Single-cell virus sequencing of influenza infections that trigger innate immunity

Alistair B. Russell,[a] **Elizaveta Elshina,**[b] **Jacob R. Kowalsky,**[a] **Aartjan J. W. te Velthuis,**[b] **Jesse D. Bloom**[a,c,d,*]

[a] Basic Sciences and Computational Biology, Fred Hutchinson Cancer Research Center, Seattle, United States
[b] Division of Virology, Department of Pathology, University of Cambridge, Cambridge, United Kingdom
[c] Department of Genome Sciences, University of Washington, Seattle, United States
[d] Howard Hughes Medical Institute, Seattle, United States

ABSTRACT   Influenza-infected cells vary widely in their expression of viral genes, and only occasionally activate innate immunity. Here we develop a new method to assess how the genetic variation in viral populations contributes to this heterogeneity. We do this by determining the transcriptome and full-length sequences of all viral genes in single cells infected with a nominally "pure" stock of influenza virus. Most cells are infected by virions with defects, some of which increase the frequency of innate-immune activation. These immunostimulatory defects are diverse, and include mutations that perturb the function of the viral polymerase protein PB1, large internal deletions in viral genes, and failure to express the virus's interferon antagonist NS1. However, immune activation remains stochastic in cells infected by virions with these defects, and occasionally is triggered even by virions that express unmutated copies of all genes. Our work shows that the diverse spectrum of defects in influenza virus populations contributes to—but does not completely explain—the heterogeneity in viral gene expression and immune activation in single infected cells.

IMPORTANCE   Because influenza virus has a high mutation rate, many cells are infected by mutated virions. But so far, it has been impossible to fully characterize the sequence of the virion infecting any given cell, since conventional techniques like flow cytometry and single-cell RNA-seq only detect if a protein or transcript is present—not its sequence. Here we develop a new approach that uses long-read PacBio sequencing to determine the sequences of virions infecting single cells. We show that viral genetic variation explains some but not all of the cell-to-cell variability in viral gene expression and innate-immune induction. Overall, our study provides the first complete picture of how viral mutations affect the course of infection in single cells.

KEYWORDS: influenza virus; interferon; PacBio; single-cell RNAseq; 10X Chromium; NS1; PB1; defective virus; heterogeneity

INTRODUCTION

Infection with an acute virus such as influenza initiates a race between the virus and immune system. As the virus spreads, some cells detect infection and begin producing interferon (IFN). This IFN directs expression of anti-viral interferon-stimulated genes (ISGs) in the infected cell and its neighbors via autocrine and paracrine signaling, as well as helping launch a broader immune response (1, 2). If innate immunity is activated sufficiently rapidly, it can reduce viral replication and disease (3, 4, 5, 6, 7)—although excessive immune responses later in infection can actually be associated with immunopathology and severe disease (8, 9).

Unfortunately for the host, influenza initially only rarely triggers IFN production by infected cells ([10], [11]). This rareness of IFN induction is just one form of the extreme cell-to-cell heterogeneity that characterizes infection: cells also vary widely in their production of viral mRNA, proteins, and progeny virions ([12], [13], [14], [15], [16]). Because viral growth and the IFN response both amplify themselves, early variation in the initiation of these events could have significant downstream consequences for the race between virus and immune system—especially since natural human infections are typically founded by just a few virions entering a few cells ([17], [18], [19]).

It is unclear why only some infected cells trigger innate-immune responses. Two possible contributors are pure stochasticity and pre-existing variation in cellular state. For instance, only some cells induce IFN even upon treatment with synthetic innate-immune ligands ([20], [21], [22]), and the frequency of IFN induction may depend on a cell's pre-existing chromatin state ([23]). But for influenza, a third possible contributor also looms large: viral genetic diversity. The virus has evolved mechanisms to avoid IFN induction, including expressing proteins that interfere with innate-immune induction ([24], [25], [26], [27], [28]) and sequestering immunogenic viral RNA ([29]). However, because influenza has a high mutation rate ([30], [31], [32], [33], [34]), individual virions often have genetic defects that could impair these immune-evasion strategies. Indeed, many studies have identified mutations that increase IFN induction when engineered into a viral population ([11], [35], [36], [37]), and viral stocks that are rich in internal deletions in the polymerase genes induce more IFN ([16], [38], [39], [40], [41], [42]).

However, existing techniques are inadequate to determine how viral genetic diversity contributes to cell-to-cell heterogeneity during infection. Flow cytometry and fluorescent reporters only measure protein levels ([14], [43], [44]), and current single-cell transcriptomic techniques measure abundance of transcripts and provide only fragmentary information on their sequences ([12], [13], [16], [45], [46], [47], [48]). None of these techniques reliably reveal if the virion infecting a specific cell has some idiosyncratic mutation.

Here we develop a new approach to measure both the full transcriptome and sequences of all viral genes in single influenza-infected cells. To do this, we perform both standard Illumina-based transcriptomics and full-length PacBio sequencing of viral genes from single cells. Two-thirds of cells are infected by virions that have a mutation or defect in gene expression. We identify several types of viral defects that increase IFN induction, but also show that viral diversity is insufficient to fully explain cell-to-cell heterogeneity during influenza infection.

RESULTS

A system to identify and enrich rare IFN+ cells. Influenza virus only rarely triggers IFN expression in infected cells ([11], [12], [10])—a fact that poses a challenge for the study IFN induction in single cells. Therefore, we developed a method to identify and enrich rare IFN+ cells by creating A549 cells that carried IFN reporters consisting of a type I (*IFNB1*) or type III (*IFNL1*) promoter driving expression of a cell-surface protein (LNGFRΔC ([49], [50])) followed by a fluorescent protein (Fig. 1A). Cells that activate the IFN reporters can be enriched by magnetic-activated cell sorting (MACS) or identified by flow cytometry. The reporters were efficiently activated by infection with a strain of Sendai virus ([51]) that potently induces IFN (Fig. S1A), and activation of the type I and type III IFN reporters was highly correlated in our cells (Fig. S1B; further validated by the single-cell transcriptomics below). Therefore, for the rest of this paper, we use "IFN expression" to refer to combined expression of type I and III IFNs.

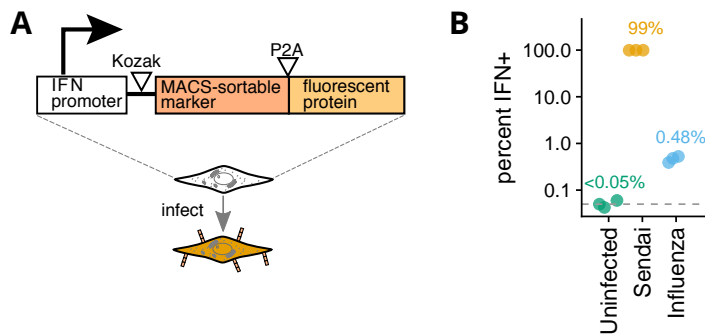We generated a stock of A/WSN/1933 (H1N1) influenza (hereafter referred to as

FIG 1 Reporter cells to identify and enrich infections that activate IFN expression. (A) The reporter consists of an IFN promoter that drives expression of a cell-surface protein amenable to MACS and a fluorescent protein. We created reporters with type I and type III IFN promoters (File S1). In A549 cells, the reporters were efficiently activated by an IFN-inducing strain of Sendai virus (Fig. S1A). (B) Frequency of IFN induction upon infection with the influenza virus stock used in the single-cell studies in this paper, as quantified using the type III IFN reporter (see Fig. S2 for full flow cytometry data). The plot also shows uninfected cells, and cells infected with Sendai virus. The limit of detection of 0.05% is indicated with a dashed line, and numbers show the median of three measurements.

"WSN") directly from reverse-genetics plasmids (52), and passaged this stock at low multiplicity of infection (MOI). This process ensures that the viral stock that is relatively "pure", with only low levels of the large internal deletions and other defects that arise in stocks passaged at high MOI (53). As described in the next subsection, our stock actually consisted of a mix of two viruses: wild-type WSN and a variant of this virus that carries synonymous viral "barcodes" near the termini of each gene. This viral stock activated the IFN reporter in ~0.5% of infected cells (Fig. 1B), a frequency roughly comparable to that reported in prior studies (11, 12). We also validated that MACS for the cell-surface protein driven by the IFN reporter enriched the IFN+ cells by >50-fold (Fig. S3).

Combined transcriptomics and virus-sequencing of single infected cells. We developed the approach in Fig. 2 to obtain the entire transcriptome *and* the full sequences of all viral genes in single cells. First, we generated the viral stock described in the previous subsection, which consisted of a mix of wild-type WSN and a "synonymously barcoded" variant that contained two engineered synonymous mutations near each termini of each gene (File S2). These viral barcodes allow us to identify co-infections from single-cell transcriptomic data (12), and provide a control for PCR artifacts during full-length sequencing of viral transcripts (see below). We used this viral stock to infect A549 IFN reporter cells (Fig. 2A) at a dose that led to detectable viral transcription in ~25% of cells (this moderately low MOI reasonably balances our desire to limit the number of co-infections with the cost of performing transcriptomics on uninfected cells). From 12 to 13 hours post-infection, we used MACS to enrich cells that activated the IFN reporter. To ensure the presence of IFN-negative cells, we added back non-enriched cells to ~10% of the total. We also added uninfected canine cells to ~5% of the total as a control for multiplets and to estimate the background amount of viral mRNA detected in truly uninfected cells.

We processed the cells on a commercially available platform (54) that isolates cells in droplets and reverse transcribes polyadenylated mRNAs to append a unique cell barcode to all cDNAs in each droplet, and a unique molecular identifier (UMI) to each cDNA molecule (Fig. 2B). Because influenza virus mRNAs are polyadenylated (55), this
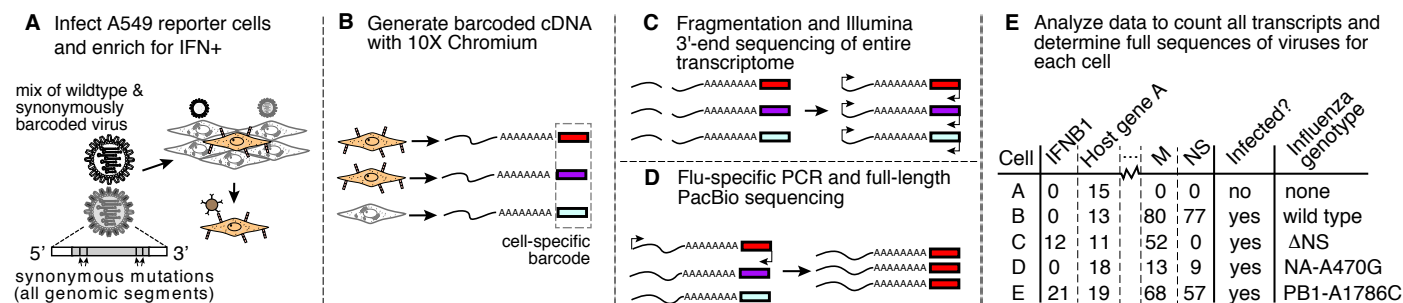
FIG 2 Approach for combined transcriptomics and viral sequencing of single influenza-infected cells that express IFN. (A) IFN reporter A549 cells are infected with a mix of wild-type and synonymously barcoded viruses. IFN+ cells are enriched by MACS, and pooled with non-enriched cells and uninfected canine cells that serve as a control for multiplets and mRNA leakage. (B) The mRNAs from individual cells are converted to cDNAs tagged with cell-specific barcodes. (C) Cellular transcriptomes are quantified using standard single-cell 3'-end Illumina sequencing, and (D) viral genes are enriched by influenza-specific PCR and fully sequenced by PacBio (in this schematic, only the cell labeled by the red barcode is infected and has viral transcripts that are sequenced by PacBio). (E) The result is a matrix giving the expression of each gene in each cell, as well as the full sequences of the viral genes in infected cells.

process appends cell barcodes to viral as well as cellular mRNAs. Furthermore, because virtually the entire influenza genome is transcribed, the cell-barcoded cDNA spans almost all 13,581 nucleotides in the segmented viral genome: the only portions not covered are one universally conserved nucleotide upstream of the transcription start site ([56]) and 17 to 22 highly conserved nucleotides downstream of the polyadenylation site ([55]) in each of the eight viral gene segments.

We used a portion of the cell-barcoded cDNA for standard single-cell transcriptomics by Illumina 3'-end sequencing (Fig. 2C). But we also took a portion and enriched for full-length viral molecules by PCR (Fig. 2D). We performed PacBio sequencing on these full-length viral cDNAs to generate high-accuracy circular consensus sequences (CCSs) ([57]). These CCSs retain the cell barcodes, and with sufficient sequencing depth we obtain CCSs from multiple unique UMI-tagged cDNAs for each viral gene in each cell. Because most cells are infected by just one or two virions, we can build a consensus of CCSs for each viral gene in each cell to determine the sequence(s) of these virions. Combining this information with the 3'-end sequencing determines the entire transcriptome and the full sequences of the infecting virions in single cells (Fig. 2E).

Transcriptomic analyses of single IFN+ and IFN- influenza-infected cells. We obtained transcriptomes for 1,614 human (A549) cells, and 50 of the uninfected canine cells that were spiked into the experiment as a control (Fig. 3A). We also obtained 12 transcriptomes with a mix of human and canine transcripts; from the number of such mixed cell-type transcriptomes, we estimate ([58]) that ~11% of the transcriptomes are derived from multiple cells. To remove some of these multiplets along with low-quality droplets, we filtered transcriptomes with unusually high or low numbers of cellular transcripts as is commonly done in analysis of single-cell RNA-seq data ([59]). After this filtering, we retained 1,490 human cells for further analysis (Fig. 3B)

To identify infected cells, we examined the fraction of each transcriptome derived from virus (Fig. 3C). As expected, only a small fraction (~0.7%) of transcripts in the uninfected canine cells were viral; this low-level background is likely from lysed cells that release ambient viral mRNA. We tested whether each cell contained significantly more viral transcripts than expected under a Poisson model given this background fraction, and classified 290 human cells as definitively infected with influenza (Fig. 3C). We classified the other cells as uninfected, although it is possible that some were infected with virions that produced very little mRNA. The distribution of the amount
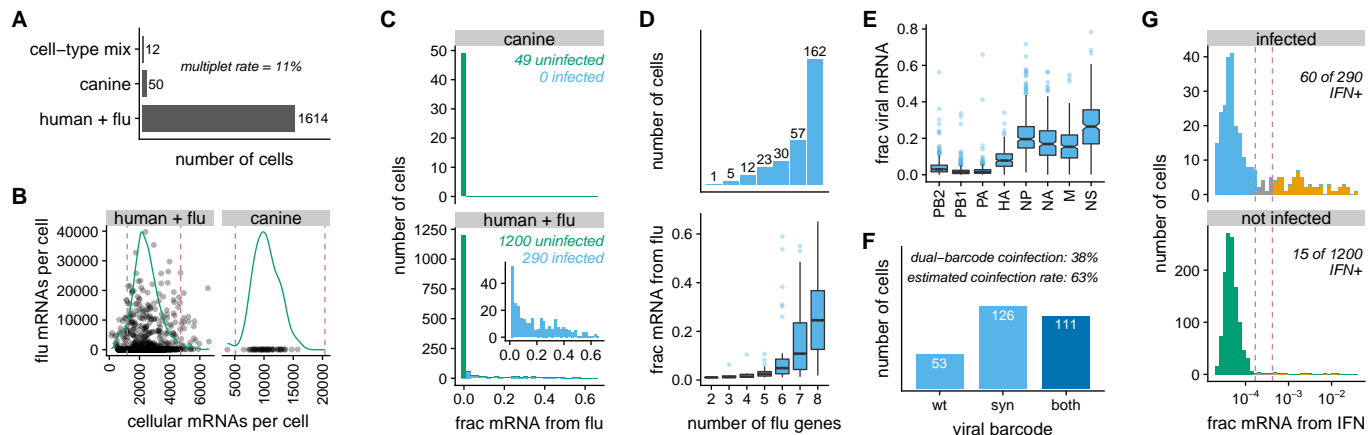
FIG 3 Single-cell transcriptomics of IFN-enriched influenza-infected cells. (A) Number of cells for which transcriptomes were obtained. From these numbers, we estimate (58) that ≈11% of the transcriptomes are derived from multiple cells. (B) The number of cellular and viral mRNAs detected for each cell is plotted as a point. Green lines show the distribution of cellular mRNAs per cell. Cells outside the dashed magenta lines have unusually low or high amounts of cellular mRNA (likely low-quality emulsions or multiplets), and are excluded from subsequent analyses. (C) Distribution across cells of the fraction of all mRNA derived from influenza. Cells called as infected are in blue, while other cells are in green. The inset shows the amount of viral mRNA in the human cells that are called as infected. (D) Number of influenza genes detected per infected cell, and the amount of viral mRNA in cells expressing each number of viral genes. Fig. S4 shows the frequency that each viral gene is detected. (E) Relative expression of viral genes, quantified as the fraction of all viral mRNA in each infected cell derived from each gene. (F) Number of cells infected with wild-type virus, synonymously barcoded virus, or both. From the cells infected with both viral barcodes, we estimate (58) that 63% of infected cells are co-infected. (G) Fraction of cellular mRNA from IFN across cells, faceted by whether the cells are infected. Cells to the left of the first dashed magenta line are classified as IFN-, and cells to the right of the second line as IFN+. A pseudocount is added to the number of IFN transcripts detected in each cell, which is why none of the fractions are zero.

of viral mRNA across infected cells is shown in the inset in Fig. 3C. As in our prior work (12), the distribution is extremely heterogeneous: many infected cells have only a few percent of mRNA derived from virus, but viral mRNA comprises over half the transcriptome of a few cells.

We called the presence or absence of each viral gene in each infected cell, again using a Poisson model parameterized by background fractions estimated from uninfected canine cells. We called presence / absence of genes rather than transcripts, since the two genes that encode multiple transcripts (M1 / M2 from the M gene, and NS1 / NS2 from the NS gene) do so via alternative splicing that leaves both isoforms with the same termini, making them indistinguishable by 3'-end sequencing. Fig. 3D (top panel) shows that 162 of 290 infected cells express all eight genes (see Fig. S4 for frequencies for individual genes). This measured frequency of infected cells expressing all eight genes is slightly higher than in our own prior work using the WSN strain (12), and slightly to substantially higher than that reported in studies by others using different viral strains or methodologies (15, 43, 60, 61).

The amount of viral mRNA was lower in cells that failed to express viral genes (Fig. 3D, bottom panel). However, viral burden remained highly variable even after conditioning on the number of viral genes: some cells that failed to express one or even two genes still derived >50% of their mRNA from virus, while other cells that expressed all genes had only a few percent of their mRNA from virus (Fig. 3D, bottom panel). Consistent with our prior work (12), despite the wide variation in absolute expression of viral genes, their *relative* expression was fairly consistent (Fig. 3E) and generally matched values from older bulk studies (62).

By examining the synonymous viral barcodes near the 3' termini of transcripts,

we determined that 38% of cells were co-infected with wild-type and synonymously barcoded virions (Fig. 3F; cells called as co-infected if a binomial test rejected null hypothesis that ≥95% of viral mRNA is from one viral barcode variant). From Fig. 3F, we estimate (58) that 63% of infected cells are co-infected. Interestingly, this co-infection rate is higher than expected from the relative numbers of infected and uninfected cells (Fig. 3C) if infection is Poisson. This discrepancy could arise if the MACS for IFN+ cells also enriches co-infected cells, if infection is not truly Poisson, or if co-infection complements otherwise transcriptionally defective virions to increase the likelihood that we identify a cell as infected. The first explanation seems unlikely, as there is no tendency for co-infected cells to express more IFN (Fig. S5). Therefore we favor the latter two explanations, both of which have been demonstrated for other viruses (63, 64). The moderately high rate of co-infection may also explain why more cells in our experiment express all eight viral genes compared to some prior studies, as a co-infecting virion can complement a missing viral gene.

We next examined expression of IFN and ISGs (Fig. 3G and Fig. S6). Over 20% of infected cells were IFN+, indicating that the MACS successfully enriched IFN+ cells far beyond their initial frequency. The expression of type I and type III IFN was highly correlated in single cells, justifying our decision to collapse both classes under the single label of "IFN" in the analyses that follow (Fig. S7). Few (~1.3%) uninfected cells were IFN+; the few that were present might be because the MACS enriched for rare cells that activated IFN in response to non-viral ligands (65, 66, 67) or because some cells that we classified as uninfected were actually infected at low levels. The difference in the frequency of IFN positivity among infected and uninfected cells in Fig. 3G was highly significant ($P < 10^{-5}$, Fisher's exact test). Many more cells expressed ISGs than IFN itself (Fig. S6A). The IFN+ cells were a subset of the ISG+ cells: IFN+ cells always expressed ISGs, but many ISG+ cells did not express IFN (Fig. S6B). These results are consistent with the established knowledge that IFN is expressed primarily in cells that directly detect infection, but that ISGs are also expressed via paracrine signaling in other cells (1, 2).

Finally, we qualitatively examined how expression of viral genes, IFN, and ISGs relate to the overall structure of the high-dimensional transcriptomic data. Fig. S8 shows unsupervised t-SNE clustering (68) of the cells. Cells expressing high levels of viral genes, IFN, and ISGs cluster together—and most of the structure in the t-SNE plot that is not associated with these genes involves uninfected and IFN- cells.

Full genotypes of viruses infecting single IFN+ and IFN- cells. We next used PacBio sequencing (Fig. 2D) to determine the full sequences of the viral genes in single infected cells. We obtained >200,000 high-quality PacBio CCSs that mapped to an influenza gene and contained a cell barcode and UMI (Fig. S9). The synonymous viral barcodes at both termini of each gene enabled us to confirm that PCR strand exchange was rare (Fig. S10), meaning that the vast majority of CCSs correctly link the sequence of the transcript to cell barcodes and UMIs that identify the cell and molecule of origin.

After calling the presence / absence of each viral gene in each cell as described in the previous section, we called mutations if they were found in at least two CCSs originating from different mRNAs (unique UMIs) and at least 30% of all CCSs for that gene in that cell. For cells co-infected with both viral barcode variants, we called mutations separately for each viral variant. This strategy reliably identifies mutations in virions that initiate infection of cells infected with at most one virion of each viral barcode variant (~75% of infected cells), as well as high-abundance mutations in cells co-infected with multiple virions of the same viral barcode. It will not identify mutations that arise within a cell after the first few rounds of viral genome replication, since such
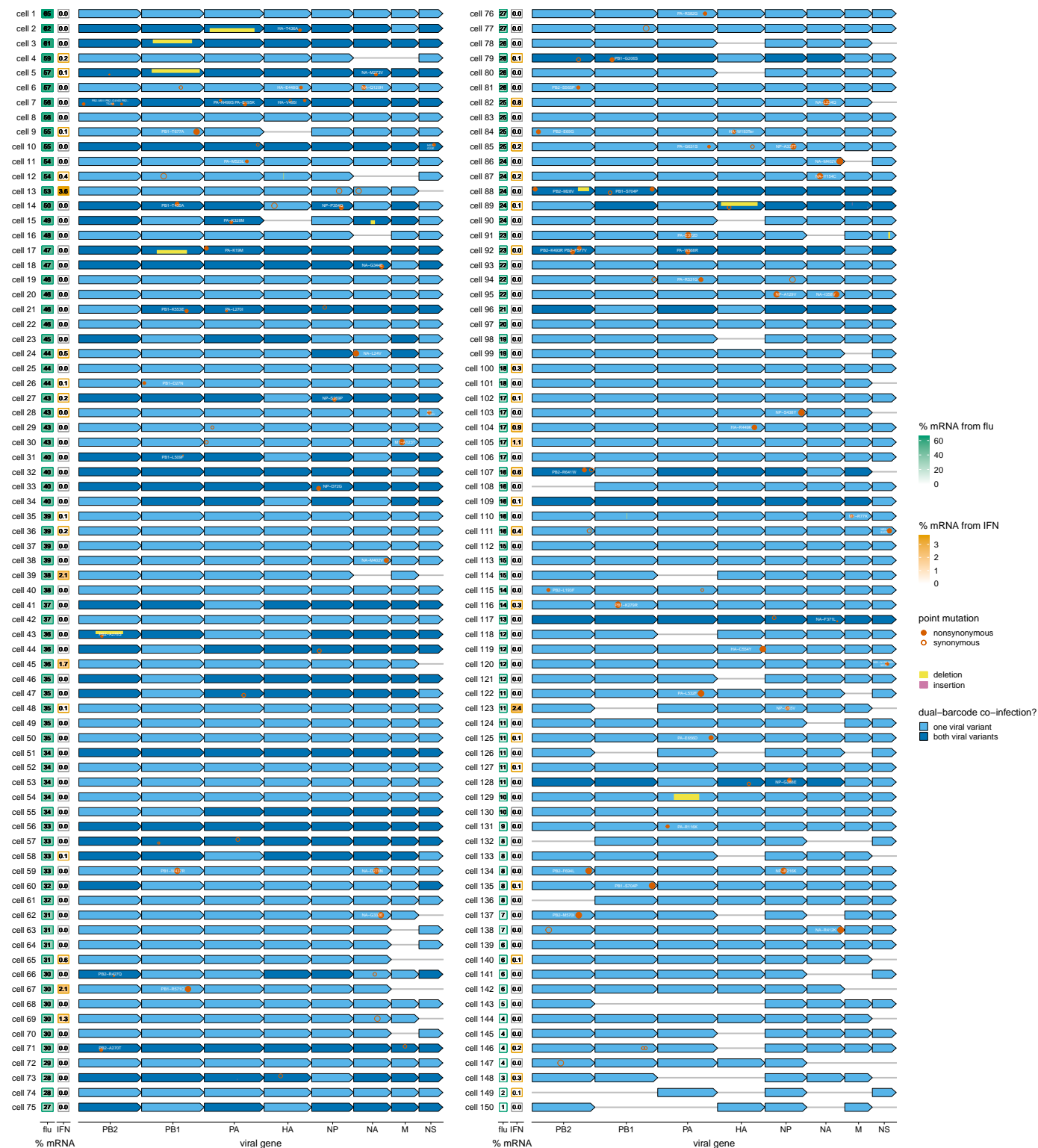
Single-cell virus sequencing of influenza infections that trigger innate immunity

FIG 4 Viral genotypes and infection outcomes in single cells. Green and orange boxes at the left show the percent of all mRNA in that cell derived from virus and the percent of all cellular mRNA derived from IFN, respectively. The second box is framed in orange for cells classified as IFN+ in Fig. 3G. Blue arrows indicate the presence of a viral gene from one (light blue) or both (dark blue) viral barcode variants; a dark blue arrow therefore means that a cell was co-infected. Circles and boxes on the arrows indicate mutations or indels as described in the legend at right. The circle areas and box heights are proportional to the fraction of CCSs with that mutation. For dual-barcode infections, mutations / indels for the wild-type and synonymously barcoded viral variants are shown in the top and bottom half of the arrows, respectively. For instance, *cell 5* was co-infected by a virion with one unmutated and one internally deleted copy of PB1.

mutations will not reach 30% frequency in that cell. Therefore, analogous to somatic variant calling in tumor sequencing (69, 70), there is a limit to our detection threshold: we cannot identify mutations that occur on just a small fraction of transcripts in a cell.

We could call the sequences of all expressed viral genes in the majority of infected cells (Fig. S11). We were most effective at calling full viral genotypes in cells that expressed high amounts of viral mRNA and were infected by only one viral barcode variant (Fig. S11). But we also called full genotypes for many cells that had low viral burden or were co-infected by both viral barcode variants.

The 150 cells for which we called the full viral genotypes are shown in Fig. 4. Visual inspection of this figure reveals a wealth of information. For instance, the cell with the highest viral burden (*cell 1* in Fig. 4, which has 65% of its mRNA from virus) was infected by a virion that expressed unmutated copies all eight genes and did not induce detectable IFN. But 12 of the other 13 cells with at least 50% of their mRNA from virus were infected by virions that had a mutation or failed to express a gene, and five of these cells produced IFN. As expected, all cells infected by virions that failed to express a component of the viral polymerase complex (PB2, PB1, PA, or NP) expressed low amounts of viral mRNA since they are limited to primary transcription off the incoming proteins (e.g., *cell 132* and *cell 143*). The two cells that expressed the most IFN (*cell 13* and *cell 123*) lacked the viral NS gene that encodes the virus's primary IFN antagonist (24, 25). Many other IFN+ cells had different defects such as large internal deletions (e.g., *cell 5* and *cell 89*) or amino-acid mutations (e.g., *cell 9*, *cell 28*, and many others).

However, Fig. 4 also reveals stochasticity that is independent of viral genotype. This stochasticity sometimes acts to the detriment of the virus, and sometimes to the detriment of the cell. As an example of the former case, expressing unmutated copies of all eight genes did not guarantee high viral gene expression and successful innate-immune evasion: for instance, the unmutated virion that infected *cell 139* only managed to express viral mRNA to 6% of the total transcriptome, and the unmutated virion that infected *cell 105* still induced IFN. But in other cases, the stochasticity allows a defective virus to still escape immune recognition. For instance, there are a number of cells (e.g., *cell 62* and *cell 78*) that do not activate IFN despite being infected by virions that fail to express NS.

Viral defects associated with viral gene expression and IFN induction in single cells. To systematically assess viral features associated with infection outcome, we divided the 150 cells in Fig. 4 into those that expressed unmutated copies of all eight genes (disregarding synonymous mutations) and those that did not. Fig. 5A shows that the 49 cells infected by full unmutated virions had a significantly tighter distribution of the amount of viral mRNA per cell than the other 101 cells as quantified by the Gini index (71). Therefore, viral defects are a major contributor to the heterogeneity in viral transcriptional burden.

Some viral defects also contribute to IFN induction. Specifically, cells infected by incomplete or mutated virions expressed IFN more frequently than cells infected by virions that expressed unmutated copies of all genes (Fig. 5B), although this difference was not statistically significant ($P = 0.12$, Fisher's exact test). However, the association was significant for certain classes of viral defects: absence of NS and amino-acid mutations in PB1 were significantly enriched in IFN+ cells, and amino-acid mutations in NS and deletions in HA were weakly enriched (Fig. 5C). The only trend that remained significant at a false discovery rate (FDR) of 10% was absence of NS. This lack of statistical significance after FDR correction could be due to the relatively modest number of fully sequenced infected cells (just 150). The validation experiments in the
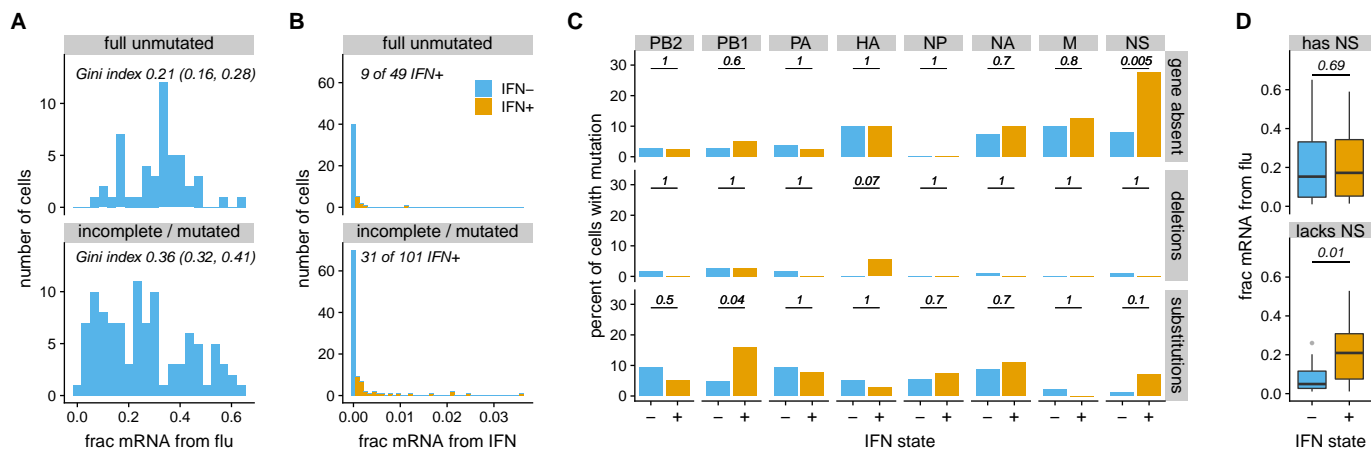
FIG 5   Viral features associated with heterogeneity in infection outcome among cells for which we determined viral genotypes. (A) Percent of all mRNA derived from virus, faceted by whether cells express unmutated copies of all eight genes. Cells infected by fully unmutated virions exhibit less heterogeneity in viral burden as quantified by the Gini index (95% confidence intervals are indicated). (B) IFN expression among cells expressing unmutated copies of all genes, and among cells with mutations or missing genes. (C) Specific viral defects associated with IFN induction. The top panel show the percent of IFN- and IFN+ cells that fail to express each viral gene. The middle and bottom panels show the percent of IFN- and IFN+ cells that have a deletion or amino-acid substitution in each gene, conditioned on the cell expressing that gene. Numbers give P-values (Fisher's exact test) for rejecting the null hypothesis that percents are equal among IFN- and IFN+ cells. (D) There is no association between IFN induction and the amount of viral mRNA in cells that express NS, but viral burden is associated with IFN induction among cells that lack NS. Throughout this figure, we only consider substitutions that are non-synonymous.

next section show that many of the viral mutations in IFN+ cells do in fact increase the rate of IFN induction.

One other interesting trend emerges from the single-cell data. There is no difference in the amount of viral mRNA between IFN+ and IFN- cells that express NS (Fig. 5D). But among cells that lack NS, cells with more viral mRNA are significantly more likely to be IFN+ (Fig. 5D); this finding is elaborated on in the validation experiments below. Overall, the lack of reduced viral gene expression in IFN+ cells suggests that autocrine IFN signaling typically occurs too late to suppress viral transcription, and the well-known inhibitory effect of IFN against influenza depends mainly on paracrine signaling.

Validation that viral defects in single IFN+ cells often increase IFN induction. To test if the viral defects identified in single IFN+ cells cause increased IFN expression, we used reverse genetics to generate bulk stocks of viruses with some of these defects.

The viral defect most strongly associated with IFN induction was failure to express the NS gene (Fig. 4, Fig. 5C). Although it is sometimes possible to use complementing cells to generate influenza viruses lacking a specific gene (72, 73), we were unable to generate viruses that lacked NS. The NS gene encodes two proteins (NS1 and NS2), the first of which is influenza's primary innate-immune antagonist (24, 25). We therefore mimicked the absence of NS by creating a mutant virus (which we term "NS1stop") that had multiple stop codons early in the NS1 coding sequence.

The single-cell data also showed that amino-acid substitutions in proteins encoded by the PB1 and NS genes were enriched in IFN+ cells (Fig. 4, Fig. 5C), so we created mutant viruses with some of these substitutions: PB1-D27N, PB1-G206S, PB1-K279R, PB1-T677A, NS1-A122V, and NS2-E47G.

Finally, prior work has suggested that virions with internal deletions in the polymerase genes can induce higher levels of IFN (16, 38, 39, 40, 41, 42). Although such
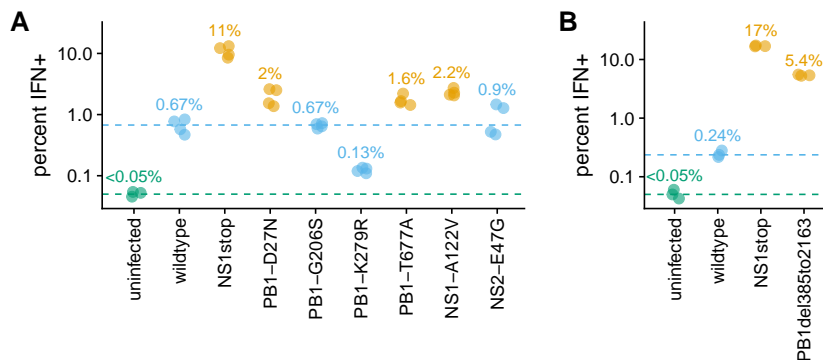
FIG 6 Validation that IFN induction is increased by some of the mutations identified in the single-cell virus sequencing of IFN+ cells. (A) Percent of infected cells that become IFN+ after infection with a bulk stock of the indicated viral mutant, as determined using a reporter cell line. The numbers indicate the median of four measurements for each viral mutant. The limit of detection of 0.05% is indicated with a dashed green line, and the median value for the wild type viral stock is indicated with a dashed blue line. Points are colored orange if the mutant virus stock induces IFN more frequently than the wild-type viral stock (one-sided t-test, $P < 0.01$), and blue otherwise. (B) Similar to the first panel, but validates increased IFN induction for a large internal deletion in the PB1 gene, and normalizes infecting virion dose rather than calling IFN+ percentage only among infected cells. See Fig. S12 and Fig. S13 for details. The experiments in the two panels were performed on different days, and so numerical values can be reliably compared within panels but not between panels.

deletions are not significantly enriched among IFN+ cells in our single-cell data (Fig. 5C), there is a co-infected IFN+ cell where one viral variant has a deletion in PB1 spanning nucleotides 385 to 2163 (*cell 5* in Fig. 4). We therefore created a virus carrying this deletion, and propagated it in cells constitutively expressing PB1 protein.

We tested the rate of IFN induction by each viral stock using the reporter cells. Fig. 6 shows that five of the eight mutant viral stocks induced IFN more frequently than a wild-type viral stock. The strongest IFN induction was by the NS1stop virus, but the PB1 internal deletion and three of the point-mutant viruses (PB1-D27N, PB1-T677A, and NS1-A122V) also induced IFN significantly more frequently than wild type. The other three point mutants (PB1-G206S, PB1-K279R, and NS2-E47G) did not increase IFN induction—an unsurprising finding, since we expect some mutations without an IFN-enhancing effect to be found in IFN+ cells by chance. Overall, the results in Fig. 6 validate that the viral defects in single IFN+ cells often cause increased IFN production.

However, IFN induction remains stochastic even for the most potently IFN-inducing viral mutants. Fig. 6 shows flow cytometry data (see also Fig. S12), which is itself a single-cell measurement, albeit one that does not report the viral genotype. As can be seen from these data, none of the mutant viral stocks induce IFN in more than 20% of infected cells. Of course, these mutant virus stocks are themselves genetically heterogeneous, as many virions will have additional defects similar to that revealed by our single-cell sequencing of the "wild-type" viral stock. But our single-cell data show that IFN induction is stochastic even for infections that share the same defect, such as absence of NS (e.g., compare *cell 62* and *cell 69* in Fig. 4). Therefore, the experiments in this section not only validate some specific viral defects that increase IFN induction, but also show that induction remains stochastic even with these defects.

The IFN-inducing viral defects act by diverse mechanisms. Some of the viral defects in IFN+ cells are easy to reconcile with existing knowledge: for instance, NS1
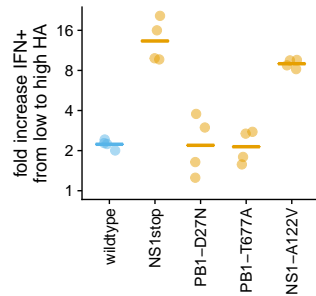
FIG 7 Infected cells that express higher levels of HA protein are much more likely to induce IFN expression only if they are infected by virus with defects in NS1. The y-axis shows the ratio of the percent of IFN+ cells in the highest HA-expression quartile relative to the lowest HA-expression quartile. Points indicate replicates, and lines indicate the mean. This figure is based on joint analysis of the IFN reporter and HA staining for all infected cells in the flow cytometry data in Fig. S12; see Fig. S14 for more details.

is the virus's primary IFN antagonist (24, 25), and internal deletions are prevalent in immunostimulatory viral stocks (16, 38, 39, 40, 41, 42). Other defects are more surprising: for instance, it is not obvious why amino-acid mutations in PB1 increase IFN induction. We therefore designed experiments to interrogate some of these defects in more detail.

We first focused on one of the strongest trends from the single cell data: increased viral gene expression is associated with increased IFN induction when the infecting virion fails to express NS, but not otherwise (Fig. 5D). To confirm this observation, we performed a flow cytometry analysis of the reporter cells infected by different immunostimulatory viral mutants to examine the association between expression of a viral gene product (HA protein) and IFN induction. Consistent with the single-cell data, cells that expressed more HA were much more likely to turn IFN+ when infected with the NS1stop or NS1-A122V mutants, but not when infected with any of the other viral variants (Fig. 7). This fact suggests that when there are high levels of viral transcription, NS1 becomes more important as a buffer against detection of viral products.

We hypothesized that the immunostimulatory mutations to PB1 might cause the viral polymerase to produce aberrant products, in line with recent work showing that mutations to PB2 can lead to the generation of aberrant RNAs that trigger RIG-I (35, 36). To investigate if the PB1 mutations might perturb polymerase activity, we examined their location in a structural model of the polymerase complex (Fig. 8A). The IFN-enhancing PB1 mutation T677A occurs at the tip of a helix that interacts with the 3' terminus of the RNA template as it enters the channel above the active site, whereas the IFN-enhancing D27N mutation is deeper in the polymerase close to the binding pocket of the 5' terminus of the template. Therefore, both mutations could plausibly alter the polymerase's interactions with the RNA template.

To test if the PB1 mutations affect activity, we transfected 293T cells with plasmids that express wild-type or mutant PB1 protein along with the other proteins in the polymerase complex (PB2, PA, and NP) and full-length viral RNA (vRNA) for the NA segment. Both polymerase mutations increased IFN expression in this assay (Fig. 8B), indicating that they have an immunostimulatory effect in the context of an active viral polymerase even when other viral components are absent. We next directly measured polymerase activity on the full-length vRNA template by extracting total RNA and quantifying replication (vRNA) and transcription (mRNA) products by primer extension. Both immunostimulatory PB1 mutations had activities that were significantly
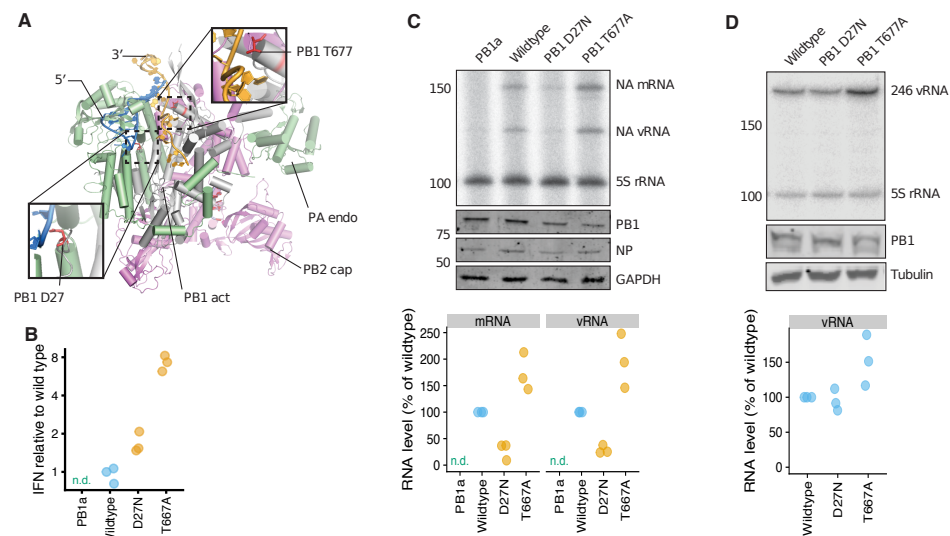
FIG 8 IFN-inducing mutations D27N and T677A in the PB1 protein affect polymerase activity. (A) Model of bat influenza A virus polymerase (PDB 4WSB) (74) superposed with the influenza B virus polymerase (PDB 5MSG) (75). The locations of PB1 D27 and T677 (both red) relative to the 5' (blue) and 3' (orange) termini of the RNA template and the PB1 active site (grey; PB1 act) are indicated. The PA endonuclease (green; PA endo) and PB2 cap binding domain (pink; PB2 cap) are also indicated. Part of the fingers subdomain of PB1 is hidden to reveal the template in the entry channel. (B) IFN-beta promoter activity measured using a dual luciferase reporter assay in 293T cells transfected with plasmids expressing the indicated PB1 protein, the other polymerase complex proteins (PB2, PA, and NP), and a full-length NA vRNA template. PB1a is a catalytically inactive PB1 active site control. In this panel and the next two panels, points show three biological replicates; "n.d." indicates not detectable, and orange indicates a variant was significantly different than wild type by a two-sided t-test. (C) Polymerase activity on full-length vRNA template in 293T cells transfected as in panel (B). Steady-state RNA levels were measured by primer extension, denaturing PAGE, and phosphorimaging. PB1a was used as negative control and background correction. The 5S rRNA signal was used as loading control. Other panels show Western blot analysis of PB1, NP and GAPDH protein expression, and the graph at the bottom shows quantification by phosphoroimaging. (D) Polymerase activity on a short 246-nucleotide vRNA template. The top panel shows the steady state levels of vRNA template as determined by primer extension and denaturing PAGE. The other two panels show the PB1 and tubulin expression levels analyzed by Western blot, and the graph shows quantification.

different from wild type, despite being expressed at wild-type protein levels (Fig. 8B). Specifically, T677A had higher levels of both activities, whereas D27N had reduced levels of both—although D27N still retained activity far in excess of a control active-site mutant (Fig. 8C). We speculated that the mutations might alter polymerase processivity, leading to accumulation of aberrant RNA products that activate the innate immune system (35, 38, 39, 40, 41, 42). We therefore repeated the activity assays using a short 246-nucleotide template (35) in place of the full-length NA vRNA (Fig. 8D). On this shorter template, the activity of the D27N mutant was now similar to wild type, while the activity of the T677A mutant remained higher than wildtype (although not significantly so in three biological repeats). Therefore, the two immunostimulatory PB1 mutations have distinct effects on the polymerase: D27N reduces processivity thereby favoring shorter RNA products, whereas T677A increases overall activity which could also lead to accumulation of aberrant RNA products.

Overall, the results in this section show that the diverse range immunostimulatory viral defects identified in single cells act by diverse processes, showing that viral varia-

tion influences not only the rate of IFN induction but also the factors that contribute to this induction.

DISCUSSION

We have determined the full sequences of all viral genes in single influenza-infected cells. Methodologically, our major advance is to measure the *genotypes* of viruses in addition to the abundance of viral components (i.e., transcripts, proteins, or progeny virions) as has been done by prior single-cell studies (12, 13, 15, 16, 14, 43, 45, 46, 47, 48, 76, 77, 78). Our method builds on the observation that fragmentary viral genetic information can be obtained by more standard single-cell transcriptomic techniques (16, 46, 47). To make this information complete, we have coupled single-cell transcriptomics with long-read PacBio sequencing of viral genes, a strategy analogous to that used by (79) to obtain full-length isoforms of some cellular genes in single cells.

This viral genetic information helps explain cell-to-cell variation in viral gene expression and innate-immune induction. Despite the fact that we used a low-passage viral stock generated from plasmids, most infected cells do not express unmutated copies of all viral genes. Although our study is certainly not the first to note that influenza has a high mutation rate (30, 31, 32, 33, 34) and sometimes fails to express genes (12, 15, 43, 60, 61), it is the first to directly observe the full spectrum of these defects across single cells. Visual inspection of Fig. 4 shows how any experiment that does not sequence viral genes in single cells is averaging across a diverse spectrum of viral defects.

We identified four types of defects that we validated to increase IFN induction. Two types of defects—absence of the NS gene and amino-acid mutations to the NS1 protein—presumably impair NS1's well-known ability to antagonize innate immunity (24, 25). But although the general role of NS in innate-immune antagonism has long been appreciated, our work represents the first direct demonstration that stochastic absence and mutations to this gene are a major contributor to IFN induction in single cells. A third type of defect, amino-acid mutations in PB1, was more surprising since this protein has not been described as a major player in innate-immune detection. We characterized two IFN-inducing PB1 mutations, and showed that one impaired polymerase processivity whereas the other increased overall activity. We speculate that these alterations increase production of aberrant immunostimulatory RNA products (35). Finally, we found an internal deletion in PB1 that enhances IFN induction, consistent with prior work showing such deletions are immunostimulatory (38, 39, 40, 41, 42). In fact, given the extensive prior work on deletions, we were surprised not to identify more of them in our IFN+ cells. There may be several reasons: we used pure viral stocks (53) at modest MOI; our experiments preferentially captured cells with higher viral transcriptional load; and most prior studies have used techniques that can detect large deletions but not subtle point mutations. Additionally, the relative importance of different defects likely varies across infection conditions, viral strains, and cell types: it is an open question which defects are most relevant for immune detection during actual human infections.

However, the greatest value of our work is not as a screen for IFN-inducing defects, but rather as a relatively unbiased survey of the breadth of viral variation in individual infected cells. This survey shows that no single type of viral defect determines whether a cell induces IFN: even the most immunostimulatory defect (absence of NS) occurs in only about a quarter of IFN+ cells. Therefore, innate-immune detection of influenza is a multi-faceted process that cannot be ascribed a single dominant viral genetic cause.

Our results further show that viral genetic defects do not fully explain the hetero-

geneity among influenza-infected cells. There is substantial breadth in viral transcriptional burden and occasional IFN induction even among cells infected with unmutated virions. Additionally, no viral defect induces IFN deterministically: every type of immunostimulatory defect that we characterize is also observed in IFN- cells in our single-cell dataset. Therefore, stochasticity or pre-existing cellular states also play a major role in affecting innate-immune induction—a finding that concords with the fact that IFN induction is heterogeneous even among cells treated with synthetic innate-immune ligands (20, 21, 22, 23), as well as for other viruses (48).

Perhaps the most intriguing question is how the heterogeneity that we have described ultimately affects the macroscopic outcome of infection. Natural human influenza infections are established by just a handful of virions (17, 18, 19) that then undergo exponential growth, and early IFN responses are amplified by paracrine signaling (1, 2). It is therefore plausible that early heterogeneity in innate-immune induction could affect the entire course of infection. Extending our approaches to more complex systems could shed further light on how viral genetic variation and stochasticity interact to shape the race between virus and immune system.

MATERIALS AND METHODS

IFN reporter cell lines. We created IFN reporter variants of the A549 human lung epithelial cell line (Fig. 1A). The parental A549 cell line used to create these reporters was obtained from ATCC (CCL-185), and was tested as negative for mycoplasma contamination by the Fred Hutch Genomics Core and authenticated using the ATCC STR profiling service. The 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 and 5% carbon dioxide.

To create the type I interferon reporters, a 1kb promoter region upstream of the human IFNB1 gene were cloned into the pHAGE2 lentiviral vector (80), with a NotI site immediately downstream of the promoter serving as an artificial Kozak sequence. Downstream of this NotI site, each of the following reporter constructs was cloned: mCherry, mNeonGreen, and low-affinity nerve growth factor lacking the C-terminal signaling domain (LNGFRΔC) (49, 50) linked to mNeonGreen by a P2A linker (81). The sequence of the last of these constructs is provided in File S1.

To create the type III interferon reporters, a 1.2kb region upstream of the human IL29 (IFNL1) gene was cloned into the pHAGE2 vector, with the native Kozak sequence retained at the 3' end. Downstream of this promoter we cloned LNGFRΔC linked to ZsGreen via a P2A linker. The sequence of this construct is provided in File S1.

We used these constructs to generate lentiviral vectors and transduce of A549 cells in the presence of 5 μg polybrene. We then sorted single transduced cells and expanded them. A portion of the expanded cells were tested for reporter activity by transfecting poly(I:C) (a potent agonist of the RIG-I pathway), and we retained clones with strong activation. Importantly, the cells that we retained for further use were not the same portion that were tested by poly(I:C) treatment, but rather a separate split of the same population—this avoids any selection on the cells from transient activation of IFN. For the dual type I / type III reporter used in Fig. S1B, a single-cell clone of the type III reporter cell line was transduced with the type I reporter bearing the mCherry fluorescent marker, and then isolated and propagated as a single cell clone for the other cell lines. All reporter lines tested negative for mycoplasma contamination by the Fred Hutch Genomics Core.

Fig. S1A shows validation of the reporter cell lines using infection with saturating amounts of the Cantell strain of Sendai virus (obtained from Charles River Laboratories). For detection of the cell-surface bound LNGFRΔC, cells were stained with PE-conjugated anti-LNGFR (CD271) antibody from Miltenyi Biotec.

Viruses for single-cell experiments. We performed the single-cell experiments using the A/WSN/1933 (H1N1) strain of influenza virus. We used both the wild-type virus and a variant of the virus where synonymous mutations were added within a few 100 nucleotides of each

termini of each gene segment. We have used a similar synonymous viral barcoding strategy in our prior single-cell work ([12](#)) as it allows us to detect about half of co-infected cells based on the expression of both viral barcode variants. In the current work, we extended this approach by placing synonymous barcodes near *both* termini of the gene segments in order to quantify strand exchange during PacBio sequencing ([Fig. S10](#)). The sequences of all gene segments from the wild-type and synonymously barcoded viral strains are in [File S2](#). These genes were cloned into the pHW2000 ([52](#)) reverse-genetics plasmid.

Both viral strains were generated by reverse genetics using the pHW18* series of bi-directional plasmids ([52](#)). We controlled the durations and MOI during viral passaging since these factors can greatly affect the accumulation of defective viral particles ([53](#)). The viruses were generated by reverse genetics in co-cultures of 293T and MDCK-SIAT1 cells in 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) and then propagated in MDCK-SIAT1 cells in influenza growth media using the same basic procedures detailed in ([12](#)). Specifically, after generation by reverse genetics, the wild-type variant was expanded at an MOI of 0.001 for 72 hours twice in MDCK-SIAT1 cells, and the synonymously barcoded variant was expanded once at an MOI of 0.01 for 60 hours. The MOIs for this passaging are based on titers determined using TCID50 assays via the formula of Reed and Muench ([82](#)) as implemented at https://github.com/jbloomlab/reedmuenchcalculator. After being passaged independently, the two viral stocks were combined at equivalent numbers of infectious units to make the stock used in the single-cell experiments.

Flow cytometry analyses for HA expression. For the single-cell experiments (which only examine the transcriptional results of a single cycle of infection), we were most interested in the titer of viral particles that are transcriptionally active for a single round of infection of A549 cells. We estimated titers of transcriptionally active virions by staining for HA expression in virus-infected A549 cells. Specifically, we infected A549 cells (or one of the A549 reporter cell line variants as indicated) in influenza growth medium, and at 13 to 14 hours post-infection, we trypsinized cells, re-suspended in phosphate-buffered saline (PBS) supplemented with 2% heat-inactivated fetal bovine serum (FBS), and stained with 10 μg/ml of H17-L19, a mouse monoclonal antibody previously shown to bind to the HA from the A/WSN/1933 strain of virus ([83](#)). After washing in PBS supplemented with 2% FBS, the cells were stained with a goat anti-mouse IgG antibody conjugated to APC, washed, fixed in 1% formaldehyde in PBS, washed again, and then analyzed by flow cytometry to determine the fraction expressing detectable HA protein.

Single-cell transcriptomics of IFN-enriched infected cells using 10X Chromium. The single-cell transcriptomics and virus sequencing was performed using the A549 cells with the *IFNB1* LNGFRΔC-P2A-mNeonGreen reporter. A schematic of the experiment is shown in [Fig. 2](#).

The wild-type and synonymously barcoded viruses were mixed with the goal of adding equal numbers of transcriptionally active HA-expressing virions of each virus strain. The cells were then infected with this mixture at a dose designed to infect about half the cells ([Fig. 3](#)C suggests that the actual rate of detectable infection was slightly lower). Infections were allowed to proceed for 12 hours. The cells were then trypsinized, the trypsin was quenched with D10 media, and cells were resuspended in de-gassed PBS supplemented with 0.5% bovine serum albumin and 5 mM EDTA. To enrich IFN+ cells, the cells were then incubated with anti-LNGFR MACSelect Microbeads (Miltenyi Biotec) and twice passed over an MS magnetic column (Miltenyi Biotec), retaining the bound (and presumably IFN-enriched) population each time. This MACS sorting is expected to give approximately the enrichment for IFN+ cells shown in [Fig. S3](#). The original, unsorted, population was then added back in to ~10% of the final cell fraction in order to ensure the presence of interferon negative cells. At this point, uninfected canine (MDCK-SIAT1) cells were also added to ~5% of the final cell fraction to enable quantification of the cell multiplet rate ([Fig. 3](#)A) and background viral mRNA in uninfected cells ([Fig. 3](#)C). We began this entire process of cell collection and enrichment at 12 hours post-infection, but the process (which was performed at room temperature) took about an hour, and thus we consider the cells to have been analyzed at 13 hours post-infection. The final cell suspension was counted using a disposable hemocytometer and loaded on the 10x Genomics Chromium instrument ([54](#)), targeting capture of ~1,500 cells.

This sample was then processed to create libraries for Illumina 3′-end sequencing according to the 10X Genomics protocol using the Chromium Single Cell 3′ Library and Gel Bead Kit v2 with one important modification: rather than process all full-length cDNA through enzymatic fragmentation, several nanograms were retained for targeted full-length viral cDNA sequencing as described below. The single-cell transcriptomics library was sequenced on an Illumina HiSeq 2500, and the data analyzed as described below.

Enrichment and preparation of viral cDNA for PacBio sequencing. We amplified virus-derived molecules from cDNA retained from the 10X Genomics protocol for PacBio sequencing of the full-length cDNA. These cDNA have at their 3′ end the cell barcode and UMI plus the adaptor sequence that is added during the 10X protocol (see Fig. 2 for simple schematic, and File S7 for more details). We only wanted to PacBio sequence cDNA molecules derived from virus. We therefore needed to enrich for the viral molecules while retaining the 10X adaptor / UMI / cell barcode at the 3′ end.

We first performed a multiplex PCR reaction on 1 ng of the full-length 10X cDNA using a 3′ primer complementary to the common 10X adaptor, and a multiplex mix of eight 5′ primers, one specific for the mRNAs from each of the eight viral gene segments (File S3). A major concern during these PCRs is strand exchange (see Fig. S10) which would scramble the cell barcodes and mutations on viral cDNAs. To reduce strand-exchange and obtain more even PCR amplification across segments, we performed emulsion PCRs using the Micellula DNA Emulsion Kit (Roboklon), which physically separates disparate template molecules, preventing strand exchange and allowing each molecule to be amplified to exhaustion of its droplet's reagents (84). We performed the PCRs using Kapa HiFi Hotstart ReadyMix, supplementing the reactions with additional BSA to a final concentration of 0.1 mg/ml and using a volume of 100μl. Both the common 3′ primer and the multiplex mix of eight 5′ primers were added to a final concentration of 0.5μM. We performed 30 cycles of PCR, using an extension time of 2 minutes 15 seconds at 67°C, and a melting temperature of 95°C. This melting temperature is lower than the standard 98°C melting step suggested by the manufacturer for Kapa HiFi because we wanted to avoid collapse of emulsion integrity at high temperature.

The product of this multiplex PCR was subjected to eight additional individual emulsion PCR reactions, each using only a single segment-specific 5′ primer as well as the common 3′ primer, using 1 ng of material in each reaction. The material from these eight segment-specific PCRs was then pooled with the goal of obtaining a equimolar ratio of segments, and sequenced on one SMRT Cell in a PacBio RS II and one SMRT Cell of a PacBio Sequel. Detailed results from the analysis of these first two sequencing runs is shown in File S7. These results showed that although the PCRs substantially enriched for influenza molecules, the relative coverage of the different viral genes was still uneven, with the longer genes under-sampled.

To improve coverage of the polymerase genes, we produced two new sequencing pools: one consisting of the five shortest viral segments (HA, NP, NA, M, and NS) from the afore-mentioned segment-specific emulsion PCRs, and the other consisting of the three longer polymerase segments (PB2, PB1, and PA). The former was sequenced on one cell of a single SMRT Cell of a PacBio Sequel, and the latter on two additional SMRT Cells of a PacBio Sequel. As is shown File S7, the coverage remained relatively low for the polymerase genes—and most of the reads we did obtain were dominated by shorter internally deleted variants of the polymerase genes (53) which are preferentially amplified during PCR.

To obtain more reads for longer full-length polymerase variants, we therefore subjected 10 ng of our amplified material for each polymerase segment to a bead selection using SPRIselect beads at a volume ratio of 0.4. This selection removes most low-molecular weight DNA species including internally-deleted defective segments. Material from this selection was amplified using 16 (PB1) or 14 (PB2 and PA) cycles of a non-emulsion PCR using the standard conditions recommended by the Kapa HiFi Hotstart ReadyMix (extension at 67°C for 2 minutes 15 seconds, and melting at 98°C). The use of relatively few PCR cycles was designed to prevent the occurrence of the artifacts (including strand exchange) that occur in non-emulsion PCRs. We pooled the products of these reactions from this size-selection and sequenced on a SMRT Cell of a PacBio Sequel. As is shown in File S7, this sequencing yielded more full-length polymerase variants, but they were still undersampled compared to other viral genes.

To further to improve recovery of full-length PB1, PB2, and PA, we use an approach that

allowed us to perform a specific PCR for full-length polymerase variants. We circularized the template molecules, and then used two segment-specific primers that annealed in apposition near the center of each polymerase gene to linearize these circular molecules. Only molecules that contain the middle of the polymerase genes (which are typically full-length) are linearized by this process. In the downstream computational analysis, we can then determine the full sequence of the gene as well as the cell barcode of the initial molecule from which the linearized molecule is derived. Specifically, we first used 2.5 ng of our already-amplified segment-specific material in a 10-cycle PCR to append circularization adapters (see File S3 for sequences), and cleaned the resultant mixture using SPRIselect beads at a volume ratio of 0.4. We then used 10 ng of this amplified material in a 20μl NEBuilder reaction using an extended reaction time of 50 minutes in order to circularize the molecules. We next incubated these reactions for 1 hour at 37°C with exonuclease V and additional ATP to a final increase in concentration of 1 mM to digest all non-circularized molecules. The circularized and digested material was then cleaned using SPRIselect beads at a volume ratio of 0.4. This material was then used as template for three non-emulsion PCRs specific to PB2, PB1, or PA, using two segment-specific primers that align to the central portion of each gene but in apposition to each another (see File S3 for sequences). These linearization reactions used 20 (PB2) or 26 (PB1 and PA) PCR cycles, and the resulting products were cleaned using SPRIselect beads at a volume ratio of 1.0. This material was pooled to produce an equimolar mixture of full-length PB1, PA, and PB2 and sequenced in an additional SMRT Cell of PacBio Sequel. As is shown in File S7, this process yielded many full-length polymerase variants.

The computational analyses of the full-length viral gene sequences described below combined the data from all these reactions. The number of sequences obtained for each gene after pooling the data from all reactions is shown in Fig. S9, which also indicates that the net rate of strand exchange is very low (see Fig. S10 for an illustration of how this is determined). A detailed breakdown of the coverage of each gene and PacBio run is in File S7. Importantly, the PCR biases mean that the coverage of molecules by the PacBio sequencing is not proportional to their abundance in the starting mRNA. However, as described in the computational analysis section below, the final analyses use the cell barcodes and UMIs in conjunction with the standard 10X Illumina sequencing to ensure that none of the conclusions are affected by the disproportionate amplification of some molecules during the PacBio library preparation (for instance, duplicate UMIs are removed from the PacBio data, and all conclusions about gene abundance or absence are based on the Illumina data).

qPCR for viral genes and IFN. We performed qPCR on reverse-transcribed mRNA for influenza HA (to quantify viral transcription), IFNB1 (to quantify IFN induction), and L32 (a cellular housekeeping gene for normalization). For the qPCR, we used the SYBR Green PCR Master Mix (Thermo Fisher) according to the manufacturer's protocol using oligo-dT primers. The qPCR primers were: HA primer 1, 5′-GGCCCAACCACACATTCAAC-3′; HA primer 2, 5′-GCTCATCACTGCTAGACGGG-3′; IFNB1 primer 1, 5′-AAACTCATGAGCAGTCTGCA-3′; IFNB1 primer 2, 5′-AGGAGATCTTCAGTTTCGGAGG-3′; L32 primer 1, 5′-AGCTCCCAAAAATAGACGCAC-3′; L32 primer 2, 5′-TTCATAGCAGTAGGCACAAAGG-3′.

For the qPCR in Fig. S13, A549 cells were seeded at a density of 10^4 cells/well in a 96-well plate in D10 media 24 hours prior to infection, with four independent wells seeded per experimental treatment. Immediately prior to infection D10 media was removed and replaced with influenza growth media and infected with the indicated influenza strains at a MOI of 0.4 based on TCID50 in MDCK-SIAT1 cells. For the cells with cycloheximide added to block protein expression (and hence secondary transcription), cycloheximide was added to a final concentration of 50 μg/ml (a concentration sufficient to block secondary transcription ([85])) at the time of infection. After 8 hours, mRNA was harvested using the CellAmp Direct RNA Prep Kit for RT-PCR, reverse-transcribed using an oligo-dT primer, and qPCR was performed as described above.

Viruses and experiments for validation experiments. In Fig. 6, we tested the IFN inducing capacity of a variety of viral mutants identified in the single-cell experiments. For point-mutant viruses, we created variants for all amino-acid substitutions found in PB1 and NS among IFN+ cells that did not also lack NS. One of these mutants (amino-acid substitution S704P in PB1) did not reach sufficient titers in a single attempt to generate it by reverse ge-

netics, and so was dropped from the experiment (note that we did not attempt replicates of the reverse genetics for this mutant, and so are *not* confident in drawing strong conclusions about its actual attenuation). This left six point-mutant viruses: four with point mutations in PB1, and two with point mutations in NS. We also created a mutant virus that contained the internal deletion in PB1 found in an IFN+ cell. In addition, we created a virus with an inactivated NS1 to mimic the infections that failed to express NS (we were unable to use complementing cells to generate a viral stock that completely lacked the NS segment). This NS1stop virus contained six nucleotide changes resulting in the addition of five in-frame stop codons in NS1 starting 10 nucleotides downstream of the 5′ splice donor site, thereby disrupting NS1 while leaving NS2 (NEP) intact. All of these mutants were cloned into the pHW2000 bi-directional reverse-genetics plasmid (52) in order to enable generation of viruses encoding the mutant genes. File S6 provides the full sequences for all of these plasmids.

We generated the wild-type and point-mutant viruses for the validation experiments in Fig. 6A by reverse genetics using the pHW18* series of WSN reverse genetics plasmids (52), but substituting the appropriate mutant plasmid listed in File S6 for the wild-type plasmid for that gene. To generate the viruses from these plasmids, we transfected an equimolar mix of all eight plasmids into co-cultures 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. At 50 hours post-transfection (for the replicate 1 viruses in Fig. S12A) or 72 hours (for the replicate 2 viruses in Fig. S12), we harvested the virus-containing supernatant, clarified this supernatant by centrifugation at 300×g for 4 min, and stored aliquots of the clarified viral supernatant at -80°C. We then thawed aliquots and titered by TCID50 on MDCK-SIAT1 cells. For the infections in Fig. S12A, we wanted to use equivalent particle counts, so we normalized all viruses to an equivalent hemagglutination titer on turkey red blood cells (86). Briefly, a solution of 10% v/v red blood cells (LAMPIRE Biological Laboratories, Fisher Scientific catalogue number 50412942) was washed in PBS and diluted to a final concentration of 0.5% v/v. Two-fold serial dilutions of virus were added to an equal volume of diluted red blood cells, and titer was measured as the highest dilution of viral stock at which complete hemagglutination of red blood cells was observed. We then performed infections of the A549 reporter cell line at equivalent hemagglutination titer and analyzed the data as described in Fig. S12A.

To generate the NS1stop mutant virus and the wild-type and PB1del385to2163 mutant viruses in Fig. S12B, we used slightly different procedures. The wild-type virus was generated by reverse genetics as described for the point-mutant viruses above, harvested at 48 hours post-transfection, and then passaged on MDCK-SIAT1 cells for 36 hours at an MOI of 0.05— conditions that we previously validated to lead to relatively little accumulation of defective particles (12). The NS1stop virus was similarly generated, but was passaged for 48 rather than 36 hours, since it had slower growth kinetics and so needed a longer period of time to reach high titers. The viruses with deletions in the PB1 segment could not be generated in normal 293T and MDCK-SIAT1 cells, since they required the exogenous expression of the PB1 protein. Therefore, these viruses were generated in previously described 293T and MDCK-SIAT1 cells that had been engineered to constitutively express PB1 (87). These viruses were harvested from transfections at 72 hours, and passaged twice in the MDCK-SIAT1 cells constitutively expressing PB1 at a MOI of 0.001 for 72 hours and 0.01 for 48 hours. This passaging was necessary as viral titers from transfections were too low to generate sufficient virus from a single passage. The wild-type and NS1stop viruses were titered by TCID50 on MDCK-SIAT1 cells, and the PB1 deletion viruses were titered on the MDCK-SIAT1 cells constitutively expressing PB1. The infections in Fig. S12B were performed at equivalent TCID50s as described in the legend to that figure. That these equivalent TCID50s were also roughly equivalent in terms of particles capable of undergoing primary transcription is shown in Fig. S13.

Computational analysis of single-cell transcriptomic and viral sequence data. A computational pipeline that performs all steps in the data analysis is available at https://github.com/jbloomlab/IFNsorted_flu_single_cell. This pipeline is orchestrated by Snakemake (88), and begins with the raw sequencing data and ends by generating the figures shown in this paper. The sequencing data and annotated cell-gene matrix are available on the GEO repository under accession GSE120839 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE120839).

Briefly, the raw deep sequencing data from the Illumina 3′-end sequencing were processed

using the 10X Genomics software package cellranger (version 2.2.0). We built a multi-species alignment reference consisting of a concatenation of the human and influenza virus transcriptomes (the first "species") and the canine transcriptome (the second "species"). The human transcriptome was generated by filtering genome assembly GRCh38 for protein-coding genes defined in GTF file GRCh38.87. The influenza virus transcriptome consisted of the mRNAs for the wild-type A/WSN/1933 virus strain in File S2 (the cellranger alignment is sufficiently permissive that it aligns sequences from both the wild-type and synonymously barcoded viral variants to this transcriptome). The canine transcriptome was generated by filtering genome assembly CanFam3.1 for protein-coding genes defined in GTF file CanFam3.1.87. The cellranger software was used to align the Illumina 3'-end sequencing reads to this multi-species transcriptome, call human+influenza and canine cells (Fig. 3A), and generate a matrix giving the expression of each gene in each single cell. We used a custom Python script to determine the number of influenza virus reads that could be assigned to the wild-type or synonymously barcoded virus, and added this information to the annotated the cell-gene matrix.

The PacBio sequences of the full-length viral genes were analyzed as follows. First, we used version 3.1.0 of PacBio's ccs program (https://github.com/PacificBiosciences/unanimity) to build circular consensus sequences (CCSs) from the subreads files, requiring at least 3 passes and a minimum accuracy of 0.999. We further processed these CCSs using custom Python code and the minimap2 (89) long-read aligner (version 2.11-r797). The Python code has been implemented in the API of dms_tools2 (https://jbloomlab.github.io/dms_tools2/ (90)) package (version 2.3.0). A Jupyter notebook that performs these analyses is at https://github.com/jbloomlab/IFNsorted_flu_single_cell/blob/master/pacbio_analysis.ipynb, and is also provided in HTML form as File S7. We refer the reader to this notebook for a detailed description and extensive plots showing the results at each step. Here is a brief summary: we filtered for CCSs that had the expected 5' termini (from the influenza-specific primers) and 3' termini (corresponding to the 10X adaptor), and for which we could identify the cell barcode, UMI, and polyA tail. We aligned the cDNAs flanked by these termini to the influenza transcriptome, and performed a variety of quality control steps. At this point, we examined whether cDNAs had the synonymous viral barcodes at both ends or neither end as expected in the absence of strand exchange (Fig. S10), and reassuringly found that strand exchange was rare (Fig. S9). The small number of CCSs with identifiable strand exchange were filtered from further analysis. We then further filtered for CCSs that contained valid cell barcodes as identified by the cellranger pipeline, and kept just one CCS per UMI (preferentially retaining high-quality CCSs that aligned to full-length cDNAs). We then removed from the CCSs the barcoding synonymous mutations that we had engineered into one of the two viral variants. Finally, we used the CCSs to call the sequence of the viral gene in each cell, calling mutations separately for each viral barcode variant. We called mutations (insertions, deletions, and substitutions) in the viral gene sequences as follows:

1. Mutations with accuracies less than 0.999 (which constitute <0.5% of all mutations) were ignored.
2. If all CCSs for a particular viral-barcode variant of a gene in a cell were wild-type, it was called as wild type.
3. If any CCSs for a particular viral-barcode variant of gene in a cell had a mutation, then require at least two CCSs to call the sequence.
4. If at least two and >30% of the CCSs had a specific mutation, then call that mutation as present and note its frequency among the CCSs. The exception was single-nucleotide indels in homopolymers, for which we required three CCSs to call a mutation (the reason is that the main mode of PacBio sequencing errors is short indels in homopolymers).

The plots in File S7 indicate that these are reasonable mutation-calling criteria. We could call the sequences of all expressed viral genes in about half of the infected cells (Fig. S11). The mutations called using this pipeline are shown in Fig. 4, and File S4 gives the number of CCSs supporting each mutation call. The called sequences of the viral genes were added to the annotated cell-gene matrix.

Finally, we process the annotated cell-gene matrix in R to generate the plots shown in this paper. This analysis utilized a variety of R and Bioconductor (91) packages, including Monocle (92, 93) and ggplot2. A Jupyter notebook that performs these analyses is at https:

Russell et al.

//github.com/jbloomlab/IFNsorted_flu_single_cell/blob/master/monocle_analysis.ipynb, and is also provided in HTML form as File S8. We refer the reader to this notebook for a detailed description and a variety of additional plots not included in the paper. Briefly, we first filtered cells that were extreme outliers in the amount of mRNA as shown in Fig. 3B. We used the uninfected canine cells to estimate the percentage of total mRNA in a cell that would come from influenza purely due to background (e.g., from cell lysis) in the absence of infection, and called as infected the human cells for which significantly more than this amount of mRNA was derived from influenza under a Poisson model (Fig. 3C). We next used a Poisson model parameterized by the amount of expected background mRNA for each influenza gene to call the presence or absence of each influenza gene in each infected cell (Fig. 3D and Fig. S4). To identify cells that were co-infected with both viral barcodes (Fig. 3F), we used a binomial test to identify cells for which we could reject the null hypothesis that at least 95% of viral mRNA was derived from the more common viral barcode. We called IFN+ and ISG+ cells using the heuristic thresholds shown in Fig. 3G and Fig. S6, respectively. We counted IFN mRNAs as any IFN-α, IFN-β, or IFN-λ transcripts. We counted ISG mRNAs as any of CCL5, IFIT1, ISG15, or Mx1. The plot in Fig. 4 summarizes all of the genotypic information, and was created in substantial part using gggenes (https://github.com/wilkox/gggenes).

Structural analysis of PB1 mutants. To locate the PB1 mutations in the influenza A virus RNA polymerase structure relative to the template and active site in Fig. 8A, we superposed the bat influenza A virus RNA polymerase structure (PDB 4WSB) (74), which shows the 3′ terminus of the template on the surface of the RNA polymerase, with the influenza B virus transcription initiation complex (PDB 5MSG) (75), which shows the 3′ terminus of the template in the template entry channel that leads towards the active site. The structural alignment was performed in Pymol 1.8.7 using motifs A and C.

Experimental analysis of PB1 mutants. For the experimental analysis of the PB1 mutants in Fig. 8, we used plasmids pcDNA-PB1, pcDNA-PA, pcDNA-PB2, and pcDNA-NP, which encode the WSN proteins that compose the polymerase complex (94); pPolI-NA, which encodes the viral RNA for the WSN NA (94); and pcDNA-PB1a, which encodes an inactive version of the WSN PB1 polymerase protein (95). To construct plasmids expressing mutant PB1 proteins D27N and T677A, the plasmid pcDNA-PB1 was subjected to site-directed mutagenesis. PB1 expression was analysed by western blot using antibody GTX125923 (GeneTex).

To analyze the activity of the PB1 mutants in cell culture, the plasmids expressing the WSN PA, PB2, NP and PB1 proteins were transfected into 293T cells together with the plasmid expressing the wildtype NA vRNA or a 246-nt long segment NP-based template (35). Twenty-four hours post transfection, the RNA was extracted using Trizol (Invitrogen), and the steady state RNA levels assessed using reverse transcription with 32P-labelled oligonucleotides against the viral RNA species and ribosomal 5S RNA as described previously (35, 96). 32P-derived signals were imaged using phosphorimaging on a Typhoon scanner and analysed using Prism (GraphPad). In all experiments, the apparent RNA levels were background corrected using the PB1a mutant and normalized to the 5S rRNA loading control.

To measure the induction of the IFN-beta promoter during these RNP reconstitution assays, they were carried out in the presence of a plasmid expressing Renilla luciferase from a CMV promoter and a plasmid expressing Firefly luciferase from the IFN-beta promoter (35). Twenty-four hours post transfection, cells were harvested, lysed and analysed using a DualGlo luciferase kit (Promega) according to the manufacturer's instructions. Samples were analysed using a GloMax (Promega).

REFERENCES

1. Stetson DB, Medzhitov R. Type I interferons in host defense. Immunity 2006; 25(3):373–381.

2. Honda K, Takaoka A, Taniguchi T. Type I inteferon gene induction by the interferon regulatory factor family of transcription factors. Immunity 2006; 25(3):349–360.

3. Solov'ev V. The results of controlled observations on the prophylaxis of influenza with interferon. Bull. World Heal. Organ. 1969; 41(3-4-5):683.

4. Treanor JJ, Betts RF, Erb SM, Roth FK, Dolin R. Intranasally administered interferon as prophylaxis against experimentally induced influenza A virus infection in humans. The J. Infect. Dis. 1987; 156(2):379–383.

5. Beilharz MW, Cummins JM, Bennett AL. Protection from lethal influenza virus challenge by oral type 1 interferon. Biochem. Biophys. Res. Commun. 2007; 355(3):740–744.

6. Kugel D, Kochs G, Obojes K, Roth J, Kobinger GP, Kobasa D, Haller O, Staeheli P, Von Messling V. Intranasal administration of alpha interferon reduces seasonal influenza A virus morbidity in ferrets. J. Virol. 2009; 83(8):3843–3851.

7. Steel J, Staeheli P, Mubareka S, García-Sastre A, Palese P, Lowen AC. Transmission of pandemic H1N1 influenza virus and impact of prior exposure to seasonal strains or interferon treatment. J. Virol. 2010; 84(1):21–26.

8. La Gruta NL, Kedzierska K, Stambas J, Doherty PC. A question of self-preservation: immunopathology in influenza virus infection. Immunol. Cell Biol. 2007; 85(2):85–92.

9. Iwasaki A, Pillai PS. Innate immunity to influenza virus infection. Nat. Rev. Immunol. 2014; 14(5):315.

10. 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. J. Virol. 2013; 87(12):6925–6930.

11. Killip MJ, Jackson D, Pérez-Cidoncha M, Fodor E, Randall RE. Single-cell studies of IFN-β promoter activation by wild-type and NS1-defective influenza A viruses. J. Gen. Virol. 2017; 98(3):357–363.

12. Russell AB, Trapnell C, Bloom JD. Extreme heterogeneity of influenza virus infection in single cells. eLife 2018; 7:e32303.

13. Steuerman Y, Cohen M, Peshes-Yaloz N, Valadarsky L, Cohn O, David E, Frishberg A, Mayo L, Bacharach E, Amit I, Gat-Viks I. Dissection of Influenza Infection In Vivo by Single-Cell RNA Sequencing. Cell systems 2018; 6:679–691.e4.

14. Sjaastad LE, Fay EJ, Fiege JK, Macchietto MG, Stone IA, Markman MW, Shen S, Langlois RA. Distinct antiviral signatures revealed by the magnitude and round of influenza virus replication in vivo. Proc. Natl. Acad. Sci. 2018; 115(38):9610–9615.

15. Heldt FS, Kupke SY, Dorl S, Reichl U, Frensing T. Single-cell analysis and stochastic modelling unveil large cell-to-cell variability in influenza A virus infection. Nat. Commun. 2015; 6:8938.

16. Wang C, Forst CV, Chou Tw, Geber A, Wang M, Hamou W, Smith M, Sebra R, Zhang B, Zhou B, et al. Cell-to-cell variation in defective virus expression and effects on host responses during influenza virus infection. bioRxiv 2018; DOI: 10.1101/364950.

17. McCrone JT, Woods RJ, Martin ET, Malosh RE, Monto AS, Lauring AS. Stochastic processes constrain the within and between host evolution of influenza virus. eLife 2018; 7:e35962.

18. Xue KS, Bloom JD. Reconciling disparate estimates of viral genetic diversity during human influenza infections. Nat. Genet. 2019; DOI: 10.1038/s41588-019-0349-3.

19. Varble A, Albrecht RA, Backes S, Crumiller M, Bouvier NM, Sachs D, García-Sastre A, et al. Influenza A virus transmission bottlenecks are defined by infection route and recipient host. Cell Host & Microbe 2014; 16(5):691–700.

20. Shalek AK, Satija R, Adiconis X, Gertner RS, Gaublomme JT, Raychowdhury R, Schwartz S, Yosef N, Malboeuf C, Lu D, et al. Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. Nature 2013; 498(7453):236–241.

21. Shalek AK, Satija R, Shuga J, Trombetta JJ, Gennert D, Lu D, Chen P, Gertner RS, Gaublomme JT, Yosef N, et al. Single cell RNA Seq reveals dynamic paracrine control of cellular variation. Nature 2014; 510(7505):363.

22. Wimmers F, Subedi N, van Buuringen N, Heister D, Vivié J, Beeren-Reinieren I, Woestenenk R, Dolstra H, Piruska A, Jacobs JF, et al. Single-cell analysis reveals that stochasticity and paracrine signaling control interferon-alpha production by plasmacytoid dendritic cells. Nat. Commun. 2018; 9(1):3317.

23. Bhushal S, Wolfsmüller M, Selvakumar T, Kemper L, Wirth D, Hornef M, Hauser H, Köster M. Cell Polarization and Epigenetic Status Shape the Heterogeneous Response to Type III Interferons in Intestinal Epithelial Cells. Front. Immunol. 2017; 8:671–671.

24. García-Sastre A, Egorov A, Matassov D, Brandt S, Levy DE, Durbin JE, Palese P, Muster T. Influenza A virus lacking the NS1 gene replicates in interferon-deficient systems. Virology 1998; 252(2):324–330.

25. Hale BG, Randall RE, Ortín J, Jackson D. The multifunctional NS1 protein of influenza A viruses. J. Gen. Virol. 2008; 89(10):2359–2376.

26. Hayashi T, MacDonald LA, Takimoto T. Influenza A virus protein PA-X contributes to viral growth and suppression of the host antiviral and immune responses. J. Virol. 2015; 89(12):6442–6452.

27. Vreede FT, Chan AY, Sharps J, Fodor E. Mechanisms and functional implications of the degradation of host RNA polymerase II in influenza virus infected cells. Virology 2010; 396(1):125–134.

28. Dudek SE, Wixler L, Nordhoff C, Nordmann A, Anhlan D, Wixler V, Ludwig S. The influenza virus PB1-F2 protein has interferon antagonistic activity. Biol. Chem. 2011; 392(12):1135–1144.

29. Killip MJ, Fodor E, Randall RE. Influenza virus activation of the interferon system. Virus research 2015; 209:11–22.

30. Parvin JD, Moscona A, Pan W, Leider J, Palese P. Measurement of the mutation rates of animal viruses: influenza A virus and poliovirus type 1. J. Virol. 1986; 59(2):377–383.

31. Suárez P, Valcárcel J, Ortín J. Heterogeneity of the mutation rates of influenza A viruses: isolation of mutator mutants. J. Virol. 1992; 66(4):2491–2494.

32. Suárez-López P, Ortín J. An estimation of the nucleotide substitution rate at defined positions in the influenza virus haemagglutinin gene. J.

Gen. Virol. 1994; 75(2):389–393.

33. Bloom JD. An experimentally determined evolutionary model dramatically improves phylogenetic fit. Mol. Biol. Evol. 2014; 31(8):1956–1978.

34. Pauly MD, Procario MC, Lauring AS. A novel twelve class fluctuation test reveals higher than expected mutation rates for influenza A viruses. eLife 2017; 6:e26437.

35. te Velthuis A, Long J, Bauer DL, Fan R, Yen HL, Sharps J, Siegers J, Killip M, French H, Oliva-Martin MJ, Randall R, de Wit E, van Riel D, Poon L, Fodor E. Mini viral RNAs act as innate immune agonists during influenza virus infection. Nat. Microbiol. 2018; 3(11):1234–1242.

36. Du Y, Xin L, Shi Y, Zhang TH, Wu NC, Dai L, Gong D, Brar G, Shu S, Luo J, Reiley W, Tseng YW, Bai H, Wu TT, Wang J, Shu Y, Sun R. Genome-wide identification of interferon-sensitive mutations enables influenza vaccine design. Science 2018; 359(6373):290–296.

37. Pérez-Cidoncha M, Killip M, Oliveros J, Asensio V, Fernández Y, Bengoechea J, Randall R, Ortín J. An unbiased genetic screen reveals the polygenic nature of the influenza virus anti-interferon response. J. Virol. 2014; 88(9):4632–4646.

38. Baum A, Sachidanandam R, García-Sastre A. Preference of RIG-I for short viral RNA molecules in infected cells revealed by next-generation sequencing. Proc. Natl. Acad. Sci. 2010; 107(37):16303–16308.

39. Tapia K, Kim Wk, Sun Y, Mercado-López X, Dunay E, Wise M, Adu M, López CB. Defective viral genomes arising in vivo provide critical danger signals for the triggering of lung antiviral immunity. PLoS Pathog. 2013; 9(10):e1003703.

40. Boergeling Y, Rozhdestvensky TS, Schmolke M, Resa-Infante P, Robeck T, Randau G, Wolff T, Gabriel G, Brosius J, Ludwig S. Evidence for a novel mechanism of influenza virus-induced type I interferon expression by a defective RNA-encoded protein. PLoS Pathog. 2015; 11(5):e1004924.

41. Dimmock NJ, Easton AJ. Cloned defective interfering influenza RNA and a possible pan-specific treatment of respiratory virus diseases. Viruses 2015; 7(7):3768–3788.

42. Liu G, Lu Y, Liu Q, Zhou Y. Inhibition of Ongoing Influenza A Virus Replication Reveals Different Mechanisms of RIG-I Activation. J. Virol. 2019; 93:DOI 10.1128/JVI.02066–18.

43. Brooke CB, Ince WL, Wrammert J, Ahmed R, Wilson PC, Bennink JR, Yewdell JW. Most influenza A virions fail to express at least one essential viral protein. J. Virol. 2013; 87:3155–3162.

44. Guo F, Li S, Caglar MU, Mao Z, Liu W, Woodman A, Arnold JJ, Wilke CO, Huang TJ, Cameron CE. Single-cell virology: on-chip investigation of viral infection dynamics. Cell Reports 2017; 21(6):1692–1704.

45. Zanini F, Pu SY, Bekerman E, Einav S, Quake SR. Single-cell transcriptional dynamics of flavivirus infection. eLife 2018; 7:e32942.

46. Zanini F, Robinson M, Croote D, Sahoo MK, Sanz AM, Ortiz-Lasso E, Albornoz LL, Suarez FR, Montoya JG, Pinsky BA, Quake S, Einav S. Virus-inclusive single cell RNA sequencing reveals molecular signature predictive of progression to severe dengue infection. PNAS 2018; DOI: 10.1073/pnas.1813819115.

47. Saikia M, Burnham P, Keshavjee SH, Wang MF, Heyang M, Moral-Lopez P, Hinchman MM, Danko CG, Parker JS, De Vlaminck I. Simultaneous multiplexed amplicon sequencing and transcriptome profiling in single cells. Nat. Methods 2019; 16:59–62.

48. O'Neal JT, Upadhyay AA, Wolabaugh A, Patel NB, Bosinger SE, Suthar MS. West Nile virus-inclusive single-cell RNA sequencing reveals heterogeneity in the type I interferon response within single cells. J. Virol. 2019; DOI: 10.1128/JVI.01778-18.

49. Bonini C, Ferrari G, Verzeletti S, Servida P, Zappone E, Ruggieri L, Ponzoni M, Rossini S, Mavilio F, Traversari C, Bordignon C. HSV-TK gene transfer into donor lymphocytes for control of allogeneic graft-versus-leukemia. Science 1997; 276(5319):1719–1724.

50. Ruggieri L, Aiuti A, Salomoni M, Zappone E, Ferrari G, Bordignon C. Cell-surface marking of CD34+-restricted phenotypes of human hematopoietic progenitor cells by retrovirus-mediated gene transfer. Hum. Gene Ther. 1997; 8(13):1611–1623.

51. Strahle L, Garcin D, Kolakofsky D. Sendai virus defective-interfering genomes and the activation of interferon-beta. Virology 2006; 351(1):101–111.

52. Hoffmann E, Neumann G, Kawaoka Y, Hobom G, Webster RG. A DNA transfection system for generation of influenza A virus from eight plasmids. Proc. Natl. Acad. Sci. 2000; 97(11):6108–6113.

53. Xue J, Chambers BS, Hensley SE, López CB. Propagation and characterization of influenza virus stocks that lack high levels of defective viral genomes and hemagglutinin mutations. Front. Microbiol. 2016; 7:326.

54. Zheng GX, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, Ziraldo SB, Wheeler TD, McDermott GP, Zhu J, et al. Massively parallel digital transcriptional profiling of single cells. Nat. Commun. 2017; 8:14049.

55. Robertson J, Schubert M, Lazzarini R. Polyadenylation sites for influenza virus mRNA. J. Virol. 1981; 38(1):157–163.

56. Koppstein D, Ashour J, Bartel DP. Sequencing the cap-snatching repertoire of H1N1 influenza provides insight into the mechanism of viral transcription initiation. Nucleic Acids Res. 2015; 43(10):5052–5064.

57. Travers KJ, Chin CS, Rank DR, Eid JS, Turner SW. A flexible and efficient template format for circular consensus sequencing and SNP detection. Nucleic Acids Res. 2010; 38(15):e159.

58. Bloom JD. Estimating the frequency of multiplets in single-cell RNA sequencing from cell-mixing experiments. PeerJ 2018; 6:e5578.

59. Haque A, Engel J, Teichmann SA, Lönnberg T. A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. Genome Medicine 2017; 9(1):75.

60. Dou D, Hernández-Neuta I, Wang H, Östbye H, Qian X, Thiele S, Resa-Infante P, Kouassi NM, Sender V, Hentrich K, Mellroth P, Henriques-Normark B, Gabriel G, Nilsson M, Daniels R. Analysis of IAV replication and co-infection dynamics by a versatile RNA viral genome labeling method. Cell Reports 2017; 20(1):251–263.

61. Jacobs NT, Onuoha NO, Antia A, Antia R, Steel J, Lowen AC. Incomplete influenza A virus genomes are abundant but readily complemented during spatially structured viral spread. bioRxiv 2019; DOI 10.1101/529065.

62. Hatada E, Hasegawa M, Mukaigawa J, Shimizu K, Fukuda R. Control of influenza virus gene expression: quantitative analysis of each viral RNA species in infected cells. The J. Biochem. 1989; 105(4):537–546.

63. Aguilera ER, Erickson AK, Jesudhasan PR, Robinson CM, Pfeiffer JK. Plaques formed by mutagenized viral populations have elevated coinfection frequencies. mBio 2017; 8(2):e02020–16.

64. Combe M, Garijo R, Geller R, Cuevas JM, Sanjuán R. Single-cell analysis of RNA virus infection identifies multiple genetically diverse viral genomes within single infectious units. Cell Host & Microbe 2015; 18(4):424–432.

65. Dhir A, Dhir S, Borowski LS, Jimenez L, Teitell M, Rötig A, Crow YJ, Rice GI, Duffy D, Tamby C, et al. Mitochondrial double-stranded RNA triggers antiviral signalling in humans. Nature 2018; 560(7717):238–242.

66. Benitez AA, Panis M, Xue J, Varble A, Shim JV, Frick AL, López CB, Sachs D, et al. In vivo RNAi screening identifies MDA5 as a significant contributor to the cellular defense against influenza A virus. Cell reports 2015; 11(11):1714–1726.

67. Chiang J, Sparrer K, Lässig C, Huang T, Osterrieder N, Hopfner K, Gack M, et al. Viral unmasking of cellular 5S rRNA pseudogene transcripts induces RIG-I-mediated immunity. Nat. Immunol. 2018; 19(1):53–62.

68. Maaten Lvd, Hinton G. Visualizing data using t-SNE. J. Mach. Learn. Res. 2008; 9(Nov):2579–2605.

69. Xu H, DiCarlo J, Satya RV, Peng Q, Wang Y. Comparison of somatic mutation calling methods in amplicon and whole exome sequence data. BMC Genomics 2014; 15(1):244.

70. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, Gabriel S, Meyerson M, Lander ES, Getz G. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat. Biotechnol. 2013; 31(3):213–222.

71. Gini C. Measurement of inequality of incomes. The Econ. J. 1921; 31(121):124–126.

72. Fujii Y, Goto H, Watanabe T, Yoshida T, Kawaoka Y. Selective incorporation of influenza virus RNA segments into virions. Proc. Natl. Acad. Sci. 2003; 100(4):2002–2007.

73. Marsh GA, Hatami R, Palese P. Specific residues of the influenza A virus hemagglutinin viral RNA are important for efficient packaging into budding virions. J. Virol. 2007; 81(18):9727–9736.

74. Reich S, Guilligay D, Pflug A, Malet H, Berger I, Crépin T, Hart D, Lunardi T, Nanao M, Ruigrok RW, et al. Structural insight into cap-snatching and RNA synthesis by influenza polymerase. Nature 2014; 516(7531):361–366.

75. Reich S, Guilligay D, Cusack S. An in vitro fluorescence based study of initiation of RNA synthesis by influenza B polymerase. Nucleic Acids Res. 2017; 45(6):3353–3368.

76. Zhu Y, Yongky A, Yin J. Growth of an RNA virus in single cells reveals a broad fitness distribution. Virology 2009; 385(1):39–46.

77. Schulte M, Andino R. Single-cell analysis uncovers extensive biological noise in poliovirus replication. J. Virol. 2014; 88(11):6205–6212.

78. Akpinar F, Timm A, Yin J. High-throughput single-cell kinetics of virus infections in the presence of defective interfering particles. J. Virol. 2016; 90(3):1599–1612.

79. Gupta I, Collier PG, Haase B, Mahfouz A, Joglekar A, Floyd T, Koopmans F, Barres B, Smit AB, Sloan S, et al. Single-cell isoform RNA sequencing (ScISOr-Seq) across thousands of cells reveals isoforms of cerebellar cell types. Nat. Biotechnol. 2018; 36:1197–1202.

80. O'Connell RM, Balazs AB, Rao DS, Kivork C, Yang L, Baltimore D. Lentiviral vector delivery of human interleukin-7 (hIL-7) to human immune system (HIS) mice expands T lymphocyte populations. PLoS One 2010; 5(8):e12009.

81. Kim JH, Lee SR, Li LH, Park HJ, Park JH, Lee KY, Kim MK, Shin BA, Choi SY. High cleavage efficiency of a 2A peptide derived from porcine teschovirus-1 in human cell lines, zebrafish and mice. PloS One 2011; 6(4):e18556.

82. Reed LJ, Muench H. A simple method of estimating fifty per cent endpoints. Am. J. Epidemiol. 1938; 27(3):493–497.

83. Doud MB, Hensley SE, Bloom JD. Complete mapping of viral escape from neutralizing antibodies. PLoS Pathog. 2017; 13(3):e1006271.

84. Boers SA, Hays JP, Jansen R. Micelle PCR reduces chimera formation in 16S rRNA profiling of complex microbial DNA mixtures. Sci. Reports 2015; 5:14181.

85. 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. J. Virol. 2014; 88(8):3942–3952.

86. Hirst GK. The quantitative determination of influenza virus and antibodies by means of red cell agglutination. J. Exp. Medicine 1942; 75(1):49–64.

87. Bloom JD, Gong LI, Baltimore D. Permissive Secondary Mutations Enable the Evolution of Influenza Oseltamivir Resistance. Science 2010; 328(5983):1272–1275.

88. Köster J, Rahmann S. Snakemake—a scalable bioinformatics workflow engine. Bioinformatics 2012; 28(19):2520–2522.

89. Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 2018; 1:7.

90. Bloom JD. Software for the analysis and visualization of deep mutational scanning data. BMC Bioinforma. 2015; 16(1):168.

91. Huber W, Carey VJ, Gentleman R, Anders S, Carlson M, Carvalho BS, Bravo HC, Davis S, Gatto L, Girke T, Gottardo R, Hahne F, Hansen KD, Irizarry RA, Lawrence M, Love MI, MacDonald J, Obenchain V, Oles AK, Pages H, Reyes A, Shannon P, Smyth GK, Tenenbaum D, Waldron L, Morgan M. Orchestrating high-throughput genomic analysis with Bioconductor. Nat. Methods 2015; 12(2):115.

92. Qiu X, Mao Q, Tang Y, Wang L, Chawla R, Pliner HA, Trapnell C. Reversed graph embedding resolves complex single-cell trajectories. Nat. Methods 2017; 14(10):979.

93. Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, Lennon NJ, Livak KJ, Mikkelsen TS, Rinn JL. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat. Biotechnol. 2014; 32(4):381.

94. Fodor E, Crow M, Mingay LJ, Deng T, Sharps J, Fechter P, Brownlee GG. A single amino acid mutation in the PA subunit of the influenza virus RNA polymerase inhibits endonucleolytic cleavage of capped RNAs. J. Virol. 2002; 76(18):8989–9001.

95. Vreede FT, Jung TE, Brownlee GG. Model suggesting that replication of influenza virus is regulated by stabilization of replicative intermediates. J. Virol. 2004; 78(17):9568–9572.

96. te Velthuis AJ, Robb NC, Kapanidis AN, Fodor E. The role of the priming loop in influenza A virus RNA synthesis. Nat. Microbiol. 2016; 1(5):16029.