

# 1 Extreme heterogeneity of influenza 2 virus infection in single cells

3 Alistair B. Russell<sup>1</sup>, Cole Trapnell<sup>2</sup>, Jesse D. Bloom<sup>1,2\*</sup>

\*For correspondence:  
jbloom@fredhutch.org

4 <sup>1</sup>Basic Sciences Division and Computational Biology Program, Fred Hutchinson Cancer  
5 Research Center, Seattle, United States; <sup>2</sup>Department of Genome Sciences, University of  
6 Washington, Seattle, United States

7

8 **Abstract** Viral infection can dramatically alter a cell's transcriptome. However, these changes  
9 have mostly been studied by bulk measurements on many cells. Here we use single-cell mRNA  
10 sequencing to examine the transcriptional consequences of influenza virus infection. We find  
11 extremely wide cell-to-cell variation in the productivity of viral transcription – viral transcripts  
12 comprise less than a percent of total mRNA in many infected cells, but a few cells derive over half  
13 their mRNA from virus. Some infected cells fail to express at least one viral gene, but this gene  
14 absence only partially explains variation in viral transcriptional load. Despite variation in viral load,  
15 the relative abundances of viral mRNAs are fairly consistent across infected cells. Activation of  
16 innate immune pathways is rare, but some cellular genes co-vary in abundance with the amount of  
17 viral mRNA. Overall, our results highlight the complexity of viral infection at the level of single cells.

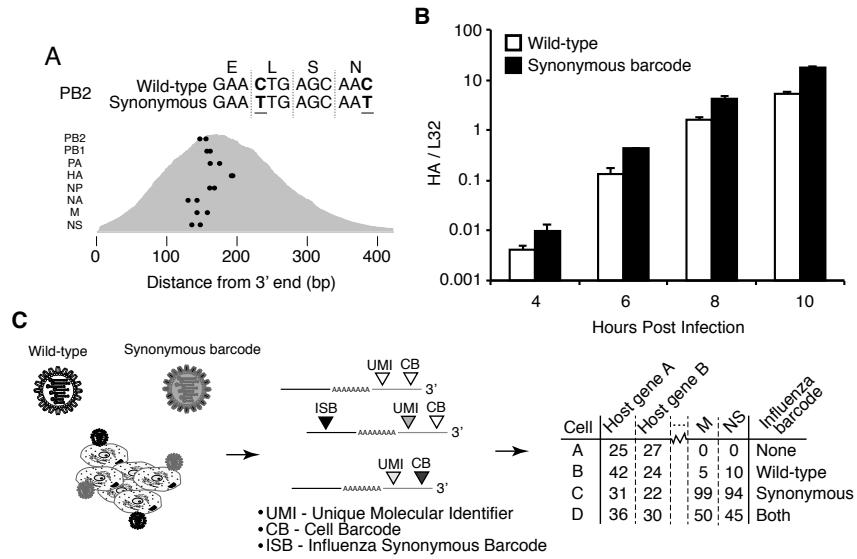
18

## 19 Introduction

20 Viruses can cause massive and rapid changes in a cell's transcriptome as they churn out viral mRNAs  
21 and hijack cellular machinery. For instance, cells infected with influenza virus at high multiplicity  
22 of infection (MOI) express an average of 50,000 to 100,000 viral mRNAs per cell, corresponding  
23 to 5 to 25% of all cellular mRNA (Hatada et al., 1989). Infection can also trigger innate-immune  
24 sensors that induce the expression of cellular anti-viral genes (Killip et al., 2015; Iwasaki and Pillai,  
25 2014; Crotta et al., 2013). This anti-viral response is another prominent transcriptional signature of  
26 high-MOI influenza virus infection in bulk cells (Geiss et al., 2002).

27 However, initiation of an actual influenza infection typically involves just a few virions infecting  
28 a few cells (Varble et al., 2014; Poon et al., 2016; Leonard et al., 2017; McCrone et al., 2017). The  
29 dynamics of viral infection in these individual cells may not mirror bulk measurements made  
30 on many cells infected at high MOI. Over 70 years ago, Max Delbrück showed that there was a  
31 ~100-fold range in the number of progeny virions produced per cell by clonal bacteria infected  
32 with clonal bacteriophage (Delbrück, 1945). Subsequent work has shown similar heterogeneity  
33 during infection with other viruses (Zhu et al., 2009; Schulte and Andino, 2014; Combe et al., 2015;  
34 Akpinar et al., 2016), including influenza virus (Heldt et al., 2015).

35 In the case of influenza virus infection, targeted measurements of specific proteins or RNAs  
36 have shed light on some factors that contribute to cell-to-cell heterogeneity. The influenza virus  
37 genome consists of eight negative-sense RNA segments, and many infected cells fail to express  
38 one more of these RNAs (Heldt et al., 2015; Dou et al., 2017) or their encoded proteins (Brooke  
39 et al., 2013). In addition, activation of innate-immune responses is inherently stochastic (Shalek  
40 et al., 2013, 2014; Bhushal et al., 2017; Hagai et al., 2017), and only some influenza-infected cells  
41 express anti-viral interferon genes (Perez-Cidoncha et al., 2014; Killip et al., 2017). However, the  
42 extent of cell-to-cell variation in these and other host and viral factors remains unclear, as does the



**Figure 1.** Experimental design. **(A)** We engineered a virus that carried two synonymous mutations near the 3' end of each mRNA. At top are the mutations for PB2. At bottom are locations of the synonymous mutations relative to the typical distribution of read depth for our 3'-end sequencing. **(B)** The wild-type and synonymously barcoded viruses transcribe their genes with similar kinetics. The abundance of the viral hemagglutinin (HA) transcript relative to the cellular housekeeping gene L32 was assessed by qPCR in A549 cells infected at an MOI of 0.5 (as determined on MDCK-SIAT1 cells). Error bars  $\pm$  S.D., n=3. **(C)** For the single-cell mRNA sequencing, A549 cells were infected with an equal mixture of wild-type and synonymously barcoded virus. Immediately prior to collection, cells were physically separated into droplets and cDNA libraries were generated containing the indicated barcodes. The libraries were deep sequenced, and the data processed to create a matrix that gives the number of molecules of each transcript observed in each cell. Infected cells were further annotated by whether their viral mRNAs derived from wild-type virus, synonymously barcoded virus, or both.

**Figure 1-source data 1.** Sequences of wild-type and barcoded viruses are in [viralsequences.fasta](#).

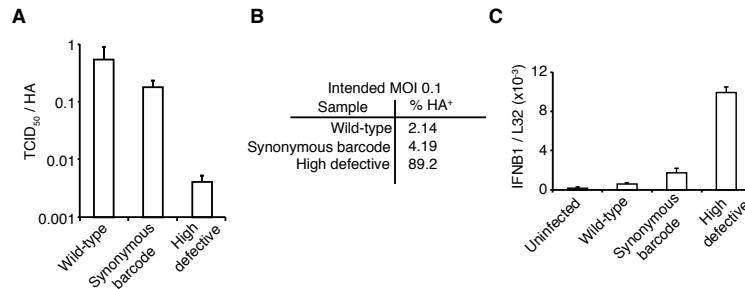
43 association among them in individual infected cells.

44 Here we use single-cell mRNA sequencing to quantify the levels of all cellular and viral mRNAs  
45 in cells infected with influenza virus at low MOI. We find extremely large variation in the amount  
46 of viral mRNA expressed in individual cells. Both co-infection and activation of innate-immune  
47 pathways are rare in our low-MOI infections, and do not appear to be the major drivers of cell-  
48 to-cell heterogeneity in viral transcriptional load. Individual infected cells often fail to express  
49 specific viral genes, and such gene absence explains some but certainly not all of the cell-to-cell  
50 heterogeneity. A variety of cellular genes, including ones involved in the oxidative-stress response,  
51 co-vary with viral transcriptional load. Overall, our work demonstrates remarkable heterogeneity in  
52 the transcriptional outcome of influenza virus infection among nominally identical cells infected  
53 with a relatively pure population of virions.

## 54 Results

### 55 Strategy to measure mRNA in single virus-infected cells.

56 We performed single-cell mRNA sequencing using a droplet-based system that physically isolates  
57 individual cells prior to reverse transcription ([Zheng et al., 2017; Macosko et al., 2015; Klein et al., 2015](#)).  
58 Each droplet contains primers with a unique *cell barcode* that tags all mRNAs from that  
59 droplet during reverse-transcription. Each primer also contains a *unique molecular identifier (UMI)*  
60 that is appended to each mRNA molecule during reverse transcription. The 3' ends of the mRNAs are  
61 sequenced and mapped to the human and influenza virus transcriptomes to determine transcript  
62 identities. This information is combined with that provided by the UMIs and cell barcodes to



**Figure 2.** The viral stocks in our experiments are relatively pure of defective particles. **(A)** Our viral stocks have a higher ratio of infectious particles to HA virion RNA compared to a high-defective stock propagated at high MOI. HA viral RNA was quantified by qPCR on virions. Error bars  $\pm$  S.D., n=6 (qPCR replicates). **(B)** Our viral stocks have a higher ratio of infectious particles to particles capable of expressing HA protein. A549 cells were infected at an MOI of 0.1, and the percentage of cells expressing HA protein at 9 hours post-infection was quantified by antibody staining and flow cytometry. **(C)** Our viral stocks are less immunostimulatory than virus propagated at high MOI. Measurements of *IFNB1* transcript by qPCR normalized to the housekeeping gene *L32* in A549 cells at 10 hours post infection at an MOI of 0.5. Error bars  $\pm$  S.D., n=3. Note that MOIs were calculated by TCID50 on MDCK-SIAT1 cells, whereas the experiments in this figure involved infection of A549 cells.

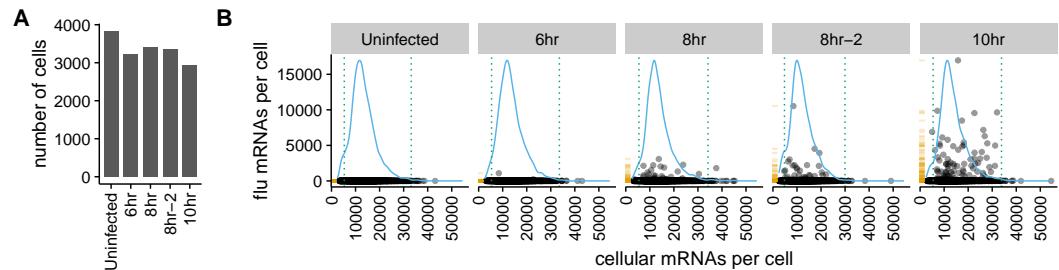
**Figure 2-Figure supplement 1.** Full flow cytometry data for panel B .

63 quantify the number of molecules of each mRNA species that have been captured for each cell.  
 64 Infected cells will express viral as well as cellular mRNAs – however the cell barcodes and UMs  
 65 cannot distinguish whether a cell was initially infected by one or multiple viral particles. We therefore  
 66 engineered an influenza virus (strain A/WSN/1933) that additionally carried *viral barcodes* consisting  
 67 of synonymous mutations near the 3' end of each transcript (Figure 1A). Critically, these synonymous  
 68 mutations did not greatly impact viral growth kinetics (Figure 1B). We infected A549 human lung  
 69 carcinoma cells with an equal mix of the wild-type and synonymously barcoded viruses. Cells  
 70 infected by a single virion will exclusively express mRNAs from either wild-type or synonymously  
 71 barcoded virus, whereas cells that are co-infected with multiple virions will often express mRNAs  
 72 from both the wild-type and synonymously barcoded viruses (Figure 1C).

73 We took care to generate stocks of virus that were relatively “pure” of defective particles.  
 74 Stocks of viruses typically contain an array of biologically active viral particles, some of which are  
 75 defective for replication owing to mutations or deletions in essential viral genes (*von Magnus, 1954;*  
*Huang et al., 1970; Brooke, 2014; Fonville et al., 2015; Lauring and Andino, 2010; Dimmock et al., 2014; Saira et al., 2013*). These defective particles become prevalent when a virus is grown at  
 76 high MOI, where complementation permits the growth of otherwise deleterious genotypes. To  
 77 minimize the levels of defective particles, we propagated our viral stocks at low MOI for a relatively  
 78 brief period of time (*Xue et al., 2016*). We validated that our stocks exhibited greater purity of  
 79 infectious particles than a stock propagated at high MOI by verifying that they had a higher ratio  
 80 of infectious particles to virion RNA (Figure 2A) and to particles capable of inducing expression of  
 81 a single viral protein (Figure 2B). In addition, viral stocks with many defective particles are more  
 82 immunostimulatory (*Tapia et al., 2013; Lopez, 2014*). We confirmed that our viral stocks induced  
 83 less interferon than a stock propagated at higher MOI (Figure 2C).

#### 86 Single cells show an extremely wide range of expression of viral mRNA.

87 We infected A549 cells at low MOI with a mixture of the wild-type and synonymously barcoded  
 88 viruses, and collected cells for sequencing at 6, 8, and 10 hours post-infection, including a replicate  
 89 a 8-hours. We replaced the infection inoculum with fresh media at one-hour post-infection, thereby  
 90 ensuring that most infection was initiated during a narrow time window. The exception was the  
 91 replicate 8-hour sample, for which we did *not* perform this media change and instead left the



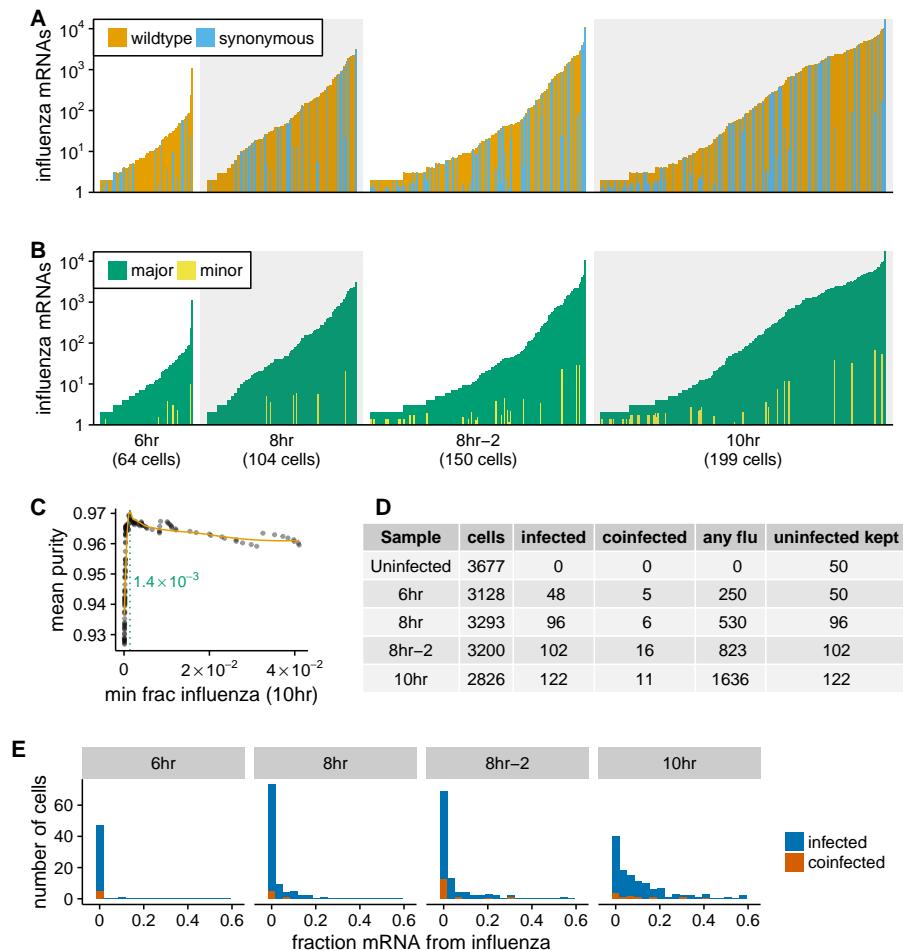
**Figure 3.** There is a very wide distribution in the amount of viral mRNA per cell. **(A)** Number of cells sequenced for each sample. **(B)** The number of cellular and viral mRNAs detected for each cell is plotted as a point. The blue lines show the overall distribution of the number of cellular mRNAs per cell. The orange rug plot at the left of each panel shows the distribution of the number of viral mRNAs per cell. Cells outside the dotted green lines were considered outliers with suspiciously low or high amounts of cellular mRNA (possibly derived from two cells per droplet), and were excluded from all subsequent analyses. Figure 3-Figure supplement 1 shows the exact distributions of the fraction of viral mRNA per cell.

**Figure 3-Figure supplement 1.** Cumulative fraction plot of proportion of total mRNA from virus.

92 cells in the original infection inoculum. We recovered between 3,000 and 4,000 cells for each  
 93 sample (Figure 3A). As expected for a low-MOI infection, most cells expressed little or no viral mRNA  
 94 (Figure 3B and Figure 3-Figure supplement 1). Also as expected, the amount of viral mRNA per cell  
 95 among infected cells increased over time (Figure 3B and Figure 3-Figure supplement 1). But what  
 96 was most notable was how widely the number of viral mRNA molecules varied among infected cells.  
 97 While the fraction of mRNA derived from virus was <0.1% for most cells, viral mRNA constituted  
 98 half the transcriptome in a few cells at 8 and 10 hours (Figure 3B).

99 A complicating factor is that uninfected cells could have small amounts of viral mRNA due  
 100 to leakage of transcripts from lysed cells. It is therefore important to establish a threshold for  
 101 identifying truly infected cells. We can do this by taking advantage of the fact that roughly half the  
 102 infecting virions bear synonymous barcodes. Reads derived from lysed cells will be drawn from  
 103 both wild-type and synonymous barcoded viral transcripts. However, most cells are infected by at  
 104 most one virion, and so the reads from truly infected cells will usually derive almost entirely from  
 105 one of the two viral variants. Figure 4A shows the fraction of viral reads in individual cells from each  
 106 viral variant, and Figure 4B indicates the fraction of viral reads from the most abundant variant in  
 107 that cell. Most cells with large amounts of viral mRNA have viral transcripts exclusively derived from  
 108 one viral variant – indicating non-random partitioning as expected from viral infection. However,  
 109 cells with a small amount of viral mRNA often have viral transcripts from both variants, as expected  
 110 from the random partitioning associated with simple mRNA leakage. Finally, a few cells with large  
 111 amounts of viral mRNA have viral transcripts from both variants, likely reflecting co-infection.

112 We determined the threshold amount of viral mRNA per cell for each sample at which the  
 113 barcode partitioning clearly resulted from infection rather than leakage (Figure 4C and Figure 4-  
 114 Figure supplement 2), and used these thresholds to annotate cells that we were confident were  
 115 truly infected. We also annotated as co-infected cells above this threshold that had mRNA from  
 116 both viral variants. Figure 4D shows the number of cells annotated as infected and co-infected for  
 117 each sample – these cells are just a small fraction of the number of cells with any viral read. These  
 118 annotation thresholds are conservative, and may miss some true low-level infections. However, it is  
 119 important that the analyses below are restricted to cells that are truly infected with virus, so we  
 120 accepted the possible loss of some low-level infections in order to avoid false positives. In addition,  
 121 the synonymous viral barcodes only identify co-infections by viruses with different barcodes – since  
 122 the barcodes are at roughly equal proportion, we expect to miss about half of the co-infections.  
 123 Since we annotate about ~10% of the infected cells as co-infected by viruses with different barcodes  
 124 (Figure 4D), we expect another ~10% of the infected cells to also be co-infected but not annotated  
 125 as so by our approach. Because most cells are not infected, we subsampled the uninfected cells to



**Figure 4.** Synonymous barcodes on the viral mRNAs distinguish true infections from cells that contain viral mRNAs derived from leakage of lysed cells. **(A)** Cells with at least two viral mRNAs for which the barcode could be called, arranged in order of increasing influenza transcript counts. Bar heights denote the number viral mRNAs on a  $\log_{10}$  scale, bar coloring is linearly proportional to the fractions of viral mRNAs derived from wild-type and synonymously barcoded virus. **(B)** Same as (A), but each bar is colored according to the relative fraction of the more common (major) and less common (minor) virus variant. At low levels of viral mRNA there is often a roughly equal mix, suggesting contamination with viral mRNAs leaked from lysed cells. At higher levels of viral mRNA, cells generally have only one viral variant, suggesting infection initiated by a single virion. A few cells are also obviously co-infected with both viral variants. **(C)** We determined a threshold for calling “true” infections by finding the amount of viral mRNA per cell at which the viral barcode purity no longer increases with more viral mRNA. The purity is the fraction of all viral mRNA in a cell derived from the most abundant viral barcode in that cell. We fit a curve (orange line) to the mean purity of all cells with more than the indicated amount of viral mRNA, and drew the cutoff (dotted green line) at the point where this curve stopped increasing with the fraction of total mRNA derived from virus. This plot illustrates the process for the 10-hour sample, see Figure 4-Figure supplement 2 for similar plots for other samples. See the Methods for details. **(D)** The number of cells identified as infected and co-infected for each sample, as well as the number of cells with any viral read. For all subsequent analyses, we subsampled the number of uninfected cells per sample to the greater of 50 or the number of infected cells. **(E)** Distribution of the fraction of mRNA per cell derived from virus for both infected and co-infected cells.

**Figure 4-Figure supplement 1.** Number of viral barcodes called.

**Figure 4-Figure supplement 2.** Thresholds for calling infected cells.

**Figure 4-Figure supplement 3.** Fraction of total viral mRNA derived from a given fraction of infected cells.

**Figure 4-Figure supplement 4.** Flow cytometry analysis of viral protein production of cells contingent on infectious dose or coinfection state.

126 the numbers shown in Figure 4D to balance the proportions of infected and uninfected cells for all  
 127 subsequent analyses.

128 Strikingly, the extreme variation in the number of viral transcripts per cell remains even after we  
 129 apply these rigorous criteria for annotating infected cells (Figure 4E). The fraction of viral mRNA per  
 130 infected cell follows a roughly exponential distribution, with many cells having few viral transcripts  
 131 and a few cells having many. At 6 and 8 hours <10% of infected cells are responsible for over half the viral transcripts  
 132 (Figure 4-Figure supplement 3). Notably, Figure 4E shows that there are co-infected cells with both  
 133 low and high amounts of viral mRNA, suggesting that the initial infectious dose does not drive a  
 134 simple continuous increase in viral transcript production. In support of this view, we used flow  
 135 cytometry to quantify the levels of individual viral proteins in cells infected at various MOIs or for  
 136 which we could delineate co-infection status (Figure 4-Figure supplement 4). This analysis shows  
 137 that sub-populations of cells that express similarly low and high levels of viral proteins persist  
 138 across a wide range of infectious doses, although co-infection can influence the relative proportion  
 139 of infected cells that fall into these sub-populations (Figure 4-Figure supplement 4).

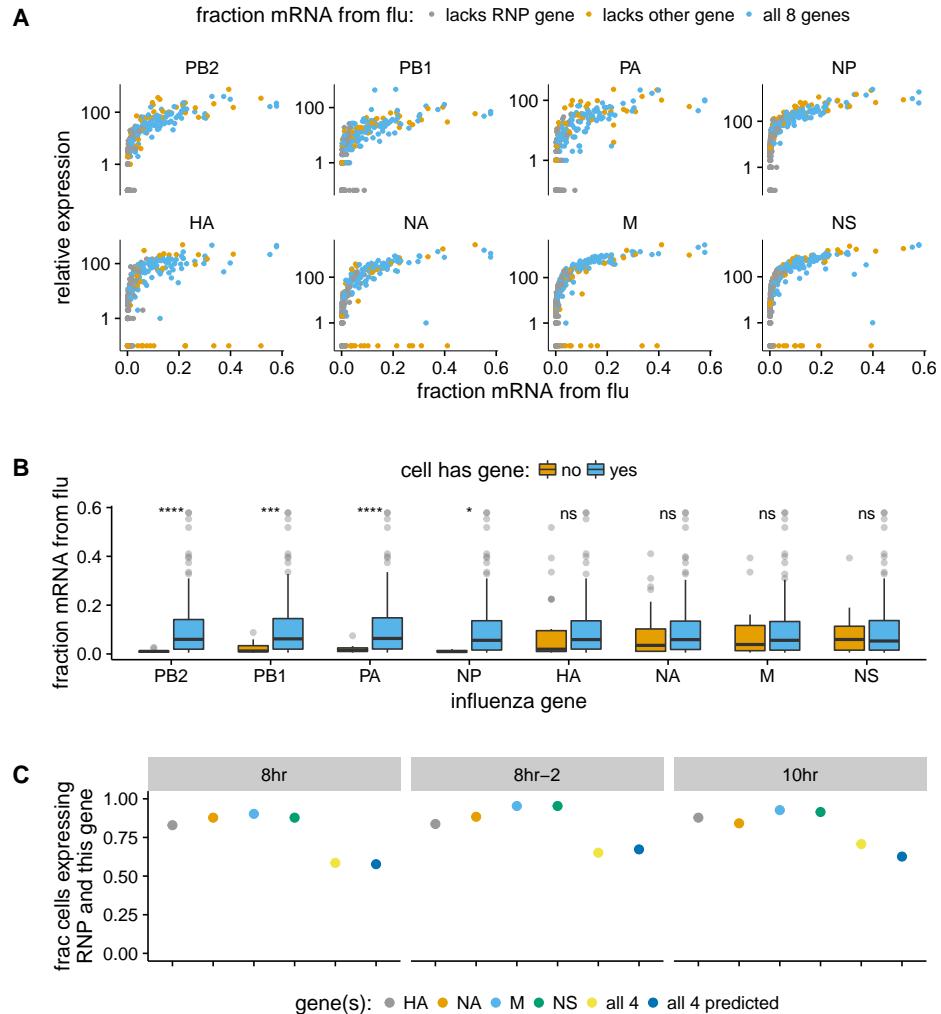
#### 141 **Absence of viral genes partially explains cell-to-cell variability in viral load.**

142 The influenza genome is segmented, and cells can fail to express a viral mRNA if the encoding  
 143 gene segment is not packaged in the infecting virion or fails to initiate transcription after infection.  
 144 Indeed, several groups have reported that the majority of infected cells fail to express at least  
 145 one viral gene (*Brooke et al., 2013; Heldt et al., 2015; Dou et al., 2017*). We wondered if the  
 146 absence of specific viral genes might be associated with reduced amounts of viral mRNA within  
 147 single infected cells. In particular, transcription of influenza virus mRNAs is performed by the viral  
 148 ribonucleoprotein (RNP) complex, which consists of the three proteins that encode the tripartite  
 149 polymerase (PB2, PB1, and PA) as well as nucleoprotein (NP) (*Huang et al., 1990*). Each viral gene  
 150 segment is associated with one RNP in incoming infecting virions, but secondary transcription by  
 151 newly synthesized RNPs requires the presence of the viral genes encoding each of the four RNP  
 152 proteins (*Vreede et al., 2004; Eisfeld et al., 2015*). This secondary transcription is a major source  
 153 of viral mRNAs, as evidenced by the fact that blocking synthesis of the RNP proteins reduces the  
 154 amount of viral mRNA by several orders of magnitude in bulk cells (Figure 5-Figure supplement 1).

155 We examined the total amount of viral mRNA versus the expression of the genes from each viral  
 156 segment (Figure 5A, Figure 5-Figure supplement 2). Note that influenza virus expresses ten major  
 157 gene transcripts from its eight gene segments, as the M and NS segments are alternatively spliced  
 158 to produce the M1 / M2 and NS1 / NEP transcript, respectively (*Dubois et al., 2014*). However, an  
 159 inherent limitation of current established single-cell mRNA sequencing techniques is that they only  
 160 sequence the 3' end of the transcript (*Zheng et al., 2017; Macosko et al., 2015; Klein et al., 2015*;  
 161 *Cao et al., 2017*). Since the alternative spliceoforms M1 / M2 and NS1 / NEP share the same 3' ends,  
 162 we cannot distinguish them and therefore will refer simply to the combined counts of transcripts  
 163 from each of these alternatively spliced segments as the M and NS genes.

164 Cells that lack an RNP gene never derive more than a few percent of their mRNAs from virus,  
 165 confirming the expected result that all four RNP genes are essential for high levels of viral transcrip-  
 166 tion (Figure 5A, Figure 5-Figure supplement 2). However, we observe cells that lack each of the other  
 167 non-RNP genes but still derive ≈40% of their mRNAs from virus, suggesting that none of the other  
 168 genes are important for high levels of viral transcription. These results are statistically supported by  
 169 Figure 5B, which shows that absence of any RNP gene but *not* any other viral gene is associated  
 170 with reduced amounts of viral mRNA. However, gene absence clearly does not explain all of the  
 171 variability in viral gene expression, since even cells expressing all viral genes exhibit a very wide  
 172 distribution in the amount of viral mRNA that they express. Specifically, at both 8 and 10 hours, the  
 173 amount of viral mRNA in individual cells expressing all eight viral genes still ranges from <1% to  
 174 >50% (Figure 5A, Figure 5-Figure supplement 2).

175 We also quantified the fraction of infected cells that completely failed to express a given gene. We



**Figure 5.** The absence of viral genes explains some of the variability in the amount of viral mRNA per cell. **(A)** The normalized expression of each viral gene as a function of the total fraction of mRNA in each infected cell derived from virus, taken over all time points. Cells with high viral burden always express all RNP genes, but some cells with high viral burden lack each of the other genes. **(B)** Box and whisker plots showing the per-cell viral burden among cells with >0.5% of their mRNA from virus, binned by whether or not the cells express each gene. A Wilcoxon signed-rank test was used to test the null hypothesis that absence of each gene does not affect viral burden: \*\*\*\* =  $P < 10^{-4}$ , \*\*\* =  $P < 10^{-3}$ , \* =  $P < 0.05$ , ns = not significant. **(C)** The fraction of cells that express each of the four other genes among cells that express all RNP genes, as well as the fraction that express *all* four of the other genes. The fraction that express all four genes is well predicted by simply multiplying the frequencies of cells that express each gene individually, indicating that gene absence is approximately independent across these genes.

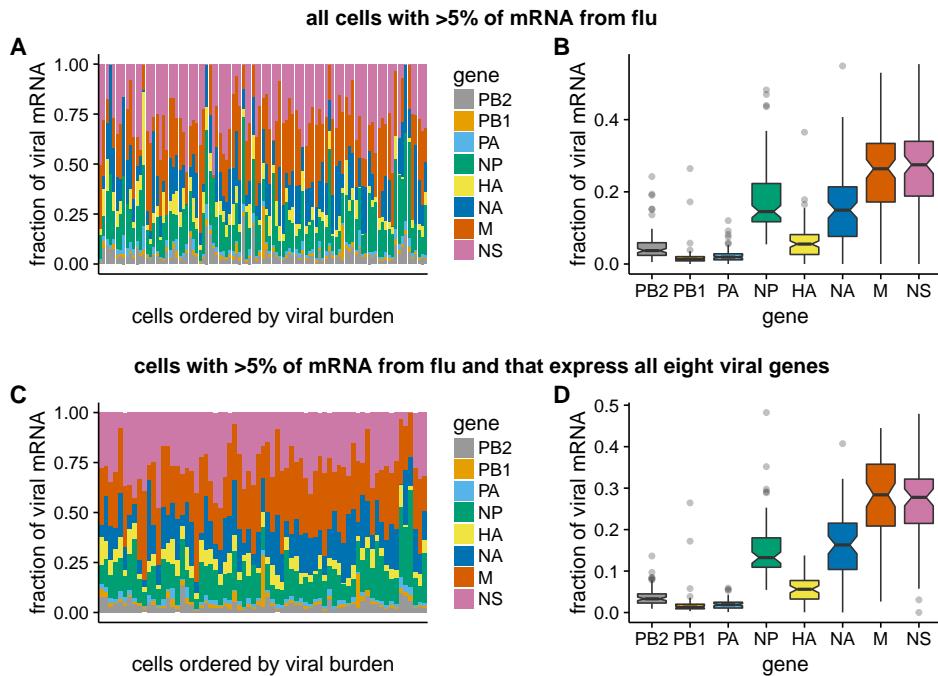
**Figure 5-Figure supplement 1.** Secondary transcription is a major source of viral mRNA during bulk infections.

**Figure 5-Figure supplement 2.** Like panel (A), but shows samples individually.

**Figure 5-Figure supplement 3.** Like panel (B) but for the 10-hr sample only.

**Figure 5-source data 1.** The numerical data for panel (C) are in p\_missing\_genes.csv.

176 limited this analysis to examining the presence / absence of the non-RNP genes in cells expressing  
 177 all four RNP genes, since we might fail to detect viral transcripts that are actually present at low  
 178 levels in RNP-deficient cells due to the lower viral burden in these cells. At the 8- and 10-hour time  
 179 points, between 5% and 17% of cells fail to express any one of the four non-RNP genes (Figure 5C  
 180 and Figure 5-source data 1). The absence of a given gene appears to be an independent event, as



**Figure 6.** [add expression wedges] Relative expression of influenza virus genes in highly infected cells (>5% of total mRNA from virus). **(A)** The fraction of viral mRNA from each viral gene for each cell. **(B)** Box plots showing the distribution of the fraction of viral mRNA per cell from each viral gene. The black lines at the notches are the medians, and the tops and bottoms of boxes indicate the first and third quartiles. Whiskers extend to the highest or lowest data point observed within 1.5x the interquartile range, outliers shown as circles. Notches extend 1.58x the interquartile range divided by the square root of the number of observations. **(C), (D)** The same plots, but only including cells for which we observed at least one molecule of each viral gene.

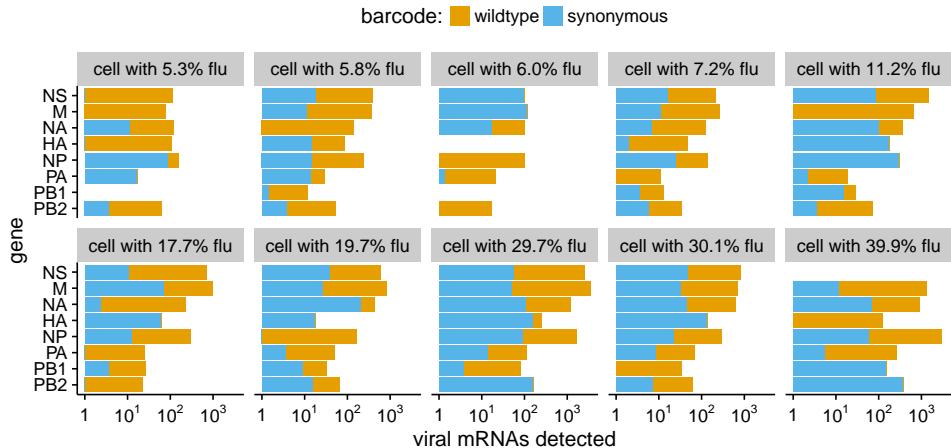
**Figure 6-source data 1.** The raw data for all cells are in `p_flu_expr_all.csv`.

**Figure 6-source data 2.** The raw data for fully infected cells are in `p_flu_expr_fullyinfected.csv`.

181 the probability of observing all four non-RNP genes in a cell is well predicted by simply multiplying  
 182 the probabilities of observing each gene individually (Figure 5C and Figure 5-source data 1). If we  
 183 extrapolate the frequencies at which cells lack non-RNP genes to the RNP genes, then we would  
 184 predict that 35-50% of infected cells express mRNAs from all eight genes. This estimate of the  
 185 frequency at which infected cells express mRNAs from all eight gene segments is slightly higher than  
 186 previous estimates of 13% (*Brooke et al., 2013*) and 20% (*Dou et al., 2017*). At least one difference  
 187 is that *Brooke et al. (2013)* stained for proteins whereas we examined the expression of mRNAs – it  
 188 is likely that some cells contain mutated viral genes that fail to produce stable protein even when  
 189 mRNA is expressed.

190 **The relative amounts of different viral mRNAs are more consistent across cells.**  
 191 The results above show that the amount of viral mRNA in infected cells varies over several orders  
 192 of magnitude. Does the relative expression of viral genes exhibit similar cell-to-cell variability?  
 193 To address this question, we focused on cells that derived >5% of their mRNA from virus, since  
 194 estimates of relative viral gene expression will be less noisy in cells with more viral mRNAs.

195 In contrast to the extreme variability in the total viral mRNA per cell, the fraction of this mRNA  
 196 derived from each gene is much more consistent across cells (Figure 6A). Total viral mRNA varies by  
 197 orders of magnitude, but the fraction from any given viral gene is fairly tightly clustered around the  
 198 median value for all cells (Figure 6B). The relative levels of each viral mRNA in our cells are similar  
 199 to prior bulk measurements made by Northern blots (*Hatada et al., 1989*), which also found an  
 200 expression hierarchy of M > NS > NP > NA > HA > PB2 ~ PB1 ~ PA. The cell-to-cell consistency in



**Figure 7.** The abundance of each viral transcript in cells that are co-infected with the two viral variants and have >5% of their mRNA derived from virus. The bars show the logarithms of the numbers of each viral mRNA detected, and are colored in linear proportion to the fraction of that mRNAs derived from wild-type or synonymously barcoded virus.

**Figure 7-Figure supplement 1.** Co-infected cells express roughly equal amounts of a gene from each infecting viral variant.

**Figure 7-source data 1.** The raw data plotted in this figure are in `p_co-infection.csv`.

**Figure 7-source data 2.** The sequence of the HA viral RNA carrying the GFP gene is in `HAflank-eGFP.fasta`.

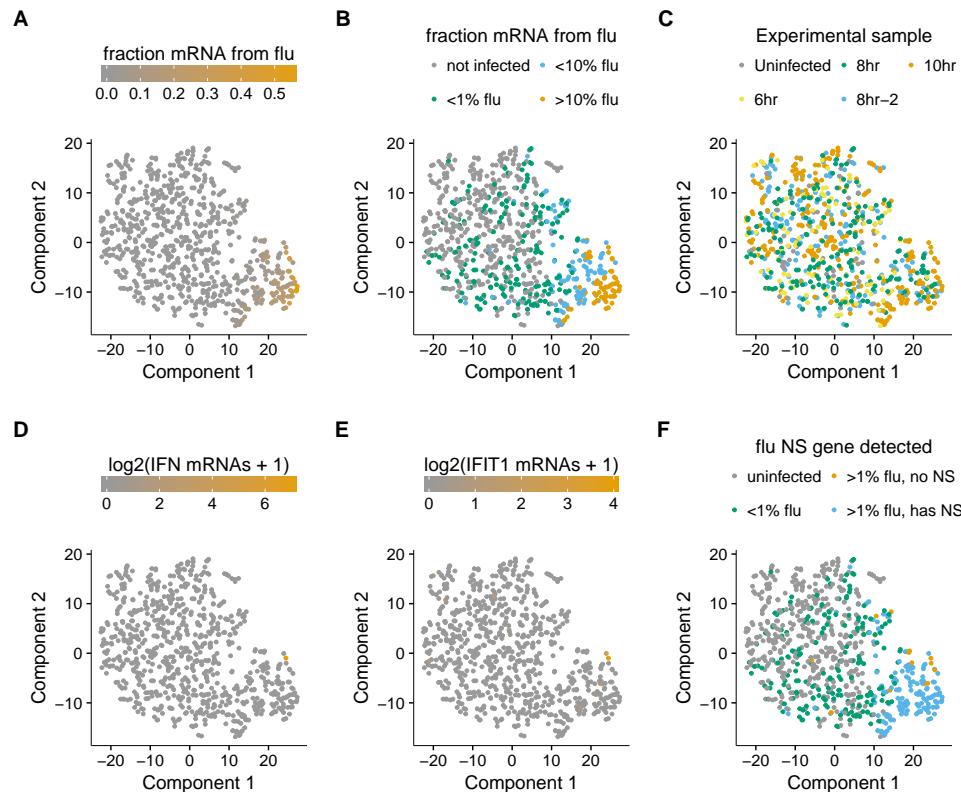
201 the relative expression of different viral genes is even tighter if we limit the analysis only to cells  
 202 that express all eight viral genes (Figure 6C,D). Therefore, with the exception of complete gene  
 203 absence, the factors that drive the dramatic cell-to-cell variability in the amount of viral mRNA have  
 204 roughly similar effects on all viral genes in a given cell. This finding is consistent with prior work  
 205 showing positive correlations among the abundance of several viral genome segments in individual  
 206 cells ([Heldt et al., 2015](#)).

#### 207 **Co-infection can provide infected cells with the full complement of viral genes.**

208 Our sequencing enables us to identify the rare cells that were co-infected with both wild-type and  
 209 synonymously barcoded viral variants. Overall, we captured 10 such co-infected cells that had >5%  
 210 of their mRNA derived from virus (Figure 7). Seven of these 10 cells expressed all eight viral genes.  
 211 Remarkably, the majority (4 of 7) of these cells would *not* have expressed all the viral genes in the  
 212 absence of co-infection, since they have at least one gene exclusively derived from each viral variant.  
 213 For instance, the cell with 11.2% of its mRNA from virus in the upper right of Figure 7 expresses  
 214 M only from the wildtype viral variant, and NP and HA only from the synonymously barcoded  
 215 variant. Our data therefore provide direct single-cell support for the idea that co-infection can  
 216 rescue missing viral genes ([Brooke et al., 2013, 2014; Fonville et al., 2015; Aguilera et al., 2017](#)).

217 Another observation from Figure 7 is that co-infected cells usually express roughly equal  
 218 amounts of transcripts from each of the two viral variants. This observation is consistent with the  
 219 finding by [Dou et al. \(2017\)](#) and [Huang et al. \(2008\)](#) that the temporal window for co-infection is  
 220 short – if both viral variants infect a cell at about the same time, then neither will have a headstart  
 221 and so each will have a roughly equal opportunity to transcribe its genes.

222 To support this idea with a larger dataset albeit at lower resolution, we generated a virus  
 223 in which the HA coding sequence was replaced by GFP. We then co-infected cells with a mix of  
 224 wildtype and ΔHA-GFP virus and used flow cytometry to score cells for the presence of HA only  
 225 (infection by wildtype virus), GFP only (infection by ΔHA-GFP virus), or both (co-infection) as shown  
 226 in Figure 7 Figure supplement 1. As in our single-cell sequencing data, we found that expression  
 227 of HA and GFP were highly correlated, indicating that co-infected cells typically expressed roughly



**Figure 8.** [add arrow] A t-SNE plot created by semi-supervised clustering using genes that co-vary with viral infection status. Each point is a single cell, and each panel shows an identical layout but colors the cells according to a different property. **(A), (B)** Cells colored by the fraction of their mRNA derived from virus. **(C)** Cells colored by the experimental sample. **(D), (E)** Cells colored by the number of detected transcripts from type I and III interferons (IFN). Only one cell has detectable interferon expression (in orange, indicated with arrow). **(E)** Cells colored by the expression of the interferon-stimulated gene IFIT1. **(F)** Cells colored by whether they express the viral NS gene. The one interferon-positive cell is lacking NS, but so are many interferon-negative cells.

228 equal amounts of transcript from each viral variant.

### 229 Activation of the interferon response is rare in single infected cells.

230 Because our sequencing captured all polyadenylated transcripts, we can examine whether there  
 231 are prominent changes in the host-cell transcriptome in sub-populations of infected cells. Influenza  
 232 virus infection can trigger innate-immune sensors that lead to the transcriptional induction of  
 233 type I and III interferons, and subsequently of anti-viral interferon-stimulated genes (*Killip et al.,*  
 234 *2015; Iwasaki and Pillai, 2014; Crotta et al., 2013*). However, activation of the interferon response  
 235 is stochastic and bi-modal at the level of single cells (*Chen et al., 2010; Shalek et al., 2013, 2014;*  
 236 *Perez-Cidoncha et al., 2014; Bhushal et al., 2017; Hagai et al., 2017*). We therefore hypothesized  
 237 that we might see two sub-populations of infected cells: one in which the interferon response  
 238 inhibited viral transcription, and another in which the virus was able to express high levels of its  
 239 mRNA by evading or blocking this response.

240 To examine whether there were distinct sub-populations of virus-infected cells, we used a  
 241 semi-supervised t-SNE approach (*Van der Maaten and Hinton, 2008*) to cluster cells by genes that  
 242 co-varied with viral infection status. As shown in Figure 8A,B, this approach effectively grouped cells  
 243 by the amount of viral mRNA that they expressed. Sample-to-sample variation was regressed away  
 244 during the clustering, as cells did not obviously group by time-point, with expected exception that

245 the uninfected and 6-hour samples had few cells in the region of the plot corresponding to large  
246 amounts of viral mRNA (Figure 8C).

247 But to our surprise, we did not see a prominent clustering of infected cells into sub-populations  
248 as expected if the interferon response was strongly activated in some cells. To investigate fur-  
249 ther, we annotated each cell by the total number of type I and III interferon transcripts detected.  
250 Remarkably, only a single cell expressed detectable interferon (Figure 8D). We also examined  
251 interferon-stimulated genes, which are induced by autocrine and paracrine interferon signaling.  
252 Figure 8E shows expression of one such gene, IFIT1 (*Fensterl and Sen, 2011*). As with interferon  
253 itself, expression of IFIT1 was rare and most prominent in the single interferon-positive cell, pre-  
254 sumably due to the higher efficiency of autocrine versus paracrine signaling. Notably, interferon  
255 and interferon-stimulated genes were also relatively ineffective at blocking viral transcription in the  
256 single cell in which they were potently induced, since >10% of the mRNA in this cell was derived  
257 from virus (Figure 8A,B,D,E).

258 We posited that the paucity of interferon induction might be due to the activity of influenza  
259 virus's major interferon antagonist, the NS1 protein (*García-Sastre et al., 1998; Hale et al., 2008*).  
260 We therefore identified cells that expressed substantial amounts of viral mRNA but lacked the  
261 NS gene (Figure 8F). Consistent with the idea that NS1 is important for suppressing interferon,  
262 the one interferon-positive cell lacked detectable expression of the NS gene. But other cells that  
263 lacked NS expression still failed to induce a detectable interferon response, despite often having a  
264 substantial amount of their mRNA derived from virus (Figure 8). This result is in line with other work  
265 showing that NS1-deficient influenza virus does not deterministically induce interferon (*Killip et al.,  
266 2017; Kallfass et al., 2013*). Therefore, many individual infected cells fail to activate innate-immune  
267 responses even when the virus lacks its major interferon antagonist.

### 268 **Some host genes co-vary with viral gene expression.**

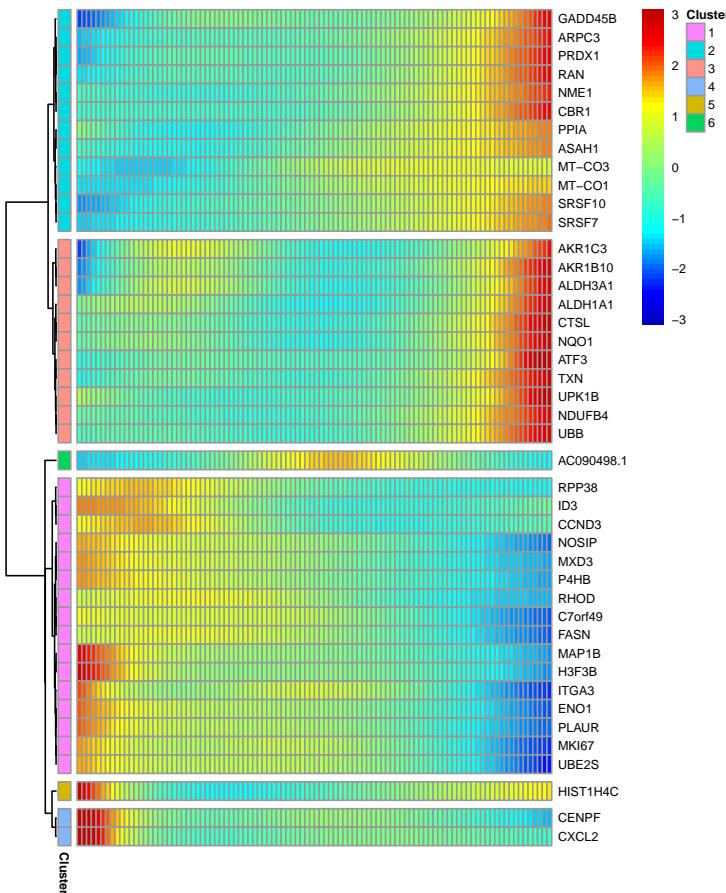
269 We examined whether any host genes were differentially expressed in cells with more viral mRNA.  
270 We restricted this analysis to infected cells with all eight viral genes in order to focus on cellular  
271 genes that were associated with viral mRNA burden independent of effects due to the presence or  
272 absence of particular viral transcripts. We identified 45 cellular genes that co-varied with viral gene  
273 expression at a false discovery rate of 0.1 (Figure 9).

274 Many of the genes with increased expression in cells with more viral mRNA are known or  
275 suspected to be regulated by the Nrf2 master regulator in response to oxidative stress. These genes  
276 produce proteins that are involved in detoxification of reactive oxygen species or resultant products,  
277 the management of misfolded proteins, the electron transport chain, or a general stress response  
278 (Figure 9-Figure supplement 1). We additionally see reduced expression of the nitric oxide synthase  
279 interacting protein (NOSIP). Transient oxidative stress is known to occur during viral infection, and  
280 may act in a proviral fashion via MAPK activation driving vRNP export (*Amatore et al., 2014*). The  
281 antioxidant response is thought to be largely antiviral, potentially through inhibition of MAPK activity  
282 (*Lin et al., 2016; Sgarbanti et al., 2014*). Our data do not reveal whether the expression of genes  
283 involved in the response to oxidative stress are a cause or a symptom of higher levels of viral mRNA,  
284 and further investigation of this topic is an interesting area for future work.

### 285 **Discussion**

286 We have quantified the total transcriptome composition of single cells infected with influenza virus.  
287 While we observe a general increase in the amount of viral mRNA over time as expected from  
288 bulk measurements (*Hatada et al., 1989; Shapiro et al., 1987*), there is wide variation in viral gene  
289 expression among individual infected cells.

290 The most obvious form of heterogeneity is the complete failure of some infected cells to express  
291 one or more viral genes, which we estimate occurs in about half the infected cells in our experiments.  
292 The absence of some viral genes in some infected cells has been noted previously (*Brooke et al.,  
293 2013; Heldt et al., 2015; Dou et al., 2017*), and our work provides a holistic view by quantifying the



**Figure 9.** [add gray arrow wedge] Cellular genes that co-vary in expression with the amount of viral mRNA in cells expressing all eight viral genes. The columns are cells, ordered from left to right by the fraction of mRNA derived from virus. Each row is a gene that is differentially expressed as a function of the fraction of mRNA derived from virus at a false discovery rate of 0.1. Genes for which the color goes from blue at left to red at right are expressed at higher levels in cells with more viral mRNA. The scale bar indicates the number of standard deviations above or below the mean expression, truncated at 3-fold on both sides.

**Figure 9-source data 1.** The full results of the differential expression test is in `p_sig_cellular_genes.csv`.

**Figure 9-source data 2.** The results of a gene-set analysis are in `p_pathway_enrichment.tsv`.

**Figure 9-Figure supplement 1.** Many genes that co-vary with viral load are involved in the oxidative stress response.

total viral transcriptional load as a function of the level of each mRNA. We find that cells lacking expression of any of the four genes that encode the viral RNP express much less total viral mRNA, consistent with prior bulk studies (*Vreede et al., 2004; Eisfeld et al., 2015*). Interestingly, the reason some cells fail to express some viral genes remains unclear. The prototypical influenza virion packages one copy of each of the eight gene segments (*Noda et al., 2006; Hutchinson et al., 2010*), but some virions surely package fewer (*Brooke et al., 2014*). However, it is also possible that much of the viral gene absence is due to stochastic loss of viral RNPs after infection but prior to the initiation of viral transcription in the nucleus.

The absence of viral genes only partially explains the cell-to-cell variation in amount of viral mRNA, which still varies from <1% to >50% among cells expressing all the viral genes. It is likely that other viral genetic factors explain some of this remaining heterogeneity. The 3'-end sequencing strategy used in our experiments detects the presence of a viral gene, but does not identify whether that gene contains a mutation that might hinder viral replication. However, viral mutations are also unlikely to explain all the observed heterogeneity, since current consensus estimates of

308 influenza virus's mutation rate suggest that the typical virion in a stock such as the one used in our  
 309 experiment should contain less than one mutation per genome (*Parvin et al., 1986; Suárez et al.,  
 310 1992; Suárez-López and Ortín, 1994; Nobusawa and Sato, 2006; Bloom, 2014; Pauly et al., 2017*).

311 The rest of the heterogeneity must be due to some combination of cellular factors and inherent  
 312 stochasticity. Some features of the cellular transcriptome co-vary with the amount of influenza  
 313 mRNA. In particular, the viral load in individual cells is associated with the expression of genes  
 314 involved in response to cellular stresses, including oxidative stress. It will be interesting to determine  
 315 if these cellular transcriptional signatures are simply a consequence of the stress imposed by viral  
 316 replication, or if their stronger activation in some cells is a causative factor that promotes viral  
 317 transcription. However, it also would not be surprising if a substantial amount of the cell-to-cell  
 318 heterogeneity cannot be ascribed to pre-existing features of either the viral genome or cellular state.  
 319 Apparently stochastic heterogeneity is a common feature of many processes at a single-cell level (*Cai  
 320 et al., 2006; Raj et al., 2006; Buganim et al., 2012; Shalek et al., 2013; Avraham et al., 2015*) –  
 321 especially when those processes are initiated by very small numbers of initial molecules (*Elowitz  
 322 et al., 2002*), as is the case for low-MOI viral infection.

323 Our data do suggest that the factors driving the heterogeneity in viral transcriptional load exert  
 324 relatively concordant effects on all viral genes in a given cell. Specifically, despite the extreme  
 325 heterogeneity in total viral mRNA per cell, the relative levels of the viral mRNAs are reasonably  
 326 consistent across cells, and generally reflective of classical bulk measurements (*Hatada et al., 1989*).  
 327 Therefore, despite the stochasticity inherent in initiating transcription and replication of each gene  
 328 from a single copy carried by the incoming virion, as long as a gene is not completely lost then the  
 329 virus possesses mechanisms to control its relative expression (*Shapiro et al., 1987; Hatada et al.,  
 330 1989; Perez et al., 2010; Heldt et al., 2012; Chua et al., 2013*).

331 One factor that surprisingly does *not* appreciably contribute to the heterogeneity in our ex-  
 332 periments is activation of innate-immune interferon pathways. Only one of the hundreds of  
 333 virus-infected cells expresses any detectable interferon, despite the fact that a number of cells fail  
 334 to express the influenza-virus interferon antagonist NS1. It is known that interferon activation is  
 335 stochastic at the level of single cells in response to both synthetic ligands (*Shalek et al., 2013, 2014;  
 336 Bhushal et al., 2017; Hagai et al., 2017*) and actual infection (*Rand et al., 2012; Perez-Cidoncha  
 337 et al., 2014; Avraham et al., 2015; Killip et al., 2017*). But interferon expression is a prominent  
 338 transcriptional signature of high-MOI influenza virus infection of bulk cells, including in the epithelial  
 339 cell line and at the time-points used in our experiments (*Geiss et al., 2002; Sutejo et al., 2012*). So  
 340 it is notable how rarely single cells express interferon. Interferon expression would surely be more  
 341 common at later times or with a viral stock passaged at higher MOI, since paracrine interferon  
 342 signaling (*Crotta et al., 2013*) and accumulation of defective viral particles enhance innate-immune  
 343 detection (*Tapia et al., 2013; Lopez, 2014*). However, the early events of physiological influenza  
 344 infection involve just a few virions (*Varble et al., 2014; McCrone et al., 2017*), and so it is interesting  
 345 to speculate whether rare events such as interferon activation during the first few cycles of viral  
 346 replication could contribute to heterogeneity in the eventual outcome of infection.

347 Overall, our work shows the power and importance of characterizing cellular infection at the  
 348 level of single cells (*Avraham et al., 2015*). The dynamics of viral infection in any given cell is shaped  
 349 by the genetic composition of the incoming virion, the host-cell state, the bi-modality of innate-  
 350 immune activation, and the inherent stochasticity of molecular processes initiated by a single copy  
 351 of each viral gene. We have shown how the confluence of these factors leads to extreme cell-to-cell  
 352 heterogeneity in the transcriptional outcome of influenza virus infection. Further deconstruction of  
 353 the contributions of each factor will enable a deeper understanding of how the bulk features of  
 354 infection emerge from the processes occurring within individual virus-infected cells.

355 **Methods and Materials**

356 **Cell lines and viruses**

357 The following cell lines were used in this study: the human lung epithelial carcinoma line A549  
 358 (ATCC CCL-185), the MDCK-SIAT1 variant of the Madin Darby canine kidney cell line overexpressing  
 359 human SIAT1 (Sigma-Aldrich 05071502), and the human embryonic kidney cell line 293T (ATCC  
 360 CRL-3216). All cells were maintained in D10 media (DMEM supplemented with 10% heat-inactivated  
 361 fetal bovine serum, 2 mM L-glutamine, 100 U of penicillin/ml, and 100 µg of streptomycin/ml) at 37  
 362 °C at 5 % CO<sub>2</sub>.

363 Wildtype A/WSN/1933 (H1N1) influenza virus was generated by reverse genetics using the  
 364 plasmids pHW181-PB2, pHW182-PB1, pHW183-PA, pHW184-HA, pHW185-NP, pHW186-NA, pHW187-  
 365 M, and pHW188-NS (**Hoffmann et al., 2000**). The sequences of the viral RNAs encoded in these  
 366 plasmids are in Figure 1-source data 1. Reverse-genetics plasmids encoding the synonymously  
 367 barcoded WSN virus were created by using site-directed mutagenesis to introduce two synonymous  
 368 mutations near the 3' end of the mRNA for each viral gene. The sequences of the synonymously  
 369 barcoded viral RNAs are in Figure 1-source data 1.

370 To generate viruses from these plasmids, we transfected an equimolar mix of all eight plasmids  
 371 into cocultures of 293T and MDCK-SIAT1 cells seeded at a ratio of 8:1. At 24 hours post-transfection,  
 372 we changed media from D10 to influenza growth media (Opti-MEM supplemented with 0.01% heat-  
 373 inactivated FBS, 0.3% BSA, 100 U of penicillin/ml, 100 µg of streptomycin/ml, and 100 µg of calcium  
 374 chloride/ml). At 48 hours post-transfection we harvested the virus-containing supernatant, pelleted  
 375 cellular material by centrifugation at 300 x g's for 4 minutes, and stored aliquots of the clarified  
 376 viral supernatant at -80 °C. We then titered thawed aliquots of viral by TCID50 on MDCK-SIAT1 cells,  
 377 computing titers via the formula of **Reed and Muench (1938)**. To generate our "high-purity" stocks of  
 378 viruses for the single-cell sequencing experiments, we then infected MDCK-SIAT1 cells at an MOI of  
 379 0.01, and let the virus replicate for 36 hours prior to harvesting aliquots that were again clarified by  
 380 low-speed centrifugation, aliquoted, stored at -80 °C, and titered by TCID50. The high-MOI passage  
 381 (high-defective particle) stock used in Figure 2 was generated by instead passaging in MDCK-SIAT1  
 382 cells twice at an MOI of 1 for 48 hours.

383 For the experiments in Figure 7-Figure supplement 1, we created a virus that carried an HA gene  
 384 segment in which GFP replaced most of the HA coding sequence, following a scheme first described  
 385 by **Marsh et al. (2007)**. Briefly, we created a plasmid encoding a viral RNA with GFP in place of the  
 386 HA coding sequence in the context of the pH21 (**Neumann et al., 1999**) reverse-genetics plasmid,  
 387 removing potential start codons upstream of the GFP (see Figure 7-source data 2 for the sequence  
 388 of the viral RNA). We then generated GFP-carrying virus by reverse-genetics in cells constitutively  
 389 expressing HA (**Doud and Bloom, 2016**). To obtain sufficient titers, this HA-eGFP virus was expanded  
 390 for 44 rather than 36 hours after initiating infection at an MOI of 0.01.

391 **qPCR**

392 For the qPCR in Figure 2 and Figure 5-Figure supplement 1, A549 cells were seeded at 3x10<sup>5</sup>  
 393 cells per well in a 6-well tissue culture plate in D10 the day prior to infection. On the day of  
 394 infection, a single well was trypsinized and the cells were counted in order to determine the  
 395 appropriate amount of virus to use to achieve the intended MOI. Immediately before infection,  
 396 D10 was replaced with influenza growth media. For cells incubated with cyclohexamide, the  
 397 compound was added to a final concentration of 50 µg/ml at the time of infection – previously  
 398 confirmed to be sufficient to halt viral protein production (**Killip et al., 2014**). RNA was purified  
 399 using the QIAGEN RNeasy plus mini kit following manufacturer's instructions. cDNA was syn-  
 400 thetized using an oligoDT primer and the SuperScript™ III first-strand synthesis supermix from  
 401 ThermoFisher using the manufacturer's protocol. Transcript abundance was measured using  
 402 SYBR™ green PCR master mix, using a combined anneal/extension step of 60 °C for one minute  
 403 with the following primers: HA: 5'-GGCCAACCACACATTCAAC-3', 5'-GCTCATCACTGCTAGACGGG-

404 3', *IFNB1*: 5'-AAACTCATGAGCAGTCTGCA-3', 5'-AGGAGATCTTCAGTTCGGAGG-3', L32: 5'-  
 405 AGCTCCCAAAAATAGACGCAC-3', 5'-TTCATAGCAGTAGGCACAAAGG-3'. Biological triplicates were per-  
 406 formed for all samples.

407 For the measurements of viral genomic HA content in Figure 2A, vRNA was harvested from 80  
 408 µl of viral supernatant by the addition of 600 µl of RLT plus before proceeding with the standard  
 409 QIAGEN RNeasy Plus Mini kit protocol. The cDNA was generated using SuperScript™ III first-strand  
 410 synthesis supermix using the manufacturer's protocol, and using the universal vRNA primers of  
 411 **Hoffmann et al. (2001)** with the modifications described in **Xue et al. (2017)**. The qPCR was then  
 412 performed as for mRNA measurements. A standard curve was generated from three independent  
 413 dilutions of the HA-encoding reverse genetics plasmid. All vRNA values represent three independent  
 414 RNA extractions with two replicate qPCR measurements.

#### 415 **Flow cytometry titering and analyses**

416 To determine viral titers in terms of HA-expressing units and for the flow cytometry in Figure 7,  
 417 Figure supplement 1, and Figure supplement 4 A549 cells were seeded in a 6-well plate and infected  
 418 as described above for the qPCR analyses. Cells were harvested by trypsinization, resuspended in  
 419 phosphate-buffered saline supplemented with 2% heat-inactivated FBS, and stained with 10 µg/ml  
 420 of H17-L19, a mouse monoclonal antibody confirmed to bind to WSN HA in a prior study (**Doud**  
 421 **et al., 2017**). After washing in PBS supplemented with 2% FBS, the cells were stained with a goat  
 422 anti-mouse IgG antibody conjugated to APC. Cells were then washed, fixed in 1% formaldehyde,  
 423 and washed further before a final resuspension and analysis. We then determined the fraction of  
 424 cells that were HA positive and calculated the HA-expressing units. For NS1 staining in Figure sup-  
 425 plement 4, cells stained for HA as described above were permeabilized using BD Cytofix/Cytoperm  
 426 following manufacturer's instructions, stained with anti-NS1 (GTX125990, Genetex) at 4.4 µg/ml,  
 427 washed, stained with a goat anti-rabbit IgG antibody conjugated to Alexa Fluor 405, washed, and  
 428 analyzed. For Figure 7, Figure supplement 1, and Figure supplement 4), after gating to exclude  
 429 multiplets in FlowJo, data were extracted using the R package flowCore (**Le Meur et al., 2007**) and  
 430 analyzed using a custom Python script. For Figure supplement 4) channels were additionally com-  
 431 pensated in FlowJo. Guassian kernel density estimates were obtained using the scipy stats package  
 432 method, guassian\_kde, using automatic bandwidth determination (**van der Walt et al., 2017**). For  
 433 Figure supplement 4), the percentage of influenza-infected cells was determined by HA staining  
 434 alone, and the top quantile of NS1-stained cells matching that percentage were taken as the NS1  
 435 positive population.

#### 436 **Infections for single-cell mRNA sequencing**

437 Single-cell sequencing libraries were generated using the 10x Chromium Single Cell 3' plat-  
 438 form (**Zheng et al., 2017**) using the V1 reagents.

439 All time points except for the 8-hour replicate were prepared on the same day. For the infections,  
 440 A549 cells were seeded in a 6-well plate, with two wells per time point. A single well of cells was  
 441 trypsinized and counted prior to initiation of the experiment for the purposes of calculating MOI.  
 442 Wild-type and synonymously barcoded virus were mixed to an estimated ratio of 1:1 based on prior,  
 443 exploratory, single-cell analyses (data not shown). At the initiation of our experiment, the wells  
 444 for all time points were changed from D10 to influenza growth media. Cells were then infected  
 445 with 0.3 HA-expressing units of virus per cell (a determined by flow cytometry). The infections  
 446 were performed in order of time point: first the 10-hour time point, then the 8-hour, and then the  
 447 6-hour time point. At one hour after infection, the media for each time point was changed to fresh  
 448 influenza growth media. Note that the HA-expressing units were calculated without this additional  
 449 washing step, and so likely represent an overestimate of our final infectious dose (consistent  
 450 with the fact that fewer than 30% of cells appear infected in the single-cell sequencing data). All  
 451 cells were then harvested for single-cell analysis concurrently – ensuring all had spent equivalent  
 452 time in changed media . For our replicate 8-hour time point, cells were infected as above except

453 that the cells were infected at 0.1 HA-expressing units of virus per cell but no wash step was  
 454 performed, and the sample was prepared on a different day. After harvest, cells were counted using  
 455 disposable hemocytometers and diluted to equivalent concentrations with an intended capture  
 456 of 3000 cells/sample following the manufacturer's provided by 10x Genomics for the Chromium  
 457 Single Cell platform. All subsequent steps through library preparation followed the manufacturer's  
 458 protocol. Samples were sequenced on an Illumina HiSeq. The deep sequencing data are available  
 459 on the Sequence Read Archive under accession [*[Upload in progress.](#)*]

#### 460 Computational analysis of single-cell mRNA sequencing data

461 Jupyter notebooks that perform all of the computational analyses are available in Supplemen-  
 462 tary file 1 and at [https://github.com/jbloomlab/flu\\_single\\_cell](https://github.com/jbloomlab/flu_single_cell) [*This GitHub repository will be made*  
*463 public upon acceptance of the manuscript. If you are a reviewer who needs access, please contact the*  
*464 editor.*].

465 Briefly, the raw deep sequencing data were processed using the 10X Genomics software package  
 466 CellRanger (version 2.0.0). The reads were aligned to a concatenation of the human and influenza  
 467 virus transcriptomes. The human transcriptome was generated by filtering genome assembly  
 468 GRCh38 for protein coding genes defined in the GTF file GRCh38.87. The influenza virus transcrip-  
 469 tome was generated from the reverse-complement of the wildtype WSN viral RNA sequences as  
 470 encoded in the reverse-genetics plasmids (Figure 1-source data 1). CellRanger calls cells based on  
 471 the number of observed cell barcodes, and creates a cell-gene matrix. We used custom Python  
 472 code to annotate the cells in this matrix by the number of viral reads that could be assigned to  
 473 the wildtype and synonymously barcoded virus. Only about half of the viral reads overlapped  
 474 the barcoded regions of the genes (Figure 1A) and could therefore be assigned to a viral barcode  
 475 (Figure 4-Figures supplement 1). So for calculations of the number of reads in a cell derived from  
 476 each viral barcode for each viral gene, the total number of detected molecules of that gene are  
 477 multiplied by the fraction of those molecules with assignable barcodes that are assigned to that  
 478 barcode. This annotated cell-gene matrix is in Supplementary file 2. A Jupyter notebook that  
 479 performs these analyses is in Supplementary file 1.

480 The annotated cell-gene matrix was analyzed in R, primarily using the Monocle package (version  
 481 2.4.0) (*Qiu et al., 2017; Trapnell et al., 2014*). A Jupyter notebook that performs these analyses is in  
 482 Supplementary file 1. For each sample, cell barcodes that had >2.5-fold fewer or more UMI counts  
 483 mapping to cellular transcripts than the sample mean were excluded from downstream analyses  
 484 (see red vertical lines in Figure 3B).

485 In order to determine an appropriate cutoff for how many reads a cell needed to contain in  
 486 order to be classified as infected, we calculated the mean viral barcode purity across all cells that  
 487 contained at least a given fraction of viral mRNA and had multiple viral reads that could be assigned  
 488 a barcode (Figure 4B,C and Figure 4-Figure supplement 2). We then determined the threshold  
 489 fraction of viral mRNA at which the mean purity no longer increased as a function of the amount  
 490 of viral mRNA. This threshold represents the point at which we have effectively eliminated cells  
 491 that have low barcode purity simply due to lysis-acquired reads sampled randomly from both viral  
 492 barcodes. As is apparent from Figure 4B, only the 10-hour sample and the 8-hour-2 sample have  
 493 the excess of mixed barcodes among cells with low amounts of viral mRNA. The likely reason is that  
 494 these samples have more total viral mRNA (and so there is more available mRNA to be acquired  
 495 from lysed cells); in addition, there is always some experimental variability in the amount of cell  
 496 lysis during the 10X sequencing process, and these samples may simply have the most. So the  
 497 above threshold procedure is appropriate for those two samples. For the other samples, we simply  
 498 set a minimum threshold of requiring at least a fraction  $2 \times 10^{-4}$  reads to come from viral mRNA as  
 499 explained in the legend to Figure 4-Figure supplement 2. The thresholds for each sample are shown  
 500 in Figure 4C and Figure 4-Figure supplement 2. This procedure is expected to be conservative, and  
 501 may miss some truly infected cells with very low amounts of viral mRNA. For subsequent analyses,  
 502 we retained all infected cells and a subsample of uninfected cells (the greater of 50 or the number

503 of infected cells for that sample). The rationale for subsampling the uninfected cell is that the  
 504 vast majority of cells are uninfected, and we did not want these cells to completely dominate the  
 505 downstream analyses. Cells were classified as co-infected if both viral variants had an RNA level  
 506 that exceeded the threshold, and if the minor variant contributed at least 5% of the viral mRNA.

507 For the semi-supervised t-SNE clustering, we used Monocle's cell hierarchy function to bin cells  
 508 into those with no viral mRNA, <2% viral mRNA, between 2% and 20% viral mRNA, and >20%.  
 509 Candidate marker genes for t-SNE dimensionality reduction were then determined using the  
 510 Monocle function markerDiffTable, excluding the effects of sample variation and the number of  
 511 genes identified in a given cell, using a q-value cutoff of 0.01. The specificity of these markers was  
 512 determined using the function calculateMarkerSpecificity – the top 50 markers were retained, and  
 513 used to place populations in a two-dimensional plane based on tSNE dimensionality reduction.

514 For the analyses of cellular genes that differed in expression as a function of the amount of viral  
 515 mRNA, we only considered cells that expressed all 8 viral mRNAs to avoid effects driven simply by  
 516 viral gene absence. We also only considered cellular genes in the differential gene analysis, since  
 517 viral gene expression will tautologically co-vary with the amount of viral mRNA. Additionally, because  
 518 influenza virus has the capacity to degrade or prevent the synthesis of host mRNAs (*Bercovich-*  
*519 Kinori et al., 2016*) and contributes significantly to the total number UMIs in some cells, we calculate  
 520 size factors (a scalar value representing efficiency of UMI capture) based on cellular transcripts alone.  
 521 Finally, we assigned all cells a ceiling fraction of mRNA from virus of 25% so that a few extremely  
 522 high-expressing cells did not dominate. Cellular genes with expression that co-varied with the  
 523 fraction of viral mRNAs in a cell were then determined using the Monocle differentialGeneTest, after  
 524 removing variance explained by sample to sample variation. Figure 9 shows all genes that were  
 525 significantly associated with the fraction of mRNA from virus at a false discovery rate of 0.1.

## 526 Acknowledgments

527 We thank Xiaojie Qiu for advice about use of the Monocle software package, David Bacsik and  
 528 Robert Bradley for comments on the manuscript, and the Fred Hutch Genomics Core for performing  
 529 the Illumina deep sequencing.

## 530 References

- 531 **Aguilera ER, Erickson AK, Jesudhasan PR, Robinson CM, Pfeiffer JK.** Plaques formed by mutagenized viral  
 532 populations have elevated coinfection frequencies. *mBio*. 2017; 8(2):e02020.
- 533 **Akpınar F, Timm A, Yin J.** High-Throughput Single-Cell Kinetics of Virus Infections in the Presence of Defective  
 534 Interfering Particles. *Journal of Virology*. 2016 Jan; 90(3):1599–1612.
- 535 **Amatore D, Sgarbanti R, Aquilano K, Baldelli S, Limongi D, Civitelli L, Nencioni L, Garaci E, Ciriolo MR, Palamara  
 536 AT.** Influenza virus replication in lung epithelial cells depends on redox-sensitive pathways activated by  
 537 NOX4-derived ROS. *Cellular microbiology*. 2014 Oct; 17(1):131–145.
- 538 **Avraham R, Haseley N, Brown D, Penaranda C, Jijon HB, Trombetta JJ, Satija R, Shalek AK, Xavier RJ, Regev A, et al.**  
 539 Pathogen cell-to-cell variability drives heterogeneity in host immune responses. *Cell*. 2015; 162(6):1309–1321.
- 540 **Banning A, Deubel S, Kluth D, Zhou Z, Brigelius-Flohe R.** The Gl-GPx Gene Is a Target for Nrf2. *Molecular and  
 541 Cellular Biology*. 2005 May; 25(12):4914–4923.
- 542 **Bercovich-Kinori A, Tai J, Gelbart IA, Shitrit A, Ben-Moshe S, Drori Y, Itzkovitz S, Mandelboim M, Stern-Ginossar  
 543 N.** A systematic view on influenza induced host shutoff. *eLife*. 2016 Aug; 5:e18311.
- 544 **Bhushal S, Wolfsmüller M, Selvakumar TA, Kemper L, Wirth D, Hornef MW, Hauser H, Köster M.** Cell polarization  
 545 and epigenetic status shape the heterogeneous response to Type III interferons in intestinal epithelial cells.  
 546 *Frontiers in immunology*. 2017; 8:671.
- 547 **Bloom JD.** An experimentally determined evolutionary model dramatically improves phylogenetic fit. *Molecular  
 548 Biology and Evolution*. 2014; 31(8):1956–1978.

- 549   **Brooke CB**, Ince WL, Wrammert J, Wrammert J, Ahmed R, Wilson PC, Bennink JR, Yewdell JW. Most Influenza A  
550    Virions Fail To Express at Least One Essential Viral Protein. *Journal of Virology*. 2013 Feb; 87(6):3155–3162.
- 551   **Brooke CB**. Biological activities of 'noninfectious' influenza A virus particles. *Future Virology*. 2014 Jan; 9(1):41–51.
- 552   **Brooke CB**, Ince WL, Wei J, Bennink JR, Yewdell JW. Influenza A virus nucleoprotein selectively decreases  
553    neuraminidase gene-segment packaging while enhancing viral fitness and transmissibility. *Proceedings of  
554    the National Academy of Sciences*. 2014; 111(47):16854–16859.
- 555   **Buganim Y**, Faddah DA, Cheng AW, Itsikovich E, Markoulaki S, Ganz K, Klemm SL, van Oudenaarden A, Jaenisch R.  
556    Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierachic  
557    phase. *Cell*. 2012; 150(6):1209–1222.
- 558   **Cai L**, Friedman N, Xie XS. Stochastic protein expression in individual cells at the single molecule level. *Nature*.  
559    2006; 440(7082):358–362.
- 560   **Cao J**, Packer JS, Ramani V, Cusanovich DA, Huynh C, Daza R, Qiu X, Lee C, Furlan SN, Steemers FJ, et al.  
561    Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science*. 2017; 357(6352):661–  
562    667.
- 563   **Chen S**, Chen S, Short JAL, Young DF, Killip MJ, Schneider M, Schneider M, Goodbourn S, Randall RE. Heterocellular  
564    induction of interferon by negative-sense RNA viruses. *Virology*. 2010 Nov; 407(2):247–255.
- 565   **Chua MA**, Schmid S, Perez JT, Langlois RA, et al. Influenza A virus utilizes suboptimal splicing to coordinate the  
566    timing of infection. *Cell reports*. 2013; 3(1):23–29.
- 567   **Combe M**, Garijo R, Geller R, Cuevas JM, Sanjuán R. Single-cell analysis of RNA virus infection identifies multiple  
568    genetically diverse viral genomes within single infectious units. *Cell host & microbe*. 2015; 18(4):424–432.
- 569   **Crotta S**, Davidson S, Mahlakov T, Desmet CJ, Buckwalter MR, Albert ML, Staeheli P, Wack A. Type I and type III  
570    interferons drive redundant amplification loops to induce a transcriptional signature in influenza-infected  
571    airway epithelia. *PLoS Pathogens*. 2013; 9(11):e1003773.
- 572   **Delbruck M**. The burst size distribution in the growth of bacterial viruses (bacteriophages). *Journal of bacteriology*.  
573    1945 Aug; 50(2):131–135.
- 574   **Dimmock NJ**, Dimmock NJ, Easton AJ, Easton AJ. Defective Interfering Influenza Virus RNAs: Time To Reevaluate  
575    Their Clinical Potential as Broad-Spectrum Antivirals? *Journal of Virology*. 2014 Apr; 88(10):5217–5227.
- 576   **Dou D**, Hernández-Neuta I, Wang H, Östbye H, Qian X, Thiele S, Resa-Infante P, Kouassi NM, Sender V, Henrich K,  
577    Mellroth P, Henriques-Normark B, Gabriel G, Nilsson M, Daniels R. Analysis of IAV Replication and Co-infection  
578    Dynamics by a Versatile RNA Viral Genome Labeling Method. *Cell reports*. 2017 Jul; 20(1):251–263.
- 579   **Doud M**, Bloom J. Accurate Measurement of the Effects of All Amino-Acid Mutations on Influenza Hemagglutinin.  
580    *Viruses*. 2016 Jun; 8(6):155–17.
- 581   **Doud MB**, Hensley SE, Bloom JD. Complete mapping of viral escape from neutralizing antibodies. *PLoS  
582    Pathogens*. 2017 Mar; 13(3):e1006271.
- 583   **Doyle V**, Virji S, Crompton M. Evidence that cyclophilin-A protects cells against oxidative stress. *The Biochemical  
584    journal*. 1999 Jul; 341 ( Pt 1)(Pt 1):127–132.
- 585   **Dubois J**, Terrier O, Rosa-Calatrava M. Influenza viruses and mRNA splicing: doing more with less. *MBio*. 2014;  
586    5(3):e00070–14.
- 587   **Duong HQ**, You K, Oh S, Kwak SJ, Seong YS. Silencing of NRF2 Reduces the Expression of ALDH1A1 and ALDH3A1  
588    and Sensitizes to 5-FU in Pancreatic Cancer Cells. *Antioxidants*. 2017 Sep; 6(3):52–10.
- 589   **Eisfeld AJ**, Neumann G, Kawaoka Y. At the centre: influenza A virus ribonucleoproteins. *Nature reviews  
590    Microbiology*. 2015; 13(1):28.
- 591   **Elowitz MB**, Levine AJ, Siggia ED, Swain PS. Stochastic gene expression in a single cell. *Science*. 2002;  
592    297(5584):1183–1186.
- 593   **Fensterl V**, Sen GC. The ISG56/IFIT1 gene family. *Journal of Interferon & Cytokine Research*. 2011 Jan; 31(1):71–  
594    78.

- 595 **Fonville JM**, Marshall N, Tao H, Steel J, Lowen AC. Influenza Virus Reassortment Is Enhanced by Semi-  
596 infectious Particles but Can Be Suppressed by Defective Interfering Particles. *PLoS Pathogens*. 2015 Oct;  
597 11(10):e1005204.
- 598 **García-Sastre A**, Egorov A, Matassov D, Brandt S, Levy DE, Durbin JE, Palese P, Muster T. Influenza A virus lacking  
599 the NS1 gene replicates in interferon-deficient systems. *Virology*. 1998; 252(2):324–330.
- 600 **Geiss GK**, Salvatore M, Tumpey TM, Carter VS, Wang X, Basler CF, Taubenberger JK, Bumgarner RE, Palese  
601 P, Katze MG, García-Sastre A. Cellular transcriptional profiling in influenza A virus-infected lung epithelial  
602 cells: the role of the nonstructural NS1 protein in the evasion of the host innate defense and its potential  
603 contribution to pandemic influenza. *Proceedings of the National Academy of Sciences of the United States of  
604 America*. 2002 Aug; 99(16):10736–10741.
- 605 **Gorrini C**, Harris IS, Mak TW. Modulation of oxidative stress as an anticancer strategy. *Nature Publishing Group*.  
606 2013 Dec; 12(12):931–947.
- 607 **Hagai T**, Chen X, Miragaia RJ, Gomes T, Rostom R, Kunowska N, Proserpio V, Donati G, Bossini-Castillo L, Naamati  
608 G, Emerton G, Trynka G, Kondova I, Denis M, Teichmann SA. Gene expression variability across cells and  
609 species shapes innate immunity. *bioRxiv*. 2017; doi: 10.1101/137992.
- 610 **Hale BG**, Randall RE, Ortín J, Jackson D. The multifunctional NS1 protein of influenza A viruses. *Journal of  
611 General Virology*. 2008; 89(10):2359–2376.
- 612 **Hatada E**, Hasegawa M, Mukaigawa J, Shimizu K, Fukuda R. Control of influenza virus gene expression: quanti-  
613 tative analysis of each viral RNA species in infected cells. *The Journal of Biochemistry*. 1989; 105(4):537–546.
- 614 **Heldt FS**, Frensing T, Reichl U. Modeling the intracellular dynamics of influenza virus replication to understand  
615 the control of viral RNA synthesis. *Journal of Virology*. 2012; 86(15):7806–7817.
- 616 **Heldt FS**, Kupke SY, Dorl S, Reichl U, Frensing T. Single-cell analysis and stochastic modelling unveil large  
617 cell-to-cell variability in influenza A virus infection. *Nature communications*. 2015 Nov; 6:1–12.
- 618 **Hoffmann E**, Stech J, Guan Y, Webster RG, Perez DR. Universal primer set for the full-length amplification of all  
619 influenza A viruses. *Archives of virology*. 2001 Dec; 146(12):2275–2289.
- 620 **Hoffmann E**, Neumann G, Kawaoka Y, Hobom G, Webster RG. A DNA transfection system for generation of  
621 influenza A virus from eight plasmids. *Proceedings of the National Academy of Sciences*. 2000; 97(11):6108–  
622 6113.
- 623 **Huang AS**, Baltimore D, et al. Defective viral particles and viral disease processes. *Nature*, London. 1970;  
624 226:325–7.
- 625 **Huang IC**, Li W, Sui J, Marasco W, Choe H, Farzan M. Influenza A Virus Neuraminidase Limits Viral Superinfection.  
626 *Journal of Virology*. 2008 Apr; 82(10):4834–4843.
- 627 **Huang T**, Palese P, Krystal M. Determination of influenza virus proteins required for genome replication. *Journal  
628 of Virology*. 1990; 64(11):5669–5673.
- 629 **Hutchinson EC**, von Kirchbach JC, Gog JR, Digard P. Genome packaging in influenza A virus. *Journal of general  
630 virology*. 2010; 91(2):313–328.
- 631 **Iwasaki A**, Pillai PS. Innate immunity to influenza virus infection. *Nature Reviews Immunology*. 2014 May;  
632 14(5):315–328.
- 633 **Jiang X**, An Z, Lu C, Chen Y, Du E, Qi S, Yang K, Zhang Z, Xu Y. The protective role of Nrf2-Gadd45b against  
634 antimony-induced oxidative stress and apoptosis in HEK293 cells. *Toxicology Letters*. 2016 Aug; 256:11–18.
- 635 **Jung KA**, Lee S, Kwak MK. NFE2L2/NRF2 Activity Is Linked to Mitochondria and AMP-Activated Protein Kinase Sig-  
636 naling in Cancers Through miR-181c/Mitochondria-Encoded Cytochrome C Oxidase Regulation. *Antioxidants  
637 & Redox Signaling*. 2017 Apr; p. ars.2016.6797–3.
- 638 **Kallfass C**, Lienenklaus S, Weiss S, Staeheli P. Visualizing the beta interferon response in mice during infec-  
639 tion with influenza A viruses expressing or lacking nonstructural protein 1. *Journal of Virology*. 2013 Jun;  
640 87(12):6925–6930.
- 641 **Killip MJ**, Jackson D, Perez-Cidoncha M, Fodor E, Randall RE. Single-cell studies of IFN- $\beta$  promoter activation by  
642 wild-type and NS1-defective influenza A viruses. *Journal of General Virology*. 2017 Mar; 98(3):357–363.

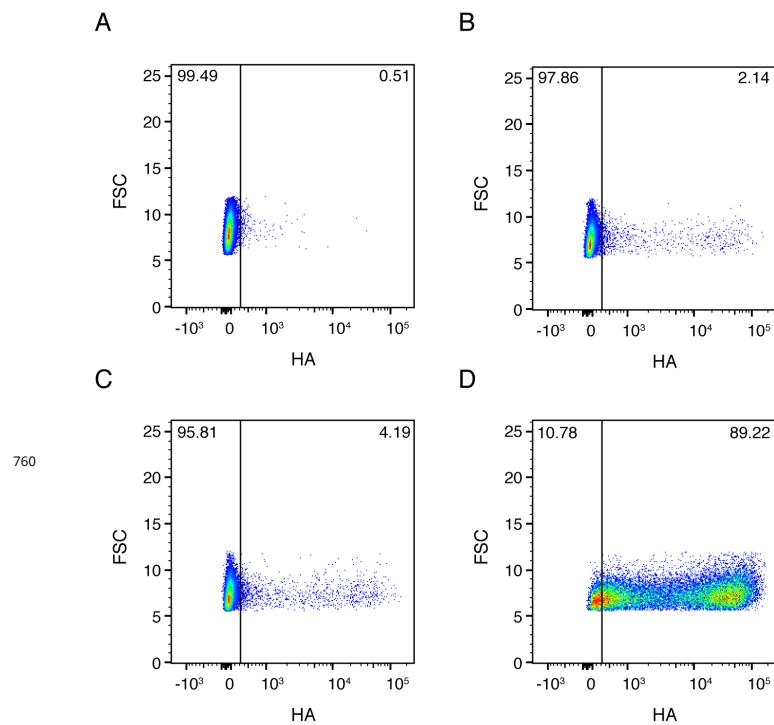
- 643   **Killip MJ**, Smith M, Jackson D, Randall RE. Activation of the Interferon Induction Cascade by Influenza A Viruses  
644   Requires Viral RNA Synthesis and Nuclear Export. *Journal of Virology*. 2014 Mar; 88(8):3942–3952.
- 645   **Killip MJ**, Fodor E, Randall RE. Influenza virus activation of the interferon system. *Virus research*. 2015 Feb; .
- 646   **Kim KH**, Jeong JY, Surh YJ, Kim KW. Expression of stress-response ATF3 is mediated by Nrf2 in astrocytes. *Nucleic  
647   Acids Research*. 2009 Oct; 38(1):48–59.
- 648   **Klein AM**, Mazutis L, Akartuna I, Tallapragada N, Veres A, Li V, Peshkin L, Weitz DA, Kirschner MW. Droplet  
649   Barcodeing for Single-Cell Transcriptomics Applied to Embryonic Stem Cells. *Cell*. 2015 May; 161(5):1187–1201.
- 650   **Lauring AS**, Andino R. Quasispecies Theory and the Behavior of RNA Viruses. *PLoS Pathogens*. 2010 Jul;  
651   6(7):e1001005–8.
- 652   **Le Meur N**, Hahne F, Ellis B, Haaland P. FlowCore: data structures package for flow cytometry data. *Bioconductor  
653   Project*; 2007.
- 654   **Lee D**, Ryu KY. Effect of cellular ubiquitin levels on the regulation of oxidative stress response and proteasome  
655   function via Nrf1. *Biochemical and Biophysical Research Communications*. 2017 Apr; 485(2):234–240.
- 656   **Leonard AS**, Weissman D, Greenbaum B, Ghedin E, Koelle K. Transmission Bottleneck Size Estimation from  
657   Pathogen Deep-Sequencing Data, with an Application to Human Influenza A Virus. *Journal of Virology*. 2017;  
658   p. JVI-00171–17.
- 659   **Lin X**, Wang R, Zou W, Sun X, Liu X, Zhao L, Wang S, Jin M. The Influenza Virus H5N1 Infection Can Induce ROS  
660   Production for Viral Replication and Host Cell Death in A549 Cells Modulated by Human Cu/Zn Superoxide  
661   Dismutase (SOD1) Overexpression. *Viruses*. 2016 Jan; 8(1):13–16.
- 662   **Lopez CB**. Defective Viral Genomes: Critical Danger Signals of Viral Infections. *Journal of Virology*. 2014 Jul;  
663   88(16):8720–8723.
- 664   **Van der Maaten L**, Hinton G. Visualizing data using t-SNE. *Journal of Machine Learning Research*. 2008;  
665   9:2579–2605.
- 666   **MacLeod AK**, Acosta-Jimenez L, Coates PJ, McMahon M, Carey FA, Honda T, Henderson CJ, Wolf CR. Aldo-keto  
667   reductases are biomarkers of NRF2 activity and are co-ordinately overexpressed in non-small cell lung cancer.  
668   *British journal of cancer*. 2016 Dec; 115(12):1530–1539.
- 669   **Macosko EZ**, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, Tirosh I, Bialas AR, Kamitaki N, Martersteck EM,  
670   et al. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell*. 2015;  
671   161(5):1202–1214.
- 672   **von Magnus P**. Incomplete forms of influenza virus. *Advances in Virus Research*. 1954; 2:59–79.
- 673   **Marsh GA**, Hatami R, Palese P. Specific residues of the influenza A virus hemagglutinin viral RNA are important  
674   for efficient packaging into budding virions. *Journal of Virology*. 2007; 81(18):9727–9736.
- 675   **McCrone JT**, Woods RJ, Martin ET, Malosh RE, Monto AS, Lauring AS. Stochastic processes dominate the within  
676   and between host evolution of influenza virus. *bioRxiv*. 2017; p. 176362.
- 677   **Miura T**, Taketomi A, Nishinaka T, Terada T. Regulation of human carbonyl reductase 1 (CBR1, SDR21C1) gene  
678   by transcription factor Nrf2. *Chemico-Biological Interactions*. 2013 Feb; 202(1-3):126–135.
- 679   **Murray J**, Taylor SW, Zhang B, Ghosh SS, Capaldi RA. Oxidative Damage to Mitochondrial Complex I Due to  
680   Peroxynitrite. *Journal of Biological Chemistry*. 2003 Sep; 278(39):37223–37230.
- 681   **Neumann G**, Watanabe T, Ito H, Watanabe S, Goto H, Gao P, Hughes M, Perez DR, Donis R, Hoffmann E, Hobom  
682   G, Kawaoka Y. Generation of influenza A viruses entirely from cloned cDNAs. *Proceedings of the National  
683   Academy of Sciences of the United States of America*. 1999 Aug; 96(16):9345–9350.
- 684   **Nobusawa E**, Sato K. Comparison of the mutation rates of human influenza A and B viruses. *Journal of Virology*.  
685   2006; 80(7):3675–3678.
- 686   **Noda T**, Sagara H, Yen A, Takada A, Kida H, Cheng RH, Kawaoka Y. Architecture of ribonucleoprotein complexes  
687   in influenza A virus particles. *Nature*. 2006; 439(7075):490.
- 688   **Parvin JD**, Moscona A, Pan W, Leider J, Palese P. Measurement of the mutation rates of animal viruses: influenza  
689   A virus and poliovirus type 1. *Journal of Virology*. 1986; 59(2):377–383.

- 690 **Pauly MD**, Procario MC, Lauring AS. A novel twelve class fluctuation test reveals higher than expected mutation  
691 rates for influenza A viruses. *eLife*. 2017; 6:e26437.
- 692 **Perez JT**, Varble A, Sachidanandam R, Zlatev I, Manoharan M, García-Sastre A, et al. Influenza A virus-generated  
693 small RNAs regulate the switch from transcription to replication. *Proceedings of the National Academy of  
694 Sciences*. 2010; 107(25):11525–11530.
- 695 **Perez-Cidoncha M**, Killip MJ, Oliveros JC, Asensio VJ, Fernandez Y, Bengoechea JA, Randall RE, Ortín J. An  
696 Unbiased Genetic Screen Reveals the Polygenic Nature of the Influenza Virus Anti-Interferon Response.  
697 *Journal of Virology*. 2014 Apr; 88(9):4632–4646.
- 698 **Peuchant E**, Bats ML, Moranvillier I, Lepoivre M, Guittot J, Wendum D, Lacombe ML, Moreau-Gaudry F, Boissan  
699 M, Dabernat S. Metastasis suppressor NM23 limits oxidative stress in mammals by preventing activation of  
700 stress-activated protein kinases/JNKs through its nucleoside diphosphate kinase activity. *The FASEB Journal*.  
701 2017 Mar; 31(4):1531–1546.
- 702 **Poon LL**, Song T, Rosenfeld R, Lin X, Rogers MB, Zhou B, Sebra R, Halpin RA, Guan Y, Twaddle A, DePasse JV,  
703 Stockwell TB, Wentworth DE, Holmes EC, Greenbaum B, Peiris JSM, Cowling BJ, Ghedin E. Quantifying influenza  
704 virus diversity and transmission in humans. *Nature Genetics*. 2016; 48(2):195–200.
- 705 **Qiu X**, Mao Q, Tang Y, Wang L, Chawla R, Pliner HA, Trapnell C. Reversed graph embedding resolves complex  
706 single-cell trajectories. *Nature Methods*. 2017 Aug; p. 1–10.
- 707 **Raj A**, Peskin CS, Tranchina D, Vargas DY, Tyagi S. Stochastic mRNA synthesis in mammalian cells. *PLoS Biology*.  
708 2006; 4(10):e309.
- 709 **Rand U**, Rinas M, Schwerk J, Nöhren G, Linné M, Kröger A, Flossdorf M, Kullai KK, Hauser H, Höfer T, Köster M.  
710 Multi-layered stochasticity and paracrine signal propagation shape the type-I interferon response. *Molecular  
711 Systems Biology*. 2012 Jan; 8(1):584.
- 712 **Reed LJ**, Muench H. A simple method of estimating fifty percent endpoints. *American journal of epidemiology*.  
713 1938; 27(3):493–497.
- 714 **Saira K**, Lin X, DePasse JV, Halpin R, Twaddle A, Stockwell T, Angus B, Cozzi-Lepri A, Delfino M, Dugan V, Dwyer  
715 DE, Freiberg M, Horban A, Losso M, Lynfield R, Wentworth DN, Holmes EC, Davey R, Wentworth DE, Ghedin E.  
716 Sequence analysis of in vivo defective interfering-like RNA of influenza A H1N1 pandemic virus. *Journal of  
717 Virology*. 2013; 87(14):8064–8074.
- 718 **Schulte MB**, Andino R. Single-Cell Analysis Uncovers Extensive Biological Noise in Poliovirus Replication. *Journal  
719 of Virology*. 2014 May; 88(11):6205–6212.
- 720 **Sgarbanti R**, Amatore D, Celestino I, Marcocci ME, Fraternale A, Ciriolo MR, Magnani M, Saladino R, Garaci E,  
721 Palamara AT, Nencioni L. Intracellular redox state as target for anti-influenza therapy: are antioxidants always  
722 effective? *Current topics in medicinal chemistry*. 2014; 14(22):2529–2541.
- 723 **Shalek AK**, Satija R, Adiconis X, Gertner RS, Gaublomme JT, Raychowdhury R, Schwartz S, Yosef N, Malboeuf C,  
724 Lu D, et al. Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature*.  
725 2013; 498(7453):236–241.
- 726 **Shalek AK**, Satija R, Shuga J, Trombetta JJ, Gennert D, Lu D, Chen P, Gertner RS, Gaublomme JT, Yosef N, et al.  
727 Single cell RNA Seq reveals dynamic paracrine control of cellular variation. *Nature*. 2014; 510(7505):363.
- 728 **Shapiro GI**, Gurney T, Krug RM. Influenza virus gene expression: control mechanisms at early and late times of  
729 infection and nuclear-cytoplasmic transport of virus-specific RNAs. *Journal of Virology*. 1987 Mar; 61(3):764–  
730 773.
- 731 **Suárez P**, Valcarcel J, Ortín J. Heterogeneity of the mutation rates of influenza A viruses: isolation of mutator  
732 mutants. *Journal of Virology*. 1992; 66(4):2491–2494.
- 733 **Suárez-López P**, Ortín J. An estimation of the nucleotide substitution rate at defined positions in the influenza  
734 virus haemagglutinin gene. *Journal of general virology*. 1994; 75(2):389–393.
- 735 **Sutejo R**, Yeo DS, Zu Myaing M, Hui C, Xia J, Ko D, Cheung PC, Tan BH, Sugrue RJ. Activation of Type I and III  
736 Interferon Signalling Pathways Occurs in Lung Epithelial Cells Infected with Low Pathogenic Avian Influenza  
737 Viruses. *PLoS ONE*. 2012; 7(3).

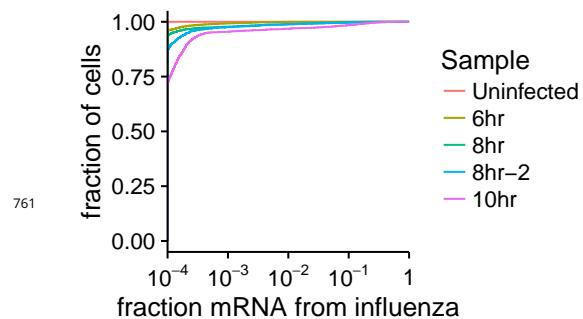
- 738    **Tapia K**, Kim Wk, Sun Y, Mercado-López X, Dunay E, Wise M, Adu M, López CB. Defective viral genomes arising  
739    in vivo provide critical danger signals for the triggering of lung antiviral immunity. *PLoS Pathogens*. 2013;  
740    9(10):e1003703.
- 741    **Trapnell C**, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, Lennon NJ, Livak KJ, Mikkelsen TS, Rinn JL. The  
742    dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nature  
743    Biotechnology*. 2014 Mar; 32(4):381–386.
- 744    **Varble A**, Albrecht RA, Backes S, Crumiller M, Bouvier NM, Sachs D, García-Sastre A, et al. Influenza A virus  
745    transmission bottlenecks are defined by infection route and recipient host. *Cell host & microbe*. 2014;  
746    16(5):691–700.
- 747    **Vreede FT**, Jung TE, Brownlee GG. Model Suggesting that Replication of Influenza Virus Is Regulated by  
748    Stabilization of Replicative Intermediates. *Journal of Virology*. 2004 Aug; 78(17):9568–9572.
- 749    **van der Walt S**, Colbert SC, Varoquaux G. The NumPy Array: A Structure for Efficient Numerical Computation.  
750    *Computing in Science & Engineering*. 2017 Aug; 13(2):22–30.
- 751    **Xue J**, Chambers BS, Hensley SE, López CB. Propagation and Characterization of Influenza Virus Stocks That  
752    Lack High Levels of Defective Viral Genomes and Hemagglutinin Mutations. *Frontiers in microbiology*. 2016  
753    Mar; 7(Pt 10):e1004924–15.
- 754    **Xue KS**, Stevens-Ayers T, Campbell AP, Englund JA, Pergam SA, Boeckh M, Bloom JD. Parallel evolution of  
755    influenza across multiple spatiotemporal scales. *eLife*. 2017 Jun; 6:e26875.
- 756    **Zheng GX**, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, Ziraldo SB, Wheeler TD, McDermott GP, Zhu J, et al.  
757    Massively parallel digital transcriptional profiling of single cells. *Nature communications*. 2017; 8:14049.
- 758    **Zhu Y**, Yongky A, Yin J. Growth of an RNA virus in single cells reveals a broad fitness distribution. *Virology*. 2009;  
759    385(1):39–46.

**Supplementary file 1.** Computer code for the analyses. This ZIP file contains a Jupyter notebook that runs CellRanger to align and annotate the reads, and a Jupyter notebook that uses Monocle to analyze the cell-gene matrix. The ZIP file also includes associated custom scripts.

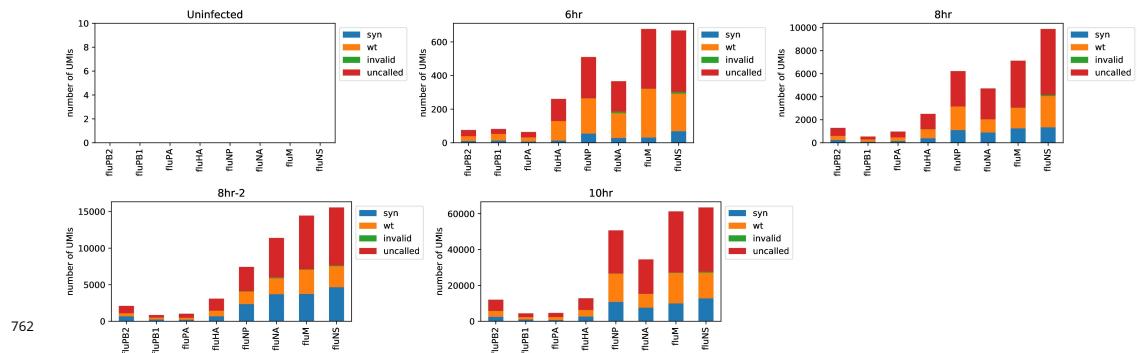
**Supplementary file 2.** The annotated cell-gene matrix in Matrix Market Format. *[This file is too large for the eLife submission system. We will communicate with the editors to get it uploaded for a final accepted version, or we will post it on DataDryad. If you are a reviewer and need access, please contact the editor.]*



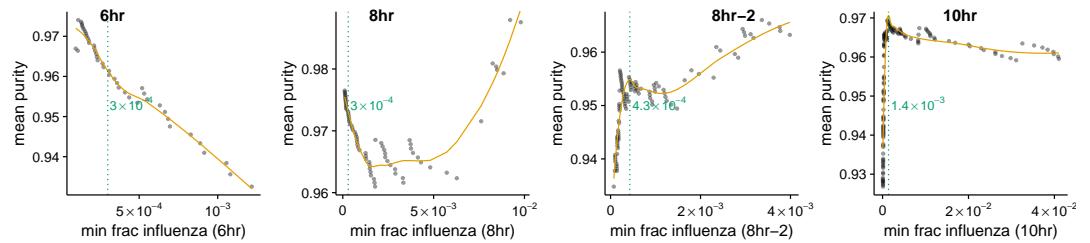
**Figure 2–Figure supplement 1.** Full flow cytometry data for Figure 2B. A549 cells were infected at an MOI of 0.1 as calculated by TCID<sub>50</sub> on MDCK-SIAT1 cells. **(A)** Uninfected gating control. **(B)** Cells infected with the wild-type virus stock used in our experiments. **(C)** Cells infected with synonymously barcoded virus stock used in our experiments. **(D)** Cells infected with a stock of wild-type virus propagated at a high MOI, and therefore enriched in defective particles.



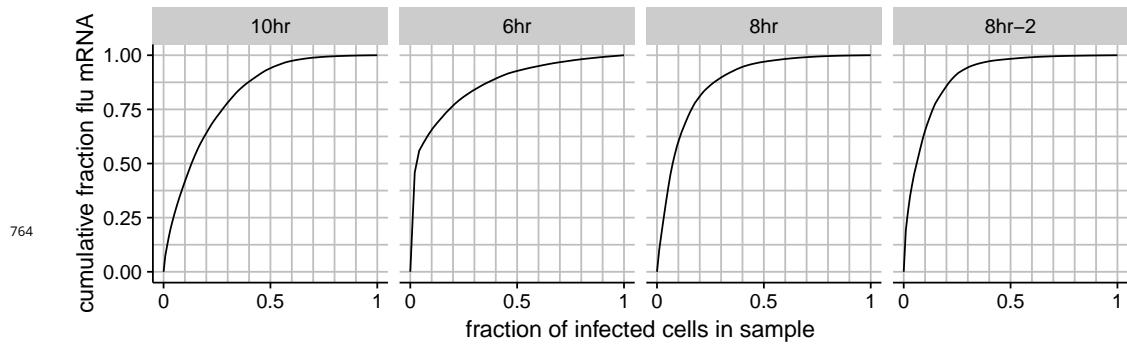
**Figure 3–Figure supplement 1.** For each sample, this plot shows the fraction of all cells that derive at least the indication fraction of their mRNA from influenza virus.



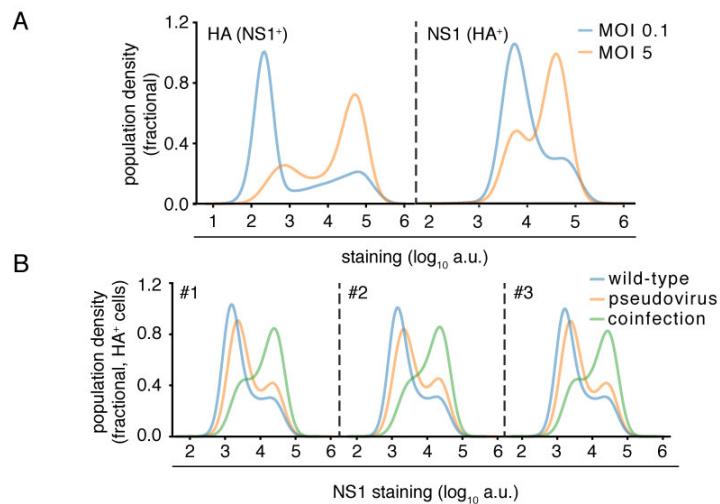
**Figure 4-Figure supplement 1.** The number of viral barcodes called for each sample and gene segment. Viral transcripts are classified as *syn* if they mapped to a synonymously barcoded influenza transcript, *wt* if they mapped to a wild-type influenza transcript, *invalid* if multiple reads for the same UMI differed on the status of the viral barcode, and as *uncalled* if none of the reads for that UMI overlapped the region of the viral transcript containing the viral barcode. For calculations of the number of reads in a cell derived from each viral barcode for each viral gene, the total number of detected molecules of that gene are multiplied by the fraction of those molecules with assignable barcodes that are assigned to that barcode.



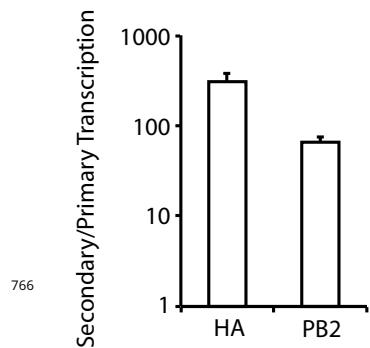
**Figure 4-Figure supplement 2.** Cell lysis can lead cells to the spurious association of small amounts of extraneous mRNA with individual cells. We wanted to avoid classifying as infected cells that had simply acquired such lysis-derived mRNA. The amount of lysis-derived mRNA will vary among samples as a function of both the lysis rate during the cell preparation (which always varies slightly from sample to sample in the 10X procedure) and with the amount of total viral mRNA for that sample (the more viral mRNA, the more there is to be acquired from lysed cells). As is shown in Figure 4B, the 8hr-2 and 10hr sample clearly have an enrichment of mixed barcodes in cells with small numbers of viral mRNA. For each sample, we calculated the mean purity of all cells with at least the indicated amount of viral mRNA, and determined the threshold amount of viral mRNA where purity no longer increased by finding the first maxima in a loess curve fit (orange line). We called the threshold at this point of maximum purity (dotted green line). For the 6hr and 8hr samples there is no indication of contamination from lysis-derived reads, as Figure 4B shows no increase in mixed barcodes in low viral mRNA cells. Therefore, for these samples we simply set a threshold of requiring at least  $2 \times 10^{-4}$  of the total mRNA to come from virus, which corresponds to ~2 viral mRNAs for the typical cell with  $10^4$  total reads (Figure 3B).



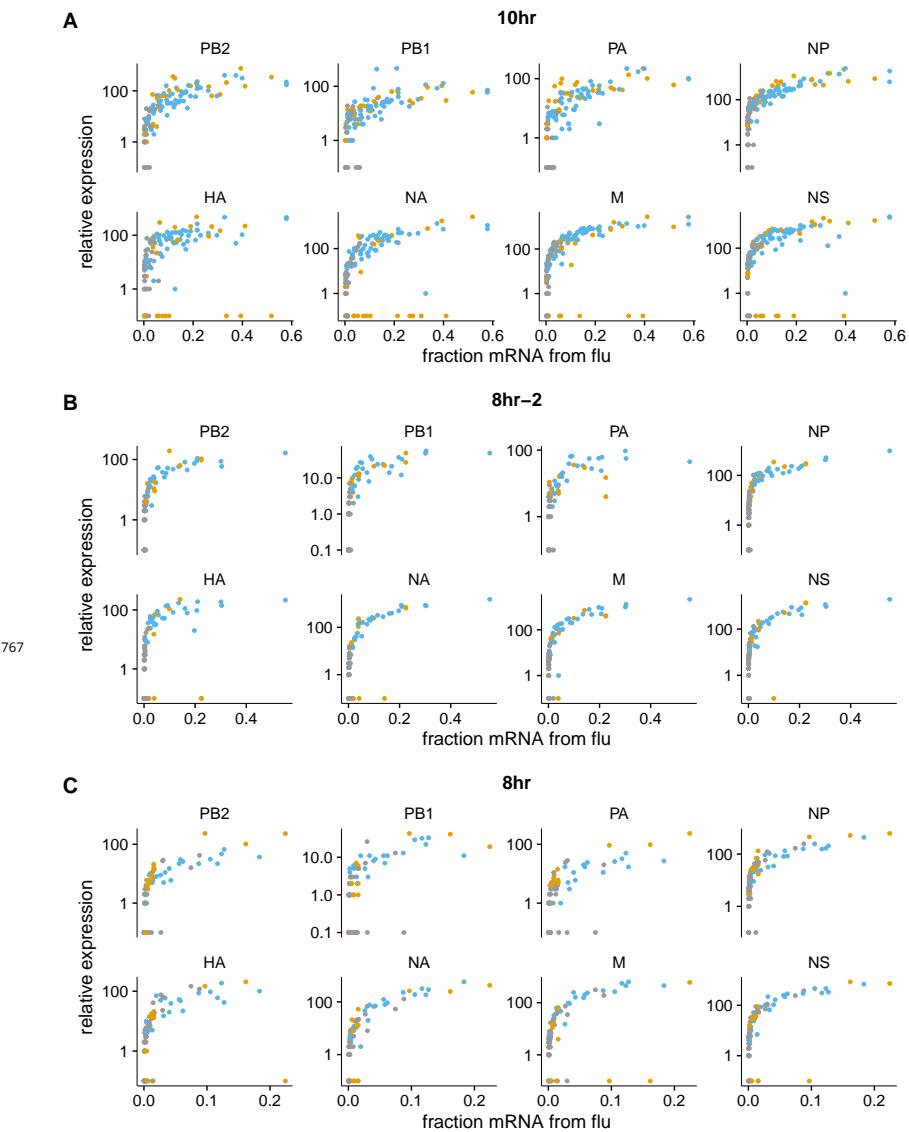
**Figure 4-Figure supplement 3.** The total fraction of all viral mRNA among infected cells that is attributable to a given fraction of these cells. For instance, the plot for the 8-hour sample shows that roughly 50% of all viral mRNA is derived from 10% of the infected cells.



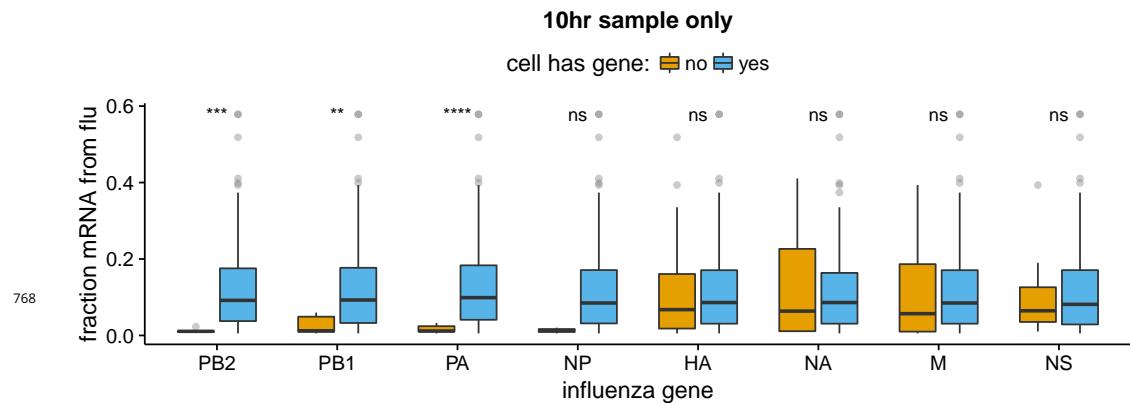
**Figure 4-Figure supplement 4. (A)** Expression of viral proteins in cells infected at high (MOI 5) as calculated by TCID50 on MDCK-SIAT1 cells) or low (MOI 0.1) initial infectious dose. Cells were concurrently stained for HA and NS1 proteins 10 hours post infection. HA staining was analyzed in cells positive for NS1 (left), and NS1 staining was analyzed in cells positive for HA (right). While a higher dose leads to more cells expressing high amounts of viral protein, it does not greatly increase the amount of viral protein in either the low-expressing or high-expressing cells. Therefore, higher viral dose does not lead to a large continuous increase in viral protein production among all cells – rather, it mostly changes the proportions of cells that fall in different parts of the highly heterogeneous distribution. **(B)** Cells were co-infected with a mix of wild-type virus and virus in which the HA gene was replaced by GFP flanked by the terminal regions of the HA gene segment at an MOI of 0.1 for each virus. At 10 hours post-infection, cells were stained for NS1 and HA expression and analyzed by flow cytometry. Cells could be annotated as infected by one or more infectious particles (wild-type or pseudovirus alone) or at least two infectious particles (coinfection). Coinfected cells, like cells infected at a higher infectious dose, occupy different positions in the distribution of viral protein production but do not exhibit a continuous increase in viral protein production. Replicates shown.



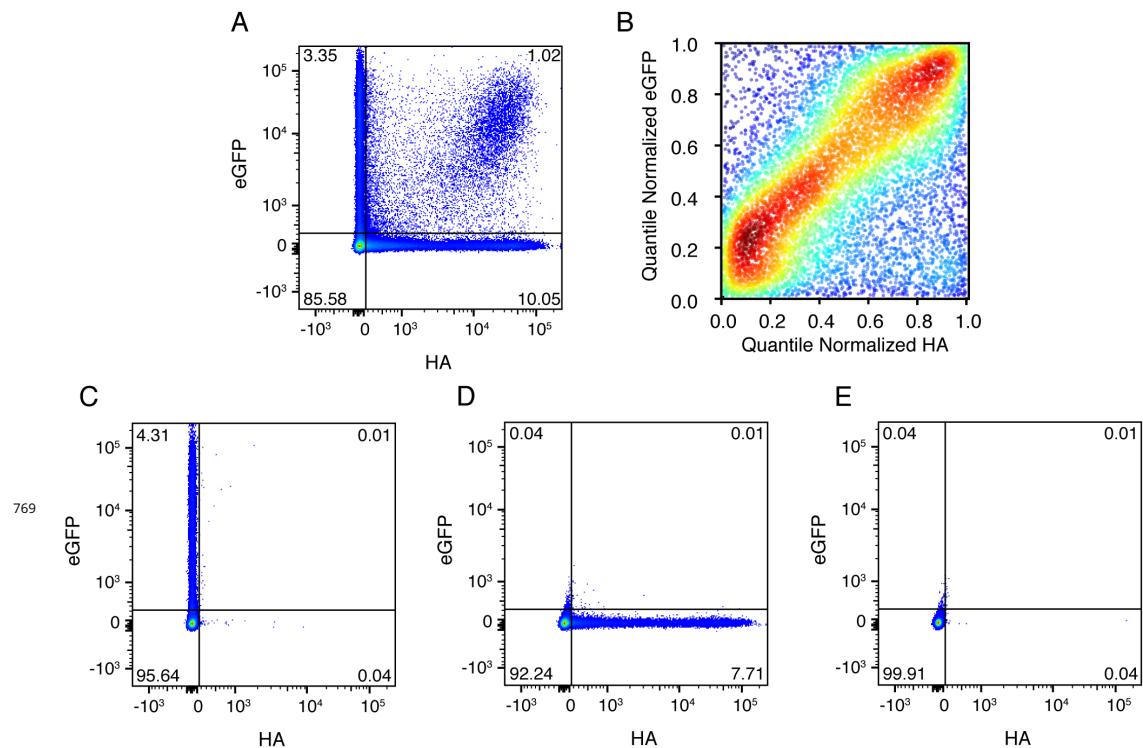
**Figure 5–Figure supplement 1.** A549 cells were infected at an MOI of 0.2 as calculated on MDCK-SIAT1 cells in either the presence or absence of the protein-translation inhibitor cyclohexamide, and viral mRNA was quantified at 8 hours post-infection by qPCR. The cyclohexamide prevents translation of new PB2, PB1, PA, and NP protein, and so prevents the formation of the new RNPs needed for secondary transcription. The bars show the relative amount of HA and PB2 mRNA in the absence versus the presence of cyclohexamide. Error  $\pm$  S.D. n=3.



**Figure 5–Figure supplement 2.** The fraction of mRNA in each infected cell derived as a function of the normalized expression of each viral gene, shown for the 10-hour and 8-hour samples individually (the other samples had too few infected cells for this analysis to be useful). Points are colored as in Figure 5A.



**Figure 5–Figure supplement 3.** The absence of viral RNP genes but *not* non-RNP genes remains significantly associated with reduced viral burden when we examine only the 10-hr sample, which is the single time point with the most data points. The difference for NP is no longer statistically significant due to low counts of infected cells lacking NP, but the trend remains. We do not show statistical analyses for other samples, as the number of infected cells is too low.



**Figure 7–Figure supplement 1.** **(A)** Cells were co-infected with a mix of wild-type virus and virus in which the HA gene was replaced by GFP flanked by the terminal regions of the HA gene segment. At 10 hours post-infection, cells were analyzed by flow cytometry for HA and eGFP expression. **(B)** The expression of HA and GFP are correlated in co-infected cells. Shown are the quantile-normalized HA and eGFP signals for double-positive cells. Cells are colored by density, using a Gaussian kernel density estimate. **(C),(D),(E)** Gating controls, single infection with eGFP virus, single infection with wild-type virus, and uninfected cells, respectively.

Category	Genes
Detoxification	AKR1C3, AKR1B10, GPX2, ALDH3A1, ALDH1A1, NQO1, CBR1, PRDX1
Protein folding	TXN, PPIA
Electron transport chain	NDUFB4, MT-CO1, MT-CO3
Regulators	ATF3, GADD45B
ROS-responsive relevance complex/unknown	UBB, NME1

**Figure 9–Figure supplement 1.** Table delineating genes in Figure 9 that are associated with the response to oxidative stress (*Duong et al., 2017; Jung et al., 2017; Lee and Ryu, 2017; Peuchant et al., 2017; MacLeod et al., 2016; Jiang et al., 2016; Gorrini et al., 2013; Miura et al., 2013; Kim et al., 2009; Banning et al., 2005; Murray et al., 2003; Doyle et al., 1999*).