**Reviewer 1 more conclusion**  
In this study, the authors conducted a multivariate taxonomic and functional microbiome comparison of publicly available human bronchoalveolar lavage fluid (BALF) metatranscriptome samples amongst COVID-19, community acquired pneumonia and uninfected samples with a stratified analysis based on mortality amongst the COVID-19 cohort with known outcomes of deceased versus survived, identified significant taxonomic and functional differences in BALF metatranscriptomes associated with COVID-19, community acquired pneumonia and uninfected samples and also indentified significant taxonomic and functional differences in BALF metatranscriptomes associated with COVID-19 disease and death. It is an intersting study, which provide an insight of the role of the human microbiome in pneumonia. However, there is one question confused me.  
The authors compared the BLF metatranscriptome samples amongst COVID-19, community acquired pneumonia and uninfected samples and got some results, but did not mentioned in the conclusion part.  
  
**Reviewer 2 – more discussion**  
This paper, "Analysis of Bronchoalveolar Lavage Fluid Metatranscriptomes Among Patients with COVID-19 Disease," performed a multivariate analysis to compare the BALF metatranscriptome samples from COVID-19 patients and controls.

However, similar studies have been published elsewhere, as noted by authors. The analysis of the current study is still important and worth sharing in the form of a publication.

The article is easy to follow, but the writing can be improved.

The results need to be discussed more.

The authors should expand on the discussion section and explain what their findings mean in a broader context.

The authors should also elaborate on the relevance of this analysis and how such analysis can be used in evaluating microbiome and pathogen-related associations.  
  
**Reviewer 3**  
In this manuscript by Jochum et al, the authors re-analyze metatranscriptome data from nine published datasets examining bronchoalveolar lavage fluid (BALF) from COVID-19 patients, community acquired pneumonia (CAP) patients and uninfected volunteers. The first section of the findings relates to identification of differential GO terms associated with the transcripts, while a second part of the analysis examines other microbial species found enriched in the BALF samples from COVID-19 patients compared to other conditions. Finally, the authors look at the differences in GO terms and microbial population abundances between COVID-19 survivors and those deceased due to the disease.  
  
The study premise is valuable, and the findings are interesting. Overall, I think the paper has merit. I have a few suggestions for the authors that I think will not only provide important additional details but will also improve the appeal of the findings. Additionally, there is need for more attention to detail in putting together figure/table legends.  
  
1.      Could the authors examine the code used to generate Figure 3? Why are the two samples with outcome= Deceased (fourth and fifth from the right side) clustered within Column cluster 2, when their trends are entirely different from samples in Cluster2?

*Thank you for pointing this out. The clustering configuration for this figure was originally calculated with variance stabilizing transformed (VST ) read counts for all 13,534 of the GO terms. We have amended the code and reconfigured the column clustering using the default parameters in the pheatmap package.*

2.      Table 4: Why are only negative log2Median ratios for COVID-19 discussed in this table? I think the manuscript would benefit from the examination of taxa that are enriched (positive log2median ratio) in COVID-19 compared to both CAP and uninfected. These include *Actinomyces, Cytophagales, Lactobacillaceae, Staphylococcus, Clostridia, Vibrionaceae* and the node representing *Comamonadaceae*. It is also interesting that these taxa are highly relevant to the taxa that are significant in the analysis of deceased vs survived COVID-19 patients.

*We would like to thank the reviewer for this specific comment regarding the unique presence of only depleted taxa being present in Table 4. In order to summarize the findings for the full list of 298 taxa with species level specificity were included in Supplementary Table 6 , we only reported values in table 4 based on the mean values at the genus level for taxa that containing >10 species level significant comparisons with a log2 median ratio >1.0 and a qvalue < 0.05 using Wilcoxon rank sum test adjusted for multiple test comparison, which resulted in only the depleted taxa. Of the 40 species that were significantly enriched as compared to the COVID-19 cohort, they were only significant as compared to the uninfected cohort, with small median differences < 0.005. We have amended Supplementary table 6 by removing taxa with “Inf” log2 median ratios, separated the enriched from deplete taxa to showcase taxa that were enriched, and have amended the results section to report these results.*

3.      Can the authors include measures of the overall metatranscriptome diversity between COVID-19, CAP and uninfected? it seems that inclusion of diversity indices may allow them to make the point that across the various studies from which data is taken, there is a change in microbial species richness in BALF samples in COVID-9 vs other conditions and uninfected.

*We would like to thank the authors for this inquiry and have modified the Supplementary data to showcase the alpha diversity metrics. We did not observe any statistically significant comparisons in alpha or beta diversity, and have ammeded the results section to include these findings alongside the addition of a supplementary table containing diversity metrics.*   
4.      Methods: it would be useful to provide more explanation about why the Dirichlet Multinomial Mixtures model was used, particularly since it's guides the authors decision to show the data based on DMM clustering (line 142)  
Thank you for this recommendation: We have expanded on the rationale for selecting DMM clustering in the recommended section of the manuscript:

*In order to identify and describe any variability for the observed taxonomic and functional features distinctive by case type or COVID-19 mortality, we employed Dirichlet multinomial mixture (DMM) probabilistic modelling. DMM modeling was selected as the means for identifying community clusters due to the algorithm’s ability to generate mixture component vectors based on unique hyperparameters in a multinomial fashion. By design, this methodology intrinsically incorporates dynamic features with ranging sample sizes and species rareity when clustering communities of similar composition, therein making it an optimal tool for this meta-analysis.*

5.      Continuation of previous point: Could the authors explain what they mean by depth=1

*Kristen will take care of this*  
6.      Figure and table legends should be expanded upon, and each term/ abbreviation shown in the figure/ table should eb explained in the legend. For e.g.  
a.      Figs. 1 and 3- what values do the heatmaps and their scales show?

*Thank you for pointing this out. The scales in heatmap figures1 and 3 represent z-score calculations of the total read counts matrices. We have amended the figure annotations and added labels to the heatmap scales in figures 1 and 3.*

b.       Figs. 2 and 4, what does the scale represent? Why is the scale labeled "Nodes"?

*The scale for figures 2 and 4 labeled as Nodes represents the log2 median ratio of the taxonomic proportions for each taxon scaled depicted as color gradient and the number of observations as sizes for the nodes. We have amended the figure by removing the label “Nodes” to clear up the confusion and amended the figure annotations to more clearly describe this explanation.*

c.      All tables: What is N.not.0

*N . not 0 represents the number of samples that do not have zero as the count for the reported taxa. We have amended this table by changing the annotations from N.not.0 to Samples w/ >0 counts to appropriately reflect this.*  
  
  
Minor points:  
1.      Line 130: What do the authors mean by negative controls? healthy volunteers? If so that should be made clear.  
2.      Line 62: Not sure what the authors mean by "openly" published metatranscriptome sequences... I would re-consider the use of "openly" in this sentence.

*“can you rephrase this to say something like Published in the open domain”*

3.      Figure 4: No depiction of the color codes for survived vs deceased.  
*Thank you for point this out. We have made ammendements to the figure annotation to more clearly depict the findings portrayed in this figure.*  
  
**Reviewer 4**  
Title: Analysis of Bronchoalveolar Lavage Fluid Metatranscriptomes Among Patients with COVID-19 Disease  
  
With the observation of potentially relation between the BALF microbiome and the severity of COVID-19 disease onset and progression, Jochum et al evaluate significant differences between BALF metatranscriptomes from COVID-19, CAP, and uninfected cohorts, as well as COVID-19 morbidity and mortality.

Though few metatranscriptome sequences from BALF of patients with COVID-19 disease have been published earlier as per several references, this paper specifically evaluated the microbial taxonomic and functional profiles of the BALF metatranscriptomes.

Overall, the attempt is good but the writing part is confusing especially in result section. There are some issues that need to be clarified and some corrections are recommended for providing clearer information.

Some of the sentence structures in the results when comparison was done between COVID-19 cohort and CAP/uninfected can be enhanced to improve the flow of the paper and to help readers grasp the content better.  
  
In particular, I have listed the following comments in detail here:  
1.      All of the names and terms should be completely mentioned for the first time along with abbreviations, for example, line 31: CAP should also be written as it's not cleared in line 37.  
2.      Line no-63-64 "Early in the SARS-CoV-2 outbreak........... from taking place". Add few references  
3.      Table 1: under data for gender i.e. female, male and unspecified, what is the denominator? The analysed data is not clarified.  
4.      Method section seems to be confusing. Rewrite it and make other sub headings too.  
5.      Line 133 "and n=32 COVID-19 samples from 18 subjects were available for comparison". It is not cleared.

**“can you rephase this to say something like *(n samples were derived from the same subject)”***  
6.      In result section: Line 161-162 "This revealed 35 out of 13,534 GO terms were associated with patients with COVID-19 when compared to patients with CAP or uninfected control subjects (Figure 1, Table 2, Table 3)" This means 35 GO terms were associated with COVID-19 patients as compared to patients with CAP and 35 GO term as compared to uninfected control subjects. This sentence is not clarified. I suggest to rewrite a sentence mentioning results as per table 2 and table 3 separately and then comparing data of COVID-19 with both CAP and Uninfected control.

**Can you break this into two different sentences and use the number of rows from tables 2 and 3 to report the *n* of sig. GO terms for each cohort comparison?**

7.      Line 192-194 "Taxonomic comparisons of the COVID-19 cohort to uninfected and CAP cohorts revealed a statistically significant decrease of several microbial genera within the phylum of Proteobacteria, including those of the families Sphingomonadaceae I assumed from this line that there was a significant decrease of several microbial genera like Sphingomonas in COVID-19 cohort than CAP/Uninfected? It is confusing. As per the discussion, significant increase of microbial genera among BALF specimens from COVID-19 patients than CAP. ***I will look into this***

8.      As the author observed, the distinct taxonomic features of BALF specimens from the COVID-19 vs. uninfected vs. CAP analysis included an increase in the genus Sphingomonas, among COVID-19 patients.

This non fermenting Gram-negative bacilli is known as the most widespread cause of nosocomial infections and known to be of minor clinical significance.

The significant increase of this pathogen in COVID -19 patients as observed should be correlated to clinical conditions before driving such results. As many factors like Hospitalization of COVID-19 patients and CAP depends on mechanical ventilation, poor infection control practices in hospitals, length of stays in hospital, medication like steroids etc affect the microbial genera and this increase may be due to hospital acquired infection. Do the authors have more thoughts on this field?

***Can you state something about how we are unable to differentiate the findings associated with the disease or its treatment***

9.      Line 313-315: This serves to protect cells from the toxic effects of ....... undergoing great amounts of oxidative stress. Any study published to support this?

***Can you find and input a couple citations that support this statement. I will also look***

10.     Author should also correlated the GO terms associated with **COVID-19 mortality** with the pathways, whether any modulation of metabolic pathway observed.

11.     A precise conclusion is needed, as it's too short in its current form. Hence, please add a significant statement that must be structured as, "what was offered by authors? Presently it looks like a section of limitations.  
12.     Some typographical errors such as two words without spaces in-between are seen throughout the manuscript.