Cell Genomics Transparent Peer Review Record

Altered and allele-specific open chromatin landscape reveals epigenetic and genetic regulators of innate immunity in COVID-19

Bowen Zhang[1,2,3] †, Zhenhua Zhang[1,2,4,5] †, Valerie A. C. M. Koeken[1,2,6] †, Saumya Kumar[1,2], Michelle Aillaud[7], Hsin-Chieh Tsay[1,2], Zhaoli Liu[1,2], Anke R.M. Kraft[1, 2, 8, 9, 10], Chai Fen Soon[2,8], Ivan Odak[11], Berislav Bošnjak[11], Anna Vlot[12], Deutsche COVID-19 OMICS Initiative (DeCOI), Morris A. Swertz[4,5], Uwe Ohler[12], Robert Geffers[13], Thomas Illig[14,15], Jochen Huehn[10,16], Antoine-Emmanuel Saliba[17], Leif Erik Sander[14,18], Reinhold Förster[9,10,11], Cheng-Jian Xu[1,2,6,8], Markus Cornberg[1, 2, 8, 9, 10], Leon N. Schulte[7,14], Yang Li[1, 2, 6, 10, 19] *

Summary

Initial submission:	Received : 5/31/2022
	Scientific editor: Laura Zahn
First round of review:	Number of reviewers: 2 Revision invited : 7/17/2022 Revision received : 10/21/2022
Second round of review:	Number of reviewers: 1 Accepted : 11/17/2022
Data freely available:	Yes
Code freely available:	Yes

This transparent peer review record is not systematically proofread, type-set, or edited. Special characters, formatting, and equations may fail to render properly. Standard procedural text within the editor's letters has been deleted for the sake of brevity, but all official correspondence specific to the manuscript has been preserved.

Cell Genomics Transparent Peer Review Record

Referees' reports, first round of review

Referee #1:

The work from Bowen Zhang et al describes epigenetic and biological features associated with sars-cov2 infection in patients developing severe covid-19 syndrome. The work reports about the identification of transcription factors involved in the inflammatory response against the coronavirus infection and identifies chromatin structure modification associated with gene expression in immune cells responsible for cytokines production. In addition the role of the lncRNA LUCAT1 is investigated and found functionally associated with the inflammatory activation of monocytes in the presence of reduced oxygen pressure.

This work is state of art. It has been well designed and performed and, although probably of specialist interest, it provides original information about the epigenetic changes occurring in immune cells of Covid-19 patients developing severe symptoms.

I am enthusiast about this work and I have not further questions for the authors.

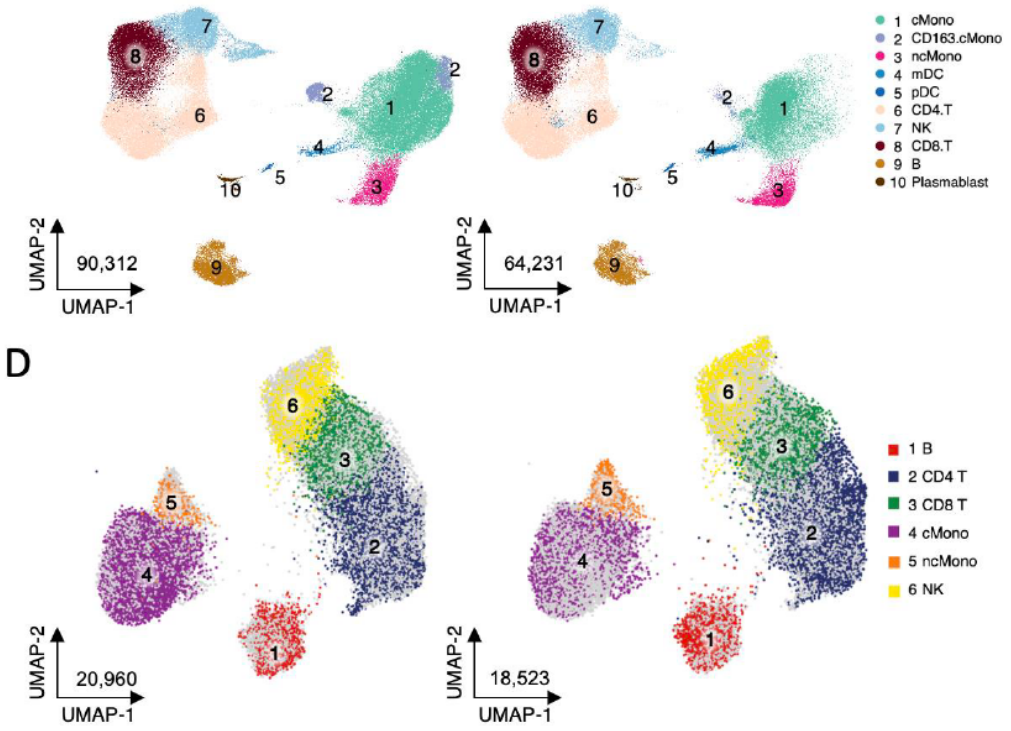
Referee #2:

Comments enter in this field will be shared with the author; your identity will remain anonymous.
Comments to the Author：
This paper comprehensively described an altered and allele-specific open chromatin landscape of COVID-19 in innate immunity. Based on a multi-omics integration, the authors did comparative analysis of PBMCs isolated from hospitalized (severe and mild) and convalescent COVID-19 patients using single-cell RNA-seq, single-cell ATAC-seq and genotype integration. The author's research objectives were clearly defined. Some of the findings are novel and valuable for better understanding of the genetic and transcriptional regulation of dysregulated immune responses in COVID-19. Strikingly, this study can provide a potential intervention strategy in detrimental immunity to reduce viral loads in lung and risk of COVID-19 hospitalization.
However, the authors might still need consider the following points to improve the manuscript even further:
Major
1. For Fig 1C, Fig 1D, Fig 3A, Fig 3D and Supplementary Figures, cell types or cell clusters should be labeled in UAMP, since it is hard to distinguish these cell populations due to low discrimination of colors.
2. There are some descriptions about Fig 1E that confuse me. Given that transcriptome and epigenome represent different biological processes, the authors should describe differences in scRNA-seq data and scATAC-seq in detail, respectively. In addition, T cells (in line 164) should be replaced to CD4+T or CD8+T cells because NKT and γδ T cells are not analyzed.
3. In line 175-177, "the highest" should represent one cell type rather than three cell types. To avoid ambiguity, number of DE genes should be labeled in Fig S4B.
4. In line 184-189, some genes were not shown in the figures such as IFI6, IFI30 and IFI44L. The resulting description should be in keeping with the original figures.
5. In line 287-288, this title does not sum up or summarize the main idea in Fig 3 because most of the results describe scRNA and scATAC profiles. Therefore, I suggest the title of Fig 3 in Results should be redrawn.
6. In line 298-300, there is no evidence to suggest that LUCAT1 is highly expressed in R2 and R8. Additionally, a featureplot of IFI30 should be showed to follow the original faithfully.
7. To show the cell proportions across different disease conditions more concretely, UMAP plots showing three disease conditions should be added.

8. In line 360-363, a regulon analysis of C4 might be performed to predict a combination of TFs. scRNA-seq regulons enrichment and scATAC-seq TFs-motif enrichment can identify some key TFs between convalescent and hospitalized COVID-19 patients.
9. In line 379-381, risk variants on chr3, 12, 17, and 21 were found to be located in several open chromatin peaks. The methodology for identifying risk variants on these chromes should be detailed in Methods section.
10. In line 414-415, "suggesting that infection response genes are regulated by chromatin accessibilities of genetic variants." This conclusion is very confusing, it is really hard to understand the rationale in this conclusion. Same as line 421-422 and Figure S9D-9F.
11. In line 474-477, more details on identifying DPP9 as a candidate gene for COVID-19 severity should be added.
Minor
1. The author should reorganize the logic of this paper to increase readability.
2. The format of pictures should be uniform and neat, including font size, image resolution and image resolution.

Authors' response to the first round of review

Reviewer #1: The work from Bowen Zhang et al describes epigenetic and biological features associated with sars-cov2 infection in patients developing severe covid-19 syndrome. The work reports about the identification of transcription factors involved in the inflammatory response against the coronavirus infection and identifies chromatin structure modification associated with gene expression in immune cells responsible for cytokines production. In addition the role of the lncRNA LUCAT1 is investigated and found functionally associated with the inflammatory activation of monocytes in the presence of reduced oxygen pressure.
This work is state of art. It has been well designed and performed and, although probably of specialist interest, it provides original information about the epigenetic changes occurring in immune cells of Covid-19 patients developing severe symptoms.
I am enthusiast about this work and I have not further questions for the authors.
Answer. We thank the reviewer for the positive assessment.
Reviewer #2: Comments enter in this field will be shared with the author; your identity will remain anonymous.
Comments to the Author :
This paper comprehensively described an altered and allele-specific open chromatin landscape of COVID-19 in innate immunity. Based on a multi-omics integration, the authors did comparative analysis of PBMCs isolated from hospitalized (severe and mild) and convalescent COVID-19 patients using single-cell RNA-seq, single-cell ATAC-seq and genotype integration. The author's research objectives were clearly defined. Some of the findings are novel and valuable for better understanding of the genetic and transcriptional regulation of dysregulated immune responses in COVID-19. Strikingly, this study can provide a potential intervention strategy in detrimental immunity to reduce viral loads in lung and risk of COVID-19 hospitalization.
Answer. We thank the reviewer for the positive assessment.
However, the authors might still need consider the following points to improve the manuscript even further:
Major
1. For Fig 1C, Fig 1D, Fig 3A, Fig 3D and Supplementary Figures, cell types or cell clusters should be labeled in UAMP, since it is hard to distinguish these cell populations due to low discrimination of colors.
Answer. We thank the reviewer for the suggestion and apologize for the ambiguity. We have now added the numbers of each cluster on the UMAP and the legends (also with numbering) to help distinguish the cell populations. The updated figure is also shown below:



2. There are some descriptions about Fig 1E that confuse me. Given that transcriptome and epigenome represent different biological processes, the authors should describe differences in scRNA-seq data and scATAC-seq in detail, respectively. In addition, T cells (in line 164) should be replaced to CD4+T or CD8+T cells because NKT and γδ T cells are not analyzed.

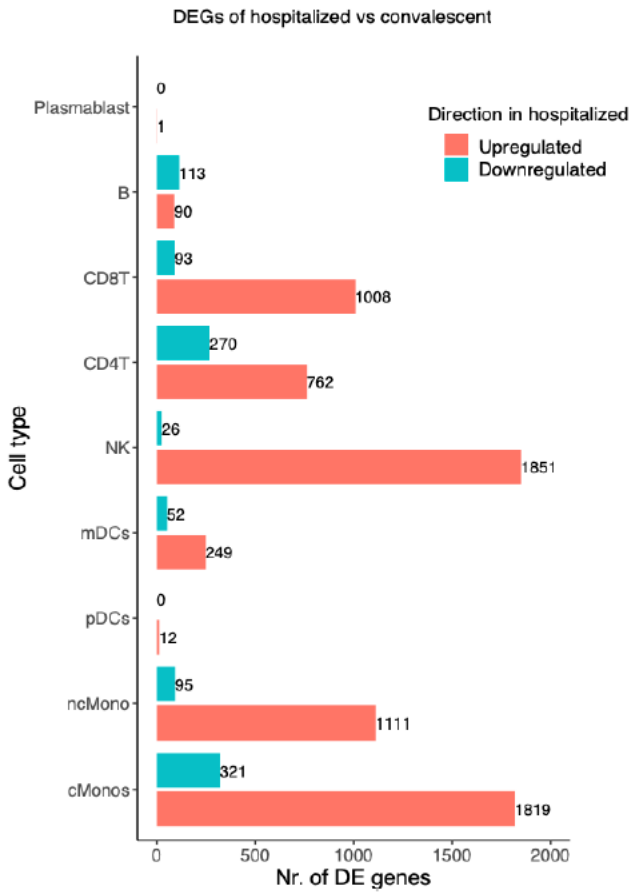
Answer. We apologize for the confusion and many thanks for the suggestions. We have now updated the text of Fig 1E to better describe the CD4+ and CD8+ T cells on page 5, and also shown below:

"…. a lower abundance of non-classical monocytes as well as CD4+ and CD8+ T cells in hospitalized COVID-19 compared to convalescent patients in both datasets (Figure 1E, Figure S4A, Dirichlet regression test, false discovery rate (FDR) adjusted P < 0.05) …."

Since the cell proportions described here represent the relative abundance of these cells in the PBMC samples, the estimates from two technologies (i.e. scRNA-seq and scATAC-seq) are expected to show a similarity, and they are both in line with the other studies on COVID-19 patients (Patterson et al. Immunology, 2021). We agree with the reviewer that the differences in the estimated cell numbers between scRNA-seq data and scATAC-seq data could be due to biological variation, to some extent. However, technical variation can be one of important factors, too. The differential contribution of biological and technical factors to the cell counts estimation is beyond the scope of this study, as we here focus on a robust identification of the cell types that differ in abundance between patient groups (e.g. hospitalized vs non-hospitalized).

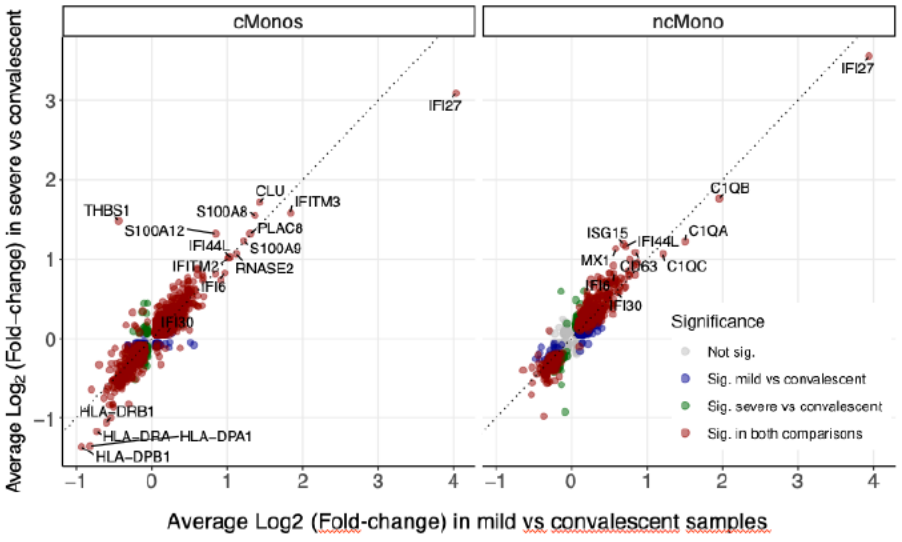
3. In line 175-177, "the highest" should represent one cell type rather than three cell types. To avoid ambiguity, number of DE genes should be labeled in Fig S4B.

Answer. We thank for the suggestion. We have replaced "the highest number of" with "large number of", and we have added number of DE genes into Figure S4B (also shown below)



4. In line 184-189, some genes were not shown in the figures such as IFI6, IFI30 and IFI44L. The resulting description should be in keeping with the original figures.

Answer. We apologize for the missing information and thank for the suggestion. We have now added the label of these genes in Figure 1F, which is also shown below.



5. In line 287-288, this title does not sum up or summarize the main idea in Fig 3 because most of the results describe scRNA and scATAC profiles. Therefore, I suggest the title of Fig 3 in Results should be redrawn.

Answer. Many thanks for the suggestion and we fully agree with it. We have revised the title of this part as following:
"Single cell RNA and ATAC profiles revealed altered C/EBP regulation in a monocyte subset associated with oxygen supply of COVID-19 patients".
6. In line 298-300, there is no evidence to suggest that LUCAT1 is highly expressed in R2 and R8. Additionally, a featureplot of IFI30 should be showed to follow the original faithfully.
Answer. We apologize for the confusion. We fully agree with the reviewer that LUCAT1 expression in Fig.3B, is not obviously high. This is due to the fact the scale of the colors was set for all genes (including LUCAT1) and the expression of LUCAT1 was relatively low compared to the other genes shown in Fig.3B. Actually, LUCAT1 is statistically higher expressed in R2 and R8 compared to the other clusters (data shown in the original Table S6). We have now updated the description on page 8 to avoid misunderstanding as follow:
"Through the DE tests comparing expression of gene between one cluster to the rest of clusters and visualization of selected marker gene expression in UMAPs (Table S6, Figure 3B), we identified....."
7. To show the cell proportions across different disease conditions more concretely, UMAP plots showing three disease conditions should be added.
Answer. We thank the reviewer for the suggestion. We have now added the UMAP plots of three disease conditions in Fig. S8A, which are also shown below:

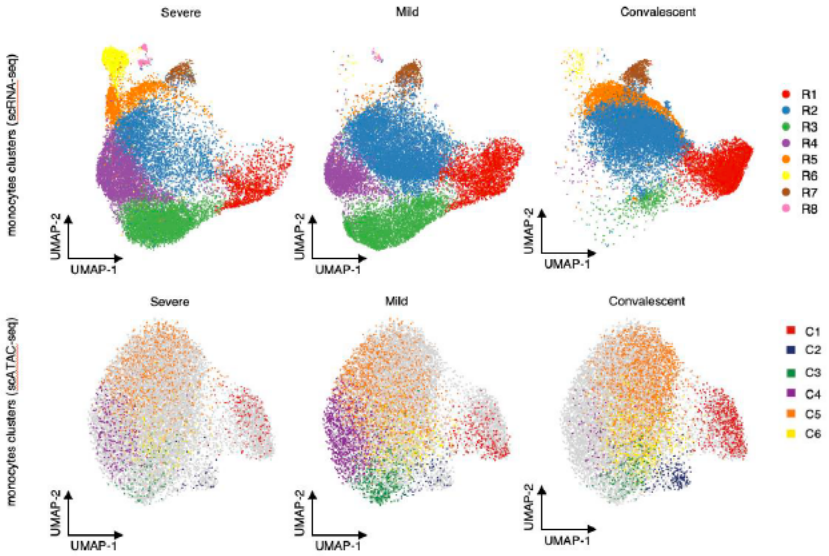


Figure S8. A. UMAPs showing the distribution of monocytes subclusters across severe, mild and convalescent patients from the scRNA-seq (top row) and scATAC-seq (bottom row), respectively.

Figure S8. A. UMAPs showing the distribution of monocytes subclusters across severe, mild and convalescent patients from the scRNA-seq (top row) and scATAC-seq (bottom row), respectively.
8. In line 360-363, a regulon analysis of C4 might be performed to predict a combination of TFs. scRNA-seq regulons enrichment and scATAC-seq TFs-motif enrichment can identify some key TFs between convalescent and hospitalized COVID-19 patients.
Answers. We appreciate the reviewers for raising this interesting point. Since the R4/C4 cluster was specific to hospitalized COVID-19 patients, we have performed the TFs-motif enrichment analysis across the scATAC-seq cell clusters in the original manuscript and identified CEBPA, CEBPB, CEBPD, CEBPG, CEBPE, and ATF4 as key TFs specifically open chromatin in C4 cluster (original Figure 3G, updated Figure S8B).
In order to confirm this in scRNA-seq data. We firstly followed the suggestion of the Reviewer to apply a regulon enrichment analysis on R4 cluster cells via SCENIC [Aibar, S. et al., Nat. Methods. 2017]. This independent analysis based on scRNA-seq data identified both C/EBP TFs and ATF4 as the enriched TFs with high confidence, together with IRF4, FOS, JUNB, JUND, BACH1, etc. (Please see Table S7 for all identified enriched TFs from this analysis). These identified TFs agree with the enriched motifs in C4 cluster cells based on scATAC-seq data (red star marked in updated Figure 3G, also see below).

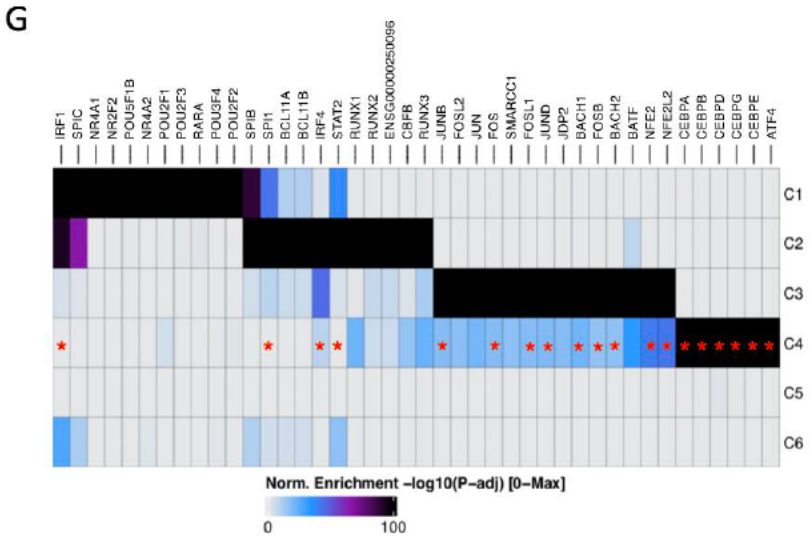


Figure 3G. Heatmap showing the significantly enriched TF motifs in the open chromatin peaks of each monocytes sub-cluster; TFs that were also enriched as regulon in R4 cluster cells by SCENIC are marked with red stars.

Next, we estimated the percentage of R4 cluster cells with each of the motif-enrichment-identified TFs expressed from hospitalized COVID-19 patients, and obs

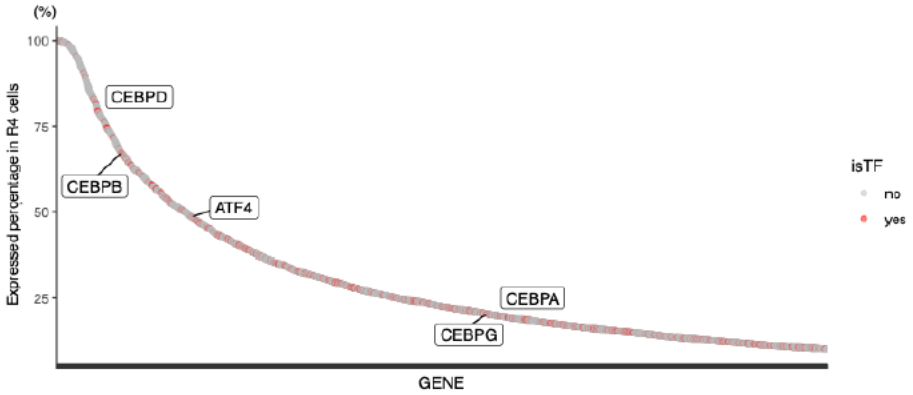


Figure S8B. Gene expressed percentage of cells from hospitalized COVID-19 patients in R4 clusters. TF genes are marked in RED.

Based on the above results, we have added the following text to the Results part of the revised manuscripts on page 9-10:
From the scRNA-seq data, we found that among these TFs, CEBPD, CEBPB, and ATF4 were also widely expressed in R4 cells (Figure S8B). Through an independent TF regulon enrichment analysis in R4 cluster cells (Aibar, S. et al., Nat. Methods. 2017), we've confirmed that the identified C/EBPs and ATF4 were also highconfidently
enriched TFs, together with IRF4, FOS, JUNB, JUND, BACH1 etc (red stars marked in Figure 3G, see all enriched TFs in Table S7).
The description of above analysis has been added to the STAR Method part of the revised manuscript on page 24:
TF gene expression and regulon enrichment analysis
In order to estimate the expression of the motif-enriched TFs in R4/C4 monocyte cluster, we ranked the genes based on their expressed percentages of R4 cells from hospitalized COVID-19 patients in scRNA-seq dataset and marked out TF genes among them. Next, we applied a regulon enrichment analysis across genes that were expressed at least 10% of R4 cells with SCENIC (Aibar, S. et al., Nat. Methods. 2017). Then, we intersected the enriched TFs, that were marked with 'high confidence' annotations by the algorithm, with the TF-motif

enrichment results and marked the overlap TFs on the Heatmap.

9. In line 379-381, risk variants on chr3, 12, 17, and 21 were found to be located in several open chromatin peaks. The methodology for identifying risk variants on these chromes should be detailed in Methods section.

Answer. We thank the reviewer for this suggestion. We have now added the following information in the STAR Methods:

Chromatin accessibility of COVID-19 risk variants

The genetic variants with a reported P-value < $5x10^{-8}$ from the COVID-19 GWAS summary statistics (Hospitalized covid vs. population" release 6) by HGI (COVID-19 Host Genetics Initiative 2021) were considered as risk variants of COVID-19. An open-chromatin peak was regarded to be associated with risk variants if its genomic location overlaps with at least one significant variant. The over-representation of "riskvariants-overlapping" peaks was estimated by the Fisher's exact test comparing between peaks found in hospitalized patients and convalescent samples in each cell type, respectively.

10. In line 414-415, "suggesting that infection response genes are regulated by chromatin accessibilities of genetic variants." This conclusion is very confusing, it is really hard to understand the rationale in this conclusion. Same as line 421-422 and Figure S9D-9F.

Answer. We apologize for the confusion here and we have now revised this sentence as follow:

"..... suggesting that the genetic risk variants have an impact on the transcriptional responses to SARS-CoV-2 infection through allele-specific chromatin accessibilities".

11. In line 474-477, more details on identifying DPP9 as a candidate gene for COVID-19 severity should be added.

Answer. Thanks for this suggestion and we have added detailed information on identifying DPP9 as a candidate gene and rephrase the paragraph into the following on page 12-13:

11. In line 474-477, more details on identifying DPP9 as a candidate gene for COVID-19 severity should be added.

Answer. Thanks for this suggestion and we have added detailed information on identifying DPP9 as a candidate gene and rephrase the paragraph into the following on page 12-13:

"Another interesting example of potential regulation programs for DPP9, a candidate gene for COVID-19 severity, is depicted in Figure S9I-K. The DPP9 genes harbors SNPs associated with COVID-19 ($P < 5 \times 10^{-8}$) and was prioritized as a candidate gene that is involved in host-driven inflammatory lung injury in severe patients (Pairo-Castineira et al. 2021). Also, this locus has been previously reported to be associated with fibrotic idiopathic interstitial pneumonias (Fingerlin et al. 2013), which suggests the potential role of dipeptidyl peptidase 9 (the enzyme encoded by DPP9 gene) in severe COVID-19 patients. Moreover, early studies reported the enzyme involves in antiviral signaling pathways (H. Zhang et al. 2015), antigen presentation (Geiss-Friedlander et al. 2009), and the activation of inflammosome (Griswold et al. 2019). Taken together, these data depict an epigenetic regulation effect of risk allele of DPP9 locus in severe COVID-19 patients."

Minor

1. The author should reorganize the logic of this paper to increase readability.

Answer. Thanks to the reviewer for the suggestion. We have revised the manuscripts to be more logical, e.g. on page 5, 6 and page 12.

2. The format of pictures should be uniform and neat, including font size, image resolution and image resolution.

Answer. We agree with the reviewer and thanks for the suggestion. We have now uniformed the texts and labels in the main figures.

Referees' report, second round of review

Referee #1:

None

Referee #2:

The authors have answered all my concerns, I have no further comments.

Cell Genomics Transparent Peer Review Record

Authors' response to the second round of review