of the non-parametric tests Wilcoxon rank-sum test (Mann-Whitney U test) to compare median values among groups in scRNA-seq, we met with the Cornell Statistical Consulting Unit and discussed this concern with Stephen Parry (https://www.stephenparryconsulting.com , stephen.parry@cornell.edu). Mr. Parry confirmed that the use of the non-parametric test to compare median values across groups, as we have presented, aligns with standard practices for analyzing scRNA-

seq data. He advised incorporating effect size measures and explicitly noting the number of cells analyzed (n) in our plots for greater transparency. Accordingly, we have added Cliff's Delta as an effect size measure to our plots. Cliff's Delta is especially suitable for gene expression data, which may be skewed or contain outliers, as it provides a standardized measure of the likelihood that a randomly selected value from one group will exceed a randomly selected value from another group, adjusted for the opposite probability.

Mr. Parry also suggested performing bootstrap analysis to address concerns about potential bias from data pooled from multiple mice influencing the results. Specifically, he recommended testing the robustness of our findings by downsampling our data at varying sample sizes. We conducted bootstrap analyses (with replacement) for 10,000 iterations at both the actual sample sizes and reduced sample sizes (250, 200, 150, 100, and 50 cells) for data presented in plots 2E, 3E, and Supp. Figure 4E. The results from our bootstrap analysis show that for larger sample sizes (200 cells and above), the confidence intervals remained strictly positive. This outcome reinforces the robustness of our findings, suggesting that the observed differences are consistent and not influenced by outliers. These positive confidence intervals imply that the median differences between groups are statistically significant and not zero. For smaller subsets, particularly at 50 cells, we observed that some confidence intervals included negative values. It is essential to clarify that by 'confidence intervals included negative values,' we mean the intervals span from negative to positive values, crossing zero. This indicates a loss of statistical power rather than a substantive change in the underlying biological phenomenon.

Confidence intervals that do not include zero and remain strictly positive or negative across all sample sizes signify stable and robust findings. In our case, the consistency of positive confidence intervals even at small sample sizes supports the conclusion that the significant differences observed in our original analysis are robust and not driven by outliers.

For plot 2E:

Bootstrap on full data:
Level Percentile
#95% (0.6095, 0.6912)

[1] "Confidence intervals for size 250 : 0.331050559196309, 0.968606164599493"
[1] "Confidence intervals for size 200 : 0.280665586012956, 1.01576609606046"
[1] "Confidence intervals for size 150 : 0.233189315741881, 1.06802215813194"
[1] "Confidence intervals for size 100 : 0.143553744122481, 1.15375256982277"
[1] "Confidence intervals for size 50 : -0.0985701914168844, 1.36179760642309"

For plot 3E:

print(bootstrap_results_full)

	Timepoint	CI_lower	CI_upper
2.5%	2_Weeks	0.1107920	0.1503642
2.5%1	3_Weeks	0.6724898	0.7262444
2.5%2	4_Weeks	1.0608289	1.0963339
2.5%3	6_Weeks	0.8672758	0.9147439

print(results_list[["250"]])

	Timepoint	CI_lower	CI_upper	Size
2.5%	2_Weeks	0.03226741	0.2376736	250
2.5%1	3_Weeks	0.52337462	0.8656597	250
2.5%2	4_Weeks	0.95060322	1.1965776	250
2.5%3	6_Weeks	0.74827460	1.0232920	250

> print(results_list[["200"]])

	Timepoint	CI_lower	CI_upper	Size

```
2.5%    2_Weeks 0.01302261 0.2560701  200
2.5%1   3_Weeks 0.49628297 0.8981813  200
2.5%2   4_Weeks 0.94051406 1.2063011  200
2.5%3   6_Weeks 0.72862551 1.0409013  200

> print(results_list[["150"]])
         Timepoint CI_lower        CI_upper    Size
2.5%    2_Weeks 0.0001236478 0.2705420  150
2.5%1   3_Weeks 0.4666055518 0.9041325  150
2.5%2   4_Weeks 0.9087516006 1.2363787  150
2.5%3   6_Weeks 0.6984117229 1.0723197  150

> print(results_list[["100"]])
         Timepoint  CI_lower     CI_upper    Size
2.5%    2_Weeks -0.03288358 0.2898970  100
2.5%1   3_Weeks  0.41817850 0.9689444  100
2.5%2   4_Weeks  0.88336423 1.2647558  100
2.5%3   6_Weeks  0.66308672 1.1103879  100

> print(results_list[["50"]])
         Timepoint  CI_lower     CI_upper    Size
2.5%    2_Weeks -0.09163614 0.3933878   50
2.5%1   3_Weeks  0.31448853 1.0741024   50
2.5%2   4_Weeks  0.81002851 1.3583986   50
2.5%3   6_Weeks  0.56488177 1.2062364   50
```

For plot Supp. Figure 4E

```
Gene                CI_lower                  CI_upper
 [1,] "Dennd1a"  "1.1005683479628"   "1.19405745884172"
 [2,] "Rapgef2"  "0.777911534174067" "0.886585744694527"
 [3,] "Vav3"     "0.575436801740864" "0.646017009428453"
 [4,] "Rasgef1b" "0.910394279925982" "1.00687078780294"
 [5,] "Rabgef1"  "0.564613451842999" "0.654912656466352"
 [6,] "Rapgef6"  "0.432504664906889" "0.496718498685333"
 [7,] "Rab8b"    "1.18242108442741"  "1.2957004573287"
 [8,] "Arhgap10" "0.572662817056634" "0.643184716580974"
 [9,] "Arhgap24" "0.896043244151536" "0.975007620291822"
[10,] "Dock1"    "0.327621798827824" "0.377039444138202"
[11,] "Tbc1d5"   "0.844655170163298" "0.915266593838302"
[12,] "Asap1"    "0.951760715157773" "1.02869858471935"
[13,] "Arl15"    "0.580685628710375" "0.641515099373777"
```

While it would be too long to attach all the different results for downsampling for each gene, we confirmed the confidence intervals to be significantly positive for each gene represented in Supp. Figure 4E at the different sample sizes.

2. The authors should address if BCG persists in vaccinated mice up to the 2 month timepoint. This is important because persistence of BCG would be somewhat analogous to a chronic infection, which reflects a very different scenario ITO macrophage activation and protective effects, than a scenario where BCG has been cleared and trained innate “memory” is observed. For a vaccine to be realistically feasible as a public health intervention it should have long-term effects that last years beyond vaccine clearance. Even if BCG does get effectively cleared, the authors should discuss the likely duration of the effects they observe and how this may have implications for vaccination.

Intranasal and IV BCG vaccination have been well established in the field as experimental protocols to drive protective responses against Mtb re-infection in both mice and NHP (References 11, 12, 13, 14, 15, 16, 17). There is no intention to present this as a vaccine strategy for humans. In this study we use intranasal BCG vaccination merely as a tool to understand the compositional shifts and transcriptional modifications that intranasal BCG causes on the AM population, in the light of the newly identified $CD38^{+/-}$ subpopulations.
By studying these compositional shifts and transcriptional modifications, we can in future work screen for chemotherapeutics that drive similar desirable activation responses in AMs. Accordingly, we have updated the discussion section to emphasize that our focus is not on trained immunity or the use of BCG as a long-term vaccination strategy. Instead, we are exploring the immediate changes in AM composition and phenotype that are driving protection against Mtb infection.
It is also crucial to acknowledge, as the reviewer pointed out, that determining the long-term persistence of BCG's effects on AM activation would necessitate additional longitudinal studies. Our experimental focus is on identifying the immediate, rather than long-term, changes, that can then be used to our advantage in the development of new host-directed chemotherapeutics.

3. The discussion makes claims about the results informing rational vaccine design and host directed therapies. It would be useful for the authors to comment on how likely it is that the effects they see in SPF mice may translate to humans. In countries where interventions against TB are most urgently needed infants receive BCG vaccination at birth (i.e. most people are not “naïve”) and people are continuously exposed to a variety of organisms, while many environmental exposures are experienced. Among older individuals infection with Mtb is also common. How would these factors affect lung AM activation, trained immunity and the phenotypes reported in the paper? What does this mean for the interpretation of the results?

We appreciate the reviewer's inquiry into the translatability of our results from SPF mice to human populations, particularly in the context of vaccine design and host-directed therapies. Our study primarily utilizes SPF mice to elucidate fundamental mechanisms by which AM activation induced by BCG infection can potentially inhibit Mtb spread. This model is instrumental in identifying specific phenotypes and responses that could inform the development of therapeutics aimed at eliciting similar protective responses in humans.

In revising the discussion section, we have clarified that our references to rational vaccine design and host-directed therapies are specifically tied to identifying phenotypes that prevent Mtb spread into the lungs. This, in turn, could guide the screening of therapeutics that promote similar responses in AMs. While the SPF model provides a controlled environment to study these mechanisms, we acknowledge the complexities involved in translating these findings to humans. Factors such as pre-existing immunity from BCG vaccination, continuous exposure to diverse pathogens, and genetic variability significantly affect the immune responses in human populations.

While we don’t currently have data to address these complex issues in humans, we cite similar activation phenotypes observed in NHPs post-BCG vaccination up to 6 months (references 12 and 16), which correlate with protection against Mtb re-infection, suggesting potential translational relevance.

Reviewer #3

In this work the authors take a systematic approach to examining the time course of lung macrophage response to an intracellular lung pathogen. The data sets are intriguing and exciting for those interested in the kinetics and responsiveness of macrophages in the lung. The data sets provide new knowledge about the behaviour of different macrophage populations and the paper informs us how the different populations interact and respond - and how this response relates to the viability of the intracellular pathogen.
I think the authors were wise to focus on the lung and macrophage biology, their link through to eventual markers of immune status to be cautious but meaningful.

We thanks the reviewer for their comments and believe that we have presented our data, and its interpretation, with care.

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

The authors made efforts to address the comments and updated the manuscript by restructuring the figures (Fig. 7), adding some new data (Suppl. 2A, E) and rephrasing the manuscript text (results part Fig. 7, discussion part). However, a causality has not been established for CD38 expression on AMs and control of Mtb. The early depletion studies of AMs and IMs (ref. 6) do not answer the question related to control of Mtb replication in CD38+ AMs. Further, coincidence of adaptive immune response's onset and the heightened activity of CD38+ AMs and IMs does not imply direct regulation of expression of the surface marker. These temporal patterns do not imply causality. The authors explain that the aim is "presenting CD38+ as a marker that correlates with protective macrophage responses in tuberculosis infection, not as a mechanism directly linked to antimicrobial function". According to this statement, current manuscript data, and to avoid confusions, they should consider revising the title of the manuscript.
The explanation of Ly6c expression in T and NK cells, or iNOS in T and B cells, notably immune synapses, indicates presence of doublets. Doublets should be excluded from analysis as they may introduce artefacts beyond these obvious ones.
Overall, the investigations remain largely descriptive.

Reviewer #2 (Remarks to the Author):

The authors have addressed my concerns

Reviewer #1 (Remarks to the Author):

Reviewer comment: The authors made efforts to address the comments and updated the manuscript by restructuring the figures (Fig. 7), adding some new data (Suppl. 2A, E) and rephrasing the manuscript text (results part Fig. 7, discussion part). However, a causality has not been established for CD38 expression on AMs and control of Mtb. The early depletion studies of AMs and IMs (ref. 6) do not answer the question related to control of Mtb replication in CD38+ AMs. Further, coincidence of adaptive immune response's onset and the heightened activity of CD38+ AMs and IMs does not imply direct regulation of expression of the surface marker. These temporal patterns do not imply causality. The authors explain that the aim is "presenting CD38+ as a marker that correlates with protective macrophage responses in tuberculosis infection, not as a mechanism directly linked to antimicrobial function". According to this statement, current manuscript data, and to avoid confusions, they should consider revising the title of the manuscript.
The explanation of Ly6c expression in T and NK cells, or iNOS in T and B cells, notably immune synapses, indicates presence of doublets. Doublets should be excluded from analysis as they may introduce artefacts beyond these obvious ones.
Overall, the investigations remain largely descriptive.

Response:

1. With respect to the title. We agree with the reviewer and did not mean to overstate our results. We have modified the title to: "The frequency of CD38+ alveolar macrophages correlates with early control of M. tuberculosis in the lung."

2. In response to the concerns regarding the inclusion of doublets. We identified doublets expressing *Ly6c*, *Nos2*, or *Mafb* in T-cells, NK cells, and B-cells. These comprised 1.61% of our dataset (811 out of 49,600 cells), and we have removed them from our datasets. Their exclusion is reflected in the revised figures 1 and 2. More specifically removing the doublets led to the loss of *Mafb* and Ly6c2 signal in the lymphocytes/NK cells in Figure 1F, and loss of *Nos2* signal in the same cell population in Figure 2E. The doublets did not appear anywhere else in the figures and were not mentioned in the text.

We hope that these resolve the remaining concerns expressed by reviewer 1.