Response to reviews of "Extreme heterogeneity of influenza virus infection in single cells" for *eLife*

Alistair B. Russell, Cole Trapnell, and Jesse D. Bloom

January 13, 2018

Below, the original comments *are in blue, and our responses are in black.*

SUMMARY

Your manuscript describes the use of a novel single cell RNAseq method to examine heterogeneity in viral mRNA production within single cells infected with single influenza A virions. You report that viral mRNA levels, as a proportion of total cellular mRNA, can vary significantly between infected cells, and this variability cannot be simply explained by the activity of innate anti-viral defenses. Further, correlating the levels of specific host transcripts with viral transcript levels allows you to identify pathways that tend to correlate with viral levels.

The strengths of the paper include the description of a novel method for leveraging single-cell RNAseq to quantify heterogeneity and stochasticity in viral gene expression during infection. This is a really important question for the field, and the approaches detailed here will likely by adopted by other groups studying other viruses. The authors do a good job of accounting for many of the pitfalls inherent in the experimental execution and data analysis. The manuscript is well written, and does an especially effective job of visually presenting complex datasets.

The main shortcoming of the paper is that it provides very little in the way of new information. As the authors clearly and honestly point out, most of the findings in this paper are simply confirming observations made in older papers. Additionally, the findings are purely descriptive, and provide little insight into the mechanisms that may give rise to the observations.

We believe that addressing the points below would help ameliorate some of the shortcomings of the paper.

This is a fair summary of our work. Thank you for recognizing the power of the approach and our efforts to fully describe pitfalls and effectively present the complex datasets.

As you note, our approach does not uncover any truly new biological phenomenon, and (as you also note) we honestly describe how our results end up mostly confirming older observations made using more indirect methods. However, there is substantial value to explicitly observing these processes in single cells, even if the findings are mostly in line with existing circumstantial evidence from bulk studies or less comprehensive single-cell approaches (e.g., flow cytometry). Our explicit measurements of transcript counts in single cells provide a more definitive and quantitative description of influenza virus infection than prior inferences from other approaches. Also, the extremeness of the heterogeneity across cells is surprising even in light of all that was previously known.

We have revised our manuscript in response to the many helpful suggestions made by the reviewers below. Below we detail these revisions. As you will see, we have been able to address all of the major questions and suggestions.

MAJOR COMMENTS

1. There are some obvious sources of variability that haven't been suitably discussed.

- The first is variation in the timing of the early stages of infection. No steps appear to have been taken to synchronize infection, or to limit secondary spread of the virus within the cultures. Could heterogeneity in the timing of binding/entry/fusion/trafficking explain a lot of the variation observed?

Our argument here would be that washed cells in 8-2 essentially do this. Alistair is generating data by flow with heterogeneity with washing +/- synchronization of infection. The thought is this will show that synchronization doesn't matter with respect to HA protein heterogeneity. This would be a new supporting figure.

The co-infection barcodes provide good evidence against much secondary spread. We can explain that.

- Another potential contributor is the cell cycle status of the infected cells. Did the host transcript data shed light on whether cell cycle status influenced viral transcription levels? This is especially important to address given that some of the genes showing association with the viral burden are cell-cycle related.

Cole had some suggestions of how we could do a few analyses that address this:

This is a super tricky thing to do - I have not seen computational methods that I trust described in the literature for detecting genes that co-vary with cell cycle. Aviv has a

method that seems reasonable at first blush but I have some concerns with in terms of the details. I have had better luck testing for genes that are differentially expressed between post-mitotic/quiescent and those that are actively cycling

What would first do is add up the transcript counts for well known markers of actively proliferating cells (CCNB2 is my favorite, but I you could sum a bunch of cyclins and CDKs) to get a ?proliferation index? for each cell. Then look at the distribution of this - I would expect it to be lognormal for a population of cells that are robustly growing. For cells that are mostly quiescent (e.g. a confluent epithelial plate), I would guess the transcript levels for the cells would be low, maybe with a long tail to the right. Then you would add the proliferation index values as a new column in the pData table in Monocle. Then you can use the index as a variable in the model formula for differentialGeneTest. This will give you genes that positively correlate with the proliferation index. The trouble is establishing what values of the index really correspond to actively cycling cells. Ideally, you?d have a bimodal distribution and you?d be able to set some clear threshold, but in practice this doesn?t usually happen. In general, this is pretty tricky to do without built-in controls (e.g. sets of cells that are sorted on DNA content, arrested with some drug, or what have you). So I am generally incline to take these sort of analyses with a huge grain of salt.

- In assessing to what extent lack of RNP expression accounts for the viral mRNA expression variability (page 6, paragraph 2), we think it is important to take into account potential extracellular contamination. Specifically, it would be more convincing if the analysis omitted cells with mixed wildtype/synonymous clones.

If we didn't already do this, we can do it easily — it shouldn't have much of an effect because mixed cells are rare.

We can potentially add a supplemental figure based on wildtype / HA-GFP co-infection.

- As the variability of the viral RNA load appears to be the central result of the manuscript, it would be useful to:

a. Employ a statistic to quantify the variability (e.g. entropy or Gini index)

b. Use simple models to illustrate how much variability can be accounted for by simple effects, such as expected Poisson co-infection frequency, or the likelihood of attaining full complement of RNP genes.

Gini index is something we could definitely calculate. Not clear what we could compare it to, but we can at least calculate it and report it. We then calculate Gini index under some simple model where everything is due to decreased transcription when missing segments. Any such model is probably pretty bogus and oversimplified, so we'd want to do this in

2. While the silent tagging method used to address co-infection is clever and appreciated, the issue is not fully settled. The dismissal of co-infection as a factor influencing cell-to-cell variation in lines 127-9 is based on too few cells to draw any conclusions, and thus needs to be tempered. Also, there is likely to be a significant amount of cryptic co-infection with identical barcode viruses (expected to be similar to that of mixed barcodes, 10%) that could influence measured heterogeneity. These points should be made in the text.

These are very good points. We have added the following text that clearly draws out the fact that our approach will only identify half the co-infections:

> In addition, the synonymous viral barcodes only identify co-infections by viruses with different barcodes – since the barcodes are at roughly equal proportion, we expect to miss about half of the co-infections. Since we annotate about ∼10% of the infected cells as co-infected by viruses with different barcodes (Figure 4D), we expect another ∼10% of the infected cells to also be co-infected but not annotated as so by our approach.

The reviewers are correct that our statement about the role of co-infection in influencing cell-to-cell variation was too strong. We have clarified this statement. We have also added a figure supplement (Figure 4-Figure supplement 4) which uses an independent method (flow cytometry staining of individual viral proteins) to support the fact that even at high infectious dose both low- and high-producing cells remain, although their relative proportions change. This figure supplement also supports our clarified statement that viral mRNA (or protein) production is not a simple continuous function of infectious dose, but is also affected by other sources of heterogeneity. The new text reads:

> Notably, Figure 4E shows that there are co-infected cells with both low and high amounts of viral mRNA, suggesting that the initial infectious dose does not drive a simple continuous increase in viral transcript production. In support of this view, we used flow cytometry to quantify the levels of individual viral proteins in cells infected at various MOIs or for which we could delineate co-infection status (Figure 4-Figure supplement 4). This analysis shows that sub-populations of cells that express similarly low and high levels of viral proteins persist across a wide range of infectious doses, although co-infection can influence the relative proportion of infected cells that fall into these sub-populations (Figure 4-Figure supplement 4).

3. The analysis of host cell transcripts positively or negatively associated with high viral mRNA expression is pretty minimal. Do the host genes identified here match up with the results of previous studies that screened for host pro- and anti-viral factors (reviewed in Watanabe et al. DOI: 10.1016/j.chom.2010.05.008)? Also, targeted gene knockdown or

4

over-expression experiments could help establish the causality underlying these relationships.

In response, gently point out that this isn't actually the point.

It's not that analysis is minimal, it's that there isn't anything.

Then we can somehow work in the Watanabe reference.

We have the oxidative stress thing to add as a figure supplement.

4. All analysis of viral gene expression seems to be at the segment level. How do different transcripts expressed from the same gene segment compare (i.e. NS1 and NEP)?

This is an important point that we failed to adequately discuss in the original manuscript. As the reviews note, two of the influenza transcripts (M1 / M2 and NS1 / NS2) are generated from alternative splicing of gene segments. However, each pair of these transcripts share the same 3' end. Most current single-cell mRNA sequencing strategies, including the 10X platform that we use, can only sequence the 3' end of the transcript. The reason is that the cell barcode is appended to this end, and so Illumina sequencing can only extend to regions proximal to the 3' end if they are still to capture the barcode (as is required for single-cell mRNA sequencing). Therefore, these techniques cannot accurately identify alternative splicing – this is simply an inherent limitation of the current techniques. We now clearly explain this fact in the following added text:

> Note that influenza virus expresses ten major gene transcripts from its eight gene segments, as the M and NS segments are alternatively spliced to produce the M1 / M2 and NS1 / NEP transcript, respectively (Dubois et al, 2014). However, an inherent limitation of current established single-cell mRNA sequencing techniques is that they only sequence the 3' end of the transcript (Zheng et al., 2017; Macosko et al., 2015; Klein et al., 2015; Cao et al, 2017). Since the alternative spliceoforms M1 / M2 and NS1 / NEP share the same 3' ends, we cannot distinguish them and therefore will refer simply to the combined counts of transcripts from each of these alternatively spliced segments as the M and NS genes.

5. The determination of the minimum required influenza fraction (Figure 4) is based on sound logic; however, we are concerned that the observed results do not fully align with this model. Specifically, while the 10hr experiment looks reasonable, the distributions of wildtype/synonymous mixed cells in other experiments do not appear to show the same trend (most notable for 6hr and 8hr samples, where almost no mixed cells are observed at low fractions). In that regard, using the 10hr dataset to estimate thresholds for all of the other samples does not seem appropriate.

This is a good point. We now use the procedure in Figure 4C to call thresholds *separately*

for each sample as the reviewers suggest. The sample-specific calling is shown in Figure 4-Figure supplement 3. These new thresholds lead to a modest increase in the number of cells called as virally infected for the earlier timepoint samples. These newly called cells therefore slightly shift the quantitative values in all subsequent analyses, although the changes are very small and all results remain unchanged.

The reason that the earlier timepoints have less mixed barcodes is because these earlier samples have much less total viral mRNA – and when their is less viral mRNA in total, there is also less available to be acquired from lysed cells.

MINOR COMMENTS

- Fig 2: Normalizing high versus low DI stocks based on equal TCID50 to compare IFN induction is not appropriate since these stocks will differ in total particle number. In this experiment, the cells would receive a much higher overall dose of viral RNAs with the high DI stock compared to the others.

The reviewers are completely correct. High DI stocks have more physical virions per infectious unit (e.g., TCID50), as indeed our own data in Figure 2A,B show. Therefore, the fact that Figure 2C shows that the high DI stocks induce more IFN per TCID50 does *not* imply that the high DI stocks are more IFN inducing per physical virion.

We have clarified that we are *only* using Figure 2C to show that the high DI stocks induce more IFN per TCID50, and that this could be due entirely to more physical virions. This is an important point for readers less versed in how viral passaging / DI content can influence IFN induction, since many studies simply report on the TCID50s used and provide no information on stock purity. This fact likely explains the discrepancy in IFN induction in our study and prior bulk studies (e.g., *eLife*, 5:e18311) on A549 cells that used viruses passaged at high MOIs that almost certainly contained many DI particles.

The revised text clearly explains how Figure 2C is comparing stocks with equal TCID50 but not equal physical virions. Specifically, we have added the following statement to the legend for Figure 2C, and similar sentences in the manuscript main text:

> *Our viral stocks are less immunostimulatory than virus propagated at high MOI when used at the same number of infectious units as calculated by TCID50. Note that this fact does not necessarily imply that they are more immunostimulatory per virion, as the high-MOI stocks also have more virions per infectious unit as shown in the first two panels (Figure 2A,B).*

- Fig. 3B: While I appreciate the density of information presented by this figure, the authors' main point concerning heterogeneity in the viral proportion of total mRNA reads

6

The reviewer points out that Figure 3B is a complicated way to show the heterogeneity in total mRNA. In response to this comment, we have also added a more straightforward presentation in Figure 3 - Figure supplement 1, which shows the distribution of the viral proportion as a cumulative fraction plot. This new plot should address the reviewer's point.

Incidentally, the most obvious way to plot these data would initially simply seem to be a histogram or kernel density plot with fraction of mRNA from virus on the x-axis, and proportion of cells on the y-axis. However, such a plot turns out to be visually un-interpretable, because so much of the density is at zero (un-infected cells) that it is impossible to see any information for other parts of the plot for any reasonable y-axis scaling. This problem can be remedied by *only* showing infected cells (getting rid of the huge bump in density at zero due to un-infected cells) – but that can only be done after we have introduced a criterion to define which cells are infected, which doesn't happen until later in the text (and we want to show all the data before this censoring). Therefore, we also plot a more straightforward histogram display in Figure 4E for only the infected cells after censoring un-infected cells. We believe the combination of Figure 3B, the new Figure 3 - Figure - supplement 1, and Figure 4E effectively capture all the effects given these complexities of visual display for data distributed in such an asymmetric way.

- Line 110: Should say "likely reflecting co-infection"

Thank you for pointing this out, we have made this fix.

- Lines 186-8: This statement doesn't make much sense since you are talking about variation in mRNA proportion. Global effects such as cell size should not affect proportion.

This is a good point. Since the statement in question is tangential to our central conclusions, we have just removed the lines in question.

- Lines 193-8: Is this really remarkable? As the authors point out, there is an extensive body of work demonstrating complementation during IAV infection.

We have removed the word "Remarkably" and just pointed out that in some cells, co-infection is responsible for providing the full gene complement. This new text more clearly explains that it has long been inferred that co-infection can complement missing segments. Our contribution is to directly observe single cells where we can identify specific genes that come from different co-infecting viruses, thereby providing an explicit single-cell characterization of this phenomenon.

- Line 329: "characterizing" misspelled.

This typo has been fixed.

7

- Lines 330-4: This statement about a confluence of factors leading to extreme cell-to-cell heterogeneity should be reworded since the data here actually doesn't show a clear role for innate immunity in promoting heterogeneity and there is no significant evaluation of the effect of host cell state on heterogeneity.

needs to be re-worded

- The "replicate" 8-hr sample isn't really a replicate since a different MOI was used. Also, the purpose of the differences in treatment for this sample compared with the 8-hr sample are not well explained.

Good point that we'll address already in first comment.

-Figure 3 is not effective in supporting the observations made in the text (that most cells have little viral RNA, and that there's a wide variability in the amount of viral RNA). Separate distributions for the viral RNA would be more telling. Log scale would be more appropriate for showing counts per cell.

We have added Figure 3-Figure supplement 1, which uses a log scale to show the distribution of the amount of viral mRNA per cell. We still also retain Figure 3 to show the non-transformed data. Many cells have no viral mRNA, and it is not possible to show these non-infected cells on a logged scale, so the new combination of Figure 3 and Figure 3-supplement 1 provides both linear and log-scale views. Simple histograms or kernel density plots of the distribution of viral mRNA are not effective until we have removed non-infected cells (as is done in Figure 4E after setting the thresholds for calling infection) since the distribution is so strongly peaked at the low end.

-Figure 4C: the threshold line appears to be a bit to the right of the global maximum - going through the description, it was not clear why that should be the case.

This is a good observation. We were determining the threshold as the maximum of a loess fit to the data. Previously, the loess fitting parameters caused the fit line to overshoot, putting the threshold a bit to the right. We have adjusted span parameter in the loess fitting, which fixes the overshoot. The revised Figure 4C shows the loess fit in orange as well as the threshold in dotted green – and the threshold now goes right through the maximum.

-Why was media change skipped in the replicate experiment?

Needs to be explained

-Consider using different color schemes for Fig 4A and B, as different variables are being shown.

This is a good suggestion. We now use different color schemes for Figures 4A and 4B. We also changed the color scheme for Figure 4E to be different than those for 4A and 4B since

8

it also shows a different variable.

Thank you for pointing out this error in the legend, it has been fixed.