nature research

Peer Review Information

Journal: Nature Microbiology
Manuscript Title: **Detection of Respiratory syncytial virus defective genomes in nasal secretions is associated with distinct clinical outcomes**
Corresponding author name(s): Carolina Lopez

Reviewer Comments & Decisions:

Decision Letter, initial version:

Dear Dr. López,

Thank you for your patience while your manuscript "Detection of defective viral genomes in nasal secretions predicts RSV disease severity in children and adults" was under peer-review at Nature Microbiology. It has now been seen by 4 referees, whose expertise and comments you will find at the of this email. Although they find your work of some potential interest, they have raised a number of concerns that will need to be addressed before we can consider publication of the work in Nature Microbiology.

In this particular case, referee #1 and #4 ask why the PCR products have 2 distinct sizes but the RNAseq shows many different break and rejoin points. Referee #1, #2 and #4 point out that there is no quantitative correspondence between RNAseq and PCR-based detection of DVGs. Referee #1 wants you to be more conservative about statements implying causation in the results section and clarifying the evidence for causation in the discussion section, and referee #2 also thinks that some statements need to be toned down. Referee #2 wants to know whether DVGs levels increase with viral load, wants you to show data showing whether DVGs level correlate with total viral reads in the DVG+ patients whose samples were analysed via RNAseq, thinks that is critical that you determine a cutoff to distinguish DVG positive form negative sample, wants you to show DNA gel results of all DVG+ and DVG- samples, and points out that there is a negative control missing. Referee #2 also thinks that the conclusion that outcome and/or altered host response is linked to DVGs is problematic, asks about the expression levels of the host factors and cytokines in the patient population, wants you to provide an explanation about the scores for clinical outcome, asks about quantitative differences in 4 samples and asks whether viral load and DVGs level of L5 correlate, asks what is the level of DVGs in the prolonged DVGs group and the transient DVGs group, asks what was the RSV subtype of the patients in the various cohorts, and wants you to state the importance/impact of the study in the abstract, discuss the clinical implications in the discussion and how you view the use of this candidate biomarker in clinical practice. Referee #3 thinks that it may be the active virus as well as or instead of the DVG that is driving the pathogenesis and wants you to provide further investigation into the mechanism between these hospitalized and non-hospitalized groups, more data to resolve how the DVGs may provide for an altered response and to clarify how the ability of CVG level would further extend prediction past viral titer levels. Referee #4 wants you to provide kinetic information about cohort I, and asks whether all patients from cohort 1 were RSV naïve.

Please be aware that we will not be able to evaluate whether a revised version will be sent back to the original reviewers until we have read it.

Should further experimental data allow you to address these criticisms, we would be happy to look at a revised manuscript.

We are committed to providing a fair and constructive peer-review process. Please do not hesitate to contact us if there are specific requests from the reviewers that you believe are technically impossible or unlikely to yield a meaningful outcome.

We strongly support public availability of data. Please place the data used in your paper into a public data repository, if one exists, or alternatively, present the data as Source Data or Supplementary Information. If data can only be shared on request, please explain why in your Data Availability Statement, and also in the correspondence with your editor. For some data types, deposition in a public repository is mandatory - more information on our data deposition policies and available repositories can be found at https://www.nature.com/nature-research/editorial-policies/reporting-standards#availability-of-data.

Please include a data availability statement as a separate section after Methods but before references, under the heading "Data Availability". This section should inform readers about the availability of the data used to support the conclusions of your study. This information includes accession codes to public repositories (data banks for protein, DNA or RNA sequences, microarray, proteomics data etc…), references to source data published alongside the paper, unique identifiers such as URLs to data repository entries, or data set DOIs, and any other statement about data availability. At a minimum, you should include the following statement: "The data that support the findings of this study are available from the corresponding author upon request", mentioning any restrictions on availability. If DOIs are provided, we also strongly encourage including these in the Reference list (authors, title, publisher (repository name), identifier, year). For more guidance on how to write this section please see: http://www.nature.com/authors/policies/data/data-availability-statements-data-citations.pdf

If revising your manuscript:

* Include a "Response to referees" document detailing, point-by-point, how you addressed each referee comment. If no action was taken to address a point, you must provide a compelling argument. This response will be sent back to the referees along with the revised manuscript.

* If you have not done so already we suggest that you begin to revise your manuscript so that it conforms to our Article format instructions at http://www.nature.com/nmicrobiol/info/final-submission. Refer also to any guidelines provided in this letter.

* Include a revised version of any required reporting checklist. It will be available to referees (and, potentially, statisticians) to aid in their evaluation if the manuscript goes back for peer review. A revised checklist is essential for re-review of the paper.

Please use the link below to submit a revised paper:

{REDACTED}

Note: This url links to your confidential homepage and associated information about manuscripts you may have submitted or be reviewing for us. If you wish to forward this e-mail to co-authors, please delete this link to your homepage first.

Nature Microbiology is committed to improving transparency in authorship. As part of our efforts in this direction, we are now requesting that all authors identified as 'corresponding author' on published papers create and link their Open Researcher and Contributor Identifier (ORCID) with their account on the Manuscript Tracking System (MTS), prior to acceptance. This applies to primary research papers only. ORCID helps the scientific community achieve unambiguous attribution of all scholarly contributions. You can create and link your ORCID from the home page of the MTS by clicking on 'Modify my Springer Nature account'. For more information please visit please visit www.springernature.com/orcid.

If you wish to submit a suitably revised manuscript we would hope to receive it within 6 months. If you cannot send it within this time, please let us know. We will be happy to consider your revision, even if a similar study has been accepted for publication at Nature Microbiology or published elsewhere (up to a maximum of 6 months). Should your manuscript be substantially delayed without notifying us in advance and your article is eventually published, the received date would be that of the revised, not the original, version.

In the meantime we hope that you find our referees' comments helpful.

****
Reviewer Expertise:

Referee #1: Defective viral genomes
Referee #2: RNA seq
Referee #3: Respiratory syncytial virus and human studies
Referee #4: Paramyxovirus pathogenesis

Reviewer Comments:

Reviewer #1 (Remarks to the Author):

With the goal of evaluating the role of defective viral genomes (DVGs) in respiratory syncytial virus infection and disease, Sun et al. examined three cohorts of individuals, comprising hospitalized, non-hospitalized, and experimentally infected humans. Nasal swab specimens were analyzed for the presence of DVGs and results were considered in the context of viral loads and disease severity. The results show an association between DVG presence and severe outcomes in hospitalized children. The relationship between DVG presence and pathogenesis was more nuanced in non-hospitalized patients and in experimentally infected individuals: Individuals with early, transient DVG presence tended to have low viral loads and mild disease. Late or prolonged DVG presence was associated with more severe disease. The authors conclude that the kinetics of DVG appearance, as well as their duration, impact RSV disease severity in humans.

This is an exciting paper. The demonstration that diverse DVGs are routinely generated de novo in the context of human RSV infection clearly lays to rest any lingering belief that DVGs are an artifact of virus culture in the lab. Furthermore, evidence that the timing of DVG production impacts the course of disease is novel and important for two reasons. First, the clinical impact of DVGs needs to be understood with the goal of improving treatment of RSV. Second, the biological implications of DVGs generated by diverse viruses has long been elusive and advances in this area are critical for a full understanding of the within-

host dynamics of acute viral infections.

Specific comments:

1. Fig. 1B: Why are the PCR products of two distinct sizes (1000 and 300 bp) when the RNAseq shows many different break and rejoin points?

2. Fig. 1: Although the two methods agree in terms of presence or absence, quantitative correspondence between RNAseq and PCR-based detection of DVGs is lacking. The authors should comment on this – which assay is more quantitative?

3. Figure 2: Labeling of clusters in panel F is very subtle and it therefore took me a while to figure out what this panel was showing.

4. Page 12: "In all three cutoffs clearance must have occurred within the first 6 days pi to meet the Early criteria." At this point in the text it is not clear whether "clearance" refers to clearance of virus or DVGs. It would also be helpful to explain the rationale for including this criterion in the definition for Early.

5. The data show interesting associations between DVG patterns and disease severity / viral load, but causation is not clear looking solely at the data presented here. The causative link between DVGs and viral growth or immune responses comes from prior work in mice or culture systems. This could be better handled by being more conservative about statements implying causation in the results section and clarifying the evidence for causation in the discussion section.

6. Fig. 6F is missing a Y axis label.

7. Demonstration that DVGs were generated de novo in experimentally infected subjects is important. These data (Fig 7) are strong - indicating very little DVG content in the viral stock and the presence of unrelated and highly diverse DVGs in subjects.

8. As mentioned in the discussion, the data reported provide a strong impetus for prospective studies designed to monitor DVG levels in naturally infected individuals over time. Indeed, I expect the findings reported here will stimulate many follow-up studies.

Reviewer #2 (Remarks to the Author):

In the manuscript Sun et al. investigate the association between defective viral genomes (DVGs) generated during RSV infection with disease severity and host immune responses in different cohorts. The authors used PCR and RNA-seq on 7 samples and then PCR only, for detection of DVGs in patient samples. They correlate presence and kinetics of DVGs with viral loads and diseases severity. It is concluded that the kinetics of DVGs is associated with disease severity, and specifically that earlier appearance of DVGs predicts less severe disease.

There is a large need for biomarkers to predict disease severity in RSV infected individual, and thus the search for such factors could be of a high impact. The strength of the manuscript is the utilization of different pediatric and adult cohorts. However, there are several important limitations including

overstated conclusions, lack of quantitative measurements, and inability to demonstrate that DVGs is independent of viral load. Moreover, the study is descriptive and lacks mechanistic studies showing how DVGs modify expression of host factors and disease severity.

Major comments:

1. It is possible that the presence and level of DVGs is a confounder of viral load, which was shown in some studies to be associated with severity. Does DVGs level increase with viral load? Did DVGs level correlate with total viral reads in the DVG+ patients whose samples were analyzed via RNA-seq data? It will be important to show these data.

2. Multiple conclusions in the manuscript are overstatements, as the authors are not looking at causality but rather association between DVGs and severity. For example, the conclusion stated in line 104: "Early appearance of cbDVGs after infection results in low viral loads and mild disease". Similarly, the statement: To directly test the impact of cbDVGs on clinical outcomes in hospitalized children.." in line 154 is wrong. You are only looking at associations. Line 164: "These results suggest that RSV-infected hospitalized pediatric patients that test positive for cbDVGs fail to control the viral infection at the time of sampling leading to more severe disease outcomes." All these and other statement have to be toned down.

3. Figure 1 – the determination of a cutoff to distinguish DVG positive from negative sample is critical, yet not rigorously addressed.
- The authors set a cutoff based on the RNA-seq data, but then use PCR as the measurement modality in the remaining experiments. How do the authors decide whether a sample is DVG+ or not without a quantitative control? It would be helpful to show DNA gel results of all DVG+ and DVG- samples in supplemental files.
- In Fig 1 B, two bands are shown, yet it is unclear, which band(s) represents DVGs? Also, this figure is missing a critical negative control showing that the PCR worked, which makes the interpretation of the data difficult.
- This is further complicated by the fact that there is lack of correlation between the level of DVGs seen in Fig 1B and DVG reads shown in Fig 1C (e.g.H83 showed high DVGs reads in RNA-seq but very low on gels).

3. Fig 2 – since viral load in DVG+ patients is significantly higher than in the DVG- group, the conclusion that outcome and/or altered host response is linked to DVGs is problematic.

4. Fig 3 – What were the expression levels of the host factors and cytokines shown in Fig 2 in this patient population? This will be important to determine, since in these patients viral load did not correlate with severity, yet presence of DVGs did.

5. Disease severity scores are being used in figs 3 and 4 to measure clinical outcome (vs. length of hospitalization and ICU time in fig 2). The authors should provide an explanation what these scores are. Moreover, the scores in figure 3 are all below 8, whereas in figure 4C the scores can be as high as 20. How are these calculated and why is there such a difference?

6. Figs 4E, 4F and 5A - Are there quantitative differences in the level of DVGs between L5 and the 3 other samples (L35, L1, L16)? It would be important to show DNA gels of DVGs for these samples. Specifically DVGs levels at day 2,3 and 7 for L5 relative to the other samples. Do viral load and DVGs

level of L5 correlate?

7. Fig 5 - The reason for choosing cutoff II is unclear and seems arbitrary. Also, the error bar in Fig. 5B is confusing, since the graph is plotted on a log scale in the y axis for virus load. The other half of the error bar should be added.

8. Fig. 6 – Again, an association between DVGs and viral load is demonstrated. The prolonged DVGs group has a higher viral load than transient DVGs. What is the level of DVGs in the prolonged DVGs group and the transient DVGs group?

9. What was the RSV subtype of the patients in the various cohorts? Since subtype A has been linked with disease severity, it will be important to determine whether presence/level of DVGs is greater with this specific subtype.

10. The authors should state the importance/impact of the study in the abstract. Moreover, they should discuss the clinical implications in the discussion and how they view the use of this candidate biomarker in clinical practice.

11. Figs 1 and 7 were missing. While Fig 1 was then uploaded, I didn't realize Fig 7 was also missing, until later.

Reviewer #3 (Remarks to the Author):

In these studies the results examine the correlation of DVGs as a predictor of severe RSV disease and provide data comparing both hospitalized and non-hospitalized individuals in 3 different cohorts. Furthermore, the studies used an experimental infection with M37 to examine the correlation of CVG with disease. In several studies it has been shown that DVGs can be an indication of higher RSV load during infection. In these studies the data indicate that higher DVGs can be found in both more severely infected hospitalized and in patients not hospitalized with non-severe infection. As suggested in the Discussion, these latter data suggest that it is not simply the viral titer, but how the individual responds to the higher viral load and that both groups, higher and lower DVG may respond poorly with the higher DVG hospitalized group having more "bad" response to the higher viral titer. Some of the logic to explain the results of finding the higher DVG in hospitalized and non-hospitalized patients is confusing and not supported by a mechanistic explanation, rather a more logical explanation that CVG may protect or promote pathogenesis depending upon the response of the individual. This reviewer would agree but could add that it may be the active virus as well as or instead of the CVG that is driving the pathogenesis. However, no further investigation into what the mechanism might between these hospitalized and non-hospitalized groups has been provided.
The Discussion offers a logical explanation of how the DVGs may provide for an altered response based upon previous data, but unfortunately do not provide any data to resolve these ideas further within these studies. Altogether, these studies provide interesting and potentially important data that demonstrate that CVG levels during RSV infection may predict the outcome of an infection but really don't bring any additional clarity as to why they may or may not drive a detrimental response and what the mechanism might be for the good or poor response. In addition, it is not clear with these data how the ability of CVG level would further extend prediction past viral titer levels.

Reviewer #4 (Remarks to the Author):

In this manuscript, samples from three clinical datasets (RSV-infected hospitalized pediatric patients, hospitalized and non-hospitalized children with RSV, and adults experimentally challenged with RSV) are analyzed for presence of defective-interfering particles (cbDVGs). Overarching goal is the identification of predictive biomarkers for progression of RSV disease to severe lower respiratory infection and viral pneumonia.
The current literature is not fully consistent on whether RSV load at the time of hospitalization can serve as a predictor of outcome. Additional and/or more predictive biomarkers that may impact disease management decisions are urgently needed. The study therefore aims to address an important problem. cbDVGs emerge spontaneously during mononegavirus replication, triggering activation of host innate immune response pathways. However, the correlation of cause and effect of cbDVGs in severe RSV disease is unclear. This study provides first insight into a potentially important biomarker, but revealed that the correlation is complex and dependent on the kinetic of cbDVG appearance during the cause of RSV infection must be considered. Unfortunately, kinetic information appears to be lacking for cohort 1, confounding interpretation of this important dataset. Information on relative time of patient sample collection for this cohort needs to be provided analogous to that shown, for instance, in figure S1 for cohort 2. Can cohort 1 be analyzed by subgroups with comparable sampling times?

Specific comments:
PCR-based and RNAseq-based detection of cbDVGs in samples from the same patients shows no quantitative correlation (figures 1B and 1C, relative amplicon intensity vs DVG reads). That is surprising and challenges the claimed reliability of the PCR-based method and thus a central experimental approach of the study.

Where the 122 banked nasal washes (cohort I) collected at comparable time points (i.e. prior to patient hospitalization, at the time of hospitalization, or a consistent number of days after hospitalization? Without standardization of sampling time, samples from patients with longer duration of hospitalization may have been collected at later stages of severe disease, enriching for presence of cbDVGs and compromising usefulness as biomarker for progression to LRI.

Were all patients from cohort 1 RSV naïve, or can this group by analyzed by RSV experienced and inexperienced patients?

Page 3, top paragraph: note that ribavirin has been approved for the treatment of RSV disease.

Reason for 2 different size amplicons in figure 2B? Does this reflect different break and/or rejoin points? Please clarify.

Author Rebuttal to Initial comments

We want to thank the reviewers for their helpful comments on our original submission. We have responded to all the suggestions and commentaries, and this has certainly resulted in an improved manuscript. We have marked in burgundy font those sections in the manuscript that directly answer

to the reviewers' concerns. Below is a detailed point by point response:

Reviewer #1(Remarks to the Author):

With the goal of evaluating the role of defective viral genomes (DVGs) in respiratory syncytial virus infection and disease, Sun et al. examined three cohorts of individuals, comprising hospitalized, non-hospitalized, and experimentally infected humans. Nasal swab specimens were analyzed for the presence of DVGs and results were considered in the context of viral loads and disease severity. The results show an association between DVG presence and severe outcomes in hospitalized children. The relationship between DVG presence and pathogenesis was more nuanced in non-hospitalized patients and in experimentally infected individuals: Individuals with early, transient DVG presence tended to have low viral loads and mild disease. Late or prolonged DVG presence was associated with more severe disease. The authors conclude that the kinetics of DVG appearance, as well as their duration, impact RSV disease severity in humans.

This is an exciting paper. The demonstration that diverse DVGs are routinely generated de novo in the context of human RSV infection clearly lays to rest any lingering belief that DVGs are an artifact of virus culture in the lab. Furthermore, evidence that the timing of DVG production impacts the course of disease is novel and important for two reasons. First, the clinical impact of DVGs needs to be understood with the goal of improving treatment of RSV. Second, the biological implications of DVGs generated by diverse viruses has long been elusive and advances in this area are critical for a full understanding of the within-host dynamics of acute viral infections.
A: We thank the reviewer for the positive comments on the impact of our work.

Specific comments:

1. Fig. 1B: Why are the PCR products of two distinct sizes (1000 and 300 bp) when the RNAseq shows many different break and rejoin points?
A: The PCR is designed in such a way that no PCR products can be amplified for the standard viral genome due to the location of the primer sequences as shown in Fig 1A. Previously published controls and validation of the PCR (Sun et al. 2015) include identification of major cbDVGs in viral stocks and RSV infected cells and confirmation by Sanger's sequencing and RNAseq. For negative controls, we used an RSV full-length viral backbone plasmid to do a PCR using P1 and P2 (Fig S1D, (Sun et al. 2015). Using the same PCR conditions, we did not detect any amplicons, indicating that the bands from positive samples were not from viral genome.

We observed from the VODKA output that cbDVGs that accumulate in patients are much larger than those that accumulate in tissue culture. Upon sequencing the 1000 and 300 bp bands obtained from these samples we noticed that the 1000bp and 300 bp bands represent two different fragments within cbDVGs, but we can't associate them to any specific cbDVG species. Moreover, each band represents a fragment that is present in most cbDVG species. Regrettably, these long cbDVGs contain long complementary structures that makes the amplification of their entire sequence very difficult by PCR. Therefore, we only use the PCR as a binary tool to indicate presence or absence of cbDVGs, and not to identify individual species. ***We have now included this explanation in the relevant results section of the manuscript (lines 295-310 and 496-499).***

To further validate our method, we have also added ***two new supplementary figures (S1B and S1C***) showing that specific primers can be designed based on RNAseq data to target junction regions (containing break/rejoin point) of specific cbDVG species present in human samples. We designed 3 primers that targeted 3 different cbDVG species. The amplicon sizes were as expected based on primer design and we confirmed the presence of the junction region for one of the primers by Sanger sequencing.

2. Fig. 1: Although the two methods agree in terms of presence or absence, quantitative correspondence between RNAseq and PCR-based detection of DVGs is lacking. The authors should comment on this – which assay is more quantitative?

A: This is a very good point. First to further confirm that the two methods agree in terms of presence and absence of cbDVGs, we have now performed RNA-Seq/VODKA to ***13 additional samples from Cohort 1*** and again observed that PCR+ samples contained more DVG reads than PCR- samples ***(new supplementary figure S1A)***. We think several reasons are leading to the inconsistency between total cbDVG counts (RNA-seq/VODKA) and intensity of cbDVG bands (PCR). In addition to intrinsic differences in the quantitative power of both methods, as explained above, our RT-PCR largely detects populations of cbDVGs rather than single species. Furthermore, factors such as different amplification efficiency for different cbDVG amplicons or different cbDVG populations in samples can cause this discrepancy between the intensity of specific cbDVG amplicons and total cbDVG reads. For these reasons we did not attempt to correlate both methods as quantitative tools. ***We have now added these explanations to the discussion of the manuscript (lines 306-310).***

3. Figure 2: Labeling of clusters in panel F is very subtle and it therefore took me a while to figure out what this panel was showing.

A: We are sorry for this oversight. ***We have now added the label "Gene Clusters" on top of figure 2F***.

4. Page 12: "In all three cutoffs clearance must have occurred within the first 6 days pi to meet the Early criteria." At this point in the text it is not clear whether "clearance" refers to clearance

of virus or DVGs. It would also be helpful to explain the rationale for including this criterion in the definition for Early.

A: Clearance refers to cbDVGs. We noticed that patients L5, L41 and L2 which had cbDVGs appearing early (day 2, 5 and 6 respectively) still had cbDVGs at the peak of infection (day 7) and sometimes even later. These patients had a very different clinical outcome than the other patients whose cbDVGs were detected early on but were not detected after D6 post infection indicating

that not only cbDVG appearance matters but also their disappearance. As a consequence, ***we defined "early cbDVG" as patients with cbDVGs appearing and disappearing before the peak of infection (day 7). We have added these clarifications to the results section (line 220).***

5. The data show interesting associations between DVG patterns and disease severity / viral load, but causation is not clear looking solely at the data presented here. The causative link between DVGs and viral growth or immune responses comes from prior work in mice or culture systems. This could be better handled by being more conservative about statements implying causation in the results section and clarifying the evidence for causation in the discussion section.
A: We thank the reviewer for these suggestions. These are valid points. In the results section, ***we have now made appropriate changes to the statements implying causation (multiple places in the manuscript)***. Furthermore, in the discussion section, to improve understanding of the link between cbDVGs and diseases severity, we now incorporated into our rational not only data from our current 3 cohort study but also data from previous studies establishing cbDVGs as the primary triggers for innate immune responses and a major driver of RSV pathogenesis *in vivo.*

6. Fig. 6F is missing a Y axis label.
A: We apologize for this oversight. The Y axis label has now been added.

7. Demonstration that DVGs were generated de novo in experimentally infected subjects is important. These data (Fig 7) are strong - indicating very little DVG content in the viral stock and the presence of unrelated and highly diverse DVGs in subjects.
A: We appreciate the reviewer's positive comment.

While processing new data to include in the manuscript we improved our RNAseq/VODKA pipeline by analyzing the entire viral genome for cbDVGs (previously last 3000bp) and have increased filtering stringency as to decrease false-positive cbDVG reads. ***All data included in this new version of the manuscript was re-analyzed using our improved bioinformatics methods***. Unfortunately, with our new analysis we did not find cbDVG reads in the 6 adult patient samples that we sequenced, most probably due to our newly improved filtering stringency. Although disappointing, these data is not surprising, given the much lower overall viral reads in adults patients compared to children.

Our new analysis concluded (when normalized to 10^8 total raw reads) that our laboratory virus HD (High DVG) stock contained 5007 cbDVG reads, however the RSV stock used to infect patients in cohort 3 contained only 135 cbDVG reads, which is 37 times less. Furthermore, 0.0787% of viral reads were cbDVG reads in our HD stock, whereas only 0.0006% of viral reads were cbDVG reads in the cohort 3 stock. Overall, this analysis still strongly suggests that the patients in

cohort 3 were infected with a virus stock with extremely few cbDVGs and that the cbDVGs detected in patients by PCR were very likely generated *de novo*. Because we cannot compare anymore the cbDVG species from the virus stock to the patient samples, we removed Figure 7 from this manuscript. We think that removal of this figure does not diminish or alter any of the major conclusions of this manuscript, including that cbDVG kinetics shape RSV disease outcome.

8. As mentioned in the discussion, the data reported provide a strong impetus for prospective studies designed to monitor DVG levels in naturally infected individuals over time. Indeed, I

expect the findings reported here will stimulate many follow-up studies.
A: Thank you.

Reviewer #2 (Remarks to the Author):

In the manuscript Sun et al. investigate the association between defective viral genomes (DVGs) generated during RSV infection with disease severity and host immune responses in different cohorts. The authors used PCR and RNA-seq on 7 samples and then PCR only, for detection of DVGs in patient samples. They correlate presence and kinetics of DVGs with viral loads and diseases severity. It is concluded that the kinetics of DVGs is associated with disease severity, and specifically that earlier appearance of DVGs predicts less severe disease.

There is a large need for biomarkers to predict disease severity in RSV infected individual, and thus the search for such factors could be of a high impact. The strength of the manuscript is the utilization of different pediatric and adult cohorts. However, there are several important limitations including overstated conclusions, lack of quantitative measurements, and inability to demonstrate that DVGs is independent of viral load. Moreover, the study is descriptive and lacks mechanistic studies showing how DVGs modify expression of host factors and disease severity.
A: We have placed close attention to the reviewer comments and did our best to address all the issues brought up. We are in agreement that this is not a mechanistic study per se. Instead, the goal of this study was to investigate whether the large number of previous studies from our laboratory and others addressing the function and mechanism of action of cbDVGs in vitro and in animal models have a functional correlate in natural infections of humans. The unique cohorts utilized in this study allowed to reveal the complexity of the impact of cbDVGs on infection outcome. Remarkably, the data presented here does not in any way invalidate previous mechanistic studies, and instead allow us to elaborate on the critical roles that the standard virus, cbDVGs, and host factors have in determining the clinical outcome of infection. In addition, we now provide new data that further support the predicted mechanisms behind DVG activity in humans.

Major comments:

1. It is possible that the presence and level of DVGs is a confounder of viral load, which was shown in some studies to be associated with severity. Does DVGs level increase with viral load? Did DVGs level correlate with total viral reads in the DVG+ patients whose samples were analyzed via RNA-seq data? It will be important to show these data.
A: We understand the reviewer's concerns. As shown in fig.4, we can detect cbDVGs in patients before the peak of viral replication and even in patients that never had detectable viral titer, indicating that cbDVG presence is not a direct consequence of viral load. In cohort 2 and cohort 3, cbDVG+ subjects had higher viral loads than cbDVG- subjects with a tendency of lower severity scores, indicating that differences in clinical outcomes among these groups cannot be explained by viral load (fig. 3 and fig. S5A). Furthermore, viral load was not significantly different between

non-hospitalized and hospitalized patients in cohort 2 (fig.S5C) and viral load did not correlate with disease severity in cbDVG+ patients from cohort 3 (fig.S5B). These data indicate that high viral load does not always associate with worse disease. A lack of correlation between viral load and clinical outcome during RSV infection has also been observed independently by other groups (Piedra et al. 2017). Although the reviewer asked for total viral reads as the estimation for viral

load, total viral reads contain also reads from cbDVG non-junction regions, and we cannot distinguish them from virus reads. Hence, unfortunately, it would be difficult to make any conclusions. ***More discussion has been included regarding the relationship between cbDVG, virus and disease in the manuscript (lines 311-314 and 325-327)***.

2. Multiple conclusions in the manuscript are overstatements, as the authors are not looking at causality but rather association between DVGs and severity. For example, the conclusion stated in line 104: "Early appearance of cbDVGs after infection results in low viral loads and mild disease". Similarly, the statement: To directly test the impact of cbDVGs on clinical outcomes in hospitalized children.." in line 154 is wrong. You are only looking at associations. Line 164: "These results suggest that RSV-infected hospitalized pediatric patients that test positive for cbDVGs fail to control the viral infection at the time of sampling leading to more severe disease outcomes." All these and other statement have to be toned down.

A: We thank the reviewer for these thoughtful comments. We have revised these statements as suggested.

3. Figure 1 – the determination of a cutoff to distinguish DVG positive from negative sample is critical, yet not rigorously addressed.

- The authors set a cutoff based on the RNA-seq data, but then use PCR as the measurement modality in the remaining experiments. How do the authors decide whether a sample is DVG+ or not without a quantitative control? It would be helpful to show DNA gel results of all DVG+ and DVG- samples in supplemental files.

A: Our apology for not explaining this clearly. We used PCR to screen for presence/absence of cbDVGs in our 3 cohorts and to categorize PCR negative samples as cbDVG- and PCR positive samples as cbDVG+. As we were very aware that cbDVG- samples could contain too few cbDVGs to be detected by PCR, we decided to examine the sensitivity of our PCR method by using our RNA-seq/VODKA pipeline. As indicated in fig. 1C and fig. S1A, we found that PCR negative samples did have fewer cbDVG junction reads than PCR positive samples. These data allowed us to determine that our PCR method can detect as low as 4 cbDVG junction reads per 10^8 total reads. As we have analyzed a total of 251 samples at least 2 times (more than 100 DNA gels), we thought it was unrealistic to present all DNA gels in one supplementary figure. However, we do agree with the reviewer and ***we have now added representative DNA gels for each cohort (Fig. S2, two from Cohort 1, 1 from Cohort 2, 1 from Cohort 3)*** to help visualize PCR positive and negative results. We have added these clarifications to the result and discussion section.

- In Fig 1 B, two bands are shown, yet it is unclear, which band(s) represents DVGs?

A: The PCR is designed in such a way that no PCR products can be amplified for the standard viral genome due to the location of the primer sequences as shown in Fig 1A. Previously published controls and validation of the PCR (Sun et al. 2015) include identification of major cbDVGs in

viral stocks and RSV infected cells and confirmation by Sanger's sequencing and RNAseq. For negative controls, we used an RSV full-length viral backbone plasmid to do a PCR using P1 and P2 (Fig S1D, (Sun et al. 2015). Using the same PCR conditions, we did not detect any amplicons, indicating that the bands from positive samples were not from viral genome.

We observed from the VODKA output that cbDVGs that accumulate in patients are much larger than those that accumulate in tissue culture. Upon sequencing the 1000 and 300 bp bands obtained

from these samples we noticed that the 1000bp and 300 bp bands represent two different fragments within cbDVGs, but we can't associate them to any specific cbDVG species. Moreover, each band represents a fragment that is present in most cbDVG species. Regrettably, these long cbDVGs contain long complementary structures that makes the amplification of their entire sequence very difficult by PCR. Therefore, we only use the PCR as a binary tool to indicate presence or absence of cbDVGs, and not to identify individual species. ***We have now included this explanation in the relevant results section of the manuscript (lines 295-310 and 496-499).***

To further validate our method, we have also added ***two new supplementary figures (S1B and S1C***) showing that specific primers can be designed based on RNAseq data to target junction regions (containing break/rejoin point) of specific cbDVG species present in human samples. We designed 3 primers that targeted 3 different cbDVG species. The amplicon sizes were as expected based on primer design and we confirmed the presence of the junction region for one of the primers by Sanger sequencing.

Also, this figure is missing a critical negative control showing that the PCR worked, which makes the interpretation of the data difficult.

A: We appreciate the reviewer's comment as the negative control is really important for our cbDVG specific PCR method. We have published the establishment of this PCR strategy and various negative controls in our previous publication (Sun et al. 2015). This paper is now cited both in the results section, as well as in the materials and methods of the paper. Using this PCR method, we successfully identified major cbDVGs in viral stocks and RSV infected cells, which were confirmed by Sanger's sequencing and RNAseq-VODKA. For negative controls, we used an RSV full-length viral backbone plasmid to do a PCR using P1 and P2 (Fig S1D, (Sun et al. 2015). Using the same PCR conditions, we did not detect any amplicons, indicating that the bands from positive samples were not from viral genome. For clinical samples, we have used the same PCR method to detect RSV cbDVGs from RSV- pediatric patient samples and did not observe any bands for the control samples infected with Adenovirus (Fig 5A, (Sun et al. 2015). In addition, every time we performed a PCR for this study, we included one sample in which the cDNA template was replaced with H2O (shown in fig. S2A).

- This is further complicated by the fact that there is lack of correlation between the level of DVGs seen in Fig 1B and DVG reads shown in Fig 1C (e.g.H83 showed high DVGs reads in RNA-seq but very low on gels).

A: We understand the reviewer's concern regarding our cbDVG specific PCR method. However, PCR bands do not always represent unique cbDVG species as explained above. Therefore, we did not use PCR as a quantitative measurement for cbDVG levels, but only as a qualitative measurement (cbDVG- or cbDVG+). A direct quantitative correlation between PCR and RNA-Seq is not expected.

3. Fig 2 – since viral load in DVG+ patients is significantly higher than in the DVG- group, the

conclusion that outcome and/or altered host response is linked to DVGs is problematic.
A: We understand the reviewer's concern. As explained above, our data indicate that the relation of cbDVGs with viral load is complex and that clinical outcomes do not always associate with viral loads. We recognize that for Cohort 1 alone we cannot eliminate the possibility that the role of cbDVGs in clinical outcome/host responses is independent of viral load. However, results from

Cohort 2 showing that cbDVG+ patients did not have worse symptoms in spite of higher viral load imply that in certain circumstances cbDVGs impact clinical outcome independently of viral load. In addition, using the longitudinal Cohort 3 we demonstrate that the appearance of early cbDVGs is independent of viral load (discussed in the response to this reviewer question 1). ***We have expanded the discussion of this point in the discussion of the manuscript (lines 311-314 and 325-327).***

4. Fig 3 – What were the expression levels of the host factors and cytokines shown in Fig 2 in this patient population? This will be important to determine, since in these patients viral load did not correlate with severity, yet presence of DVGs did.

A: We assume that the reviewer wants to see the cytokine profiles in Fig. 3, not Fig. 2. ***We have added in fig. 3D and 3E the cytokine profiles from Cohort 2 comparing cbDVG+/cbDVG- groups***. The profiles for this cohort were similar to those from Cohort1 showing that in cbDVG+ patients IFNA and a series of pro-inflammatory cytokines were induced. Interestingly, cbDVGs+ non-hospitalized patients were sampled earlier than cbDVG+ hospitalized patients in this cohort (fig. S3), suggesting that cbDVGs could induce immune responses earlier in non-hospitalized patients than in hospitalized patients. This early induction may lead to a decrease in viral replication later on and thus these patients do not end up being hospitalized. This has been added to results and materials and methods.

5. Disease severity scores are being used in figs 3 and 4 to measure clinical outcome (vs. length of hospitalization and ICU time in fig 2). The authors should provide an explanation what these scores are. Moreover, the scores in figure 3 are all below 8, whereas in figure 4C the scores can be as high as 20. How are these calculated and why is there such a difference?

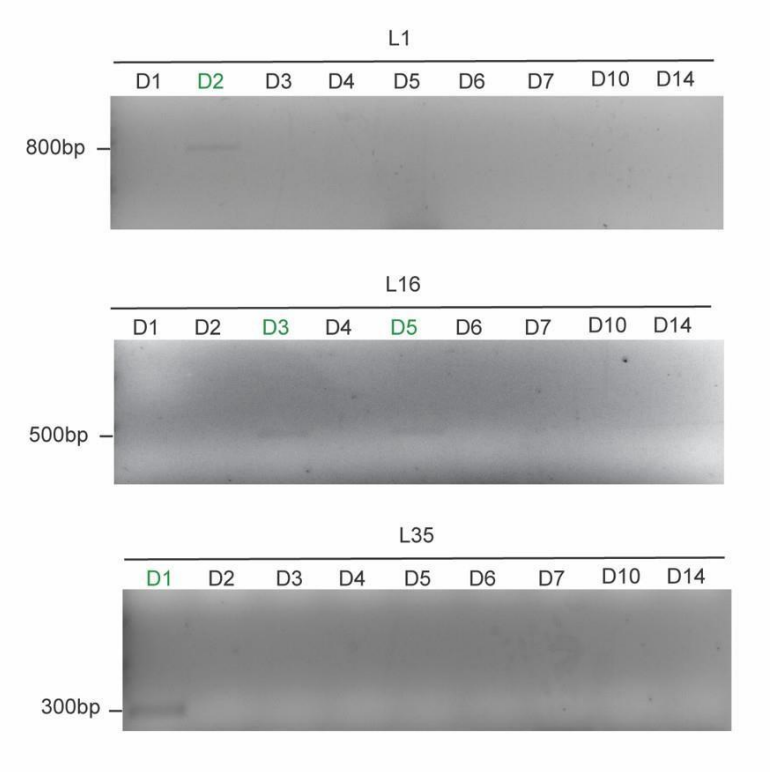
A: Cohorts 2 and 3 belong to two different studies and their score matrixes were slightly different. For cohort 2, the respiratory severity score (RSS) was used, which is an ordinal scale based on respiratory rate, flaring or retractions, heart rate, and wheezing that was slightly modified from other scores derived for acute respiratory illnesses. The RSS ranges from 0 to 12 (Rodriguez et al. 2016). For cohort 3, common cold symptoms were evaluated by self-reported symptom diary using a modified Jackson score. Eight symptoms (nasal discharge, nasal congestion, sneezing, cough, sore throat, headache, feverishness and fatigue) were assessed, with subjects asked to score each symptom 0 (absent), 1 (mild), 2 (moderate), 3 (severe). Maximum daily symptom score was therefore 24 and total maximum score over the 14 days of assessment was 384 (Jackson et al. 1958). ***We have added this information to the materials and methods with citations (lines 421-424 and 441-443).***

6. Figs 4E, 4F and 5A - Are there quantitative differences in the level of DVGs between L5 and the 3 other samples (L35, L1, L16)? It would be important to show DNA gels of DVGs for these samples. Specifically, DVGs levels at day 2,3 and 7 for L5 relative to the other samples. Do viral

load and DVGs level of L5 correlate?

A: As RT-PCR is a binary tool to screen for presence or absence of cbDVGs we think it would not be appropriate to make any quantitative conclusions between these 4 samples using PCR. However, we have added now the DNA gel picture of patient L5 as a representative result (fig. S2C) and also L1, L16 and L35 below (cbDVG+ days are labeled in green). In combination with fig. 4F, it can be concluded for L5 that cbDVGs are present at undetectable viral load levels (Day 2 and 3) and in the presence of detectable viral load levels (Day 7). Furthermore, cbDVGs were present in L1,

L16 and L35 at undetectable viral load levels. These data support our statement that cbDVG presence is independent of viral load.



7. Fig 5 - The reason for choosing cutoff II is unclear and seems arbitrary. Also, the error bar in Fig. 5B is confusing, since the graph is plotted on a log scale in the y axis for virus load. The other half of the error bar should be added.

A: We analyzed the data based on 3 different cutoffs to make sure our results are NOT arbitrary. The reason we chose cutoff II as a representative result is because with this cutoff the early and late groups had the most similar numbers of samples (7 vs 11) and the overall conclusion did not change no matter which cutoff we chose. ***We now added the other two cutoffs as a supplementary figure (fig. S4). We have also changed the error bars in Figure 5B as suggested.***

8. Fig. 6 – Again, an association between DVGs and viral load is demonstrated. The prolonged DVGs group has a higher viral load than transient DVGs. What is the level of DVGs in the prolonged DVGs group and the transient DVGs group?

A: This is a very good point and question. As explained in the response to question 1 of the

reviewer, cbDVG presence is not always dependent on high viral load and high viral load does not always associate with disease severity. Figure 6 actually shows that the observed association of cbDVGs and viral load depends on the kinetics of cbDVGs. We agree that prolonged cbDVGs associate with higher viral loads compared to no cbDVGs (fig. 6C). However, transient cbDVGs had almost identical viral load compared to no cbDVGs (fig 6C). Furthermore, early cbDVGs (fig. S5, cut off I) had no detectable viral load. As RT-PCR is a binary tool to screen for presence or absence of cbDVGs, we think it would not be appropriate to make any quantitative conclusions between DVG groups based on the intensity of their gel bands. However, based on data from

previous studies and the current study, we think that cbDVGs in the prolonged group were not effective in controlling the virus which would lead to uncontrolled replication of virus and further amplification of cbDVGs (fig. 6C).

9. What was the RSV subtype of the patients in the various cohorts? Since subtype A has been linked with disease severity, it will be important to determine whether presence/level of DVGs is greater with this specific subtype.

A: We apologize for not making this clear. All clinical samples analyzed in this manuscript are confirmed to be RSV subtype A. ***This has been added in the introduction (line 83) and in multiple other places in the manuscript***.

10. The authors should state the importance/impact of the study in the abstract. Moreover, they should discuss the clinical implications in the discussion and how they view the use of this candidate biomarker in clinical practice.

A: These data establish cbDVGs as an important factor determining the clinical outcome of RSV infection **and potentially as a prognostic tool to identify patients who are at high risk for longer hospitalization and admission to the ICU**. We have added the bold sentence as the last sentence of the abstract and added more information on this in the discussion.

11. Figs 1 and 7 were missing. While Fig 1 was then uploaded, I didn't realize Fig 7 was also missing, until later.

A: Our apology for the reviewer that could not access all figures. We have now re-uploaded all figures again and tried them on several computers.

Reviewer #3 (Remarks to the Author):

In these studies the results examine the correlation of DVGs as a predictor of severe RSV disease and provide data comparing both hospitalized and non-hospitalized individuals in 3 different cohorts. Furthermore, the studies used an experimental infection with M37 to examine the correlation of CVG with disease. In several studies it has been shown that DVGs can be an indication of higher RSV load during infection. In these studies the data indicate that higher DVGs can be found in both more severely infected hospitalized and in patients not hospitalized with non-severe infection. As suggested in the Discussion, these latter data suggest that it is not simply the viral titer, but how the individual responds to the higher viral load and that both groups, higher and lower DVG may respond poorly with the higher DVG hospitalized group having more "bad" response to the higher viral titer. Some of the logic to explain the results of finding the higher DVG in hospitalized and non-hospitalized patients is confusing and not supported by a mechanistic explanation, rather a more logical explanation that CVG may protect or promote pathogenesis

depending upon the response of the individual. This reviewer would agree but could add that it may be the active virus as well as or instead of the CVG that is driving the pathogenesis. However, no further investigation into what the mechanism might between these hospitalized and non-hospitalized groups has been provided.

The Discussion offers a logical explanation of how the DVGs may provide for an altered response based upon previous data, but unfortunately do not provide any data to resolve these ideas further within these studies. Altogether, these studies provide interesting and potentially important data that demonstrate that CVG levels during RSV infection may predict the outcome of an infection

but really don't bring any additional clarity as to why they may or may not drive a detrimental response and what the mechanism might be for the good or poor response.

A: We appreciate the reviewer's positive comments and agree that the major focus of this study was not mechanistic. Our previous *in vitro*, *in vivo*, *and ex vivo* published study has demonstrated that cbDVGs are the primary trigger of antiviral responses during RSV infection (Sun et al. 2015). In that publication, we reported elevated IFNb production when cells were supplemented with purified defective viral particles (containing cbDVGs) during RSV infection *in vitro*. In addition, we observed a rapid induction of IFNb expression as early as 6h in mice after infection with a viral stock enriched with cbDVGs (HD). However, this was not observed in mice infected with viral stocks lacking cbDVGs (LD). At this early time point, the standard viral replication was undetectable mimicking the situation of three patients in Cohort 3 having cbDVGs before D3. HD infected mice had reduced lung pathology and less viral replication compared to LD infected mice due to the early induction of antiviral responses and faster control of virus spread. In addition, in human lung explants, we noticed that different kinetics of cbDVG generation associated with distinct antiviral response induction and virus control. The earlier cbDVGs were detected, the earlier IFNL1 was expressed and the fewer standard viruses were produced, suggesting the kinetics of cbDVG generation is critical for their function and host intrinsic factors impact the rate of cbDVG generation.

In this manuscript our goal was to investigate the impact of cbDVGs in clinical outcomes. Due to the limitation of cohort studies, we cannot do the above mechanistic experiments in humans. However, in this revision, ***we have added the cytokine profiles from Cohort 2 and Cohort 3 comparing cbDVG+/cbDVG- groups and prolonged/transient cbDVG groups, respectively***.

cbDVG+ patients from non-hospitalized pediatric cohort (Cohort 2) had elevated IFNA and a series of pro-inflammatory cytokines compared to cbDVG- patients (***new fig. 3D and E***), which is similar as observed in hospitalized cohort (Cohort 1). Interestingly, cbDVGs+ non-hospitalized patients were sampled earlier than hospitalized patients in Cohort 2 (fig. S3), suggesting that cbDVGs could induce immune responses earlier in non-hospitalized patients than in hospitalized patients. This early induction may lead to decreased viral replication later on (like the situation in HD infected mice) and thus these patients did not end up being hospitalized. This is pure speculation and it has been included as such in the discussion of the manuscript.

Using the longitudinal Cohort 3, we noticed different kinetics of cbDVG generation were observed in different patients. Due to the poor viral infection in adults and limited available early time samples, we did not notice significant differences in IFN and IFN stimulated gene expression between early and late cbDVG groups at D3 post infection. However, at D7 post infection (peak of viral replication), late/prolonged cbDVGs were associated with higher viral replication and induced significantly more pro-inflammatory cytokine expression (***new fig.6G and 6H***), mimicking LD infected mice and Cohort 1 shown in fig.2. We think those cbDVGs induced the

immune response too late to control the virus, leading to more severe symptoms.

Taken together, our data suggests that kinetics of cbDVGs affect virus replication, which in turn determines pathogenesis. However, we do agree with the reviewer that, because of the complex relationship between the virus and its cbDVGs, it is difficult to determine if pathogenesis is directly affected by cbDVGs, indirectly by the effect of cbDVGs on virus replication or a combination of

both. The deeper mechanism of how different kinetics of cbDVGs impact clinical outcome, including what controls the kinetics of cbDVG generation is one of our major interests and is under investigation. Unfortunately, those questions cannot be answered using our current cohort information.

In addition, it is not clear with these data how the ability of CVG level would further extend prediction past viral titer levels.
A: Based on current literature, during RSV infection, viral load does not always correlate with disease severity (Piedra et al. 2017) and there are mixed messages in regard to using viral load as a predictor for RSV clinical outcomes. We think we can now provide a possible explanation for these conflicts. Data from cohort 1 and 2 suggest that viral load is a good predictor within hospitalized cohorts but not in non-hospitalized cohorts. Furthermore, data from cohort 3 suggests that in addition to viral load and detection of cbDVGs, when cbDVGs appear and how long they are maintained are important factors impacting RSV disease severity. To our knowledge, our paper is the first one to identify these factors. ***These points have been added in the discussion (line 311-314 and 325-327).***

Reviewer #4 (Remarks to the Author):

In this manuscript, samples from three clinical datasets (RSV-infected hospitalized pediatric patients, hospitalized and non-hospitalized children with RSV, and adults experimentally challenged with RSV) are analyzed for presence of defective-interfering particles (cbDVGs). Overarching goal is the identification of predictive biomarkers for progression of RSV disease to severe lower respiratory infection and viral pneumonia.
The current literature is not fully consistent on whether RSV load at the time of hospitalization can serve as a predictor of outcome. Additional and/or more predictive biomarkers that may impact disease management decisions are urgently needed. The study therefore aims to address an important problem.
cbDVGs emerge spontaneously during mononegavirus replication, triggering activation of host innate immune response pathways. However, the correlation of cause and effect of cbDVGs in severe RSV disease is unclear. This study provides first insight into a potentially important biomarker, but revealed that the correlation is complex and dependent on the kinetic of cbDVG appearance during the cause of RSV infection must be considered. Unfortunately, kinetic information appears to be lacking for cohort 1, confounding interpretation of this important dataset. Information on relative time of patient sample collection for this cohort needs to be provided analogous to that shown, for instance, in figure S1 for cohort 2. Can cohort 1 be analyzed by subgroups with comparable sampling times?
A: We value the reviewer's suggestion. Unfortunately, kinetics information is not available for Cohort 1. We only know the exact date of sampling, close to or at the time of admission to the hospital, but not in relation to symptoms appearance or infection onset. Therefore, we cannot assess the relative infection stage of RSV infection for those pediatric patients in Cohort 1.

However, looking at the information of all three cohorts as a whole, we can conclude that the timing of cbDVG appearance determines their impact on clinical outcome. Prospective studies are of high interest to us, but out of reach with currently approved protocols.

Specific comments:

PCR-based and RNAseq-based detection of cbDVGs in samples from the same patients shows no quantitative correlation (figures 1B and 1C, relative amplicon intensity vs DVG reads). That is surprising and challenges the claimed reliability of the PCR-based method and thus a central experimental approach of the study.
A: We understand the reviewer concern. PCR methods are semi-quantitative but remain widely used in the clinics to detect viral sequences. As explained to reviewer 1 and 2, the PCR bands do not represent specific cbDVG species but instead populations of cbDVGs. Therefore, we did not use PCR as a quantitative measurement for cbDVG levels, but only as a qualitative measurement. A direct quantitative correlation between PCR and RNA-Seq is not expected. ***We have added a new supplementary figure (fig. S1A) that confirms on 13 additional samples that the two methods agree in terms of presence and absence of cbDVGs***. We have added clarifications to the manuscript.

Where the 122 banked nasal washes (cohort I) collected at comparable time points (i.e. prior to patient hospitalization, at the time of hospitalization, or a consistent number of days after hospitalization? Without standardization of sampling time, samples from patients with longer duration of hospitalization may have been collected at later stages of severe disease, enriching for presence of cbDVGs and compromising usefulness as biomarker for progression to LRI.
A: Thanks for the reviewer's suggestion and this is a valuable point. All samples in Cohort 1 were collected at comparable time points, either at the time of or shortly after (1-2 days) admission to hospital. However, we do not have the information regarding the time of symptom onset and thus it is difficult for us to assess at what stage of infection patients were hospitalized and produced cbDVGs from Cohort 1. We have added this information in the material and methods.

Were all patients from cohort 1 RSV naïve, or can this group by analyzed by RSV experienced and inexperienced patients?
A: Thanks for the reviewer's comment. Unfortunately, we do not have this information for Cohort 1.

Page 3, top paragraph: note that ribavirin has been approved for the treatment of RSV disease.
A: We appreciate reviewer's suggestion and are aware of the usage of ribavirin in clinics. One of us, a practicing infectious disease neonatologist, commented that ribavirin was no longer used, except in exceptional circumstances in transplant patients. Therefore, we would like to exclude ribavirin here.

Reason for 2 different size amplicons in figure 2B? Does this reflect different break and/or rejoin points? Please clarify.
A: The PCR is designed in such a way that no PCR products can be amplified for the standard viral genome due to the location of the primer sequences as shown in Fig 1A. Previously published

controls and validation of the PCR (Sun et al. 2015) include identification of major cbDVGs in viral stocks and RSV infected cells and confirmation by Sanger's sequencing and RNAseq. For negative controls, we used an RSV full-length viral backbone plasmid to do a PCR using P1 and P2 (Fig S1D, (Sun et al. 2015). Using the same PCR conditions, we did not detect any amplicons, indicating that the bands from positive samples were not from viral genome.

We observed from the VODKA output that cbDVGs that accumulate in patients are much larger than those that accumulate in tissue culture. Upon sequencing the 1000 and 300 bp bands obtained from these samples we noticed that the 1000bp and 300 bp bands represent two different fragments within cbDVGs, but we can't associate them to any specific cbDVG species. Moreover, each band represents a fragment that is present in most cbDVG species. Regrettably, these long cbDVGs contain long complementary structures that makes the amplification of their entire sequence very difficult by PCR. Therefore, we only use the PCR as a binary tool to indicate presence or absence of cbDVGs, and not to identify individual species. ***We have now included this explanation in the relevant results section of the manuscript (lines 295-310 and 496-499).***

To further validate our method, we have also added ***two new supplementary figures (S1B and S1C***) showing that specific primers can be designed based on RNAseq data to target junction regions (containing break/rejoin point) of specific cbDVG species present in human samples. We designed 3 primers that targeted 3 different cbDVG species. The amplicon sizes were as expected based on primer design and we confirmed the presence of the junction region for one of the primers by Sanger sequencing.

Decision Letter, first revision:

Dear Carolina,

Thank you for your patience while your manuscript "Detection of copy-back defective viral genomes in nasal secretions predicts RSV disease severity in children and adults" was under peer review at Nature Microbiology.

I am sorry for the delay in getting back to you with a decision but at this stage we routinely run several editorial checks on all the files and figures so that we can highlight all the final revisions that are needed prior to formal acceptance, as your manuscript has now been seen by our referees, and in the light of their advice I am delighted to say that we can in principle offer to publish it.

First, however, we would like you to revise your paper to address the points made by the reviewers, and to ensure that it is in Nature Microbiology format.
Editorially, we will need you to make some changes so that the paper complies with our Guide to Authors at http://www.nature.com/nmicrobiol/info/gta.

I appreciate this email is long and recommend that you print it and use it as a checklist, reading it carefully to the end, in order to avoid delays to publication down the line.

Please note that we will be considering your paper for publication as an ARTICLE in our pages.

Specific points:
In particular, while checking through the manuscript and associated files, we noticed the following specific points which we will need you to address:

1. Length.
We estimate that your manuscript currently exceeds our normal length limit for Articles of about 3000 words. We have some flexibility, and can allow a revised manuscript at 3500 words, but please consider this a firm upper limit. You could achieve some shortening by moving some details to the Methods section that should follow the main text (the length of the Methods section is unlimited and does not count towards the main text length).

2. Title.
Titles should give an idea of the main finding of the paper and ideally not exceed 120 characters (including spaces). We discourage the use of active verbs and do not allow punctuation. In this case, we felt that the title should have the virus name in the beginning of the title and therefore suggest revising the title along the lines of "Respiratory syncytial virus disease severity is associated with detection of copy-back defective viral genomes".

3. Abstract/ Introductory paragraph.
Per journal guidelines, the abstract of Articles should not exceed 200 words, this would need to be adjusted accordingly. In addition, we would like to suggest a few changes that can be seen in the attached "NMICROBIOL-20041377A_Title and Abstract suggestions.docx" file.

4. Main text display items and supplementary information.
Please note that manuscripts in the Article format are limited to 6 display items in the main text (figures and tables), while the rest must be presented as supplementary information. All Supplementary Information must be submitted in accordance with the instructions in the attached Inventory of Supporting Information, and should fit into one of three categories: EXTENDED DATA (ED); SUPPLEMENTARY INFORMATION (SI); and SOURCE DATA. Below are detailed instructions on how to format each category. For your paper, we suggest that you do the following:

a. Main figures: please maintain the current 6 main figures that illustrate the main findings of the paper. The table can remain as main item.

Please leave more space between 5B and 5C to avoid confusion between the Early cutoff categories and the data in Figure 5C (that only uses the cutoff II). In the manuscript the panels within Figure 5C are described as Figure 5C, 5D and 5E, please adjust accordingly.

b. Extended data (ED): please convert the 6 SI figures into ED figures. These are an integral part of the paper (presented online in the online version) and are meant to be multi-panel A4 size figures. More information on file formats and how the legends should be supplied can be found below and in the attached Inventory of Supporting Information.

c. Supplementary information (SI): please maintain the Rmarkdown code as SI, which will mostly be featured in a single, flattened PDF. Please see below for additional information on how to properly format SI files.

d. Source data: we now ask authors to provide as much source data as possible. In addition, requested source data include, when applicable, raw versions of all gels shown, or the numerical data used to generate graphs and statistics. For gels, these should be raw, uncropped, unmanipulated versions of all gels, ideally showing the original molecular weight markers and using text boxes to indicate which sections of the full gels were cropped to generate the figures shown. Therefore, please include such full length blots for the gels shown in Figure 1B, S1B and S2 (which later should be Ext. Data 1B and 2) as Source Data. Raw data will be linked to specific main figures and ED figures. Please see below for additional information on how to provide source data.

e. Graph axes. Per style guidelines, all graph axes must be labelled and carefully explained in accompanying legend. Currently, the axes in Figure 5B are illegible. Please revise.

5. Replicates and statistics.
While carefully checking the figures, we noted a few things that need to be revised so that they comply with our style guidelines for data presentation and accurately report on the number of replicates, statistical testing, etc. Please find these points detailed in the attached "NMICROBIOL-20041377A_Extended_comments" file and revise the highlighted figure legends accordingly.

Wherever statistics have been derived (e.g., error bars, box plots, statistical significance), the legend needs to provide and define the n number (i.e., the sample size used to derive statistics) as a precise value (not a range), using the wording "n=X biologically independent samples/animals/independent experiments," etc. as applicable.

All error bars need to be defined in the legends (e.g., SD, SEM) together with a measure of centre (e.g. mean, median), and should be accompanied by their precise n number defined as noted above.

All violin plots need to be defined in the legends in terms of minima, maxima, centre, and percentiles, and should be accompanied by their precise n number defined as noted above.

The figure legends must indicate the statistical test used and if applicable, whether the test was one- or two-sided. A description of any assumptions or corrections such as tests of normality and adjustment for multiple comparisons must also be included.

For null hypothesis testing, please indicate the test statistic (e.g., F, t, r) with confidence intervals, effect sizes, degrees of freedom and P values noted.

Test results (e.g., p-values, q-values) should be given as exact values whenever possible and appropriate, and confidence intervals noted.

Please indicate how estimates of effect sizes were calculated (e.g., Cohen's d, Pearson's r).

Please state in the legends how many times each experiment was repeated independently with similar results. This is needed for all experiments but is particularly important wherever representative

experiments are shown. If space in the legends is limiting, this information can be included in a section titled "Statistics and Reproducibility".

For all bar graphs, the corresponding dot plot must be overlaid.

6. ORCID.
As mentioned previously, we ask all corresponding authors to provide their unique ORCID identifiers at the time of final submission. This information is currently missing for all corresponding authors. Please see below for additional information on how to sign up to ORCID and link your account to the one in our manuscript tracking system.

7. Data availability. Please include a data availability section at the end of the methods. Note that this section should refer to all source data and include all accession codes for relevant data deposited to databases. For example, access to the structural data should be detailed here. See below for additional details on how to format this section and some useful examples can be found here (https://www.nature.com/articles/s41564-019-0614-3#data-availability or https://www.nature.com/articles/s41564-019-0609-0#data-availability).

8. Code availability. Please include a code availability section at the end of the methods. Note that this section should accession details for customized scripts, which should ideally be deposited to relevant databases (such as GitHub). All accession codes must be live by the time of publication of the piece. Please describe access to the "Rmarkdown files" in this section.

9. Reporting checklist.
Note that a final version of the reporting checklist will be published with your manuscript. Therefore, please revise this document according to the instructions found in the annotated PDF attached to this message (NMICROBIOL-20041377A_ReportingSummary.pdf).

General points:
We will also need you to check through all of the following general points when preparing the final version of your manuscript:

Please read carefully through all of the following general formatting points when preparing the final version of your manuscript, as submitting the manuscript files in the required format will greatly speed the process to final acceptance of you work.

Titles should give an idea of the main finding of the paper and ideally not exceed 90 characters (including spaces). We discourage the use of active verbs and do not allow punctuation.

The paper's summary paragraph (about 150-200 words; no references) should serve both as a general introduction to the topic, and as a brief, non-technical summary of your main results and their implications. It should start by outlining the background to your work (why the topic is important) and the main question you have addressed (the specific problem that initiated your research), before going on to describe your new observations, main conclusions and their general implications. Because we hope that scientists across the wider microbiology community will be interested in your work, the first paragraph should be as accessible as possible, explaining essential but specialised terms concisely. We suggest you show your summary paragraph to colleagues in other fields to uncover any

problematic concepts.

Please include a data availability statement as a separate section after Methods but before references, under the heading "Data Availability". This section should inform readers about the availability of the data used to support the conclusions of your study. This information includes accession codes to public repositories (data banks for protein, DNA or RNA sequences, microarray, proteomics data etc…), references to source data published alongside the paper, unique identifiers such as URLs to data repository entries, or data set DOIs, and any other statement about data availability. At a minimum, you should include the following statement: "The data that support the findings of this study are available from the corresponding author upon request", mentioning any restrictions on availability. If DOIs are provided, we also strongly encourage including these in the Reference list (authors, title, publisher (repository name), identifier, year). For more guidance on how to write this section please see:
http://www.nature.com/authors/policies/data/data-availability-statements-data-citations.pdf

Choosing the right electronic format for your figures at this stage will speed up the processing of your paper. We would like the figures to be supplied as vector files - EPS, PDF, AI or postscript (PS) file formats (not raster or bitmap files), preferably generated with vector-graphics software (Adobe Illustrator for example). Please try to ensure that all figures are non-flattened and fully editable. All images should be at least 300 dpi resolution (when figures are scaled to approximately the size that they are to be printed at) and in RGB colour format. Please do not submit Jpeg or flattened TIFF files. Please see also 'Guidelines for Electronic Submission of Figures' at the end of this letter for further detail.
Please view http://www.nature.com/authors/editorial_policies/image.html for more detailed guidelines.

We will edit your figures/tables electronically so they conform to Nature Microbiology style. If necessary, we will re-size figures to fit single or double column width. If your figures contain several parts, the parts should be labelled lower case a, b, and so on, and form a neat rectangle when assembled.

Please check the PDF of the whole paper and figures (on our manuscript tracking system) VERY CAREFULLY when you submit the revised manuscript. This will be used as the 'reference copy' to make sure no details (such as Greek letters or symbols) have gone missing during file-transfer/conversion and re-drawing.

All Supplementary Information must be submitted in accordance with the instructions in the attached Inventory of Supporting Information, and should fit into one of three categories:

1. EXTENDED DATA: Extended Data are an integral part of the paper and only data that directly contribute to the main message should be presented. These figures will be integrated into the full-text HTML version of your paper and will be appended to the online PDF. There is a limit of 10 Extended Data figures, and each must be referred to in the main text. Each Extended Data figure should be of the same quality as the main figures, and should be supplied at a size that will allow both the figure and legend to be presented on a single legal-sized page. Each figure should be submitted as an individual .jpg, .tif or .eps file with a maximum size of 10 MB each. All Extended Data figure legends must be provided in the attached Inventory of Accessory Information, not in the figure files

themselves.

2. SUPPLEMENTARY INFORMATION: Supplementary Information is material that is essential background to the study but which is not practical to include in the printed version of the paper (for example, video files, large data sets and calculations). Each item must be referred to in the main manuscript and detailed in the attached Inventory of Accessory Information. Tables containing large data sets should be in Excel format, with the table number and title included within the body of the table. All textual information and any additional Supplementary Figures (which should be presented with the legends directly below each figure) should be provided as a single, combined PDF. Please note that we cannot accept resupplies of Supplementary Information after the paper has been formally accepted unless there has been a critical scientific error.

All Extended Data must be called you in your manuscript and cited as Extended Data 1, Extended Data 2, etc. Additional Supplementary Figures (if permitted) and other items are not required to be called out in your manuscript text, but should be numerically numbered, starting at one, as Supplementary Figure 1, not SI1, etc.

3. SOURCE DATA: We encourage you to provide source data for your figures whenever possible. Full-length, unprocessed gels and blots must be provided as source data for any relevant figures, and should be provided as individual PDF files for each figure containing all supporting blots and/or gels with the linked figure noted directly in the file. Statistics source data should be provided in Excel format, one file for each relevant figure, with the linked figure noted directly in the file. For imaging source data, we encourage deposition to a relevant repository, such as figshare (https://figshare.com/) or the Image Data Resource (https://idr.openmicroscopy.org).

Nature Research journals encourage authors to share their step-by-step experimental protocols on a protocol sharing platform of their choice. Nature Research's Protocol Exchange is a free-to-use and open resource for protocols; protocols deposited in Protocol Exchange are citable and can be linked from the published article. More details can found at www.nature.com/protocolexchange/about.

Please note that after the paper has been formally accepted you can only provide amended Supplementary Information files for critical changes to the scientific content, not for style. You should clearly explain what changes have been made if you do resupply any such files.

Figure legends must provide a brief description of the figure and the symbols used, within 350 words. This must include definitions of any error bars employed in the figures.

It is a condition of publication that you include a statement before the acknowledgements naming the author to whom correspondence and requests for materials should be addressed.

Finally, we require authors to include a statement of their individual contributions to the paper -- such as experimental work, project planning, data analysis, etc. -- immediately after the acknowledgements. The statement should be short, and refer to authors by their initials. For details please see the Authorship section of our joint Editorial policies at http://www.nature.com/authors/editorial_policies/authorship.html

We will not send your revised paper for further review if, in the editors' judgement, the referees' comments on the present version have been addressed. If the revised paper is in Nature Microbiology format, in accessible style and of appropriate length, we shall accept it for publication immediately.

Please resubmit electronically

* the final version of the text (not including the figures) in either Word or Latex.

* publication-quality figures. For more details, please refer to our Figure Guidelines, which is available here: https://mts-nmicrobiol.nature.com/letters/Figure_guidelines.pdf

* Extended Data & Supplementary Information, as instructed

* a point-by-point response to any issues raised by our referees and to any editorial suggestions.

* any suggestions for cover illustrations, which should be provided at high resolution as electronic files. Please note that such pictures should be selected more for their aesthetic appeal than for their scientific content. I am sure you will understand that we cannot make any promise as to whether any of your suggestions might be selected for the cover of Nature Microbiology.

Please use the following link to access your home page:

{REDACTED}

* This url links to your confidential homepage and associated information about manuscripts you may have submitted or be reviewing for us. If you wish to forward this e-mail to co-authors, please delete this link to your homepage first.

Nature Microbiology offers a transparent peer review option for new original research manuscripts submitted from 1st December 2019. We encourage increased transparency in peer review by publishing the reviewer comments, author rebuttal letters and editorial decision letters if the authors agree. Such peer review material is made available as a supplementary peer review file. Please state in the cover letter 'I wish to participate in transparent peer review' if you want to opt in, or 'I do not wish to participate in transparent peer review' if you don't. Failure to state your preference will result in delays in accepting your manuscript for publication.
Please note: we allow redactions to authors' rebuttal and reviewer comments in the interest of confidentiality. If you are concerned about the release of confidential data, please let us know specifically what information you would like to have removed. Please note that we cannot incorporate redactions for any other reasons. Reviewer names will be published in the peer review files if the reviewer signed the comments to authors, or if reviewers explicitly agree to release their name. For more information, please refer to our FAQ page.

Nature Microbiology has now transitioned to a unified Rights Collection system which will allow our Author Services team to quickly and easily collect the rights and permissions required to publish your

work. Once your paper is accepted, you will receive an email in approximately 10 business days providing you with a link to complete the grant of rights.

Please note that you will not receive your proofs until the publishing agreement has been received through our system.
If you have any questions please contact ASJournals@springernature.com.

* Should your Article contain any items (figures, tables, images, videos or text boxes) that are the same as (or are adaptations of) items that have previously been published elsewhere and/or are owned by a third party, please note that it is your responsibility to obtain the right to use such items and to give proper attribution to the copyright holder. This includes pictures taken by professional photographers and images downloaded from the internet. If you do not hold the copyright for any such item (in whole or part) that is included in your paper, please complete and return this Third Party Rights Table, and attach any grant of rights that you have collected.

For more information on our licence policy, please consult http://npg.nature.com/authors.

<B>ORCID</B>

Nature Microbiology is committed to improving transparency in authorship. As part of our efforts in this direction, we are now requesting that all authors identified as 'corresponding author' create and link their Open Researcher and Contributor Identifier (ORCID) with their account on the Manuscript Tracking System (MTS) prior to acceptance. ORCID helps the scientific community achieve unambiguous attribution of all scholarly contributions. For more information please visit http://www.springernature.com/orcid

For all corresponding authors listed on the manuscript, please follow the instructions in the link below to link your ORCID to your account on our MTS before submitting the final version of the manuscript. If you do not yet have an ORCID you will be able to create one in minutes.
https://www.springernature.com/gp/researchers/orcid/orcid-for-nature-research

IMPORTANT: All authors identified as 'corresponding author' on the manuscript must follow these instructions. Non-corresponding authors do not have to link their ORCIDs but are encouraged to do so. Please note that it will not be possible to add/modify ORCIDs at proof. Thus, if they wish to have their ORCID added to the paper they must also follow the above procedure prior to acceptance.

To support ORCID's aims, we only allow a single ORCID identifier to be attached to one account. If you have any issues attaching an ORCID identifier to your MTS account, please contact the Platform Support Helpdesk.

Nature Research journals encourage authors to share their step-by-step experimental protocols on a protocol sharing platform of their choice. Nature Research's Protocol Exchange is a free-to-use and open resource for protocols; protocols deposited in Protocol Exchange are citable and can be linked from the published article. More details can found at <a

href="https://www.nature.com/protocolexchange/about"
target="new">www.nature.com/protocolexchange/about.

We hope that you will support this initiative and supply the required information. Should you have any query or comments, please do not hesitate to contact me.

I think that's it. Sorry for sending you such a long list of points to address, but resolving these issues at this stage should help to ensure that everything runs smoothly once the papers is passed on to our production team, which should help to speed up publication of the article.

We hope to hear from you within two weeks; please let us know if the revision process is likely to take longer.

****

Reviewer Expertise:

Referee #1: Defective viral genomes
Referee #2: RNA-Sequencing
Referee #3: Respiratory syncytial virus and clinical studies
Referee #4: Paramyxovirus pathogenesis

Reviewer Comments:

Reviewer #1 (Remarks to the Author):

The authors have addressed my prior comments well in their revisions.

Reviewer #2 (Remarks to the Author):

The revised manuscript is improved. Most of my comments have been addressed.

Reviewer #3 (Remarks to the Author):

The Authors have appropriately addressed concerns.

Reviewer #4 (Remarks to the Author):

The authors have satisfactorily addressed my previous concerns. Overall, this is an important study that may contribute to establishing much-needed biomarkers predicting RSV disease severity.

Final Decision Letter:

Dear Carolina,

I am pleased to accept your Article "Detection of Respiratory syncytial virus defective genomes in nasal secretions is associated with distinct clinical outcomes" for publication in Nature Microbiology. Thank you for having chosen to submit your work to us and many congratulations.

Before your manuscript is typeset, we will edit the text to ensure it is intelligible to our wide readership and conforms to house style. We look particularly carefully at the titles of all papers to ensure that they are relatively brief and understandable.

Once your manuscript is typeset and you have completed the appropriate grant of rights, you will receive a link to your electronic proof via email with a request to make any corrections within 48 hours. If, when you receive your proof, you cannot meet this deadline, please inform us at rjsproduction@springernature.com immediately. Once your paper has been scheduled for online publication, the Nature press office will be in touch to confirm the details.

Acceptance of your manuscript is conditional on all authors' agreement with our publication policies (see www.nature.com/nmicrobiolate/authors/gta/content-type/index.html). In particular your manuscript must not be published elsewhere and there must be no announcement of the work to any media outlet until the publication date (the day on which it is uploaded onto our website).

Several funders require deposition the accepted manuscript (AM) to PubMed Central or Europe PubMed Central. To enable compliance with these requirements, Nature Research therefore offers a free manuscript deposition service for original research papers supported by a number of PMC/EPMC participating funders. If you do not choose to publish immediate open access, we can deposit the accepted manuscript in PMC/Europe PMC on your behalf, if you authorise us to do so.

In approximately 10 business days you will receive an email with a link to choose the appropriate publishing options for your paper and our Author Services team will be in touch regarding any additional information that may be required.

You will not receive your proofs until the publishing agreement has been received through our system.

If you have any questions about our publishing options, costs, Open Access requirements, or our legal forms, please contact ASJournals@springernature.com

An online order form for reprints of your paper is available at https://www.nature.com/reprints/author-reprints.html. All co-authors, authors' institutions and authors' funding agencies can order reprints using the form appropriate to their geographical region.

We welcome the submission of potential cover material (including a short caption of around 40 words) related to your manuscript; suggestions should be sent to Nature Microbiology as electronic files (the image should be 300 dpi at 210 x 297 mm in either TIFF or JPEG format). Please note that such pictures should be selected more for their aesthetic appeal than for their scientific content, and that colour images work better than black and white or grayscale images. Please do not try to design a cover with the Nature Microbiology logo etc., and please do not submit composites of images related to your work. I

am sure you will understand that we cannot make any promise as to whether any of your suggestions might be selected for the cover of the journal.

You can now use a single sign-on for all your accounts, view the status of all your manuscript submissions and reviews, access usage statistics for your published articles and download a record of your refereeing activity for the Nature journals.

To assist our authors in disseminating their research to the broader community, our SharedIt initiative provides you with a unique shareable link that will allow anyone (with or without a subscription) to read the published article. Recipients of the link with a subscription will also be able to download and print the PDF.

As soon as your article is published, you will receive an automated email with your shareable link.

Congratulations again and I look forward seeing your article published!