

## RESEARCH ARTICLE

## Library-based analysis reveals segment and length dependent characteristics of defective influenza genomes

Marisa Mendes , Alistair B. Russell \*

Division of Biological Sciences, University of California, San Diego, La Jolla, California, United States of America

\* [a5russell@ucsd.edu](mailto:a5russell@ucsd.edu) OPEN ACCESS

**Citation:** Mendes M, Russell AB (2021) Library-based analysis reveals segment and length dependent characteristics of defective influenza genomes. *PLoS Pathog* 17(12): e1010125. <https://doi.org/10.1371/journal.ppat.1010125>

**Editor:** Christopher B. Brooke, University of Illinois at Urbana-Champaign, UNITED STATES

**Received:** September 12, 2021

**Accepted:** November 17, 2021

**Published:** December 9, 2021

**Copyright:** © 2021 Mendes, Russell. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** Pipeline can be found at [https://github.com/a5russell/Defective\\_Library\\_Mendes\\_Russell](https://github.com/a5russell/Defective_Library_Mendes_Russell). All sequencing data can be found under the GEO accession GSE183442.

**Funding:** This work was supported by the NIAID of the NIH under grant AI141678 and by the Damon Runyon Cancer Research Foundation Dale F. Frey Award (damonrunyon.org), DFS 36-19 awarded to ABR. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of this manuscript.

## Abstract

Found in a diverse set of viral populations, defective interfering particles are parasitic variants that are unable to replicate on their own yet rise to relatively high frequencies. Their presence is associated with a loss of population fitness, both through the depletion of key cellular resources and the stimulation of innate immunity. For influenza A virus, these particles contain large internal deletions in the genomic segments which encode components of the heterotrimeric polymerase. Using a library-based approach, we comprehensively profile the growth and replication of defective influenza species, demonstrating that they possess an advantage during genome replication, and that exclusion during population expansion reshapes population composition in a manner consistent with their final, observed, distribution in natural populations. We find that an innate immune response is not linked to the size of a deletion; however, replication of defective segments can enhance their immunostimulatory properties. Overall, our results address several key questions in defective influenza A virus biology, and the methods we have developed to answer those questions may be broadly applied to other defective viruses.

## Author summary

Viral populations can generate defective interfering particles, parasitic variants that decrease, rather than increase, the fitness of the population. There has been considerable interest in use of such particles as therapeutics or vaccine adjuvants. However, for many viruses, data are currently lacking regarding the features of these particles that allow them to outcompete their replication-competent progenitors, as well as why and how some trigger host immune pathways. Exploring this phenomenon in influenza A virus, wherein defective particles predominantly consist of deletions in the three viral genomic segments that encode the viral polymerase, we generated a range of deletions and duplications in a segment wherein such particles are commonly observed, PB1, and one in which deletions are only rarely seen in sequencing studies, HA. In doing so, we find that, for both HA and PB1 deletions result in small molecules that are greatly advantaged during genome replication, yet are at least partially excluded from the population at the step of viral packaging.

**Competing interests:** The authors have declared that no competing interests exist.

Exploring the use of such particles to stimulate cell intrinsic innate immunity, we find, as others before us, that deletions are associated with the production of interferons, but this process is largely length-independent and is modulated by viral replication.

## Introduction

It has long been known that the total number of particles in a viral population far exceeds the number of infectious particles. [1]

Early work on this phenomenon revealed that, in addition to fully-infectious particles, there are semi-infectious, interferon-suppressive, and defective-interfering particles (DIPs). [2–6] The latter is marked by their capacity to inhibit, rather than enhance, infection. Interference by DIPs can arise from direct competition for resources within an infected cell, exclusion of replication-competent viruses during packaging, or increasing the production of interferons, key signaling components of cell-intrinsic innate immunity. [7–10] While the genetic basis of DIPs is idiosyncratic, differing from virus to virus, a defining characteristic is their capacity to, despite their defects, rise to high frequencies in viral populations when complemented by wild-type virus in coinfections.

Due to their association with coinfections, DIPs were once largely thought to be artifacts of laboratory-grown viruses. [11] However, clinical sequencing has revealed that they are a common component of viral infections and can be associated with disease outcome. [12–14] Their potential ability to modulate the course of disease has now sparked a renewed interest in their therapeutic potential. [15–18]

For influenza A virus, a (-) sense, segmented, RNA virus, defective particles, interfering or otherwise, largely bear genomes with large internal deletions in the three polymerase-encoding segments. [19–22] These deletions generally reduce the size of each segment from ~2.3kb to ~400nt. [13, 23] It is unknown what drives this size and segment preference. Understanding deletions, from formation through selection, is challenging. Studying *de novo* deletion formation, a stochastic event, requires single-cell resolution, as was performed by Kupke *et al.* (2020) [24], or a highly sensitive sequencing approach in order to disambiguate rare production of deletions from artifacts such as template-switching. [25] Even with those controls, observed deletions remain subject to selective pressures from genome replication. Rather than focus on their formation, and the challenges that presents, we instead chose to study selection on deletions *after* they form in order to understand if these pressures may nevertheless provide an understanding of the ecology of these particles.

Using a combination of PCR and Gibson assembly we seeded an initial population with a variety of deletions and duplications that generated a broad range of lengths, and measured changes in this distribution as a function of progression through the viral life cycle. While validated full-length, amplicon-based, approaches exist for sequencing influenza A virus, they exhibit length-dependent biases, which can artifactually generate apparent differences in deletion abundance due to differences in saturation kinetics. [26] To overcome this limitation, we instead used an approach wherein individual barcoded adapters were associated with a particular deletion, or duplication. Our libraries were similar to those generated using RanDeL-Seq, a recent effort wherein deletions were generated in HIV and Zika genomes using random transposon insertions and exonuclease digestion. [27]

Using our artificial libraries, we demonstrate that deletions in influenza segments lead to significant advantages during genome replication, explaining their abundance. Certain deletions are excluded as populations expand, both in terms of size and segment identity,

explaining, at least in part, the observed distribution in final viral populations. Lastly, we find that deletions can generate aberrant sequences that can drive interferon induction, but their stimulatory behavior does not appear to be correlated with their length, and is enhanced by genome replication.

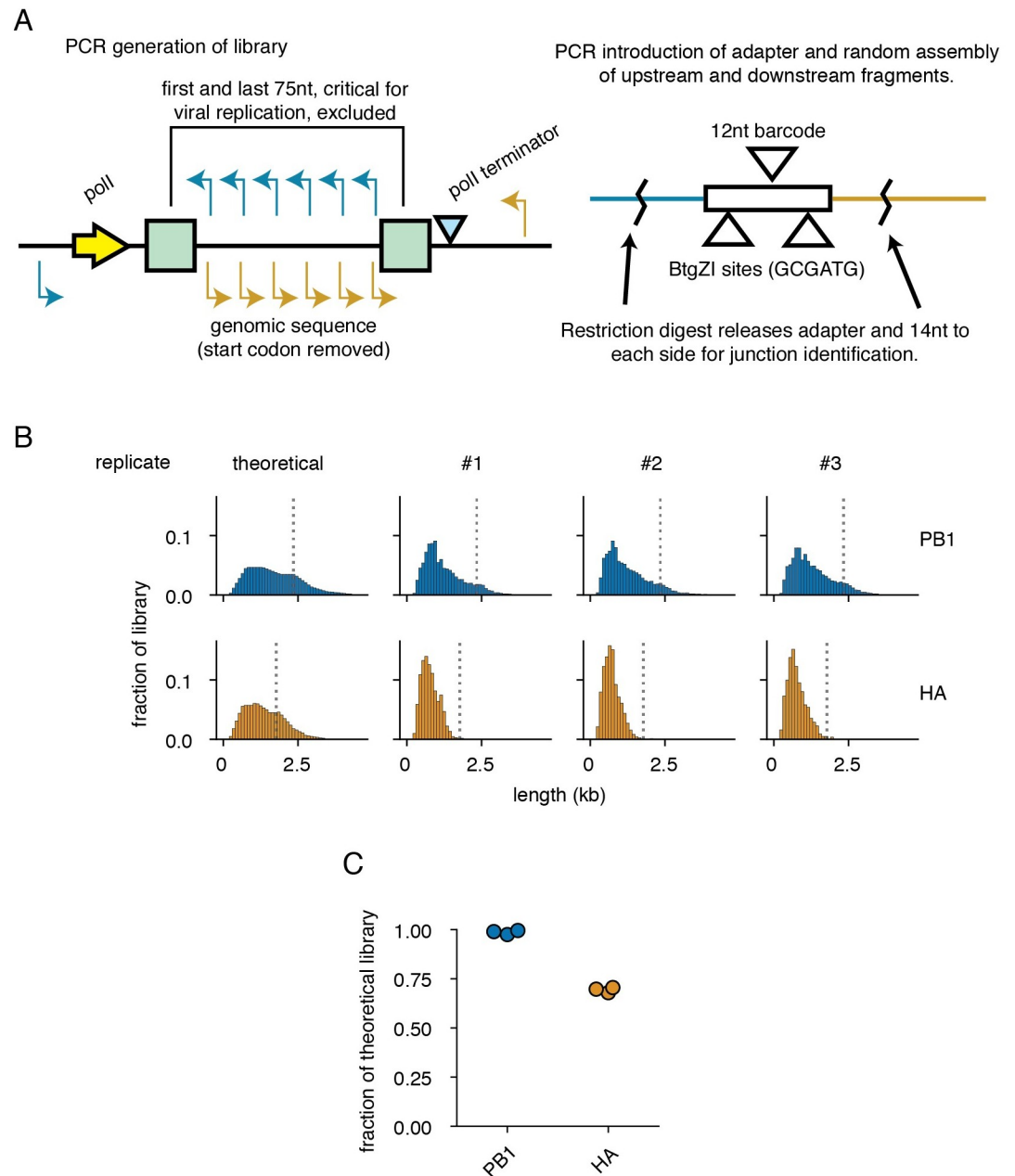
## Results

### PCR-based generation of length polymorphic deletion/duplication libraries

The observed frequencies of variants within an initially clonal population are not only reflective of the fitness benefits, or drawbacks, of any given variant, but also the timing with which they are formed. [28] In other words, the distribution of defective segments is highly confounded by whether a deletion forms early in population expansion or late. To better understand how defective influenza species may compete amongst one-another, we designed a library-based approach wherein we introduce, contemporaneously, a distribution of more than 100 possible lengths representing over 1,000 possible junction combinations into an influenza genomic segment. We thereafter track the distribution of variants within each library as a function of progression through the viral life cycle. Our library approach is described in Fig 1A.

In brief, we computationally designed primers in both the forward and reverse orientations across target influenza segments staggered every 25nt for the first third of the segment, 50nt for the second, and 100nt for the last. The first and last 75nt of each segment were excluded as they contain the viral promoter and critical packaging sequences. [9, 29–32] Each individual primer with the same orientation was paired with a specific, shared, primer, and individual PCRs were performed to generate forward and reverse flanks of staggered sizes. An additional round of PCR was performed on reverse flanks to add a 57bp adapter containing a 12bp random barcode, flanked by cut sites for the type IIS restriction enzyme BtgZI. Gibson assembly was then used to randomly pair forward and reverse flanks, reconstructing a library consisting of both deletions and duplications. [33] Taking advantage of the capacity of type IIS restriction enzymes to cleave outside of their restriction sites, we released barcoded adapters with 14/10nt of flanking sequence with a BtgZI digest. By sequencing these adapters, we may link specific deletions, or duplications, to their corresponding barcode. Thereafter, to infer junction abundance we can sequence across our variable barcodes from constant regions in our adapter. Our barcoding schema avoids the length-bias of target amplicon approaches to sequencing influenza, which overestimate the abundance of shorter species relative to longer. [26]

Influenza A virus defective particles largely consist of those bearing large deletions in the three polymerase-encoding segments, as deletions in the remaining five segments appear only rarely in viral populations. [19–22] There is currently no single explanation for this observation. Possibilities include differences between segments in the rate of deletion formation, replication, or even packaging. [10, 34] To explore whether features other than biases in formation may contribute to the observed differences in abundance, we chose to generate libraries in a single polymerase segment, PB1, encoding the core viral polymerase, and a single non-polymerase segment, HA, encoding the viral receptor-binding and fusion protein hemagglutinin. In order to focus solely on length and sequence composition, and not on whether a given segment produces an essential protein, we used variants of PB1 and HA with mutated start codons. While these variants will still generate products from alternative start codons (such as PB1-F2), as well as “UFO” proteins from upstream start codons generated from cap-snatching, the absence of the essential PB1 or HA proteins means these viruses are not viable, and thus we should not see selection on viability. [35, 36] Using the schema described above, we expect 1,936 possible unique junctions for PB1 and 1,225 possible junctions for HA. After library



**Fig 1. A PCR-based strategy for exploring length-dependent phenomena in viral libraries.** (A) As templates for each library, we used vectors encoding each segment under control of a polymerase I-dependent system for generating authentic viral RNAs. For each library, a set of staggered primers were used in individual PCR reactions with an external primer. Further PCR was used to append a 57nt adapter with a random, 12nt, barcode. Products were then randomly combined in a pooled Gibson assembly reaction. To link barcodes to duplication or deletion junctions, barcodes and flanking regions were freed from the plasmid pool with a BtgZI digest and sequenced. Primers used to generate the library are in *S1 File*. (B) Theoretical and actual length distributions of libraries. Lengths do not include the 57nt adapter. Bins in this and all other panels are 100nt in length, numbers represent the right edge of each bin. Dotted lines show the length of wild-type PB1 and HA. (C) Libraries contain most of the theoretical diversity. Each point represents a single library.

<https://doi.org/10.1371/journal.ppat.1010125.g001>

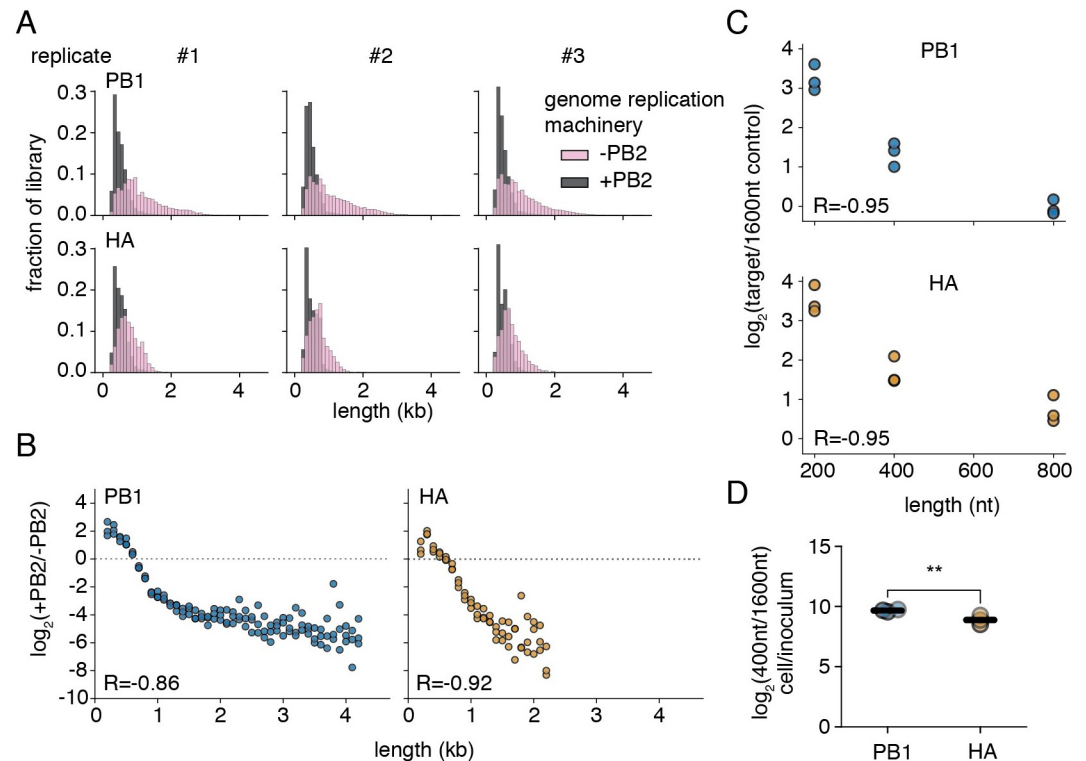
generation, we find a slight bias towards shorter lengths, potentially due to higher molar yield in PCR, and relatively complete representation of unique junctions ( $\sim 98\%$  of theoretical PB1 junctions and  $\sim 70\%$  of theoretical HA junctions) (Fig 1B and 1C).

### Shorter segments are preferentially amplified during genome replication

The first question we wished to address with these libraries was whether length, independent of other features of the viral life cycle, influences genome replication. There are conflicting reports concerning whether shorter segments replicate faster, and thus outcompete their longer counterparts. [10, 37, 38] To make these measurements, we co-transfected libraries with plasmids encoding the minimal influenza genome replication machinery—PB2, PB1, PA, and NP—and measured barcode abundance in viral genomes after 24 hours. To determine enrichment, or depletion, of a given barcode, we generated an orthogonal dataset consisting of the same plasmid libraries transfected with PB1, PA, and NP, but not PB2 (S1 Fig). By generating a dataset free from PB2, we corrected for length-dependent transcription and degradation biases from the polymerase I-dependent viral RNA transcription system; without PB2 these RNA species are neither replicated nor stabilized by the remaining components of the viral replication machinery. [39]

Comparing the composition of libraries between replication-incompetent, and -competent, conditions, there is a noticeable shift towards smaller variants (Fig 2A). Plotting enrichment as a function of length, these two features are negatively correlated, with Spearman correlation coefficients of  $-0.86$  and  $-0.92$  for PB1 and HA respectively (Fig 2B). Smaller segments that arise due to deletions would therefore be expected to possess a considerable replication advantage over their longer counterparts. This observation is in line with prior work from Widjaja *et al.* (2012) [40] showing shorter, *Gussia*, luciferase was preferentially replicated by viral replication machinery over longer, firefly, luciferase. To confirm this conclusion, we generated PB1 and HA segments of defined length (200, 400, 800, and 1600nt) consisting of equal numbers of nucleotides from the 5' and 3' termini of each segment. As in our libraries, we introduced an additional 57nt adapter with a barcode; however, rather than a random barcode we instead used a defined barcode to delineate each size. A control qPCR demonstrates our ability to disambiguate each of these barcodes from one-another, with greater than 1000-fold discrimination for target versus off-target amplification (S2 Fig). As expected, there remains a clear inverse relationship between enrichment in a genome replication assay and length for both segments at a Spearman correlation coefficient of  $-0.95$  for both PB1 and HA (Fig 2C).

However, replication using the minimal viral polymerase machinery does not truly recapitulate the full suite of feedback mechanisms and inter-segment competition that occurs during bona fide viral infection. To confirm if length plays a role in genome replication during actual viral infection we generated two viral populations containing our marked 400nt and 1600nt variants of either HA or PB1—rescuing such populations by coinfection with a wild-type virus to provide missing essential viral genes. We then measured how the ratio of these two species changed during genome replication in an infection of the human lung epithelial carcinoma cell line, A549, at an MOI of 25. This high MOI increases the probability that an infected cell possesses both marked species competing with one-another, as well as our complementing, wild-type, virus. This high MOI also excludes the effects of second rounds of infection due to superinfection exclusion (S3 Fig). [41, 42] At 14 hours we measured the shift in ratio of these two species in these cells and found, as in our transfection assays, a considerable advantage to shorter, over longer, species. Curiously, this effect was slightly stronger for PB1 than for HA (Fig 2D). Therefore a length-dependent replication advantage explains how these parasitic,



**Fig 2. Segment length and genome replication efficiency are inversely correlated.** (A) Size distributions of libraries in transfections with (+PB2) and without (-PB2) the full viral genome replication machinery 24 hours post-transfection in 293t cells. Polymerase length biases are shown in S1 Fig (B) Smaller species are enriched during genome replication. The fraction of variants falling within each 100nt bin was compared between replication-competent (+PB2), and -incompetent (-PB2) selections. Points above the dotted line represent sizes in each individual library which were enriched in the +PB2 dataset. Points were only shown if represented in all three libraries under both conditions R is the Spearman correlation coefficient.  $n = 3$ . (C) qPCR supports the conclusion that genome replication is inversely correlated with length. Vectors encoding variants of four separate lengths (200, 400, 800, and 1600nt) were transfected into 293t cells with and without the full viral genome replication machinery and analyzed by qPCR 24 hours post transfection. The fraction of each variant was calculated relative to the longest variant (1600nt) and this value compared between replication-competent and -incompetent datasets to determine the relative enrichment attributable to genome replication. R is the Spearman correlation coefficient.  $n = 3$ . (D) Validation of qPCR in S2 Fig. (D) Shorter defective viral genomes are enriched during genome replication in viral infection. A549 cells were infected with mixed populations consisting of 400 and 1600nt variants of either HA or PB1 and a complementing wild-type strain at a combined MOI of 25, as measured in genomic equivalents by qPCR. Cells were then washed to remove any additional virus at 2 hours post infection, and, at 14 hours post infection supernatant was removed and cellular RNA harvested. The initial ratio of these two variants was measured in the viral inoculum by qPCR, and compared with the ratio in the cellular fraction. Both PB1 and HA 400nt were enriched over 1600nt by one-tailed t-test, Benjamini-Hochberg corrected FDR  $< 0.05$ . The effect was stronger for PB1 than HA, two-tailed t-test,  $p < 0.05$ . Each point is a single replicate, lines are the mean.  $n = 3$ . Validation that second rounds of infection are excluded is in S3 Fig.

<https://doi.org/10.1371/journal.ppat.1010125.g002>

deleterious, variants rise to frequencies exceeding their longer, replication-competent, progenitors.

### Length-dependent selection reshapes PB1 and HA distributions as viral populations expand

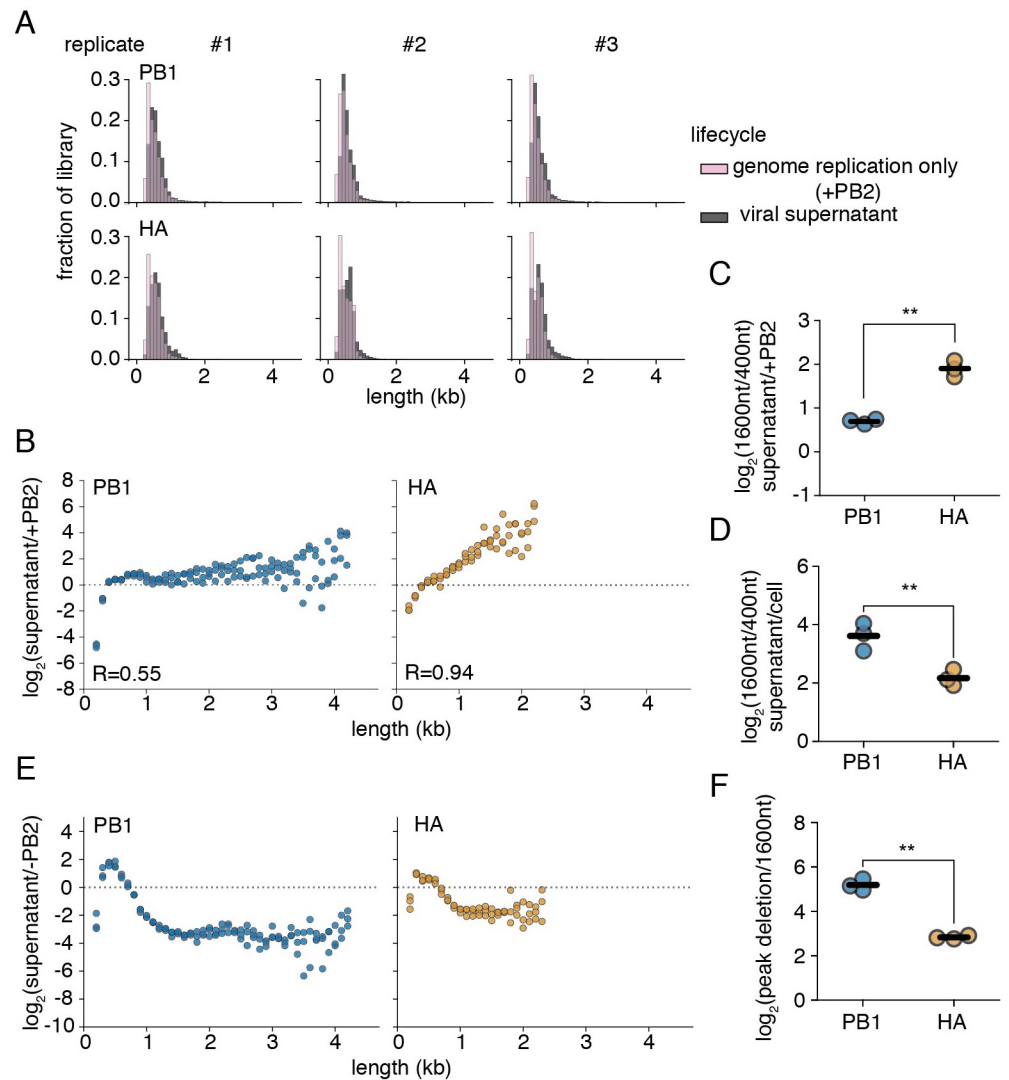
Genome replication is only one step in the viral life cycle at which defective viral genomes may be favored or disfavored. To expand our library study to aspects of the viral life cycle uncaptured in our genome replication assay, we transfected cells with our libraries and the viral replication machinery and waited 24 hours. Thereafter, we infected these transfected cells with

wild-type virus and allowed infection to proceed for 72 hours. During this time, variants within each population could be complemented by coinfection with wild-type virus and compete with one-another. We then harvested clarified supernatant and sequenced those variants that had successfully become incorporated into the viral population.

Comparing the viral supernatant to genome replication-only controls, there are divergent patterns of enrichment between HA and PB1 libraries (Fig 3A). For HA, there is a direct, positive, relationship between length and enrichment in the viral supernatant, with considerable advantage to longer, over shorter, segments (Spearman correlation coefficient of 0.94) (Fig 3B). For PB1, small segments, <400nt of natural sequence, are excluded from the viral population but otherwise preferences are relatively flat, indicating little change in these populations after they were shaped during genome replication. As such there is a much more modest correlation between length and enrichment at a Spearman correlation coefficient of 0.55. To confirm that PB1 and HA exhibit divergent relationships between length and competition as viral populations expand, we used defined 400nt and 1600nt variants and repeated selection as performed for our more extensive libraries. As expected, qPCR analysis finds more robust enrichment of a 1600nt length segment over a 400nt counterpart in HA than in PB1 (Fig 3C). Therefore, features other than genome replication appear to act to largely exclude, or at least blunt the advantages of, deletions in HA relative to those in PB1.

The enrichments in Fig 3B are the results of selection pressure during several steps in the viral life cycle, including additional rounds of genome replication, a requirement for complementation during coinfection with wild-type virus, competition for packaging into viral particles, and the timing and magnitude of viral release. If genome replication alone were the only selection pressure acting on our libraries, we would anticipate that our enrichment curves would be similar to those procured in our initial experiments. To explain our observation, there must exist a length-dependent negative selection that counteracts the advantage of smaller, over larger, variants during genome replication. While initially thought to be a step at which defective viral genomes are preferred, recent data indicate that packaging may instead exclude these species from the population. [34, 37] To determine whether exclusion during packaging may benefit larger, over smaller, variants, we performed experiments as in Fig 2D, but now measured how variant composition differs between the supernatant and the cell fraction at 14 hours post infection. Similar to results from Alnaji *et al.* (2021) [37], we conclude that this step exerts a potent negative selection on our 400nt, relative to our 1600nt, variant (Fig 3D). This effect was slightly stronger for PB1 than for HA, and thus cannot explain the advantages to PB1 deletions over HA that we observe over the course of viral expansion. Other pressures, as yet unexplored, must also contribute to the observed shifts in population length distribution.

Assaying the entire process, from polI transcription to successful encapsidation in viral particles in the supernatant, we see a highly stereotyped enrichment curve for both PB1 and HA, with maxima close to 400nt in length (Fig 3E). Variants with a final length centering around this size become enriched in our library, whereas those that are smaller or larger become depleted. The magnitude of this effect, as measured from the peak enrichment to 1600nt, where preferences flatten but our measurements remain relatively robust, was consistently greater for PB1 than for HA, indicating an increased competitive fitness for smaller species in the former over the latter (Fig 3F). This process was highly replicable between independently-generated libraries with inter-replicate Pearson correlation coefficients of  $\sim 0.75$  when comparing across the enrichment or depletion of individual junctions (S5 Fig). These maxima correlate well with previously-identified “sweet-spot” lengths from continuous culture propagation of influenza defective particles. [23] Taken together, our data indicate that selection on defective segments, both positive and negative, is sufficient to recapitulate their



**Fig 3. PB1 and HA variants exhibit divergent dynamics during population expansion.** (A) Variant distribution after replication alone (+PB2), or in the supernatant after 72h of infection with wild-type, virus. (B) Association of length versus enrichment in the supernatant for PB1 and HA. The fraction of variants falling within each 100nt bin was compared between replication only and supernatant selections. Lengths above the dotted line are enriched during viral expansion. Points were only shown if represented in all three libraries under both conditions R is the Spearman correlation coefficient.  $n = 3$ . (C) qPCR confirmation of data from (B). The ratio of a 1600nt to a 400nt product was compared between replication only and supernatant samples as in (B).  $n = 3$  (D) Packaging excludes smaller variants from the population. Experiments performed as in Fig 2D, with supernatant harvested and compared to the cell fraction at 14 hours.  $n = 3$ . (E) Conflicting pressures during population expansion shape the distribution of defective particles. The fraction of variants falling within each 100nt bin was compared between polymerase I (-PB2), and viral supernatant. Data plotted as in (B)  $n = 3$ . Size distributions for each state shown in S4 Fig. Correlation between replicates shown in S5 Fig (F) Enrichment for smaller PB1 variants was stronger than for HA variants in (E). Maximally-enriched size bin for each library replicate was compared to matched values at 1600nt.  $n = 3$ . For panels (C), (D), and (F), each point represents a single measurement, lines represent means. All measurements significantly differed from no enrichment of the measured species by one-tailed t-test, with Benjamini-Hochberg corrected FDR  $< 0.05$  within each panel. Significant differences between HA and PB1 noted with asterisks,  $p < 0.05$ , two-sample t-test.

<https://doi.org/10.1371/journal.ppat.1010125.g003>



observed natural length and segment variation, regardless of their initial distribution upon formation.

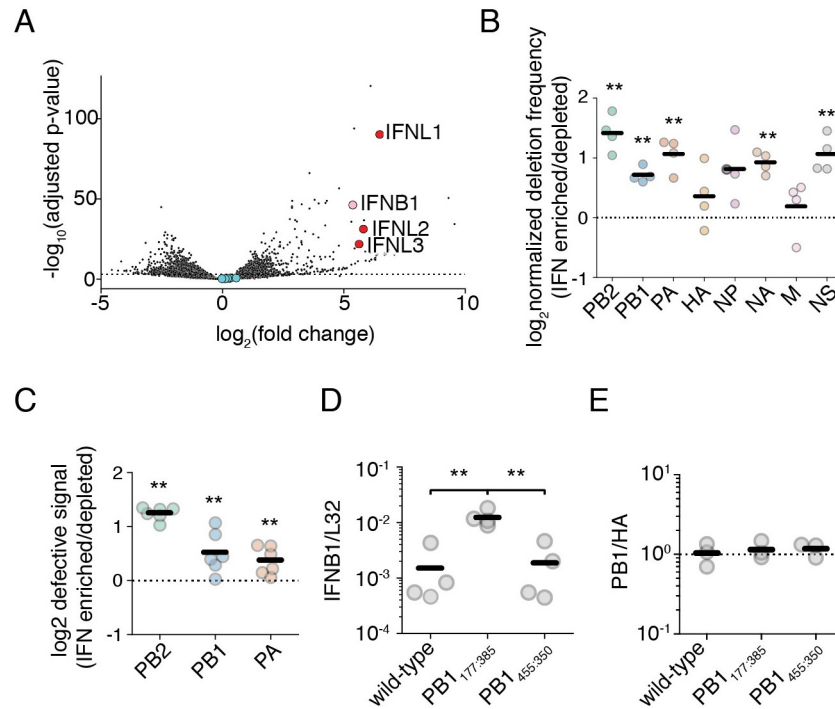
### Deletions in the influenza genome are associated with interferon-producing cells

The work of Tapia *et al.* (2013) [8], since repeated by others, demonstrated that populations replete in defective influenza particles exhibit enhanced interferon induction. [43] To explore further, we wished to probe whether our libraries capture the variation responsible for this increase in the innate immune response. Unlike our previous selections, where viral replication and packaging served to shape variant distribution, understanding how variants may influence innate immune detection is less straightforward. Potentially solving this problem, prior work indicates that observed increases in interferon induction in response to influenza variants are frequently due to more cells producing interferon rather than an increase in the amount of interferon each infected cell produces. [44–46] As such, if we sort infected cells based on interferon production, we would anticipate that we may be able to co-enrich stimulatory variants in our libraries.

Before applying this selection to our libraries, we first sought to understand whether it could recapitulate the earlier findings that deletions broadly associate with interferon induction. To this end, we first generated two wild-type influenza A virus populations that contained an intermediate burden of defective particles, independently, from reverse-genetics onwards (S6 Fig). By working with a population where defective particles are neither rare nor overwhelming, we allow the potential for enrichment of deletions in interferon-producing cells; otherwise, if all cells are infected by defective viruses we may be unable to achieve any additional enrichment, and if few to none are we may lack the sensitivity to detect their presence. We next infected an A549 cell line expressing the cell-surface marker LNGFR under control of the *IFNB1* promoter and sorted populations into interferon-enriched, and -depleted, subsets using magnetic-activated cell sorting. [46] Prior to sequencing viral genomes from these populations, we first validated our sorting efficiency with total mRNA sequencing of both populations (Fig 4A). As expected, there is significant enrichment of the gene encoding the most highly-transcribed type I interferon in A549 cells, interferon beta (*IFNB1*). We also find co-enrichment of type III interferon transcripts (*IFNL1*, *IFNL2*, and *IFNL3*). Additional enriched genes include a number of interferon-stimulated genes, suggesting either direct transcription of a subset of ISG loci by IRE3, or a high level of autocrine signaling experienced by interferon-producing cells.

We next used a full-length PCR-based strategy to sequence influenza genomes. [47] While length-biased, we sought to reduce this bias by initiating reverse transcription with equivalent viral RNA, and limiting cycles of PCR. We identified reads containing deletion junctions by mapping continuous reads with STAR, identifying discontinuous mapping in unmapped reads with BLAST, collapsing all identified junctions into a single annotation file for each biological replicate, and then remapping discontinuous reads using this BLAST-annotated STAR index. [48, 49]. We required that a deletion had three mapped bases to either side of a junction and that read pairmates were concordant. As we used paired-end reading, deletion junction counts were collapsed to a per fragment value rather than per read value. As expected from the highly stochastic nature of deletion formation, while junction counts between infection replicates using the same biological population are highly correlated, those from divergent biological replicates exhibit no detectable correlation to one-another (S7 and S8 Figs).

Deletion-spanning fragments are predominantly derived from the polymerase segments, regardless of whether a sample was enriched or depleted for interferon (S9 Fig). When we



**Fig 4. Interferon-producing cells are enriched for defective influenza segments.** (A) Differential gene expression analysis on interferon-enriched and -depleted populations. An A549 type I interferon reporter cell line was infected at an MOI of 0.1 and sorted using magnetic-activated cell sorting 14 hours post infection. Fold-change calculated as mRNA abundance in interferon-enriched/interferon-depleted datasets. Type I interferon gene *IFNB1* labeled in pink, the type III interferon genes *IFNL1–3* in red, and influenza transcripts in blue.  $n = 4$ . Populations used are described in S6 Fig. Full table of differentially expressed genes in S2 File (B) The fraction of junction-spanning fragments per influenza segment was compared between interferon-enriched and depleted populations. Values above the dotted line indicate segments where deletions are co-enriched with interferon. Asterisks represent significant enrichment, one-sample one-tailed t-test. Stochastic detection of deletions demonstrated in S7 and S8 Figs. Raw data in S9 Fig, mapping depth in S10 Fig  $n = 4$ . (C) qPCR validation of (B) for the three polymerase segments using the assay as described in S11 Fig. Values above the dotted line indicate co-enrichment of deletions with interferon. Asterisks represent significant enrichment, one-sample one-tailed t-test. Triplicate measurements were made of two biological replicates.  $n = 6$ . (D) Different PB1 deletions induce different amounts of interferon. A549 cells were infected by the indicated variants at an MOI of 0.5. *IFNB1* transcripts normalized to the housekeeping control *L32* were measured 8h post infection by qPCR. Asterisks represent significantly different values in all pairwise comparisons, two-tailed t-test.  $n = 4$ . Measurements of influenza transcription shown in S11 Fig. (E) Different PB1 deletions package to similar ratios as wild-type PB1. Ratio of PB1 to HA, normalized to wild-type, in viral stocks as measured by qPCR. Dotted line represents average wild-type ratio. No values significantly diverged from wild-type, two-tailed t-test,  $p < 0.05$ .  $n = 3$ . For panels (B), (C), (D), and (E) individual points are individual measurements, lines are mean values. For panels (B), (C), and (D) significance was calculated using Benjamini-Hochberg correction at an FDR of 0.05.

<https://doi.org/10.1371/journal.ppat.1010125.g004>

compare the fraction of deletion-containing fragments between interferon-enriched and depleted datasets, there is statistically significant enrichment across most segments (Fig 4B). Mapping depth across the polymerase segments is consistent with this enrichment, showing a slight, but replicable, increase in depth at the 5' and 3' ends of the three polymerase segments relative to the central portion of each gene segment (S10 Fig). Other deletions, while enriched, were too rare to produce visible changes in sequencing depth. To confirm these results were not driven by PCR bias, we developed a qPCR sensitive to increased defective content (S11 Fig). Our design leverages the short extension time common in most qPCR protocols, and so, by designing primers flanking the shared limits of deletion breakpoints between biological replicates we generate a signal that is responsive to the presence of shortened products while producing little to no signal on full-length polymerase segments. We can then correct this gestalt

“DI-specific” signal to “full-length” signal generated from qPCR primers that lie nested within most observed deletions. [50] Using this qPCR, we confirm our deep-sequencing results (Fig 4C).

As deletions preferentially associate with interferon induction, we may ask whether all deletions exhibit an identical relationship to this process. That is, is the mere presence of a deletion sufficient to induce high levels of interferon, or do different deletions exhibit differential induction of interferon pathways? Using complementing cells, we grew a previously-cloned stimulatory PB1 deletion containing 177nt from the 5' end of the vRNA, and 385nt from the 3' end of the vRNA, PB1<sub>177:385</sub>, and a newly-identified deletion from our dataset, PB1<sub>455:350</sub>. [46] Comparing the two, we find that the former retains its capacity in infections to induce high levels of type I interferon transcripts, whereas the latter is no more stimulatory than a nominally wild-type population (Fig 4D). This difference appears to be linked to the identity of the deletion itself, and not due to differences in packaging rates of these two segments. Prior work on influenza A viruses suggests that even upon substantial manipulation, these viruses maintain a canonical 1+7 configuration in viral genomic packaging, and likely package only a single copy of each genomic segment. [51–53] Suggesting that this is largely true for these defective viruses, and demonstrating equivalent dose not just of biologically active particles, but even of defective segments themselves, we find that the ratio of the HA genomic segment to the PB1 genomic segment in viral preparations is unchanged between wild-type and either PB1<sub>177:385</sub> or PB1<sub>455:350</sub> (Fig 4E).

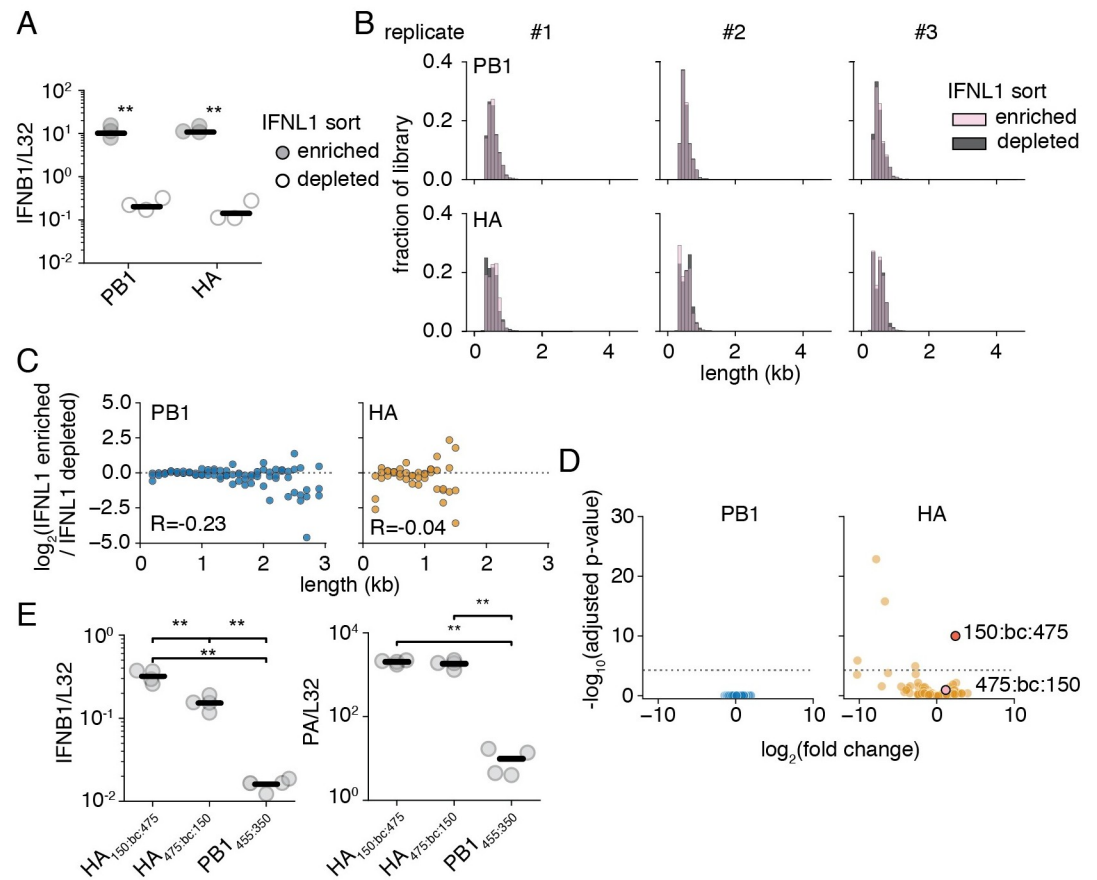
### Segment length and interferon induction do not appear to be correlated

As different deletions can cause divergent levels of interferon stimulation, and as we may sort by interferon production to classify immunostimulatory variation, we may use our sortable cell lines and our libraries to generate comprehensive measurements of how length of a defective segment might impact interferon induction. The increase in interferon observed in infections with defective influenza species has been posited to arise from a preference for shorter molecular species by RIG-I, the major sensor of this virus. [54, 55] To first establish whether, as expected, RIG-I is detecting the stimulatory defective species that we are studying, we introduced either of two siRNAs targeting RIG-I to A549 cells and infected with PB1<sub>177:385</sub>. As expected, either RIG-I-targeting siRNA reduced not only levels of RIG-I, but also the production of interferon transcripts in response to PB1<sub>177:385</sub>, consistent with its general role as the predominant sensor of this virus in non-immune cells (S13 Fig). To take things a step further, we wished to generally demonstrate whether, as expected, smaller defective species tend to evoke a stronger innate immune response. Therefore, we infected a type III interferon reporter A549 cell line with our libraries and sorted on interferon production.

We changed to a cell line expressing LNGFR and eGFP under control of a type III interferon promoter over a type I reporter as the sort described in Fig 4 only produced a yield of ~5000 cells per sort from a starting population of 20 million cells. Notably, length of a defective segment, if it does influence interferon induction, would only be one variable within our experiment doing so. Timing of infection, autocrine signaling, paracrine signaling, and likely other features, including those intrinsic to the cells we are infecting, will influence the probability a cell produces interferon. Therefore we must sample a sufficient number of cells to average out such behavior over our whole population in order to measure how length of a segment may, or may not, influence innate immune detection. As the theoretical yield based on the fraction of interferon-positive cells ranges from ~100,000–200,000 in our previous experiment, our interferon-beta reporter sorting resulted in a greater than 95% loss of material, and would produce far too low a yield to analyze a comprehensive set of variants.

Based on our mRNAseq data, and the prior observation that a type III interferon reporter line produces a stronger signal that still correlates with type I interferon production, we opted to change strategies. [46] Consistent with an improved resolution, we were able to recover ~100,000 cells in each of our interferon-enriched sorts, and, by qPCR, our sorting efficiency as measured by type I interferon expression was comparable to that achieved with a type I interferon reporter (Fig 5A and S14 Fig).

Comparing length distributions of our library in interferon-enriched versus interferon-depleted populations, we notice no obvious shifts as we did in our other selections (Fig 5B). To explore in greater depth, we looked for evidence for enrichment with interferon as a function of length to look for more subtle shifts that may drive interferon expression (Fig 5C). We see



**Fig 5. Variant length and interferon induction appear uncorrelated.** (A) An A549 type III interferon reporter line was infected at an MOI of 0.05 and sorted using magnetic-activated cell sorting 14 hours post infection. qPCR against the type I interferon, *IFNB1*, as corrected by the housekeeping gene *L32*. Asterisks indicate significantly increased *IFNB1* transcript, two-tailed t-test. n = 3. qPCR of *IFNB1* sort performed in Fig 4A for comparison in S14 Fig (B) Size distributions of libraries from the sorting in (A). (C) Interferon induction and length are not correlated. Each point represents a 100nt bin. The fraction of variants falling within each bin was compared between interferon-enriched, and -depleted selections. Points above the dotted line are co-enriched with interferon. Points were only shown if represented in all three libraries under both conditions. R is the Spearman correlation coefficient. n = 3. (D) Enrichment or depletion of individual variants within libraries. Each individual point represents a distinct 5' and 3' junction with <100 average observations. An adjusted p-value threshold of 0.05 is shown by the dotted line. Variants chosen for further analysis highlighted in red and pink. Full data of differentially enriched variants found in S3 File. (E) Variants of identical length can exhibit different levels of interferon induction. (left) qPCR measuring levels of *IFNB1* normalized by *L32* in A549 cells infected at an MOI of 0.5 at 14 hours post-infection. (right) qPCR measuring levels of *PA* to show similar levels of infection between HA variants. Asterisks represent significantly different values in all pairwise comparisons, two-tailed t-test. n = 4. For panels (A), and (D), individual points are individual measurements, lines are mean values. For panels (A), and (D) significance was calculated using Benjamini-Hochberg correction at an FDR of 0.05.

<https://doi.org/10.1371/journal.ppat.1010125.g005>

no clear relationship between size and interferon state, with no strong enrichment at any size, at Spearman correlation coefficients of -0.23 and -0.04, for PB1 and HA, respectively. Overall, we fail to find significant evidence that length is a determinant of interferon production.

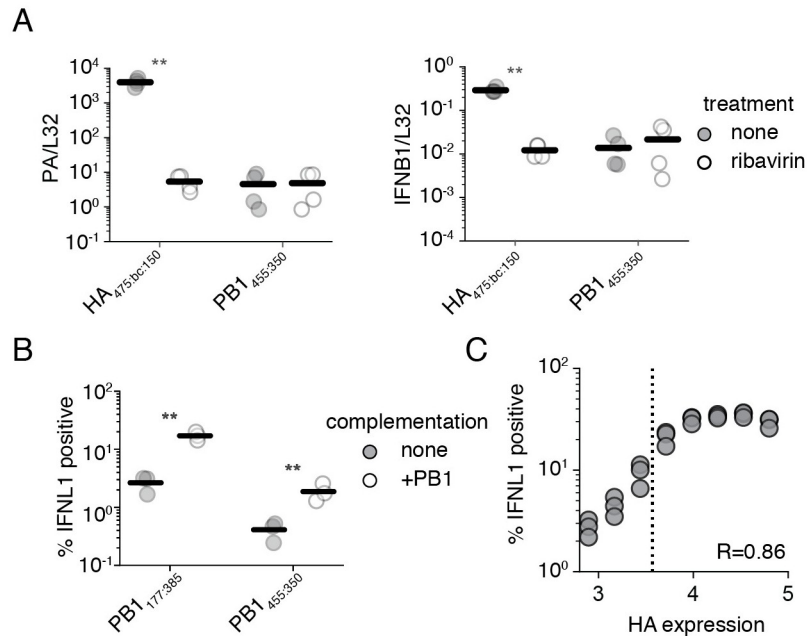
To probe whether we captured any variation in interferon induction within our libraries, we explored how individual junctions rather than binned sizes behaved in our interferon enrichment sequencing (Fig 5D). Using DESeq2 to analyze our data, there was little detectable variation between interferon-enriched and -depleted datasets in our PB1 library. [56] However, we find slightly more variation within our HA library, and identify a single variant with statistically significant enrichment in our interferon-producing cells. This variant consists of a junction lying 150nt from the 5' end of the HA vRNA, then our 57nt adapter, and lastly a downstream junction lying 475nt from the 3' end of the HA vRNA, for a total length of 682nt (HA<sub>150:bc:475</sub>). Curiously, the mirror-image deletion, HA<sub>475:bc:150</sub> exhibits mild, not statistically significant, enrichment.

This finding provided us with an opportunity to demonstrate that length is not the sole contributor to whether any given deletion is stimulatory or not. We cloned both HA<sub>150:bc:475</sub> and HA<sub>475:bc:150</sub>, and grew each independently using complementing cells. Thereafter, we tested the capacity of viral populations bearing these mutations to stimulate interferon (Fig 5E). As predicted by our more comprehensive analysis, HA<sub>150:bc:475</sub> induces approximately 2-fold higher transcription of *IFNB1* than HA<sub>475:bc:150</sub>. As our methodology had the capacity to detect this two-fold difference, it suggests that length plays no more than a modest role, at most, in contributing to defective variation in interferon induction. While a surprise, this conclusion is highly consistent with recent experiments in 293t cells expressing viral replication machinery, which found that while length of an influenza segment could influence innate immune detection, the size at which this occurs, <125nt, is far smaller than defective influenza genomes, indicating that observed stimulation by defective viral particles likely arises from some other feature of deletions such as perturbations to RNA secondary structure. [57]

### Interferon induction by defective segments is sensitive to genome replication

If our differentially-stimulatory variants exhibit different affinities for RIG-I or engage in aberrant behavior that indirectly impacts RIG-I signaling, then we might expect that these behaviors would be exacerbated by genome replication. While deletions in the polymerase segments produce defective species that, on their own, cannot engage in genome replication due to a requirement for *de novo* polymerase synthesis, deletions in non-polymerase segments would not be subject to such restrictions. [58] Therefore, we must ask, is the immune stimulation in our HA deletions contingent upon their ability to engage in genome replication? Interestingly, both HA<sub>150:bc:475</sub> and HA<sub>475:bc:150</sub> exhibit induction of *IFNB1* above that of our less stimulatory PB1 variant, PB1<sub>445:bc:350</sub>, suggesting that both act as efficient triggers of innate immunity, even if their potency differs.

To determine whether genome replication may play a role in this process, we used a suppressor of this process, the nucleoside analog, ribavirin. [59, 60] While prior studies have used the translation inhibitor cycloheximide, cycloheximide enhances interferon signaling to even pure innate immune agonists, thus complicating conclusions using this inhibitor. [61–63] To confirm that we are investigating the effects of genome replication alone, and not pleiotropic impacts on cellular pathways, we compare the impacts of ribavirin treatment concurrently on infections with our less stimulatory PB1 deletion, PB1<sub>445:bc:350</sub>, which is unable to engage in genome replication, and infections with HA<sub>475:bc:150</sub>. Upon ribavirin treatment, HA<sub>475:bc:150</sub> exhibits a massive drop in viral transcription of the PA gene, consistent with reduced template



**Fig 6. Genome replication is a key component of interferon induction.** (A) (left) qPCR against *PA*, as corrected by the housekeeping gene *L32*, was used to validate the suppression of viral replication by the antiviral nucleoside ribavirin. (right) qPCR against *IFNB1* as corrected by the housekeeping gene *L32* demonstrates the suppression of interferon induction by a deletion capable of genome replication. A549 cells were pretreated with 200  $\mu$ M ribavirin for 2h and then infected at an MOI of 0.5. RNA was harvested for analysis after 14 hours of infection. Astrisks indicate strains wherein signal was significantly impacted by treatment. Two-tailed t-test.  $n = 4$ . Similar data with neuraminidase shown in S15 Fig. (B) Complementation of PB1 defective species by PB1 expression *in trans* increases interferon induction. The fraction of cells in a population expressing a type III interferon reporter when infected with the indicated influenza variant at an MOI of 0.5 for 13h. Gating schema shown in S16 Fig. (C) The increase in interferon-induction observed in (B) is linked to influenza replication. Complemented cells were infected as in (B) with an MOI of 0.5 of PB1<sub>177:385</sub>, stained for the viral protein HA and analyzed after 13h. Points represent binned values of HA expression, each point represents the right-most edge of a bin. The dotted line is the 99<sup>th</sup> percentile of PB1<sub>177:385</sub> HA staining without complementation. R represents the Spearman correlation coefficient.  $n = 3$ . Gating schema shown in S17 Fig. For panels (A), and (B), individual points are individual measurements, lines are mean values. For panels (A), and (B) significance was calculated using Benjamini-Hochberg correction at an FDR of 0.05.

<https://doi.org/10.1371/journal.ppat.1010125.g006>

vRNA molecules for mRNA synthesis (Fig 6A). PB1<sub>445:350</sub> exhibits no drop as any mRNA transcribed in this virus must be from the initiating viral genome. With matched transcription levels between our deletions in HA and PB1, we now see a decrease in interferon induction for HA<sub>475:bc:150</sub> but none for PB1<sub>445:350</sub>, demonstrating that genome replication is a key determinant of interferon induction even for defective influenza species.

Wishing to demonstrate that replication and identity are features more critical than size in interferon induction, and that this effect is likely not restricted to HA alone, we cloned an incredibly small defective species in the neuraminidase segment (NA), NA<sub>189:76</sub>, only 265nt in size. As complementation of neuraminidase is difficult to do *in trans*, we replaced the coding sequence for hemagglutinin with neuraminidase, and grew this variant on cells which complement for hemagglutinin, similar to a previous construction of an eGFP pseudovirus. [64] We then compared the capacity of this variant to stimulate interferon with, and without, ribavirin to our stimulatory PB1 variant, PB1<sub>177:385</sub> (S15 Fig). In doing so we find that NA<sub>189:76</sub> matches, but does not exceed, stimulation by PB1<sub>177:385</sub> in the absence of ribavirin. Moreover, when we add ribavirin, as with our HA variants, we see a suppression of interferon induction by

NA<sub>189:76</sub>, but not PB1<sub>177:385</sub>, indicating that when genome replication is restricted, even though NA<sub>189:76</sub> is half the size of PB1<sub>177:385</sub>, it induces far less of an innate immune response.

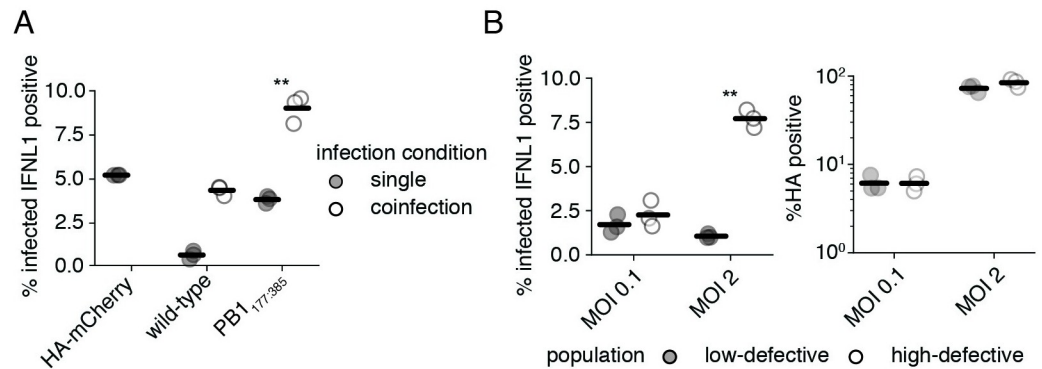
This leads to an interesting question; would PB1<sub>445:350</sub>, noted to have approximately the same capacity to stimulate interferon as wild-type influenza, actually be potently stimulatory if it was capable of replicating its genome? To answer this question, we integrated a PB1 expression cassette into our type III interferon reporter line. This allowed us to measure the effects of complementation on interferon induction by flow cytometry, permitting a more refined measurement delineating whether complementation not only drives increased interferon induction, but also whether it fundamentally does so by increasing the number of cells that successfully detect influenza.

As expected from our model derived from ribavirin treatment, complementation of either PB1 deletion, PB1<sub>445:350</sub> or PB1<sub>177:385</sub>, with PB1 expression *in trans* greatly increases the fraction of cells participating in the production of interferon, although each variant retains its relative level of stimulation to one-another (Fig 6B). This supports our conclusions that relative interferon induction is not only contingent on the amount, but also the sequence, of the aberrant segment. Combined with our prior observations, we can see that even those deletions that drive modest interferon induction, such as PB1<sub>445:350</sub>, could be quite dangerous to the virus were they to occur in non-polymerase segments.

Following from these data, we wanted to conclusively demonstrate that enhanced replication drives a higher probability an infected cell detects viral infection, and that our complementation experiments weren't simply leading to additional rounds of infection. Using the highly stimulatory PB1<sub>177:385</sub>, we repeated the infection of complemented interferon reporter cells and concurrently stained for the viral protein HA. As transcription and replication are fundamentally linked in influenza viruses, increased protein production is highly correlated with genome replication. [65, 66] Therefore, we may use HA staining as a rough proxy for the level of replication within an infected cell. In doing so, we see a positive relationship between the level of HA staining of infected cells and the probability that those cells produce interferon, with a Spearman coefficient of 0.89. This demonstrates that our results were not merely driven by an increased fraction of infected cells, but rather by an increased probability of response in cells with active viral replication (Fig 6C). Indeed, at higher levels of HA staining we find that PB1<sub>177:385</sub> can induce interferon in as many as 35% of infected cells, rivaling prior measurements of interferon expression in viral infections lacking the critical interferon antagonist NS1. [45]

### Coinfection increases interferon induction by defective particles

Seeing that complementation appears to impact innate immune detection of defective polymerase species, we wondered whether this observation could be extended to coinfections, wherein intact polymerase genes are delivered by other viral particles, rather than an integrated expression cassette. In a coinfection we would expect some level of competition between our defective segment and its full-length counterpart, potentially reducing levels of RIG-I ligand. To explore whether coinfection reproduces data consistent with our complementation experiments, we turned to a pre-existing pseudovirus system in which the coding sequence of influenza HA has been replaced with mCherry. By infecting reporter cells with this pseudovirus alongside either wild-type influenza, or PB1<sub>177:385</sub>, we can identify coinfections as those that stain positive for both HA and mCherry. [64, 67] Curiously, our pseudovirus alone appears to induce much higher levels of interferon than wild-type influenza, potentially due to the presence of non-native sequence. Regardless, when present in a coinfection with PB1<sub>177:385</sub> a greater fraction of cells express interferon than when infected with either



**Fig 7. Coinfection increases interferon induction by defective particles.** (A) Coinfection with a complementing virus increases interferon induction by PB1<sub>177:385</sub>. Type III interferon reporter A549 cells were infected with PB1<sub>177:385</sub> or wild-type influenza at an MOI of 0.1, and HA<sub>mCherry</sub> pseudovirus at an MOI of 0.5. Interferon expression was measured in HA-stained cells at 13h post infection. Cells from single infections were called as infected if they stained positive for HA or expressed mCherry. Coinfections required co-staining for both HA and mCherry. Asterisks indicate conditions under which coinfection led to a significant increase in interferon from both HA<sub>mCherry</sub> and the indicated variant alone, one-tailed t-test. Gating schema shown in S18 Fig for HA and mCherry, and S19 Fig for interferon. (B) Populations replete in defective particles exhibit much higher interferon induction under high MOI infections where coinfection is common. (left) The fraction of HA-positive cells expressing interferon. (right) The total fraction of HA-positive cells. Type III interferon reporter A549 cells were infected with a low-defective or high-defective population as described in S6 Fig. Cells were then infected at a genome-normalized MOI as calculated from the low-defective population and stained for HA and analyzed for interferon reporter expression after 13h. Gating schema shown in S20 Fig. For panels (A), and (B), individual points are individual measurements, lines are mean values. For panels (A), and (B) significance was calculated using Benjamini-Hochberg correction at an FDR of 0.05.

<https://doi.org/10.1371/journal.ppat.1010125.g007>

virus alone (Fig 7A). A similar effect is not seen for wild-type influenza, demonstrating the specificity of this phenomenon to defective influenza species. Thus complementation during coinfection can drive higher levels of interferon induction by deletions in polymerase segments.

Following the logic that complementation from co-infecting viruses can lead to increased detection of defective influenza particles, we might expect that deletions in the polymerase segments drive a proportionately greater innate immune response at higher infectious doses where coinfection is more common. To see if this assumption is true we infected our interferon reporter line at an MOI of either 0.1 or 2 of a low-defective population, and equivalent genome copies of a high-defective population, and measured the fraction of HA-positive cells that were expressing interferon lambda (Fig 7B). With a low-defective particle population, the average rate at which infected cells produce interferon remains largely unchanged between these two MOIs. However, with a high defective population, there is a massive increase in the fraction of interferon-producing cells at a higher MOI. This was not due to a difference in infection rate, as the fraction of HA-expressing cells in each condition between the low and the high defective content viral populations was similar.

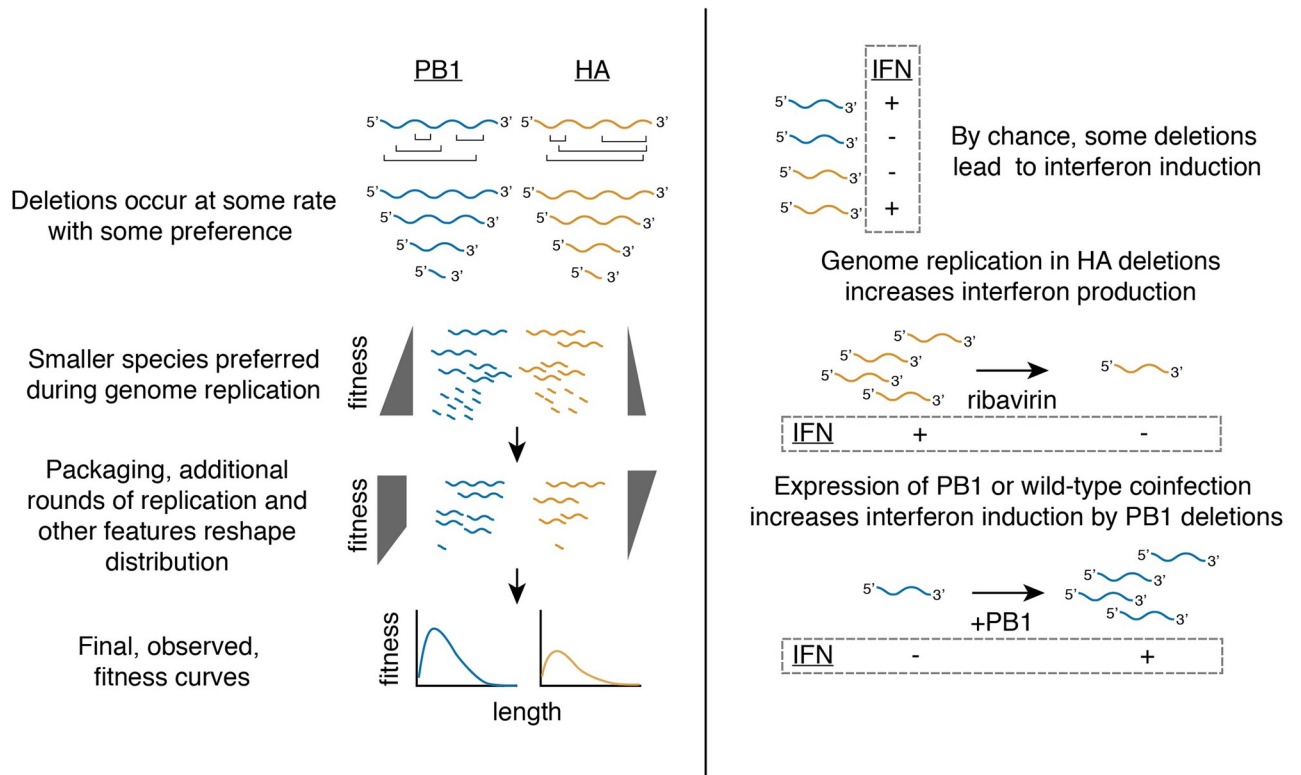
This suggests that interferon induction by the vast majority of defective species is a nonlinear process. Their relative importance is dependent upon not only their frequency, but also the rate of complementation as driven by coinfection by viruses bearing full-length segments. Critically, these processes modulate interferon induction, but genome replication is not absolutely required for detection of defective species as demonstrated by stimulation from PB1<sub>177:385</sub> alone. This nicely explains why some groups have concluded that genome replication is wholly dispensible for interferon induction, others have concluded it is modulatory, and yet still more have concluded it is absolutely essential. [43, 54, 68–70] Each group was undoubtedly correct



for their experimental setup, but the contribution of genome replication to the observed induction of interferons is highly sensitive to the rate of coinfection and the fraction of the population which is defective, and thus different results may be derived even from the same underlying process.

**Discussion**

We have used a library-based approach to comprehensively explore the relationship between length, replication, competition for participation in the viral population, and interferon induction in influenza A virus for the polymerase segment, PB1, and the non-polymerase segment, HA. Additionally, we describe how genome replication can contribute to interferon induction by defective species, and demonstrate that both the population composition and the rate of coinfection influence the probability with which these viruses are detected by cell intrinsic innate immunity (Fig 8). Interestingly, during replication our artificial libraries converge on a near-identical length as natural defective populations. [23]. This is the result of positive selection for small segments, particularly in PB1, during genome replication, negative selection during packaging, and, likely, other features that remain to be explored. This suggests that the contrasting pressures during viral growth are sufficient to shape defective influenza populations to their final, observed, lengths, from potentially disparate initial distributions.



**Fig 8. Summary of major conclusions.** After deletions form, they are subject to selection during genome replication, which favors smaller products. As viral populations expand, different selection pressures reshape viral populations. In part, selection against small variants during packaging, observed for both PB1 and HA, contributes to this reshaping. During viral growth and expansion, very small (<400nt) PB1 variants are excluded, whereas the HA segment exhibits length dependence throughout, favoring larger products. Combining pressures produces a curve with a deletion fitness maximum of ~ 400 nt, but with a lower overall preference for HA deletions than PB1 deletions. Interferon induction by deletions appears to be largely length-independent. Genome replication modulates inteferon induction by deletions, either occurring in single infections for non-polymerase deletions, or during coinfection for polymerase deletions.

<https://doi.org/10.1371/journal.ppat.1010125.g008>

The advantage to small segments during genome replication could be due to several mechanisms. One possibility is that during genome expansion, after carefully regulated steps of conversion of negative-sense vRNA to positive-sense cRNA and subsequent massive expansion of the viral vRNA pool, that re-initiation of replication upon a template must wait for completion of current rounds of replication. Therefore, smaller segments may simply be able to initiate subsequent rounds of replication more rapidly. Another possibility may be found in studies of nucleoprotein. Nucleoprotein (NP) is required both as a processivity factor in polymerase extension and to stabilize nascent cRNA antigenomes. Smaller RNA species do not require NP, and in its absence mvRNA <125nt may still be readily produced [57] While not quite as small, defective genome segments appear to require less NP, potentially allowing for more rapid progression through the viral life cycle. [71]

The disadvantage to small segments during packaging is wholly consistent with recent RNA-RNA interaction experiments performed on intact influenza particles. [72, 73] It has now been demonstrated that while the minimal packaging requirements are restricted to the 5' and 3' termini of influenza genomic segments, intersegmental contacts occur throughout and can dramatically impact packaging efficiency. [29, 30]. Therefore the loss of many low-affinity contacts that contribute to packaging may serve to exclude defective species from the population.

The combination of genome replication, packaging, and other features we have yet to explore produces a highly replicable maximum fitness, even between PB1 and HA, at 400nt, not smaller, nor larger, with a preference for PB1 deletions over HA. Given recent interest in highly-fit defective particles as potential therapeutics, our results demonstrate that the consideration of multiple, contrasting, selection pressures is required to rationally design defective genomes that may compete and propagate throughout an otherwise replication-competent population.

In contrast to the other features we explored, we were surprised to find that interferon induction by influenza A virus defective particles is, at least largely, length-independent. This finding is consistent with similar work on *Paramyxoviridae*, which generate short, copy-back defective genomes. Work on these viruses found that it was not the length of these genomes, nor their double-strandedness, that leads to their detection, but rather changes in their RNA secondary structure. [74] It seems likely that deletions in influenza likewise perturb otherwise non-stimulatory sequence composition, and, by chance, these perturbations can generate potent RIG-I ligands. How can we explain our results in light of the finding that small, likely defective, genomes associate with RIG-I? [55] With the following logic we may show that this apparent contradiction may be explained: 1) perturbation in natural sequence may generate potent ligands for RIG-I, 2) most deviations from wild-type sequence, particularly those that would drive interferon induction, would be purged from the population, and 3) largely due to advantages during genome replication, defective particles, despite their defects, can nevertheless rise to high levels in the viral population. Therefore, sequence differences that generate potent RIG-I ligands would generally be expected to be found in defective genomes, simply as they represent a rare instance where a largely deleterious mutation (at least in terms of the population) still rises to high frequency. This does not preclude the role of mvRNAs <125nt in interferon induction, and these incredibly small RNA species likely contribute alongside defective genomes to the observed interferon response. [57, 71]

Lastly, we find that replication contributes to, but is not necessary for, the innate immune response to defective influenza particles. This is consistent with the idea that deletions, by chance, can generate more potent RIG-I ligands. RIG-I is known to exhibit a dose-responsive activation, and therefore for any given ligand, an increased concentration of that ligand would be expected to increase the probability of RIG-I engagement and interferon production. [75]

This is in contrast to some reports which found that genome replication fails to modulate innate immune detection of these populations, however experimental choices such as the use of the inhibitor cycloheximide, or high MOIs wherein replication may not meaningfully contribute additional RIG-I ligands, may explain these discrepancies. [43, 70]

The finding that genome replication modulates interferon induction in turn leads us to reconsider the observed distribution of defective species across the viral genome, largely consisting of deletions in the polymerase segments, in a new light. Under conditions early in infection, where coinfection may be more limited, occasional packaging of deletions in the polymerase segment may be tolerated as the likelihood these deletions induce interferon remains low. However, should the virus package non-polymerase deletions, these viruses would exhibit an increased probability of interferon induction even under conditions where coinfection is rare. We therefore postulate that, rather than defective particles preferentially consisting of deletions in the polymerase segments due to some selection for their presence, we should also consider that there may be a strong selection against inclusion of deletions in non-polymerase segments.

Overall our study addresses several critical gaps in our understanding of influenza A virus defective particles. Our library-based approach was quite successful at generating models of how defective influenza populations are shaped after their initial formation, demonstrating that pressures other than generation are sufficient to explain their observed distribution. While there still remain features left to be explored to procure a full explanation of the kinetic features underlying their propagation, this is perhaps unsurprising. Prior experiments tracking defective content in long-term culture showed oscillatory dynamics, consistent with the behavior of parasitic elements in many social systems, implying frequency-dependent negative selection in addition to the features discussed here. [76] Therefore continued work to understand these particles will need to couple substantial modeling with tractable experimental systems, such as the libraries we generated in this report. Although we were only able to make a few inferences regarding the precise nature of deletions that drive interferon induction, namely that this phenomenon is length-independent and impacted by genome replication, our study was not designed to probe sequence-dependent phenomena as well as length-dependent owing to the introduction of barcodes and other artificial sequences which themselves likely impacted RNA structure. Future work with a modified library design will be required for further exploration of those features with an ultimate goal of being able to perform *in silico* prediction of stimulatory behavior. Finally, beyond this work on influenza A virus, we note that our library design is also broadly extensible to explore any length-dependent phenomena wherein the lengths are compatible with PCR amplification, and hope that providing this method will serve as a resource to the broader community.

## Materials and methods

### Cell lines and viruses

The following cell lines were used in this study: HEK293T (ATCC CRL-3216), MDCK-SIAT1 (variant of the Madin Darby canine kidney cell line overexpressing SIAT1 (Sigma-Aldrich 05071502)) and A549 (human lung epithelial carcinoma cell line, ATCC CCL-185). Variants of MDCK-SIAT1 overexpressing influenza HA and PB1 were generated using a lentiviral vector as previously described. [77–79] A549 type I and type III interferon reporter lines were previously described. [46] A549 cells overexpressing PB1 were generated using lentiviral transduction using the same vector for MDCK-SIAT1. Cell lines were tested for mycoplasma using the LookOut Mycoplasma PCR Detection Kit (Sigma-Aldrich) using JumpStart Taq DNA Polymerase (Sigma-Aldrich). A549 cells used in all sequencing experiments had their

identity confirmed via STR profiling by ATCC. All cell lines were maintained in D10 media (DMEM supplemented with 10% heat-inactivated fetal bovine serum and 2 mM L-Glutamine) in a 37°C incubator with 5% CO<sub>2</sub>.

Wild-type A/WSN/1933 (H1N1) virus was created by reverse genetics using plasmids pHW181-PB2, pHW182-PB1, pHW183-PA, pHW184-HA, pHW185-NP, pHW186-NA, pHW187-M, pHW188-NS. [80] Genomic sequence of this virus provided in [S4 File](#). HEK293T to MDCK-SIAT1 cells were seeded in an 8:1 coculture and transfected using BioT (Bioland Scientific, LLC) 24 hours later with equimolar reverse genetics plasmids. 24 hours post transfection, D10 media was changed to Influenza Growth Medium (IGM, Opti-MEM supplemented with 0.04% bovine serum albumin fraction V, 100 µg/ml of CaCl<sub>2</sub>, and 0.01% heat-inactivated fetal bovine serum). 48 hours post-transfection, viral supernatant was collected, centrifuged at 300g for 4 minutes to remove cellular debris, and aliquoted into cryovials to be stored at -80°C. Thawed aliquots were titered by TCID<sub>50</sub> on MDCK-SIAT1 cells and calculated using the Reed and Muench formula or titered by qPCR against the viral segment HA. [81] To create wild-type viral stocks with a low defective content, MDCK-SIAT1 cells were infected at an MOI of 0.01 and harvested 30 hours post-infection. To generate intermediate defective populations, MDCK-SIAT1 cells were serially infected at an MOI of 0.01 twice, harvested at 72h each passage. To generate high-defective populations, these passages were instead performed at an MOI of 1.

Indicated mutations were introduced by inverse-PCR followed by single-molecule Gibson ligation and confirmed by Sanger sequencing. To generate viruses bearing large deletions in PB1 or HA, HEK293T cells were co-transfected with all reverse-genetics plasmids encoding A/WSN/1933 except for either HA, or PB1, the indicated variant, and a construct expressing either HA or PB1 under control of a constitutive promoter in the pHAGE vector as used to transduce MDCK-SIAT1 cells. [79] To generate the viruses bearing a large deletions in NA, HEK293T cells were co-transfected with all reverse-genetics plasmids encoding A/WSN/1933 except for HA and NA, the indicated variant, a copy of the coding sequence of NA flanked by the packaging sequence of HA, and a construct expressing either HA under control of the a constitutive promoter in the pHAGE vector as used to transduce MDCK-SIAT1 cells. MDCK-SIAT1 cells in co-culture either constitutively expressed HA or PB1. Viruses were titered on complementing MDCK-SIAT1 cells, or by qPCR against the viral segment PA. For the barcoded HA deletions, HA<sub>150:bc:475</sub> and HA<sub>475:bc:150</sub>, 5' AGCTCCTATTAG 3', and 5' TATAGTCCCAAT 3', were used instead of random sequence.

## Library cloning and initial sequencing

Up (amplifying the 5' end of the gene) and Down primer sets were used in individual PCR reactions alongside the common primers 5' GGTAAGTGCAGACCAAGTTTACTC 3' (Up) or 5' GAGTAAACTTGGTCTGACAGTTACC 3' (Down) using Q5 Hot Start High-Fidelity 2x Master Mix (New England Biolabs, M0494). Primers were designed as described in [https://github.com/a5russell/Defective\\_Library\\_Mendes\\_Russell](https://github.com/a5russell/Defective_Library_Mendes_Russell), using annealing temperature calculations as described by Breslauer *et al.* (1986) [82]. Each reaction was performed using 2.5 ng of template DNA consisting of a pHH21 vector encoding either HA or PB1 with the ATG start codon mutated to GTA. [39] Reaction conditions consisted of a 55°C annealing temperature, a 2 minute extension, and 11 total cycles of amplification. Up and Down amplicons were independently pooled and subjected to DpnI digest at 37°C for 1 hour. Products were then purified using SPRI beads from the ProNex Size-Selective Purification System (Promega, NG2001) at a 1x bead volume following manufacturer's instructions.

Down amplicons were then subjected to a second round of PCR using the same common PCR primer and an additional primer consisting of an artificial adapter containing a 12nt random nucleotide sequence, 5' GCCATCGCCTACACGACGCTTCNNNNNNNNNNNNNNNNNNNN CGTTGACCACTGCTTGCGATGAT 3'. This primer was acquired from integrated DNA Technologies, with a hand-mixed randomer. Conditions for this amplification consisted of 25ng of template DNA subjected to 8 cycles of amplification at a 50°C annealing temperature and 2 minute extension. This amplicon was then purified once more using a 1x bead volume.

25 ng of each amplicon pool, Up and Down with adapter, were combined in a Gibson reaction using NEBuilder HiFi DNA Assembly Master Mix (New England Biolabs, E2621) using manufacturer's protocols save for a 1 hour incubation step rather than 15 minutes. The resultant reaction was then cleaned using 1x bead volumes, and electroporated into electrocompetent Stbl3 cells. Plates were counted, scraped, and plasmids prepared for each step. This whole process, from PCR onwards, was performed 3 independent times for each library.

To sequence junction regions associated with each barcode, 1 $\mu$ g of each library was digested with BtgZI overnight and released products were purified by gel electrophoresis followed by gel extraction. Bands were then subjected to ligation with a DNA splint containing an NNNN region to facilitate ligation to the released overhang. This splint was generated by annealing 5' AAGCAGTGGTATCAACGCAGAGTACAT 3' to 5' NNNNATGTACTCTGCGTTGATACACTGCTT 3'. Ligation to each splint was performed overnight at 16°C using T4 DNA ligase (New England Biolabs).

Ligation reactions were used directly as templates for PCR to infer Up junctions and Down junctions, separately, using the common primer 5' CGGGGTACGATGAGACACCATATTGGTCTCAAGCAGTGGTATCAACGCAG 3', and 5' GAAGCAGAAGACGGCATAACGATAGAGCCATCGCCTACACGACGCT 3' (Up) or 5' GAAGCAGAAGACGGCATAACGATAGATCATCGCAAGCAGTGGTCA 3' (Down) with the following reaction conditions: 60°C annealing temperature, 15s extension, for 12 cycles. A second round of PCR was then used to add partial Illumina adapters (5' TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCGGGTACGATGAGACACCA 3' / 5' GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGAAGCAGAAGACGGCATAACGAGAT 3', with a 55°C annealing temperature, 15s extension, and 6 cycles. Samples were then barcoded using Illumina UD indices with a 62°C annealing temperature, 20s extension, and 10 cycles. HA and PB1 libraries were pooled in indexing as they could be disambiguated during sequencing. Final samples were bead cleaned with a 3x bead volume and sizes confirmed by tapestation prior to sequencing.

### Barcoded library selections

To analyze rates of genome replication, in an 8:1 coculture of HEK293T and MDCK-SIAT1 cells, each individual deletion/duplication library was cotransfected with plasmids that individually encode the mRNA for PB2, PB1, PA, and NP in an equimolar ratio using BioT according to the manufacturer's protocol. An orthogonal dataset was generated with transfections consisting of a library, PB1, PA, and NP but not PB2. Transfections were then allowed to proceed for 24 hours and RNA was purified using the Monarch Total RNA Miniprep Kit from New England Biolabs (T2010). To analyze other steps in the viral life cycle, after 24 hours media was changed from D10 to IGM and cells were infected with a low-defective wild-type WSN population at an MOI of 0.25, and infection was allowed to progress for 72h. Virus-containing supernatant was clarified by centrifugation at 300g for 4 minutes, and RNA was purified from 100  $\mu$ l of supernatant using the Monarch Total RNA Miniprep Kit with 600  $\mu$ l of lysis buffer.

To associate barcodes with the induction of interferon, 21 million interferon-lambda reporter cells were seeded across 3 15cm plates. 24 hours later, the media was changed to IGM

and cells were infected with library supernatant at an MOI of 0.05. 14 hours post-infection, cells were separated using the MACSelect LNGFR System (Miltenyi Biotec). The supernatant of the samples were aspirated before being trypsinized for 10 minutes at 37°C and quenched with D10 media. Samples were centrifuged for 4 minutes at 300g, had their supernatant removed, and were resuspended in 320  $\mu$ L of 1x phosphate buffered saline (PBS). 80  $\mu$ L of MACSelect LNGFR MicroBeads were mixed into the samples thoroughly and then placed in a 4°C fridge for 15 minutes. Samples were then applied to LS columns (Miltenyi Biotec, 130–042–401) according to manufacturer's instructions. In addition to our enriched population, five million interferon-depleted cells from each sample were retained for analysis. RNA was thereafter purified from both sets of samples using the Monarch Total RNA Miniprep Kit from New England Biolabs.

To grow mixed 400nt and 1600nt variant populations, transfection was performed as for our more complex libraries, and thereafter populations were permitted to expand as described above over 72 hours.

### Barcoded library sequencing

All RNA samples (minigenome +PB2, minigenome -PB2, viral supernatant, and interferon-enriched and -depleted sorts) were converted to cDNA using the universal vRNA primers of Hoffmann *et al.* (2001) [47] and the SuperScript III First-Strand Synthesis SuperMix kit (Invitrogen, 18080400) according to the manufacturer's protocol. Before converting the interferon sorted RNA samples to cDNA, a qPCR against the HA segment or PB1 segment was performed to equilibrate the number of RNA molecules to be reverse transcribed between each enriched and depleted biological sample. Primers for this qPCR were, for PB1, 5' CCGTC TGAGCTCTTCAATGG 3'/5' GATGCACGAATTGATTCGAATCTG 3', and for HA, 5' GCAAACACTACTGGTCCTGTTATATGC 3'/5' GTGTCAACAGTGTCCGGTTGAG 3'.

cDNA samples were amplified and partial i5 and i7 adapters added with the following, sequential, amplifications using Q5 Hot Start 2x Master Mix (New England Biolabs), each consisting of 10 cycles and a 55°C annealing temperature with the following primer pairs: reaction 1, 5' TGTGCTCTTCCGGCCATCGCCTACACGACGCT-3'/5' AAGCAGTGGTATCAACG CAGAGTACATATCATCGCAAGCAGTGGTCA 3'; reaction 2, 5' GAAGCAGAAGACGG CATAAGATAGTGTGCTCTTCCGGCCATCGC 3'/5' CGGGGTACGATGAGACACCA TATTGGTCTCAAGCAGTGGTATCAACGCAG 3'; reaction 3, 5' TCGTCGGCAGCGTCA GATGTGTATAAGAGACAGCGGGTACGATGAGACACCA 3'/5' GTCTCGTGGGC TCGGAGATGTGTATAAGAGACAGGAAGCAGAAGACGGCATAAGAGAT 3'. A final PCR was performed to index all samples individually with a 62°C annealing temperature and 20 second extension time for 7 cycles. Samples were purified by gel electrophoresis and agarose gel extraction. Products were pooled and cleaned by a 3x volume bead clean-up.

### Sequencing of naturally-occurring deletions in influenza A virus

Interferon-beta A549 reporter cell lines were infected at an MOI of 0.1 and sorted using magnetic activated cell sorting as described in the previous section as for interferon-lambda reporters, save for MS columns were used instead of LS columns. RNA was purified using either RNeasy plus mini kit (Qiagen, 74134), for interferon-depleted samples, or the RNeasy plus micro kit (Qiagen, 74034) for interferon-enriched samples. As above, RNA was converted to cDNA using the universal vRNA primers of Hoffmann *et al.* (2001) [47] and the SuperScript III First-Strand Synthesis SuperMix kit (Invitrogen, 18080400) according to the manufacturer's protocol. Samples were normalized by qPCR against HA. PCR was then performed as per Hoffmann *et al.* (2001) [47] for 16 cycles.

To generate cDNA from mRNA, normalized RNA samples were converted to cDNA using the SMART-Seq2 method and subjected to 14 cycles of amplification. [83] Both mRNA and vRNA samples were fragmented by tagmentation using the Nextera XT DNA library preparation kit from Illumina (FC-131), indexed, and sequenced.

### Computational analysis of sequencing data

The entire analysis pipeline is found at [https://github.com/a5russell/Defective\\_Library\\_Mendes\\_Russell](https://github.com/a5russell/Defective_Library_Mendes_Russell). To analyze barcodes, a custom Python script was used to parse sequencing files. Upstream and downstream junctions were assigned to barcodes separately and combined to generate a barcode-junction map. Likely due to PCR chimerism, highly abundant barcodes ended up assigning to multiple junctions, but these secondary assignments were rare. Therefore, an empirical threshold of 75% of reads must confidently assign to a single junction to assemble a barcode. A single mismatch was allowed within junction sequence for assembly. Thereafter, for each individual sample the number of times a barcode was identified within sequencing was analyzed without any further discrimination, no score values needed to be considered, and mismatched barcodes were discarded. This resulted in an assembly for the majority of observed barcodes, suggesting little benefit could be procured using a more complex analysis pipeline allowing for a greater hamming distance. As we used paired-end reads for all sequencing, barcodes or junctions were collapsed to a consensus sequence, if a mismatch was observed whichever read possessed the higher quality score at that position was presumed to be the correct nucleotide.

To analyze mRNA sequencing, reads were first trimmed using Trimmomatic providing the Nextera adapter sequences and the following variables: 2 seed mismatches, a palindrome clip threshold of 30, a simple clip threshold of 10, a minimum adapter length of 2, keep both reads, a lead of 20, a sliding window from 4 to 15, and a minimum retained length of 36. [84] Reads were then mapped against a concatenated GRCh38 human genome assembly and the A/WSN/1933 genome using STAR with default settings. [48] HTseq-count was used to count occurrences of a given transcript within our read mapping. [85] Output from HTseq-count was then used in DeSeq2 using a short R script, and results parsed using Python scripts.

To analyze vRNA sequencing, reads were also trimmed using Trimmomatic as above. Next read1 and read2 were mapped separately using STAR against the A/WSN/1933 genome with a requirement to map without any gaps by enforcing mapping to a custom gtf file consisting of the full length of each vRNA. Mismatches were excluded with the command `-outFilterMismatchNmax` of 3, deletions were largely excluded by providing a deletion open penalty of 6, and deletion per base penalty of 6. Unmapped reads were retained, and were then used in a BLASTn search against a custom BLAST database consisting of just the A/WSN/1933 genome. [49]. Input for this search used a percent identity of 90, a word size of 10, a gap open penalty of 5, extend penalty of 2, and an evalue cutoff of 0.000001. BLAST output was then parsed and reads were identified that mapped to the same segment, discontinuously. Such reads must map completely, with no “missing” bases, and must map with the same polarity, with no inversion. Repetitive elements, such as 2nt repeated at the 5' and 3' ends of the junction, which could, in theory, be assigned to either, were assigned to the 5' end arbitrarily.

BLAST output was then used to initialize a new gtf file, using all data from all technical replicates for a given biological replicate. Therefore two gtf files were generated, one for each biological replicate. Unmapped reads were then mapped using this updated file using STAR with identical parameters as above. The resulting four files, two from the first mapping and two from the second, updated, mapping, were then combined and bam flags were fixed for appropriate, paired-end, analysis. Junctions were thereafter counted in the bamfile with a

requirement that both reads mapped, neither is inconsistent with the deletion (ie if both map over the region a deletion is contained within, both show a deletion within the read), and that at least three bases are mapped on either side of the junction when considering the consensus produced by the paired-end reads.

Output was then parsed using custom Python scripts and Samtools. [86]

### qPCR

Code generating graphs in manuscript can be found at [https://github.com/a5russell/Defective\\_Library\\_Mendes\\_Russell](https://github.com/a5russell/Defective_Library_Mendes_Russell). Primers for all qPCR analyses listed in S5 File. For validation of natural diversity, purified RNA was used as described above. For all other qPCR analyses, the method of Shatzkes *et al.* (2014) [87] was used. In brief, adherent cells were incubated, after a PBS wash, with a gentle permeabilization buffer consisting of 10 mM Tris-HCl, 150 mM NaCl, 0.25% IGEPAL, and 1% DNaseI in DNase/RNase-free water and incubated at 37°C for 5–10 minutes. DNase and infectious virus were then inactivated with a 5 minute 85°C incubation. For supernatant measurements, 10  $\mu$ l of supernatant was incubated with 90  $\mu$ l buffer instead.

cDNA from lysate was generated using the High Capacity First Strand Synthesis Kit (Applied Biosystems, 4368814) with random hexamers according to the manufacturer's protocol using a final lysate concentration of 10% of the reaction volume. qPCR was thereafter performed using Luna Universal qPCR Master Mix (New England Biolabs, M3003) with manufacturer's suggested reaction conditions. For pure plasmid controls, plasmids were used at the indicated concentrations instead of cDNA in qPCR.

### Flow cytometry

Indicated cells were seeded 24 hours prior to infection, and, at the indicated time points, trypsinized and resuspended in PBS supplemented with 2% of heat-inactivated fetal bovine serum (FBS). For HA staining, cells were stained with 10  $\mu$ g/ml of H17-L19, a mouse monoclonal antibody confirmed to bind to WSN HA in a prior study. [88] Cells were refrigerated with the primary antibody for one hour at 4°C before being washed with PBS supplemented with 2% FBS and then stained with a goat anti-mouse IgG antibody conjugated to allophycocyanin and refrigerated for another hour at 4°C. Cells were washed with PBS supplemented with 2% FBS and fixed with 1% formaldehyde (BD Cytotfix) for 30 minutes at room temperature. Cells were washed with PBS and thereafter run on a flow cytometer. Data processing consisted of first generating a debris gate in FlowJo prior to export to a tsv or csv and analysis using custom Python scripts.

### RIG-I siRNA experiments

Each transfection received 50  $\mu$ L of Opti-Mem, 1.5  $\mu$ L of Lipofectamine 3000 (Invitrogen, L3000001), and 0.25  $\mu$ L of 10  $\mu$ M siRNA or 10  $\mu$ M non-targeting control. All siRNAs used were Ambion Silencer select siRNAs from Life Technologies Corporation, catalog number s223614, and s223615 for RIG-I, and 4390843 for the non-targeting control. Transfection mixes were combined with A549 cells diluted to 200,000 cells/mL in D10 media. 100,000 transfected cells were seeded in each well of a 24-well plate. 24 hours post-transfection, the media was replaced with fresh D10. 48 hours post-transfection, the media was changed to IGM and cells were infected with PB1<sub>177:385</sub>. 14 hours post-infection, cells were harvested using the method of Shatzkes *et al.* (2014) [87].



## Supporting information

### S1 File. Primers used to generate libraries.

(TSV)

### S2 File. Full differential gene expression data from Fig 4A.

(TSV)

### S3 File. Full differential junction enrichment data from Fig 5D.

(TSV)

### S4 File. Sequence of A/WSN/1933 used in this study.

(FA)

### S5 File. Sequences of all qPCR primers used in this study.

(TSV)

**S1 Fig. Polymerase I length biases from Fig 2A.** Biases present from polymerase I transcription alone. (A) Size distributions of original plasmid libraries and matched length distribution after transfection in the absence of the full viral genome replication machinery (-PB2). (B) Polymerase I transcription alone leads to a slight length bias in library composition. To analyze enrichment, or depletion, the fraction of variants falling within each 100nt bin was compared between replication-incompetent selection and the original plasmid library. Points above the dotted line represent lengths which were overrepresented in our polI transcribed library relative to a plasmid control. Points were only shown if represented in all three libraries under both conditions. R-value shown is the Spearman correlation coefficient.  $n = 3$ .

(TIF)

**S2 Fig. Validation of qPCR in Fig 2C.** (A) All four qPCR reactions exhibit linearity when tested against a plasmid control.  $R^2$  from linear regression against Ct versus the log-transformed plasmid concentration. Linear regression line shown. (B) Each qPCR is highly-specific to its cognate target, and discriminatory against non-cognate targets. Target versus non-target specificity was calculated as the relative signal of each qPCR reaction on the indicated off-target control when compared to a target control at 0.01 ng per qPCR reaction.  $n = 3$  for both panels.

(TIF)

**S3 Fig. Experiments in Fig 2D exclude second rounds of infection.** Infections were performed as in Fig 2D, save for, at 8 hours post infection, cells were infected with a pseudovirus wherein the coding sequence for HA has been replaced by eGFP, with packaging signals maintained. [64] Micrographs were taken 16 hours post infection with HA-flanked eGFP pseudovirus. A lack of fluorescent cells in our variant infections indicates that, at least largely, second rounds of replication are excluded under these high MOI conditions via superinfection exclusion.

(TIF)

### S4 Fig. Length distributions for Fig 3E.

(TIF)

**S5 Fig. Data in Fig 3E are highly replicable between libraries.** Inter-replicate enrichment values as calculated in Fig 3E. R is the Pearson correlation coefficient.

(TIF)

**S6 Fig. Confirmation of defective burden of populations used in Fig 4A, 4B and 4C.** A549 cells were infected with the indicated populations, and, at 9h post-infection, stained with a

monoclonal antibody against viral HA protein. Virus 1 and Virus 2 were the two populations used in experiments in Fig 4. Populations were grown as described in Materials and methods. The dotted line indicates the gate at which positivity was called, in this case the 99.5<sup>th</sup> percentile of the uninfected control.

(TIF)

**S7 Fig. Sampling stochasticity for interferon-enriched samples for Fig 4B.** Numbers represent raw counts of deletion-spanning junctions. Each point is a unique deletion. Virus 1 and Virus 2 were the two biologically-independent populations analyzed for Fig 4B.

(TIF)

**S8 Fig. Sampling stochasticity for interferon-depleted samples for Fig 4B.** Numbers represent raw counts of deletion-spanning junctions. Each point is a unique deletion. Virus 1 and Virus 2 were the two biologically-independent populations analyzed for Fig 4B.

(TIF)

**S9 Fig. Raw data for Fig 4B.** Total number of junction-spanning fragments normalized to the total number of fragments mapped per influenza genomic segment.  $n = 4$ .

(TIF)

**S10 Fig. Mapping depth for data from Fig 4B.** Depth per base corrected to fractional depth, area under each line sums to 1. Increased depth at 5' and 3' ends of polymerase segments consistent with increased fractions of deletions.

(TIF)

**S11 Fig. Validation of qPCR in Fig 4C.** (A) Plasmids containing full-length PB2, PB1, and PA, and those containing PB2<sub>316:267</sub>, PB1<sub>177:385</sub>, and PA<sub>328:137</sub> were mixed at the indicated molar ratios and 0.02 ng of the resultant mixture was analyzed by qPCR. Values represent the ratio of signal of our defective-spanning qPCR corrected for full-length-only signal. (B) Same as (A) but only intermediate values to show capacity to discriminate between extremes.

(TIF)

**S12 Fig. Validation of infections in Fig 4D, confirming levels of influenza transcript expression.** qPCR measuring levels of HA transcript as normalized to the housekeeping control *L32* for the indicated influenza mutants infecting A549 cells at an MOI of 0.5 at 8h post-infection. The significant difference between PB1 deletion variants is not due to incorrect dosage, as shown by similar levels of influenza transcripts. Neither PB1 mutant is replication-competent, as shown by the reduced influenza transcript levels compared to a wild-type control. Asterisks represent significantly different values in all pairwise comparisons, two-tailed t-test, using Benjamini-Hochberg multiple-hypothesis correction at an FDR of 0.05.  $n = 4$ .

(TIF)

**S13 Fig. RIG-I detects PB1<sub>177:385</sub>.** Cells were transfected with the indicated siRNA. 48 hours after transfection, cells were infected with PB1<sub>177:385</sub> at an MOI of 0.5, and RNA was harvested at 14 hours and analyzed by qPCR. Asterisks represent significantly different values in all pairwise comparisons, two-tailed t-test, using Benjamini-Hochberg multiple-hypothesis correction at an FDR of 0.05.

(TIF)

**S14 Fig. qPCR of sort in Fig 4A measuring levels of *IFNB1* transcript compared to the housekeeping control *L32*, for comparison with Fig 5A.** Asterisks indicate significantly increased *IFNB1* transcript, two-tailed t-test  $p < 0.05$ .  $n = 3$ .

(TIF)

**S15 Fig. NA<sub>189:76</sub>, when not replicating, is less stimulatory than PB1<sub>177:385</sub>.** (left) qPCR against *IFNB1* as corrected by the housekeeping gene *L32* demonstrates that, even with genome replication NA<sub>189:76</sub> is no more stimulatory than PB1<sub>177:385</sub>, and when replication is suppressed, it is actually less stimulatory. (right) qPCR against *PA*, as corrected by the housekeeping gene *L32*, was used to validate the suppression of viral replication by the antiviral nucleoside ribavirin, and matched dose between our variants. A549 cells were pretreated with 200  $\mu$ M ribavirin for 2h and then infected at an MOI of 0.5. RNA was harvested for analysis after 14 hours of infection. Astrisks indicate conditions wherein signal was significantly different between variants. Two-tailed t-test. n = 4.

(TIF)

**S16 Fig. Gating for Fig 6B.** Individual replicates shown. Dotted line represents positivity threshold Gating drawn on the uninfected control before application to infected samples, set at the 99.9<sup>th</sup> percentile of uninfected. Data shown as kernel density estimates with Gaussian distributions. n = 3.

(TIF)

**S17 Fig. Gating for Fig 6C.** Individual replicates shown. Dotted line represents gating. Gating drawn on the uninfected control before application to infected samples, set at the 99.9<sup>th</sup> percentile of uninfected. Data were subsampled to 5000 events and shown as individual points. Points in orange were called as interferon-positive. n = 3.

(TIF)

**S18 Fig. Gating for Fig 7A for HA and mCherry.** Individual replicates shown. Dotted line represents gating. Gating drawn on the uninfected control before application to infected samples, set at the 99.95<sup>th</sup> percentile of uninfected. Data were subsampled to 5000 events and shown as individual points. n = 3.

(TIF)

**S19 Fig. Gating for Fig 7A for the interferon lambda reporter.** Individual replicates shown. Dotted line represents gating. Gating drawn on the uninfected control before application to infected samples, set at the 99.95<sup>th</sup> percentile of uninfected. Data shown as kernel density estimates with Gaussian distributions. n = 3.

(TIF)

**S20 Fig. Gating for Fig 7B.** Individual replicates shown. Dotted line represents gating. Gating drawn on the uninfected control before application to infected samples, set at the 99.9<sup>th</sup> percentile of uninfected. Data were subsampled to 5000 events and shown as individual points. Points in orange were called as interferon-positive. n = 3.

(TIF)

## Acknowledgments

We thank the members of the Russell laboratory, and Jacquelyn Bragg, for comments on the manuscript.

## Author Contributions

**Conceptualization:** Marisa Mendes, Alistair B. Russell.

**Data curation:** Marisa Mendes, Alistair B. Russell.

**Formal analysis:** Marisa Mendes, Alistair B. Russell.

**Funding acquisition:** Alistair B. Russell.

**Investigation:** Marisa Mendes, Alistair B. Russell.

**Methodology:** Marisa Mendes, Alistair B. Russell.

**Project administration:** Marisa Mendes, Alistair B. Russell.

**Software:** Alistair B. Russell.

**Supervision:** Alistair B. Russell.

**Validation:** Alistair B. Russell.

**Visualization:** Alistair B. Russell.

**Writing – original draft:** Marisa Mendes, Alistair B. Russell.

**Writing – review & editing:** Marisa Mendes, Alistair B. Russell.

## References

1. Donald HB, Isaacs A. Counts of Influenza Virus Particles. *Journal of General Microbiology*. 1954; 10(3):457–464. <https://doi.org/10.1099/00221287-10-3-457> PMID: 13174769
2. Brooke CB. Biological activities of ‘noninfectious’ influenza A virus particles. *Future Virology*. 2014; 9(1):41–51. <https://doi.org/10.2217/fvl.13.118> PMID: 25067941
3. Ngunjiri JM, Sekellick MJ, Marcus PI. Clonogenic Assay of Type A Influenza Viruses Reveals Noninfectious Cell-Killing (Apoptosis-Inducing) Particles. *Journal of Virology*. 2008; 82(6):2673–2680. <https://doi.org/10.1128/JVI.02221-07> PMID: 18184709
4. Marcus PI, Rojek JM, Sekellick MJ. Interferon Induction and/or Production and Its Suppression by Influenza A Viruses. *Journal of Virology*. 2005; 79(5):2880–2890. <https://doi.org/10.1128/JVI.79.5.2880-2890.2005> PMID: 15709007
5. Brooke CB, Ince WL, Wrammert J, Ahmed R, Wilson PC, Bennink JR, et al. Most Influenza A Virions Fail To Express at Least One Essential Viral Protein. *Journal of Virology*. 2013; 87(6):3155–3162. <https://doi.org/10.1128/JVI.02284-12> PMID: 23283949
6. Akpınar F, Timm A, Yin J. High-Throughput Single-Cell Kinetics of Virus Infections in the Presence of Defective Interfering Particles. *Journal of Virology*. 2016; 90(3):1599–1612. <https://doi.org/10.1128/JVI.02190-15> PMID: 26608322
7. Genoyer E, López CB. The Impact of Defective Viruses on Infection and Immunity. *Annual Review of Virology*. 2019; 6(1):1–20. <https://doi.org/10.1146/annurev-virology-092818-015652> PMID: 31082310
8. Tapia K, Kim Wk, Sun Y, Mercado-López X, Dunay E, Wise M, et al. Defective Viral Genomes Arising In Vivo Provide Critical Danger Signals for the Triggering of Lung Antiviral Immunity. *PLoS Pathogens*. 2013; 9(10):e1003703. <https://doi.org/10.1371/journal.ppat.1003703> PMID: 24204261
9. Odagiri T, Tashiro M. Segment-specific noncoding sequences of the influenza virus genome RNA are involved in the specific competition between defective interfering RNA and its progenitor RNA segment at the virion assembly step. *Journal of Virology*. 1997; 71(3):2138–2145. <https://doi.org/10.1128/jvi.71.3.2138-2145.1997> PMID: 9032347
10. Laske T, Heldt FS, Hoffmann H, Frensing T, Reichl U. Modeling the intracellular replication of influenza A virus in the presence of defective interfering RNAs. *Virus Research*. 2016; 213:90–99. <https://doi.org/10.1016/j.virusres.2015.11.016> PMID: 26592173
11. Vignuzzi M, López CB. Defective viral genomes are key drivers of the virus–host interaction. *Nature Microbiology*. 2019; 4(7):1075–1087. <https://doi.org/10.1038/s41564-019-0465-y> PMID: 31160826
12. Felt SA, Sun Y, Jozwik A, Paras A, Habibi MS, Nickle D, et al. Detection of respiratory syncytial virus defective genomes in nasal secretions is associated with distinct clinical outcomes. *Nature Microbiology*. 2021; p. 1–10. PMID: 33795879
13. Saira K, Lin X, DePasse JV, Halpin R, Twaddle A, Stockwell T, et al. Sequence Analysis of In Vivo Defective Interfering-Like RNA of Influenza A H1N1 Pandemic Virus. *Journal of Virology*. 2013; 87(14):8064–8074. <https://doi.org/10.1128/JVI.00240-13> PMID: 23678180
14. Vasilijevic J, Zamarreño N, Oliveros JC, Rodriguez-Frandsen A, Gómez G, Rodriguez G, et al. Reduced accumulation of defective viral genomes contributes to severe outcome in influenza virus

- infected patients. *PLOS Pathogens*. 2017; 13(10):e1006650. <https://doi.org/10.1371/journal.ppat.1006650> PMID: 29023600
15. Rezelj VV, Carrau L, Merwaiss F, Levi LI, Erazo D, Tran QD, et al. Defective viral genomes as therapeutic interfering particles against flavivirus infection in mammalian and mosquito hosts. *Nature Communications*. 2021; 12(1):2290. <https://doi.org/10.1038/s41467-021-22341-7> PMID: 33863888
  16. Rand U, Kupke SY, Shkaret H, Hein MD, Hirsch T, Marichal-Gallardo P, et al. Antiviral activity of influenza A virus defective interfering particles against SARS-CoV-2 replication in vitro through stimulation of innate immunity. *bioRxiv*. 2021; p. 2021.02.19.431972. <https://doi.org/10.3390/cells10071756> PMID: 34359926
  17. Tanner EJ, Jung SY, Glazier J, Thompson C, Zhou Y, Martin B, et al. Discovery and Engineering of a Therapeutic Interfering Particle (TIP): a combination self-renewing antiviral. *bioRxiv*. 2019; p. 820456.
  18. Dimmock NJ, Easton AJ. Defective Interfering Influenza Virus RNAs: Time To Reevaluate Their Clinical Potential as Broad-Spectrum Antivirals? *Journal of Virology*. 2014; 88(10):5217–5227. <https://doi.org/10.1128/JVI.03193-13> PMID: 24574404
  19. Nayak DP, Sivasubramanian N, Davis AR, Cortini R, Sung J. Complete sequence analyses show that two defective interfering influenza viral RNAs contain a single internal deletion of a polymerase gene. *Proceedings of the National Academy of Sciences*. 1982; 79(7):2216–2220. <https://doi.org/10.1073/pnas.79.7.2216> PMID: 6954536
  20. Nakajima K, Ueda M, Sugiura A. Origin of small RNA in von Magnus particles of influenza virus. *Journal of Virology*. 1979; 29(3):1142–1148. <https://doi.org/10.1128/jvi.29.3.1142-1148.1979> PMID: 448797
  21. Akkina RK, Chambers TM, Nayak DP. Mechanism of interference by defective-interfering particles of influenza virus: Differential reduction of intracellular synthesis of specific polymerase proteins. *Virus Research*. 1984; 1(8):687–702. [https://doi.org/10.1016/0168-1702\(84\)90059-5](https://doi.org/10.1016/0168-1702(84)90059-5)
  22. Alnaji FG, Holmes JR, Rendon G, Vera CJ, Fields CJ, Martin BE, et al. Sequencing Framework for the Sensitive Detection and Precise Mapping of Defective Interfering Particle-Associated Deletions across Influenza A and B Viruses. *Journal of Virology*. 2019; 93(11). <https://doi.org/10.1128/JVI.00354-19> PMID: 30867305
  23. Pelz L, Rüdiger D, Alnaji FG, Genzel Y, Brooke CB, Kupke SY, et al. Semi-continuous propagation of influenza A virus and its defective interfering particles: analyzing the dynamic competition to select candidates for antiviral therapy. *bioRxiv*. 2021; p. 2021.02.08.430251. PMID: 34550771
  24. Kupke S, Ly LH, Börno S, Ruff A, Timmermann B, Vingron M, et al. Single-Cell Analysis Uncovers a Vast Diversity in Intracellular Viral Defective Interfering RNA Content Affecting the Large Cell-to-Cell Heterogeneity in Influenza A Virus Replication. *Viruses*. 2020; 12(1):71. <https://doi.org/10.3390/v12010071> PMID: 31936115
  25. Cocquet J, Chong A, Zhang G, Veitia RA. Reverse transcriptase template switching and false alternative transcripts. *Genomics*. 2006; 88(1):127–131. <https://doi.org/10.1016/j.ygeno.2005.12.013> PMID: 16457984
  26. Boussier J, Munier S, Achouri E, Meyer B, Crescenzo-Chaigne B, Behillil S, et al. RNA-seq accuracy and reproducibility for the mapping and quantification of influenza defective viral genomes. *RNA*. 2020; 26(12):1905–1918. <https://doi.org/10.1261/rna.077529.120> PMID: 32929001
  27. Notton T, Glazier JJ, Saykally VR, Thompson CE, Weinberger LS. RanDeL-Seq: a High-Throughput Method to Map Viral cis—and trans -Acting Elements. *mBio*. 2021; 12(1). <https://doi.org/10.1128/mBio.01724-20> PMID: 33468683
  28. Luria SE, Delbrück M. MUTATIONS OF BACTERIA FROM VIRUS SENSITIVITY TO VIRUS RESISTANCE. *Genetics*. 1943; 28(6):491–511. <https://doi.org/10.1093/genetics/28.6.491> PMID: 17247100
  29. Liang Y, Hong Y, Parslow TG. cis-Acting Packaging Signals in the Influenza Virus PB1, PB2, and PA Genomic RNA Segments. *Journal of Virology*. 2005; 79(16):10348–10355. <https://doi.org/10.1128/JVI.79.16.10348-10355.2005> PMID: 16051827
  30. Goto H, Muramoto Y, Noda T, Kawaoka Y. The Genome-Packaging Signal of the Influenza A Virus Genome Comprises a Genome Incorporation Signal and a Genome-Bundling Signal. *Journal of Virology*. 2013; 87(21):11316–11322. <https://doi.org/10.1128/JVI.01301-13> PMID: 23926345
  31. Azzeh M, Flick R, Hobom G. Functional Analysis of the Influenza A Virus cRNA Promoter and Construction of an Ambisense Transcription System. *Virology*. 2001; 289(2):400–410. <https://doi.org/10.1006/viro.2001.1107> PMID: 11689061
  32. Flick R, Neumann G, Hoffmann E, Neumeier E, Hobom G. Promoter elements in the influenza vRNA terminal structure. *RNA (New York, NY)*. 1996; 2(10):1046–57. PMID: 8849780
  33. Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA, Smith HO. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods*. 2009; 6(5):343–345. <https://doi.org/10.1038/nmeth.1318> PMID: 19363495

34. Duhaut SD, McCauley JW. Defective RNAs Inhibit the Assembly of Influenza Virus Genome Segments in a Segment-Specific Manner. *Virology*. 1996; 216(2):326–337. <https://doi.org/10.1006/viro.1996.0068> PMID: 8607262
35. Chen W, Calvo PA, Malide D, Gibbs J, Schubert U, Bacik I, et al. A novel influenza A virus mitochondrial protein that induces cell death. *Nature Medicine*. 2001; 7(12):1306–1312. <https://doi.org/10.1038/nm1201-1306> PMID: 11726970
36. Ho JSY, Angel M, Ma Y, Sloan E, Wang G, Martinez-Romero C, et al. Hybrid Gene Origination Creates Human-Virus Chimeric Proteins during Infection. *Cell*. 2020; 181(7):1502–1517.e23. <https://doi.org/10.1016/j.cell.2020.05.035> PMID: 32559462
37. Alnaji FG, Reiser WK, Velthuis At, Brooke CB. Influenza A virus defective viral genomes are inefficiently packaged into virions relative to wild-type genomic RNAs. *bioRxiv*. 2021. <https://doi.org/10.1128/mBio.02959-21> PMID: 34809454
38. Arora P, Bdeir N, Gärtner S, Reiter S, Pelz L, Felgenhauer U, et al. Interferon induction and not replication interference mainly determines anti-influenza virus activity of defective interfering particles. *bioRxiv*. 2021; p. 2021.09.20.461172.
39. Neumann G, Watanabe T, Ito H, Watanabe S, Goto H, Gao P, et al. Generation of influenza A viruses entirely from cloned cDNAs. *Proceedings of the National Academy of Sciences*. 1999; 96(16):9345–9350. <https://doi.org/10.1073/pnas.96.16.9345> PMID: 10430945
40. Widjaja I, Vries Ed, Rottier PJM, Haan CAMd. Competition between Influenza A Virus Genome Segments. *PLoS ONE*. 2012; 7(10):e47529. <https://doi.org/10.1371/journal.pone.0047529> PMID: 23071819
41. Huang IC, Li W, Sui J, Marasco W, Choe H, Farzan M. Influenza A Virus Neuraminidase Limits Viral Superinfection. *Journal of Virology*. 2008; 82(10):4834–4843. <https://doi.org/10.1128/JVI.00079-08> PMID: 18321971
42. Sun J, Brooke CB. Influenza A Virus Superinfection Potential Is Regulated by Viral Genomic Heterogeneity. *mBio*. 2018; 9(5):e01761–18. <https://doi.org/10.1128/mBio.01761-18> PMID: 30377278
43. Liu G, Lu Y, Liu Q, Zhou Y. Inhibition of Ongoing Influenza A Virus Replication Reveals Different Mechanisms of RIG-I Activation. *Journal of Virology*. 2019; 93(6). <https://doi.org/10.1128/JVI.02066-18> PMID: 30602605
44. Chen S, Short JAL, Young DF, Killip MJ, Schneider M, Goodbourn S, et al. Heterocellular induction of interferon by negative-sense RNA viruses. *Virology*. 2010; 407(2):247–255. <https://doi.org/10.1016/j.virol.2010.08.008> PMID: 20833406
45. Killip MJ, Jackson D, Pérez-Cidoncha M, Fodor E, Randall RE. Single-cell studies of IFN- $\beta$  promoter activation by wild-type and NS1-defective influenza A viruses. *Journal of General Virology*. 2017; 98(3):357–363. <https://doi.org/10.1099/jgv.0.000687>
46. Russell AB, Elshina E, Kowalsky JR, Velthuis AJWT, Bloom JD. Single-Cell Virus Sequencing of Influenza Infections That Trigger Innate Immunity. *Journal of virology*. 2019; 93(14). <https://doi.org/10.1128/JVI.00500-19> PMID: 31068418
47. Hoffmann E, Stech J, Guan Y, Webster RG, Perez DR. Universal primer set for the full-length amplification of all influenza A viruses. *Archives of Virology*. 2001; 146(12):2275–2289. <https://doi.org/10.1007/s007050170002> PMID: 11811679
48. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013; 29(1):15–21. <https://doi.org/10.1093/bioinformatics/bts635> PMID: 23104886
49. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *Journal of Molecular Biology*. 1990; 215(3):403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2) PMID: 2231712
50. Schwartz SL, Lowen AC. Droplet digital PCR: A novel method for detection of influenza virus defective interfering particles. *Journal of Virological Methods*. 2016; 237:159–165. <https://doi.org/10.1016/j.jviromet.2016.08.023> PMID: 27590979
51. Noda T, Murakami S, Nakatsu S, Imai H, Muramoto Y, Shindo K, et al. Importance of the 1+7 configuration of ribonucleoprotein complexes for influenza A virus genome packaging. *Nature Communications*. 2018; 9(1):54. <https://doi.org/10.1038/s41467-017-02517-w> PMID: 29302061
52. Chou Yy, Vafabakhsh R, Doğanay S, Gao Q, Ha T, Palese P. One influenza virus particle packages eight unique viral RNAs as shown by FISH analysis. *Proceedings of the National Academy of Sciences*. 2012; 109(23):9101–9106. <https://doi.org/10.1073/pnas.1206069109> PMID: 22547828
53. Noda T, Sagara H, Yen A, Takada A, Kida H, Cheng RH, et al. Architecture of ribonucleoprotein complexes in influenza A virus particles. *Nature*. 2006; 439(7075):490–492. <https://doi.org/10.1038/nature04378> PMID: 16437116

54. Rehwinkel J, Tan CP, Goubau D, Schulz O, Pichlmair A, Bier K, et al. RIG-I detects viral genomic RNA during negative-strand RNA virus infection. *Cell*. 2010; 140(3):397–408. <https://doi.org/10.1016/j.cell.2010.01.020> PMID: 20144762
55. Baum A, Sachidanandam R, García-Sastre A. Preference of RIG-I for short viral RNA molecules in infected cells revealed by next-generation sequencing. *Proceedings of the National Academy of Sciences*. 2010; 107(37):16303–16308. <https://doi.org/10.1073/pnas.1005077107> PMID: 20805493
56. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*. 2014; 15(12):550. <https://doi.org/10.1186/s13059-014-0550-8> PMID: 25516281
57. Velthuis AJT, Long JC, Bauer DL, Fan RL, Yen HL, Sharps J, et al. Mini viral RNAs act as innate immune agonists during influenza virus infection. *Nature Microbiology*. 2018; 3(11):1234–1242. <https://doi.org/10.1038/s41564-018-0240-5>
58. Vreede FT, Jung TE, Brownlee GG. Model Suggesting that Replication of Influenza Virus Is Regulated by Stabilization of Replicative Intermediates†. *Journal of Virology*. 2004; 78(17):9568–9572. <https://doi.org/10.1128/JVI.78.17.9568-9572.2004> PMID: 15308750
59. Scholtissek C. Inhibition of influenza RNA synthesis by virazole (ribavirin). *Archives of Virology*. 1976; 50(4):349–352. <https://doi.org/10.1007/BF01317961> PMID: 1275709
60. Vanderlinden E, Vrancken B, Houdt JV, Rajwanshi VK, Gillemot S, Andrei G, et al. Distinct Effects of T-705 (Favipiravir) and Ribavirin on Influenza Virus Replication and Viral RNA Synthesis. *Antimicrobial Agents and Chemotherapy*. 2016; 60(11):6679–6691. <https://doi.org/10.1128/AAC.01156-16> PMID: 27572398
61. Tan YH, Armstrong JA, Ke YH, Ho M. Regulation of Cellular Interferon Production: Enhancement by Antimetabolites. *Proceedings of the National Academy of Sciences*. 1970; 67(1):464–471. <https://doi.org/10.1073/pnas.67.1.464> PMID: 5272327
62. Raj NB, Pitha PM. Two levels of regulation of beta-interferon gene expression in human cells. *Proceedings of the National Academy of Sciences*. 1983; 80(13):3923–3927. <https://doi.org/10.1073/pnas.80.13.3923> PMID: 6575385
63. Ringold GM, Dieckmann B, Vannice JL, Trahey M, McCormick F. Inhibition of protein synthesis stimulates the transcription of human beta-interferon genes in Chinese hamster ovary cells. *Proceedings of the National Academy of Sciences*. 1984; 81(13):3964–3968. <https://doi.org/10.1073/pnas.81.13.3964> PMID: 6330726
64. Russell AB, Trapnell C, Bloom JD. Extreme heterogeneity of influenza virus infection in single cells. *eLife*. 2018; 7. <https://doi.org/10.7554/eLife.32303> PMID: 29451492
65. Bean WJ, Simpson RW. Primary transcription of the influenza virus genome in permissive cells. *Virology*. 1973; 56(2):646–651. [https://doi.org/10.1016/0042-6822\(73\)90067-6](https://doi.org/10.1016/0042-6822(73)90067-6) PMID: 4796552
66. Velthuis AJ, Long JS, Barclay WS. Assays to Measure the Activity of Influenza Virus Polymerase. *Methods in molecular biology* (Clifton, NJ). 2018; 1836:343–374. [https://doi.org/10.1007/978-1-4939-8678-1\\_17](https://doi.org/10.1007/978-1-4939-8678-1_17)
67. Martínez-Sobrido L, Cadagan R, Steel J, Basler CF, Palese P, Moran TM, et al. Hemagglutinin-Pseudotyped Green Fluorescent Protein-Expressing Influenza Viruses for the Detection of Influenza Virus Neutralizing Antibodies. *Journal of Virology*. 2009; 84(4):2157–2163. <https://doi.org/10.1128/JVI.01433-09> PMID: 19939917
68. Österlund P, Strengell M, Sarin LP, Poranen MM, Fagerlund R, Melén K, et al. Incoming Influenza A Virus Evades Early Host Recognition, while Influenza B Virus Induces Interferon Expression Directly upon Entry. *Journal of Virology*. 2012; 86(20):11183–11193. <https://doi.org/10.1128/JVI.01050-12> PMID: 22855501
69. Fay EJ, Aron SL, Macchietto MG, Markman MW, Esser-Nobis K, Gale M, et al. Cell type- and replication stage-specific influenza virus responses in vivo. *PLOS Pathogens*. 2020; 16(8):e1008760. <https://doi.org/10.1371/journal.ppat.1008760> PMID: 32790753
70. Killip MJ, Smith M, Jackson D, Randall RE. Activation of the Interferon Induction Cascade by Influenza A Viruses Requires Viral RNA Synthesis and Nuclear Export. *Journal of Virology*. 2014; 88(8):3942–3952. <https://doi.org/10.1128/JVI.03109-13> PMID: 24478437
71. Nilsson-Payant BE, Blanco-Melo D, Uhl S, Escudero-Pérez B, Olschewski S, Thibault P, et al. Reduced Nucleoprotein Availability Impairs Negative-Sense RNA Virus Replication and Promotes Host Recognition. *Journal of Virology*. 2021; 95(9). <https://doi.org/10.1128/JVI.02274-20> PMID: 33568513
72. Sage VL, Kanarek JP, Snyder DJ, Cooper VS, Lakdawala SS, Lee N. Mapping of Influenza Virus RNA-RNA Interactions Reveals a Flexible Network. *Cell Reports*. 2020; 31(13):107823. <https://doi.org/10.1016/j.celrep.2020.107823> PMID: 32610124

73. Dadonaite B, Gilbertson B, Knight ML, Trifkovic S, Rockman S, Laederach A, et al. The structure of the influenza A virus genome. *Nature Microbiology*. 2019; 4(11):1781–1789. <https://doi.org/10.1038/s41564-019-0513-7> PMID: 31332385
74. Xu J, Mercado-López X, Grier JT, Kim Wk, Chun LF, Irvine EB, et al. Identification of a Natural Viral RNA Motif That Optimizes Sensing of Viral RNA by RIG-I. *mBio*. 2015; 6(5):e01265–15. <https://doi.org/10.1128/mBio.01265-15> PMID: 26443454
75. Zhao M, Zhang J, Phatnani H, Scheu S, Maniatis T. Stochastic Expression of the Interferon- $\beta$  Gene. *PLoS Biology*. 2012; 10(1):e1001249. <https://doi.org/10.1371/journal.pbio.1001249> PMID: 22291574
76. Frensing T, Heldt FS, Pflugmacher A, Behrendt I, Jordan I, Flockerzi D, et al. Continuous Influenza Virus Production in Cell Culture Shows a Periodic Accumulation of Defective Interfering Particles. *PLoS ONE*. 2013; 8(9):e72288. <https://doi.org/10.1371/journal.pone.0072288> PMID: 24039749
77. Doud MB, Bloom JD. Accurate Measurement of the Effects of All Amino-Acid Mutations on Influenza Hemagglutinin. *Viruses*. 2016; 8(6):155. <https://doi.org/10.3390/v8060155> PMID: 27271655
78. Bloom JD, Gong LI, Baltimore D. Permissive Secondary Mutations Enable the Evolution of Influenza Oseltamivir Resistance. *Science*. 2010; 328(5983):1272–1275. <https://doi.org/10.1126/science.1187816> PMID: 20522774
79. O'Connell RM, Balazs AB, Rao DS, Kivork C, Yang L, Baltimore D. Lentiviral Vector Delivery of Human Interleukin-7 (hIL-7) to Human Immune System (HIS) Mice Expands T Lymphocyte Populations. *PLoS ONE*. 2010; 5(8):e12009. <https://doi.org/10.1371/journal.pone.0012009> PMID: 20700454
80. Hoffmann E, Neumann G, Kawaoka Y, Hobom G, Webster RG. A DNA transfection system for generation of influenza A virus from eight plasmids. *Proceedings of the National Academy of Sciences*. 2000; 97(11):6108–6113. <https://doi.org/10.1073/pnas.100133697> PMID: 10801978
81. Reed LJ, Muench H. A SIMPLE METHOD OF ESTIMATING FIFTY PER CENT ENDPOINTS. *American Journal of Epidemiology*. 1938; 27(3):493–497. <https://doi.org/10.1093/oxfordjournals.aje.a118408>
82. Breslauer KJ, Frank R, Blöcker H, Marky LA. Predicting DNA duplex stability from the base sequence. *Proceedings of the National Academy of Sciences*. 1986; 83(11):3746–3750. <https://doi.org/10.1073/pnas.83.11.3746> PMID: 3459152
83. Picelli S, Björklund AK, Faridani OR, Sagasser S, Winberg G, Sandberg R. Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nature Methods*. 2013; 10(11):1096–1098. <https://doi.org/10.1038/nmeth.2639> PMID: 24056875
84. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014; 30(15):2114–2120. <https://doi.org/10.1093/bioinformatics/btu170> PMID: 24695404
85. Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics*. 2015; 31(2):166–169. <https://doi.org/10.1093/bioinformatics/btu638> PMID: 25260700
86. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009; 25(16):2078–2079. <https://doi.org/10.1093/bioinformatics/btp352> PMID: 19505943
87. Shatzkes K, Teferedegne B, Murata H. A simple, inexpensive method for preparing cell lysates suitable for downstream reverse transcription quantitative PCR. *Scientific Reports*. 2014; 4(1):4659. <https://doi.org/10.1038/srep04659> PMID: 24722424
88. Doud MB, Hensley SE, Bloom JD. Complete mapping of viral escape from neutralizing antibodies. *PLoS Pathogens*. 2017; 13(3):e1006271. <https://doi.org/10.1371/journal.ppat.1006271> PMID: 28288189