A statistical framework for quantifying the nuclear export rate of influenza viral mRNAs.

Michi Miura[a#], Naho Kiuchi[a], Siu-Ying Lau[b], Bobo Wing-Yee Mok[b], Hiroshi Ushirogawa[a], Tadasuke Naito[a], Honglin Chen[b], Mineki Saito[a#]

[a] Department of Microbiology, Kawasaki Medical School, Japan

[b] State Key Laboratory for Emerging Infectious Diseases, Department of Microbiology, The University of Hong Kong, Hong Kong SAR, People's Republic of China

[#]Correspondence

Michi Miura: michi.miura@med.kawasaki-m.ac.jp; michi.miura.res@gmail.com

Mineki Saito: mineki@med.kawasaki-m.ac.jp

Abstract

Influenza A virus acutely transcribes viral mRNAs from the eight segmented viral genome when it infects. The kinetics of viral transcription, nuclear export of viral transcripts, and their potential variation between the eight segments are poorly characterised. Here we introduce a statistical framework for estimating the nuclear export rate of each segment from a snapshot of mRNA *in situ* localisation at a single time point, exploiting the cell-to-cell variation observed by an imaging-based *in situ* transcriptome assay. A mathematical modelling indicated that the relationship between the nuclear ratio and the total count of mRNAs in single cells is dictated by a proxy for the nuclear export rate. Using this model, we showed that the two influenza viral antigens hemagglutinin and neuraminidase were the slowest segments in the nuclear export, suggesting that influenza A virus uses the nuclear retention of viral transcripts to delay the expression of antigenic molecules. Our framework presented in this study can be widely used for investigating the nuclear retention of nascent transcripts produced in a transcription burst.

Introduction

Temporal gene expression is of vital importance for viral pathogens to proliferate in the host. Human exogenous retroviruses insert their RNA genome into the host chromatin and use host factors to minimise the viral protein production to evade the immune response and persist in the host [1]. In contrast, influenza A virus acutely infects the respiratory tract and produces thousands of viral progenitors in a single round of infection to spread in the surrounding epithelial cells by transcribing and replicating its RNA genome [2]. Influenza A virus genome is segmented into eight single-stranded RNAs, and expression of each gene is essential for productive infection [3]. Unlike the human exogenous retroviruses, influenza A virus does not seem to have a sophisticated mechanism of controlling the viral gene expression at the transcription level. However, it has been known for decades that hemagglutinin (HA) and neuraminidase (NA), the two influenza viral antigens targeted by the host immune response, are produced late during the course of infection [4, 5, 6], a mechanism of which is currently unclear.

Eight viral RNA segments are transcribed by the same set of molecular machinery [7]. Viral genome is contained in the viral envelope in association with viral proteins. The 5′ and 3′ ends of each viral genomic RNA hybridise to form an intramolecular loop [8]. The RNA-dependent RNA polymerase of influenza A virus, comprising PB1 (polymerase basic 1), PB2 (polymerase basic 2), and PA (polymerase acidic) subunits, is associated with the 5′ and 3′ joint end; the rest of the loop is bound by the oligomeric viral nucleoprotein (NP) [9]. The viral ribonucleoprotein complex (vRNPs) initiates the viral transcription in the nucleus by deriving a 5′-capped short RNA fragment from the host nascent RNA on the stalled RNA polymerase II (Pol II) [10]. At the end of transcription, the viral polymerase adds a poly-A tail to the viral mRNA by stuttering on a short uridine repeat in the viral RNA template [11]. Thus, eight viral

segments sharing the common mechanism of viral transcription with no additional layers of transcriptional controls e.g., transcription-factor binding, the transcription of eight influenza viral genes is likely to occur at the same rate across the segments that is determined by the collision of the vRNP and the stalled Pol II.

Another possible factor that would determine the rate of transcription is the length of viral RNA template [12, 13], given that the duration of transcription is rate-limiting. However, this does not agree with the order of protein expression in the influenza virus infection: PB2 (2.4 kb), PB1 (2.3 kb) and PA (2.2 kb), the three longest viral genes, and NP (1.8 kb) are expressed first, followed by the expression of three segments HA (1.8 kb), NA (1.5 kb) and M (1.0 kb) [4, 6].

A third possibility is the nuclear retention of viral mRNAs. Nuclear retention of mRNAs has gained an attention as a point of control on the protein expression in the context of transcription burst [14]. Previous studies reported the accumulation of M mRNAs in the nuclear speckles [15], NXF1-dependent nuclear export of HA, NA and M mRNAs [16, 17], and the nuclear export inhibition by the host factor hnRNPAB for M, HA, NP and NS mRNAs [18]. It is becoming clear that the influenza viral segments show differential dependencies on the nuclear export factors; and this dependency is not accounted for by the RNA structure (i.e., whether they are intron-less, unspliced or spliced). Because an incomplete set of viral segments was studied in each of these studies, the influenza viral mRNA transport remains elusive.

In this paper, we introduce a statistical framework for systematically quantifying the nuclear transport across all the eight viral segments. We localised eight viral mRNAs simultaneously in single cells *in situ*. Intracellular distribution of viral mRNAs indicated that the nuclear export rate of viral transcripts varied according to the segments. A statistical model, exploiting cell-to-cell variation in the abundance of transcripts due to the stochastic nature of multiple virological

processes prior to the onset of viral transcription, allowed for the estimation of nuclear export rate from the intracellular distribution of mRNAs observed at a single time point. With this model, we show that hemagglutinin and neuraminidase mRNAs were the slowest in exiting the nucleus, suggesting that influenza A virus uses nuclear retention of mRNAs to delay the production of antigenic molecules in the course of viral infection. This study demonstrates that the nuclear export rate of nascent transcripts can be estimated from a snapshot of RNA *in situ* measurement in single cells, and our result serves as a reference for the nuclear export rate of viral transcripts across the segments, shedding light on the comprehensive understanding on the nuclear export of viral transcripts.

Results

Diffraction-limited, single-colour imaging of viral mRNAs.

First, we used single-molecule RNA FISH (smFISH) to study the viral transcription of PB1, HA, NP and NS. The lung carcinoma cell line A549 was infected with the WSN strain of influenza A virus at the multiplicity of infection (M.O.I.) of ~2 for two hours, and the viral mRNAs were visualised with a tile of short oligonucleotide probes, each labelled with a single fluorophore [19] (Fig. 1A). Viral mRNAs were detected as single diffraction-limited spots in the three-dimensional image stacks, allowing for absolute mRNA quantification (Fig. 1B).

We localised the mRNAs in 3D and counted the total number of mRNAs in each cell (Fig. 1C). The median number of mRNAs per cell was 145, 231, 310 and 303 molecules for PB1, HA, NP and NS, respectively (Fig. 1D), in line with the theoretical prediction that the rate of mRNA production is partly determined by the gene length.

Figure 1

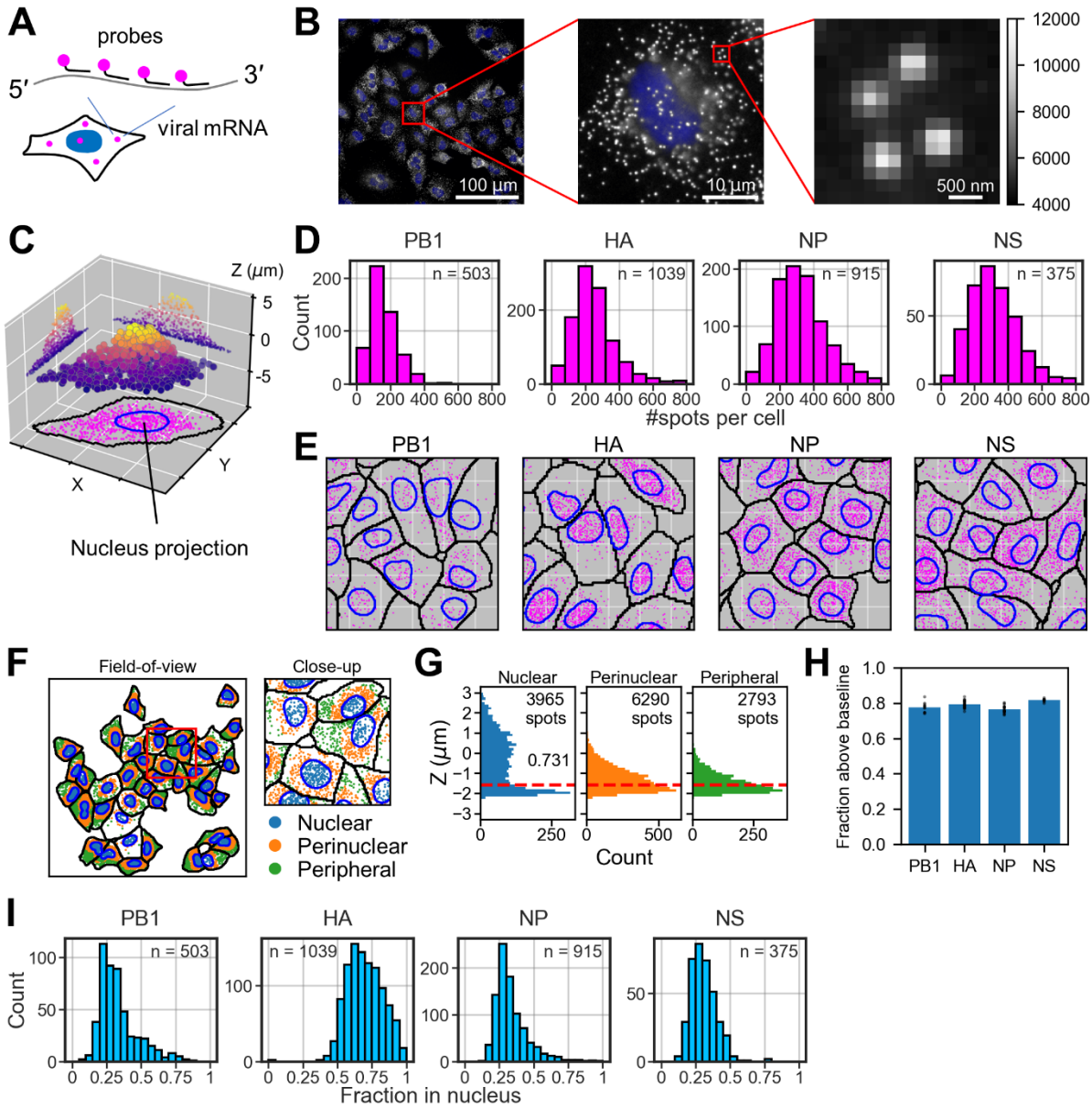


Fig. 1 Viral mRNA detection by single-colour RNA FISH. (A) A schematic diagram of mRNA labelling for the diffraction-limited imaging. Each mRNA was hybridised with a tile of oligonucleotide, each labelled with a single fluorophore at the 3′ end. **(B)** A representative field-of-view of the microscope image in which NP mRNAs were stained with Quasar 670. White, smFISH signal; blue, nucleus stained with DAPI. The red square indicates the area for the sequential close-up of the image. The signal intensity is shown on the same grey scale across the three panels, indicated on the right. **(C)** Spot detection in 3D. Spots are projected to either the XZ or YZ plane to indicate the distribution of spots in a typical 3D cell volume. The colour of each spot indicates the Z position of that spot. Spot projection to the XY plane at the bottom (magenta) highlights the XY positions of spots relative to the nucleus boundary (blue), that is, nucleus projection on the XY plane. The grid interval on the X and Y axis is ca. 18.3 µm. **(D)** Total count of mRNAs in each cell. The sample size (n) indicates the number of cells segmented in the microscope images. **(E)** Intracellular distribution of viral mRNAs. Spots in the 3D volume were projected on the XY plane, together with the cytoplasmic boundary (black solid line) and the nucleus (blue solid line). The grid interval indicates ca. 18.3 µm. **(F)** Spot segmentation within cells. The panel on the left shows the entire field-of-view (measures ~294 µm), and the panel on the right shows the close-up of the area indicated in the red square (~73 µm). Spots within the 2D nuclear projection are indicated in blue (nuclear); spots within the expanded nuclear boundary twofold from the centre of mass of the nucleus in 2D are indicated in orange (perinuclear); and spots outside this expansion are indicated in green (peripheral). **(G)** Distribution of

spots on the Z axis in the three segments (nuclear, perinuclear and peripheral) presented in panel F. The red dashed line shows the median Z coordinates of the peripheral spots, used to indicate the cell baseline. The number in the histogram for nuclear spots (0.731) indicates the fraction of spots that were above the cell base line in the field-of-view presented in panel F. **(H)** Fraction of spots above the cell baseline in the nuclear segment. The black dots indicate the fraction obtained in each field-of-view, and the blue bar indicates the mean. **(I)** Fraction of spots within the 2D boundary of the nucleus. The sample size is the same as in panel D.

What came intriguingly was the intracellular distribution of mRNAs of the four segments. The intracellular distribution of mRNAs projected on the XY plane (Fig. 1C) indicated that the HA mRNAs were more abundant in the nucleus compared to the other three segments (Fig. 1E).

We wished to quantify the nuclear ratio of mRNAs in each cell. The mRNA count within the 2D nuclear projection is considered to be a good approximation for the number of mRNAs contained in the nucleus [20, 21]. We validated this approximation for the lung epithelial A549 cells used in this study by looking at the intracellular Z positions of mRNAs: The mRNAs within the nuclear projection would have higher Z positions over the cell baseline, considering that the nucleus has a diameter of several microns, much thicker than the cytoplasm in the cell periphery. To this end, mRNAs were segmented into three regions in the 2D projection, that is, nuclear, perinuclear and peripheral, depending on the relative location to the nucleus (Fig. 1F). In contrast to the perinuclear (orange) and peripheral (green) mRNAs, mRNAs within the nuclear projection (blue) were distributed around the nuclear focal plane Z = 0 (Fig. 1G); and about 80% of mRNAs in the nuclear projection (blue) were above the cell baseline (Fig. 1H), defined by the median of the Z distribution of peripheral mRNAs (Fig. 1G, red dashed line). Thus, in this paper, we consider the spots within the nuclear 2D projection to be the nuclear fraction for the sake of simplicity.

With this approximation, about 30% of the PB1, NP and NS mRNAs were nuclear, while ~70% of HA mRNAs were in the nuclear fraction (Fig. 1I), confirming the visual inspection that the HA mRNAs are more abundant in the nucleus at the early time

point of infection (Fig. 1E). This result indicates that the velocity of viral mRNA export from the nucleus varies according to the viral segments. Note that our fluorescent probes could also detect complementary RNA (cRNA), an intermediate RNA that is produced for the viral genome replication, as it has the same polarity as viral mRNA. However, the fluorescent signals detected at 2 hours post-infection are deemed to be from the mRNAs, because cRNA production occurs much later than the mRNA synthesis [13, 22].

Mapping eight viral mRNAs in single cells by multiplex RNA FISH.

Next, we wish to simultaneously quantify all the eight viral segments *in situ*. To this end, we applied MERFISH (multiplex error-robust RNA FISH) [23], an imaging-based transcriptome assay in which multiple rounds of probe hybridisation and imaging are performed to identify hundreds to thousands of mRNA species *in situ* by decoding binary fluorescent signals. For example, hundreds of distinct mRNAs can be encoded in 16-bit binary signals, in which case eight rounds of hybridisation and imaging would be required to read out the signals with two fluorophores [23]. In our case of detecting eight influenza A virus segments, 6 bits were sufficient to accommodate all the eight mRNAs, with each code being separated by two or more binary flips from others (Fig. 2A). Thus, only three rounds of hybridisation and two-colour imaging were required. We sought for an economical way of performing sequential rounds of hybridisation and imaging, instead of setting up a custom-built, automated liquid pumping system used elsewhere [23]. Inspired by an *in vitro* reconstitution assay of molecular motors [24], we assembled a flow cell on a slide glass, comprising two stripes of double-sided tape and the coverslip, with cells on the coverslip facing inwards (Fig. 2A). Liquids were introduced from the right open side and wicked from the left. During image acquisition, the two open sides were

Figure 2

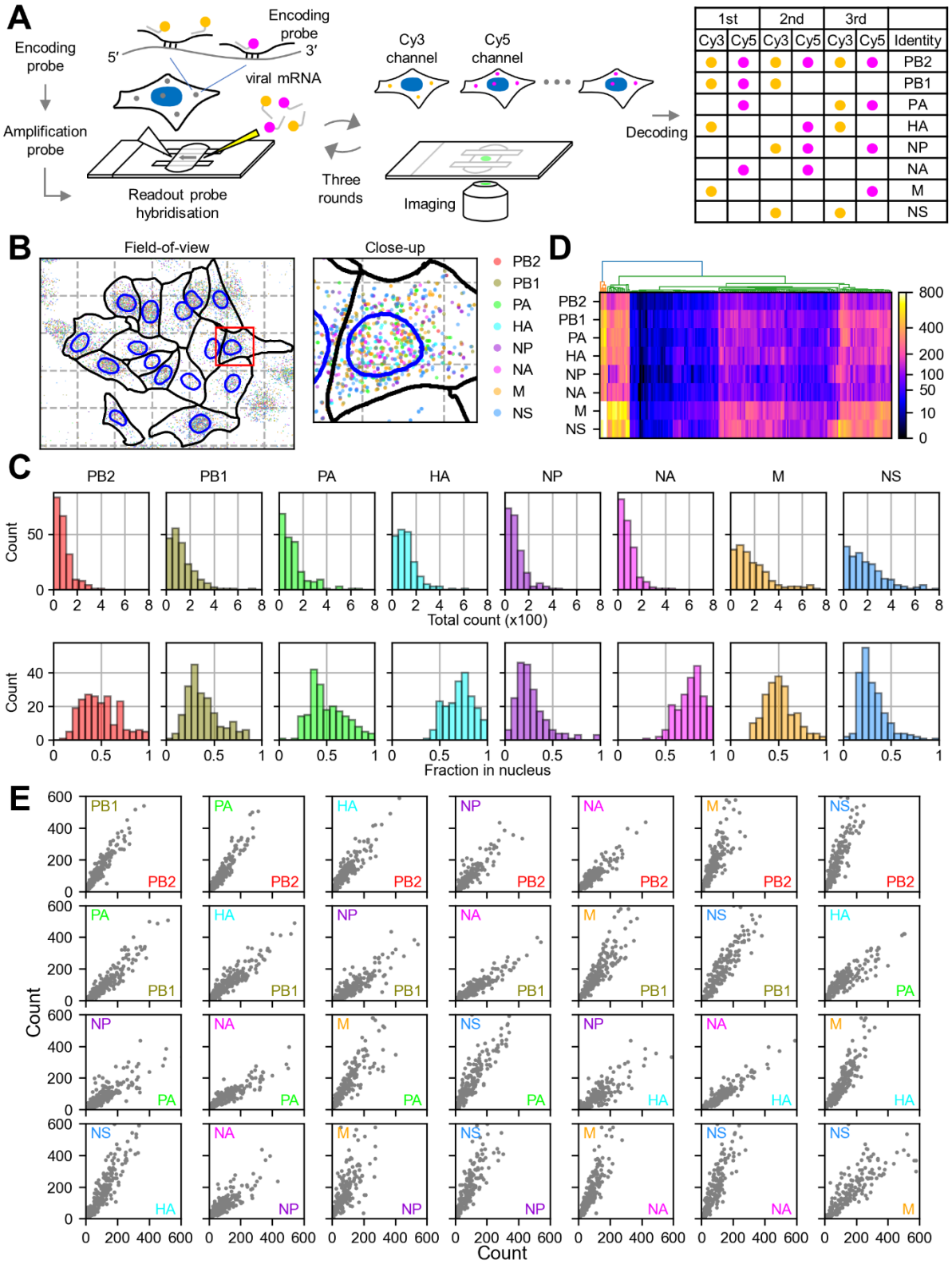


Fig. 2 Viral mRNA detection using multiplex RNA FISH. (A) An overview of the assay. The viral mRNAs were hybridised with the encoding probes, and then the amplification probes were annealed (not shown in the diagram for simplicity). A flow cell was assembled in which the readout probes were hybridised and the cells were imaged in the three sequential rounds of hybridisation and imaging. Subsequently fluorescent signals were decoded according to the encodings listed on the right. **(B)** A representative mapping of identified viral mRNAs on the segmented cell and nuclear boundary in the entire field-of-view (left) and the close-up in the red square (right). The mesh grids indicate ~34-μm

interval on the left, and ~17-μm interval on the right. **(C)** Total count of mRNAs in 204 single cells (top row) and the nuclear fraction (bottom row) for each segment. **(D)** Heatmap representation of mRNA counts in single cells. The viral segments are in rows, cells in columns. The cells were clustered column-wise to highlight the inequality in the abundance of transcripts carried per cell. The colour code for the absolute count is indicated in the vertical bar on the right. **(E)** A pair-wise analysis of the transcript count between segments.

sealed with rubber cement. This easy implementation of flow cell allowed us to comfortably perform three rounds of hybridisation and two-colour imaging to read out the 6-bit binary tags; and by decoding the 6-bit encodings, the eight viral segments were identified in each cell (Fig. 2B). On average, about 100-200 molecules of each segment were contained per cell (Fig. 2C, top row). The difference in the total number of spots from the result presented in Fig. 1D is likely due to the batch effect (experimental variation).

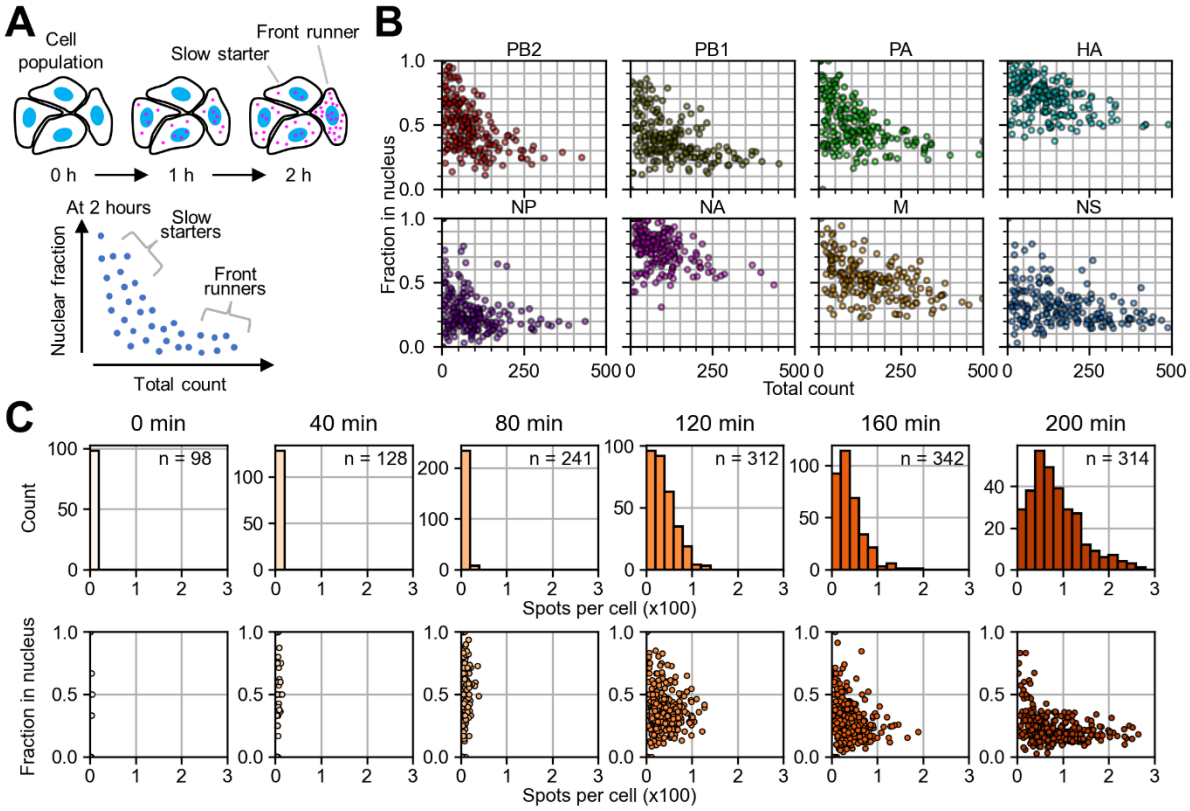
The nuclear fraction of mRNAs varied between the viral segments (Fig. 2C, bottom row): While ~30% of mRNAs were nuclear for PB1, NP and NS segments, about 70% of mRNAs remained in the nucleus for HA and NA segments; for the other segments (PB2, PA and M), around half of mRNAs were in the nucleus. This reiterates the idea that the velocity of viral mRNA export from the nucleus varies according to the viral segments. These results were confirmed by smFISH using probes for each segment that were synthesised economically in-house (Supplementary figure 1A, B and C).

Single-cell clustering according to the abundance of eight viral segments revealed that the cell population was heterogeneous in terms of the abundance of viral mRNAs (Fig. 2D). A pair-wise analysis of the abundance of viral segments indicated inequal distribution of viral mRNAs, regardless of the segments (Fig. 2E): Cells that carry abundant mRNAs from one segment tend to also carry abundant mRNAs from the other segments. These observations suggest that the viral transcription is susceptible to the stochastic variation in each infection [25, 26], which we subsequently used in our statistical model to estimate the nuclear export rate.

221 **Estimating the nuclear export rate using population statistics.**

222 We suspected that the higher nuclear ratio of HA and NA mRNAs is due to the
223 slower nuclear export rate for these mRNAs compared to other segments such as
224 NP and NS. Thus, we wished to quantify the nuclear export rate of eight viral mRNA
225 species from the observed nuclear-to-cytoplasmic distribution. To this end, we
226 conceptualised a statistical model using the cell-to-cell variation in the abundance
227 of viral transcripts of each segment (Fig. 3A). In this model, influenza viral particles
228 are added to the cell population at time 0 h, and are allowed to undergo viral
229 attachment to the cell surface, membrane fusion, and vRNP migration to the

230 **Figure 3**



231
232 **Fig. 3 Cell-to-cell heterogeneity for time-course inference. (A)** A theoretical diagram of how viral
233 mRNAs emerge in single cells in the population. The font runners designate ones in which the viral
234 transcription was initiated earlier than others, and the slow starters are ones that started transcription
235 lately. The bottom plot shows the theoretical relationship between the nuclear ratio and the
236 abundance of transcripts in which the front runners and slow starters (and many more along the
237 spectrum in between) compose the clusters as indicated. **(B)** Scatter plots of nuclear fraction against
238 the total count in each cell, derived from the dataset presented in Fig. 2C. **(C)** Time-course of the
239 distribution of NP mRNA count per cell, measured by smFISH, at the indicated time points (top row);
240 and the change in the nuclear fraction size, plotted against the total number of transcripts (bottom
241 row). The sample size (n) indicates the number of cells observed at each time point.

242 nucleus. The waiting time for these processes vary in each infection due to the
243 stochastic nature of biochemical events [25, 27], and therefore, the onset of viral
244 transcription varies from cell to cell in the population. In other words, each cell will
245 have a varying duration for viral transcription by the time the observation is made
246 at time 2 h (Fig. 3A, top). As a result, the population contains front runners in
247 which the viral transcription was initiated early, and therefore abundant viral
248 transcripts are carried; slow starters in which the transcription began late,
249 containing fewer transcripts; and many other cells that lie along the spectrum
250 between the two. The cell population at the single point in time, therefore, forms a
251 trajectory of the time development of viral transcription and subsequent nuclear
252 export on the scale of nuclear fraction against the abundance of transcripts (Fig. 3A,
253 bottom): In a spontaneous transcription burst, the nascent transcripts emerge in
254 the nucleus at the early stage of the burst, and they are exported to the cytoplasm
255 as more transcripts are being synthesised [20].

256 Indeed, when the nuclear ratio of viral mRNA was plotted against the total quantity
257 of mRNA molecules observed by multiplex RNA FISH in each cell, the plots formed a
258 curve that originated from top left (i.e., nearly 100% of the nuclear ratio when a cell
259 contained a few mRNA molecules), and descended as the number of mRNA
260 molecules increased (30% to 70%, depending on the segments) (Fig. 3B and
261 Supplementary figure 2).

262 We further validated this concept by performing smFISH measurements along the
263 time-course with 40-min interval (Fig. 3C). A549 was infected with WSN at the M.O.I
264 of ~0.02. At each time point, the NP mRNAs were stained and quantified. At the
265 beginning (0 and 40 min), the mRNA count was almost zero. As the mRNA count
266 increased (80 min to 200 min post-infection) (Fig. 3C, top), the nuclear ratio against
267 the total mRNA count derived from the single cells progressed along a curve, from

the top left to the bottom right (Fig. 3C, bottom), supporting the concept of our statistical model.

The curvature of the trace appeared to vary between segments (Fig. 3B). We explored to see if the curvature could represent a proxy for the nuclear export rate. Thus, we formulated a simple kinetic model in which mRNAs are synthesised at the rate k in the nucleus, and are exported with the reaction constant ε (Fig. 4A). In this model, the molar concentrations of mRNAs in the nuclear and cytoplasmic compartments are described, respectively,

Figure 4

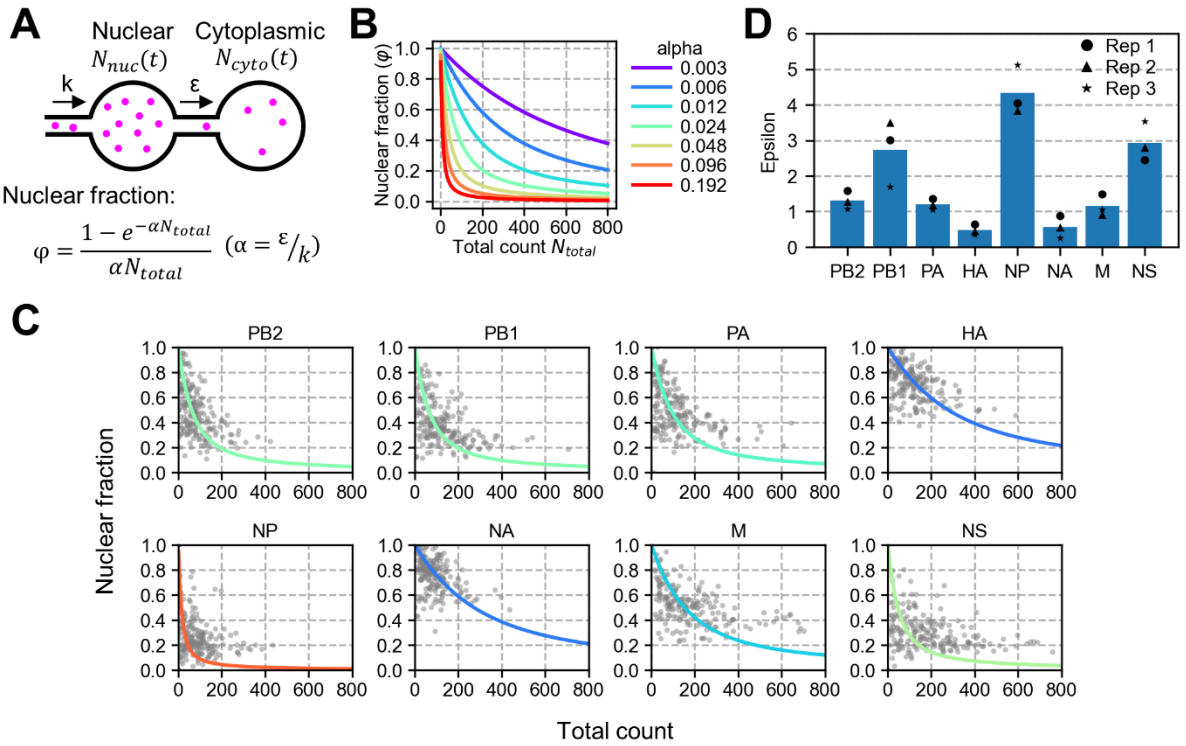


Fig. 4 Kinetic model for estimating the nuclear export rate. (A) A schematic diagram representing the production of viral transcripts in the nucleus and the export to the cytoplasm. The two spherical compartments are for the nucleus and the cytoplasm. Viral transcripts (magenta) are produced at a rate constant k in the nucleus, and exported at a rate constant ε to the cytoplasm. **(B)** The relationship between the nuclear fraction and the total count in the kinetic model, according to the formula presented in panel A. The parameter α determines the curvature of the plot. **(C)** The parameter fit to the observed datapoints. The parameter α was estimated by fitting the equation to the observed datapoints presented in Fig. 3B. The colour representation in each line is the same as the ones used in panel B, according to the parameter α. **(D)** Estimated nuclear export rate. The parameter ε was estimated from the three independent assays (indicated in black for each segment); the blue bar indicates the mean of the three measurements.

$$\frac{d[Nuclear]}{dt} = k - \varepsilon[Nuclear] \quad (1)$$

$$\frac{d[Cytoplasmic]}{dt} = \varepsilon[Nuclear] \quad (2)$$

Omitting the degradation of mRNAs, the analytical solutions of these equations were:

$$[Nuclear] = \frac{k}{\varepsilon}(1 - e^{-\varepsilon t}) \quad (3)$$

$$[Cytoplasmic] = kt - \frac{k}{\varepsilon}(1 - e^{-\varepsilon t}) \quad (4)$$

Thus, the mRNA nuclear fraction φ was described as a function of total mRNA molecules N_{total},

$$Nuclear\ fraction\ \varphi = \frac{1}{\alpha N_{total}}(1 - e^{-\alpha N_{total}}) \quad (5)$$

where

$$\alpha = \frac{\varepsilon}{k} \quad (6)$$

The nuclear fraction φ described in equation (5) is time-independent, and the parameter α determines the curvature of the line that the single cells would trace in the time-course (Fig. 4B). Thus, we estimated the parameter α by fitting to the observed data points derived at the time in which cells were fixed (Fig. 4C). By obtaining the synthesis rate k from the total number of mRNAs produced by the time of fixation (Fig. 2C, top row), the nuclear export rate ε for each segment was derived according to equation (6) (Fig. 4D). The nuclear export rate ε agreed with the historical classification of the early and late influenza A virus genes [4, 6], indicating that the timing of influenza A virus protein expression in the initial stage of infection is determined by the velocity of the nuclear export. Notably, the HA and NA mRNAs had the slowest nuclear export rate, suggesting that influenza A virus delays the viral antigen production by nuclear retention of viral transcripts.

Discussion

Based on the stochastic nature of influenza viral transcription studied by single-cell sequencing [25, 26], we introduced a quantitative framework for estimating the nuclear export rate of viral mRNAs to advanced our current understanding on the kinetics of viral gene expression. This was achieved by *in situ* localisation of eight influenza A virus segments simultaneously, coupled with the statistical model that allowed for estimating the nuclear export rate from the intracellular distribution of mRNAs at a single time point. This is based on the concept that the cells in the population are at various stages of transcription at the time the observation is made, due to the multiple stochastic processes the virus needs to overcome before it begins transcription. We demonstrated that the initiation of each viral transcription is concurrent in the eight viral segments; it is instead the nuclear export rate that controls the temporal protein expression at the initial phase of infection. The early production of the polymerase subunits (PB2, PB1 and PA) and nucleoprotein would facilitate the formation of vRNP, and hence the viral transcription and genome replication. On the other hand, pushing back the expression of viral antigens HA and NA would be beneficial for the virus to delay the host immune response against the infected cells in which the virus is being replicated.

Instead of time-series measurements, the estimation of mRNA export rate was made from a snapshot of *in situ* RNA localisation data taken at a single time point. We anticipate that the statistical framework for quantifying the nuclear export rate proposed in our study can be widely used and benefits the research area on the regulation of gene expression, especially since nuclear retention has been recognised as a point of control against the transcription burst [14]. For example, it is applicable for the cell population in which *de novo* synthesis of mRNAs occurs concurrently, such as a stimuli-induced activation of immediate-early genes [21, 28]. On

339 the contrary, this approach is not appropriate for scenarios in which asynchronous
340 transcription bursts are taking place randomly in each single cell [20, 29]; cells at the
341 late stage of transcription burst, having residual amount of transcripts in the
342 cytoplasm [20], would obscure the time-course trajectory of the nuclear export.

343 The experiment in this study was conducted at the early stage of infection (up to
344 200 min post-infection). This conferred two advantages over investigating
345 influenza viral mRNA export at the late stage of infection (e.g., 6-12 hours post-
346 infection). First, our study allowed for the absolute quantification of viral
347 transcripts at the single-molecule level in each cell, and thereby enabled us to
348 formulate the model and fit the equation to the numerical data points. Otherwise
349 at the late stage of infection, accurate quantification of mRNAs by the imaging-
350 based single-molecule detection would not be feasible due to the high density of
351 mRNAs. Second, at the late stage of infection, NS1 protein would be expressed in
352 the cytoplasm. NS1 protein would then migrate into the nucleus and alter the
353 nuclear export of viral mRNAs [30], obscuring the inherent nature of nuclear export of
354 each viral segment. Thus, in our study, we elucidated the quantitative measure for
355 the influenza viral mRNA nuclear export that is intrinsic to each viral segment.

356 Host factors that relay the influenza viral mRNAs to the nuclear pore remain elusive.
357 NXF1 and CRM are two major cellular mRNA export pathways for the 5′-capped
358 mRNAs [31]. Of these two pathways, influenza A virus mRNAs, also carrying 5′ cap
359 structure derived from the host nascent mRNA and the poly-A tail as a result of
360 viral polymerase stuttering the uridine repeat, are dependent on the NXF1 pathway
361 [16, 17]. However, the eight viral segments show differential dependencies on this
362 export pathway. The segments HA, M and NS is exported by NXF1 and UAP56,
363 whereas PB2, PB1 and PA appear to be independent of the NXF1 pathway [17]. The
364 nuclear export rate derived in our study does not agree with the dependencies on

365 NXF1. According to the previous study, both HA and NS are dependent on NXF1
366 and UAP56 [16]; however, in our study, HA and NS had significantly different nuclear
367 export rate.

368 Two viral segments M and NS are subject to splicing, yielding unspliced and spliced
369 transcripts. Our current system lacks the ability of discriminating between the
370 unspliced and spliced forms of these segments; although we infer that the majority
371 of NS segments that we detected in this study were in the unspliced form because
372 we obtained similar result with probes exclusively detecting the unspliced form of
373 NS (Supplementary figure 3). An observation that the unspliced form of M and NS
374 dominates throughout the infection supports this [13]. In either case, the RNA
375 structure does not appear to accord with the NXF1 dependency, nor the nuclear
376 export rate, either.

377 Our study suggests that the nuclear export of influenza viral transcripts is intrinsic
378 to the nucleotide sequence of each segment. Recently, the human nuclear factor
379 hnRNPAB was shown to associate with the nucleoprotein of influenza A virus, and
380 inhibit the NXF1 pathway. It is tempting to speculate that the host factors (e.g.,
381 hnRNPAB) have differential binding affinities to each segment, thus determining the
382 velocity of nuclear export (or the degree of nuclear export inhibition). It is currently
383 unknown what factors binds to the early genes such as PB2, PB1 and PA. Further
384 studies will be required to identify the host factors that bind to these transcripts
385 for the nuclear export. Our study serves as a reference for the nuclear export rate
386 of each segment, and thus sheds light to comprehensive understanding on the
387 nuclear export of viral segments. Having revealed the nuclear retention of
388 influenza viral transcripts at the unprecedented resolution, this study opens up a
389 new research direction on how the nuclear retention is regulated and its implication
390 in viral pathogenesis.

Materials and Methods

Cells and virus preparation

The human lung carcinoma cell line A549 was maintained in Dulbecco's Modified Eagle's Medium (DMEM) (Sigma, D5796) supplemented with 10% FBS (Nichirei Biosciences, 175012), 100 U/ml penicillin and 100 μg/ml streptomycin (Gibco, 15140-122) at 37°C in 5% CO_2. Madin-Darby canine kidney (MDCK) cell line was maintained in Minimum Essential Medium (MEM) (Sigma, M4655) supplemented with 10% FBS, 100 U/ml penicillin and 100 μg/ml streptomycin at 37°C in 5% CO_2. Influenza A virus strain WSN was propagated in MDCK cells in a 75 cm^2 flask. The plaque-forming unit of the harvested virus was determined using MDCK cells grown confluent in a 6-well plate.

Coverslip preparation

An 18-mm round coverslip (0.17±0.005 mm thickness, Matsunami Glass) was used. For MERFISH, two straight sides were created by cutting two edges off the glass using a laser-wheel glass cutter (Nikken Dia, NC-X03 LASER 110°). The coverslip was rinsed in 70% ethanol and placed in a 12-well plate. A549 cells were seeded and attached onto the coverslip overnight. On the following day cells were rinsed twice with PBS and infected with WSN at the indicated multiplicity of infection in MEM for 2 hours (unless indicated otherwise in Fig. 3C) in the absence of FBS. After the infection, cells were fixed with 4% paraformaldehyde (Electron Microscopy Sciences, 15714-S) in PBS for 10 min and then rinsed three times with PBS. The coverslip was immersed in 70-75% ethanol and stored at −20°C until use.

416 **smFISH**

417 The probes used in Figs. 1, 3C, Supplementary figure 1 (probes for NP) and
418 Supplementary figure 3 were obtained from LGC Biosearch Technologies. Stellaris
419 RNA FISH Probe Designer (https://www. biosearchtech.com/stellaris-designer) was
420 used to design the probes against the coding region of WSN mRNAs.

421 The coverslip was equilibrated in 1 ml of Wash buffer A (Stellaris) for >5 min in a
422 12-well plate prior to hybridisation. Cells were hybridised to 125 nM probes in
423 Hybridization buffer (Stellaris) on a Hybrislip (Sigma, GBL712222) at 37°C overnight
424 in a humidified chamber. On the following day the coverslip was rinsed twice in 1
425 ml of Wash buffer A (Stellaris), each for 30 min. Cytoplasmic stain (HCS Cellmask
426 Green Stain, Invitrogen, H32714) was included in the second wash at the
427 concentration of 50 ng/ml in Figs 1, 3C and Supplementary figure 3. Then the
428 coverslip was rinsed in 1 ml of Wash buffer B (Stellaris) for 5 min. The coverslip
429 was mounted in anti-fade medium containing DAPI (Vector Laboratories, H-1200) on
430 a slide glass and sealed with nail polish.

431 For an economical reason and to allow for more flexible design of probes to
432 increase the number of fluorophores attached per mRNA molecule, the probes used
433 in Supplementary figure 1 (except the NP probes) were synthesised in-house
434 (Supplementary figure 1A). (The NP probe used in Supplementary figure 1 was
435 purchased from LGC Biosearch.) Candidate probe-binding sites in the coding region
436 of each viral segment were identified using OligoMiner [32], and were further selected
437 against the human RefSeq. The sequences were flanked by P5 (5′ AAT GAT ACG
438 GCG ACC ACC GA 3′) and P7 (5′ CAA GCA GAA GAC GGC ATA CGA GAT 3′), with P7
439 at the 5′ end and P5 at the 3′ end of the probe sequence (Supplementary figure 1A).
440 For the probes against HA, NA, M and NS, additional 20-nt sequences RS0332 (5′
441 GGG AGA ATG AGG TGT AAT GT 3′) and RS0406 (5′ GAT GAT GTA GTA GTA AGG

442 GT 3′) [23] were inserted between P7 and the probe, the probe and P5, respectively
443 (Supplementary figure 1A). These two insertions were used for additional probes to
444 bind to increase the number of fluorophores per molecule. An oligonucleotide pool
445 comprising the set of probe-binding sites with these additional nucleotide
446 sequences was synthesised at Integrated DNA Technologies (IDT). The
447 oligonucleotide pool was subjected to PCR amplification (98°C for 30 sec; 18 cycles
448 of 98°C for 5 sec, 64°C for 10 sec and 72°C for 20 sec; and 72°C for 5 min) (Phusion
449 High-Fidelity DNA Polymerase, NEB, M0530S) using the P7 and a T7-fused P5 primer
450 (5′ TAA TAC GAC TCA CTA TAG GGA ATG ATA CGG CGA CCA CCG A 3′)
451 (Supplementary figure 1A). RNA was synthesised using T7 RNA polymerase
452 (HiScribe T7 High Yield RNA Synthesis Kit, NEB, E2040S). Then RNA was reverse-
453 transcribed using a recombinant Moloney leukemia virus reverse transcriptase
454 (GeneAce cDNA Synthesis Kit, Nippon Gene, 319-08881) with an ATTO550-
455 conjugated P7 primer at the 5′ end. The single-stranded DNA probe was purified
456 using a spin column (Zymo Research, D7010) after RNA digestion with a cocktail of
457 RNase H and RNase A. The probe was further purified by Urea-PAGE to remove any
458 residual RNA and the fluorescent-labelled primer.

459 The coverslip was equilibrated in 1 ml of 30% formamide buffer composed of 2×SSC
460 and 30% deionised formamide (Sigma, F9037) in a 12-well plate for >5 min prior to
461 hybridisation. Cells were hybridised with the probes at the concentration of 5.6 nM
462 per probe in 100 µl Hybridization buffer composed of 2×SSC, 30% deionised
463 formamide, 0.1% yeast tRNA (Invitrogen, 15401-011), 1% (v/v) RNase inhibitor (NEB,
464 M0314S) and 10% (w/v) Dextran Sulfate (Merck, S4030) [23] on a Hybrislip at 37°C
465 overnight in a humidified chamber. On the following day the coverslip was rinsed
466 twice in 1 ml of 30% formamide buffer, each for 30 min. DAPI was included at 5
467 ng/ml in the second wash if additional staining was not performed with the
468 external probes (for PB2, PB1 and PA). In this case the coverslip was rinsed in 1 ml

469 of PBS after the DAPI staining, mounted in anti-fade medium (Vector Laboratories,
470 H-1900) on a slide glass and sealed with nail polish. For staining HA, NA, M and NS
471 with the extra probes, the coverslip was first equilibrated with 10% formamide
472 buffer composed of 2×SSC, 10% deionised formamide and 0.05% (v/v) RNase
473 inhibitor, and then hybridised with 10 nM each of probes for RS0332 and RS0406 [23],
474 both labelled with Quasar 570 (LGC Biosearch) at the 3′ end, in Stellaris
475 hybridisation buffer (containing 10% formamide) at 37°C in the humidified chamber
476 for 1 hour. Cells were washed in 10% formamide buffer twice, with DAPI included
477 in the second wash. The coverslip was rinsed in PBS, mounted in the anti-fade
478 medium (Vector Laboratories, H-1900), and sealed with nail polish.

479

480 **Sequential hybridisation and imaging in the flow cell**

481 An oligonucleotide pool for synthesising the encoding probes used in Fig. 2,
482 comprising the target-binding sequence, readout tags compatible with the two-step
483 amplification probes [23, 33], and the P5 and P7 PCR tags, were obtained from IDT. The
484 target-binding sites were identified in the coding region of each segment using
485 OligoMiner [32] and selected against human RefSeq. The readout tags RS0332, RS0343,
486 RS0406, RS0255, RS0015 and RS038 [23, 33] were inserted adjacent to the upstream or
487 downstream of the target-binding sites for the 6-bit encoding of each segment (Fig.
488 2A). Probes were synthesised by PCR, T7 transcription and reverse-transcription as
489 described above.

490 A coverslip was equilibrated in 1 ml of 30% formamide buffer (2×SSC and 30%
491 deionised formamide) [23]. Cells were hybridised with 1.6 μM encoding probes in
492 Hybridisation buffer composed of 2×SSC, 30% deionised formamide, 0.1% yeast
493 tRNA, 1% RNase inhibitor, 10% dextran sulfate [23] on a Hybrislip at 37°C overnight in
494 a humidified chamber. Cells were washed twice in 1 ml of 30% formamide buffer at

495 37°C, and then equilibrated in 1 ml of Readout-probe hybridisation and wash buffer
496 composed of 2×SSC, 10% deionised formamide and 20 units/ml Murine RNase
497 inhibitor [23]. The first and second amplification probes for RS0332, RS0255, RS0343,
498 RS0406, RS0015 and RS0384, described previously [33], were annealed sequentially at
499 10 nM of each probe in Readout-probe hybridisation and wash buffer for 2 hours at
500 37°C on a Hybrislip. The coverslip was washed for 30 min twice in Readout-probe
501 hybridisation and wash buffer at each of the first and second probe hybridisation.
502 At the final wash, 0.10 µm green-fluorescent carboxyl polystyrene beads (Bangs
503 Laboratories, FCDG002) were included (~0.1% v/v) as a fiducial marker for image
504 registration. The cells and beads were fixed with 3.2% paraformaldehyde in PBS for
505 10 min. The coverslip was rinsed in PBS three times and stored in 2×SSC
506 supplemented with 40 units/ml RNase inhibitor at 4°C overnight.

507 On the following day a flow cell was assembled on a slide glass, composed of two
508 parallel stripes of double-sided tape (0.23 mm thickness, ~5 mm apart) and the
509 coverslip on top, with cells facing inwards (Fig. 2A). The cells at the top and bottom
510 end of the coverslip were scraped using a cell scraper before the assembly and any
511 trace amount of liquid remaining on the exposed surface was removed with filter
512 paper to ensure that the coverslip is firmly attached to the tape. The chamber was
513 immediately flushed with 2×SSC to avoid cells drying up.

514 Readout-probe hybridisation and wash buffer was flushed in the chamber to
515 equilibrate the cells. Readout probes were introduced at 10 nM in Readout-probe
516 hybridisation and wash buffer and incubated at 37°C in a humidified chamber for
517 30 min. Subsequently, Readout-probe hybridisation and wash buffer was flushed to
518 wash away the probes. Imaging buffer, composed of 2×SSC, 50 mM Tris-HCl pH 8,
519 10% glucose (Fujifilm Wako, 049-31165), 2 mM Trolox (Sigma, 238813), 0.5 mg/ml
520 glucose oxidase (Sigma, G2133), 40 µg/ml catalase (Sigma, C30) and 0.1% RNase

521 inhibitor [23], was flushed to the flow cell prior to the image acquisition. The flow cell
522 was sealed at the both sides with Fixogum rubber cement ejected through a 200-µl
523 pipette tip during the image acquisition.

524 After the imaging, the disulfide bond linking the probe and the fluorophore was
525 cleaved with Cleavage buffer comprising 2×SSC and 50 mM TCEP (Sigma, 646547)
526 for 15 min. The chamber was flushed with Readout-probe hybridisation and wash
527 buffer to wash away the free fluorophores prior to introducing the readout probes
528 for the next round. Readout probes used in this study are listed in Table 1.

529 **Table 1**

ID	Round	Channel	Sequence (5′ to 3′) and fluorescent dye	Manufacturer
RS0332amp	1	1	TTC TTC CCT CAA TCT TCA TC -S-S- Q570	LGC Biosearch
RS0343amp	1	2	TAC TAC AAA CCC ATA ATC CC -S-S- Q670	LGC Biosearch
RS0406amp	2	1	AAT CTC ACC TTC CAC TTC AC -S-S Q570	LGC Biosearch
RS0255amp	2	2	TCA CCT CTA ACT CAT TAC CT -S-S- Q670	LGC Biosearch
RS0015amp	3	1	TCT CAC ACC ACT TTC CTC AT -S-S- ATTO550	Generay
RS0384amp	3	2	TCC TCA TCT TAC TCC CTC TA -S-S- ATTO647N	Generay

530

531 **Imaging platform**

532 Images were acquired using a Nikon Ti2-E with a Nikon Plan Apo Lambda 60x (NA
533 1.40) (Fig. 1, Supplementary figure 3) or 100x (NA 1.45) (Fig. 3C) objective lens
534 equipped with a Prime 95B 25MM back-illuminated sCMOS camera (Photometrics).
535 Otherwise, an Olympus IX83 equipped with an Olympus UPlan XApo 60x (NA 1.42)
536 objective lens and a cooled monochrome CCD camera (DP80, Olympus) was used
537 (Fig. 2 and Supplementary figure 1).

538

539 **Spot quantification**

540 Fluorescent spots detected by smFISH were quantified using either FISH-QUANT [34]
541 (Figs. 1, 3C and Supplementary figure 3) or a more recent Python package Big-FISH [35]

(Supplementary figure 1). Cytoplasmic and nuclear segmentation was performed using Cellpose [36, 37]. Objects that were on the boundary of the image was eliminated. Cytoplasmic regions with no nucleus therein due to the erroneous segmentation was also eliminated. The identification numbers assigned to each object by Cellpose were rearranged from scratch so that the cytoplasm and nucleus of the same cell correspond for the downstream analyses. In order to make the Cellpose mask readable in FISH-QUANT, the Cellpose mask were converted into x and y coordinates by tracing each object in the mask using the Moore neighbourhood edge-finding algorithm.

For MERFISH signal decoding, signals from the green-fluorescent carboxyl beads were used to register the images. Images at the corresponding field-of-view from the three rounds of imaging were registered using phase_cross_correlation from the Python package skimage.registration. The registered images were filtered using Laplacian-of-Gaussian implemented in Big-FISH, and the image intensity was manually normalised. Fluorescent signals were then decoded using PixelSpotDecoder in DetectPixels implemented in starfish (https://github.com/spacetx/starfish).

Data analyses and presentation

Microscopy images were presented using the Python package microfilm (https://github.com/guiwitz/microfilm) (Fig. 1B). Cell clustering was performed with the method centroid implemented in scipy.cluster.hierarchy.linkage from SciPy (https://scipy.org/) (Fig. 2D). The parameter α in equation (5) was estimated using curve_fit implemented in scipy.optimize from SciPy (Fig. 4C).

Acknowledgements

We thank Imaging and Flow Cytometry Core, The Centre for PanorOmic Sciences (CPOS), LKS Faculty of Medicine of The University of Hong Kong for the use of Nikon Ti2-E widefield imaging system; and Central Research Institute of Kawasaki Medical School for the use of Olympus IX83 imaging system. This study was supported by Research Project Grant R04S-003 and R03S-001 from Kawasaki Medical School, The KAWASAKI Foundation of Medical Science and Medical Welfare, and Grant-in-Aid for Scientific Research (C) (22K07091) from Japan Society for the Promotion of Science (JSPS) to MM; Grant-in-Aid for Scientific Research (B) (21H03188) from JSPS to TN; the General Research Fund (17107019) of the Research Grants Council, Hong Kong SAR to HC.

Author contributions

Michi Miura: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization, Supervision, Project administration, Funding acquisition. Naho Kiuchi: Software, Validation, Investigation, Writing - Review & Editing. Siu-Ying Lau: Resources, Writing - Review & Editing. Bobo Wing-Yee Mok: Resources, Writing - Review & Editing. Hiroshi Ushirogawa: Resources, Writing - Review & Editing. Tadasuke Naito: Resources, Writing - Review & Editing, Funding acquisition. Honglin Chen: Resources, Data Curation, Writing - Review & Editing, Supervision, Project administration, Funding acquisition. Mineki Saito: Resources, Data Curation, Writing - Review & Editing, Supervision, Project administration, Funding acquisition.

Declaration of interests

The authors declare no competing interests.

Data and code availability

The probe sequences, original images, and the custom codes used in this study are available upon request.

Reference

1. Bangham CRM, Miura M, Kulkarni A, Matsuoka M. Regulation of Latency in the Human T Cell Leukemia Virus, HTLV-1. *Annu Rev Virol* **6**, 365-385 (2019).

2. Einav T, Gentles LE, Bloom JD. SnapShot: Influenza by the Numbers. *Cell* **182**, 532-532 e531 (2020).

3. Dou D, Revol R, Ostbye H, Wang H, Daniels R. Influenza A Virus Cell Entry, Replication, Virion Assembly and Movement. *Front Immunol* **9**, 1581 (2018).

4. Lamb RA, Choppin PW. The gene structure and replication of influenza virus. *Annu Rev Biochem* **52**, 467-506 (1983).

5. Shapiro GI, Gurney T, Jr., Krug RM. Influenza virus gene expression: control mechanisms at early and late times of infection and nuclear-cytoplasmic transport of virus-specific RNAs. *J Virol* **61**, 764-773 (1987).

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

7. Te Velthuis AJ, Fodor E. Influenza virus RNA polymerase: insights into the mechanisms of viral RNA synthesis. *Nat Rev Microbiol* **14**, 479-493 (2016).

8. Fodor E, Pritlove DC, Brownlee GG. Characterization of the RNA-fork model of virion RNA in the initiation of transcription in influenza A virus. *J Virol* **69**, 4012-4019 (1995).

9. Arranz R, *et al.* The structure of native influenza virion ribonucleoproteins. *Science* **338**, 1634-1637 (2012).

10. Walker AP, Fodor E. Interplay between Influenza Virus and the Host RNA Polymerase II Transcriptional Machinery. *Trends Microbiol* **27**, 398-407 (2019).

11. Poon LL, Pritlove DC, Fodor E, Brownlee GG. Direct evidence that the poly(A) tail of influenza A virus mRNA is synthesized by reiterative copying of a U track in the virion RNA template. *J Virol* **73**, 3473-3476 (1999).

12. Enami M, Fukuda R, Ishihama A. Transcription and replication of eight RNA segments of influenza virus. *Virology* **142**, 68-77 (1985).

13. Phan T, Fay EJ, Lee Z, Aron S, Hu WS, Langlois RA. Segment-specific kinetics of mRNA, cRNA and vRNA accumulation during influenza infection. *J Virol* **95**, (2021).

641
642 14. Bahar Halpern K, *et al.* Nuclear Retention of mRNA in Mammalian Tissues.
643 *Cell Rep* **13**, 2653-2662 (2015).

644
645 15. Mor A, *et al.* Influenza virus mRNA trafficking through host nuclear speckles.
646 *Nat Microbiol* **1**, 16069 (2016).

647
648 16. Read EK, Digard P. Individual influenza A virus mRNAs show differential
649 dependence on cellular NXF1/TAP for their nuclear export. *J Gen Virol* **91**,
650 1290-1301 (2010).

651
652 17. Larsen S, Bui S, Perez V, Mohammad A, Medina-Ramirez H, Newcomb LL.
653 Influenza polymerase encoding mRNAs utilize atypical mRNA nuclear export.
654 *Virol J* **11**, 154 (2014).

655
656 18. Wang X, *et al.* Cellular hnRNPAB binding to viral nucleoprotein inhibits flu
657 virus replication by blocking nuclear export of viral mRNA. *iScience* **24**,
658 102160 (2021).

659
660 19. Raj A, van den Bogaard P, Rifkin SA, van Oudenaarden A, Tyagi S. Imaging
661 individual mRNA molecules using multiple singly labeled probes. *Nat*
662 *Methods* **5**, 877-879 (2008).

663
664 20. Billman MR, Rueda D, Bangham CRM. Single-cell heterogeneity and cell-cycle-
665 related viral gene bursts in the human leukaemia virus HTLV-1. *Wellcome*
666 *Open Res* **2**, 87 (2017).

667
668 21. Miura M, Dey S, Ramanayake S, Singh A, Rueda DS, Bangham CRM. Kinetics of
669 HTLV-1 reactivation from latency quantified by single-molecule RNA FISH
670 and stochastic modelling. *PLoS Pathog* **15**, e1008164 (2019).

671
672 22. Kawakami E, *et al.* Strand-specific real-time RT-PCR for distinguishing
673 influenza vRNA, cRNA, and mRNA. *J Virol Methods* **173**, 1-6 (2011).

674
675 23. Moffitt JR, Hao J, Wang G, Chen KH, Babcock HP, Zhuang X. High-throughput
676 single-cell gene-expression profiling with multiplexed error-robust
677 fluorescence in situ hybridization. *Proc Natl Acad Sci U S A* **113**, 11046-
678 11051 (2016).

679
680 24. Miura M, Matsubara A, Kobayashi T, Edamatsu M, Toyoshima YY. Nucleotide-
681 dependent behavior of single molecules of cytoplasmic dynein on
682 microtubules in vitro. *FEBS Lett* **584**, 2351-2355 (2010).

683
684 25. Heldt FS, Kupke SY, Dorl S, Reichl U, Frensing T. Single-cell analysis and
685 stochastic modelling unveil large cell-to-cell variability in influenza A virus
686 infection. *Nat Commun* **6**, 8938 (2015).

26. Russell AB, Trapnell C, Bloom JD. Extreme heterogeneity of influenza virus infection in single cells. *Elife* **7**, (2018).

27. Heldt FS, Frensing T, Reichl U. Modeling the intracellular dynamics of influenza virus replication to understand the control of viral RNA synthesis. *J Virol* **86**, 7806-7817 (2012).

28. Kulkarni A, Taylor GP, Klose RJ, Schofield CJ, Bangham CR. Histone H2A monoubiquitylation and p38-MAPKs regulate immediate-early gene-like reactivation of latent retrovirus HTLV-1. *JCI Insight* **3**, (2018).

29. Kiik H, Ramanayake S, Miura M, Tanaka Y, Melamed A, Bangham CRM. Time-course of host cell transcription during the HTLV-1 transcriptional burst. *PLoS Pathog* **18**, e1010387 (2022).

30. Pereira CF, Read EKC, Wise HM, Amorim MJ, Digard P. Influenza A Virus NS1 Protein Promotes Efficient Nuclear Export of Unspliced Viral M1 mRNA. *J Virol* **91**, (2017).

31. Palazzo AF, Lee ES. Sequence Determinants for Nuclear Retention and Cytoplasmic Export of mRNAs and lncRNAs. *Front Genet* **9**, 440 (2018).

32. Beliveau BJ*, et al.* OligoMiner provides a rapid, flexible environment for the design of genome-scale oligonucleotide in situ hybridization probes. *Proc Natl Acad Sci U S A* **115**, E2183-e2192 (2018).

33. Xia C, Babcock HP, Moffitt JR, Zhuang X. Multiplexed detection of RNA using MERFISH and branched DNA amplification. *Sci Rep* **9**, 7721 (2019).

34. Mueller F*, et al.* FISH-quant: automatic counting of transcripts in 3D FISH images. *Nat Methods* **10**, 277-278 (2013).

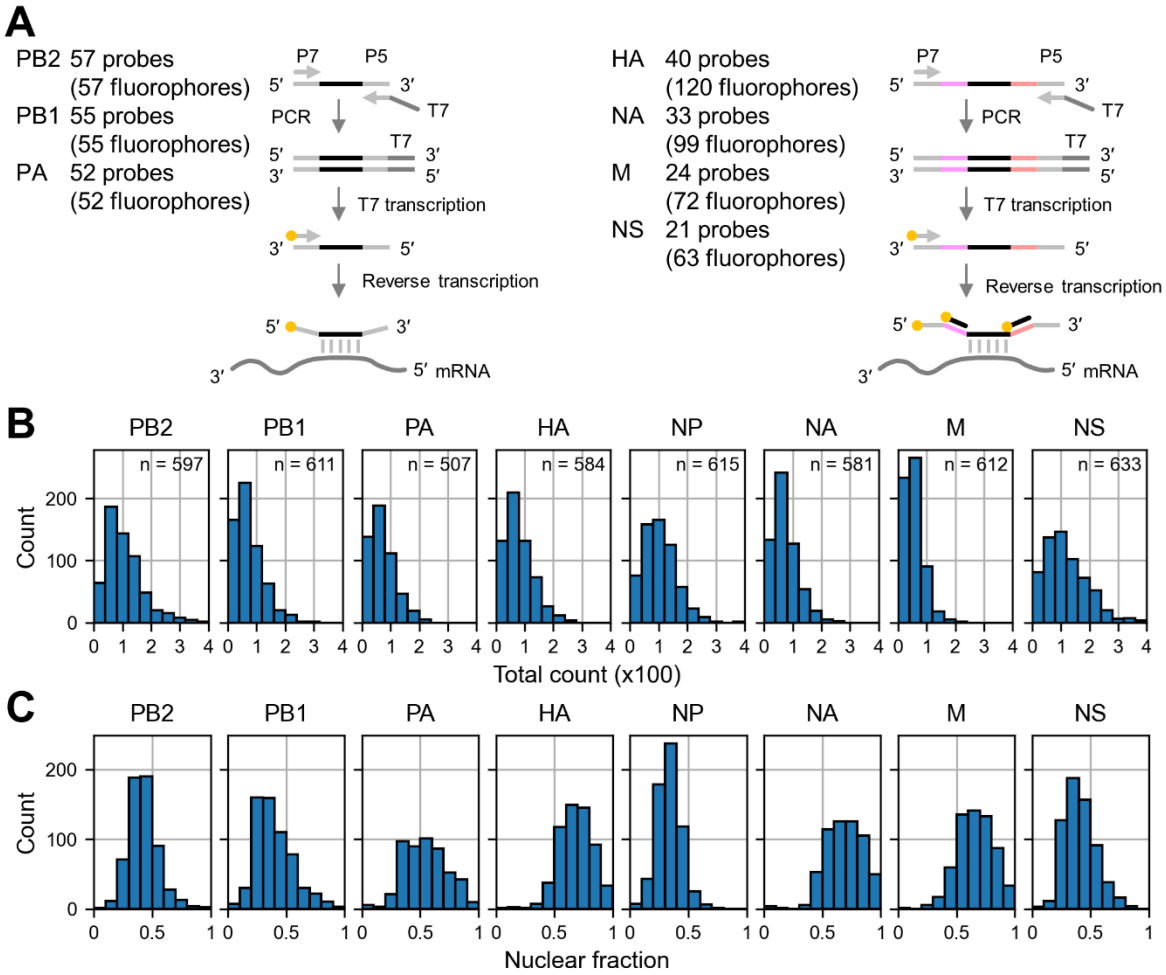
35. Imbert A*, et al.* FISH-quant v2: a scalable and modular tool for smFISH image analysis. *RNA* **28**, 786-795 (2022).

36. Stringer C, Wang T, Michaelos M, Pachitariu M. Cellpose: a generalist algorithm for cellular segmentation. *Nat Methods* **18**, 100-106 (2021).

37. Pachitariu M, Stringer C. Cellpose 2.0: how to train your own model. *Nat Methods* **19**, 1634-1641 (2022).

730 **Appendices**

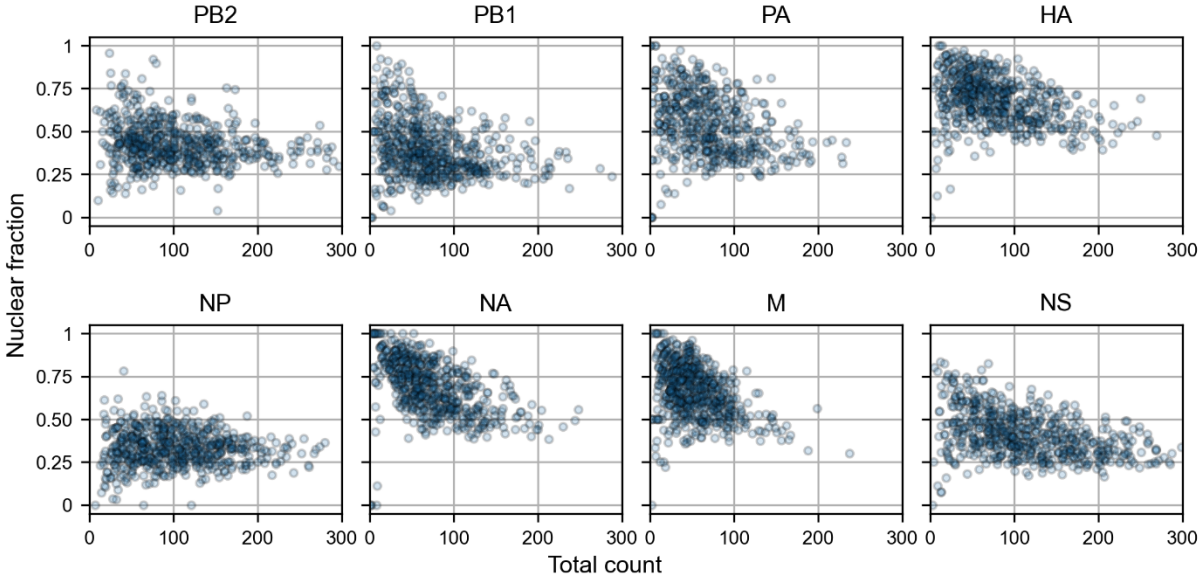
Supplementary figure 1: smFISH for eight segments



Supplementary figure 1 Independent validation of the total mRNA count and the nuclear ratio of each segment by smFISH. (A) A schematic diagram of in-house probe synthesis. The pool of single-stranded DNA templates were PCR-amplified, T7-transcribed, and reverse-transcribed using the fluorescent-labelled primer. The number of distinct probes contained in the pool is indicated for each segment. For HA, NA, M and NS segments, two additional fluorescent probes bind to each probe. As a result, the number of fluorophores bound per mRNA molecule is theoretically three times as many as the probes for those segments. **(B)** Total count of mRNAs per cell for each segment. The sample size (n) indicates the number of cells analysed. **(C)** The nuclear ratio of mRNAs in each cell. The sample size is the same as the one indicated in panel B.

731

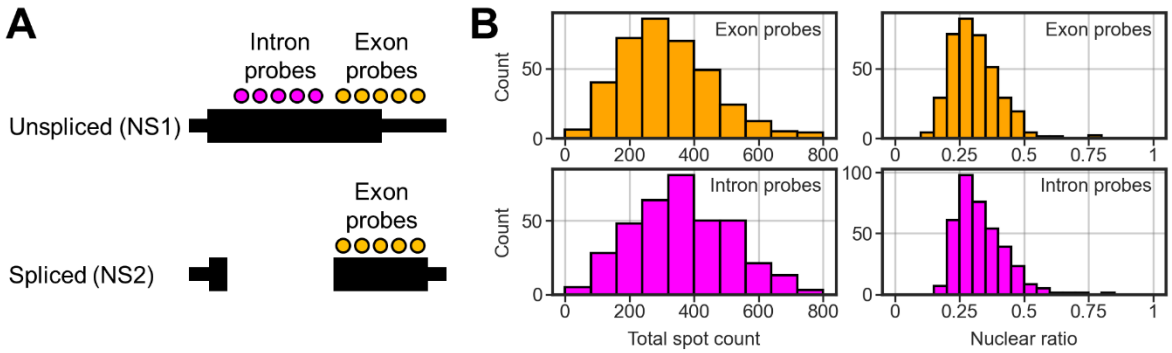
Supplementary figure 2: Nuclear ratio vs. total count



Supplementary figure 2 Independent validation of the nuclear ratio against the total count of mRNAs for each segment by smFISH. The nuclear ratio of mRNAs in each cell was plotted against the total number of mRNAs in that cell. The original data were derived from Supplementary figure 1B and C.

732

Supplementary figure 3: Unspliced form of NS segment



Supplementary figure 3 Exclusive detection of unspliced NS transcripts. (A) Probe-binding sites on the NS segment. The exon probes, used in Fig. 1, target the second exon of NS2 (spliced form), and the intron probes are against the region that is spliced out. Thus, exon probes detect both NS1 and NS2, while intron probes exclusively detect NS1. The black line represents the NS gene. The protein coding region is indicated with the wide horizontal bar. **(B)** The total spot count (left column) detected using either exon probes or intron probes. The spot nuclear ratio in each cell is indicated on the right. Note that the plots of the exon probes (orange) are the reproduction from Fig. 1 and are presented for comparison with those of the intron probes (magenta).

733