

1 Extreme heterogeneity of influenza 2 virus infection in single cells

3 Alistair B. Russell¹, Cole Trapnell², Jesse D. Bloom^{1,2*}

*For correspondence:
jbloom@fredhutch.org

4 ¹Basic Sciences Division and Computational Biology Program, Fred Hutchinson Cancer
5 Research Center, Seattle, United States; ²Department 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

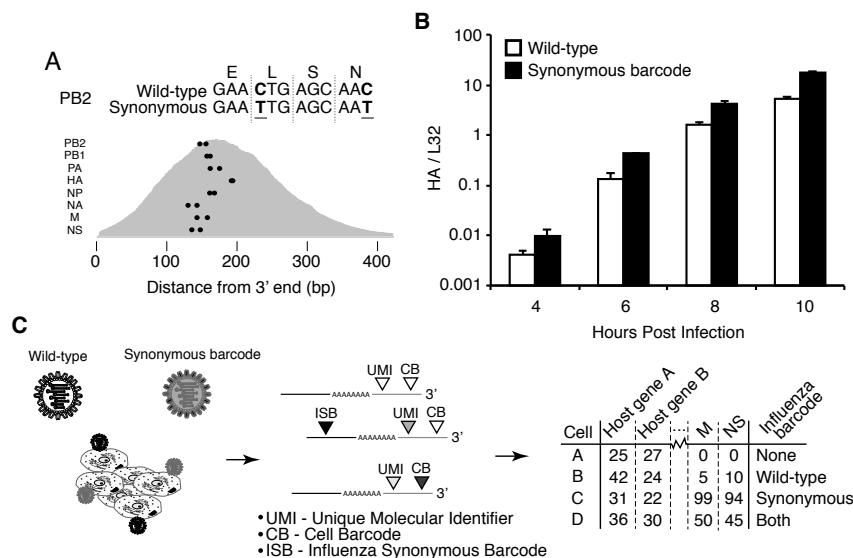


Figure 1. Experimental design. **(A)** We engineered a virus that carried two synonymous mutations near the 3' end of each mRNA. At top are the mutations for PB2. At bottom are locations of the synonymous mutations relative to the typical distribution of read depth for our 3'-end sequencing. **(B)** 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

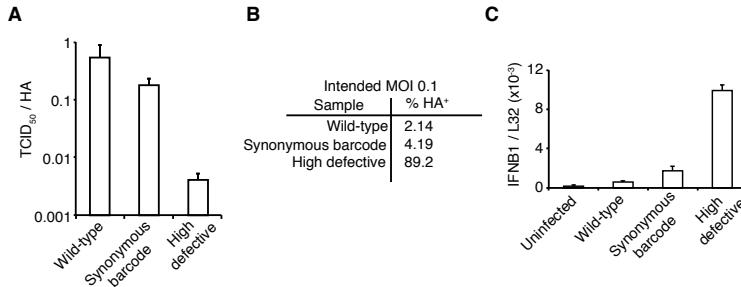


Figure 2. The viral stocks in our experiments are relatively pure of defective particles. **(A)** Our viral stocks have a higher ratio of infectious particles to HA virion RNA compared to a high-defective stock propagated at high MOI. HA viral RNA was quantified by qPCR on virions. Error bars \pm S.D., 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 *when used at the same number of infectious units as calculated by TCID₅₀. Note that this fact does not necessarily imply that they are more immunostimulatory per virion, as the high-MOI stocks also have more virions per infectious unit as shown in the first two panels.* 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
 65 UMIs cannot distinguish whether a cell was initially infected by one or multiple viral particles. We
 66 therefore engineered an influenza virus (strain A/WSN/1933) that additionally carried *viral barcodes*
 67 consisting of synonymous mutations near the 3' end of each transcript (*Figure 1A*). Critically, these
 68 synonymous mutations did not greatly impact viral growth kinetics (*Figure 1B*). We infected A549
 69 human lung carcinoma cells with an equal mix of the wild-type and synonymously barcoded
 70 viruses. Cells infected by a single virion will exclusively express mRNAs from either wild-type or
 71 synonymously barcoded virus, whereas cells that are co-infected with multiple virions will often
 72 express mRNAs 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. Stocks
 74 of viruses typically contain an array of biologically active viral particles, some of which are defec-
 75 tive 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
 78 high MOI, where complementation permits the growth of otherwise deleterious genotypes. To
 79 minimize the levels of defective particles, we propagated our viral stocks at low MOI for a rel-
 80 atively brief period of time (*Xue et al., 2016*). We validated that our stocks exhibited greater
 81 purity of infectious particles than a stock propagated at high MOI by verifying that they had a
 82 higher ratio of infectious particles to virion RNA (*Figure 2A*) and to particles capable of inducing
 83 expression of a single viral protein (*Figure 2B*). In addition, viral stocks with many defective parti-
 84 cles are more immunostimulatory (*Tapia et al., 2013; Lopez, 2014*)*per infectious unit* (e.g., TCID₅₀)
 85 *than low-defective stocks* (*Tapia et al., 2013; Lopez, 2014*), *in part simply because there are more*
 86 *physical virions per infectious unit* (*Figure 2A,B*). We confirmed that our viral stocks induced less
 87 *interferon per infectious unit* than a stock propagated at higher MOI (*Figure 2C*).

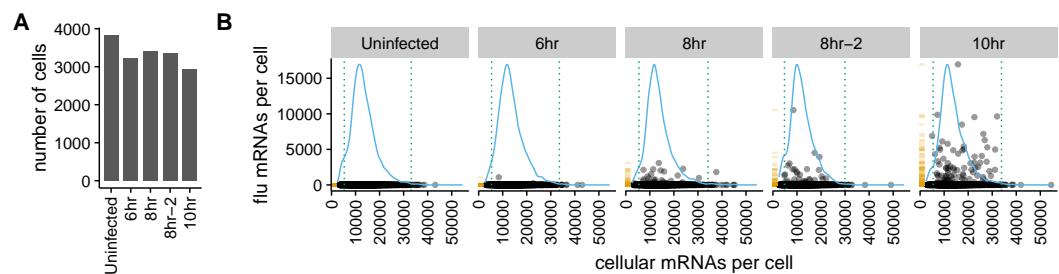


Figure 3. There is a very wide distribution in the amount of viral mRNA per cell. **(A)** Number of cells sequenced for each sample. **(B)** The number of cellular and viral mRNAs detected for each cell is plotted as a point. The blue lines show the overall distribution of the number of cellular mRNAs 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.](#)

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

89 We infected A549 cells at low MOI with a mixture of the wild-type and synonymously barcoded
 90 viruses, and collected cells for sequencing at 6, 8, and 10 hours post-infection, ~~including a replicate a~~
 91 ~~8-hours. We performing two slightly different variants of the experiment for the 8-hour timepoint.~~
 92 ~~For most of the samples, we replaced the infection inoculum with fresh media at one-hour post-~~
 93 ~~infection, thereby ensuring that most infection was initiated during a narrow time window. The~~
 94 ~~exception was the replicate However, for the second 8-hour sample , for which we (which we~~
 95 ~~denote as "8hr-2" in the figures), we did not perform this media change and instead left the cells in~~
 96 ~~the original infection inoculum. The rationale for including a sample without a media change was~~
 97 ~~to determine the importance of synchronicity of the timing of infection as discussed later in this~~
 98 ~~subsection.~~

99 We recovered between 3,000 and 4,000 cells for each sample (*Figure 3A*). As expected for a low-
 100 MOI infection, most cells expressed little or no viral mRNA (*Figure 3B*, *Figure 3–Figure Supplement 1*).
 101 Also as expected, the amount of viral mRNA per cell among infected cells increased over time
 102 (*Figure 3B*, *Figure 3–Figure Supplement 1*). But what was most notable was how widely the number
 103 of viral mRNA molecules varied among infected cells. While the fraction of mRNA derived from
 104 virus was <0.1% for most cells, viral mRNA constituted half the transcriptome in a few cells at 8 and
 105 10 hours (*Figure 3B*, *Figure 3–Figure Supplement 1*).

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

119 We determined the threshold amount of viral mRNA per cell ~~for each sample~~ at which the
 120 barcode partitioning clearly resulted from infection rather than leakage (*Figure 4C*, *Figure 4–Figure*

121 **Supplement 2**), and used ~~this threshold~~ these thresholds to annotate cells that we were confident
 122 were truly infected. We also annotated as co-infected cells above this threshold that had mRNA
 123 from both viral variants. **Figure 4D** shows the number of cells annotated as infected and co-infected
 124 for each sample – these cells are just a small fraction of the number of cells with any viral read.
 125 These annotation thresholds are conservative, and ~~will~~may miss some true low-level infections;
 126 ~~as well as any co-infections with the same viral variant~~. However, it is important that the analyses
 127 below are restricted to cells that are truly infected with virus, so we accepted the possible loss
 128 of some low-level infections in order to avoid false positives. In addition, the synonymous viral
 129 barcodes only identify co-infections by viruses with different barcodes – since the barcodes are at
 130 roughly equal proportion, we expect to miss about half of the co-infections. Since we annotate
 131 about ~10% of the infected cells as co-infected by viruses with different barcodes (Figure 4D), we
 132 expect another ~10% of the infected cells to also be co-infected but not annotated as so by our
 133 approach. Because most cells are not infected, we subsampled the uninfected cells to the numbers
 134 shown in **Figure 4D** to balance the proportions of infected and uninfected cells for all subsequent
 135 analyses.

136 Strikingly, the extreme variation in the number of viral transcripts per cell remains even after
 137 we apply these rigorous criteria for annotating infected cells (**Figure 4E**). The fraction of vi-
 138 ral mRNA per infected cell follows a roughly exponential distribution, with many cells having
 139 few viral transcripts and a few cells having many. At 6 and 8 hours <10% of infected cells are
 140 responsible for over half the viral transcripts, while at 10 hours \approx 15% of infected cells pro-
 141 duce over half the viral transcripts (**Figure 4–Figure Supplement 3**). One way to quantify the
 142 heterogeneity of a distribution is to calculate the Gini coefficient (Gini, 1921), which ranges from
 143 0 for a completely uniform distribution, to 1 for a maximally skewed distribution. **Figure 4–Figure**
 144 **Supplement 3** shows the Gini coefficients for the distribution of viral mRNA across infected cells for
 145 each sample. The Gini coefficients are >0.64 for all samples. As a fun point of comparison, these
 146 Gini coefficients indicate that the distribution of viral mRNA across infected cells is more uneven
 147 than the distribution of income in the United States (Alvaredo, 2011).

148 One possible source of heterogeneity in the amount of viral mRNA per cell is variability in the
 149 timing of infection. If some cells are infected earlier in the experiment than others, then they
 150 might have substantially more viral mRNA. However, several lines of evidence indicate that this
 151 is not the major cause of heterogeneity across cells. First, the sample for which the infection
 152 inoculum was never removed (8hr-2) only shows slightly more heterogeneity than samples for
 153 which the inoculum was washed away after one hour (**Figure 4E**, **Figure 4–Figure Supplement 3**),
 154 despite the fact that the potential time window for infection is much longer in the former sample.
 155 Second, in an independent experiment, we performed completely synchronized infections by
 156 pre-binding virus to cells on ice and then washing away unbound virus before bringing the cells
 157 to 37°C (Dapat et al., 2014). As shown in **Figure 4–Figure Supplement 4**, flow cytometry staining
 158 found that the heterogeneity in the levels of individual viral proteins was not markedly different
 159 for these synchronized infections than in the absence of pre-binding and washing. Finally, viral
 160 mRNA expression from the secondary spread of virus from infected cells does not appreciably
 161 occur during the timeframes of our experiments, since **Figure 4B** does not show the pervasive
 162 presence of mixed barcodes that would occur in this case. Therefore, variability in the timing of
 163 infection is not the dominant cause of the cell-to-cell heterogeneity in our experiments.

164 Notably, ~~co-infected cells do not express dramatically more viral mRNA than other~~
 165 ~~cells (Figure 4)~~ **E** shows that there are co-infected cells with both low and high amounts of viral
 166 mRNA, suggesting that the initial infectious dose ~~is not the major driver of cell-to-cell variation in~~
 167 ~~our low-MOI infections~~. ~~does not drive a simple continuous increase in viral transcript production.~~
 168 In support of this view, we used flow cytometry to quantify the levels of individual viral proteins in
 169 cells infected at various MOIs or for which we could delineate co-infection status (**Figure 4–Figure**
 170 **Supplement 5**). This analysis shows that sub-populations of cells that express similarly low and
 171 high levels of viral proteins persist across a wide range of infectious doses, although co-infection

172 can influence the relative proportion of infected cells that fall into these sub-populations (*Figure 4*–
 173 *Figure Supplement 5*).
 174

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

175 The influenza genome is segmented, and cells can fail to express a viral mRNA if the encoding
 176 gene segment is not packaged in the infecting virion or fails to initiate transcription after infection.
 177 Indeed, several groups have reported that the majority of infected cells fail to express at least
 178 one viral gene (*Brooke et al., 2013; Heldt et al., 2015; Dou et al., 2017*). We wondered if the
 179 absence of specific viral genes might be associated with reduced amounts of viral mRNA within
 180 single infected cells. In particular, transcription of influenza virus mRNAs is performed by the viral
 181 ribonucleoprotein (RNP) complex, which consists of the three proteins that encode the tripartite
 182 polymerase (PB2, PB1, and PA) as well as nucleoprotein (NP) (*Huang et al., 1990*). Each viral gene
 183 segment is associated with one RNP in incoming infecting virions, but secondary transcription by
 184 newly synthesized RNPs requires the presence of the viral genes encoding each of the four RNP
 185 proteins (*Vreede et al., 2004; Eisfeld et al., 2015*). This secondary transcription is a major source
 186 of viral mRNAs, as evidenced by the fact that blocking synthesis of the RNP proteins reduces the
 187 amount of viral mRNA by several orders of magnitude in bulk cells (*Figure 5–Figure Supplement 1*).
 188

189 We examined the total amount of viral mRNA versus the expression of ~~each viral gene~~ ~~the~~
 190 ~~genes from each viral segment~~ (*Figure 5A, Figure 5–Figure Supplement 2*)–*Figure 5–Figure Sup-*
191 plement 3). Note that influenza virus expresses ten major gene transcripts from its eight gene
 192 segments, as the M and NS segments are alternatively spliced to produce the M1 / M2 and NS1
 193 / NEP transcript, respectively (*Dubois et al., 2014*). However, an inherent limitation of current
 194 established single-cell mRNA sequencing techniques is that they only sequence the 3' end of the
 195 transcript (*Zheng et al., 2017; Macosko et al., 2015; Klein et al., 2015; Cao et al., 2017*). Since the
 196 alternative spliceoforms M1 / M2 and NS1 / NEP share the same 3' ends, we cannot distinguish
 197 them and therefore will refer simply to the combined counts of transcripts from each of these
 198 alternatively spliced segments as the M and NS genes.
 199

200 Cells that lack an RNP gene never derive more than a few percent of their mRNAs from virus,
 201 confirming the expected result that all four RNP genes are essential for high levels of viral trans-
 202 cription (*Figure 5A, Figure 5–Figure Supplement 2, Figure 5–Figure Supplement 3*). However, we
 203 observe cells that lack each of the other non-RNP genes but still derive ≈40% of their mRNAs from
 204 virus, suggesting that none of the other genes are important for high levels of viral transcription.
 205 These results are statistically supported by *Figure 5B*, which shows that absence of any RNP gene
 206 but *not* any other viral gene is associated with reduced amounts of viral mRNA. However, gene
 207 absence clearly does not explain all of the variability in viral gene expression, since even cells
 208 expressing all viral genes exhibit a very wide distribution in the amount of viral mRNA that they
 209 express. Specifically, at both 8 and 10 hours, the amount of viral mRNA in individual cells expressing
 210 all eight viral genes still ranges from <1% to >50% (*Figure 5A, Figure 5–Figure Supplement 2*)–*Fig-*
211 ure 5–Figure Supplement 3). Furthermore, the actual distribution of viral mRNA per infected cell
 212 (*Figure 4E*) does not match the mostly bi-modal shape expected under a simple model where RNP
 213 gene absence and Poisson co-infection are the only factors (*Figure 5–source data 2*), indicating
 214 that there are additional sources of variability beyond whether cells have full complement of RNP
 215 genes.
 216

217 We also quantified the fraction of infected cells that completely failed to express a given gene. We
 218 limited this analysis to examining the presence / absence of the non-RNP genes in cells expressing
 219 all four RNP genes, since we might fail to detect viral transcripts that are actually present at low
 220 levels in RNP-deficient cells due to the lower viral burden in these cells. At the 8- and 10-hour time
 221 points, between 5% and 17% of cells fail to express any one of the four non-RNP genes (*Figure 5C*,
Figure 5–source data 1). The absence of a given gene appears to be an independent event, as the
 222 probability of observing all four non-RNP genes in a cell is well predicted by simply multiplying
 223 the probabilities of observing each gene individually (*Figure 5C* and *Figure 5–source data 1*). If we
 224

222 extrapolate the frequencies at which cells lack non-RNP genes to the RNP genes, then we would
 223 predict that ~~40–50~~35–50% of infected cells express mRNAs from all eight genes. This estimate of the
 224 frequency at which infected cells express mRNAs from all eight gene segments is slightly higher than
 225 previous estimates of 13% (*Brooke et al., 2013*) and 20% (*Dou et al., 2017*). At least one difference
 226 is that *Brooke et al. (2013)* stained for proteins whereas we examined the expression of mRNAs – it
 227 is likely that some cells contain mutated viral genes that fail to produce stable protein even when
 228 mRNA is expressed.

229 **The relative amounts of different viral mRNAs are more consistent across cells.**

230 The results above show that the amount of viral mRNA in infected cells varies over several orders
 231 of magnitude. Does the relative expression of viral genes exhibit similar cell-to-cell variability?
 232 To address this question, we focused on cells that derived >5% of their mRNA from virus, since
 233 estimates of relative viral gene expression will be less noisy in cells with more viral mRNAs.

234 In contrast to the extreme variability in the total viral mRNA per cell, the fraction of this *viral*
 235 mRNA derived from each gene is much more consistent across cells (*Figure 6A*). Total viral mRNA
 236 varies by orders of magnitude, but the fraction from any given viral gene is fairly tightly clustered
 237 around the median value for all cells (*Figure 6B*). The relative levels of each viral mRNA in our cells
 238 are similar to prior bulk measurements made by Northern blots (*Hatada et al., 1989*), which also
 239 found an expression hierarchy of M > NS > NP > NA > HA > PB2 ~ PB1 ~ PA. The cell-to-cell
 240 consistency in the relative expression of different viral genes is even tighter if we limit the analysis
 241 only to cells that express all eight viral genes (*Figure 6C,D*). Therefore, with the exception of complete
 242 gene absence, the factors that drive the dramatic cell-to-cell variability in the amount of viral mRNA
 243 have roughly similar effects on all viral genes in a given cell. This finding is consistent with prior
 244 work showing positive correlations among the abundance of several viral genome segments in
 245 individual cells (*Heldt et al., 2015*). ~~It is also reminiscent of the observation that total cellular mRNA
 246 varies among single cells as a function of factors such as volume that affect all mRNA species
 247 concordantly (Padovan-Merhar et al., 2015).~~

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

249 Our sequencing enables us to identify the rare cells that were co-infected with both wild-type
 250 and synonymously barcoded viral variants. Overall, we captured 10 such co-infected cells that
 251 had >5% of their mRNA derived from virus (*Figure 7*). Seven of these 10 cells expressed all eight
 252 viral genes. ~~Remarkably, the~~ The majority (4 of 7) of these cells would *not* have expressed all the
 253 viral genes in the absence of co-infection, since they have at least one gene exclusively derived
 254 from each viral variant. For instance, the cell with 11.2% of its mRNA from virus in the upper
 255 right of *Figure 7* expresses M only from the wildtype viral variant, and NP and HA only from the
 256 synonymously barcoded variant. Our data therefore provide the first direct single-cell support for
 257 the idea observation of the fact that co-infection can rescue missing viral genes (*Brooke et al., 2013,*
 258 *2014; Fonville et al., 2015; Aguilera et al., 2017*).

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

264 To support this idea with a larger dataset albeit at lower resolution, we generated a virus in
 265 which the HA coding sequence was replaced by GFP. We then co-infected cells with a mix of wildtype
 266 and ΔHA-GFP virus and used flow cytometry to score cells for the presence of HA only (infection by
 267 wildtype virus), GFP only (infection by ΔHA-GFP virus), or both (co-infection) as shown in *Figure 7–*
Figure Supplement 1. As in our single-cell sequencing data, we found that expression of HA and GFP
 269 were highly correlated, indicating that co-infected cells typically expressed roughly equal amounts
 270 of transcript from each viral variant.

271 **Activation of the interferon response is rare in single infected cells.**

272 Because our sequencing captured all polyadenylated transcripts, we can examine whether there
 273 are prominent changes in the host-cell transcriptome in sub-populations of infected cells. Influenza
 274 virus infection can trigger innate-immune sensors that lead to the transcriptional induction of
 275 type I and III interferons, and subsequently of anti-viral interferon-stimulated genes (*Killip et al.,*
 276 *2015; Iwasaki and Pillai, 2014; Crotta et al., 2013*). However, activation of the interferon response
 277 is stochastic and bi-modal at the level of single cells (*Chen et al., 2010; Shalek et al., 2013, 2014;*
 278 *Perez-Cidoncha et al., 2014; Bhushal et al., 2017; Hagai et al., 2017*). We therefore hypothesized
 279 that we might see two sub-populations of infected cells: one in which the interferon response
 280 inhibited viral transcription, and another in which the virus was able to express high levels of its
 281 mRNA by evading or blocking this response.

282 To examine whether there were distinct sub-populations of virus-infected cells, we used a
 283 semi-supervised t-SNE approach (*Van der Maaten and Hinton, 2008*) to cluster cells by genes that
 284 co-varied with viral infection status. As shown in *Figure 8A,B*, this approach effectively grouped cells
 285 by the amount of viral mRNA that they expressed. Sample-to-sample variation was regressed away
 286 during the clustering, as cells did not obviously group by time-point, with expected exception that
 287 the uninfected and 6-hour samples had few cells in the region of the plot corresponding to large
 288 amounts of viral mRNA (*Figure 8C*).

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

300 We posited that the paucity of interferon induction might be due to the activity of influenza
 301 virus's major interferon antagonist, the NS1 protein (*García-Sastre et al., 1998; Hale et al., 2008*).
 302 We therefore identified cells that expressed substantial amounts of viral mRNA but lacked the
 303 NS gene (*Figure 8F*). Consistent with the idea that NS1 is important for suppressing interferon,
 304 the one interferon-positive cell lacked detectable expression of the NS gene. But other cells that
 305 lacked NS expression still failed to induce a detectable interferon response, despite often having a
 306 substantial amount of their mRNA derived from virus (*Figure 8*). This result is in line with other work
 307 showing that NS1-deficient influenza virus does not deterministically induce interferon (*Killip et al.,*
 308 *2017; Kallfass et al., 2013*). Therefore, many individual infected cells fail to activate innate-immune
 309 responses even when the virus lacks its major interferon antagonist.

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

311 We examined whether any host genes were differentially expressed in cells with more viral mRNA.
 312 We restricted this analysis to infected cells with all eight viral genes in order to focus on cellular
 313 genes that were associated with viral mRNA burden independent of effects due to the presence or
 314 absence of particular viral transcripts. We identified 45–43 cellular genes that co-varied with viral
 315 gene mRNA expression at a false discovery rate of 0.1 (*Figure 9, Figure 9-source data 1*).

316 Many of the genes with increased expression in cells with more A gene-set analysis shows that
 317 many cellular genes that are associated with the amount of viral mRNA are involved in the response
 318 to reactive oxygen species or hypoxia (*Figure 9-source data 2*). Genes known or suspected to be
 319 regulated by the Nrf2 master regulator in response to oxidative stress are often expressed at
 320 higher levels in cells with more viral mRNA (*Figure 9*). These genes produce proteins that are

321 involved in detoxification of reactive oxygen species or resultant products, the management of
 322 misfolded proteins, the electron transport chain, or a general stress response (**Figure 9–Figure**
 323 **Supplement 1**). We additionally see reduced expression of the nitric oxide synthase interacting
 324 protein (NOSIP). Transient oxidative stress is known to occur during viral infection, and may act in a
 325 proviral fashion via MAPK activation driving vRNP export (*Amatore et al., 2014*). The antioxidant
 326 response is thought to be largely antiviral, potentially through inhibition of MAPK activity (*Lin et al.,*
 327 *2016; Sgarbanti et al., 2014*). Our data do not reveal whether the expression of genes involved
 328 in the response to oxidative stress are To directly test the effect of transient oxidative stress, we
 329 compared the fraction of cells that expressed detectable viral protein when infected either with or
 330 without pre-treatment to suppress oxidative stress. **Figure 9–Figure Supplement 2** shows that the
 331 cells pre-treated with an antioxidant exhibited less frequent detectable expression of viral protein.
 332 These results, in conjunction with the differential expression test in **Figure 9** and the prior work
 333 mentioned above, suggest that oxidative stress acts in a proviral fashion.

334 The gene-set analysis also found that the amount of viral mRNA was associated with the
 335 expression of genes involved in the G2-M cell-cycle checkpoint (**Figure 9–source data 2**). The
 336 cell-cycle associated genes CCND3, MKI67, UBE2S, and CENPF are all expressed at significantly
 337 lower levels in cells with more viral mRNA (**Figure 9**). However, our data are not sufficient to
 338 determine whether the lower expression of these genes is a cause or a symptom of higher levels
 339 of viral mRNA, and further investigation of this topic is an interesting area for future work. effect
 340 of the reduction in viral mRNA.

341 Interestingly, none of the cellular genes that are significantly associated with the amount of viral
 342 mRNA in our study are among the 128 genes that *Watanabe et al. (2010)* report as having been
 343 identified multiple times in genome-wide screens for factors affecting influenza virus replication.
 344 One possible explanation is that most of the cell-to-cell heterogeneity in our experiments might
 345 arise from viral segment absence or mutations, pure stochasticity, or more subtle alterations in
 346 host-cell state – not due to changes in expression of the type of single large-effect genes that are
 347 usually identified in genome-wide knockdown / knockout studies.

348 Discussion

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

353 The most obvious form of heterogeneity is the complete failure of some infected cells to express
 354 one or more viral genes, which we estimate occurs in about half the infected cells in our experiments.
 355 The absence of some viral genes in some infected cells has been noted previously (*Brooke et al.,*
 356 *2013; Heldt et al., 2015; Dou et al., 2017*), and our work provides a holistic view by quantifying the
 357 total viral transcriptional load as a function of the level of each mRNA. We find that cells lacking
 358 expression of any of the four genes that encode the viral RNP express much less total viral mRNA,
 359 consistent with prior bulk studies (*Vreede et al., 2004; Eisfeld et al., 2015*). Interestingly, the reason
 360 some cells fail to express some viral genes remains unclear. The prototypical influenza virion
 361 packages one copy of each of the eight gene segments (*Noda et al., 2006; Hutchinson et al., 2010*),
 362 but some virions surely package fewer (*Brooke et al., 2014*). However, it is also possible that much
 363 of the viral gene absence is due to stochastic loss of viral RNPs after infection but prior to the
 364 initiation of viral transcription in the nucleus.

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

371 influenza virus's mutation rate suggest that the typical virion in a stock such as the one used in our
 372 experiment should contain less than one mutation per genome (*Parvin et al., 1986; Suárez et al.,*
 373 *1992; Suárez-López and Ortín, 1994; Nobusawa and Sato, 2006; Bloom, 2014; Pauly et al., 2017*).

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

386 Our data do suggest that the factors driving the heterogeneity in viral transcriptional load exert
 387 relatively concordant effects on all viral genes in a given cell. Specifically, despite the extreme
 388 heterogeneity in total viral mRNA per cell, the relative levels of the viral mRNAs are reasonably
 389 consistent across cells, and generally reflective of classical bulk measurements (*Hatada et al., 1989*).
 390 Therefore, despite the stochasticity inherent in initiating transcription and replication of each gene
 391 from a single copy carried by the incoming virion, as long as a gene is not completely lost then the
 392 virus possesses mechanisms to control its relative expression (*Shapiro et al., 1987; Hatada et al.,*
 393 *1989; Perez et al., 2010; Heldt et al., 2012; Chua et al., 2013*).

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

410 Overall, our work shows the power and importance of ~~characterizing~~characterizing cellular infection
 411 at the level of single cells (*Avraham et al., 2015*). ~~The dynamics of viral infection in any given cell~~
 412 is shaped by Viral infection can involve heterogeneity in the genetic composition of the incoming
 413 virion, the host-cell state, the bi-modality of innate-immune activation, and the inherent stochas-
 414 ticity of molecular processes initiated by a single copy of each viral gene. ~~We have shown how the~~
 415 confluence of these factors leads to extreme cell-to-cell In our experiments with short-timeframe
 416 and low-MOI infections with a relatively pure stock of influenza virus, we find only a minor role for
 417 innate-immune activation, but a substantial role for heterogeneity in the transcriptional outcome
 418 of influenza virus infection. Further complement of viral genes that are expressed in individual
 419 cells and at least some contribution of host-cell state. Our current experiments are not able to
 420 quantify the role of other possibly important factors such as mutations in viral genes, but we
 421 suspect that they may also contribute. Future extensions of the approaches described here should

enable further deconstruction of the contributions of each factor will enable sources of cell-to-cell heterogeneity during viral infection, thereby enabling a deeper understanding of how the bulk features of infection emerge from the processes occurring processes within individual virus-infected cells.

Methods and Materials

Cell lines and viruses

The following cell lines were used in this study: the human lung epithelial carcinoma line A549 (ATCC CCL-185), the MDCK-SIAT1 variant of the Madin Darby canine kidney cell line overexpressing human SIAT1 (Sigma-Aldrich 05071502), and the human embryonic kidney cell line 293T (ATCC CRL-3216). All cells were maintained in D10 media (DMEM supplemented with 10% heat-inactivated fetal bovine serum, 2 mM L-glutamine, 100 U of penicillin/ml, and 100 µg of streptomycin/ml) at 37 °C at 5 % CO₂.

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

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

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

qPCR

For the qPCR in *Figure 2* and *Figure 5–Figure Supplement 1*, A549 cells were seeded at 3x10⁵ cells per well in a 6-well tissue culture plate in D10 the day prior to infection. On the day of infection, a single well was trypsinized and the cells were counted in order to determine the appropriate amount of virus to use to achieve the intended MOI. Immediately before infection, D10 was replaced with influenza growth media. For cells incubated with cyclohexamide, the compound was added to a final concentration of 50 µg/ml at the time of infection – previously confirmed to be sufficient to halt viral protein production (*Killip et al., 2014*). RNA was purified

470 using the QIAGEN RNeasy plus mini kit following manufacturer's instructions. cDNA was syn-
 471 thesized using an oligoDT primer and the SuperScript™ III first-strand synthesis supermix from
 472 ThermoFisher using the manufacturer's protocol. Transcript abundance was measured using
 473 SYBR™ green PCR master mix, using a combined anneal/extension step of 60 °C for one minute
 474 with the following primers: *HA*: 5'-GGCCCAACCACACATTCAAC-3', 5'-GCTCATCACTGCTAGACGGG-
 475 3', *IFNB1*: 5'-AAACTCATGAGCAGTCTGCA-3', 5'-AGGAGATCTCAGTTGGAGG-3', *L32*: 5'-
 476 AGCTCCAAAAATAGACGCAC-3', 5'-TTCATAGCAGTAGGCACAAAGG-3'. Biological triplicates were per-
 477 formed for all samples.

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

486 **Flow cytometry titering and analyses**

487 To determine viral titers in terms of HA-expressing units and for the flow cytometry ~~in~~, A549 cells
 488 were seeded in a 6-well plate and infected as described above for the qPCR analyses. Cells were
 489 harvested by trypsinization, resuspended in phosphate-buffered saline supplemented with 2% heat-
 490 inactivated FBS, and stained with 10 µg/ml of H17-L19, a mouse monoclonal antibody confirmed
 491 to bind to WSN HA in a prior study (**Doud et al., 2017**). After washing in PBS ~~plus supplemented~~
 492 ~~with~~ 2% FBS, the cells were stained with a goat anti-mouse IgG antibody conjugated to APC. Cells
 493 were then washed, fixed in 1% formaldehyde, and washed further before a final resuspension
 494 and analysis. We then determined the fraction of cells that were HA positive and calculated the
 495 HA-expressing units.

496 ~~For, after gating to exclude multiplets~~ For NS1 staining, cells stained for HA as described
 497 above were permeabilized using BD Cytofix/Cytoperm following manufacturer's instructions,
 498 stained with anti-NS1 (GTX125990, Genetex) at 4.4 µg/ml, washed, stained with a goat anti-rabbit
 499 IgG antibody conjugated to Alexa Fluor 405, washed, and analyzed. To analyze the effect of
 500 N-acetylcysteine, the compound was added to cells in D10 9h prior to media change and infection,
 501 and included in infection media. Stocks of N-acetylcysteine were reconstituted immediately prior
 502 to use, and the pH of growth media supplemented with this compound was adjusted using sodium
 503 hydroxide. After channels were compensated and cells gated to exclude multiplets and debris in
 504 FlowJo, data were extracted using the R package flowCore (**Le Meur et al., 2007**) and analyzed using
 505 a custom Python script. Gaussian kernel density estimates were obtained using the scipy stats
 506 package method, gaussian_kde, using automatic bandwidth determination (**van der Walt et al.,**
 507 **2017**). For gating on NS1 positive cells, the percentage of influenza-infected cells was determined
 508 by HA staining alone, and the top quantile of NS1-stained cells matching that percentage were
 509 taken as the NS1 positive population.

510 **Infections for single-cell mRNA sequencing**

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

513 All time points except for the ~~second~~ 8-hour ~~replicate sample~~ (8hr-2) were prepared on the same
 514 day. For the infections, A549 cells were seeded in a 6-well plate, with two wells per time point. A
 515 single well of cells was trypsinized and counted prior to initiation of the experiment for the purposes
 516 of calculating MOI. Wild-type and synonymously barcoded virus were mixed to an estimated ratio
 517 of 1:1 based on prior, exploratory, single-cell analyses (data not shown). At the initiation of our
 518 experiment, the wells for all time points were changed from D10 to influenza growth media. Cells

519 were then infected with 0.3 HA-expressing units of virus per cell (a determined by flow cytometry).
 520 The infections were performed in order of time point: first the 10-hour time point, then the 8-hour,
 521 and then the 6-hour time point. At one hour after infection, the media for each time point was
 522 changed to fresh influenza growth media. Note that the HA-expressing units were calculated without
 523 this additional washing step, and so likely represent an overestimate of our final infectious dose
 524 (consistent with the fact that fewer than 30% of cells appear infected in the single-cell sequencing
 525 data). All cells were then harvested for single-cell analysis concurrently – ensuring all had spent
 526 equivalent time in changed media . For our replicate 8-hour time point 8hr-2 sample, cells were
 527 infected as above except that the cells were infected at 0.1 HA-expressing units of virus per cell but
 528 no wash step was performed, and the sample was prepared on a different day. After harvest, cells
 529 were counted using disposable hemocytometers and diluted to equivalent concentrations with an
 530 intended capture of 3000 cells/sample following the manufacturer's provided by 10x Genomics for
 531 the Chromium Single Cell platform. All subsequent steps through library preparation followed the
 532 manufacturer's protocol. Samples were sequenced on an Illumina HiSeq. The deep sequencing
 533 data are available on the Sequence Read Archive under accession-

534 Computational analysis of single-cell mRNA sequencing data

535 Jupyter notebooks that perform all of the computational analyses are available in Supplemen-
 536 tary file 1 and at https://github.com/jbloomlab/flu_single_cell [*This GitHub repository will be made*
 537 *public upon acceptance of the manuscript.*]

538 Briefly, the raw deep sequencing data were processed using the 10X Genomics software pack-
 539 age CellRanger (version 2.0.0). The reads were aligned to a concatenation of the human and
 540 influenza virus transcriptomes. The human transcriptome was generated by filtering genome
 541 assembly GRCh38 for protein coding genes defined in the GTF file GRCh38.87. The influenza
 542 virus transcriptome was generated from the reverse-complement of the wildtype WSN viral
 543 RNA sequences as encoded in the reverse-genetics plasmids (*Figure 1*-source data 1). The
 544 aligned deep sequencing data are available on the GEO repository under accession GSE108041
 545 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE108041>).

546 CellRanger calls cells based on the number of observed cell barcodes, and creates a cell-gene
 547 matrix. We used custom Python code to annotate the cells in this matrix by the number of viral
 548 reads that could be assigned to the wildtype and synonymously barcoded virus. Only about half of
 549 the viral reads overlapped the barcoded regions of the genes (*Figure 1A*) and could therefore be
 550 assigned to a viral barcode (*Figure 4-Figure Supplement 1*). So for calculations of the number of
 551 reads in a cell derived from each viral barcode for each viral gene, the total number of detected
 552 molecules of that gene are multiplied by the fraction of those molecules with assignable barcodes
 553 that are assigned to that barcode. This annotated cell-gene matrix is in Supplementary file 2. A
 554 Jupyter notebook that performs these analyses is in Supplementary file 1.

555 The annotated cell-gene matrix was analyzed in R, primarily using the Monocle package (version
 556 2.4.0) (*Qiu et al., 2017; Trapnell et al., 2014*). A Jupyter notebook that performs these analyses is in
 557 Supplementary file 1. For each sample, cell barcodes that had >2.5-fold fewer or more UMI counts
 558 mapping to cellular transcripts than the sample mean were excluded from downstream analyses
 559 (see red vertical lines in *Figure 3B*).

560 In order to determine an appropriate cutoff for how many reads a cell needed to contain in order
 561 to be classified as infected, we calculated the mean viral barcode purity across all cells that contained
 562 at least a given fraction of viral mRNA and had multiple viral reads that could be assigned a barcode
 563 (*Figure 4B,C and Figure 4-Figure Supplement 2*). We then determined the threshold fraction of viral
 564 mRNA at which the mean purity no longer increased as a function of the amount of viral mRNA.
 565 This threshold represents the point at which we have effectively eliminated cells that have low
 566 barcode purity simply due to lysis-acquired reads sampled randomly from both viral barcodes.
 567 The As is apparent from Figure 4B, only the 10-hour sample has the most viral mRNA, and so
 568 is expected to and the 8hr-2 sample have the excess of mixed barcodes among cells with low

569 amounts of viral mRNA. The likely reason is that these samples have more total viral mRNA (and
 570 so there is more available mRNA to be acquired from lysed cells); in addition, there is always some
 571 experimental variability in the amount of cell lysis during the 10X sequencing process, and these
 572 samples may simply have the most lysis-associated influenza reads. We therefore determined the
 573 threshold using just this sample, and then applied this threshold to all samples. The threshold is
 574 . So the above threshold procedure is appropriate for those two samples. For the other samples,
 575 we simply set a minimum threshold of requiring at least a fraction 2×10^{-4} reads to come from
 576 viral mRNA as explained in the legend to **Figure 4–Figure Supplement 2**. The thresholds for each
 577 sample are shown in **Figure 4C** and **Figure 4–Figure Supplement 2**. This procedure is expected to
 578 be conservative, and may miss some truly infected cells with very low amounts of viral mRNA. For
 579 subsequent analyses, we retained all infected cells and a subsample of uninfected cells (the greater
 580 of 50 or the number of infected cells for that sample). The rationale for subsampling the uninfected
 581 cell is that the vast majority of cells are uninfected, and we did not want these cells to completely
 582 dominate the downstream analyses. Cells were classified as co-infected if both viral variants had an
 583 RNA level that exceeded the threshold, and if the minor variant contributed at least 5% of the viral
 584 mRNA.

585 For the semi-supervised t-SNE clustering, we used Monocle's cell hierarchy function to bin cells
 586 into those with no viral mRNA, <2% viral mRNA, between 2% and 20% viral mRNA, and >20%.
 587 Candidate marker genes for t-SNE dimensionality reduction were then determined using the
 588 Monocle function markerDiffTable, excluding the effects of sample variation and the number of
 589 genes identified in a given cell, using a q-value cutoff of 0.01. The specificity of these markers was
 590 determined using the function calculateMarkerSpecificity – the top **40–50** markers were retained,
 591 and used to place populations in a two-dimensional plane based on tSNE dimensionality reduction.

592 For the analyses of cellular genes that differed in expression as a function of the amount of
 593 viral mRNA, we only considered cells that expressed all 8 viral mRNAs to avoid effects driven
 594 simply by viral gene absence. We also only considered cellular genes in the differential gene
 595 analysis, since viral gene expression will tautologically co-vary with the amount of viral mRNA.
 596 Additionally, because influenza virus has the capacity to degrade or prevent the synthesis of
 597 host mRNAs (*Bercovich-Kinori et al., 2016*) and contributes significantly to the total number UMIs
 598 in some cells, we calculate size factors (a scalar value representing efficiency of UMI capture)
 599 based on cellular transcripts alone. Finally, we assigned all cells a ceiling fraction of mRNA from
 600 virus of 25% so that a few extremely high-expressing cells did not dominate. Cellular genes with
 601 expression that co-varied with the fraction of viral mRNAs in a cell were then determined using the
 602 Monocle differentialGeneTest, after removing variance explained by sample to sample variation.
 603 **Figure 9** shows all genes that were significantly associated with the fraction of mRNA from virus
 604 at a false discovery rate of 0.1. We performed the gene set analysis using the P -values from the
 605 Monocle differentialGeneTest with piano (*Väremo et al., 2013*) using the hallmark gene set from
 606 GSEA v6 (*Subramanian et al., 2005*) and Fisher's method.

607 Acknowledgments

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

611 References

- 612 Aguilera ER, Erickson AK, Jesudhasan PR, Robinson CM, Pfeiffer JK. Plaques formed by mutagenized viral
 613 populations have elevated coinfection frequencies. *mBio*. 2017; 8(2):e02020.
- 614 Akpinar F, Timm A, Yin J. High-Throughput Single-Cell Kinetics of Virus Infections in the Presence of Defective
 615 Interfering Particles. *Journal of Virology*. 2016 Jan; 90(3):1599–1612.
- 616 Alvaredo F. A note on the relationship between top income shares and the Gini coefficient. *Economics Letters*.
 617 2011; 110(3):274–277.

- 618 **Amatore D**, Sgarbanti R, Aquilano K, Baldelli S, Limongi D, Civitelli L, Nencioni L, Garaci E, Ciriolo MR, Palamara
619 AT. Influenza virus replication in lung epithelial cells depends on redox-sensitive pathways activated by
620 NOX4-derived ROS. *Cellular microbiology*. 2014 Oct; 17(1):131–145.
- 621 **Avraham R**, Haseley N, Brown D, Penaranda C, Jijon HB, Trombetta JJ, Satija R, Shalek AK, Xavier RJ, Regev A, et al.
622 Pathogen cell-to-cell variability drives heterogeneity in host immune responses. *Cell*. 2015; 162(6):1309–1321.
- 623 **Banning A**, Deubel S, Kluth D, Zhou Z, Brigelius-Flohe R. The GI-GPx Gene Is a Target for Nrf2. *Molecular and
624 Cellular Biology*. 2005 May; 25(12):4914–4923.
- 625 **Bercovich-Kinori A**, Tai J, Gelbart IA, Shitrit A, Ben-Moshe S, Drori Y, Itzkovitz S, Mandelboim M, Stern-Ginossar
626 N. A systematic view on influenza induced host shutoff. *eLife*. 2016 Aug; 5:e18311.
- 627 **Bhushal S**, Wolfsmüller M, Selvakumar TA, Kemper L, Wirth D, Hornef MW, Hauser H, Köster M. Cell polarization
628 and epigenetic status shape the heterogeneous response to Type III interferons in intestinal epithelial cells.
629 *Frontiers in immunology*. 2017; 8:671.
- 630 **Bloom JD**. An experimentally determined evolutionary model dramatically improves phylogenetic fit. *Molecular
631 Biology and Evolution*. 2014; 31(8):1956–1978.
- 632 **Brooke CB**, Ince WL, Wrammert J, Wrammert J, Ahmed R, Wilson PC, Bennink JR, Yewdell JW. Most Influenza A
633 Virions Fail To Express at Least One Essential Viral Protein. *Journal of Virology*. 2013 Feb; 87(6):3155–3162.
- 634 **Brooke CB**. Biological activities of ‘noninfectious’ influenza A virus particles. *Future Virology*. 2014 Jan; 9(1):41–51.
- 635 **Brooke CB**, Ince WL, Wei J, Bennink JR, Yewdell JW. Influenza A virus nucleoprotein selectively decreases
636 neuraminidase gene-segment packaging while enhancing viral fitness and transmissibility. *Proceedings of
637 the National Academy of Sciences*. 2014; 111(47):16854–16859.
- 638 **Buganim Y**, Faddah DA, Cheng AW, Itsikovich E, Markoulaki S, Ganz K, Klemm SL, van Oudenaarden A, Jaenisch R.
639 Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierachic
640 phase. *Cell*. 2012; 150(6):1209–1222.
- 641 **Cai L**, Friedman N, Xie XS. Stochastic protein expression in individual cells at the single molecule level. *Nature*.
642 2006; 440(7082):358–362.
- 643 **Cao J**, Packer JS, Ramani V, Cusanovich DA, Huynh C, Daza R, Qiu X, Lee C, Furlan SN, Steemers FJ, et al.
644 Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science*. 2017; 357(6352):661–
645 667.
- 646 **Chen S**, Chen S, Short JAL, Young DF, Killip MJ, Schneider M, Schneider M, Goodbourn S, Randall RE. Heterocellular
647 induction of interferon by negative-sense RNA viruses. *Virology*. 2010 Nov; 407(2):247–255.
- 648 **Chua MA**, Schmid S, Perez JT, Langlois RA, et al. Influenza A virus utilizes suboptimal splicing to coordinate the
649 timing of infection. *Cell reports*. 2013; 3(1):23–29.
- 650 **Combe M**, Garijo R, Geller R, Cuevas JM, Sanjuán R. Single-cell analysis of RNA virus infection identifies multiple
651 genetically diverse viral genomes within single infectious units. *Cell host & microbe*. 2015; 18(4):424–432.
- 652 **Crotta S**, Davidson S, Mahlakoiv T, Desmet CJ, Buckwalter MR, Albert ML, Staeheli P, Wack A. Type I and type III
653 interferons drive redundant amplification loops to induce a transcriptional signature in influenza-infected
654 airway epithelia. *PLoS Pathogens*. 2013; 9(11):e1003773.
- 655 **Dapat C**, Saito R, Suzuki H, Horigome T. Quantitative phosphoproteomic analysis of host responses in human
656 lung epithelial (A549) cells during influenza virus infection. *Virus research*. 2014 Jan; 179:53–63.
- 657 **Delbruck M**. The burst size distribution in the growth of bacterial viruses (bacteriophages). *Journal of bacteriology*.
658 1945 Aug; 50(2):131–135.
- 659 **Dimmock NJ**, Dimmock NJ, Easton AJ, Easton AJ. Defective Interfering Influenza Virus RNAs: Time To Reevaluate
660 Their Clinical Potential as Broad-Spectrum Antivirals? *Journal of Virology*. 2014 Apr; 88(10):5217–5227.
- 661 **Dou D**, Hernández-Neuta I, Wang H, Östbye H, Qian X, Thiele S, Resa-Infante P, Kouassi NM, Sender V, Henrich K,
662 Mellroth P, Henriques-Normark B, Gabriel G, Nilsson M, Daniels R. Analysis of IAV Replication and Co-infection
663 Dynamics by a Versatile RNA Viral Genome Labeling Method. *Cell reports*. 2017 Jul; 20(1):251–263.
- 664 **Doud M**, Bloom J. Accurate Measurement of the Effects of All Amino-Acid Mutations on Influenza Hemagglutinin.
665 *Viruses*. 2016 Jun; 8(6):155–17.

- 666 **Doud MB**, Hensley SE, Bloom JD. Complete mapping of viral escape from neutralizing antibodies. PLoS
667 Pathogens. 2017 Mar; 13(3):e1006271.
- 668 **Doyle V**, Virji S, Crompton M. Evidence that cyclophilin-A protects cells against oxidative stress. The Biochemical
669 journal. 1999 Jul; 341 (Pt 1)(Pt 1):127–132.
- 670 **Dubois J**, Terrier O, Rosa-Calatrava M. Influenza viruses and mRNA splicing: doing more with less. MBio. 2014;
671 5(3):e00070-14.
- 672 **Duong HQ**, You K, Oh S, Kwak SJ, Seong YS. Silencing of NRF2 Reduces the Expression of ALDH1A1 and ALDH3A1
673 and Sensitizes to 5-FU in Pancreatic Cancer Cells. Antioxidants. 2017 Sep; 6(3):52–10.
- 674 **Eisfeld AJ**, Neumann G, Kawaoka Y. At the centre: influenza A virus ribonucleoproteins. Nature reviews
675 Microbiology. 2015; 13(1):28.
- 676 **Elowitz MB**, Levine AJ, Siggia ED, Swain PS. Stochastic gene expression in a single cell. Science. 2002;
677 297(5584):1183–1186.
- 678 **Fensterl V**, Sen GC. The ISG56/IFIT1 gene family. Journal of Interferon & Cytokine Research. 2011 Jan; 31(1):71–
679 78.
- 680 **Fonville JM**, Marshall N, Tao H, Steel J, Lowen AC. Influenza Virus Reassortment Is Enhanced by Semi-
681 infectious Particles but Can Be Suppressed by Defective Interfering Particles. PLoS Pathogens. 2015 Oct;
682 11(10):e1005204.
- 683 **García-Sastre A**, Egorov A, Matassov D, Brandt S, Levy DE, Durbin JE, Palese P, Muster T. Influenza A virus lacking
684 the NS1 gene replicates in interferon-deficient systems. Virology. 1998; 252(2):324–330.
- 685 **Geiss GK**, Salvatore M, Tumpey TM, Carter VS, Wang X, Basler CF, Taubenberger JK, Bumgarner RE, Palese
686 P, Katze MG, García-Sastre A. Cellular transcriptional profiling in influenza A virus-infected lung epithelial
687 cells: the role of the nonstructural NS1 protein in the evasion of the host innate defense and its potential
688 contribution to pandemic influenza. Proceedings of the National Academy of Sciences of the United States of
689 America. 2002 Aug; 99(16):10736–10741.
- 690 **Gini C**. Measurement of inequality of incomes. The Economic Journal. 1921; 31(121):124–126.
- 691 **Gorrini C**, Harris IS, Mak TW. Modulation of oxidative stress as an anticancer strategy. Nature Publishing Group.
692 2013 Dec; 12(12):931–947.
- 693 **Hagai T**, Chen X, Miragaia RJ, Gomes T, Rostom R, Kunowska N, Proserpio V, Donati G, Bossini-Castillo L, Naamati
694 G, Emerton G, Trynka G, Kondova I, Denis M, Teichmann SA. Gene expression variability across cells and
695 species shapes innate immunity. bioRxiv. 2017; doi: 10.1101/137992.
- 696 **Hale BG**, Randall RE, Ortín J, Jackson D. The multifunctional NS1 protein of influenza A viruses. Journal of
697 General Virology. 2008; 89(10):2359–2376.
- 698 **Hatada E**, Hasegawa M, Mukaigawa J, Shimizu K, Fukuda R. Control of influenza virus gene expression: quanti-
699 tative analysis of each viral RNA species in infected cells. The Journal of Biochemistry. 1989; 105(4):537–546.
- 700 **Heldt FS**, Frensing T, Reichl U. Modeling the intracellular dynamics of influenza virus replication to understand
701 the control of viral RNA synthesis. Journal of Virology. 2012; 86(15):7806–7817.
- 702 **Heldt FS**, Kupke SY, Dorl S, Reichl U, Frensing T. Single-cell analysis and stochastic modelling unveil large
703 cell-to-cell variability in influenza A virus infection. Nature communications. 2015 Nov; 6:1–12.
- 704 **Hoffmann E**, Stech J, Guan Y, Webster RG, Perez DR. Universal primer set for the full-length amplification of all
705 influenza A viruses. Archives of virology. 2001 Dec; 146(12):2275–2289.
- 706 **Hoffmann E**, Neumann G, Kawaoka Y, Hobom G, Webster RG. A DNA transfection system for generation of
707 influenza A virus from eight plasmids. Proceedings of the National Academy of Sciences. 2000; 97(11):6108–
708 6113.
- 709 **Huang AS**, Baltimore D, et al. Defective viral particles and viral disease processes. Nature, London. 1970;
710 226:325–7.
- 711 **Huang IC**, Li W, Sui J, Marasco W, Choe H, Farzan M. Influenza A Virus Neuraminidase Limits Viral Superinfection.
712 Journal of Virology. 2008 Apr; 82(10):4834–4843.

- 713 **Huang T**, Palese P, Krystal M. Determination of influenza virus proteins required for genome replication. *Journal of Virology*. 1990; 64(11):5669–5673.
- 715 **Hutchinson EC**, von Kirchbach JC, Gog JR, Digard P. Genome packaging in influenza A virus. *Journal of general virology*. 2010; 91(2):313–328.
- 717 **Iwasaki A**, Pillai PS. Innate immunity to influenza virus infection. *Nature Reviews Immunology*. 2014 May; 14(5):315–328.
- 719 **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 antimony-induced oxidative stress and apoptosis in HEK293 cells. *Toxicology Letters*. 2016 Aug; 256:11–18.
- 721 **Jung KA**, Lee S, Kwak MK. NFE2L2/NRF2 Activity Is Linked to Mitochondria and AMP-Activated Protein Kinase Signaling in Cancers Through miR-181c/Mitochondria-Encoded Cytochrome C Oxidase Regulation. *Antioxidants & Redox Signaling*. 2017 Apr; p. ars.2016.6797–3.
- 724 **Kallfass C**, Lienenklaus S, Weiss S, Staeheli P. Visualizing the beta interferon response in mice during infection with influenza A viruses expressing or lacking nonstructural protein 1. *Journal of Virology*. 2013 Jun; 87(12):6925–6930.
- 727 **Killip MJ**, Jackson D, Perez-Cidoncha M, Fodor E, Randall RE. Single-cell studies of IFN- β promoter activation by wild-type and NS1-defective influenza A viruses. *Journal of General Virology*. 2017 Mar; 98(3):357–363.
- 729 **Killip MJ**, Smith M, Jackson D, Randall RE. Activation of the Interferon Induction Cascade by Influenza A Viruses Requires Viral RNA Synthesis and Nuclear Export. *Journal of Virology*. 2014 Mar; 88(8):3942–3952.
- 731 **Killip MJ**, Fodor E, Randall RE. Influenza virus activation of the interferon system. *Virus research*. 2015 Feb; .
- 732 **Kim KH**, Jeong JY, Surh YJ, Kim KW. Expression of stress-response ATF3 is mediated by Nrf2 in astrocytes. *Nucleic Acids Research*. 2009 Oct; 38(1):48–59.
- 734 **Klein AM**, Mazutis L, Akartuna I, Tallapragada N, Veres A, Li V, Peshkin L, Weitz DA, Kirschner MW. Droplet Barcoding for Single-Cell Transcriptomics Applied to Embryonic Stem Cells. *Cell*. 2015 May; 161(5):1187–1201.
- 736 **Lauring AS**, Andino R. Quasispecies Theory and the Behavior of RNA Viruses. *PLoS Pathogens*. 2010 Jul; 6(7):e1001005–8.
- 738 **Le Meur N**, Hahne F, Ellis B, Haaland P. FlowCore: data structures package for flow cytometry data. *Bioconductor Project*; 2007.
- 740 **Lee D**, Ryu KY. Effect of cellular ubiquitin levels on the regulation of oxidative stress response and proteasome function via Nrf1. *Biochemical and Biophysical Research Communications*. 2017 Apr; 485(2):234–240.
- 742 **Leonard AS**, Weissman D, Greenbaum B, Ghedin E, Koelle K. Transmission Bottleneck Size Estimation from Pathogen Deep-Sequencing Data, with an Application to Human Influenza A Virus. *Journal of Virology*. 2017; p. JVI-00171–17.
- 745 **Lin X**, Wang R, Zou W, Sun X, Liu X, Zhao L, Wang S, Jin M. The Influenza Virus H5N1 Infection Can Induce ROS Production for Viral Replication and Host Cell Death in A549 Cells Modulated by Human Cu/Zn Superoxide Dismutase (SOD1) Overexpression. *Viruses*. 2016 Jan; 8(1):13–16.
- 748 **Lopez CB**. Defective Viral Genomes: Critical Danger Signals of Viral Infections. *Journal of Virology*. 2014 Jul; 88(16):8720–8723.
- 750 **Van der Maaten L**, Hinton G. Visualizing data using t-SNE. *Journal of Machine Learning Research*. 2008; 9:2579–2605.
- 752 **MacLeod AK**, Acosta-Jimenez L, Coates PJ, McMahon M, Carey FA, Honda T, Henderson CJ, Wolf CR. Aldo-keto reductases are biomarkers of NRF2 activity and are co-ordinately overexpressed in non-small cell lung cancer. *British journal of cancer*. 2016 Dec; 115(12):1530–1539.
- 755 **Macosko EZ**, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, Tirosh I, Bialas AR, Kamitaki N, Martersteck EM, et al. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell*. 2015; 161(5):1202–1214.
- 758 **von Magnus P**. Incomplete forms of influenza virus. *Advances in Virus Research*. 1954; 2:59–79.

- 759 **Marsh GA**, Hatami R, Palese P. Specific residues of the influenza A virus hemagglutinin viral RNA are important
760 for efficient packaging into budding virions. *Journal of Virology*. 2007; 81(18):9727–9736.
- 761 **McCrone JT**, Woods RJ, Martin ET, Malosh RE, Monto AS, Lauring AS. Stochastic processes dominate the within
762 and between host evolution of influenza virus. *bioRxiv*. 2017; p. 176362.
- 763 **Miura T**, Taketomi A, Nishinaka T, Terada T. Regulation of human carbonyl reductase 1 (CBR1, SDR21C1) gene
764 by transcription factor Nrf2. *Chemico-Biological Interactions*. 2013 Feb; 202(1-3):126–135.
- 765 **Murray J**, Taylor SW, Zhang B, Ghosh SS, Capaldi RA. Oxidative Damage to Mitochondrial Complex I Due to
766 Peroxynitrite. *Journal of Biological Chemistry*. 2003 Sep; 278(39):37223–37230.
- 767 **Neumann G**, Watanabe T, Ito H, Watanabe S, Goto H, Gao P, Hughes M, Perez DR, Donis R, Hoffmann E, Hobom
768 G, Kawaoka Y. Generation of influenza A viruses entirely from cloned cDNAs. *Proceedings of the National
769 Academy of Sciences of the United States of America*. 1999 Aug; 96(16):9345–9350.
- 770 **Nobusawa E**, Sato K. Comparison of the mutation rates of human influenza A and B viruses. *Journal of Virology*.
771 2006; 80(7):3675–3678.
- 772 **Noda T**, Sagara H, Yen A, Takada A, Kida H, Cheng RH, Kawaoka Y. Architecture of ribonucleoprotein complexes
773 in influenza A virus particles. *Nature*. 2006; 439(7075):490.
- 774 **Padovan-Merhar O**, Nair GP, Biaesch A, Mayer A, Scarfone S, Foley SW, Wu AR, Churchman LS, Singh A, Raj
775 A. Single mammalian cells compensate for differences in cellular volume and DNA copy number through
776 independent global transcriptional mechanisms. *Molecular Cell*. 2015; 58(2):339.
- 777 **Parvin JD**, Moscona A, Pan W, Leider J, Palese P. Measurement of the mutation rates of animal viruses: influenza
778 A virus and poliovirus type 1. *Journal of Virology*. 1986; 59(2):377–383.
- 779 **Pauly MD**, Procario MC, Lauring AS. A novel twelve class fluctuation test reveals higher than expected mutation
780 rates for influenza A viruses. *eLife*. 2017; 6:e26437.
- 781 **Perez JT**, Varble A, Sachidanandam R, Zlatev I, Manoharan M, García-Sastre A, et al. Influenza A virus-generated
782 small RNAs regulate the switch from transcription to replication. *Proceedings of the National Academy of
783 Sciences*. 2010; 107(25):11525–11530.
- 784 **Perez-Cidoncha M**, Killip MJ, Oliveros JC, Asensio VJ, Fernandez Y, Bengoechea JA, Randall RE, Ortín J. An
785 Unbiased Genetic Screen Reveals the Polygenic Nature of the Influenza Virus Anti-Interferon Response.
786 *Journal of Virology*. 2014 Apr; 88(9):4632–4646.
- 787 **Peuchant E**, Bats ML, Moranvillier I, Lepoilvre M, Guittot J, Wendum D, Lacombe ML, Moreau-Gaudry F, Boissan
788 M, Dabernat S. Metastasis suppressor NM23 limits oxidative stress in mammals by preventing activation of
789 stress-activated protein kinases/JNKs through its nucleoside diphosphate kinase activity. *The FASEB Journal*.
790 2017 Mar; 31(4):1531–1546.
- 791 **Poon LL**, Song T, Rosenfeld R, Lin X, Rogers MB, Zhou B, Sebra R, Halpin RA, Guan Y, Twaddle A, DePasse JV,
792 Stockwell TB, Wentworth DE, Holmes EC, Greenbaum B, Peiris JSM, Cowling BJ, Ghedin E. Quantifying influenza
793 virus diversity and transmission in humans. *Nature Genetics*. 2016; 48(2):195–200.
- 794 **Qiu X**, Mao Q, Tang Y, Wang L, Chawla R, Pliner HA, Trapnell C. Reversed graph embedding resolves complex
795 single-cell trajectories. *Nature Methods*. 2017 Aug; p. 1–10.
- 796 **Raj A**, Peskin CS, Tranchina D, Vargas DY, Tyagi S. Stochastic mRNA synthesis in mammalian cells. *PLoS Biology*.
797 2006; 4(10):e309.
- 798 **Rand U**, Rinas M, Schwerk J, Nöhren G, Linnes M, Kröger A, Flossdorf M, Kullai KK, Hauser H, Höfer T, Köster M.
799 Multi-layered stochasticity and paracrine signal propagation shape the type-I interferon response. *Molecular
800 Systems Biology*. 2012 Jan; 8(1):584.
- 801 **Reed LJ**, Muench H. A simple method of estimating fifty percent endpoints. *American journal of epidemiology*.
802 1938; 27(3):493–497.
- 803 **Saira K**, Lin X, DePasse JV, Halpin R, Twaddle A, Stockwell T, Angus B, Cozzi-Lepri A, Delfino M, Dugan V, Dwyer
804 DE, Freiberg M, Horban A, Losso M, Lynfield R, Wentworth DN, Holmes EC, Davey R, Wentworth DE, Ghedin E.
805 Sequence analysis of in vivo defective interfering-like RNA of influenza A H1N1 pandemic virus. *Journal of
806 Virology*. 2013; 87(14):8064–8074.

- 807 **Schulte MB**, Andino R. Single-Cell Analysis Uncovers Extensive Biological Noise in Poliovirus Replication. *Journal
808 of Virology*. 2014 May; 88(11):6205–6212.
- 809 **Sgarbanti R**, Amatore D, Celestino I, Marcocci ME, Fraternale A, Ciriolo MR, Magnani M, Saladino R, Garaci E,
810 Palamara AT, Nencioni L. Intracellular redox state as target for anti-influenza therapy: are antioxidants always
811 effective? *Current topics in medicinal chemistry*. 2014; 14(22):2529–2541.
- 812 **Shalek AK**, Satija R, Adiconis X, Gertner RS, Gaublomme JT, Raychowdhury R, Schwartz S, Yosef N, Malboeuf C,
813 Lu D, et al. Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature*.
814 2013; 498(7453):236–241.
- 815 **Shalek AK**, Satija R, Shuga J, Trombetta JJ, Gennert D, Lu D, Chen P, Gertner RS, Gaublomme JT, Yosef N, et al.
816 Single cell RNA Seq reveals dynamic paracrine control of cellular variation. *Nature*. 2014; 510(7505):363.
- 817 **Shapiro GI**, Gurney T, Krug RM. Influenza virus gene expression: control mechanisms at early and late times of
818 infection and nuclear-cytoplasmic transport of virus-specific RNAs. *Journal of Virology*. 1987 Mar; 61(3):764–
819 773.
- 820 **Suárez P**, Valcarcel J, Ortín J. Heterogeneity of the mutation rates of influenza A viruses: isolation of mutator
821 mutants. *Journal of Virology*. 1992; 66(4):2491–2494.
- 822 **Suárez-López P**, Ortín J. An estimation of the nucleotide substitution rate at defined positions in the influenza
823 virus haemagglutinin gene. *Journal of general virology*. 1994; 75(2):389–393.
- 824 **Subramanian A**, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR,
825 Lander ES, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide
826 expression profiles. *Proceedings of the National Academy of Sciences*. 2005; 102(43):15545–15550.
- 827 **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
828 Interferon Signalling Pathways Occurs in Lung Epithelial Cells Infected with Low Pathogenic Avian Influenza
829 Viruses. *PLoS ONE*. 2012; 7(3).
- 830 **Tapia K**, Kim Wk, Sun Y, Mercado-López X, Dunay E, Wise M, Adu M, López CB. Defective viral genomes arising
831 in vivo provide critical danger signals for the triggering of lung antiviral immunity. *PLoS Pathogens*. 2013;
832 9(10):e1003703.
- 833 **Trapnell C**, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, Lennon NJ, Livak KJ, Mikkelsen TS, Rinn JL. The
834 dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nature
835 Biotechnology*. 2014 Mar; 32(4):381–386.
- 836 **Varble A**, Albrecht RA, Backes S, Crumiller M, Bouvier NM, Sachs D, García-Sastre A, et al. Influenza A virus
837 transmission bottlenecks are defined by infection route and recipient host. *Cell host & microbe*. 2014;
838 16(5):691–700.
- 839 **Väremo L**, Nielsen J, Nookaew I. Enriching the gene set analysis of genome-wide data by incorporating
840 directionality of gene expression and combining statistical hypotheses and methods. *Nucleic Acids Research*.
841 2013; 41(8):4378–4391.
- 842 **Vreede FT**, Jung TE, Brownlee GG. Model Suggesting that Replication of Influenza Virus Is Regulated by
843 Stabilization of Replicative Intermediates. *Journal of Virology*. 2004 Aug; 78(17):9568–9572.
- 844 **van der Walt S**, Colbert SC, Varoquaux G. The NumPy Array: A Structure for Efficient Numerical Computation.
845 *Computing in Science & Engineering*. 2017 Aug; 13(2):22–30.
- 846 **Watanabe T**, Watanabe S, Kawaoka Y. Cellular networks involved in the influenza virus life cycle. *Cell Host &
847 Microbe*. 2010; 7(6):427–439.
- 848 **Xue J**, Chambers BS, Hensley SE, López CB. Propagation and Characterization of Influenza Virus Stocks That
849 Lack High Levels of Defective Viral Genomes and Hemagglutinin Mutations. *Frontiers in microbiology*. 2016
850 Mar; 7(Pt 10):e1004924–15.
- 851 **Xue KS**, Stevens-Ayers T, Campbell AP, Englund JA, Pergam SA, Boeckh M, Bloom JD. Parallel evolution of
852 influenza across multiple spatiotemporal scales. *eLife*. 2017 Jun; 6:e26875.
- 853 **Zheng GX**, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, Ziraldo SB, Wheeler TD, McDermott GP, Zhu J, et al.
854 Massively parallel digital transcriptional profiling of single cells. *Nature communications*. 2017; 8:14049.
- 855 **Zhu Y**, Yongky A, Yin J. Growth of an RNA virus in single cells reveals a broad fitness distribution. *Virology*. 2009;
856 385(1):39–46.

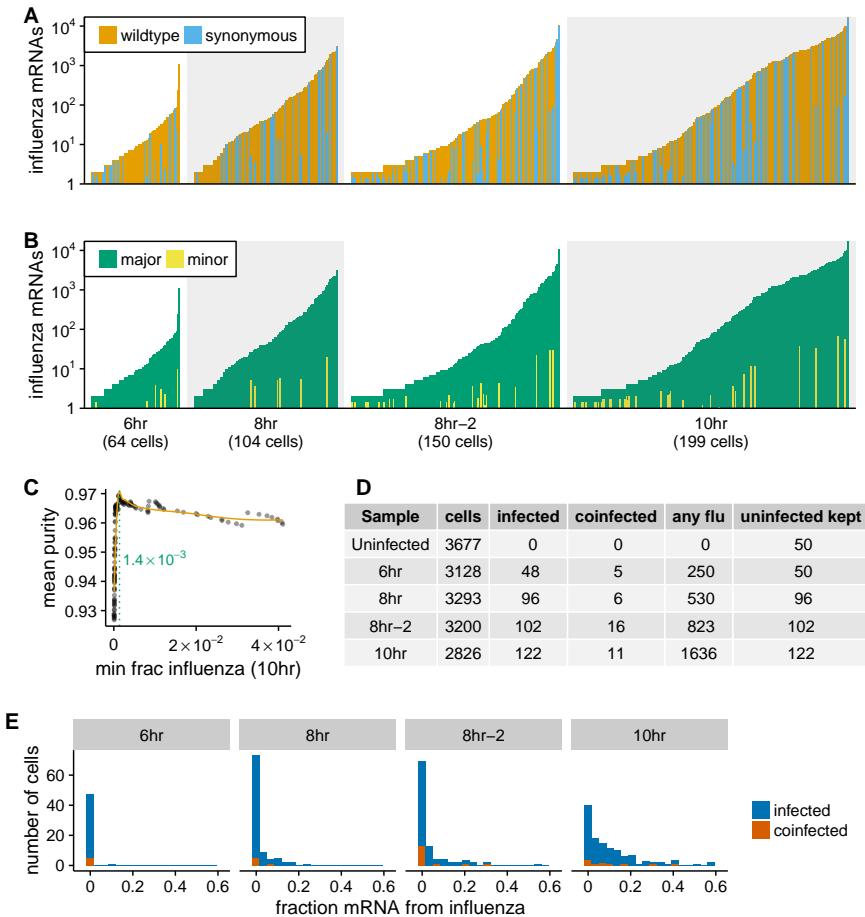


Figure 4. Synonymous barcodes on the viral mRNAs distinguish true infections from cells that contain viral mRNAs derived from leakage of lysed cells. **(A)** Cells with at least two viral mRNAs for which the barcode could be called, arranged in order of increasing influenza transcript counts. Bar heights 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 cutoff 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 3 shows these same data in a cumulative fraction plots and calculates Gini coefficients to quantify the heterogeneity in viral mRNA load.

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. Cumulative distributions of viral mRNA per cell and Gini coefficients.

Figure 4–Figure supplement 4. Synchronization of infection does not greatly affect heterogeneity.

Figure 4–Figure supplement 5. Effects of infectious dose or coinfection state.

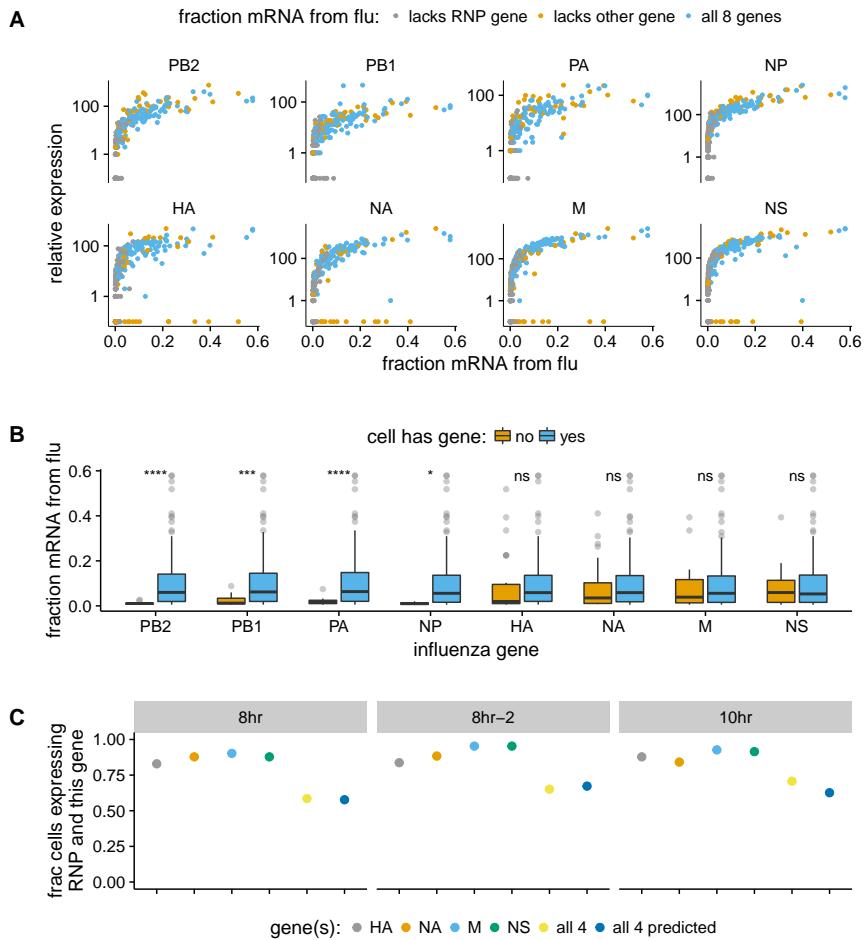


Figure 5. The absence of viral genes explains some of the variability in the amount of viral mRNA per cell. **(A)** Fraction of mRNA in each infected cell derived from virus as a function of the total fraction of each viral gene mRNA in that each infected cell derived from virus, taken over all time points. Cells with high viral burden always express all RNP genes, but some cells with high viral burden lack each of the other genes. Plots for individual samples are in Figure 5–Figure Supplement 2, and a plot that excludes known coinfecting cells is in Figure 5–Figure Supplement 3. **(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. The trends are similar if we look only at the 10-hour sample (Figure 5–Figure Supplement 4) or exclude known co-infected cells (Figure 5–Figure Supplement 5). **(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 (A), but excludes coinfecting cells with mixed viral barcodes.

Figure 5–Figure supplement 4. Like panel (B) but for the 10-hr sample only.

Figure 5–Figure supplement 5. Like panel (B) but excludes coinfecting cells with mixed viral barcodes.

Figure 5–source data 1. The numerical data for panel (C) are in p_missing_genes.csv.

Figure 5–source data 2. Simulation with a simple model for the expected heterogeneity due to Poisson co-infection and presence / absence of the full RNP is in simple_Poisson_model.html.

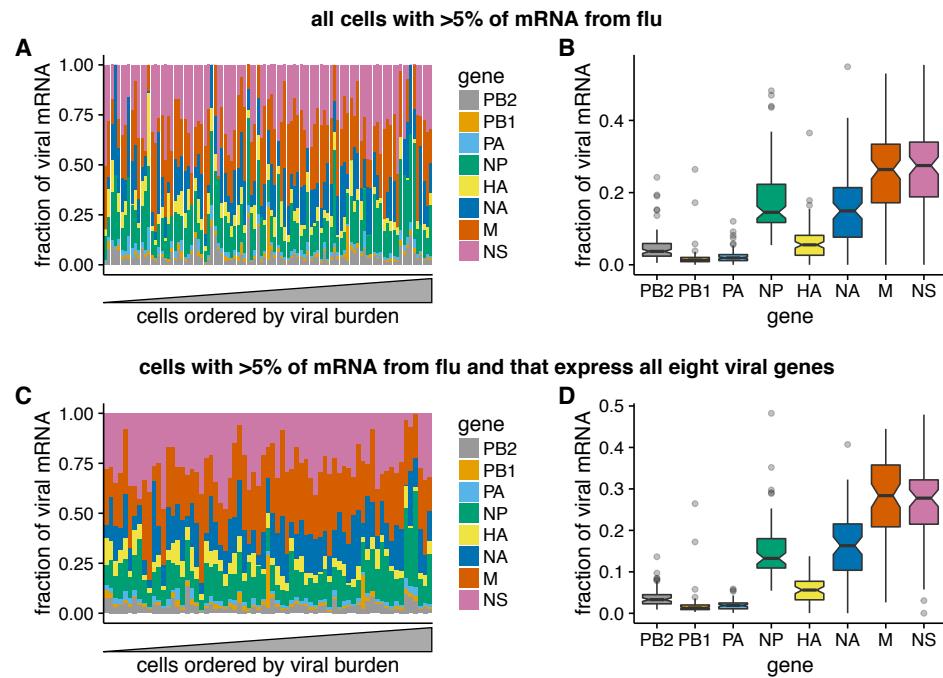


Figure 6. Relative expression of influenza virus genes in highly infected cells (>5% of total mRNA from virus). **(A)** The fraction of viral mRNA from each viral gene for each cell. **(B)** Box plots showing the distribution of the fraction of viral mRNA per cell from each viral gene. The black lines 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_fiu_expr_all.csv`.

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

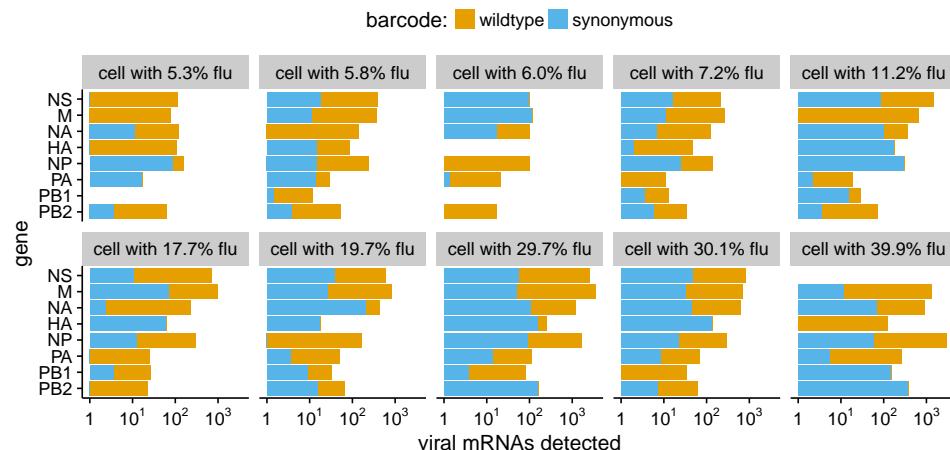


Figure 7. The abundance of each viral transcript in cells that are co-infected with the two viral variants and have >5% of their mRNA derived from virus. The bars show the logarithms of the numbers of each viral mRNA detected, and are colored in linear proportion to the fraction of that mRNAs derived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

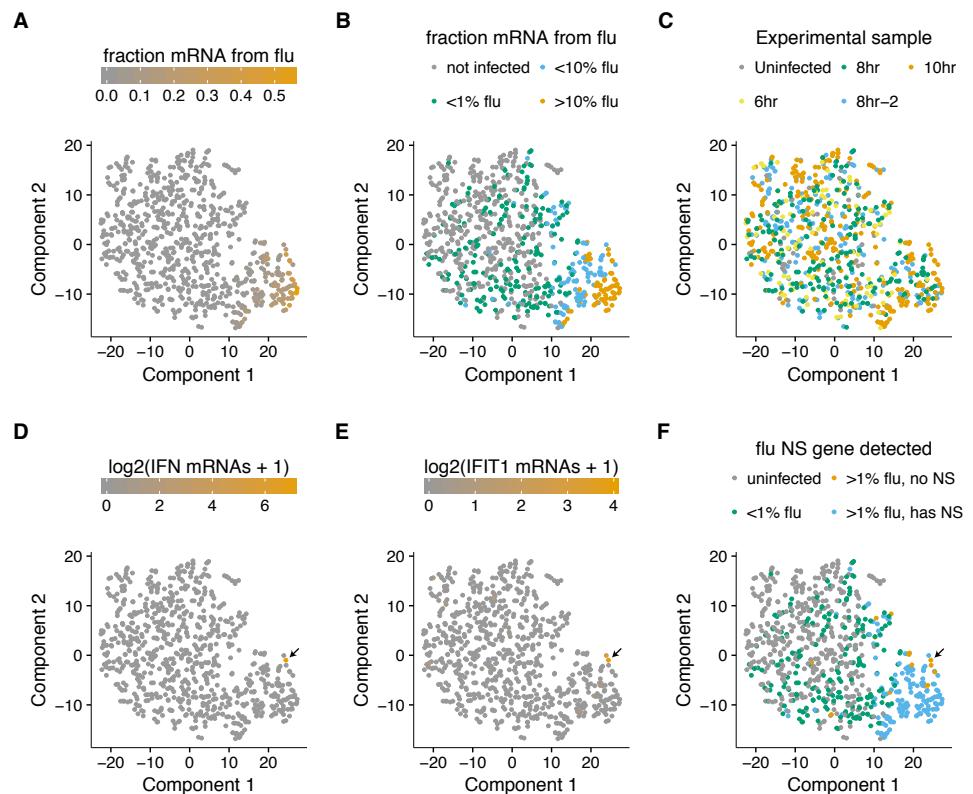


Figure 8. A t-SNE plot created by semi-supervised clustering using genes that co-vary with viral infection status. Each point is a single cell, and each panel shows an identical layout but colors the cells according to a different property. **(A), (B)** Cells colored by the fraction of their mRNA derived 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.

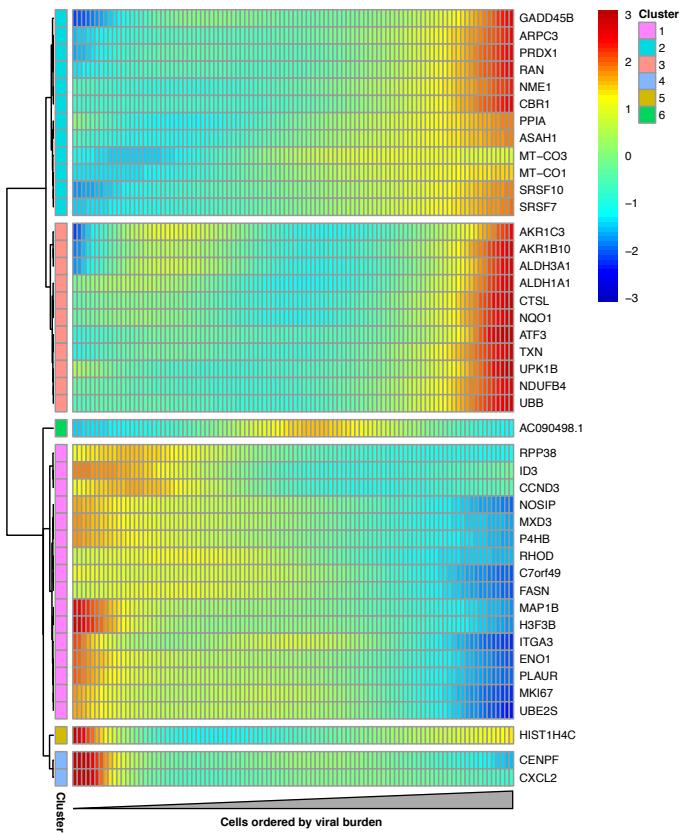


Figure 9. Cellular genes that co-vary in expression with the amount of viral mRNA in cells expressing all eight viral genes. The columns are cells, ordered from left to right by the fraction of mRNA derived from virus. Each row is a gene that is differentially expressed as a function of the fraction 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 are in `p_sig_cellular_genes.csv`.

Figure 9-source data 2. A gene-set analysis for pathways associated with the amount of viral mRNA is in `p_pathway_enrichment.csv`.

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

Figure 9-Figure supplement 2. Pre-treating to reduce oxidative stress decreases the fraction of infected cells expressing detectable viral protein.

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. [To just run the Monocle analysis in `monocle_analysis.ipynb` on a pre-generated cell-gene matrix, unpack Supplementary file 2 into `./results/cellgenecounts/`.](#)

Supplementary file 2. The annotated cell-gene matrix in Matrix Market Format. [This is the matrix generated in `./results/cellgenecounts/` by running the CellRanger analysis in `align_and_annotate.ipynb` in Supplementary file 1.](#)

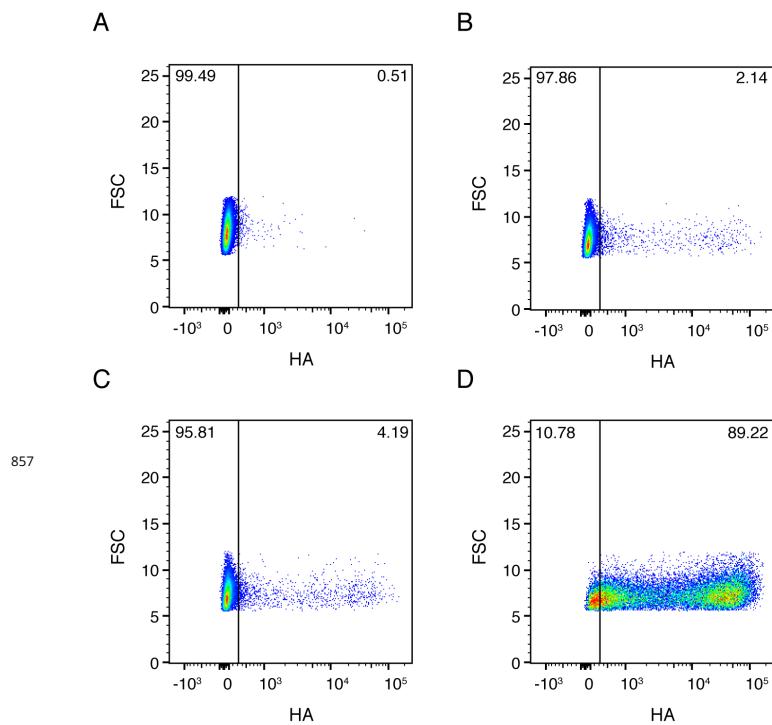


Figure 2–Figure supplement 1. Full flow cytometry data for Figure 2B. A549 cells were infected at an MOI of 0.1 as calculated by TCID₅₀ on MDCK-SIAT1 cells. **(A)** Uninfected gating control. **(B)** Cells infected with the wild-type virus stock used in our experiments. **(C)** Cells infected with synonymously 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.

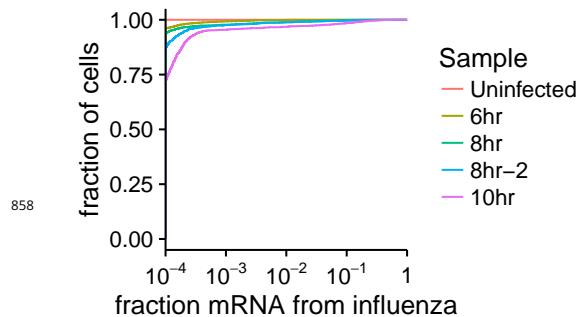


Figure 3–Figure supplement 1. For each sample, this plot shows the fraction of all cells that derive at least the indicated fraction of their mRNA from influenza virus.

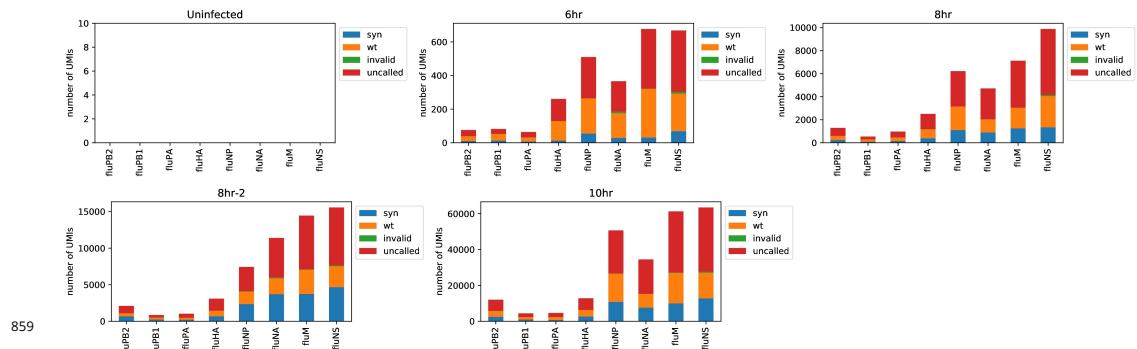


Figure 4-Figure supplement 1. The number of viral barcodes called for each sample and gene segment. Viral transcripts are classified as *syn* if they mapped to a synonymously barcoded influenza transcript, *wt* if they mapped to a wild-type influenza transcript, *invalid* if multiple reads for the same 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.

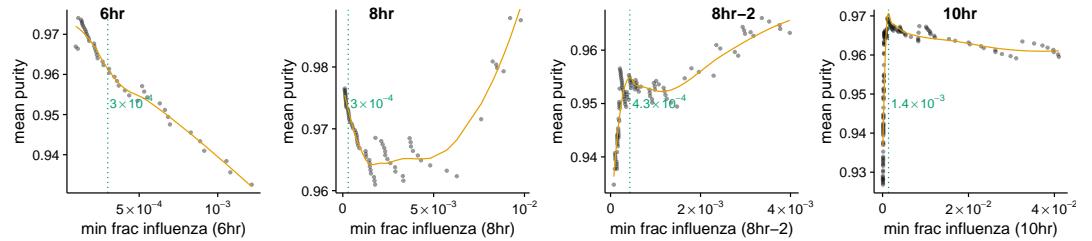


Figure 4-Figure supplement 2. Cell lysis can lead cells to the spurious association of small amounts of extraneous mRNA with individual cells. We wanted to avoid classifying as infected cells that had simply acquired such lysis-derived viral mRNA. The amount of lysis-derived viral mRNA will vary among 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 2×10^{-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).

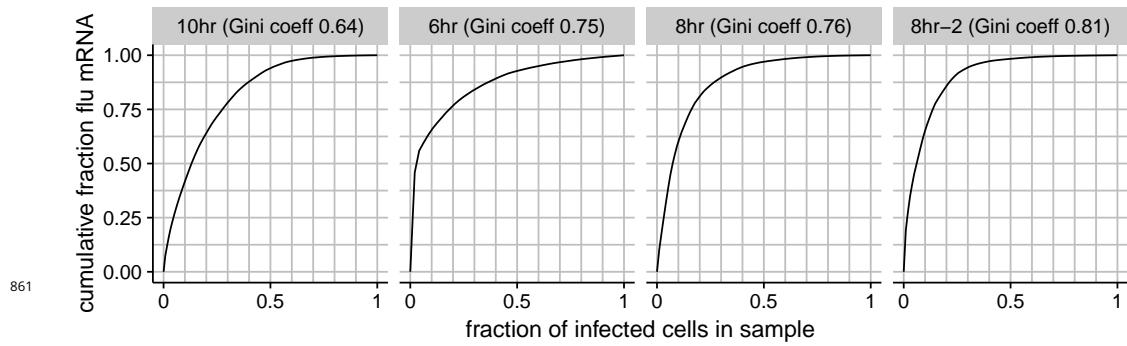


Figure 4-Figure supplement 3. The total fraction of all viral mRNA among infected cells that is attributable to a given fraction of these cells. For instance, the plot for the 8hrs sample shows that ~50% of all viral mRNA is derived from ~8% of the infected cells. [The facet titles above each plot also give the Gini coefficient \(Gini, 1921\) that calculates the heterogeneity in the distribution of viral mRNA among infected cells. Gini coefficients of 0 indicate a perfectly even distribution across cells, and Gini coefficients of 1 indicate a maximally skewed distribution.](#)

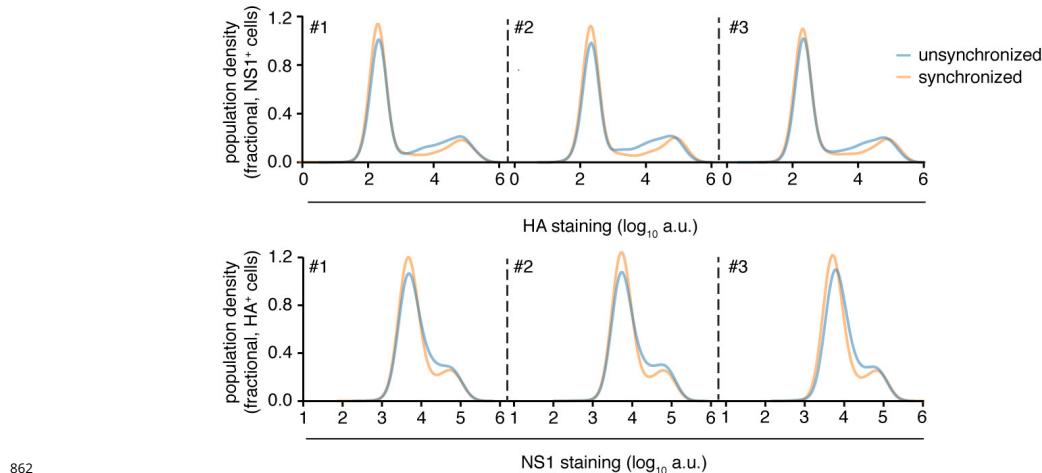


Figure 4-Figure supplement 4. [Flow cytometry analysis of expression of viral proteins in cells infected at an MOI of 0.1 \(unsynchronized\) or 0.2 \(synchronized\) as calculated by TCID50 on MDCK-SIAT1 cells. The higher MOI for synchronized infections was to attempt to account for loss of virus in washing steps. Infections were synchronized by pre-adsorbing virus at 4°C for 1h prior to initiation of infection by shifting temperature to 37°C using pre-warmed media. Cells were concurrently stained for HA and NS1 proteins at 10 hours after initiation of infection, and then analyzed by flow cytometry. The level of HA protein was quantified in cells that were identified as infected based on being positive for NS1 protein \(top\), and the level of NS1 protein was quantified in cells that were identified as infected based on being positive for HA protein \(bottom\). Synchronization resulted in a small decrease in variability in the levels of each viral protein, particularly in cells with intermediate levels. But the effects were small compared to the overall variability in the protein levels, indicating timing of infection makes only a small contribution to the observed heterogeneity. Data are shown for three independent replicates.](#)

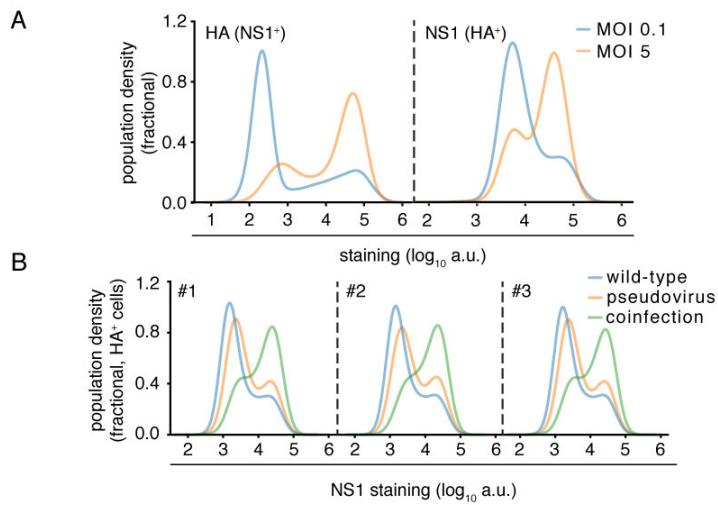


Figure 4-Figure supplement 5. (A) Flow cytometry analysis of expression of viral proteins in cells infected at high (MOI 5 as calculated by TCID₅₀ on MDCK-SIAT1 cells) or low (MOI 0.1) initial infectious dose. Cells were concurrently stained for HA and NS1 proteins at 10 hours after initiation of infection, and then analyzed by flow cytometry. The level of HA protein was quantified in cells that were identified as infected based on being positive for NS1 protein (left), and the level of NS1 protein was quantified in cells that were identified as infected based on being positive for HA protein (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 pseudovirus 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 for these proteins and GFP. Cells could be annotated as infected by virions of the same type (wild-type infection indicated by presence of HA, or pseudovirus infection indicated by the presence of GFP) or both types of virions (indicated by presence of HA and GFP). 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. Data are shown for three independent replicates.

863

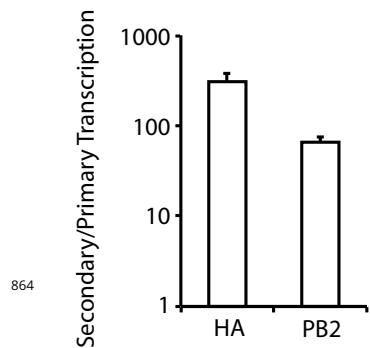


Figure 5–Figure supplement 1. A549 cells were infected at an MOI of 0.2 as calculated on MDCK-SIAT1 cells in either the presence or absence of the protein-translation inhibitor cyclohexamide, and viral mRNA was quantified at 8 hours post-infection by qPCR. The cyclohexamide prevents translation of new 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.

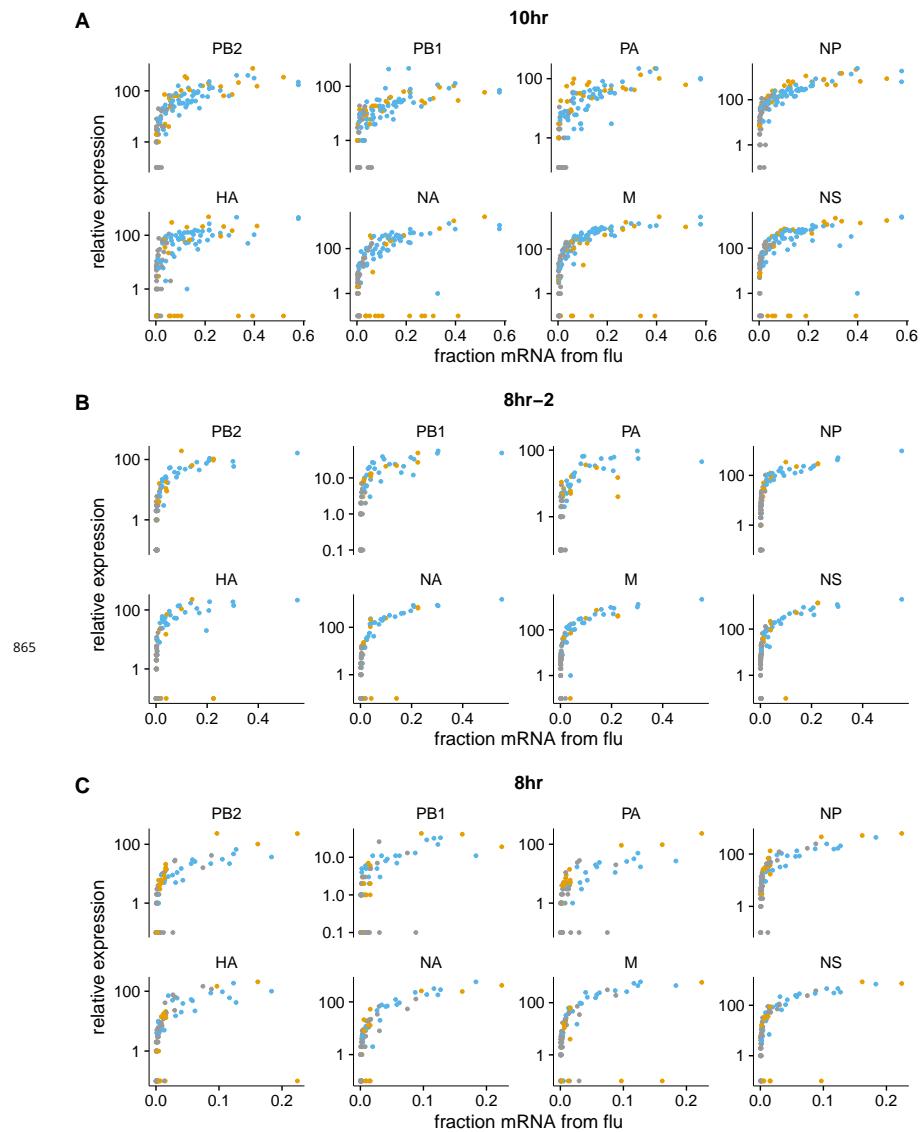


Figure 5–Figure supplement 2. The normalized expression of each viral gene as a function of the fraction of total mRNA derived from virus, shown for the 10-hour and 8-hour samples individually (the other samples had too few infected cells for this analysis to be useful). Points are colored as in **Figure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

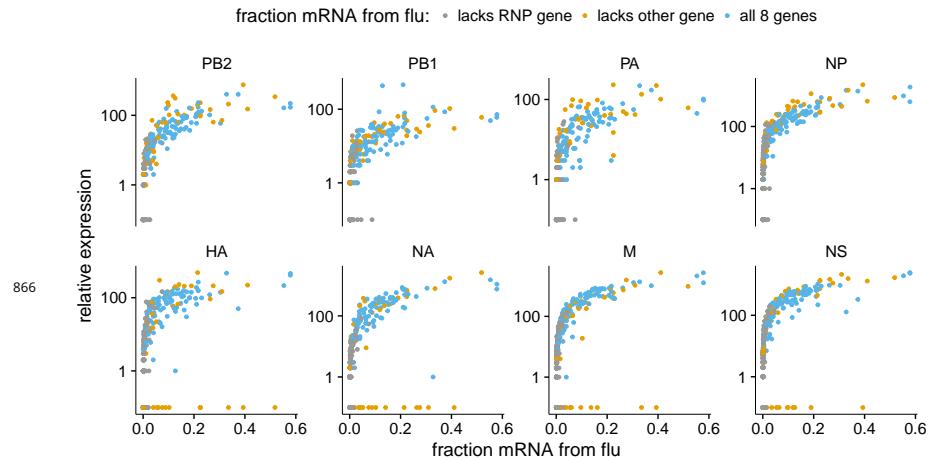


Figure 5-Figure supplement 3. The normalized expression of each viral gene as a function of the fraction of total mRNA derived from virus, excluding cells that were annotated as coinfected based on the presence of both viral barcodes.

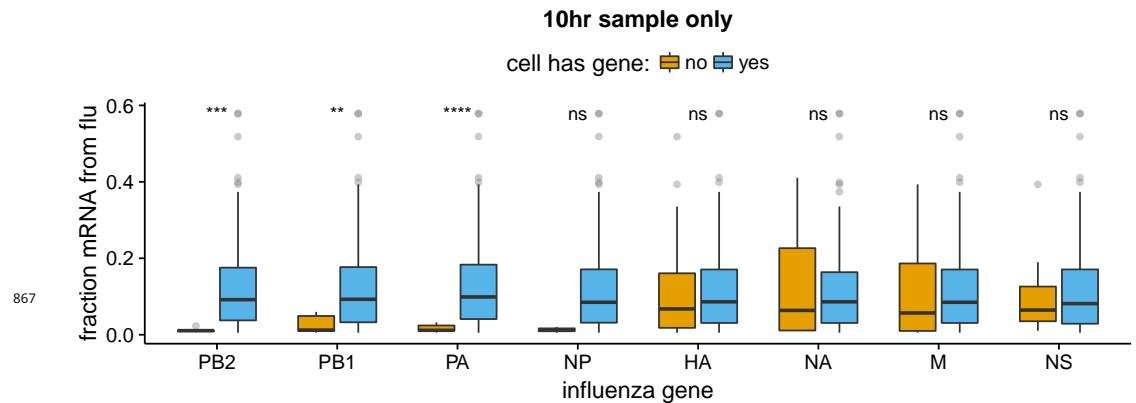


Figure 5-Figure supplement 4. The absence of viral RNP genes but *not* non-RNP genes remains significantly associated with reduced viral burden when we examine only the 10-hr sample, which is the single time point with the most data points. The difference for NP is no longer statistically significant 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.

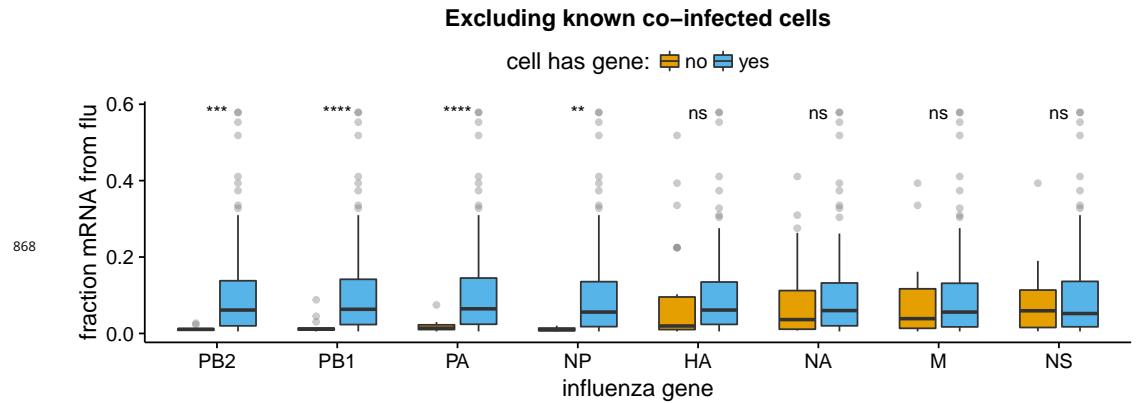


Figure 5-Figure supplement 5. All findings in Figure 5B remain unchanged if we exclude cells called as coinfecting based on the presence of mixed viral barcodes.

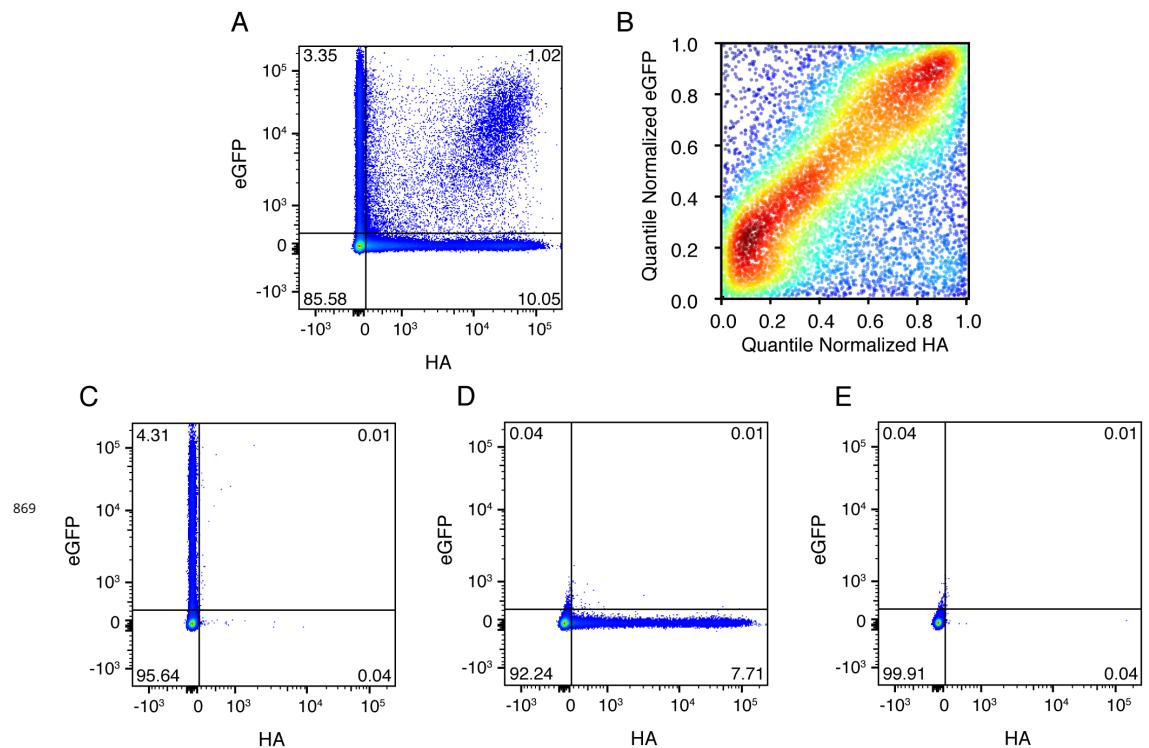


Figure 7-Figure supplement 1. **(A)** Cells were co-infected with a mix of wild-type virus and virus in which the HA gene was replaced by GFP flanked by the terminal regions of the HA gene segment. At 10 hours post-infection, cells were analyzed by flow cytometry for HA and eGFP expression. **(B)** The 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, 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*).

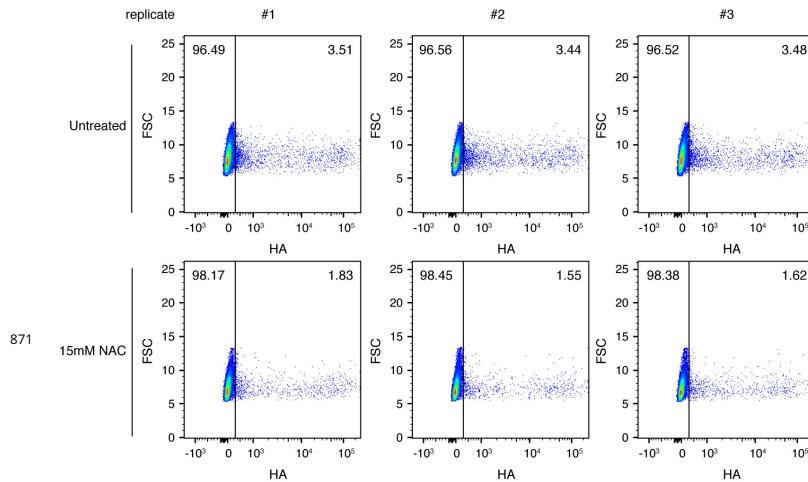


Figure 9–Figure supplement 2. Cells were treated as indicated with 15 mM N-acetylcysteine (NAC), an antioxidant, and infected at an MOI of 0.1 as calculated by TCID50 on MDCK-SIAT1 cells. At 10 hours post-infection, cells were analyzed by flow cytometry for HA expression. The percentage of HA-positive cells is indicated on the flow cytometry plots. Data are shown for three independent replicates.