



Published in final edited form as:

*Nat Microbiol.* 2020 September ; 5(9): 1158–1169. doi:10.1038/s41564-020-0749-2.

## Collective interactions augment influenza A virus replication in a host-dependent manner

Kara L. Phipps<sup>a</sup>, Ketaki Ganti<sup>a,‡</sup>, Nathan T. Jacobs<sup>a,‡</sup>, Chung-Young Lee<sup>a,‡</sup>, Silvia Carnaccini<sup>b</sup>, Maria C. White<sup>a</sup>, Miglena Manandhar<sup>c</sup>, Brett E. Pickett<sup>c,†</sup>, Gene S. Tan<sup>c,d,f</sup>, Lucas M. Ferreri<sup>b</sup>, Daniel R. Perez<sup>b,e</sup>, Anice C. Lowen<sup>a,d,\*</sup>

<sup>a</sup>Department of Microbiology and Immunology, Emory University School of Medicine, Atlanta, Georgia, United States of America

<sup>b</sup>Department of Population Health, Poultry Diagnostic and Research Center, University of Georgia, Athens, Georgia, United States of America.

<sup>c</sup>J. Craig Venter Institute, La Jolla, California, United States of America

<sup>d</sup>Emory-UGA Center of Excellence for Influenza Research and Surveillance (CEIRS), Atlanta, Georgia, United States of America

<sup>e</sup>Center of Research on Influenza Pathogenesis (CRIP) CEIRS, New York, New York, United States of America.

<sup>f</sup>Division of Infectious Disease, Department of Medicine, University of California San Diego, La Jolla, CA, United States of America

### Abstract

Infection with a single influenza A virus (IAV) is only rarely sufficient to initiate productive infection. Instead, multiple viral genomes are often required within a given cell. Here, we show that IAV reliance on multiple infection can form an important species barrier. Namely, we find that

---

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:[http://www.nature.com/authors/editorial\\_policies/license.html#terms](http://www.nature.com/authors/editorial_policies/license.html#terms)

\*Address correspondence: [anice.lowen@emory.edu](mailto:anice.lowen@emory.edu).

†Present address: Department of Microbiology and Molecular Biology, Brigham Young University, Provo, Utah 84602 USA

#### Author Contributions

KLP contributed to the conception of work, experimental design, data acquisition and analysis, interpretation of data; KG, NTJ and C-YL contributed to experimental design, data acquisition and analysis and interpretation of data; SC contributed to data acquisition; MM and MCW contributed to data acquisition and analysis; BEP contributed to data analysis and interpretation; GST contributed to conception of the work, data analysis and interpretation; LMF contributed to data acquisition; DRP contributed to experimental design, data analysis and interpretation; ACL contributed to conception of work, experimental design and data analysis and interpretation. All authors contributed to the writing of the manuscript.

‡These individuals contributed equally to the research presented.

#### Data availability

10x Genomics single-cell sequencing data is available on the GEO database with the accession number GSE135553. Other data are included as Source Data or are available from the corresponding author upon reasonable request.

#### Code availability

Code used for 10x Genomics analysis is available at: [https://github.com/njacobs627/GFHK99\\_Multiplicity](https://github.com/njacobs627/GFHK99_Multiplicity). Code used to run the agent-based model on influenza A virus reassortment was reported previously<sup>7,8</sup> and is also available at [https://github.com/njacobs627/GFHK99\\_Multiplicity](https://github.com/njacobs627/GFHK99_Multiplicity).

#### Conflict of interest statement

The authors declare no conflicts of interest.

avian H9N2 viruses representative of those circulating widely at the poultry-human interface exhibit acute dependence on collective interactions in mammalian systems. This need for multiple infection is greatly reduced in the natural host. Quantification of incomplete viral genomes showed that their complementation accounts for the moderate reliance on multiple infection seen in avian cells, but not the added reliance seen in mammalian cells. In mammals, an additional form of virus-virus interaction is needed. We find that the PA gene segment is a major driver of this phenotype and that both viral replication and transcription are affected. These data indicate that multiple distinct mechanisms underlie IAV reliance on multiple infection and underscore the importance of virus-virus interactions in IAV infection, evolution and emergence.

---

Classically, an infectious unit has been defined as a single virus particle which delivers its genome to a cell, initiates the viral reproductive program, and yields progeny viruses. Recently, however, the importance of collective interactions among viruses is being increasingly recognized<sup>1-3</sup>. The delivery of multiple viral genomes to a cell allows antagonistic and beneficial interactions which have the potential to shape transmission, pathogenicity and evolution.

In the case of IAV, several lines of evidence point to a major role in infection for complementation between genomes lacking one or more functional segments<sup>4-8</sup>. Conversely, deleteriously mutated genes act in a dominant negative fashion, suppressing production of infectious progeny<sup>9-11</sup>. Importantly, multiple infection with identical viral genomes can also alter infection outcomes<sup>8,12-14</sup> and may be particularly relevant for facilitating viral growth under adverse conditions, such as antiviral drug treatment<sup>3,15</sup>. For IAV, an important adverse condition to consider is that of a novel host environment. IAVs occupy a broad host range<sup>16,17</sup>, but host barriers to infection typically confine a given lineage to one or a small number of related species<sup>18,19</sup>. Spillovers occur occasionally, however. When a novel IAV lineage is established in humans, the result is a pandemic of major public health consequence<sup>20,21</sup>.

Here, we examined the multiplicity dependence of human and avian-origin viruses in different host systems and found that the extent to which viral progeny production is concentrated within multiply-infected cells varies greatly with virus-host context. Our data point to an important role for collective interactions in determining the potential for IAV replication in diverse hosts.

## Results

### Multiplicity dependence varies with strain and cell type

To evaluate the extent to which IAV relies on multiple infection for productive infection, we initially used reassortment and coinfection as readouts. While singly-infected cells yield only parental progeny, reassortment is highly efficient within coinfecting cells<sup>22</sup>. Monitoring reassortment therefore indicates the proportion of progeny in a bulk virus population that were produced from multiply-infected cells. Coinfections were performed under single-cycle conditions with homologous viruses that differ only by a silent mutation in each segment and the presence of either an HA or HIS epitope tag. The silent mutations allow

identification of reassortant viruses. The epitope tags enable quantification of cellular coinfection. Homologous coinfection partners, termed WT and VAR, were generated in A/ Netherlands/602/2009 (H1N1) (NL09), A/mallard/199106/99 (H3N8) (MaMN99) and A/ guinea fowl/HK/WF10/99 (GFHK99) backgrounds.

Coinfection and reassortment were examined in canine MDCK and chicken DF-1 cells (Figure 1). In MDCK cells, GFHK99 viruses showed a near linear relationship between total HA<sup>+</sup> cells and dual-HA<sup>+</sup> cells, suggesting a strict dependence on multiple infection for HA expression (Figure 1A). By contrast, HA expression resulting from infection with a single strain was more common for GFHK99 in DF-1 cells, MaMN99 in MDCK cells and NL09 in MDCK cells (Figure 1A). The more dependent expression of *any* HA protein is on coinfection, the more linear the relationship between dual-HA<sup>+</sup> cells and HA<sup>+</sup> cells becomes. Conversely, a relationship closer to the Poisson expectation indicates less dependence on coinfection. Among the viruses tested, we found that only GFHK99 virus in MDCK cells exhibits strong linearity (Figure 1B).

In line with observed levels of coinfection, GFHK99 virus exhibits high reassortment in MDCK cells, indicating that nearly all progeny virus is produced from WT-VAR coinfecting cells (Figure 1C). MaMN99 or NL09 viruses infecting MDCK cells show lower reassortment. Moreover, GFHK99 reassortment is markedly reduced in DF-1 compared to MDCK cells (Figure 1C). These results clearly implicate virus-host interactions in determining multiplicity dependence.

That all virus-cell pairings tested show multiplicity dependence is highlighted by comparison of the experimental reassortment data to a theoretical prediction<sup>7</sup> (Figure 1C). This prediction is derived from a model in which singly- and multiply-infected cells make equivalent numbers of progeny. Reassortment is therefore predicted to increase only gradually at low levels of infection where coinfection is rare. The experimental observation that reassortment rapidly reaches high levels indicates that viral progeny production is enriched in the proportion of the infected cell population that is multiply infected.

Frequent reassortment and the linear relationship between overall HA expression and HA co-expression provide two measures of multiplicity dependence. To summarize these two measures and analyze the relationships between the virus-cell pairings tested, we conducted a hierarchical clustering analysis using the parameters of the regression models shown in Figure 1A and 1C. MaMN99 and NL09 viruses in MDCK cells, and GFHK99 virus in DF-1 cells, represent one cluster, while GFHK99 virus in MDCK cells partitioned into its own cluster (Figure 1D). Although evident in all virus-cell pairings, multiple-infection dependence is particularly strong for GFHK99 virus in MDCK cells.

### Strain- and host-specific phenotypes are evident *in vivo*

To determine whether host-dependent reliance on multiple infection extended to *in vivo* models, we performed coinfections in guinea pigs and quail. GFHK99, MaMN99, and NL09 viruses were tested in guinea pigs and GFHK99 virus in quail. To ensure use of comparable effective doses for each virus/host pairing, 50% infectious dose (ID<sub>50</sub>) was determined. Guinea pigs were then infected with 10<sup>2</sup> GPID<sub>50</sub> and quail were infected with 10<sup>2</sup> QID<sub>50</sub>. To

evaluate the frequency of reassortment, plaque isolates from upper respiratory samples were genotyped for each animal on each day. The viruses collected from GFHK99-infected guinea pigs show much higher genotypic diversity throughout the course of infection than viruses isolated from MaMN99- or NL09-infected guinea pigs (Figure 1E) or GFHK99-infected quail (Figure 1F). These data indicate that the virus-host interactions which determine dependence on multiple infection in cell culture extend to *in vivo* infection.

### Multiple infection enhances viral growth

Our reassortment analyses suggest that increased multiplicity augments viral output. As a direct test of this, we infected over a range of MOIs and then measured PFU produced per cell under single-cycle conditions. Input MOIs of less than 1 PFU per cell were found to be non-saturating by flow cytometry in MDCK and DF-1 cells (Extended Data Figure 1). Under these non-saturating conditions, increasing MOI resulted in accelerated growth and higher yield for all virus/cell pairings tested (Extended Data Figure 2A–D). Enhancement of viral amplification was, however, greatest for GFHK99 virus in MDCK cells (Figure 2A).

We reasoned that the cooperative effect observed might result from i) complementation of incomplete viral genomes or ii) a benefit of increased viral genome copy number per cell. In an effort to differentiate between these possibilities, we measured growth of GFHK99 virus in MDCK and DF-1 cells infected over a range of saturating MOIs (Extended Data Figure 1). Here, incomplete viral genomes are unlikely to be prevalent and any benefit of increasing MOI would be attributable to increased genome copies per cell. In DF-1 cells, burst size is constant over an MOI range of 1 to 20 PFU per cell, while in MDCK cells, burst size declines about 4-fold, indicative of an antagonistic interaction (Figure 2B and Extended Data Figure 2E, F). Thus, increasing levels of multiple infection was found to be beneficial at lower MOIs but either neutral or detrimental to viral productivity once saturating levels of infection were achieved. Whether this threshold effect is imposed by a need for complementation or another mechanism sensitive to saturation remained unclear.

### Multiple infection enhances viral RNA replication

To evaluate the impact of multiple infection on genome replication, we measured WT viral RNA synthesis in the absence and presence of increasing amounts of homologous coinfecting virus. We also broadened the panel of viruses examined to include A/duck/HK/448/78 (H9N2) (DkHK78), a wild bird isolate that predates the establishment of H9N2 viruses in poultry, and A/quail/HK/88 (H9N2) (Qa/HK/88), an early isolate of the H9N2 poultry lineage<sup>23</sup>. Mammalian or avian cells were infected at low MOI with WT virus to ensure receipt of a single WT virus genome and, concurrently, with increasing doses of VAR virus. Replication of WT genomes was then quantified using primer/probe sets that do not detect VAR templates.

The results for NL09 and MaMN99 viruses show little to no increase in WT vRNA production with VAR virus addition, in all cell types tested (Figure 3A and B). DkHK78 virus exhibits an intermediate phenotype, with 7-fold enhancement in mammalian cells and 1.6-fold enhancement in avian cells (Figure 3C). GFHK99 and QaHK88 viruses show much stronger increases of WT vRNA replication with coinfection: in MDCK and NHBE cells,

respectively, GFHK99 virus shows increases of 60-fold and 20-fold, while QaHK88 virus shows increases of >2000-fold and 200-fold. In contrast, these two poultry isolates show markedly less enhancement in avian DF-1 cells (Figure 3D and E). Of note, addition of VAR at high doses suppresses WT vRNA production in several virus/cell pairings, likely indicating competition for limited resources in the cell. This antagonistic interaction is stronger in systems where little benefit of coinfection is detected.

In summary, for GFHK99 and QaHK88 viruses, the introduction of coinfecting virus reveals a cooperative effect acting at the level of RNA synthesis, which is particularly potent in mammalian cells. That the replication of NL09 and MaMN99 viruses is not detectably enhanced by the addition of VAR virus suggests reduced reliance on multiple infection, consistent with reassortment and dual-HA positivity data presented above.

### **The PA segment is a major determinant of multiple-infection dependence**

To identify viral determinants of multiple-infection dependence, we mapped segments responsible for the high reassortment of GFHK99 virus in MDCK cells. Segments of GFHK99 virus were placed into a MaMN99 background and, for each chimeric genotype, WT and VAR strains were generated. Coinfections were performed in MDCK cells and HA expression and reassortment were measured. Two viruses are similar to the parental GFHK99 virus in the levels of dual HA positivity observed: MaMN99:GFHK99 3PNP and MaMN99:GFHK99 PA (Figure 4A–B). Quantification of reassortment revealed that several chimeric viruses reassort at a higher frequency than MaMN99, but that the MaMN99:GFHK99 3PNP and MaMN99:GFHK99 PA viruses show the highest reassortment, comparable to that seen for the parental GFHK99 genotype (Figure 4C–D). Clustering analysis revealed that only these two genotypes group with GFHK99 parental virus, while each of the other genotypes groups with MaMN99 virus (Figure 4E). Thus, while other viral genes may contribute, PA is the primary genetic determinant of the high reassortment exhibited by GFHK99 virus in MDCK cells and defines a need for cooperation between coinfecting viruses. To verify the contribution of the PA segment by a second approach, we evaluated the impact of VAR virus coinfection on WT vRNA synthesis for the MaMN99:GFHK99 PA strain. The results confirm that the GFHK99 PA segment increases reliance on multiple infection, specifically in mammalian cells (Extended Data Figure 3).

### **Multiple infection accelerates viral replication and transcription**

Because genetic mapping implicated PA, we tested the effects of multiple infection on viral replication and transcription by measuring accumulation of vRNA and mRNA following low or high MOI infection (Extended Data Figure 4). GFHK99 virus was examined in DF-1 and MDCK cells and MaMN99 virus was tested in MDCK cells. Under low MOI conditions in MDCK cells, GFHK99 vRNA levels remain low throughout the time course. GFHK99 mRNA increases early, perhaps reflecting robust primary transcription, then accumulates relatively slowly. Both viral RNA species accumulate at a significantly higher rate for GFHK99 virus in DF-1 cells and MaMN99 virus in MDCK cells. At high MOI, accumulation of GFHK99 mRNA and vRNA in MDCK cells occurs at a similar rate to that seen for GFHK99 virus in DF-1 cells or MaMN99 virus in MDCK cells. Thus, a host-specific defect in GFHK99 polymerase activity that affects both replication and transcription

is seen at low MOI. This defect is, however, resolved under conditions where multiple infection is prevalent.

### Single-cell mRNA sequencing reveals a need for cooperation

To assess the heterogeneity of viral RNA synthesis, we evaluated GFHK99 virus-infected DF-1 and MDCK cells by single-cell mRNA sequencing. We used a thresholding approach to exclude cells that were unlikely to be infected, but may carry small amounts of viral mRNA owing to lysis of neighboring cells (Extended Data Figure 5). To allow comparison between cell types and experiments, each cell's total transcript abundance was normalized to the median number of transcripts per cell in the relevant sample. The normalized and  $\log_{10}$  transformed counts of viral transcripts per cell were then compared across conditions. We found that the amount of detected GFHK99 viral mRNA varies widely between individual DF-1 cells, consistent with previous observations<sup>24–26</sup>. In contrast, GFHK99 viral mRNA levels are uniformly low in MDCK cells, especially at the lower MOIs tested (Figure 5A, left facet).

To evaluate whether low transcript abundance corresponded with failure to detect polymerase-encoding segments, we stratified the data based on detection of PB2, PB1, PA and NP. Consistent with prior reports<sup>24</sup>, viral transcript levels are higher in cells that contained PB2, PB1, PA and NP transcripts compared to those in which one or more of these transcripts was not detected (Extended Data Figure 6). We noted that PB2, PB1 and PA mRNAs are markedly less abundant than the other viral transcripts, however. Thus, the correspondence between failure to detect polymerase transcripts and low viral mRNA abundance may simply be due to the higher likelihood of these mRNAs falling below the limit of detection in cells that have low overall viral transcription.

To measure the impact of coinfecting viruses in individual cells, we repeated the single-cell sequencing experiment with the addition of two variants, GFHK99 mVAR<sub>1</sub> and GFHK99 mVAR<sub>2</sub>. Cells were inoculated with GFHK99 WT and GFHK99 mVAR<sub>1</sub> viruses in a 1:1 ratio. Simultaneous coinfection with GFHK99 mVAR<sub>2</sub> virus was performed at the concentration found to give optimal support of WT RNA replication in Figure 3D: 0.1 PFU per cell in DF-1 cells and 1.0 PFU per cell in MDCK cells. After mRNA sequencing, only cells in which transcripts from all eight mVAR<sub>2</sub> segments were detected were analyzed further and cells that likely obtained viral RNA through lysis were excluded (Extended Data Figure 5). In cells infected with both WT and mVAR<sub>1</sub> viruses, viral transcript abundance is calculated separately for each, so that coinfecting cells contribute two distinct measurements to the dataset. Viral transcript levels per cell are shown in Figure 5A alongside data from the first experiment. In comparing the two infections, we observe that total viral transcript abundance is markedly lower in MDCK compared to DF-1 cells in the first infection, but this effect is almost entirely mitigated by the presence of mVAR<sub>2</sub> virus in the second infection. This reduction in the disparity between DF-1 and MDCK cells resulted from the fact that mVAR<sub>2</sub> virus increased transcript abundance by 149% in DF-1 cells, but 258% in MDCK cells ( $p < 10^{-4}$ , linear mixed effects model) (Figure 5A). These data underscore the significance of collective interaction to ensure productive infection in diverse hosts.



Because only a subset of a cell's transcripts is captured and therefore reliably detected<sup>27</sup>, single cell mRNA sequencing does not allow a robust determination of segment presence or absence in a cell. The addition of mVAR<sub>2</sub> virus improves the sensitivity with which WT transcripts can be detected, however, and – as noted above – levels the playing field between MDCK and DF-1 cells. We therefore evaluated the number of distinct viral gene segments detected per cell, as indicated by the presence of the corresponding mRNAs. Failure to detect a subset of segments is common in our datasets, as with others<sup>24,26</sup>. Addition of mVAR<sub>2</sub> virus increases the number of segments per cell detected, however, suggesting that failure to detect is often a result of the limit of detection (Figure 5B). Importantly, in the presence of mVAR<sub>2</sub> virus, segments detected per cell was comparable between MDCK and DF-1 cells across all MOIs tested (Figure 5B). Thus, the frequency with which incomplete GFHK99 viral genomes are expressed is similar between these mammalian and avian cell lines.

### Frequency of incomplete GFHK99 genomes is moderate

Given the limitations of single-cell mRNA sequencing for the detection of rare transcripts, we employed a second single-cell approach to quantify the frequency with which fewer than eight vRNAs are replicated in GFHK99-infected MDCK cells<sup>8</sup>. MDCK cells were coinfecting with a low MOI of GFHK99 WT virus and a high MOI of GFHK99 VAR<sub>2</sub> virus. Single cells were then sorted onto a naïve cell monolayer and multicycle replication was allowed. During this time, the GFHK99 VAR<sub>2</sub> virus propagates WT segments, even when less than the full WT genome is available. To determine which viral gene segments were present in the initially sorted cell, RT-qPCR with primers that differentiate WT and VAR<sub>2</sub> gene segments was applied. The results were then used to calculate the probability that a cell infected with a single WT virus contains a given segment. The observed Probability Present ( $P_p$ ) varies among the segments from 0.56 to 0.8 (Figure 6A). The product of the eight  $P_p$  values gives an estimate of the proportion of singular infections in which all segments are replicated. This estimate is 3.2% for GFHK99 virus in MDCK cells.

We used these data together with our model of IAV coinfection and reassortment<sup>7</sup> to evaluate whether incomplete viral genomes account for the focusing of GFHK99 virus production within multiply infected cells. In the model, the frequency of successful segment replication is governed by eight  $P_p$  parameters and an infected cell only produces virus if at least one copy of all eight segments are replicated. Importantly, the amount of virus produced from productively infected cells is constant in this model – there is no additional benefit to multiple infection. When the eight experimentally determined  $P_p$  values for GFHK99 virus in MDCK cells are used as parameters, the theoretical prediction of reassortment frequency is much lower than that observed for GFHK99 viruses in MDCK cells, but a good match with the reassortment seen in DF-1 cells (Figure 6B). Clearly, the frequency with which GFHK99 segments fail to be replicated cannot fully account for the high reassortment seen in MDCK cells, but can account for the more moderate reliance on multiple infection seen in DF-1 cells. Thus, in mammalian cells, a full GFHK99 viral genome is necessary but not sufficient to support robust replication.

## Discussion

Our data reveal that moderate reliance of IAV on multiple infection is the norm. An exceptionally high need for multiple infection can, however, occur when an IAV infects a new species. Thus, our findings identify increased reliance on multiple infection as a novel component of IAV host restriction. Since not all mismatched virus-host combinations show a high reliance on multiple infection, the data indicate that a need for collective interactions is a potential manifestation of IAV host restriction, not a universal feature. Dependence on multiple infection is of particular interest as a barrier to cross-species transfer for two reasons: first, it can be overcome in the absence of genetic adaptation, through high dose infection; and second, it leads to high levels of reassortment, which in turn can facilitate adaptation to a new host.

As has been seen for other IAVs<sup>5,8,24,26</sup>, complementation of incomplete genomes is often needed for productive GFHK99 infection. Additional cooperative interactions are, however, evident in mammalian cells. These data point to a model in which the presence of not just complete genomes, but rather multiple copies of the viral genome, are needed to overcome host-specific barriers to infection (Figure 6C). Our data suggest that this phenotype is a feature of poultry-adapted H9N2 viruses; whether a need for multiple infection in mammalian cells extends to other poultry-adapted lineages is of interest for future study. H9N2 viruses are, however, highly relevant in the context of zoonotic infection due to their prevalence at the poultry-human interface and their shared gene pool with H5N1 and H7N9 viruses that have caused hundreds of severe human infections<sup>28–33</sup>.

We previously showed that localized spread within a tissue facilitates complementation of incomplete viral genomes, allowing relatively efficient conversion of non-productive infection events to productive ones<sup>34</sup>. In the context of transmission between hosts, however, the potential for multiple infection may be limited. Indeed, the tight genetic bottleneck observed in human–human transmission is consistent with infection initiated by single particles<sup>35,36</sup>. In line with this concept, our prior work in a guinea pig transmission model revealed that a seasonal IAV engineered to be fully reliant on multiple infection replicated efficiently but did not transmit to contacts<sup>34</sup>. Thus, we favor a model in which transmission of IAV among native hosts is typically mediated by rare single infections that result in productive infection. Conversely, our present work with GFHK99 and related viruses suggests that this mechanism may not be operational in the context of zoonotic transmission, owing to more extreme inefficiency of single infections. Here, the coordinated transfer of multiple virions between hosts (e.g. in a single respiratory droplet) may be critical to overcome species barriers.

Our data implicate the PA segment, which encodes PA and PA-X<sup>37</sup>, in defining an acute reliance on cooperation. PA is an essential component of the polymerase<sup>38</sup>. It carries the endonuclease required for cap-snatching, an activity essential for viral transcription<sup>39</sup>. PA also comprises much of the interface for viral polymerase dimerization, which is needed for viral replication<sup>40</sup>. PA-X contributes to the shut-off of host protein synthesis<sup>41,42</sup>. Alignment of these proteins encoded by GFHK99 and MaMN99 viruses reveals 29 differences in PA and 9 in PA-X (Extended Data Figure 7). Three-dimensional modeling based on a recently



reported structure<sup>40</sup> did not predict marked differences. Nevertheless, polymorphisms noted in and near the dimerization loop suggest a model in which the reliance of GFHK99 virus on multiple infection may result from an inefficiency of dimerization and consequent inefficiency of vRNA synthesis<sup>40</sup>. This model is attractive in that aberrant *in vitro* initiation of vRNA synthesis by a dimerization mutant was ameliorated with increased polymerase concentration<sup>40</sup>. We speculate that multiple infection could lead to a similar effect within the cell. Finer mapping of the phenotype within PA and targeted functional studies are needed to test this model and explore other possible mechanisms.

Importantly, we show that viral cooperation is not limited to complementation of defects; rather, interactions among homologous coinfecting viruses can be highly biologically significant. Such cooperation is expected to have a strong impact on viral fitness and evolution by modulating the efficiency of spread within and between hosts and by changing how a virus population samples sequence space<sup>43–46</sup>. Thus, multiple infection dependence is likely to play an important role in determining the outcomes of IAV infection and evolution in diverse hosts.

## Methods

### Cells and cell culture media

Madin-Darby canine kidney (MDCK) cells, a gift from Peter Palese, Icahn School of Medicine at Mount Sinai, were used in all experiments. MDCK cells from Daniel Perez at University of Georgia were used for plaque assays as this variant of the MDCK line was found to yield more distinct plaques for the GFHK99 strain. Both MDCK cell lines were maintained in minimal essential medium (MEM; Gibco) supplemented with 10% fetal bovine serum (FBS; Atlanta Biologicals), penicillin (100 IU), and streptomycin (100 µg per mL) (PS; Corning). 293T cells (ATCC CRL-3216) and DF-1 cells (ATCC CRL-12203) were maintained in Dulbecco's minimal essential medium (DMEM; Gibco) supplemented with 10% FBS and PS. Duck embryo cells (ATCC CCL-141) were maintained in Eagle's minimum essential medium (EMEM; ATCC) supplemented with 10% FBS and PS. Normal human bronchial epithelial (NHBE) cells were acquired from Lonza and were amplified and differentiated into air-liquid interface cultures as recommended by Lonza and described in Danzy et al.<sup>47</sup>. All cells were cultured at 37°C and 5% CO<sub>2</sub> in a humidified incubator. Cell lines were not authenticated. All cells were tested monthly for mycoplasma contamination while in use. Medium for culture of IAV in each cell line (virus medium) was prepared by supplementing the appropriate media with 4.3% bovine serum albumin and penicillin (100 IU), and streptomycin (100 µg per mL). Ammonium chloride-containing virus medium was prepared by the addition of HEPES buffer and NH<sub>4</sub>Cl at final concentrations of 50 mM and 20 mM, respectively.

### Viruses

All viruses were generated by reverse genetics<sup>48</sup>. For avian viruses, 293T cells transfected with reverse genetics plasmids 16–24 h prior were injected into the allantoic cavity of 9–11 day old embryonated chicken eggs and incubated at 37°C for 40–48 h. The resultant egg passage 1 stocks were used in experiments. For NL09-based viruses, 293T cells transfected

with reverse genetics plasmids 16–24 h prior were co-cultured with MDCK cells at 37°C for 40–48 h. Supernatants were then propagated in MDCK cells from low MOI to generate NL09 working stocks. Defective interfering segment content of PB2, PB1, and PA segments was confirmed to be minimal for each virus stock, as described previously<sup>49</sup> (Extended Data Figure 8). All plaque assays were performed in MDCK cells; viruses were also titered by flow cytometry and/or RT qPCR-based methods where indicated. The reverse genetics system for influenza A/guinea fowl/Hong Kong/WF10/99 (H9N2) virus was reported previously<sup>50,51</sup>. This strain has been referred to as WF10 in previous publications<sup>50–52</sup>. For consistency with other strains used in the present manuscript, it is referred to herein as GFHK99. A low passage isolate of influenza A/mallard/Minnesota/199106/99 (H3N8) virus, referred to herein as MaMN99, was obtained from David Stallknecht at the University of Georgia<sup>53</sup>. The virus was passaged once in eggs and then the eight cDNAs were generated and cloned into the ambisense pDP2002 vector<sup>54</sup>. Reverse genetics systems for influenza A/duck/Hong Kong/448/78 (H9N2) and A/quail/Hong Kong/A28945/88 (H9N2), referred to herein as dkHK78 and QaHK88, were similarly generated following culture in eggs. To increase the recovery efficiency of viruses containing polymerase components from the MaMN99 virus, pCAGGS support plasmids encoding PB2, PB1, PA and NP proteins of the A/WSN/33 (H1N1) strain were supplied.

GFHK99 WT, MaMN99 VAR and NL09 VAR viruses were engineered to contain a 6XHis epitope tag plus GGGs linker at the N-terminus of the HA protein following the signal peptide. GFHK99 VAR<sub>1</sub>, MaMN99 WT and NL09 VAR viruses contain similarly modified HA genes, with an HA epitope tag plus a GGGs linker inserted at the N-terminus of the HA protein<sup>22</sup>. All assays except single-cell mRNA sequencing used viruses carrying these epitope tags.

Silent mutations were introduced into VAR viruses by site-directed mutagenesis to allow genotyping of WT and VAR segment origin. The specific changes introduced into all VAR viruses used here are listed in Supplementary Table 1. Mutations introduced into GFHK99 VAR<sub>1</sub>, MaMN99 VAR, and NL09 VAR viruses enable detection by high-resolution melt analysis or probe-based droplet digital PCR. Mutations introduced into the GFHK99 VAR<sub>2</sub> strain were designed to confer unique primer binding sites relative to GFHK99 WT virus. Viruses used for single-cell mRNA sequencing, GFHK99 mVAR<sub>1</sub> and GFHK99 mVAR<sub>2</sub>, carry mutations near the 3' end of each transcript to allow detection by sequencing following oligo-dT priming<sup>24</sup>.

### Synchronized, single-cycle infection conditions

Conditions designed to synchronize viral entry and prevent propagation of progeny viruses were used for all cell culture-based infections with the exception of those performed for measurement of  $P_p$  values. Synchronized, single-cycle infections were performed as follows: Cell monolayers were washed three times with PBS and placed on ice. Chilled virus inoculum was added to each well and incubated at 4°C for 45 minutes with occasional rocking to allow attachment without entry<sup>55</sup>. Inoculum was aspirated and cell monolayer was washed three times with cold PBS before addition of warm virus medium lacking trypsin. Cultures were incubated at 37°C. At 3 h post-infection, virus medium was replaced

with ammonium chloride-containing virus medium. Addition of ammonium chloride to the medium prevents acidification of endosomes, thereby blocking further IAV infection<sup>56</sup>. Cultures were returned to 37°C for the remainder of the incubation time.

### Infection of cultured cells for quantification of coinfection and reassortment

MDCK or DF-1 cells were seeded at a density of  $4 \times 10^5$  cells per well in 6-well dishes 24 h before inoculation. Virus inoculum was prepared by combining WT and VAR viruses at high titer in a 1:1 ratio based on PFU titers, and then diluting in PBS to achieve MOIs ranging from 10 to 0.01 PFU per cell. Synchronized, single cycle infection conditions were used. In addition, owing to the low yield of IAV in DF-1 cells, acid inactivation of inoculum virus was performed at 1 h post-infection for this cell type. This procedure is needed if residual inoculum virus would otherwise comprise an appreciable proportion of the virus sampled at 12 or 16 h post-inoculation. For acid inactivation, media was aspirated and replaced with 500  $\mu$ L of PBS-HCl, pH 3.00 and incubated 5 min at 37°C. Cells were then washed once with PBS before the addition of virus medium. GFHK99 virus-infected cells were harvested at 12 h post-infection due to high amounts of cytopathic effects at later time points. Cells infected with MaMN99 virus and MaMN99:GFHK99 chimeric viruses were harvested at 16 h post-infection. NL09 virus reassortment data shown in Figure 1C were reported previously<sup>57</sup> and are included here to allow comparison to the avian strains used.

### Determination of infection and coinfection levels based on HA surface expression.

To enumerate infected cells, surface expression of HIS and HA epitope tags was detected by flow cytometry (Extended Data Figure 9). This method was previously described in detail<sup>22</sup>. Cells were fixed following staining by resuspending in FACS buffer and 1% paraformaldehyde. The percentage of cells positive for either or both epitope tags is expressed as % cells HA<sup>+</sup>. The percentage of cells positive for both epitope tags is expressed as % cells dual-HA<sup>+</sup>. The relationship between these two parameters was evaluated by plotting % cells dual-HA<sup>+</sup> against % cells HA<sup>+</sup> and regressing the resultant curve as a polynomial:

$$\% \text{ cells dual-HA}^+ = \beta_2 * \text{Expectation}(\% \text{ cells HA}^+) + \beta_1 * (\% \text{ cells HA}^+)^2,$$

where  $\beta_2$  and  $\beta_1$  are genotype-specific. For a given fraction of cells expressing HA, the expected fraction of cells expressing both HA proteins is derived from the Poisson distribution with  $\lambda = -\ln(1 - \% \text{ cells HA}^+)$ :

$$\sum_{v=1}^{\infty} \frac{e^{-\lambda} \lambda^v}{v!} (1 - 2 * 0.5^v)$$

From the regression models, we then quantified the degree of linearity using the equation:

$$\% \text{ linearity} = \frac{|\beta_1|}{|\beta_1| + |\beta_2|}.$$

### Animal models and reassortment *in vivo*

All animal experiments were conducted in accordance with the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health. Studies were conducted under animal biosafety level 2 (ABSL-2) containment and approved by the IACUC of the University of Georgia (protocols A201506-026-Y3-A2 and A201506-026-Y3-A5) for quail studies or the IACUC of Emory University (protocol PROTO201700595) for guinea pig studies. Animals were humanely euthanized following guidelines approved by the American Veterinary Medical Association (AVMA).

Quail eggs obtained from the College of Veterinary Medicine, University of Georgia, were hatched at the Poultry Diagnostic and Research Center, University of Georgia. Two days before virus inoculation, quail sera were confirmed to be seronegative for IAV exposure by NP ELISA (IDEXX, Westbrook, ME). At 3 weeks of age, birds were moved into a HEPA in/out BSL2 facility and each group divided into individual isolator units.

Groups (n=6) of 3-week old Japanese quail (*Coturnix Japonica*) were used to determine the 50% quail infectious dose of the 1:1 GFHK99 WT and GFHK99 VAR<sub>1</sub> virus mixture. Each quail was inoculated with 500  $\mu$ L by oculo-naso-tracheal route of virus mixture in PBS, at increasing concentrations of 10<sup>0</sup> to 10<sup>6</sup> TCID<sub>50</sub> per 500  $\mu$ L. Tracheal and cloacal swab specimens were collected daily from each bird in brain heart infusion media (BHI). Swab samples were analyzed by TCID<sub>50</sub> assay and titers of tracheal swabs collected at 4 d post-inoculation were used to determine the QID<sub>50</sub> by the Reed and Muench method<sup>58</sup>. Virus was not detected in cloacal swabs. QID<sub>50</sub> was found to be equivalent to 1 TCID<sub>50</sub>.

To quantify reassortment in quail, samples collected from quail (n=6) infected with the 10<sup>2</sup> TCID<sub>50</sub> dose of the 1:1 GFHK99 WT and GFHK99 VAR<sub>1</sub> virus mixture were used. These were the same birds as used to determine QID<sub>50</sub>. Virus shedding kinetics were determined by plaque assay of tracheal swab samples and samples from days 1, 3 and 5 were chosen for genotyping of virus isolates.

Female Hartley strain guinea pigs weighing 250–350 g were obtained from Charles River Laboratories (Wilmington, MA) and housed by Emory University Department of Animal Resources. Prior to intranasal inoculation and nasal washing, guinea pigs were anesthetized by intramuscular injection with 30 mg per kg ketamine/ 4 mg per kg xylazine. The GPID<sub>50</sub> of GFHK99 WT/VAR<sub>1</sub> and MaMN99 WT/VAR virus mixtures were determined as follows. Groups of four guinea pigs were inoculated intranasally with virus mixture in PBS at doses of 10<sup>0</sup> to 10<sup>5</sup> PFU per 300  $\mu$ L inoculum. Daily nasal washes were collected in 1 mL PBS and titered by plaque assay. Results from day 2 nasal washes were used to determine the GPID<sub>50</sub> by the Reed and Muench method<sup>58</sup>. The GPID<sub>50</sub> of GFHK99 virus was found to be 2.1  $\times$  10<sup>3</sup> PFU; that of MaMN99 virus was 2.1  $\times$  10<sup>1</sup> PFU. The GPID<sub>50</sub> of NL09 virus was previously determined to be 1  $\times$  10<sup>1</sup> PFU<sup>59</sup>.

To evaluate reassortment kinetics in guinea pigs, groups of six animals were infected with 10<sup>2</sup>  $\times$  GPID<sub>50</sub> of the aforementioned GFHK99 WT / VAR<sub>1</sub>, MaMN99 WT / VAR or NL09 WT / VAR virus mixtures. Virus inoculum was given intranasally in a 300  $\mu$ L volume of PBS. Nasal washes were performed on days 1–6 post-inoculation and titered for viral

shedding by plaque assay. Viral genotyping was performed on samples collected on day 1, 3, and 5 for each guinea pig.

### Quantification of reassortment and effective diversity

Reassortment was quantified for *in vitro* coinfection supernatants, guinea pig nasal washes, and quail tracheal swabs as described previously<sup>22</sup>. Briefly, plaque assays were performed on MDCK cells in 10 cm dishes to isolate virus clones. 1 mL serological pipettes were used to collect agar plugs into 160  $\mu$ l PBS. Using a ZR-96 viral RNA kit (Zymo), RNA was extracted from the agar plugs and eluted in 40  $\mu$ l nuclease free water (Invitrogen). Reverse transcription was performed using Maxima RT (ThermoFisher) according to the manufacturer's protocol. The resulting cDNA was diluted 1:4 in nuclease free water and each cDNA was combined with segment specific primers (Supplementary Table 2) and Precision Melt Supermix (Bio-Rad) and analyzed by qPCR in a CFX384 Touch real-time PCR detection system (Bio-Rad) designed to amplify a ~100 bp region of each gene segment which contains a single nucleotide change in the VAR virus. The qPCR was followed by high-resolution melt (HRM) analysis to differentiate WT and VAR amplicons<sup>60</sup>. Precision Melt Analysis software (Bio-Rad) was used to determine the parental virus origin of each gene segment based on melting properties of the cDNAs and comparison to WT and VAR controls. Each individual plaque was assigned a genotype based on the combination of WT and VAR genome segments, with 2 variants on each of 8 segments allowing for 256 potential genotypes. We note that, despite its high reliance on multiple infection for robust replication, the GFHK99 virus does form very small, clonal plaques on MDCK cells. These plaques were confirmed to be clonal based on the unambiguous typing of each gene segment as either WT or VAR in origin.

Reassortment data were used to calculate viral genotypic diversity; that is, the diversity of the 256 possible WT/VAR genotypes present in a given sample. Diversity was quantified as reported previously<sup>61</sup>, by calculating Simpson's Index, given by  $D = \sum(p_i^2)$ , where  $p_i$  represents the proportional abundance of each genotype<sup>62</sup>. Simpson's Index accounts for both the raw number of species and variation in abundance of each, and is sensitive to the abundance of dominant species. These features are important to ensure that overrepresentation of a single reassortant genotype does not cause a viral population to appear more diverse than it actually is. Because Simpson's Index does not scale linearly, each sample's Simpson's Index value was converted to a corresponding Hill number to derive its effective diversity,  $N_2 = 1/D^{63}$ , which is defined as the number of equally abundant species required to generate the observed diversity in a sample community. Because it scales linearly, Hill's  $N_2$  allows a more intuitive comparison between communities (i.e., a community with  $N_2 = 10$  species is twice as diverse as one with  $N_2 = 5$ ) and is suitable for statistical analysis by basic linear regression methods<sup>64</sup>. Robust linear models of  $N_2$  vs. time were regressed using the R package *MASS*.

### Hierarchical clustering analysis

To compare the behavior of multiple viruses during *in vitro* coinfections, a hierarchical clustering algorithm was used. In these experiments, multiplicity dependence was measured by analyzing HA co-expression and reassortment as a function of the percentage of cells

expressing HA. Each of these regression models contain two parameters ( $\% \text{ cells dual-HA}^+ = \beta_2 * \text{Expectation}(\% \text{ cells HA}^+) + \beta_1 * (\% \text{ cells HA}^+)$ ), and  $\% \text{ reassortment} = \beta_0 + \beta_1 * \ln(\% \text{ cells HA}^+)$ ). The behavior of a given virus can therefore be expressed as a set of four parameters. These parameters were used to calculate the distance between each pair of viruses, and Ward's method of agglomerative hierarchical clustering was then used to organize viruses into clusters based on these distances<sup>65</sup>. Briefly, this algorithm combines a nearby pair of elements into a cluster with a new position that is halfway between the original individuals. This process is repeated until all elements have been incorporated into one cluster. The R package *pvclust* was used to calculate statistical support for the existence of each node by multiscale bootstrap resampling, with nodes appearing in 95% of trials being deemed statistically significant. Each dendrogram was then divided into two clusters to determine whether the behavior of GFHK99 virus in MDCK cells represented a true outgroup (Figure 1), or to determine whether each GFHK99:MaMN99 chimeric virus was more similar to GFHK99 or MaMN99 parent virus (Figure 4).

### Single-cycle viral growth kinetics

DF-1 or MDCK cells were seeded at  $4 \times 10^5$  cells per well in 6-well dishes 24 h prior to infection. Virus was serially diluted using PBS. Synchronized, single-cycle infection conditions with acid inactivation of inoculum virus as described above were used. At each time point, 120  $\mu$ l supernatant was collected. Viral titers for each sample were assessed by plaque assay in MDCK cells. Each MOI condition was used in 5–6 wells in parallel infections. Three wells served as technical replicates for growth curve sampling while the remaining wells were harvested at 24 h post-infection to enumerate HA-expressing cells via flow cytometry. Flow cytometry data for NL09 in MDCK cells are not included owing to low sensitivity of the assay for this virus. In cases where acid inactivation was inefficient, the replicate was eliminated, and data are plotted in duplicate.

### Effect of increasing multiple infection on viral RNA replication

For DF-1 and MDCK cell experiments, 12-well plates were seeded with  $3 \times 10^5$  cells per well 24 h prior to infection. For NHBE cells, cells were cultured at an air-liquid interface as previously described<sup>47</sup>. Cell surfaces were washed three times with PBS prior to inoculation. Triplicate wells were then inoculated with increasing doses of VAR virus, plus 0.005 PFU per cell of WT virus in DF-1 and MDCK cells or 0.05 PFU per cell of WT virus in NHBE cells. VAR viruses used were GFHK99 VAR<sub>2</sub>, MaMN99 VAR, NL09 VAR, MaMN99-GFHK99-PA VAR, dkHK78 VAR, and QaHK88 VAR. After 55 minutes at 37°C, inoculum was aspirated, cells were washed three times with PBS and 1 ml per well virus medium was added. Media was exchanged for ammonium chloride-containing media 3 h later. At 12 h post-infection, virus media was removed and cells were harvested using RNeasy Protect Cell Reagent (Qiagen). RNA was extracted using RNeasy columns (Qiagen) and then reverse transcribed with universal influenza primers<sup>66</sup> and Maxima RT per protocol instructions. Droplet digital PCR (ddPCR) was performed on the resultant cDNA. For GFHK99 virus, QX200™ ddPCR™ EvaGreen Supermix (Bio-Rad) was used with a combination of PB2, M, and NS primers specific for the GFHK99 WT virus (final primer concentration of 200 nM) (Supplementary Table 3). For MaMN99, MaMN99-Anhui-PA, MaMN99-GFHK99-PA,



dkHK78, QaHK88 and NL09 viruses, QX200™ ddPCR™ Supermix for Probes (Bio-Rad) was used with NP specific primers and probes (Supplementary Table 3).

### Strand-specific quantification of viral RNA species over time

Viruses used for this experiment were the same GFHK99 WT/VAR<sub>1</sub> or MaMN99 WT/VAR virus mixtures used to measure reassortment, but in this case each mixture was considered as a single virus population (i.e. the RT ddPCR assay outlined below to quantify viral m/vRNA does not differentiate between WT and VAR genotypes). To ensure viral genomes were supplied as low or single copies, a dose of 0.5 RNA copies per cell was used for low MOI infection. RNA copy numbers of virus stocks were determined by ddPCR targeting at least four segments (mean values were used). For the high MOI dose, we used 3.0 HA-expressing units per cell, as determined by flow cytometry (Extended Data Figure 10). This measure of infectivity gives a functional readout for polymerase activity, and therefore ensures the vast majority of cells carry an active viral polymerase. HA-expressing units per mL was determined by counting HA<sup>+</sup> cells by flow cytometry in the relevant cell type. Specifically, cells were infected with serial dilutions of virus under synchronized, single cycle conditions. At 24 h post-infection, cells were harvested and flow cytometry was performed targeting His and HA epitope tags. HA expressing units per mL for each virus / cell combination was calculated based on the linear range of %HA<sup>+</sup> cells plotted as a function of volume of virus added to cells<sup>67</sup> (Extended Data Figure 10).

12-well plates were seeded with  $2 \times 10^5$  cells per well of MDCK or DF-1 cells and incubated at 37°C for 24 h. Synchronized, single-cycle infection conditions were used, as described above. Chilled virus was added at a volume of 125  $\mu$ L per well. At 0, 1, 2, 4, 6, 8, and 10 h post-infection, virus medium was aspirated and cells were harvested using 400  $\mu$ L of RNAprotect Cell Reagent (Qiagen). RNA was extracted using the Qiagen RNeasy Mini kit. Two reverse transcription reactions per sample were set up with primers targeting mRNA or vRNA of the NS segment, each containing different nucleotide barcode tags (Supplementary Table 4). Maxima RT was used according to the manufacturer's instructions and combined with 300 ng MDCK or 150 ng DF-1 RNA. Absolute copy number of cDNA was determined by ddPCR. Forward and reverse primers for vRNA or mRNA of NS at a total concentration of 200 nM were combined with diluted cDNA and QX200™ ddPCR™ EvaGreen Supermix (Bio-Rad). Primer sequences were based on the method of Kawakami et al.<sup>68</sup> and are given in Supplementary Table 4. Thermocycler protocol was 95°C for 5 min, [95°C for 30s, 57°C for 60s] repeat 40x, 4°C for 5 min, 90°C for 5 min, 4°C hold. Copy number was normalized to RNA input to give final results in units of copy number per ng RNA. We did attempt to measure cRNA in this assay but found that primers designed to be specific for cRNA cross-primed on viral mRNA.

### Single-cell mRNA sequencing

For this assay, viruses were titered in DF-1 cells using flow cytometry with anti-NP antibody. DF-1 cells were used because they give more sensitive detection of GFHK99 virus infection than MDCK cells. Cells were infected with serial dilutions of virus under synchronized, single-cycle conditions. At 24 h post-infection flow cytometry was performed targeting NP: cells were processed with the BD Cytotfix/Cytoperm™ Kit (catalog no.

554714) and stained with anti-NP antibody (Abcam, clone 9G8, catalog no. ab43821) followed by donkey anti-Mouse IgG Alexa Fluor 488 (Invitrogen catalog no. A-21202). NP expression units per were calculated based on the linear range of % cells NP<sup>+</sup> plotted as a function of volume of virus added to cells<sup>67</sup> (Extended Data Figure 10).

To perform single-cell mRNA sequencing, MDCK and DF-1 cells were seeded into 6-well plates at  $5 \times 10^5$  cells per well. At 24 h post seeding, cells were infected with synchronized, single cycle conditions. In the first experiment, MDCK or DF-1 cells were inoculated with GFHK99 WT virus at an MOI of 0.07, 0.2, 0.6, or 1.8 NP units per cell. In the second experiment, MDCK or DF-1 cells were inoculated with a 1:1 ratio of GFHK99 WT virus and GFHK99 mVAR<sub>1</sub> virus that amounted to a MOI of 0.02, 0.07, 0.2 or 0.6 NP units per cell. GFHK99 mVAR<sub>2</sub> virus was added to MDCK and DF-1 cell infections at MOIs of 1 PFU per cell and 0.1 PFU per cell, respectively. Infected cells were incubated for 8 h at 37°C. Culture media was then aspirated and cells washed once with 1X PBS. Cells were then trypsinized with 200 µL of 0.25% Trypsin EDTA until all cells came off the plate and were mono-dispersed. To each well, 0.5 mL of virus medium was added and replicates were pooled (2 wells per MOI). Cells for each sample were counted. Samples were spun at 150 rcf for 3 minutes and washed with 0.5 mL of 1X PBS/0.04% BSA. Washings were performed two more times. Finally, cells were resuspended with 1X PBS/0.04% BSA to give  $7 \times 10^5$  cells per mL. Preparation for single-cell transcriptomic sequencing followed the protocol for 10x Genomics Chromium Single Cell platform.

Analysis of viral transcripts from single cells was performed with the sequencing data from all experiments using Cell Ranger software. Briefly, the Cell Ranger software assigns each read to individual cells and transcripts based on two sets of unique molecular identifiers (UMIs) that are ligated prior to amplification. This approach allows the quantification of amplification bias at both the cellular and transcript levels. The first step of the analytical workflow was to map the reads to concatenated transcriptomes of GFHK99 virus with the transcriptomes of dog or chicken to analyze MDCK and DF-1 cell infections, respectively. Protein coding regions for the dog and chicken transcriptomes were identified in the GTF file associated with genome builds CanFam3.1.98 (NCBI accession number GCA\_000002285.2) and GRCg6a (NCBI accession number GCA\_000002315.5), respectively, while GFHK99 virus coding regions were extracted from the reverse-complement sequences of the GFHK99 strains. The aligned sequencing data is available on the GEO database with the accession number GSE135553. Cell Ranger output in the form of gene and barcode counts were then analyzed in R using the packages *CellrangerRkit* and *Seurat*. To account for the issue of cellular lysis, which can allow uninfected cells to acquire viral RNA from the supernatant and thus appear infected, a preliminary analysis was conducted to exclude cells that were likely false positives. This analysis was informed by the reasoning that 1) contaminated cells were likely to contain less viral RNA than truly infected cells, and 2) the amount of viral RNA present in contaminated cells should be less consistent than the amount present in truly infected cells. Within each infection, the proportion of each cell's transcriptome that was comprised of viral RNA was calculated, and cells were ordered according to this proportion so that the marginal gain in % viral RNA from one cell to the next could be calculated. Excluding cells with no marginal gain (indicating two cells with the same % viral RNA), the marginal gain vs. % viral RNA was plotted for each infection,

which showed that this marginal gain is initially a rapidly decreasing function of % viral RNA. The first local minimum of a local regression was derived for each infection, to determine the point at which marginal gain became more consistent and less affected by % viral RNA, and cells with % viral RNA below that threshold were excluded from further analysis (Extended Data Figure 5). To enable comparisons between samples, the median number of UMIs per cell was calculated for each infection, and the UMI counts of each cell within that infection were normalized to this median value (e.g. if the total number of UMIs in a cell was 50% of that detected in the median cell, all of its UMI counts were multiplied by 2). The number of viral RNA transcripts per cell was then calculated and  $\log_{10}$ -transformed.

### Single-cell sorting assay for measurement of $P_P$ values

Segment-specific  $P_P$  values were determined as previously described for influenza A/Panama/2007/99 (H3N2) virus<sup>8</sup>, and as follows.  $4 \times 10^5$  MDCK cells were seeded into each well of a 6-well dish. 24 h later, cells were washed 3x with PBS and inoculated with 0.018 PFU per cell of GFHK99 WT virus and 1 PFU per cell of GFHK99 VAR<sub>2</sub> virus in 250  $\mu$ L of PBS. Virus was allowed to attach at 37°C for 1 h. Inoculum was then removed and cells were rinsed 3x with PBS and 2 mL of virus medium was added to the well. After 1 h at 37°C, medium was removed and cells were washed 3x with PBS and harvested by addition of Cell Dissociation Buffer (Corning). Cells were resuspended in complete medium and washed 3x with 2 mL FACS buffer (2% FBS in PBS). A final resuspension step was performed in PBS containing 1% FBS, 10 mM HEPES, and 0.1% EDTA. Cells were strained through a cell strainer cap (Falcon) and sorted on a BD Aria II cell sorter. Gating was performed to remove debris and multipllets and one event per well was sorted into each well of a 96-well plate containing MDCK monolayers at 30% confluency in 50  $\mu$ L virus medium supplemented with 1  $\mu$ g per mL TPCK-treated trypsin. Following the sort, an additional 50  $\mu$ L of virus medium plus TPCK-treated trypsin was added to each well and plates were centrifuged at 1,800 rpm for 2 minutes to promote cell attachment. Plates were incubated at 37°C for 48 h to allow propagation of virus from the sorted cell.

RNA was extracted from infected cells in the 96-well plate using a ZR-96 Viral RNA Kit (Zymo Research) per manufacturer instructions. Extracted RNA was converted to cDNA using universal influenza primers<sup>66</sup> and Maxima RT according to manufacturer instructions. After conversion, cDNA was diluted 1:4 with nuclease-free water and used as template (4  $\mu$ L per reaction) for segment-specific qPCR using SsoFast EvaGreen Supermix (Bio-Rad) in 10  $\mu$ L reactions, with 200 nM final primer concentration. Primers employed targeted each segment of GFHK99 WT virus, as well as the PB2 and PB1 segments of GFHK99 VAR<sub>2</sub> virus. Primer sequences are listed in Supplementary Table 3.

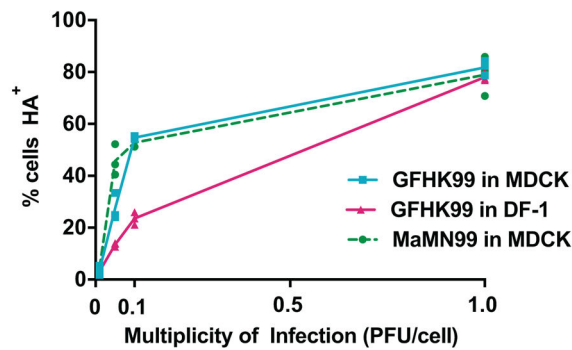
Given the MOI of GFHK99 WT virus used in the experiments, an appreciable number of wells are expected to receive two or more viral genomes, and so a mathematical adjustment is needed to estimate the probability of each genome segment being delivered by a single virion. Using the relationship between MOI and the fraction of cells infected from Poisson statistics, i.e.,  $f = 1 - e^{-\text{MOI}}$ , the probability of the  $i$ th segment being present in a singly infected cell, or  $P_{P_i}$ , can be calculated from the 96-well plate using the following equation:

$$PP_i = \frac{MOI_i}{MOI_{wt}} = \frac{-\ln(1 - f_i)}{-\ln(1 - f_{wt})} = \frac{\ln\left(1 - \frac{C_i}{A}\right)}{\ln\left(1 - \frac{B}{A}\right)}$$

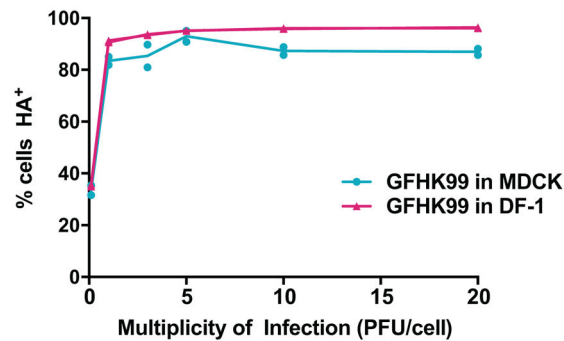
where  $A$  is the number of VAR<sub>2</sub><sup>+</sup> wells,  $B$  is the number of WT<sup>+</sup> wells (containing any WT segment), and  $C_i$  is the number of wells positive for the WT segment in question. Wells that were negative for VAR<sub>2</sub> virus segments were excluded from analysis.

## Extended Data

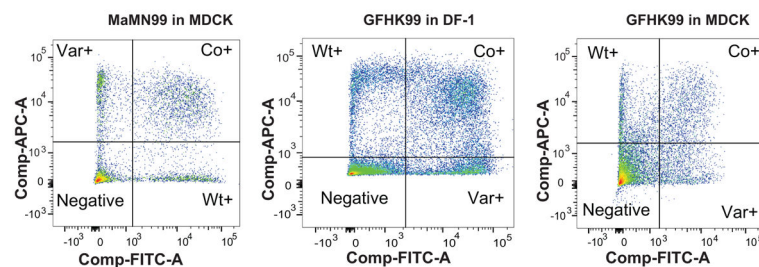
**A**



**B**

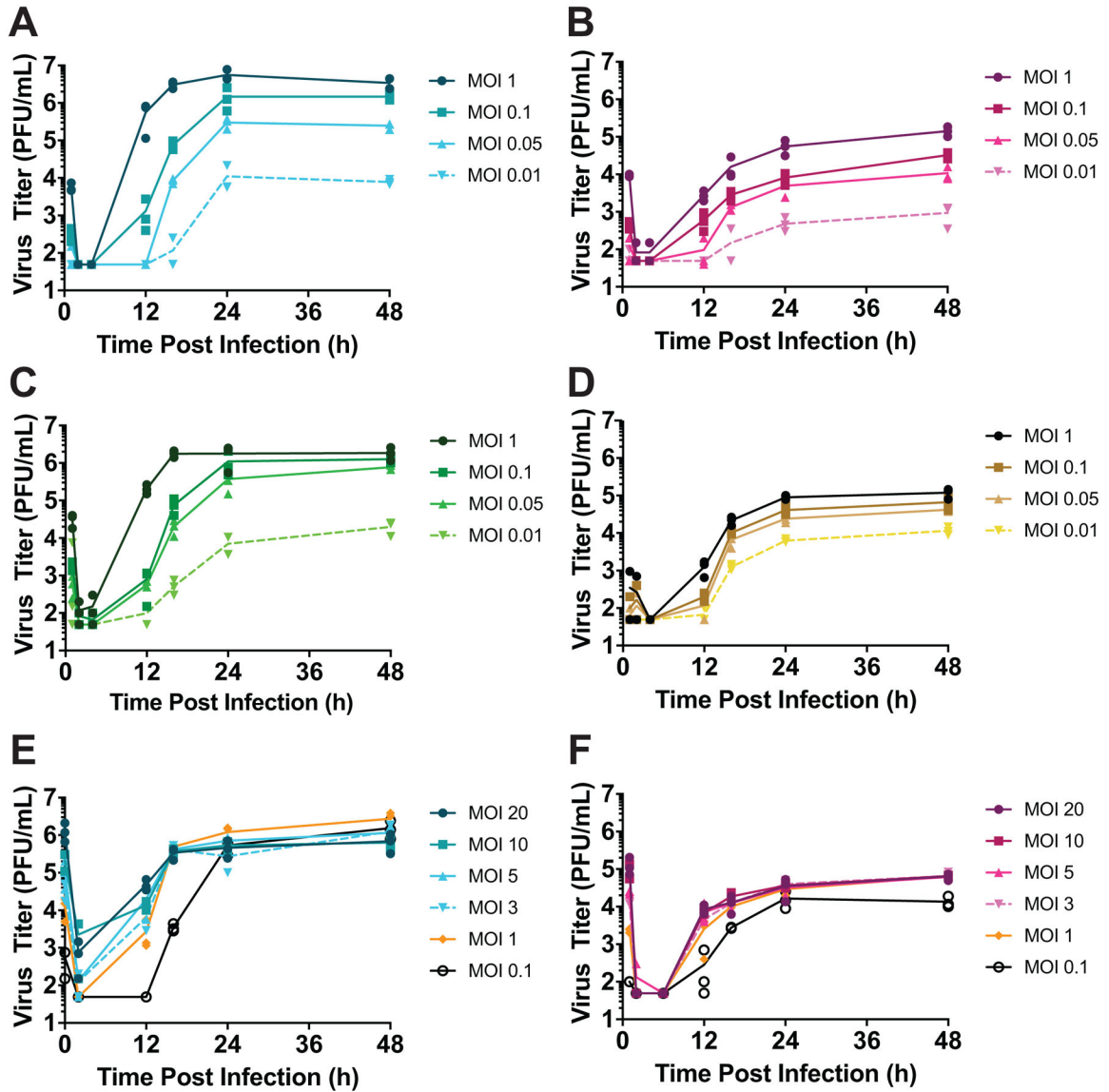


**C** Example HA<sup>+</sup> flow gates:



Extended Data Fig. 1. Level of infection achieved in single cycle growth assays, as determined by flow cytometric detection of HA protein.

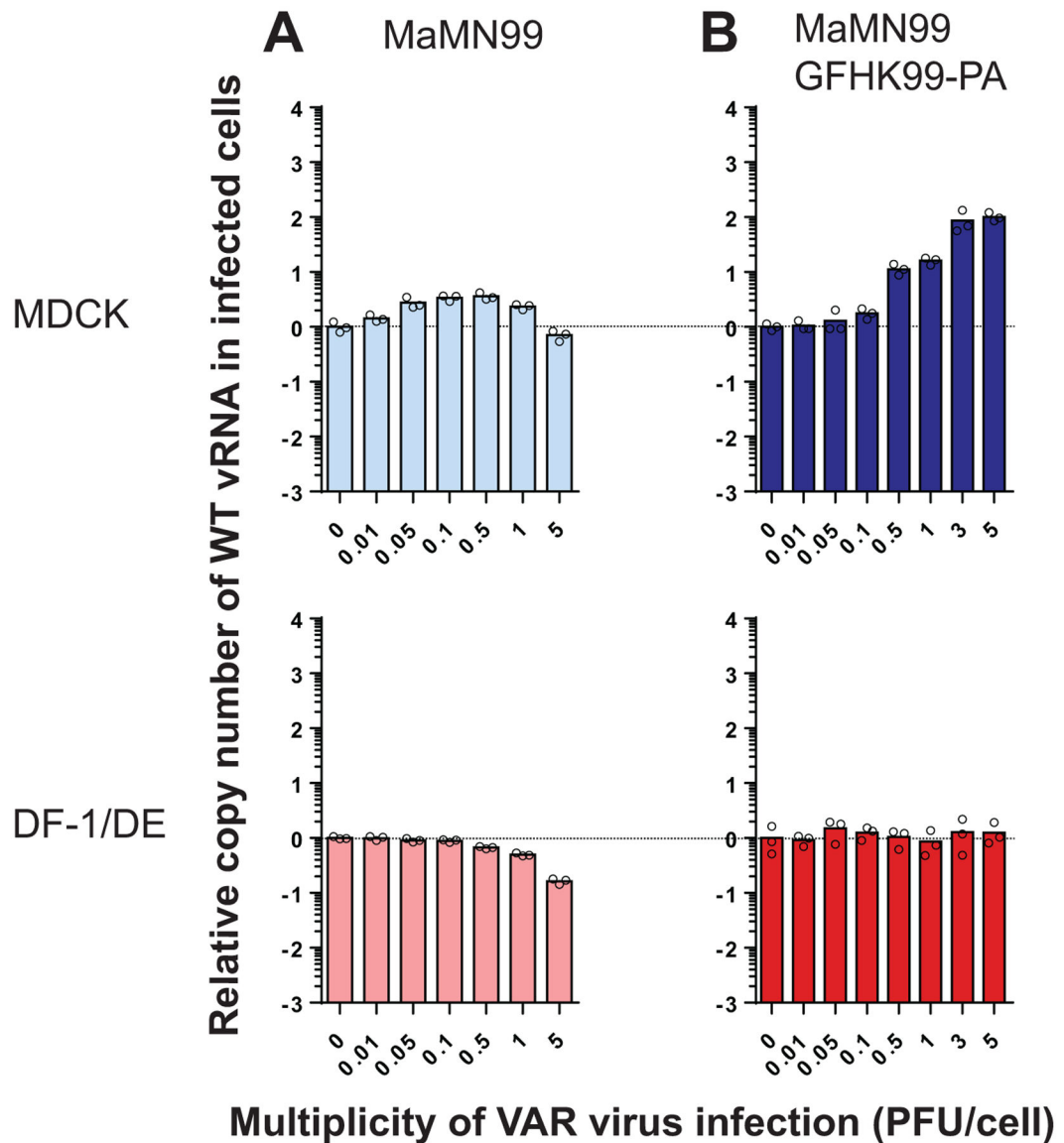
Triplicate or duplicate wells of cells were harvested at 24 h post-infection and stained to detect surface expression of HA and HIS epitope tags. Panel A) corresponds to Extended Data Figure 2 A–C and Panel B) corresponds to Extended Data Figure 2 E–F. Lines connect the means of n=2 or n=3 replicate samples. C) Flow gating was performed by excluding cell debris and multiplet cells. Quadrant gates were used to quantify each population. Flow cytometry results are representative of those obtained in three independent experiments.



**Extended Data Fig. 2. Increasing MOI increases viral productivity at sub-saturating, but not saturating MOIs.**

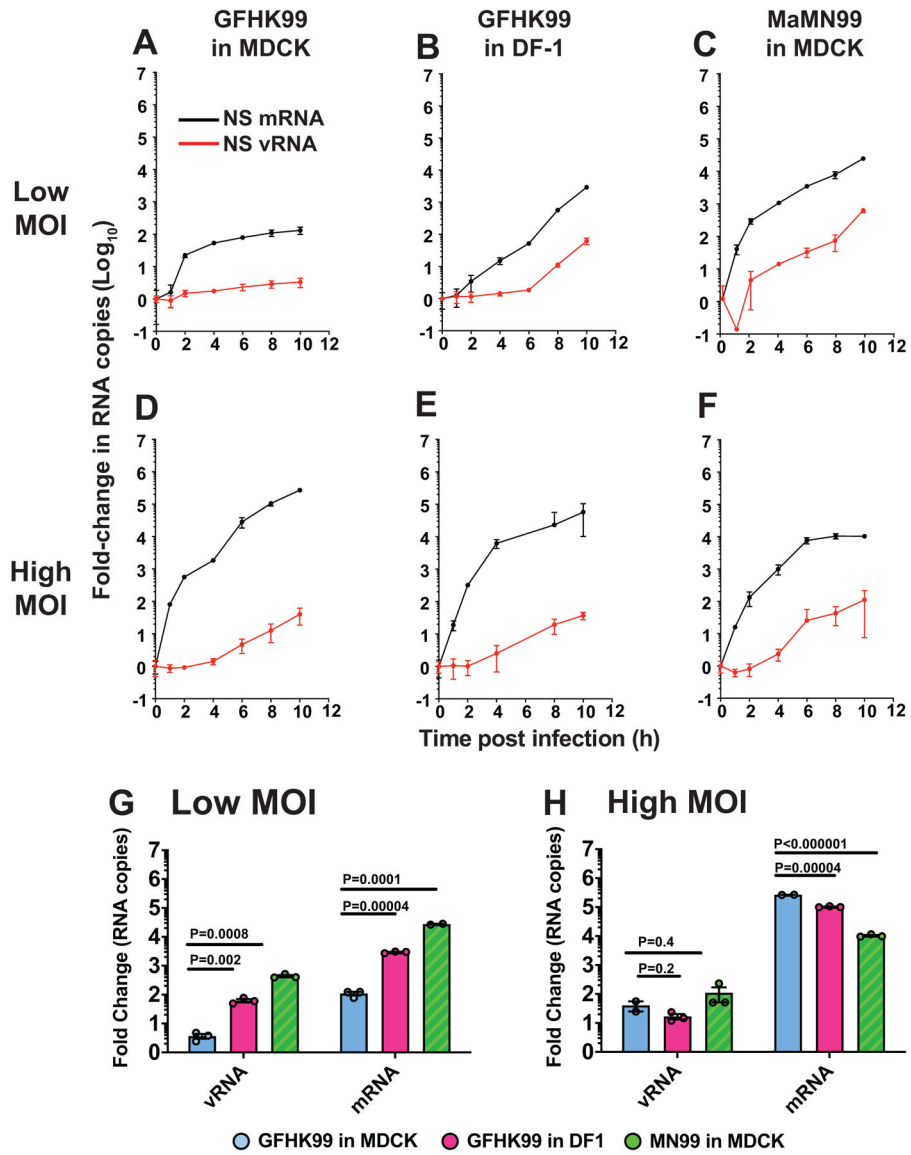
Data relate to Figure 2. MDCK or DF-1 cells were infected under single-cycle conditions at a range of MOIs. Low MOI range is shown in panels (A) to (D) and high MOI range is shown in panels (E) and (F). As shown in Extended Data Figure 1, MOIs < 1 PFU/cell were found to be sub-saturating. Viral titers observed at the indicated MOIs are plotted against time post-infection for GFHK99 virus in MDCK cells (A), GFHK99 virus in DF-1 cells (B),

MaMN99 virus in MDCK cells (C), NL09 virus in MDCK cells (D), GFHK99 virus in MDCK cells (E), and GFHK99 virus in DF-1 cells (F). Lines connect the mean values for technical replicates sampled at each time point.



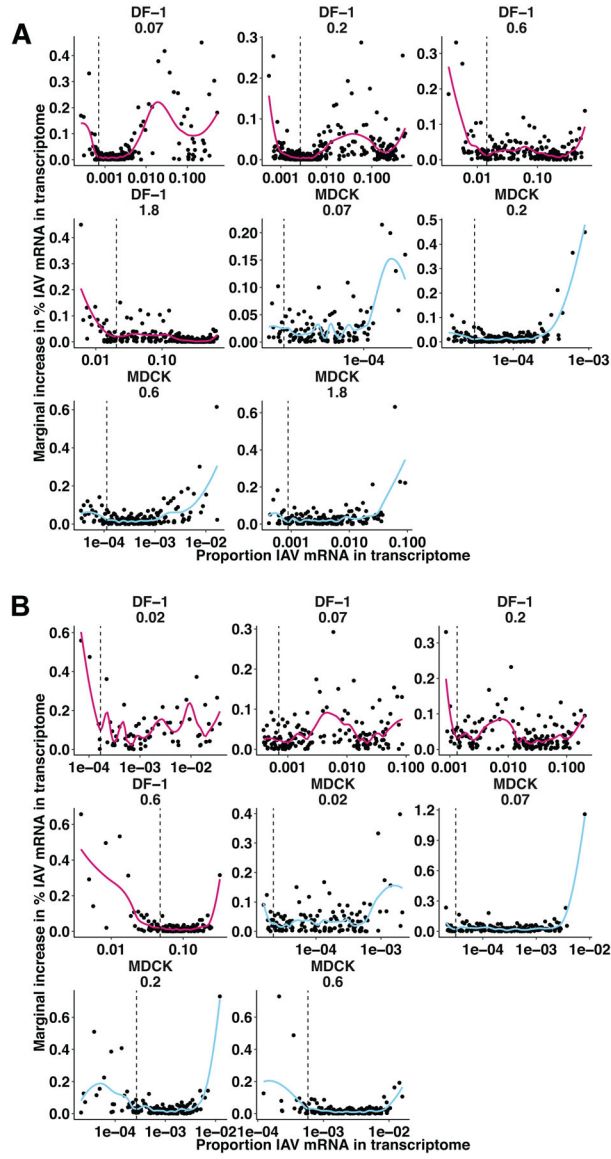
**Extended Data Fig. 3. Introduction of the PA gene segment from GFHK99 virus into MaMN99 virus confers increased dependence on multiple infection for vRNA synthesis.** Cells were coinfecting with WT virus and increasing doses of VAR virus. WT virus MOI was 0.005 PFU per cell. The fold change in WT vRNA copy number, relative to that detected in the absence of VAR virus, is plotted for MaMN99 virus (A) and MaMN99-GFHK99-PA virus (B). Bars represent the mean of n=3 replicate cell cultures per condition. Data shown in panel (A) are also shown in Figure 3. MaMN99 virus was tested in MDCK and DE cells; MaMN99 GFHK99-PA virus was tested in MDCK and DF-1 cells.





**Extended Data Fig. 4. High multiplicity of infection is needed for robust GFHK99 polymerase activity in MDCK cells.**

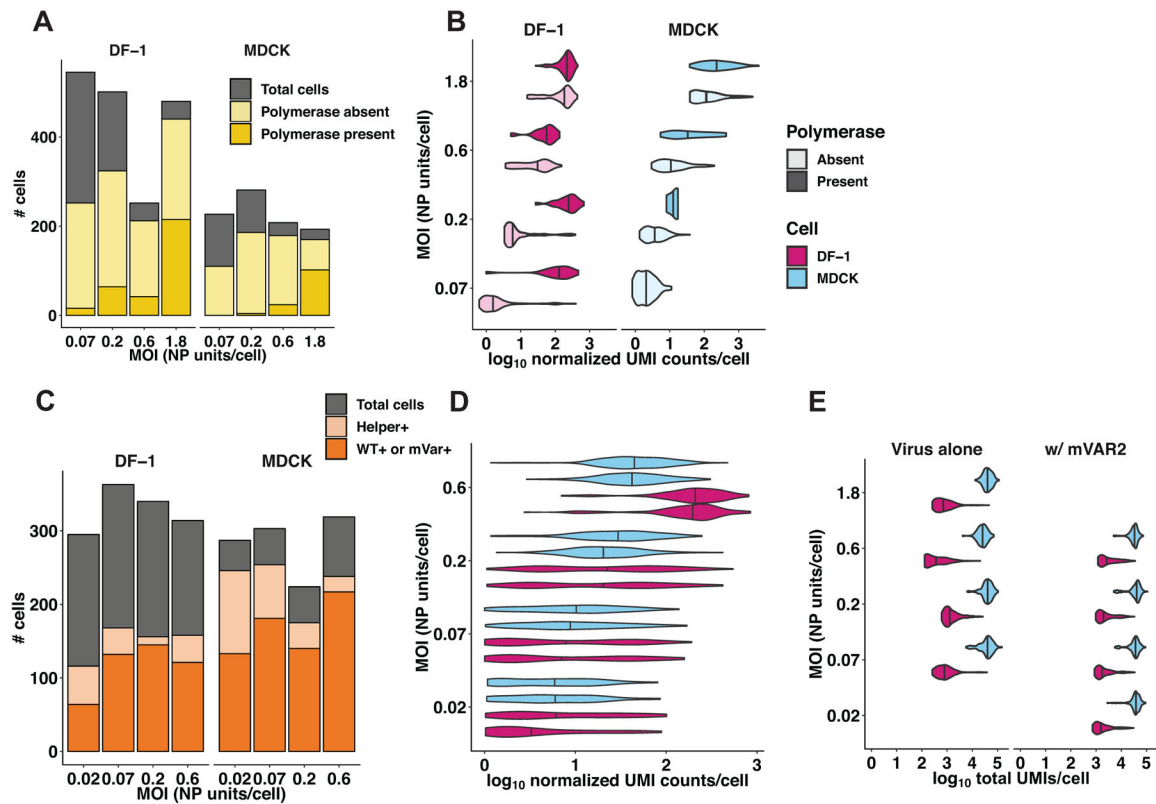
MDCK or DF-1 cells were infected with GFHK99 or MaMN99 virus at low (0.5 RNA copies per cell) or high (3 HA-expressing units per cell) MOI. NS segment vRNA and mRNA was quantified at the indicated time points (A–F). The average fold change from initial ( $t=0$ ) to peak RNA copy number is plotted for low MOI infections (G) and high MOI infections (H). Mean and standard deviation are plotted for  $n=3$  replicate cell cultures sampled sequentially. Significance was assessed by multiple unpaired, two-sided t-tests with correction for multiple comparisons using the Holm-Sidak method, with  $\alpha = 5.0\%$ . Each row was analyzed individually, without assuming a consistent SD.



**Extended Data Fig. 5. Preliminary analysis of single-cell mRNA sequencing data to exclude cells with viral mRNA that are likely uninfected.**

A) Within each infection, cells in which viral RNA was detected were rank ordered by the proportion of their transcriptome that comprised viral RNA (% viral RNA), and the relative gain in % viral RNA from one cell to the next was plotted against the proportion of viral RNA in each cell. Local regression was performed separately for each infection, and the first local minimum of the resulting functions (indicated by dashed lines) indicated the point at which the marginal gain in % viral RNA was more consistent and less sensitive to the % viral RNA of the prior cell. Cells with % viral RNA values below this threshold were deemed falsely positive and considered uninfected for the analyses shown in Figure 5 and Extended Data Figure 6. Facets indicate individual infections, with lines colored by cell type (DF-1 = pink, MDCK = blue). B) The same analysis in panel A) was applied to the data from the second experiment, in which cells were co-inoculated with a 1:1 mixture of WT and mVAR<sub>1</sub> viruses, as well as mVAR<sub>2</sub> virus at an MOI of 0.1 PFU per cell in DF-1 cells, or

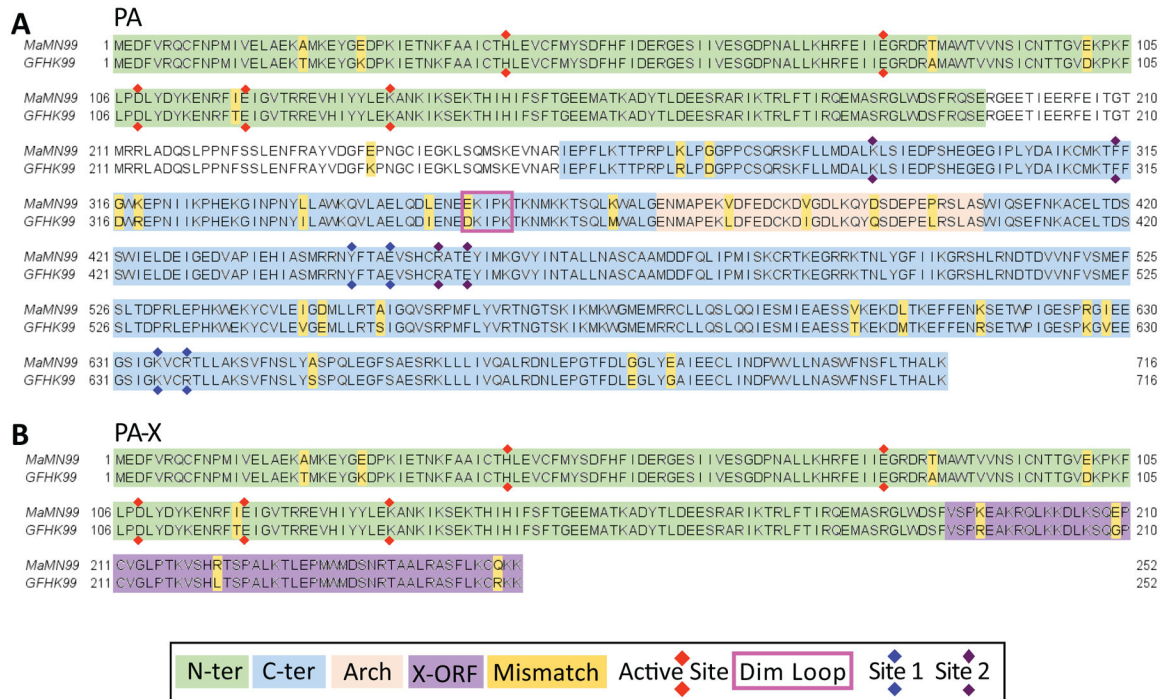
1.0 PFU per cell in MDCK cells. Only cells containing all eight mVAR<sub>2</sub> segments were analyzed in this manner.



#### Extended Data Fig. 6. Validation of single-cell mRNA sequencing data.

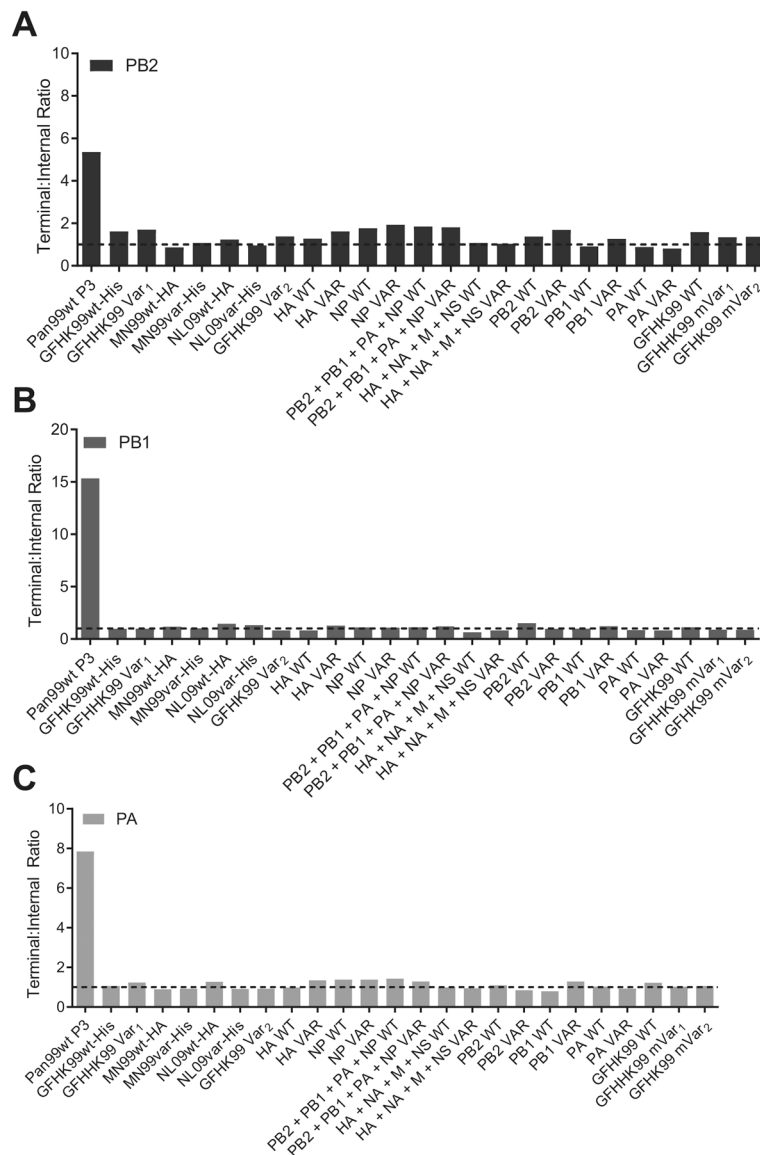
(A, B, E) In the first single-cell sequencing experiment, DF-1 or MDCK cells were infected with GFHK99 WT virus at four different MOIs (0.07, 0.2, 0.6, 1.8 NP units per cell), and the transcriptomes of individual cells were sequenced using the 10x Genomics Chromium platform ( $n = 1,228$  DF-1 cells, 645 MDCK cells, 1 sequencing replicate per infection condition). A) The total number of cells sequenced, infected, and containing PB2, PB1, PA, and NP segments are represented by the cumulative heights of the gray, light yellow, and dark yellow bars, respectively. Cells that were excluded by the analysis shown in Extended Data Figure 5 are contained within the gray bar. B) Each violin plot shows the full distribution of log<sub>10</sub>-transformed viral mRNA abundance, for all eight viral transcripts combined, in individual infected cells. Vertical lines represent the median of each distribution. The data are stratified by cell type (MDCK cells in blue, DF-1 cells in pink), MOI, and the presence of polymerase complex (light shading = cells missing PB2, PB1, PA, or NP; dark shading = cells in which PB2, PB1, PA and NP are all detected). The absence of a dark shaded distribution for MDCK cells at the lowest MOI is due to the absence of any cells in which all four of these segments were detected. (C, D, E) In the second single-cell sequencing experiment, DF-1 or MDCK cells were infected with GFHK99 WT and mVAR<sub>1</sub> viruses at total MOIs of 0.07, 0.2, 0.6, 1.8 NP units per cell, and simultaneously with a constant dose of mVAR<sub>2</sub> virus ( $n = 462$  DF-1 cells, 671 MDCK cells, 1 sequencing replicate per infection condition). C) The total number of cells sequenced, containing all eight

mVAR<sub>2</sub> genome segments, and infected with either WT or mVAR<sub>1</sub> virus are represented by the cumulative heights of the gray, light orange, and dark orange bars, respectively. As in panel (A), cells that were deemed falsely positive are contained within the gray bar. D) Distributions of viral UMIs per cell are shown separately for WT (bottom of each cell-MOI pair) and mVAR<sub>1</sub> (top of each cell-MOI pair). Vertical lines represent the median of each distribution. As expected, no significant difference was detected between WT and mVAR<sub>1</sub> transcript levels ( $p = 0.061$ , linear mixed effects model). E) The distributions of total UMIs detected per cell are shown for each cell type, MOI, and infection type, from both experiments. Vertical lines represent the median of each distribution.



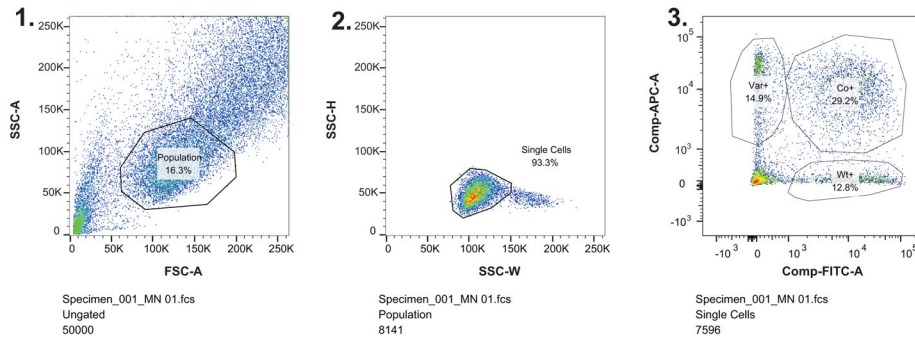
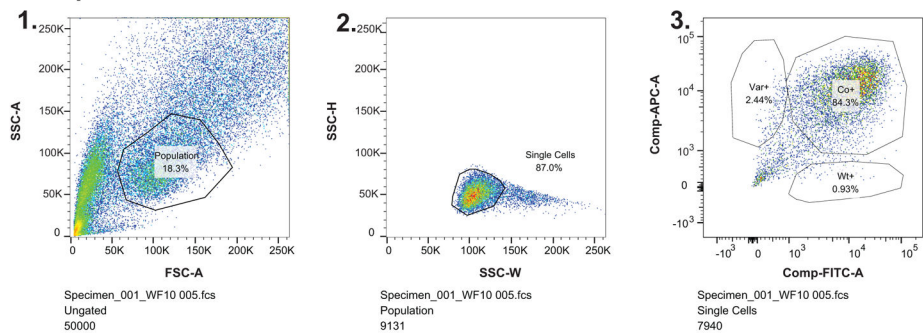
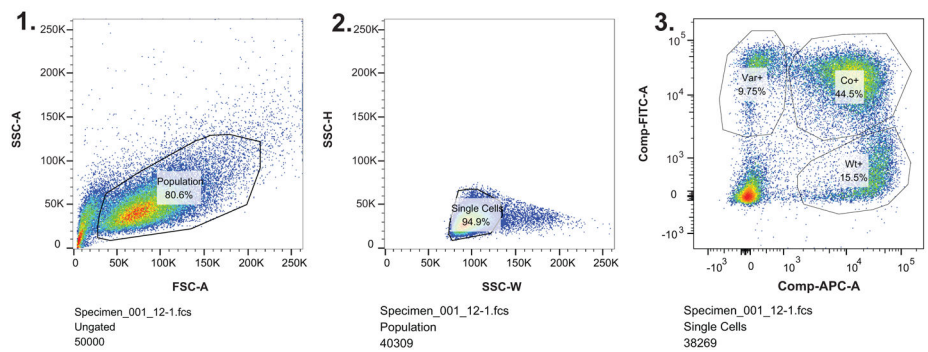
**Extended Data Fig. 7. Alignment of MaMN99 and GFHK99 virus PA and PA-X amino acid sequences.**

Sequences and functional domains of the PA protein are displayed in panel (A), and those of the PA-X protein are shown in panel (B). N-ter = the N-terminal endonuclease domain<sup>69</sup>; C-ter = C-terminal domain<sup>69</sup>; X-ORF = the 61 amino acid region of PA-X encoded in frame 2 of the PA gene<sup>37</sup>; Active site = the active site of the endonuclease<sup>39</sup>; Dim. Loop = dimerization loop important for formation of polymerase dimers<sup>40</sup>; Site 1 and Site 2 = sites mediating the interaction of PA with cellular Pol II C-terminal domain<sup>70</sup>.



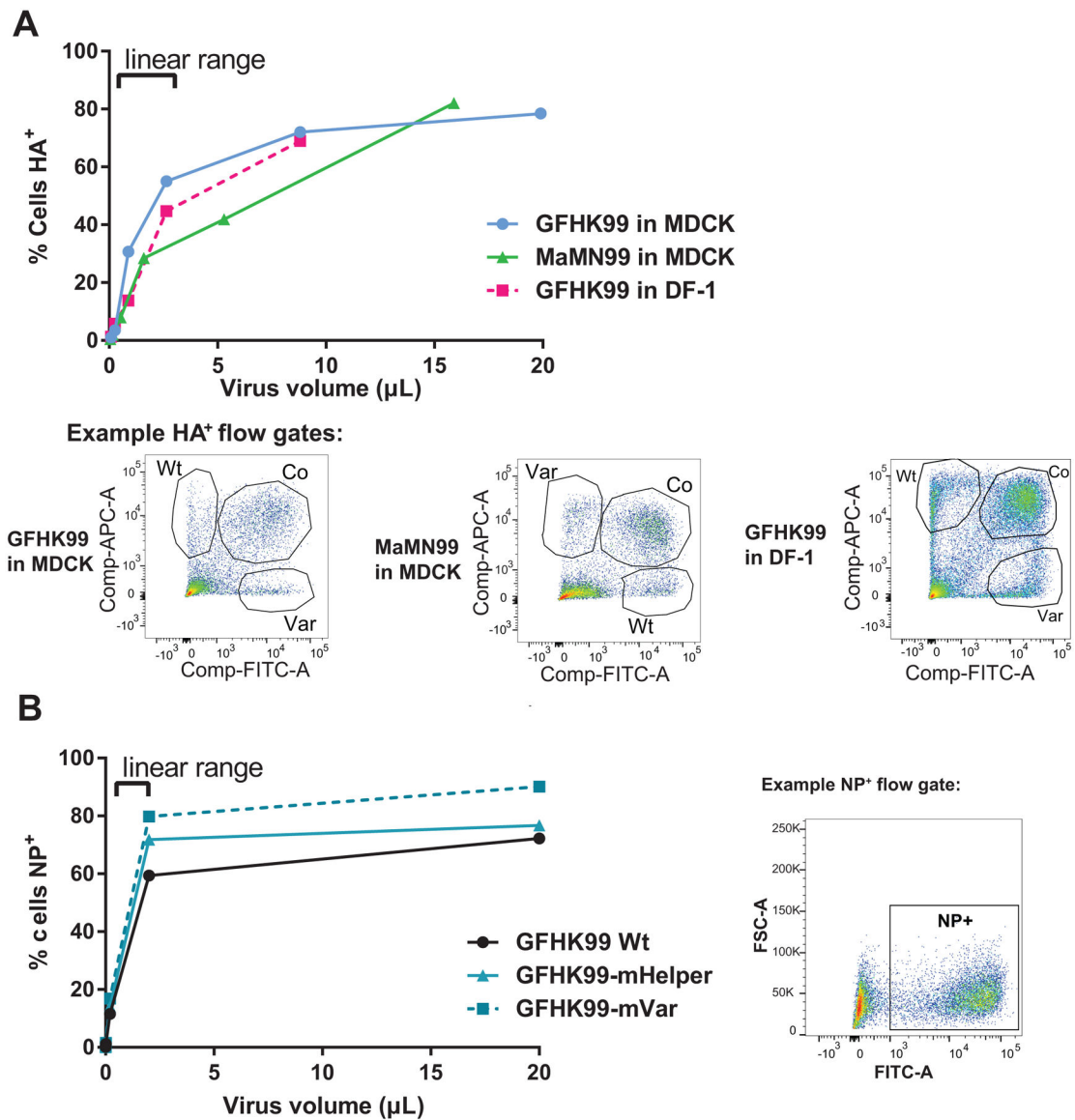
**Extended Data Fig. 8. Quantification of defective interfering RNA content in virus stocks.** Defective interfering RNA content for A) PB2, B) PB1 and C) PA segments was determined by ddPCR using primer pairs targeting terminal and internal portions of each polymerase gene segment to determine their absolute copy number and produce a ratio of terminal:internal copies. All virus stocks used in this study contained low DI content (terminal:internal ratio less than or equal to 2). A DI-rich control virus, Pan99wt P3 (A/ Panama/2007/99 [H3N2]), is included for comparison. This virus stock was passaged three times in MDCK cells at high MOI. For the MaMN99-GFHK99 chimeric viruses, the segments derived from GFHK99 virus are listed in place of the full strain names.



**HA expression of MaMN99 in MDCK cells:****HA expression of GFHK99 in MDCK cells:****HA expression of GFHK99 in DF-1 cells:**

**Extended Data Fig. 9. Example gating for flow cytometry to evaluate HA positive cell numbers.** Plots shown were generated in the course of experiments reported in Figure 1 and are representative of results obtained in at least three independent experiments. Following staining for cell-surface HA protein: 1) A population of cells was selected by gating out cell debris by SSC-A vs FSC-A. 2) Multiplets were excluded by gating for single cells in SSC-H vs SSC-W. 3) Populations of infected cells were gated based on expression of the appropriate epitope tag.





**Extended Data Fig. 10. Titration of virus stocks for HA expressing units and NP expressing units by flow cytometry.**

A) The doses to be used in RNA kinetics studies shown in Extended Data Figure 4 were determined via flow titration of HA-expressing units in the relevant cell lines. GFHK99 and MaMN99 virus mixtures were titrated in MDCK and DF-1 cells to calculate HA-expressing units/mL for each virus-cell combination. Serial dilutions of virus were used to infect cells under synchronized, single cycle conditions. Cells were harvested at 24 h post infection and stained for epitope tags. Data within the linear range were used to calculate the viral titer. B) GFHK99 viruses used in mRNA sequencing experiments shown in Figure 5 were titrated in DF-1 cells. DF-1 cells are more permissive to infection and thus give more sensitive detection of infectious virus compared to MDCK cells. As the virus strains used did not contain epitope tags, virus detection was accomplished through cell permeabilization and detection of the viral NP protein. Data within the linear range were used to calculate viral

titers. Representative flow plots show gates used following exclusion of cell debris and doublets.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgements

We thank David Stallknecht (University of Georgia) for providing the A/mallard/MN/199106/99 (H3N8) biological isolate. We are grateful to Jessica Shartouny for helpful discussion and preparation of Extended Data Figure 7. We thank Hui Tao, Shamika Danzy, and Ginger Geiger for technical assistance. This work was funded in part by NIH R01 AI127799 (to ACL and DRP); NIH/NIAID Centers of Excellence in Influenza Research and Surveillance (CEIRS), contract numbers HHSN272201400004C (to ACL and GST) and HHSN272201400008C (to DRP). Additional funds were provided by the Georgia Research Alliance and the Georgia Poultry Federation (to DRP) and NIH/NIAID Genomic Centers for Infectious Diseases (GCID), Award Number U19 AI110819 (to GST). KLP was supported by T32 AI10669.

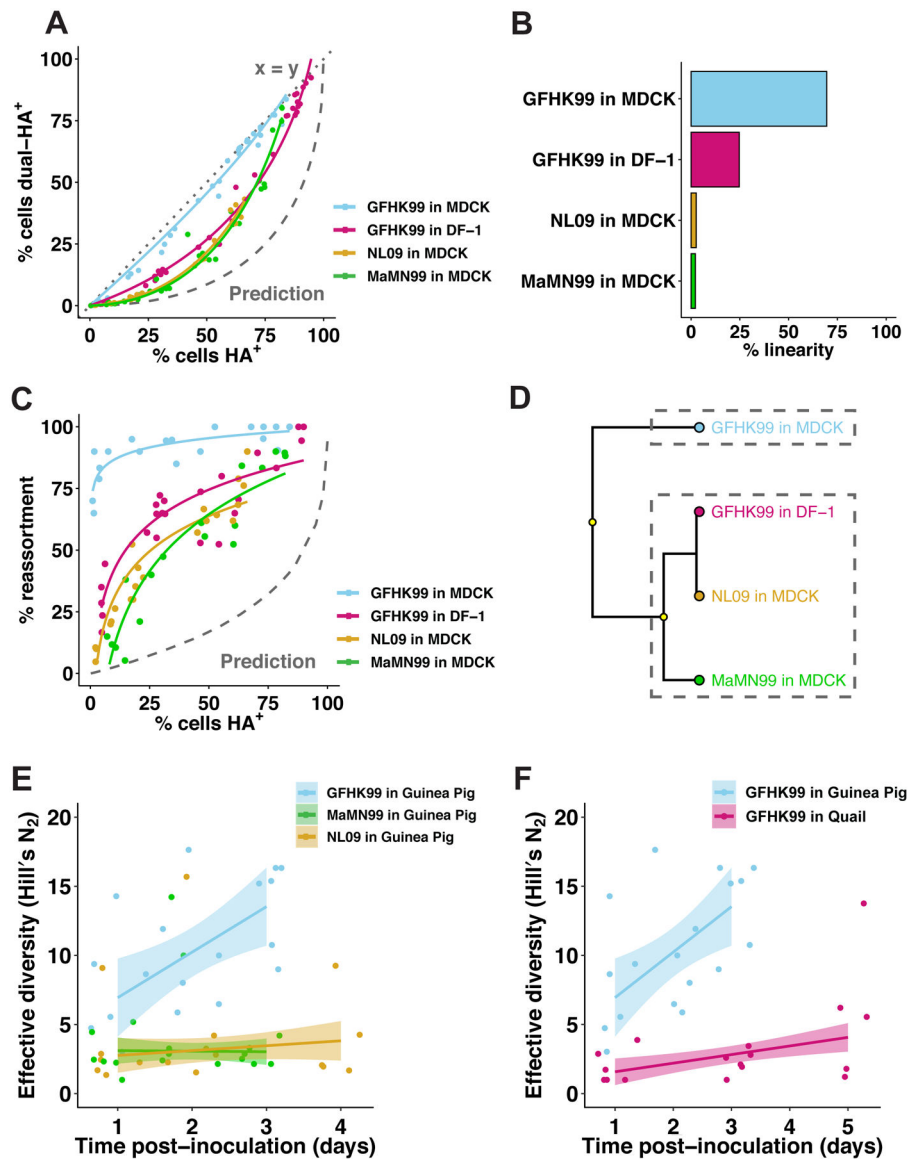
## References

1. Leeks A, Sanjuan R & West SA The evolution of collective infectious units in viruses. *Virus Res* 265, 94–101, doi:10.1016/j.virusres.2019.03.013 (2019). [PubMed: 30894320]
2. Brooke CB Population Diversity and Collective Interactions during Influenza Virus Infection. *J Virol* 91, doi:10.1128/JVI.01164-17 (2017).
3. Sanjuan R Collective Infectious Units in Viruses. *Trends Microbiol* 25, 402–412, doi:10.1016/j.tim.2017.02.003 (2017). [PubMed: 28262512]
4. Brooke CB, Ince WL, Wei J, Bennink JR & Yewdell JW Influenza A virus nucleoprotein selectively decreases neuraminidase gene-segment packaging while enhancing viral fitness and transmissibility. *Proc Natl Acad Sci U S A* 111, 16854–16859, doi:10.1073/pnas.1415396111 (2014). [PubMed: 25385602]
5. Brooke CB et al. Most influenza A virions fail to express at least one essential viral protein. *J Virol* 87, 3155–3162, doi:10.1128/JVI.02284-12 (2013). [PubMed: 23283949]
6. Sun J & Brooke CB Influenza A Virus Superinfection Potential Is Regulated by Viral Genomic Heterogeneity. *MBio* 9, doi:10.1128/mBio.01761-18 (2018).
7. Fonville JM, Marshall N, Tao H, Steel J & Lowen AC Influenza Virus Reassortment Is Enhanced by Semi-infectious Particles but Can Be Suppressed by Defective Interfering Particles. *PLoS Pathog* 11, e1005204, doi:10.1371/journal.ppat.1005204 (2015). [PubMed: 26440404]
8. Jacobs NT et al. Incomplete influenza A virus genomes occur frequently but are readily complemented during localized viral spread. *Nature Communications* (2019).
9. Nayak DP Defective interfering influenza viruses. *Annu Rev Microbiol* 34, 619–644, doi:10.1146/annurev.mi.34.100180.003155 (1980). [PubMed: 7002033]
10. Von Magnus P Incomplete forms of influenza virus. *Adv Virus Res* 2, 59–79 (1954). [PubMed: 13228257]
11. Brooke CB Biological activities of ‘noninfectious’ influenza A virus particles. *Future Virol* 9, 41–51, doi:10.2217/fvl.13.118 (2014). [PubMed: 25067941]
12. Timm C, Gupta A & Yin J Robust kinetics of an RNA virus: Transcription rates are set by genome levels. *Biotechnology and bioengineering* 112, 1655–1662, doi:10.1002/bit.25578 (2015). [PubMed: 25726926]
13. Boulle M et al. HIV Cell-to-Cell Spread Results in Earlier Onset of Viral Gene Expression by Multiple Infections per Cell. *PLoS Pathog* 12, e1005964, doi:10.1371/journal.ppat.1005964 (2016). [PubMed: 27812216]
14. Sanjuán R & Thoulouze M-I Why viruses sometimes disperse in groups?(†). *Virus Evol* 5, vez014–vez014, doi:10.1093/ve/vez014 (2019). [PubMed: 31249695]

15. Sigal A et al. Cell-to-cell spread of HIV permits ongoing replication despite antiretroviral therapy. *Nature* 477, 95–98, doi:10.1038/nature10347 (2011). [PubMed: 21849975]
16. Webster RG, Hinshaw VS, Bean WJ Jr., Turner B & Shortridge KF Influenza viruses from avian and porcine sources and their possible role in the origin of human pandemic strains. *Dev Biol Stand* 39, 461–468 (1977). [PubMed: 604130]
17. Wright PF, Neumann G & Kawaoka Y in *Fields Virology Vol. 1* (ed D. M & Howley Knipe PM) 1691–1740 (Lippincott-Raven, 2006).
18. Webster RG, Bean WJ, Gorman OT, Chambers TM & Kawaoka Y Evolution and ecology of influenza A viruses. *Microbiol Rev* 56, 152–179 (1992). [PubMed: 1579108]
19. Webster RG, Shortridge KF & Kawaoka Y Influenza: interspecies transmission and emergence of new pandemics. *FEMS Immunol Med Microbiol* 18, 275–279, doi:S0928–8244(97)00058–8 [pii] (1997). [PubMed: 9348163]
20. Taubenberger JK & Morens DM 1918 Influenza: the mother of all pandemics. *Emerg Infect Dis* 12, 15–22 (2006). [PubMed: 16494711]
21. Viboud C, Miller M, Olson D, Osterholm M & Simonsen L Preliminary Estimates of Mortality and Years of Life Lost Associated with the 2009 A/H1N1 Pandemic in the US and Comparison with Past Influenza Seasons. *PLoS Curr*, RRN1153, doi:k/-/-/35hpbbyfwdwl4n/8 [pii] (2010). [PubMed: 20352125]
22. Marshall N, Priyamvada L, Ende Z, Steel J & Lowen AC Influenza virus reassortment occurs with high frequency in the absence of segment mismatch. *PLoS Pathog* 9, e1003421, doi:10.1371/journal.ppat.1003421 (2013). [PubMed: 23785286]
23. Perez DR et al. Role of quail in the interspecies transmission of H9 influenza A viruses: molecular changes on HA that correspond to adaptation from ducks to chickens. *J Virol* 77, 3148–3156, doi:10.1128/jvi.77.5.3148-3156.2003 (2003). [PubMed: 12584339]
24. Russell AB, Trapnell C & Bloom JD Extreme heterogeneity of influenza virus infection in single cells. *Elife* 7, doi:10.7554/eLife.32303 (2018).
25. Ramos I et al. Innate Immune Response to Influenza Virus at Single-Cell Resolution in Human Epithelial Cells Revealed Paracrine Induction of Interferon Lambda 1. *J Virol* 93, doi:10.1128/JVI.00559-19 (2019).
26. Cristobal Vera J et al. A common pattern of influenza A virus single cell gene expression heterogeneity governs the innate antiviral response to infection. *bioRxiv* (2019).
27. Zheng GX et al. Massively parallel digital transcriptional profiling of single cells. *Nat Commun* 8, 14049, doi:10.1038/ncomms14049 (2017). [PubMed: 28091601]
28. Butt KM et al. Human infection with an avian H9N2 influenza A virus in Hong Kong in 2003. *J Clin Microbiol* 43, 5760–5767, doi:43/11/5760 [pii] 10.1128/JCM.43.11.5760-5767.2005 (2005). [PubMed: 16272514]
29. Peiris M et al. Human infection with influenza H9N2. *Lancet* 354, 916–917, doi:S0140673699033115 [pii] (1999). [PubMed: 10489954]
30. Guan Y et al. H9N2 influenza viruses possessing H5N1-like internal genomes continue to circulate in poultry in southeastern China. *J Virol* 74, 9372–9380 (2000). [PubMed: 11000205]
31. Guan Y, Shortridge KF, Krauss S & Webster RG Molecular characterization of H9N2 influenza viruses: were they the donors of the “internal” genes of H5N1 viruses in Hong Kong? *Proc Natl Acad Sci U S A* 96, 9363–9367 (1999). [PubMed: 10430948]
32. Lam TT et al. The genesis and source of the H7N9 influenza viruses causing human infections in China. *Nature* 502, 241–244, doi:10.1038/nature12515nature12515 [pii] (2013). [PubMed: 23965623]
33. Wu A et al. Sequential Reassortments Underlie Diverse Influenza H7N9 Genotypes in China. *Cell Host Microbe* 4, 446–452, doi:S1931–3128(13)00298–9 [pii] 10.1016/j.chom.2013.09.001 (2013).
34. Jacobs NT et al. Incomplete influenza A virus genomes occur frequently but are readily complemented during localized viral spread. *Nat Commun* 10, 3526, doi:10.1038/s41467-019-11428-x (2019). [PubMed: 31387995]
35. McCrone JT et al. Stochastic processes constrain the within and between host evolution of influenza virus. *Elife* 7, doi:10.7554/eLife.35962 (2018).

36. Valesano AL et al. Influenza B Viruses Exhibit Lower Within-Host Diversity than Influenza A Viruses in Human Hosts. *J Virol* 94, doi:10.1128/JVI.01710-19 (2020).
37. Jagger BW et al. An overlapping protein-coding region in influenza A virus segment 3 modulates the host response. *Science* 337, 199–204, doi:10.1126/science.1222213 (2012). [PubMed: 22745253]
38. Te Velthuis AJ & Fodor E Influenza virus RNA polymerase: insights into the mechanisms of viral RNA synthesis. *Nat Rev Microbiol* 14, 479–493, doi:10.1038/nrmicro.2016.87 (2016). [PubMed: 27396566]
39. Dias A et al. The cap-snatching endonuclease of influenza virus polymerase resides in the PA subunit. *Nature* 458, 914–918, doi:10.1038/nature07745 (2009). [PubMed: 19194459]
40. Fan H et al. Structures of influenza A virus RNA polymerase offer insight into viral genome replication. *Nature* 573, 287–290, doi:10.1038/s41586-019-1530-7 (2019). [PubMed: 31485076]
41. Gaucherand L et al. The Influenza A Virus Endoribonuclease PA-X Usurps Host mRNA Processing Machinery to Limit Host Gene Expression. *Cell reports* 27, 776–792 e777, doi:10.1016/j.celrep.2019.03.063 (2019). [PubMed: 30995476]
42. Khaperskyy DA, Schmalig S, Larkins-Ford J, McCormick C & Gaglia MM Selective Degradation of Host RNA Polymerase II Transcripts by Influenza A Virus PA-X Host Shutoff Protein. *PLoS Pathog* 12, e1005427, doi:10.1371/journal.ppat.1005427 (2016). [PubMed: 26849127]
43. Chao L, Tran T & Matthews C Muller’s Ratchet and the Advantage of Sex in the Rna Virus 6. *Evolution; international journal of organic evolution* 46, 289–299, doi:10.1111/j.1558-5646.1992.tb02038.x (1992). [PubMed: 28564033]
44. Froissart R et al. Co-infection weakens selection against epistatic mutations in RNA viruses. *Genetics* 168, 9–19, doi:10.1534/genetics.104.030205 (2004). [PubMed: 15454523]
45. Novella IS, Reissig DD & Wilke CO Density-dependent selection in vesicular stomatitis virus. *J Virol* 78, 5799–5804, doi:10.1128/JVI.78.11.5799-5804.2004 (2004). [PubMed: 15140977]
46. Wilke CO & Novella IS Phenotypic mixing and hiding may contribute to memory in viral quasispecies. *BMC Microbiol* 3, 11, doi:10.1186/1471-2180-3-11 (2003). [PubMed: 12795816]
47. Danzy S et al. Mutations to PB2 and NP proteins of an avian influenza virus combine to confer efficient growth in primary human respiratory cells. *J Virol* 88, 13436–13446, doi:10.1128/JVI.01093-14 (2014). [PubMed: 25210184]
48. 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 U S A* 97, 6108–6113, doi:10.1073/pnas.100133697 (2000). [PubMed: 10813369]
49. Schwartz SL & Lowen AC Droplet digital PCR: A novel method for detection of influenza virus defective interfering particles. *J Virol Methods* 237, 159–165, doi:10.1016/j.jviromet.2016.08.023 (2016). [PubMed: 27590979]
50. Perez DR, Webby RJ, Hoffmann E & Webster RG Land-based birds as potential disseminators of avian mammalian reassortant influenza A viruses. *Avian Dis* 47, 1114–1117 (2003). [PubMed: 14575124]
51. Song H, Nieto GR & Perez DR A new generation of modified live-attenuated avian influenza viruses using a two-strategy combination as potential vaccine candidates. *J Virol* 81, 9238–9248, doi:10.1128/JVI.00893-07 [pii] 10.1128/JVI.00893-07 (2007). [PubMed: 17596317]
52. Sorrell EM, Wan H, Araya Y, Song H & Perez DR Minimal molecular constraints for respiratory droplet transmission of an avian-human H9N2 influenza A virus. *Proc Natl Acad Sci U S A* 106, 7565–7570, doi:10.1073/pnas.0900877106 [pii] 10.1073/pnas.0900877106 (2009). [PubMed: 19380727]
53. Brown JD et al. Intestinal excretion of a wild bird-origin H3N8 low pathogenic avian influenza virus in mallards (*Anas platyrhynchos*). *Journal of wildlife diseases* 48, 991–998, doi:10.7589/2011-09-280 (2012). [PubMed: 23060500]
54. Chen H et al. Partial and full PCR-based reverse genetics strategy for influenza viruses. *PLoS One* 7, e46378, doi:10.1371/journal.pone.0046378 [pii] 10.1371/journal.pone.0046378 (2012). [PubMed: 23029501]
55. Matlin KS, Reggio H, Helenius A & Simons K Infectious entry pathway of influenza virus in a canine kidney cell line. *J Cell Biol* 91, 601–613, doi:10.1083/jcb.91.3.601 (1981). [PubMed: 7328111]

56. Yoshimura A & Ohnishi S Uncoating of influenza virus in endosomes. *J Virol* 51, 497–504 (1984). [PubMed: 6431119]
57. Phipps KL et al. Seasonal H3N2 and 2009 Pandemic H1N1 Influenza A Viruses Reassort Efficiently but Produce Attenuated Progeny. *J Virol* 91, doi:10.1128/JVI.00830-17 (2017).
58. Reed LJ & Muench H A SIMPLE METHOD OF ESTIMATING FIFTY PER CENT ENDPOINTS. *American Journal of Epidemiology* 27, 493–497, doi:10.1093/oxfordjournals.aje.a118408 (1938).
59. Campbell PJ et al. The M segment of the 2009 pandemic influenza virus confers increased NA activity, filamentous morphology and efficient contact transmissibility to A/Puerto Rico/8/1934-based reassortant viruses. *J Virol*, doi:JVI.03607–13 [pii] 10.1128/JVI.03607-13 (2014).
60. Wittwer CT, Reed GH, Gundry CN, Vandersteen JG & Pryor RJ High-resolution genotyping by amplicon melting analysis using LCGreen. *Clin Chem* 49, 853–860 (2003). [PubMed: 12765979]
61. Richard M, Herfst S, Tao H, Jacobs NT & Lowen AC Influenza A virus reassortment is limited by anatomical compartmentalization following co-infection via distinct routes. *J Virol*, doi:10.1128/JVI.02063-17 (2017).
62. Simpson EH Measurement of diversity. *Nature* 163, 688–688 (1949).
63. Hill MO Diversity and evenness: a unifying notation and its consequences. *Ecology* 54, 427–432 (1973).
64. Jost L Entropy and diversity. *Oikos* 113, 363–375 (2006).
65. Ward J Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association* 48, doi:10.2307/2282967 (1963).
66. Zhou B et al. Single-reaction genomic amplification accelerates sequencing and vaccine production for classical and Swine origin human influenza A viruses. *J Virol* 83, 10309–10313, doi:JVI.01109–09 [pii] 10.1128/JVI.01109-09 (2009). [PubMed: 19605485]
67. Drayman N, Kler S, Ben-nun-Shaul O & Oppenheim A Rapid method for SV40 titration. *J Virol Methods* 164, 145–147, doi:10.1016/j.jviromet.2009.12.003 (2010). [PubMed: 20005903]
68. Kawakami E et al. Strand-specific real-time RT-PCR for distinguishing influenza vRNA, cRNA, and mRNA. *J Virol Methods* 173, 1–6, doi:10.1016/j.jviromet.2010.12.014 (2011). [PubMed: 21185869]
69. Pflug A, Guilligay D, Reich S & Cusack S Structure of influenza A polymerase bound to the viral RNA promoter. *Nature* 516, 355–360, doi:10.1038/nature14008 (2014). [PubMed: 25409142]
70. Lukarska M et al. Structural basis of an essential interaction between influenza polymerase and Pol II CTD. *Nature* 541, 117–121, doi:10.1038/nature20594 (2017). [PubMed: 28002402]

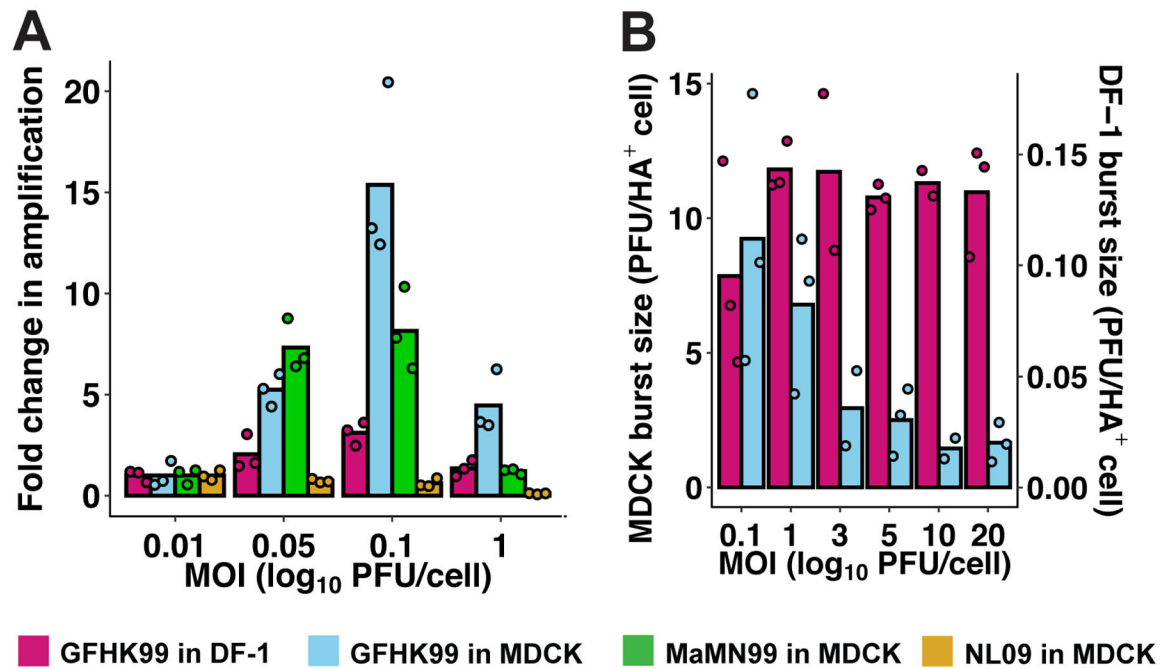


**Figure 1. Coinfection and reassortment frequencies indicate that IAV multiplicity dependence varies with virus strain and host species.**

A–D) MDCK or DF-1 cells were coinfecting with homologous WT and VAR viruses of either GFHK99, MaMN99, or NL09 strain backgrounds at MOIs ranging from 10 to 0.01 PFU/cell. The relationship between % cells HA<sup>+</sup> and % cells dual-HA<sup>+</sup> (A) varies with strain and cell type, resulting in curves of differing % linearity (B). GFHK99, MaMN99, and NL09 viruses exhibit different reassortment levels in MDCK cells, but all show high reassortment relative to a theoretical prediction in which singly infected and multiply infected cells have equivalent burst sizes (C). GFHK99 virus reassortment levels differ in MDCK and DF-1 cells, but again reassortment in DF-1 cells remains high relative to the theoretical prediction in which multiple infection confers no advantage (C). Clustering analysis of reassortment and HA co-expression regression models determines that GFHK99 virus exhibits unique behavior in MDCK cells compared to DF-1 cells or other viruses in MDCK cells (D). Yellow circles indicate nodes with >95% bootstrap support. Panels (E) and

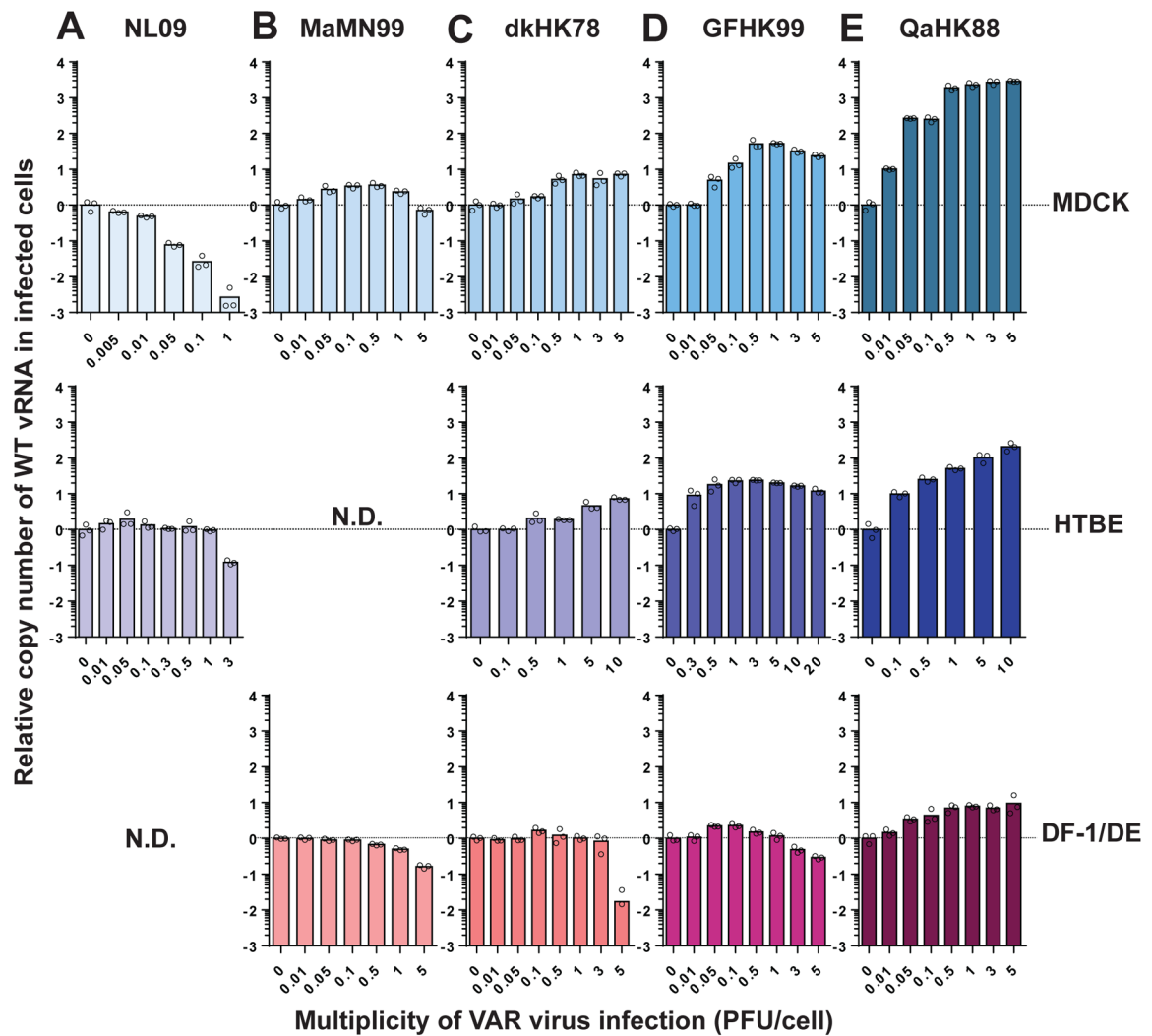


(F) show results of WT/VAR coinfections performed *in vivo*. Because multicycle replication *in vivo* allows the propagation of reassortants, analysis of genotypic diversity rather than percent reassortment is more informative for these experiments. Thus, the effective diversity was calculated for each dataset and plotted as a function of time post-inoculation. In guinea pigs (n=6), GFHK99 WT and VAR<sub>1</sub> viruses exhibit higher reassortment than MaMN99 or NL09 WT and VAR viruses, as indicated by increased genotypic diversity (E). The GFHK99 WT and VAR<sub>1</sub> viruses exhibit higher reassortment in guinea pigs than in quail (n=5) (F). Guinea pig data shown in panels E and F are the same. NL09 virus reassortment data shown in (C) were reported previously<sup>57</sup>. Lines and shading represent prediction mean and 95% prediction interval ( $\text{mean} \pm 1.96 * \text{standard error}$ ) of robust linear regression.

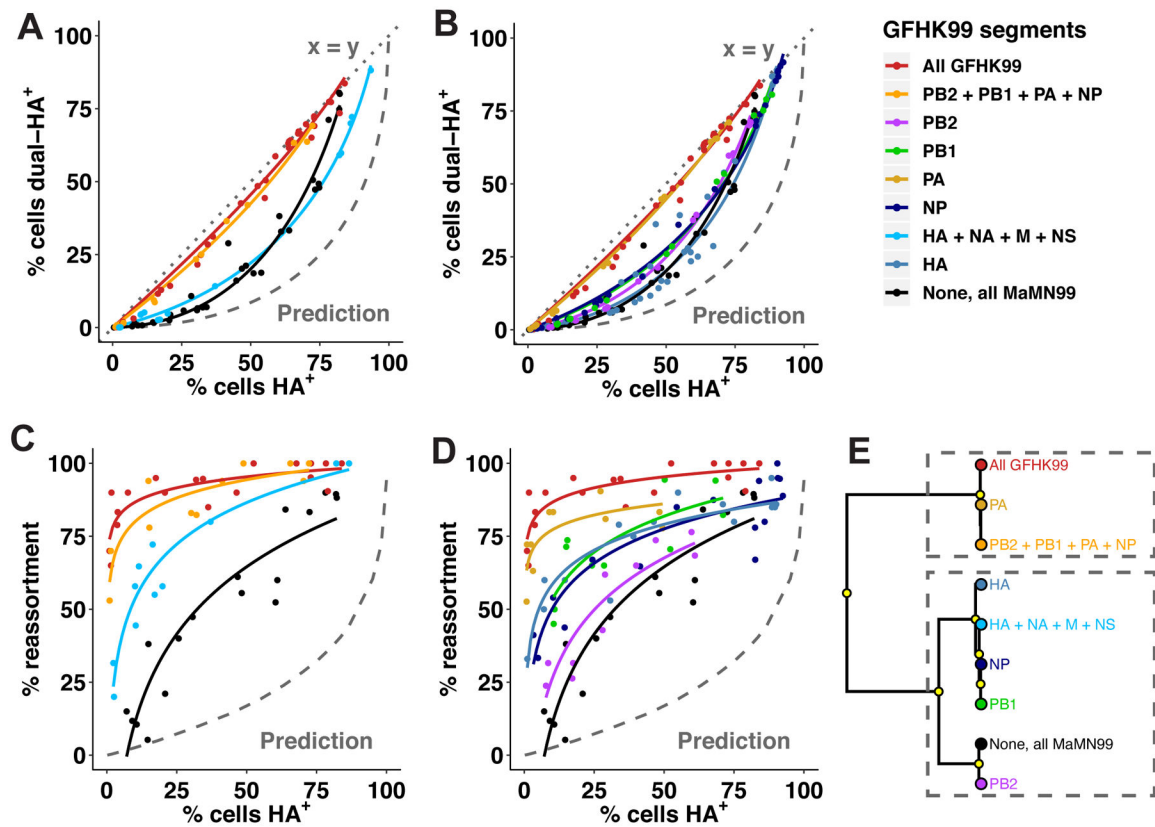


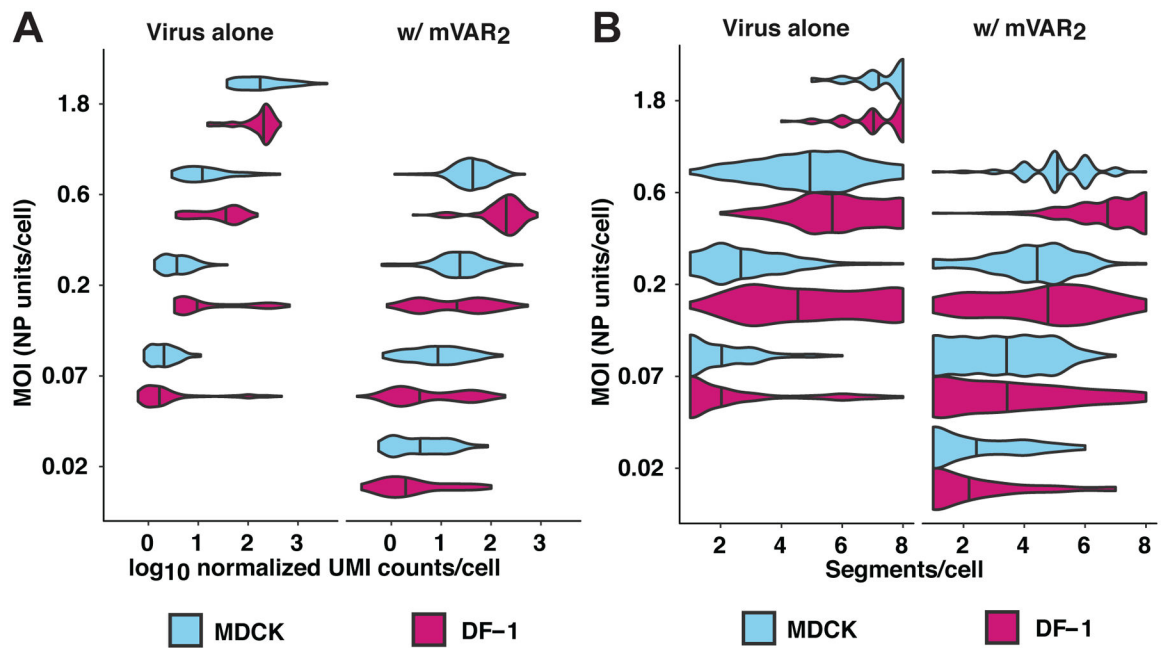
**Figure 2. Increasing MOI increases viral productivity at sub-saturating, but not saturating MOIs.**

MDCK or DF-1 cells were infected under single-cycle conditions at a range of MOIs with the viruses indicated, and released virus was sampled over the course of 24 h. A) Fold change in amplification (maximum PFU output / PFU input) relative to the MOI=0.01 PFU per cell condition is plotted for each virus-cell pairing. B) Burst size, calculated as maximum PFU output / number of HA<sup>+</sup> cells detected by flow cytometry, is plotted for each virus-cell pairing tested in the higher MOI range. MOIs shown are in units of PFU per cell, as determined in MDCK cells. Data are shown from n=3 replicate cell cultures per condition; bars show the mean.



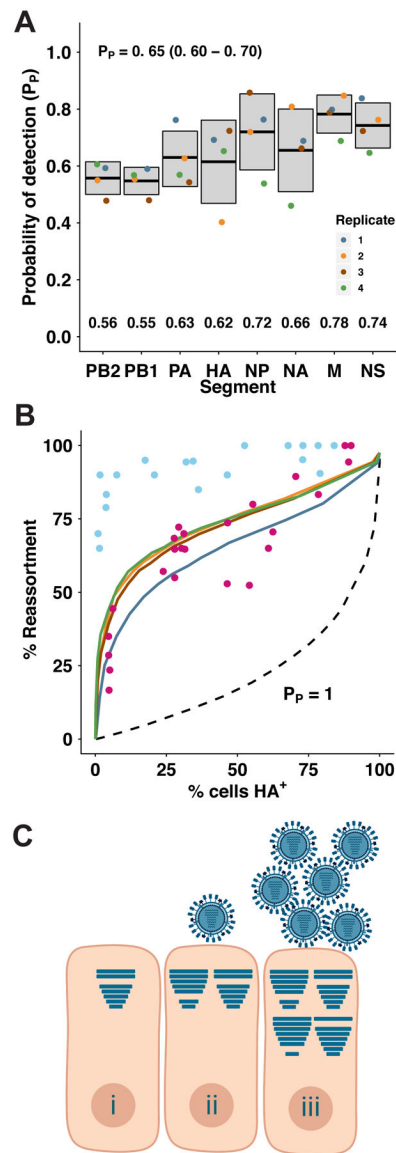
**Figure 3. Coinfection enhances GFHK99 vRNA synthesis in a dose and host dependent manner.** Cells were coinfecting with WT virus and increasing doses of VAR virus. WT virus MOI was 0.05 PFU per cell in NHBE cells and 0.005 PFU per cell in all other cell types. The fold change in WT vRNA copy number, relative to that detected in the VAR virus, is plotted for NL09 virus in MDCK and NHBE cells (A), MaMN99 virus in MDCK and duck embryo (DE) cells (B), dkHK78 virus in MDCK, NHBE, and DF-1 cells (C), GFHK99 virus in MDCK, NHBE, and DF-1 cells (D) and QaHK88 virus in MDCK, NHBE, and DF-1 cells (E). Bars represent the mean of  $n=3$  cell cultures per condition.





**Figure 5. Homologous coinfecting virus boosts GFHK99 viral transcription in single cells and reveals comparable rates of segment detection in MDCK and DF-1 cells.**

DF-1 or MDCK cells were infected with GFHK99 virus (left facet) at MOIs of 0.07, 0.2, 0.6, or 1.8 NP-expressing units per cell ( $n = 1,228$  DF-1 cells, 645 MDCK cells), or a 1:1 mixture of GFHK99 WT and GFHK99 mVAR<sub>1</sub> viruses at four different total MOIs (0.02, 0.07, 0.2, 0.6 NP-expressing units per cell,  $n = 462$  DF-1 cells, 671 MDCK cells) and a constant amount of GFHK99 mVAR<sub>2</sub> virus (0.1 PFU per cell in DF-1 cells, 1.0 PFU per cell in MDCK cells) (right facet). Per condition, two replicate wells were infected and cells from these replicates were pooled prior to analysis, giving  $n=1$  sequencing replicate per infection condition. The number of cells analyzed per condition is shown in Extended Data Figure 6. Each violin plot shows the full distribution of total viral RNA (A) or distinct viral genome segments (B) per cell in each cell-MOI infection condition. Vertical lines denote the median of each distribution. UMI = unique molecular identifier.



**Figure 6. Incomplete GFHK99 virus genomes are present but not sufficiently abundant to account for observed reassortment in MDCK cells.**

Incomplete viral genomes were quantified experimentally by a single-cell assay which relies on the amplification of incomplete viral genomes of GFHK99 WT virus (0.018 PFU per cell) by a genetically similar coinfecting virus, GFHK99 VAR<sub>2</sub>. Based on the rate of detection of GFHK99 WT virus segments in this assay, the probability that a given segment would be present and replicated in a singly infected MDCK cell is reported as  $P_p$ . A) Summary of experimental  $P_p$  data.  $n = 4$  biological replicates, distinguished by color. Horizontal bars and shading represent mean  $\pm$  standard deviation. Mean  $P_{p,i}$  values for each segment are indicated at the bottom of the plot area. Average  $P_p$  for each experiment was calculated as the geometric mean of the eight  $P_{p,i}$  values, and the mean  $\pm$  standard deviation of those four summary  $P_p$  values is shown at the top of the plot area. B) Experimentally obtained  $P_{p,i}$  values in (A) were used to parameterize a computational model<sup>7</sup>. Levels of reassortment predicted using the experimentally determined parameters are shown with



colors corresponding to points in (A). Levels of reassortment predicted if  $P_p=1.0$  are shown with the dashed line. Observed reassortment of GFHK99 WT and VAR viruses in MDCK cells are shown with blue circles. Observed reassortment of GFHK99 WT and VAR viruses in DF-1 cells are shown with pink circles. Observed data are the same as those plotted in Figure 1. C) Model: complementation of incomplete viral genomes is necessary but not sufficient for robust progeny production from mammalian cells infected with GFHK99 virus. Cells infected with single virions often replicate incomplete viral genomes and therefore do not produce viral progeny (i). In mammalian cells, complementation of incomplete viral genomes through coinfection may allow progeny production (ii), but robust viral yields require the delivery of additional genomes to the cell, beyond the levels needed for complementation (iii). Horizontal bars within cells represent segments successfully delivered and replicated.