Transcriptional dynamics of influenza virus infection at the single-cell level

Alistair B. Russell[1], **Cole Trapnell**[2], **Jesse D. Bloom**[1,2*]

***For correspondence:** jbloom@fredhutch.org

[1]Basic Sciences Division and Computational Biology Program, Fred Hutchinson Cancer Research Center, Seattle, United States; [2]Department of Genome Sciences, University of Washington, Seattle, United States

Abstract Influenza virus infection induces large changes in cellular transcription. Previously this has mostly been looked at using bulk measurements Here we examine the process at the level of single cells. We find extremely wide variation in the extent of viral gene transcription across infected cells. IFN induction is very rare. Some cellular pathways may be consistently altered in cells with high burden of viral transcripts. Overall, highlights remarkable heterogeneity in the outcome of infection.

Introduction

Heterogeneity is important in a lot of cellular processes even when isogenic (**??**).

Population (genetic) heterogeneity is also important. Viral quasispecies, cancer single-cell, etc. Salmonella paper (PhoP).

Literature on viral burst-size heterogeneity. This goes back to Delbruck, Andino polio paper (**?**), the MDCK / flu paper.

Discuss segmented nature of influenza. Maybe in the context of how this could further increase heterogeneity because there is a lot of potential for entire genes to be missing. Includes Yewdell and Lowen papers.

Results

Initial Steps to Generate Robust Single-Cell Kinetic Observations.

In order to obtain detailed measurements at a single-cell resolution of influenza growth and replication we chose to use an approach dependent on a commercially available variant of DropSeq and InSeq (Figure 1A). In brief, we infected A549 cells — a human lung carcinoma cell line frequently used in tissue-culture models of influenza infection — using a relatively low infectious dose of WSN influenza — a lab-adapted trypsin-independent strain. We then harvested cells by trypsonization, and generated libraries from total polyadenylated transcripts. These libraries consisted of a short region mapping to the 3′ end of the original mRNA transcript, a barcode identifying the cell from which the transcript was derived, and a unique molecular identifier that serves for error-correction in transcript abundance measurements. As an additional control, we generated a 3′ barcoded influenza strain containing two one base-pair synonymous mutations at the 3′ end of all transcripts, allowing us to estimate co-infection and empirically derive thresholds for influenza infection. These barcodes were designed to lie within the 3′ region sequenced in our library preparation as determined through examination of mapping distributions in uninfected A549 cells.

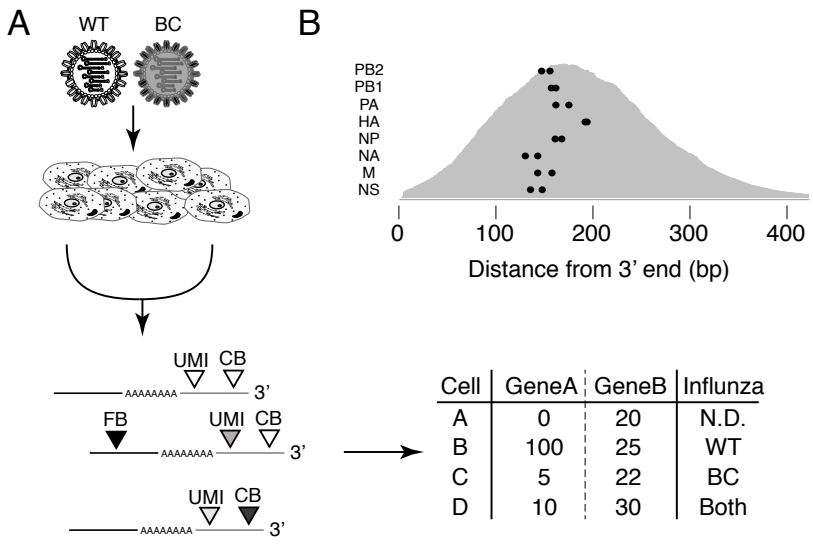


Figure 1. CAPTION

When generating viral stocks for our infections, we also considered that Influenza, like many viruses, can generate a diverse array of biologically active particles as it replicates — particles that can differ significantly from the genomic consensus sequence. One of the most abundant of these, defective-interfering (DI) particles, predominantly result from deletions accrued in the three segments encoding the tripartite polymerase during growth at high infectious doses. Generation of these particles is highly stochastic and known to influence the behavior of influenza populations both with respect to replication kinetics and interferon induction. Therefore, for the sake of generating baseline, repeatable, observations of influenza replication kinetics and host interactions at a single cell level, we first sought to propagate populations of viruses that were relatively free of these particles using similar methods to those developed by *Xue et al*. Concomitantly, we validated that both replication kinetics and relative abundance of DI particles were similar between our synonymously barcoded and wild-type populations.

First, we established that our barcoded variant progresses through infection with near-wild-type kinetics, as measured by hemagglutanin (HA) transcript abundance (Figure 2A). After establishing that our mutations did not dramatically impact viral replication, we next sought to establish a rough ratio of biologically active particles to fully-infectious particles. We therefore measured the abundance of the HA segment in viral stocks relative to their measured infectious dose (Figure 2B). Notably, this ratio is dramatically skewed in a population grown to be replete in defective particles relative to our desired experimental populations. Measurements of HA staining corroborated this finding, as the percentage of infected cells tracked well with the desired MOI, as opposed to a dramatically increased abundance of HA staining noted in our high defective control (Figure 2C). Lastly, our populations induced far less interferon than a high defective control, consistent with prior observations that such stocks are more immunostimulatory (Figure 2D).

Using these validated populations, we tracked growth kinetics in our target cell line at an infectious dose targeting a final infection rate of 5 percent as determined by flow cytometry. We observed rapid transcript accumulation through 10 hours, which then tapered off — potentially representing new rounds of infection rather than progression in initially infected cells (supplemental figure). We then chose three time points (6,8, and 10h) reflecting dramatically different influenza burden (and thus different stages in influenza replication) for this study.

(Figure 1).

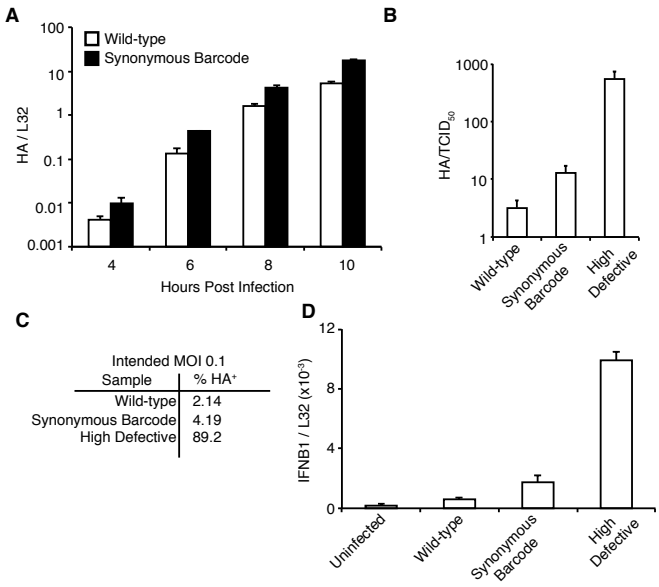


Figure 2. CAPTION

Population-Level Measurements Validate Dataset.

Before proceeding with more detailed analyses, we first examined population-level data to confirm the validity of our approach. We not only generated data for our three time points and a matched uninfected control, but also performed a biological replicate at 8 hours to determine whether our findings could be dominated by inter-sample or technical variation rather than biological processes driven by influenza. We sampled roughly equivalent number of cells per time point at an equivalent sequencing depth per cell - granting similar probability to observe rare events across datasets (Figure 3A) (ADD SUPPLEMENTAL OF DEPTH/CELL?). Overall we observed highly similar distributions of transcript abundance between samples (Figure 3B). This suggests that, in a rough global sense, transcriptome dynamics are relatively unperturbed and that influenza, if it is driving dramatic changes in transcript abundance and/or growth state, must be doing so in only a small subset of cells. More importantly, with the confidence that the average cellular state is similar between time points, we can better regard changes in subpopulations between time points as bone-fide progression in influenza replication rather than inter-sample variation.

In addition to analyses regarding global transcript abundance, we were also able to successfully detect increasing numbers of influenza transcripts as our experiment progressed (Figure 3C). Moreover, there appeared to be only a few cells at each time point making considerable contributions to the influenza transcript pool, consistent with our desired low level of infection. With these metrics, we felt confident proceeding to more detailed analyses of influenza replication kinetics and impact on these populations.

Synonymous Barcodes Provide An Empirical Means of Setting Infection Thresholds.

As a first step in understanding influenza replication at a single-cell level, we must derive a threshold for identifying infected cells. Crucial to our experimental design, we were able to detect both wild-type and synonymous influenza variants at all three timepoints in all eight segments, at levels permitting identification of cells infected with one, or both strains (Figure 4). This was true of about half of the transcripts surveyed, matching the read-density of our preliminary experiments. These barcodes partitioned non-randomly between cells, as expected if derived from

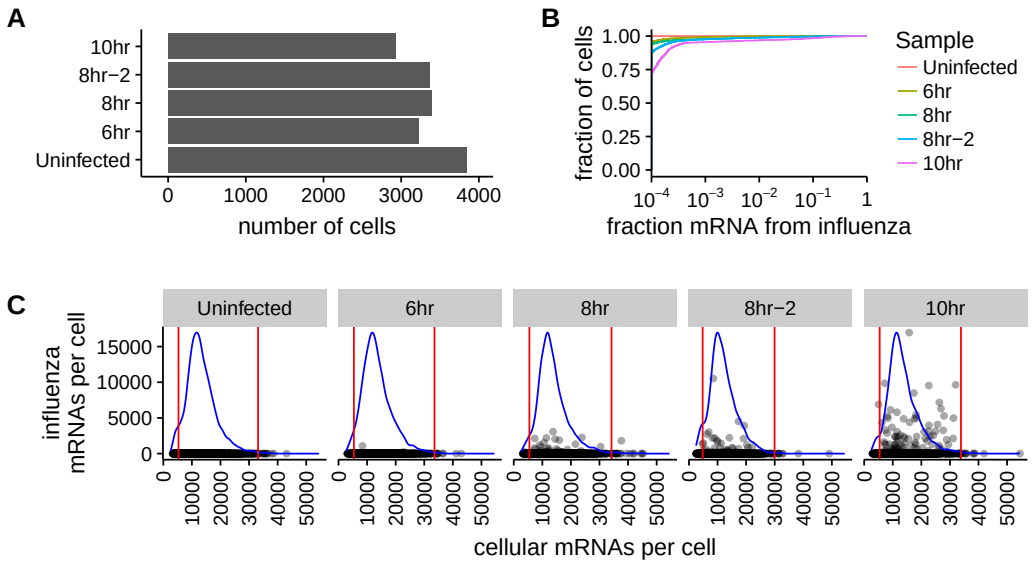


Figure 3. Overview of amounts of cellular and influenza virus mRNAs detected in each cell. **(A)** Number of cells captured for each sample. **(B)** Cumulative fraction plot showing the amount of mRNA derived from influenza for each sample. In all samples, most cells had little or no influenza mRNA. **(C)** The number of cellular and viral mRNAs for each cell is plotted as a point. The blue lines show the overall distribution of the number of cellular mRNAs per sample. Cells that fell outside the red lines were removed as outliers. At later timepoints, a small number of cells had a very high number of viral mRNAs.

Figure 3–Figure supplement 1. Shorter caption for main text.

Figure 3–Figure supplement 2. This is another supplementary figure.

Figure 3–source data 1. This is a description of a data source.

single infectious events (Figure 5A,B). This grants us the power to develop an empirical threshold for calling influenza-infected cells versus those that might have reads derived from lysis of infected cells. Reads derived from lysed cells should exhibit random partitioning ? sampling each barcode according to its overall abundance in the experiment. Therefore, when ordering all cells at all timepoints according to the number of influenza reads observed and the relative abundance of each barcode, we can set our detection threshold as the point at which barcodes reliably exhibit non-random partitioning - that is when all segments in a given cell are frequently derived from either our synonymous or wild-type virus but not both (Figure 5C). Notably, this strategy relies heavily on a relatively low MOI, as high rates of coinfection would stymie this approach (Figure 5D,E) .

Few coinfections

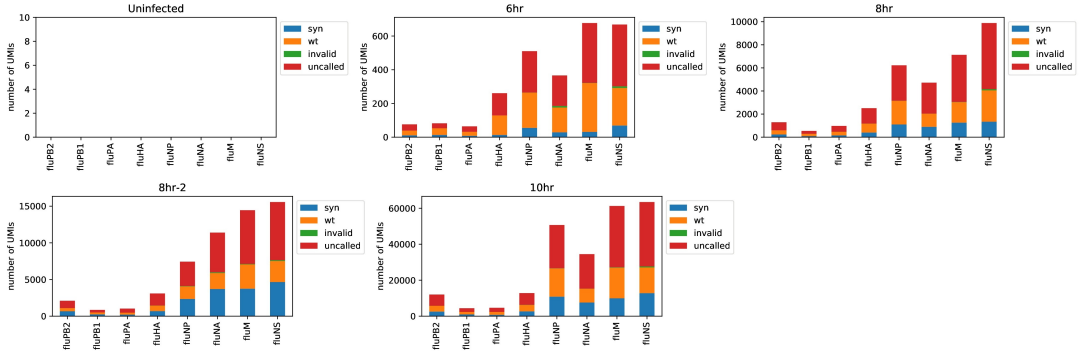


Figure 4. CAPTION

Setting threshold: Assumes rare barcode events are due to lysis (ormisannotation) and that two values SHOULD correlate Validating evidence/assumption two values correlate - flow data, flow diagram and quantile-normalized data. Include replicates in supplement.

Once defined, we have some statistics.

Presumably this subsection basically covers 5.

Influenza Exhibits A Broad Distribution of Productivity Partially Explained By Segment Absence.

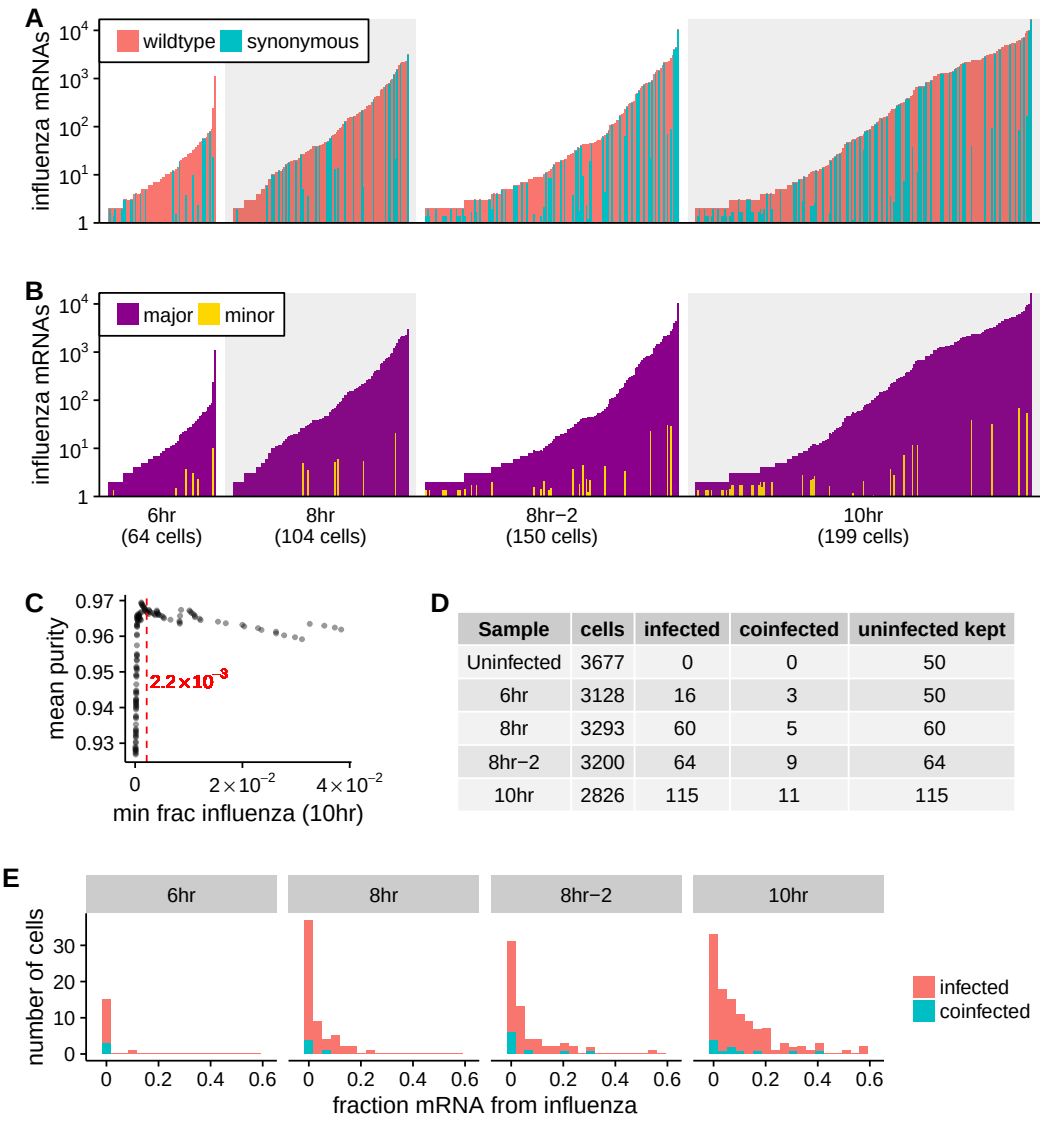
The absolute productivity of influenza infection as measured by the number of fully-infectious virions produced by a given cell is known to vary dramatically. This observation is recapitulated in our data, wherein of the small portion of infected cells, a bare few contribute the majority of influenza transcripts (Figure 3A) (MAYBE A MORE SPECIFIC SUPPLEMENTAL). As transcript abundance is highly dependent upon secondary transcript from de novo generated viral genomes, it can be presumed that when comparing between two cells, increased mRNA abundance is likely correlated with increased viral genome production, which in turn is highly predictive of viral productivity. It is thought that one driver of the observed variation in influenza productivity is the segmented nature of the influenza genome. Several groups have described that the majority of infections fail to express at least one of the eight influenza segments. Our datasets produce a similar pattern, even in cells that are otherwise expressing high levels of influenza transcripts (Figure 6A). When we bin cells according to absence of a given segment we observe that, unsurprisingly, absence of the tripartite polymerase (PB1, PA, PB2) and nuceoprotein (NP) correlates strongly with decreased influenza burden (Figure 6B).. Importantly, the gene products of all four of these segments are essential to the transition from primary mRNA transcription to replication and secondary mRNA transcription. This effect cannot be solely attributed to sampling error, as the polymerase genes exhibit the lowest expression and thus would be expected to be stochastically lost due to undersurveillance at much higher influenza transcript levels than would other genes.

Nevertheless, while segment absence can partially explain extreme components of influenza variability resulting from an inability to fully engage in secondary transcription, absence of the remaining segments does not appear to have any significant impact on influenza burden. Importantly, even cells expressing all eight segments still exhibit a wide distribution of influenza transcripts within each time point. We must instead therefore posit alternative mechanisms such as variability in host cells to explain the remainder of this distribution. Our data do differ slightly on this point than prior observations by Brooke et al, inasmuch as their report described decreased influenza burden not only associated with loss of nucleoprotein, but also of the NS segment. However, significant strain-to-strain differences exist in the behavior of NS1, a major interferon antagonist encoded by the NS segment, and these differences might explain this discrepancy.

Extrapolating these measurments to?.

Influenza Infections Are Predominated By Nonproductive Events.

We further wished to define metrics for presence/absence for a given segment and compare them to values obtained for other strains of influenza in order to understand the probability that a given infection could possibly produce infectious progeny. We decided to omit these calculations for PB1, PA, PB2, and NP. Using cyclohexamide treatment to limit cells to primary transcription alone, we find a 2-3 log reduction in mRNA levels of PB1 or HA relative to a 10h time point (supplemental figure). We therefore are not confident that our approach has the power to detect all semi-infectious events that result from a lack of polymerase or nucleoprotein segments. However, given the lack of effect absence of other segments appears to have on mRNA production, we proceeded to calculate the relative abundance of infectious events absent a given gene segment.



Sample	cells	infected	coinfected	uninfected kept
Uninfected	3677	0	0	50
6hr	3128	16	3	50
8hr	3293	60	5	60
8hr−2	3200	64	9	64
10hr	2826	115	11	115

Figure 5. Synonymous barcodes near the 3′ end of the influenza virus mRNAs were used to identify co-infection and distinguish true infections from cells that contained a few spurious viral reads. **(A)** For all cells with at least two viral mRNAs for which the synonymous barcode could be called, each line is proportional to the logarithm of the number of viral mRNAs in that cell. The bars are colored in linear proportion to the fraction of the viral mRNAs derived from either wildtype or synonymously barcoded virus. **(B)** Same as (A), but now each bar is colored according to the relative proportions of the more common (major) and less common (minor) barcoded virus variant. At low levels of viral mRNA there is often a roughly equal mix of barcodes, since many of these cells have simply picked up environment mRNA which is about equally likely to derive from either virus. But at higher levels of viral mRNA, truly infected cells are mostly one pure barcode except for a few cells that are truly co-infected. **(C)** We determined a cutoff for calling "true" infections by fitting a curve to the mean barcode purity of all cells with greater than a given fraction of their mRNA derived from virus. We called the cutoff at the point at which purity stops increasing with the fraction of viral mRNA. **(D)** The number of cells identified as infected and co-infected for each sample. For all samples, the vast majority of cells were not infected, so for subsequent analyses we subsampled to a number of uninfected cells that was the greater of 50 or the number of infected cells. **(E)** The distribution of the fraction of mRNA derived from virus for each sample for both infected and co-infected cells. For all samples, there is a very wide distribution of the amount of viral mRNA.

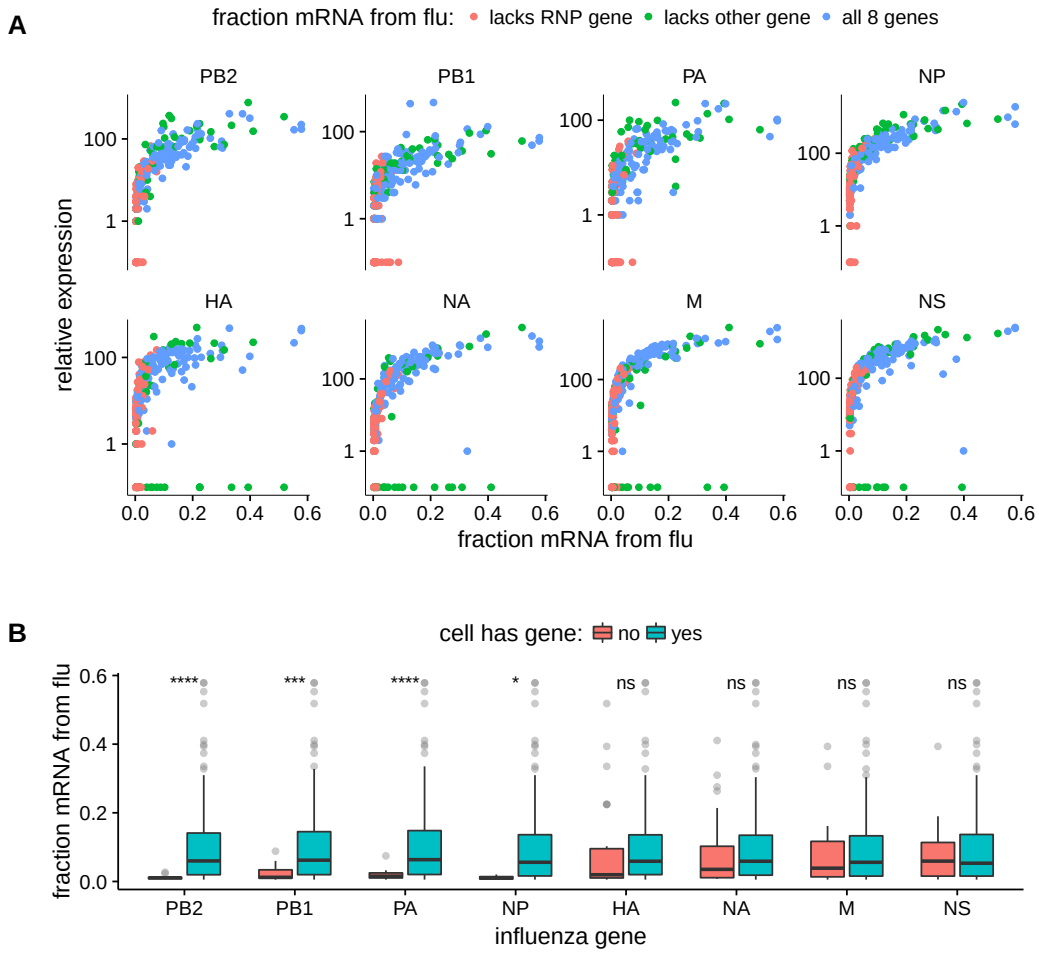


Figure 6. The viral infection burden in individual cells as a function of the amount of each viral gene detected. **(A)** Fraction of mRNAs in each cell derived from virus as a function of the *normalized* expression of each viral gene in that cell. This plot shows that all cells with very high viral burden express all of the RNP genes, but some cells with high viral burden lack each of the other four viral genes. **(B)** Statistical tests confirming that absence of viral RNP genes is significantly associated with reduced viral burden, but that the absence of the non-RNP genes does not lead to a clear decrease in viral burden.

Influenza mRNA Ratios are Tightly Controlled

This and other reports have described the considerable heterogeneity produced as influenza infection progresses in tissue culture models. However, in stark contrast to our analyses above, the ratios of mRNAs generated by the eight segments appear to behave in a highly stereotyped fashion, with values roughly consistent with prior Northern Blot measurements of Influenza mRNA abundance. Staggeringly, this is true whether or not cells are expressing mRNAs from all eight segments, at all three time points, and, in cells running the gamut of infectious burden — although we must limit these conclusions to those cells engaging in secondary transcription given the caveats described above. Therefore while either noise in influenza replication or host variability can dramatically impact the amount of Influneza mRNA produced as a whole, these processes seem to, ignoring complete absence of a segment, act equally across the Influenza genome. This is consistent with more limited measurements showing positive correlations between several segments in vRNA measurements in single cells. It would therefore be inadvisable to search for the sources of the broad distribution in influenza replication in those steps that can impact segments independently of one-another, such as noise in the transition to cRNA or vRNA production, but rather only in those processes that could impact the kinetics of all segments equivalently, such as transport to the nucleus or availability of cellular resources.

Coinfection Occurs Early and Is Likely Adaptive For Influenza

While relatively rare in our dataset, those cells positively identified as containing both wild-type and synonymously barcoded influenza provide important insight into the nature of coinfection. It has been posited that coinfection can rescue nonproductive infections, potentially by providing missing genome equivalents. Brooke et al., for example observed increased co-occurrence of a given gene-segment pair as MOI was increased. When we examine our coinfected cells, we observe a much higher rate of expression of all eight segments (metric). Consistent with this model, we also observe that if we consider only one barcode or the other, that many of these infections would instead have been non-productive. Another observation that can be immediately procured is that, on the whole, coinfections have roughly balanced amounts of wild-type and barcoded segments. Others have noted that windows for coinfection are relatively narrow, and, these data directly suggest that this window is specifically narrow with respect to cRNA and vRNA production. If significant replication of viral genomes had already occurred, one would expect a much wider distribution of ratios than those observed. To support these observations, we used a virus in which the HA segment was replaced by GFP, allowing us to detect co-infection by this modified virus and wild-type WSN by flow cytometry. When we examine strictly co-infected cells, we observe that, as in our single-cell analysis, both HA and GFP are highly correlated — indicating a tight window of coinfection — and that a peak of low HA — likely corresponding to primary transcription alone — is greatly reduced. These data give credence to the sentiments of others that non-productive events are perhaps just productive events waiting to occur, and that the true dynamics underlying descriptions of ?MOI? are far more complex than we frequently give them credence and rely on assumptions that frequently prove incorrect.

Relative expression of different viral genes

Figure 8 shows some of this. Cite **?** to show our relative expressions are consistent with that. Maybe show a panel with just the 8-segment infections.

Maybe add a figure breaking this down among timepoints.

Host stuff

Discussion

Morbi luctus, wisi viverra faucibus pretium, nibh est placerat odio, nec commodo wisi enim eget quam. Quisque libero justo, consectetuer a, feugiat vitae, porttitor eu, libero. Suspendisse sed mauris vitae elit sollicitudin malesuada. Maecenas ultricies eros sit amet ante. Ut venenatis velit.

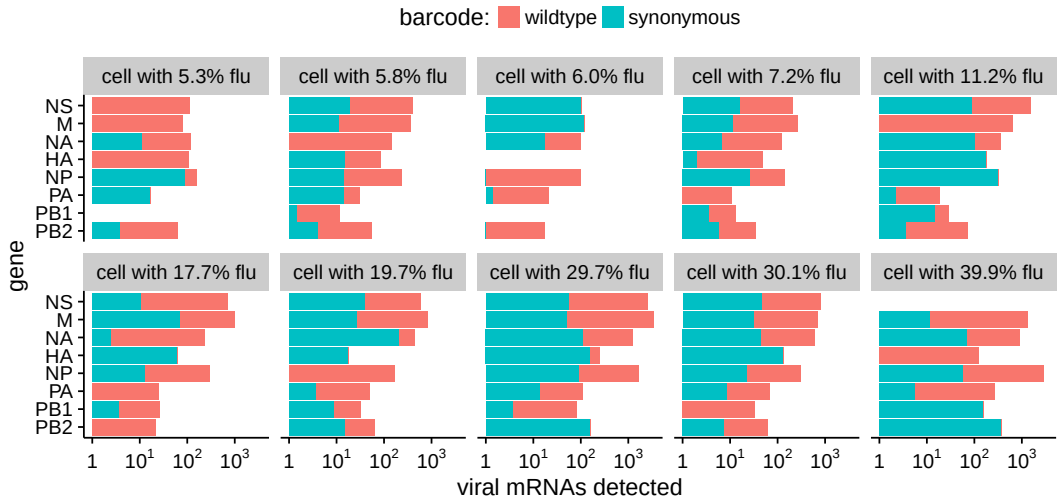


Figure 7. Frequency of each viral gene segment in co-infected cells with at least 5% of their mRNA derived from influenza. The bars indicate the logarithm of the number of each viral mRNA detected, and the bars are colored in proportion to the fraction of those mRNAs that are derived from either wildtype or synonymous barcoded virus.

Figure 7–source data 1. The raw data plotted in this figure are in `p_coinfection.csv`.

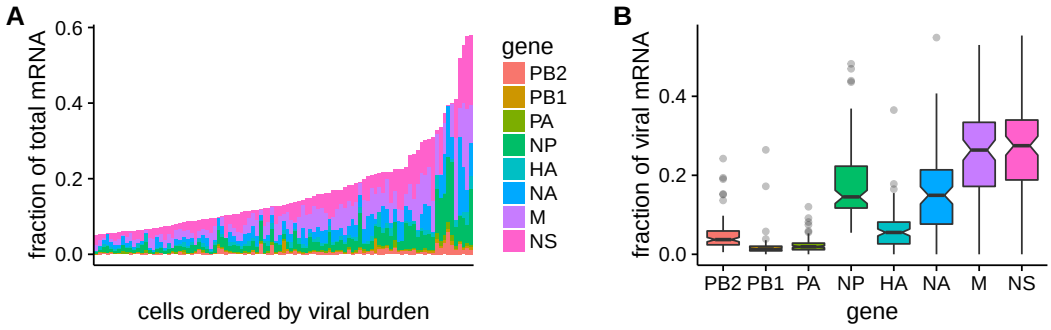


Figure 8. Expression of individual influenza genes in highly infected cells (at least 5% of total mRNA is viral). **(A)** The fraction of total mRNA from each influenza gene for each cell. **(B)** Box plots showing the fraction of viral mRNA per cell that is derived from each influenza gene taken over all highly expressed cells. The black line at the notch in each box is the median, and the top and bottom of the box indicate the first and third quartiles.

Figure 8–source data 1. The raw data are in `p_flu_expr.csv`.

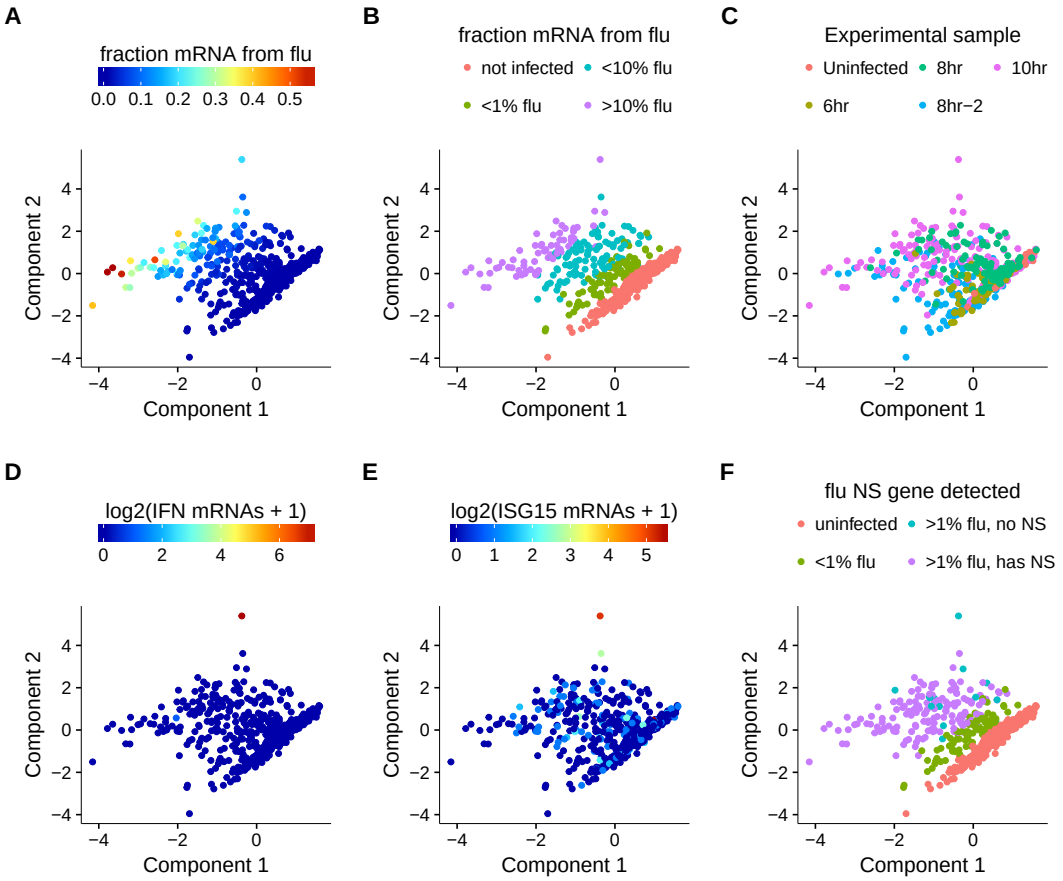


Figure 9. Visual layout of the cells according to "pseudotime". The layout is the same in all panels, but each panel colors the cells according to a different property. **(A), (B)** Cells colored by the fraction of their mRNA that is viral. **(C)** Cells colored by experimental sample. While it is clear that cells from later timepoints often have more viral RNA, there are cells from earlier timepoints with a high viral burden and cells from late timepoints with a low viral burden. **(D)** Cells colored by the number of type I and III interferon transcripts detected. Only one cell has high expression of these interferons. **(E)** Cells colored by the expression of the interferon-stimulate gene ISG15. **(F)** For cells with at least 1% of their reads from influenza, are the cells expressing the viral NS protein? The one interferon-positive cell is lacking NS, but many other cells also lack NS but do not express interferon.

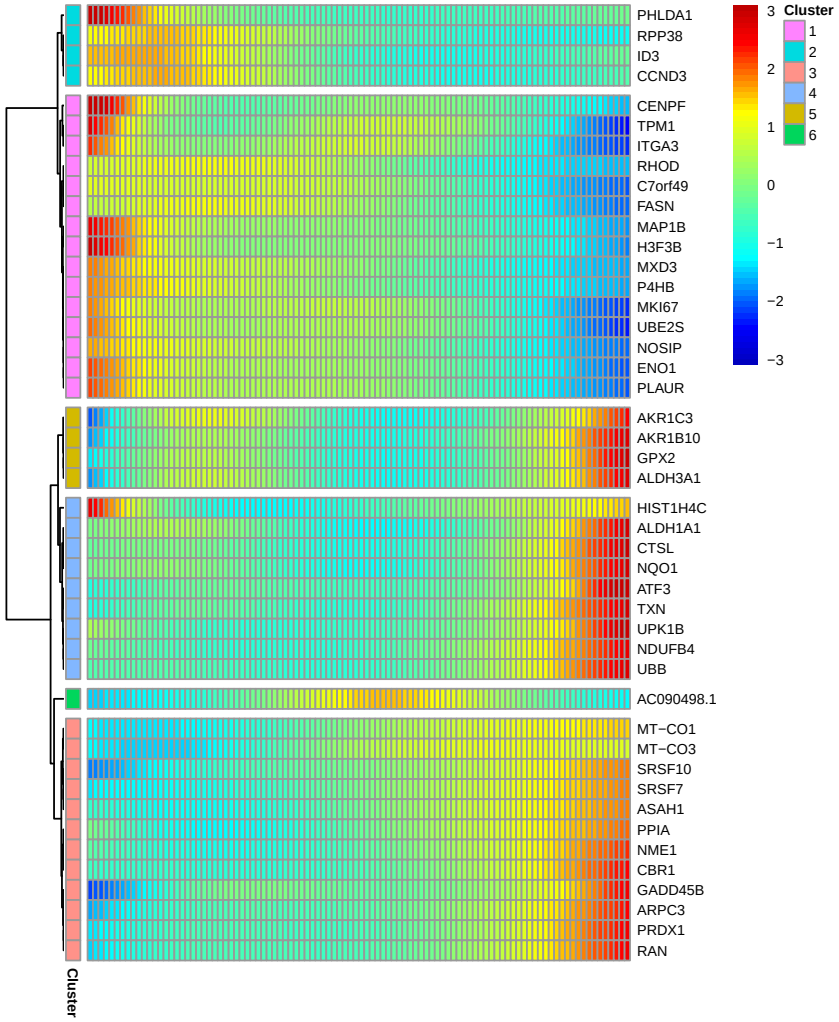


Figure 10. Cellular genes that are differentially expressed with respect to the amount of influenza mRNA in individual cells infected with full influenza virus containing all eight genes. Shown are all genes differentially expressed with $Q < 0.1$.

Figure 10–source data 1. The full results of the differential expression test is in `p_sig_cellular_genes.csv`.

Figure 10–source data 2. The results of a gene-set analysis are in `p_sig_cellular_genes.csv`.

Maecenas sed mi eget dui varius euismod. Phasellus aliquet volutpat odio. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Pellentesque sit amet pede ac sem eleifend consectetuer. Nullam elementum, urna vel imperdiet sodales, elit ipsum pharetra ligula, ac pretium ante justo a nulla. Curabitur tristique arcu eu metus. Vestibulum lectus. Proin mauris. Proin eu nunc eu urna hendrerit faucibus. Aliquam auctor, pede consequat laoreet varius, eros tellus scelerisque quam, pellentesque hendrerit ipsum dolor sed augue. Nulla nec lacus.

Methods and Materials

Guidelines can be included for standard research article sections, such as this one.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Some LaTeX Examples

Use section and subsection commands to organize your document. LaTeX handles all the formatting and numbering automatically. Use ref and label commands for cross-references.

Figures and Tables

If you use the following prefixes for your `\label`:

Figures `fig:`, e.g. `\label{fig:view}`
Tables `tab:`, e.g. `\label{tab:example}`
Equations `eq:`, e.g. `\label{eq:CLT}`
Boxes `box:`, e.g. `\label{box:simple}`

you can then use the convenience commands as in `\FIG{cells}`, to generate cross-reference **??**.

Citations

LaTeX formats citations and references automatically using the bibliography records in your .bib file, which you can edit via the project menu. Use the `\cite` command for an inline citation, like **?**, and the `\citep` command for a citation in parentheses (**?**). The LaTeX template uses a slightly-modified Vancouver bibliography style. If your manuscript is accepted, the eLife production team will re-format the references into the final published form. *It is not necessary to attempt to format the reference list yourself to mirror the final published form.*

Acknowledgments

Additional information can be given in the template, such as to not include funder information in the acknowledgments section.

Figure 3–Figure supplement 1. This is a supplementary figure's full caption, which will be used at the end of the manuscript.

Figure 3–Figure supplement 2. This is another supplementary figure.