Genotype-dependent and non-gradient patterns of RSV gene expression

Felipe-Andrés Piedra[1]*, Xueting Qiu[3], Michael N. Teng[6], Vasanthi Avadhanula[1], Annette A. Machado[1], Do-Kyun Kim[4], James Hixson[4], Justin Bahl[3,5], Pedro A. Piedra[1,2]

[1]Department of Molecular Virology & Microbiology, Baylor College of Medicine, Houston, TX, United States of America

[2]Department of Pediatrics, Baylor College of Medicine, Houston, TX, United States of America

[3]Center for the Ecology of Infectious Diseases, Department of Infectious Diseases, College of Veterinary Medicine, University of Georgia, Athens, GA, United States of America

[4]Human Genetics Center, School of Public Health, University of Texas Health Science Center, Houston, TX, United States of America

[5]Program in Emerging Infectious Diseases, Duke-National University of Singapore Graduate Medical School, Singapore

[6]Division of Allergy and Immunology, Department of Internal Medicine, University of South Florida Morsani College of Medicine, Tampa, FL, United States of America

*Corresponding author

E-mail: Felipe-Andres.Piedra@bcm.edu

Abstract

Respiratory syncytial virus (RSV) is a nonsegmented negative-strand (NNS) RNA virus and a leading cause of severe lower respiratory tract illness in infants and the elderly. Transcription of the ten RSV genes proceeds sequentially from the 3' promoter and requires conserved gene start (GS) and gene end (GE) signals. Previous studies using the prototypical GA1 genotype Long and A2 strains have indicated a gradient of gene transcription. However, recent reports show data that appear inconsistent with a gradient. To better understand RSV transcriptional regulation, mRNA abundances from five RSV genes were measured by quantitative real-time PCR (qPCR) in three cell lines and cotton rats infected with virus isolates belonging to four different genotypes (GA1, ON, GB1, BA). Relative mRNA levels reached steady-state between four and 24 hours post-infection. Steady-state patterns were genotype-specific and non-gradient, where mRNA levels from the G (attachment) gene exceeded those from the more promoter-proximal N (nucleocapsid) gene across isolates. Transcript stabilities could not account for the non-gradient patterns observed, indicating that relative mRNA levels more strongly reflect transcription than decay. While the GS signal sequences were highly conserved, their alignment with N protein in the helical ribonucleocapsid, i.e., N-phase, was variable, suggesting polymerase recognition of GS signal conformation affects transcription initiation. The effect of GS N-phase on transcription efficiency was tested using dicistronic minigenomes. Ratios of minigenome gene expression showed a switch-like dependence on N-phase with a period of seven nucleotides. Our results indicate that RSV gene expression is in part sculpted by polymerases that initiate transcription with a probability dependent on GS signal N-phase.

Author Summary

RSV is a major viral pathogen that causes significant morbidity and mortality, especially in young children. Shortly after RSV enters a host cell, transcription from its nonsegmented negative-strand (NNS) RNA genome starts at the 3' promoter and proceeds sequentially. Transcriptional attenuation is thought to occur at each gene junction, resulting in a gradient of gene expression. However, recent studies showing non-gradient levels of RSV mRNA suggest that transcriptional regulation may have additional mechanisms. We show using RSV isolates belonging to four different genotypes that gene expression is genotype-dependent and one gene (the G or attachment gene) is consistently more highly expressed than an upstream neighbor. We hypothesize that variable alignment of highly conserved gene start (GS) signals with nucleoprotein (i.e., variable GS N-phase) can affect transcription and give rise to non-gradient patterns of gene expression. We show using dicistronic RSV minigenomes wherein the reporter genes differ only in the N-phase of one GS signal that GS N-phase affects gene expression. Our results suggest the existence of a novel mechanism of transcriptional regulation that might play a role in other NNS RNA viruses.

74 # Introduction
75
76 Respiratory syncytial virus (RSV) can infect individuals repeatedly and is the
77 most common pathogen associated with severe lower respiratory tract disease in
78 children worldwide [1-5]. Numerous host-related and environmental risk factors for
79 severe disease are known [6-8] while viral factors are less clear.

80 RSV is a nonsegmented negative-strand (NNS) RNA virus classified into two
81 major subgroups, A and B, largely distinguished by antigenic differences in the
82 attachment or G protein [9, 10]. The two subgroups are estimated to have diverged from
83 an ancestral strain over 300 years ago [11] and have evolved into multiple co-circulating
84 genotypes [11-15].

85 The RNA genome of RSV is embedded in interlinking and helix-forming subunits
86 of nucleocapsid (N) protein, together forming the ribonucleocapsid (RNP) complex [16,
87 17]. Viral mRNA are not encapsidated [17, 18]. Formation of the RNP complex requires
88 high concentrations of N protein and a 5' terminal dinucleotide AC synthesized by the
89 polymerase independently of template [18, 19]. Each subunit of N protein binds a seven
90 nucleotide stretch of RNA via contacts with the sugar phosphate backbone, causing the
91 RNA to adopt a conformation with a distinct configuration of solvent-exposed and buried
92 nucleobases [16, 17]. Exposed nucleobases can presumably interact directly with viral
93 polymerases bound to the RNP complex [16]. Moreover, the alignment of the viral RNA
94 to the N protein within the RNP (N-phase) will determine its pattern of exposed and
95 buried bases. The effects of N-phase on promoter recognition have been explored in
96 RSV and some paramyxoviruses [18, 20-23]. N-phase affects RNA synthesis by

paramyxoviral RNA polymerases but, in RSV, promoter recognition is strongly determined by the proximity of the promoter sequence to the 3' terminus of the genome; replication is abolished if the core promoter starts six or more nucleotides from the 3' end [23]. Unlike its effects on promoter recognition and replication, the effects of N-phase on transcription are unexplored.

Transcription in RSV and other NNS viruses is sequential, with genes transcribed in their order of occurrence from the 3' promoter of the genome [18, 24-29]. Each of the ten genes of RSV contains essential gene start (GS) and gene end (GE) signals flanking the open reading frame (ORF) [30-32]. Transcription is initiated at the GS signal which also serves as a capping signal on the 5' end of the nascent mRNA [18, 33, 34]. The polymerase then enters elongation mode until it reaches a GE signal, where the mRNA is polyadenylated and released [18, 30]. Two genes overlap at the 5' end of the RSV genome. The GE signal of matrix 2 (M2) occurs downstream of the GS signal of the large polymerase (L) gene. The polymerase must return from the M2 GE signal for full-length L mRNA to be made [35], suggesting that transcribing polymerases scan the RSV genome bidirectionally for a new GS signal after terminating transcription. Indeed, scanning polymerase dynamics may be a universal feature of NNS virus transcription [18, 36-38].

By homology with other NNS viruses, it is widely assumed that transcription in RSV follows a gradient, where the extent to which a gene is transcribed falls with its distance from the 3' promoter [29, 39, 40]. Earlier studies reported data consistent with a gradient [39, 41, 42]; however, recent studies show mRNA abundances that peak at the G gene, which is located in the middle of the genome [40, 43]. We recently reported

120 the G gene to be the most abundant in clinical samples obtained from RSV/A- and
121 RSV/B-infected infants [44]. Thus, existing data suggest that patterns of RSV gene
122 expression are more variable than has been assumed.

123 Here we explored the natural diversity of patterns of RSV gene expression by
124 using qPCR to measure mRNA abundances of five different RSV genes (NS1, NS2, N,
125 G, F) from isolates that we sequenced belonging to both subgroups and four different
126 genotypes (RSV/A/GA1, RSV/A/ON, RSV/B/GB1, RSV/B/BA). Genotype-dependent
127 patterns were observed, all diverging from a gradient and all showing higher levels of G
128 mRNA than expected. Transcript stabilities did not account for the non-gradient patterns
129 of mRNA levels. We analyzed GS signal sequence and N-phase, and hypothesized that
130 GS signal N-phase can affect RSV gene expression on the basis of our findings. We
131 found evidence supporting this hypothesis by measuring gene expression from RSV
132 minigenomes encoding two reporter genes, where minigenomes differed only in the N-
133 phase of one GS signal.

134

Results

RSV mRNA abundances

Oligonucleotide standards of known concentration were used to convert cycle threshold (C_T) values measured by real-time PCR for mRNA targets (Fig 1A) to mRNA abundances. Twenty oligonucleotide standards and sets of reagents (primers and probe) were needed to quantify 20 mRNA targets (five genes in four isolates). All reagents gave rise to a similar range of C_T values for standards at equal concentrations (Fig 1B).

Relative mRNA levels

RSV isolates from both major subgroups (A and B) and four different genotypes (A/GA1/Tracy, A/ON/121301043A, B/GB1/18537, B/BA/80171) were used to infect HEp-2 cells (MOI = 0.01). Total mRNA abundances began to plateau at ~48 hours post-infection (pi) for all isolates (Fig 2A), consistent with the presence of significant viral cytopathic effect beyond this time-point. Relative mRNA levels were calculated for each isolate at each time-point by dividing the abundance of each mRNA by the total mRNA abundance (Fig 2B). Relative mRNA levels reached steady-state between four and 24 hours pi (Fig 2B).

Consistent with sequential transcription, the relative levels of NS1 mRNA decreased for all isolates after four hours pi (GA1: -12%; ON: -5%; GB1: -6%; BA: -22%). Percent reductions were greater for the two RSV isolates, GA1 and BA, with lower total mRNA at four hours pi (avg. of GA1 and BA = 170 ± 70 mRNA/reaction vs. avg. of ON and GB1 = 890 ± 170 mRNA/reaction).

All four sets of steady-state mRNA levels were non-gradient, with levels of G mRNA exceeding levels of N mRNA (Fig 3). Steady-state mRNA levels also showed both subgroup- and genotype-specific differences (Fig 3). Between subgroups, relative levels of NS1 and NS2 were most different (Fig 3), with the two being similar in RSV/A, and with NS1 levels exceeding NS2 by a factor of ~5 in RSV/B (Fig 3). Within RSV/A, the level of NS1 exceeded NS2 in the GA1 isolate, and was matched by NS2 in the ON isolate (Fig 3). In RSV/B, the level of G mRNA exceeded N in the BA isolate (~5-fold greater) more than it did in the GB1 isolate (~2-fold greater) (Fig 3). Furthermore, genotype-specific steady-state mRNA levels were comparable in A549, Vero, and HEp2 cell lines (Fig 4A).

We explored whether relative mRNA levels might change in the context of a fully immunocompetent host. A pair of cotton rats was infected with each virus isolate and both lung lavage (LL) and nasal wash (NW) samples were collected at four days pi. Relative mRNA levels were genotype-specific and similar in cotton rat LL and NW samples, and comparable to those measured *in vitro* (Fig 4B).

RSV mRNA stabilities

The observed divergence from a transcription gradient could be the result of differential stability of the RSV mRNAs. Therefore, we measured transcript stabilities by blocking transcription using the RSV RNA-dependent RNA polymerase (RdRp) inhibitor GS-5734 then monitoring mRNA levels by qPCR over time. Decay was measured for all five mRNAs from each of the four isolates in HEp-2 cells (Fig 5A). Exponential decay functions were fit to the data and half-lives were calculated from the decay constants. Half-lives ranged from 10 to 27 hours with a mean of 16 ± 5 hours (Fig 5B). Distributions

of mRNA stabilities varied among the isolates, with GA1 having the greatest uniformity and lowest mean (= 12 ± 1 hours) (Fig 5A). Gene expression patterns were estimated by correcting measured mRNA abundances for degradation and recalculating relative mRNA levels (mRNA expressed = measured mRNA # * $e^{(\text{decay constant} * 24\ \text{hr})}$). Estimated levels of gene expression remained non-gradient; thus, differential mRNA stabilities do not account for the non-gradient patterns observed (Fig 5C). These data indicate that relative mRNA levels are 1) more strongly shaped by gene expression than decay and 2) can safely be interpreted to reflect levels of gene expression.

GS signal sequence and N-phase

Whole genome sequences of the four RSV isolates were obtained by next-generation sequencing and analyzed for differences in GS signals that might help explain the non-gradient gene expression patterns observed. GS signals were highly conserved, with a single U to C substitution at position ten of the G gene GS signal (Fig 6A).

We analyzed GS signal sequences for their alignment with N protein, as the alignment of a GS signal with bound N protein will affect its conformation and determine its configuration of solvent-exposed and buried nucleobases [16]. The alignment of a GS signal with N protein (N-phase) might therefore affect interactions with scanning polymerases and alter the likelihood of transcription initiation. We estimated the N-phase of each GS signal by calculating the remainder resulting from dividing the number of nucleotides separating the GS signal from the L GE signal by seven (the number of nucleotides bound by one subunit of N). The L GE signal was used as a proxy as the exact 5' terminus of each RSV genotype is not known. Thus, the estimated

GS signal N-phase will differ from the actual N-phase if the nucleotide length beyond the end of the L GE signal is not equal to an integer multiple of seven. However, every GS signal N-phase within a genotype would be uniformly affected, making estimated intra-genotype differences equal to actual intra-genotype differences. GS signal N-phase was highly variable, making it a potential source of the variation observed in patterns of gene expression (Fig 6B).

Minigenomes to assess the effect of changing GS signal N-phase on gene expression

We hypothesized that changing GS signal N-phase would alter the likelihood of transcription initiation. To test our hypothesis, we designed plasmids encoding RSV minigenomes containing two reporter genes (Renilla luciferase and Firefly luciferase) each flanked by GS and GE signals [45, 46] (Fig 7A). We specifically altered the N-phase of the Firefly GS signal by introducing single nucleotide insertions within the adjoining 5' untranslated region (UTR) along with compensatory single nucleotide deletions within the adjoining intergenic region (Fig 7B). In this way, both the total length of the minigenome and the N-phase of all other sequences were fixed (Fig 7B). If gene expression occurred independently of GS signal N-phase, ratios of luciferase activity would remain constant. Measured ratios of luciferase activity showed a switch-like dependence on GS signal N-phase, with four states resulting in relatively high activity, two states with low, and one state with intermediate activity (Fig 7C). Ratios increased by as much as 50% relative to the minimum measured (Fig 7C). Thus the N-phase of the Firefly luciferase GS signal affected the relative level of gene expression, and by

231 inference, transcription initiation (Fig 7C). Furthermore, ratios of luciferase activity were
232 consistent with a periodicity of seven nucleotides (Fig 7C).

233

234 ## Discussion
235

236 We observed genotype-dependent and non-gradient patterns of RSV gene
237 expression. We hypothesize that non-gradient patterns require a mechanism to alter the
238 likelihood of transcription initiation at different GS signals. GS signal sequences were
239 highly conserved but varied in alignment with N protein, or N-phase, providing a
240 potential source of biased transcription initiation. Using RSV minigenomes, we showed
241 that varying GS signal N-phase can affect gene expression. These unexpected findings
242 highlight gaps in our knowledge of RSV transcription and raise important issues relevant
243 to future studies.

244 Accurate mRNA abundance measurements by qPCR require reagents that bind
245 target without any mismatches [47, 48]. Perfectly designed and distinct sets of reagents
246 can amplify target with variable efficiency, as the amplification efficiency depends on the
247 physicochemical properties of the reagents (the free energies of different intra- and
248 intermolecular interactions) and the qPCR conditions used. For our 20 oligonucleotide
249 standards, we found the lowest melting temperature from each set of reagents
250 correlated positively with amplification efficiencies and negatively with cycle threshold
251 values (S1 Fig). These correlations indicate that physicochemical differences in the
252 primers and probes can account for the minor variation observed in the amplification of

oligonucleotide standards, and support the accuracy of our approach to measuring viral mRNA abundances.

A gene expression gradient has been widely assumed for RSV, but supporting data come from a modest number of studies and are largely restricted to laboratory-adapted isolates (Long and A2) from the prototypic GA1 genotype of subgroup A. The first measurements were made by Collins and Wertz (1983) using an A2 strain in HEp-2 cells [28, 42, 49]. They discovered the gene order of RSV and found it was approximated by decreasing mRNA abundances measured by northern blot [28, 42, 49]. Barik later reported a gradient by dot blot hybridization of radiolabeled mRNAs produced *in vitro* using ribonucleoprotein (RNP) complex from an RSV Long strain and cell extract from uninfected HEp-2 cells [41]. Over a decade later, Boukhvalova et al. measured a gradient-like pattern by qPCR of mRNA abundances from an RSV Long strain grown in A549 cells [39]. In contrast, Aljabr et al. recently reported mRNA abundances by RNA-Seq from an A2 strain in HEp-2 cells that are inconsistent with a gradient. The most abundant mRNA they observed was associated with the G gene [40]. Levitz et al. reported the G gene to be the most highly expressed gene at later time-points in A549 cells infected with isolates from the RSV/B subgroup [43]. Thus, recent published data indicate that patterns of RSV gene expression vary and do not always follow a gradient. Here, we report data from isolates belonging to four different genotypes (GA1, ON, GB1, BA) showing variable and non-gradient patterns of gene expression and all with an apparent excess of G mRNA.

Studies of transcription in RSV and other NNS RNA viruses show that gene expression depends on a variety of factors. Among the factors affecting RSV gene

276 expression are sequences of GS and GE signals [30-32], intergenic (IG) sequences that
277 can change how efficiently transcription is terminated or initiated [50, 51], and other
278 factors including polymerase mutations and sequences of unknown function [52, 53].
279 Our minigenome experiments add GS signal N-phase to the list of factors involved in
280 RSV gene expression.

281 Our minigenome data suggest that polymerases preferentially initiate
282 transcription at GS signals with certain solvent-exposed nucleobases (3C and 10U of
283 the RSV GS signal). What accounts for this preference, and what events follow GS
284 signal recognition and lead to either transcription initiation or continued scanning is
285 unknown. It is interesting that the U to C substitution in position ten of the G gene GS
286 signal has been shown to result in less not more transcription [30]. Thus, additional
287 factor(s) beyond GS signal N-phase may account for over-expression of the G gene. It
288 is worth stating that transcription initiation, being a molecular event, must be stochastic.
289 RSV transcription is therefore sequential but likely *not obligatorily* sequential. A relative
290 excess of G gene mRNA can occur from polymerases, more often than not, failing to
291 initiate transcription at the N gene before initiating at the G gene. It is also possible that
292 the N gene is usually expressed before the G gene, but G mRNA accumulates more
293 because of polymerase scanning and increased re-initiation of transcription at G. Either
294 scenario might contribute and both are consistent with the ultraviolet (UV) transcriptional
295 mapping data underlying sequential transcription [24-26, 28].

296 Differences in luciferase activities from minigenomes are smaller than the
297 differences observed in viral mRNA levels. Several factors might help explain this
298 finding. First, the viral polymerase complex proteins (M2-1, N, P, L) that drive RNA

299 synthesis in the minigenome system are over-expressed by transfection of plasmids
300 encoding codon-optimized genes. This over-abundance of polymerase proteins might
301 not accurately represent the situation during RSV infection. Minigenomes are also
302 proxies for full-length genomes, being shorter and having a simpler genetic structure.
303 For instance, dicistronic minigenomes contain one pair of GS and GE signals straddling
304 a single intergenic sequence while genomes contain nine variably spaced pairs of GS
305 and GE signals straddling nine intergenic sequences of variable sequence and length.
306 Interactions of transcribing and scanning polymerases with the structurally complex
307 RSV genome might, in nonobvious ways, bias the expression of some genes over
308 others. There could also be differences in the stability of the mRNAs or proteins for
309 Renilla versus Firefly luciferase. Finally, concentrations of both luciferase proteins may
310 have reached saturating intracellular levels prior to measurement (24 hr pi). Saturating
311 protein concentrations would underrepresent differences in mRNA levels.

312 Our results show that transcription initiation by the RSV polymerase depends in
313 part on GS signal N-phase. This potentially helps explain our and other recent
314 observations of 1) non-gradient and 2) variable patterns of gene expression. The
315 functional importance of genotype-dependent patterns of gene expression demands
316 exploration. Finally, GS signal N-phase-regulated transcription initiation might also play
317 a role in other NNS viruses.

318

Materials and Methods

Virus strains

RSV isolates were initially genotyped as described [13, 54] by sequencing a 270 bp fragment in the second hypervariable region of the G gene. RSV/A/GA1/Tracy and RSV/B/GB1/18537 are prototypic strains isolated in 1989 and 1962, respectively [13], while RSV/A/ON/121301043A and RSV/B/BA/80171 are contemporaneous strains isolated in 2013 and 2010, respectively [55, 56].

Cell-lines and cotton rats

HEp-2 (ATCC CCL-23), A549 (ATCC CCL-185), and Vero (ATCC CCL-81) were cultured in minimal essential medium (MEM) containing 10% fetal bovine serum (FBS), 1 µg/ml penicillin, streptomycin, and amphotericin B (PSA), and supplemented with L-glutamine.

Male and female *Sigmodon hispidus* cotton rats were bred and housed in the vivarium in Baylor College of Medicine. Cotton rats were ~75 to 150 g of body weight at the start of the experiments.

Viral replication in cell culture and cotton rats

The media from 70-90% confluent HEp-2, A549, or Vero cells in 24-well plates was aspirated, and 0.2 ml of virus diluted in MEM containing 2% FBS with antibiotics, antifungal, and L-glutamine (2% FBS-MEM) was added to replicate wells for each of the time-points to be acquired. Plates were incubated at 37°C and 5% CO_2 for 1 hour. Following infection, virus-containing media was aspirated and replaced with 1 ml of pre-warmed 2% FBS-MEM. Plates were incubated at 37°C and 5% CO_2 until sample

345 collection. At each time point, the media was aspirated and infected monolayers were
346 lysed with 1X RIPA buffer and pelleted by centrifugation. The supernatant was flash
347 frozen in a mixture of dry ice and 95% ethanol then stored at -80°C.

348 Eight- to ten-week-old male and female cotton rats were sedated and inoculated
349 intranasally with 10^5 plaque forming units (pfu) of RSV as described [57]. Cotton rats
350 were euthanized on day 4 post-infection. Nasal wash (NW) samples were collected from
351 each cotton rat by disarticulating the jaw and washing with 2 ml of collection media (=
352 Iscove's media containing 15% glycerin and mixed 1:1 with 2% FBS-MEM) through
353 each nare, collecting the wash from the posterior opening of the pallet. Lung lavage (LL)
354 samples were collected after the left lung lobe was removed and rinsed in sterile water
355 to remove external blood contamination and weighed. The left lobe was transpleurally
356 lavaged using 3 mL of collection media. Both NW and LL fluids were stored at -80°C.

357 **RNA extraction and reverse transcription**
358

359 Viral RNA was extracted from clarified cell lysates or samples obtained from
360 cotton rats as described [55] by using the Mini Viral RNA Kit (Qiagen Sciences,
361 Germantown, Maryland) and automated platform QIAcube (Qiagen, Hilden, Germany)
362 according to the manufacturer's instructions. Complementary DNA (cDNA) was
363 generated using the SuperScript™ IV First-Strand Synthesis System and oligo$(dT)_{20}$
364 primers according to the manufacturer's instructions (ThermoFisher Scientific).

365 **RSV mRNA abundance measurements**
366

367 Accurate mRNA abundance measurements by qPCR require reagents that bind
368 target without any mismatches [47, 48]. Twenty sets of target-specific primers and

probes (from five mRNA targets for four virus isolates) were designed using whole genome sequences obtained by next-generation sequencing. C_T values were measured using the StepOnePlus Real-Time PCR System (ThermoFisher Scientific). Thresholding was performed according to the manufacturer's instructions [58].

Oligonucleotide standards were used to convert sample C_T values to mRNA abundances. Twenty oligonucleotide standards identical in sequence to the 20 targets of the specific primers and probes described above were purchased from IDT®, received lyophilized and resuspended in TE buffer pH 8. Each oligonucleotide standard was diluted to $4x10^6$ molecules/µl and further diluted serially to a concentration of 40 molecules/µl in TE buffer. Duplicate C_T values were measured for each dilution and an average C_T was calculated. Average C_T values and known amounts (molecules/rxn) were used to construct a standard curve for each oligonucleotide standard.

For cDNAs derived from *in vitro* cell lysate or cotton rat samples, C_T values were measured in duplicate and used to calculate an average. Each average sample C_T value was converted to an mRNA abundance using the linear relationship determined for the appropriate oligonucleotide standard C_T vs. log10 of the oligonucleotide standard amount (molecules/rxn).

RSV mRNA stability measurements

Samples of HEp-2 cells infected with virus isolates at an MOI of 0.01 were collected from single wells of 24-well plates at multiple time-points up to 48 hours after addition of 100 µM GS-5734. GS-5734 is a monophosphate prodrug of an adenosine nucleoside analog that binds a broad range of viral RNA-dependent RNA polymerases

(RdRps) and acts as an RNA chain terminator [59, 60]. Samples were collected as described above using 1X RIPA buffer to lyse infected cells, clarifying the lysate by centrifugation, and flash-freezing and storing the clarified lysate at -80°C. Viral RNA were extracted and converted to cDNA using oligo$(dT)_{20}$ primers. Transcript levels from RNase P (a host housekeeping gene) were measured using qPCR reagents acquired from the Centers for Disease Control and Prevention (CDC) and used to correct viral mRNA levels for well-to-well variation in the amount of sample obtained. Exponential decay functions were fit to the normalized data and used to calculate half-lives. Estimates of the amounts of mRNA expressed up to 24 hours pi were made by correcting the observed mRNA abundances at 24 hours pi for degradation using the exponential decay constants calculated (the number of expressed = the number of observed * $e^{(\text{decay constant} * 24\ \text{hr})}$) and assuming production of all observed mRNA at t = 0 hours post-infection. This unrealistic assumption maximizes the effect of different rates of decay on the estimated levels of total expressed mRNA.

Whole genome sequencing and assembly

cDNAs for sequencing were generated from viral RNA using the SuperScript™ VILO™ cDNA Synthesis Kit and random hexamers (ThermoFisher Scientific). cDNAs were amplified using specific primers, and PCR products of each sample were purified and pooled [61]. Pooled PCR products (1 µg) were digested with the NEBNext dsDNA fragmentase kit (New England BioLabs, Inc., Ipswich, MA). Fragmented DNA was end-repaired with the NEBNext End Repair Module (New England BioLabs, Inc.). End-repaired DNA was ligated with the Ion P1 adaptor and unique Ion Xpress barcode adaptors (KAPA Adapter Kit 1–24; KAPABiosystems). Agencourt AMPure XP beads

(Beckman Coulter, Inc., Brea, CA) were used to selectively capture DNA between 100 and 250 bp in length. All reaction products were purified with the Isolate II PCR kit (Bioline USA, Inc.). These libraries underwent nick translation and amplification. Experion Automated Electrophoresis System (Bio-Rad Laboratories, Inc., Hercules, CA) was used to confirm fragment lengths and molar concentrations. Equal molar amounts of all libraries were pooled and libraries were sequenced by Ion Proton™ System (ThermoFisher Scientific) generating 150 bp reads. Raw data, FASTQ and BAM files, were generated by the Torrent Suite™ Software (version 5.0.4; ThermoFisher Scientific).

Reads were assembled by Iterative Refinement Meta-Assembler (IRMA), which was designed for highly variable RNA viruses with more robust assembly and variant calling [62, 63]. IRMA v0.6.7 (https://wonder.cdc.gov/amd/flu/irma/) was used with an assembly module specifically designed for RSV.

Estimates of gene start (GS) signal N-phase

GS signal N-phase was estimated from whole genome sequences using the end of the L (polymerase) GE signal [30]. The L GE signal (3'-UCAAUAAUUUUU-5'; genome sense) was used as a proxy for the 5' end of the RSV genome. N-phase was calculated by determining the number of bases between the last U of the L GE and the first C of the GS (3'-CCCCGUUUAU-5') then dividing by seven, representing the number of bases encapsidated by a single N molecule.

Minigenome experiments

RSV minigenomes contained Renilla luciferase and Firefly luciferase genes flanked by the RSV A2 leader (le)-NS1 GS and L GE-trailer (Tr) sequences [45, 46]. The Renilla and Firefly luciferase genes were separated by an N/M GE signal, SH-G intergenic (IG) sequence, and F GS signal of variable N-phase. The N-phase of the F GS signal was altered by sequentially deleting the nucleotide immediately 3' to the GS and inserting the same nucleotide immediately 5' of the F GS signal. Eight constructs were constructed assuming the periodicity of seven nucleotides established by Tawar et al. [16]. Minigenomes were co-transfected into BSR-T7 cells with expression plasmids encoding codon-optimized N, P, L, and M2-1 genes [45, 46]. Firefly and Renilla luciferase activities were measured at 24 hours post-transfection using Dual Luciferase Reagent (Promega) [46].

Ethics Statement

All experimental protocols were approved by the Baylor College of Medicine's Institutional Animal Care and Use Committee (IACUC) (license # AN-2307). All experiments were conducted in accordance with the Guide for Care and Use of Laboratory Animals of the National Institutes of Health, as well as local, state and federal laws.

Accession numbers

Sequences reported in this study were deposited in GenBank database under accession numbers MG813977-MG813995.

Acknowledgements

Thanks to Michel Perron of Gilead for providing the viral RdRp inhibitor GS-5734 for use in experiments to measure RSV transcript stabilities. Thanks to Brian Gilbert from the Department of Molecular Virology and Microbiology at Baylor College of Medicine for providing the cotton rats needed to perform RSV infection studies. We thank Kim Tran for technical assistance.

References

1. American Academy of Pediatrics Subcommittee on D, Management of B. Diagnosis and management of bronchiolitis. Pediatrics. 2006;118(4):1774-93. doi: 10.1542/peds.2006-2223. PubMed PMID: 17015575.

2. Hasegawa K, Tsugawa Y, Brown DF, Mansbach JM, Camargo CA, Jr. Temporal trends in emergency department visits for bronchiolitis in the United States, 2006 to 2010. Pediatr Infect Dis J. 2014;33(1):11-8. doi: 10.1097/INF.0b013e3182a5f324. PubMed PMID: 23934206; PubMed Central PMCID: PMCPMC3984903.

3. Mansbach JM, Piedra PA, Teach SJ, Sullivan AF, Forgey T, Clark S, et al. Prospective multicenter study of viral etiology and hospital length of stay in children with severe bronchiolitis. Arch Pediatr Adolesc Med. 2012;166(8):700-6. doi: 10.1001/archpediatrics.2011.1669. PubMed PMID: 22473882; PubMed Central PMCID: PMCPMC3394902.

4. Nair H, Nokes DJ, Gessner BD, Dherani M, Madhi SA, Singleton RJ, et al. Global burden of acute lower respiratory infections due to respiratory syncytial virus in young children: a systematic review and meta-analysis. Lancet. 2010;375(9725):1545-55. doi: 10.1016/S0140-6736(10)60206-1. PubMed PMID: 20399493; PubMed Central PMCID: PMCPMC2864404.

5. Zorc JJ, Hall CB. Bronchiolitis: recent evidence on diagnosis and management. Pediatrics. 2010;125(2):342-9. doi: 10.1542/peds.2009-2092. PubMed PMID: 20100768.

6. Simoes EA. Environmental and demographic risk factors for respiratory syncytial virus lower respiratory tract disease. J Pediatr. 2003;143(5 Suppl):S118-26. PubMed PMID: 14615710.

7. Sommer C, Resch B, Simoes EA. Risk factors for severe respiratory syncytial virus lower respiratory tract infection. Open Microbiol J. 2011;5:144-54. doi: 10.2174/1874285801105010144. PubMed PMID: 22262987; PubMed Central PMCID: PMCPMC3258650.

496 8. Stein RT, Bont LJ, Zar H, Polack FP, Park C, Claxton A, et al. Respiratory syncytial virus
497 hospitalization and mortality: Systematic review and meta-analysis. Pediatr Pulmonol.
498 2017;52(4):556-69. doi: 10.1002/ppul.23570. PubMed PMID: 27740723; PubMed Central
499 PMCID: PMCPMC5396299.
500 9. Collins PL, Melero JA. Progress in understanding and controlling respiratory syncytial
501 virus: still crazy after all these years. Virus Res. 2011;162(1-2):80-99. doi:
502 10.1016/j.virusres.2011.09.020. PubMed PMID: 21963675; PubMed Central PMCID:
503 PMCPMC3221877.
504 10. Johnson PR, Spriggs MK, Olmsted RA, Collins PL. The G glycoprotein of human
505 respiratory syncytial viruses of subgroups A and B: extensive sequence divergence between
506 antigenically related proteins. Proc Natl Acad Sci U S A. 1987;84(16):5625-9. PubMed PMID:
507 2441388; PubMed Central PMCID: PMCPMC298915.
508 11. Zlateva KT, Lemey P, Moes E, Vandamme AM, Van Ranst M. Genetic variability and
509 molecular evolution of the human respiratory syncytial virus subgroup B attachment G protein. J
510 Virol. 2005;79(14):9157-67. doi: 10.1128/JVI.79.14.9157-9167.2005. PubMed PMID: 15994810;
511 PubMed Central PMCID: PMCPMC1168771.
512 12. Agoti CN, Munywoki PK, Phan MVT, Otieno JR, Kamau E, Bett A, et al. Transmission
513 patterns and evolution of respiratory syncytial virus in a community outbreak identified by
514 genomic analysis. Virus Evol. 2017;3(1):vex006. doi: 10.1093/ve/vex006. PubMed PMID:
515 28458916; PubMed Central PMCID: PMCPMC5399923.
516 13. Tapia LI, Shaw CA, Aideyan LO, Jewell AM, Dawson BC, Haq TR, et al. Gene sequence
517 variability of the three surface proteins of human respiratory syncytial virus (HRSV) in Texas.
518 PLoS One. 2014;9(3):e90786. doi: 10.1371/journal.pone.0090786. PubMed PMID: 24625544;
519 PubMed Central PMCID: PMCPMC3953119.
520 14. Schobel SA, Stucker KM, Moore ML, Anderson LJ, Larkin EK, Shankar J, et al.
521 Respiratory Syncytial Virus whole-genome sequencing identifies convergent evolution of

522 sequence duplication in the C-terminus of the G gene. Sci Rep. 2016;6:26311. doi:
523 10.1038/srep26311. PubMed PMID: 27212633; PubMed Central PMCID: PMCPMC4876326.
524 15. Bose ME, He J, Shrivastava S, Nelson MI, Bera J, Halpin RA, et al. Sequencing and
525 analysis of globally obtained human respiratory syncytial virus A and B genomes. PLoS One.
526 2015;10(3):e0120098. doi: 10.1371/journal.pone.0120098. PubMed PMID: 25793751; PubMed
527 Central PMCID: PMCPMC4368745.
528 16. Tawar RG, Duquerroy S, Vonrhein C, Varela PF, Damier-Piolle L, Castagne N, et al.
529 Crystal structure of a nucleocapsid-like nucleoprotein-RNA complex of respiratory syncytial
530 virus. Science. 2009;326(5957):1279-83. doi: 10.1126/science.1177634. PubMed PMID:
531 19965480.
532 17. Ruigrok RW, Crepin T, Kolakofsky D. Nucleoproteins and nucleocapsids of negative-
533 strand RNA viruses. Curr Opin Microbiol. 2011;14(4):504-10. doi: 10.1016/j.mib.2011.07.011.
534 PubMed PMID: 21824806.
535 18. Noton SL, Fearns R. Initiation and regulation of paramyxovirus transcription and
536 replication. Virology. 2015;479-480:545-54. doi: 10.1016/j.virol.2015.01.014. PubMed PMID:
537 25683441; PubMed Central PMCID: PMCPMC4424093.
538 19. Noton SL, Fearns R. The first two nucleotides of the respiratory syncytial virus
539 antigenome RNA replication product can be selected independently of the promoter terminus.
540 RNA. 2011;17(10):1895-906. doi: 10.1261/rna.2813411. PubMed PMID: 21878549; PubMed
541 Central PMCID: PMCPMC3185921.
542 20. Murphy SK, Ito Y, Parks GD. A functional antigenomic promoter for the paramyxovirus
543 simian virus 5 requires proper spacing between an essential internal segment and the 3'
544 terminus. J Virol. 1998;72(1):10-9. PubMed PMID: 9420195; PubMed Central PMCID:
545 PMCPMC109344.
546 21. Tapparel C, Maurice D, Roux L. The activity of Sendai virus genomic and antigenomic
547 promoters requires a second element past the leader template regions: a motif (GNNNNN)3 is

548 essential for replication. J Virol. 1998;72(4):3117-28. PubMed PMID: 9525637; PubMed Central
549 PMCID: PMCPMC109762.
550 22. Vulliemoz D, Roux L. "Rule of six": how does the Sendai virus RNA polymerase keep
551 count? J Virol. 2001;75(10):4506-18. doi: 10.1128/JVI.75.10.4506-4518.2001. PubMed PMID:
552 11312321; PubMed Central PMCID: PMCPMC114204.
553 23. Samal SK, Collins PL. RNA replication by a respiratory syncytial virus RNA analog does
554 not obey the rule of six and retains a nonviral trinucleotide extension at the leader end. J Virol.
555 1996;70(8):5075-82. PubMed PMID: 8764015; PubMed Central PMCID: PMCPMC190462.
556 24. Abraham G, Banerjee AK. Sequential transcription of the genes of vesicular stomatitis
557 virus. Proc Natl Acad Sci U S A. 1976;73(5):1504-8. PubMed PMID: 179088; PubMed Central
558 PMCID: PMCPMC430325.
559 25. Ball LA, White CN. Order of transcription of genes of vesicular stomatitis virus. Proc Natl
560 Acad Sci U S A. 1976;73(2):442-6. PubMed PMID: 174107; PubMed Central PMCID:
561 PMCPMC335925.
562 26. Collins PL, Hightower LE, Ball LA. Transcriptional map for Newcastle disease virus. J
563 Virol. 1980;35(3):682-93. PubMed PMID: 7420539; PubMed Central PMCID: PMCPMC288862.
564 27. Collins PL, Mink MA, Stec DS. Rescue of synthetic analogs of respiratory syncytial virus
565 genomic RNA and effect of truncations and mutations on the expression of a foreign reporter
566 gene. Proc Natl Acad Sci U S A. 1991;88(21):9663-7. PubMed PMID: 1946383; PubMed
567 Central PMCID: PMCPMC52778.
568 28. Dickens LE, Collins PL, Wertz GW. Transcriptional mapping of human respiratory
569 syncytial virus. J Virol. 1984;52(2):364-9. PubMed PMID: 6492254; PubMed Central PMCID:
570 PMCPMC254535.
571 29. Whelan SP, Barr JN, Wertz GW. Transcription and replication of nonsegmented
572 negative-strand RNA viruses. Curr Top Microbiol Immunol. 2004;283:61-119. PubMed PMID:
573 15298168.

574 30. Kuo L, Fearns R, Collins PL. Analysis of the gene start and gene end signals of human
575 respiratory syncytial virus: quasi-templated initiation at position 1 of the encoded mRNA. J Virol.
576 1997;71(7):4944-53. PubMed PMID: 9188557; PubMed Central PMCID: PMCPMC191725.
577 31. Kuo L, Grosfeld H, Cristina J, Hill MG, Collins PL. Effects of mutations in the gene-start
578 and gene-end sequence motifs on transcription of monocistronic and dicistronic minigenomes of
579 respiratory syncytial virus. J Virol. 1996;70(10):6892-901. PubMed PMID: 8794332; PubMed
580 Central PMCID: PMCPMC190738.
581 32. Sutherland KA, Collins PL, Peeples ME. Synergistic effects of gene-end signal mutations
582 and the M2-1 protein on transcription termination by respiratory syncytial virus. Virology.
583 2001;288(2):295-307. doi: 10.1006/viro.2001.1105. PubMed PMID: 11601901.
584 33. Barik S. The structure of the 5' terminal cap of the respiratory syncytial virus mRNA. J
585 Gen Virol. 1993;74 (Pt 3):485-90. doi: 10.1099/0022-1317-74-3-485. PubMed PMID: 8445369.
586 34. Liuzzi M, Mason SW, Cartier M, Lawetz C, McCollum RS, Dansereau N, et al. Inhibitors
587 of respiratory syncytial virus replication target cotranscriptional mRNA guanylylation by viral
588 RNA-dependent RNA polymerase. J Virol. 2005;79(20):13105-15. doi:
589 10.1128/JVI.79.20.13105-13115.2005. PubMed PMID: 16189012; PubMed Central PMCID:
590 PMCPMC1235819.
591 35. Fearns R, Collins PL. Model for polymerase access to the overlapped L gene of
592 respiratory syncytial virus. J Virol. 1999;73(1):388-97. PubMed PMID: 9847343; PubMed
593 Central PMCID: PMCPMC103844.
594 36. Barr JN, Tang X, Hinzman E, Shen R, Wertz GW. The VSV polymerase can initiate at
595 mRNA start sites located either up or downstream of a transcription termination signal but size
596 of the intervening intergenic region affects efficiency of initiation. Virology. 2008;374(2):361-70.
597 doi: 10.1016/j.virol.2007.12.023. PubMed PMID: 18241907; PubMed Central PMCID:
598 PMCPMC2593140.

599 37. Brauburger K, Boehmann Y, Krahling V, Muhlberger E. Transcriptional Regulation in
600 Ebola Virus: Effects of Gene Border Structure and Regulatory Elements on Gene Expression
601 and Polymerase Scanning Behavior. J Virol. 2016;90(4):1898-909. doi: 10.1128/JVI.02341-15.
602 PubMed PMID: 26656691; PubMed Central PMCID: PMCPMC4733972.
603 38. Kolakofsky D, Le Mercier P, Iseni F, Garcin D. Viral DNA polymerase scanning and the
604 gymnastics of Sendai virus RNA synthesis. Virology. 2004;318(2):463-73. PubMed PMID:
605 15015496.
606 39. Boukhvalova MS, Prince GA, Blanco JC. Respiratory syncytial virus infects and
607 abortively replicates in the lungs in spite of preexisting immunity. J Virol. 2007;81(17):9443-50.
608 doi: 10.1128/JVI.00102-07. PubMed PMID: 17596309; PubMed Central PMCID:
609 PMCPMC1951413.
610 40. Aljabr W, Touzelet O, Pollakis G, Wu W, Munday DC, Hughes M, et al. Investigating the
611 Influence of Ribavirin on Human Respiratory Syncytial Virus RNA Synthesis by Using a High-
612 Resolution Transcriptome Sequencing Approach. J Virol. 2016;90(10):4876-88. doi:
613 10.1128/JVI.02349-15. PubMed PMID: 26656699; PubMed Central PMCID: PMCPMC4859727.
614 41. Barik S. Transcription of human respiratory syncytial virus genome RNA in vitro:
615 requirement of cellular factor(s). J Virol. 1992;66(11):6813-8. PubMed PMID: 1404620; PubMed
616 Central PMCID: PMCPMC240184.
617 42. Collins PL, Wertz GW. cDNA cloning and transcriptional mapping of nine
618 polyadenylylated RNAs encoded by the genome of human respiratory syncytial virus. Proc Natl
619 Acad Sci U S A. 1983;80(11):3208-12. PubMed PMID: 6190173; PubMed Central PMCID:
620 PMCPMC394009.
621 43. Levitz R, Gao Y, Dozmorov I, Song R, Wakeland EK, Kahn JS. Distinct patterns of
622 innate immune activation by clinical isolates of respiratory syncytial virus. PLoS One.
623 2017;12(9):e0184318. doi: 10.1371/journal.pone.0184318. PubMed PMID: 28877226; PubMed
624 Central PMCID: PMCPMC5587315.

625 44. Piedra FA, Mei M, Avadhanula V, Mehta R, Aideyan L, Garofalo RP, et al. The
626 interdependencies of viral load, the innate immune response, and clinical outcome in children
627 presenting to the emergency department with respiratory syncytial virus-associated bronchiolitis.
628 PLoS One. 2017;12(3):e0172953. doi: 10.1371/journal.pone.0172953. PubMed PMID:
629 28267794; PubMed Central PMCID: PMCPMC5340370.
630 45. Dochow M, Krumm SA, Crowe JE, Jr., Moore ML, Plemper RK. Independent structural
631 domains in paramyxovirus polymerase protein. J Biol Chem. 2012;287(9):6878-91. doi:
632 10.1074/jbc.M111.325258. PubMed PMID: 22215662; PubMed Central PMCID:
633 PMCPMC3307299.
634 46. Teng MN, Tran KC. Use of Minigenome Systems to Study RSV Transcription. Methods
635 Mol Biol. 2016;1442:155-64. doi: 10.1007/978-1-4939-3687-8_11. PubMed PMID: 27464693.
636 47. Boyle B, Dallaire N, MacKay J. Evaluation of the impact of single nucleotide
637 polymorphisms and primer mismatches on quantitative PCR. BMC Biotechnol. 2009;9:75. doi:
638 10.1186/1472-6750-9-75. PubMed PMID: 19715565; PubMed Central PMCID:
639 PMCPMC2741440.
640 48. Suss B, Flekna G, Wagner M, Hein I. Studying the effect of single mismatches in primer
641 and probe binding regions on amplification curves and quantification in real-time PCR. J
642 Microbiol Methods. 2009;76(3):316-9. doi: 10.1016/j.mimet.2008.12.003. PubMed PMID:
643 19135484.
644 49. Collins PL, Huang YT, Wertz GW. Identification of a tenth mRNA of respiratory syncytial
645 virus and assignment of polypeptides to the 10 viral genes. J Virol. 1984;49(2):572-8. PubMed
646 PMID: 6546401; PubMed Central PMCID: PMCPMC255499.
647 50. Hardy RW, Harmon SB, Wertz GW. Diverse gene junctions of respiratory syncytial virus
648 modulate the efficiency of transcription termination and respond differently to M2-mediated
649 antitermination. J Virol. 1999;73(1):170-6. PubMed PMID: 9847319; PubMed Central PMCID:
650 PMCPMC103820.

651 51. Moudy RM, Sullender WM, Wertz GW. Variations in intergenic region sequences of
652 Human respiratory syncytial virus clinical isolates: analysis of effects on transcriptional
653 regulation. Virology. 2004;327(1):121-33. doi: 10.1016/j.virol.2004.06.013. PubMed PMID:
654 15327903.
655 52. Cartee TL, Megaw AG, Oomens AG, Wertz GW. Identification of a single amino acid
656 change in the human respiratory syncytial virus L protein that affects transcriptional termination.
657 J Virol. 2003;77(13):7352-60. PubMed PMID: 12805433; PubMed Central PMCID:
658 PMCPMC164798.
659 53. Harmon SB, Wertz GW. Transcriptional termination modulated by nucleotides outside
660 the characterized gene end sequence of respiratory syncytial virus. Virology. 2002;300(2):304-
661 15. PubMed PMID: 12350361.
662 54. Peret TC, Hall CB, Schnabel KC, Golub JA, Anderson LJ. Circulation patterns of
663 genetically distinct group A and B strains of human respiratory syncytial virus in a community. J
664 Gen Virol. 1998;79 (Pt 9):2221-9. doi: 10.1099/0022-1317-79-9-2221. PubMed PMID: 9747732.
665 55. Avadhanula V, Chemaly RF, Shah DP, Ghantoji SS, Azzi JM, Aideyan LO, et al.
666 Infection with novel respiratory syncytial virus genotype Ontario (ON1) in adult hematopoietic
667 cell transplant recipients, Texas, 2011-2013. J Infect Dis. 2015;211(4):582-9. doi:
668 10.1093/infdis/jiu473. PubMed PMID: 25156562.
669 56. Nicholson EG, Schlegel C, Garofalo RP, Mehta R, Scheffler M, Mei M, et al. Robust
670 Cytokine and Chemokine Response in Nasopharyngeal Secretions: Association With
671 Decreased Severity in Children With Physician Diagnosed Bronchiolitis. J Infect Dis.
672 2016;214(4):649-55. doi: 10.1093/infdis/jiw191. PubMed PMID: 27190183; PubMed Central
673 PMCID: PMCPMC4957440.
674 57. Stobart CC, Rostad CA, Ke Z, Dillard RS, Hampton CM, Strauss JD, et al. A live RSV
675 vaccine with engineered thermostability is immunogenic in cotton rats despite high attenuation.
676 Nat Commun. 2016;7:13916. doi: 10.1038/ncomms13916. PubMed PMID: 28000669; PubMed

677 Central PMCID: PMCPMC5187593 for the Company. M.L.M., C.C.S., A.L.H., J.M. and C.A.R.
678 are co-inventors of RSV vaccine technology subject to evaluation in this paper. The vaccine
679 technology has been optioned to Meissa by Emory University. The remaining authors declare
680 no competing financial interests.
681 58. AB StepOne and StepOnePlus Real-Time PCR Systems - Relative Standard Curve and
682 Comparative CT Experiments. 2008.
683 59. Lo MK, Jordan R, Arvey A, Sudhamsu J, Shrivastava-Ranjan P, Hotard AL, et al. GS-
684 5734 and its parent nucleoside analog inhibit Filo-, Pneumo-, and Paramyxoviruses. Sci Rep.
685 2017;7:43395. doi: 10.1038/srep43395. PubMed PMID: 28262699; PubMed Central PMCID:
686 PMCPMC5338263.
687 60. Warren TK, Jordan R, Lo MK, Ray AS, Mackman RL, Soloveva V, et al. Therapeutic
688 efficacy of the small molecule GS-5734 against Ebola virus in rhesus monkeys. Nature.
689 2016;531(7594):381-5. doi: 10.1038/nature17180. PubMed PMID: 26934220; PubMed Central
690 PMCID: PMCPMC5551389.
691 61. Agoti CN, Otieno JR, Munywoki PK, Mwihuri AG, Cane PA, Nokes DJ, et al. Local
692 evolutionary patterns of human respiratory syncytial virus derived from whole-genome
693 sequencing. J Virol. 2015;89(7):3444-54. doi: 10.1128/JVI.03391-14. PubMed PMID: 25609811;
694 PubMed Central PMCID: PMCPMC4403408.
695 62. Shepard SS, Meno S, Bahl J, Wilson MM, Barnes J, Neuhaus E. Erratum to: Viral deep
696 sequencing needs an adaptive approach: IRMA, the iterative refinement meta-assembler. BMC
697 Genomics. 2016;17(1):801. doi: 10.1186/s12864-016-3138-8. PubMed PMID: 27737640;
698 PubMed Central PMCID: PMCPMC5064889.
699 63. Shepard SS, Meno S, Bahl J, Wilson MM, Barnes J, Neuhaus E. Viral deep sequencing
700 needs an adaptive approach: IRMA, the iterative refinement meta-assembler. BMC Genomics.
701 2016;17:708. doi: 10.1186/s12864-016-3030-6. PubMed PMID: 27595578; PubMed Central
702 PMCID: PMCPMC5011931.

703

Supporting Information Legends

S1 Fig. Amplification efficiencies positively correlate and C_T values negatively correlate with the minimum melting temperature (min. T_m) of the target-specific qPCR reagents used. (a) Pearson correlation for amplification efficiencies vs. min. Tm: R=0.57, p=0.0086. **(b)** Pearson correlations for C_T values measured at the extremes of target quantity (200 and $2x10^7$ molecules / rxn) vs. min. Tm: R=−0.65, p=0.002 and R=−0.66, p=0.0015, respectively.

724 **Fig 1. qPCR-based measurements of mRNA abundances from five RSV genes and**
725 **four different virus isolates using oligonucleotide standards. (a)** Five of 10 RSV
726 genes were chosen for mRNA abundance measurements by qPCR. The 5 genes (NS1,
727 NS2, N, G, & F) span half of the nucleotide length of the 15.2 kb genome and its entire
728 gene length minus the final two genes, M2 and L. **(b)** Known amounts of different
729 oligonucleotide standards were detected over a similar range of cycle threshold (C_T)
730 values. Twenty different oligonucleotide standards at known concentrations were
731 needed (4 virus isolates x 5 mRNA targets) to convert C_T values measured for viral
732 mRNA targets into mRNA abundances. Each dot represents the mean C_T of duplicate
733 measurements of an oligonucleotide standard at a known concentration or quantity (=
734 number of molecules / qPCR rxn). Dots of like color are dilutions of the same
735 oligonucleotide standard.

736 **Fig 2. Total mRNA abundances plateau beyond 48 hours post-infection and**
737 **relative mRNA levels reach steady-state soon after the start of infection. (a)** Total
738 mRNA abundances (= NS1+NS2+N+G+F) from HEp-2 cells infected with different
739 isolates of RSV (MOI = 0.01) begin to plateau by ~48 hours post-infection (pi). Each dot
740 (RSV/A/GA1Tracy [pale blue]; RSV/A/ON/121301043A [dark blue]; RSV/B/GB1/18537
741 [light green]; RSV/B/BA/80171 [dark green]) represents the mean and error bars the
742 standard deviation of two independent experiments (n=2). For each independent
743 experiment, the mean was calculated from duplicate measurements and used in
744 subsequent calculations. **(b)** Relative mRNA levels reach steady-state sometime
745 between four and 24 hours pi. Histograms depicting relative mRNA levels are shown for
746 all measured time-points (4, 24, 48, 72 hr pi) and all four isolates (color scheme same

as (a)). Each bar depicts the mean mRNA # / total mRNA # of the indicated species and error bars show the standard deviation (n = 2). For each independent experiment, the mean was calculated from duplicate measurements and used in subsequent calculations.

Fig 3. Relative mRNA levels are genotype-specific and non-gradient. Grey bars depict relative mRNA levels expected from an expression gradient resulting from a 20% decrease in transcription at every gene junction. Each dot depicts the mean mRNA # / total mRNA # observed for the indicated species and isolate (RSV/A/GA1Tracy [pale blue]; RSV/A/ON/121301043A [dark blue]; RSV/B/GB1/18537 [light green]; RSV/B/BA/80171 [dark green]) in HEp-2 cells (MOI = 0.01) at steady-state. Steady-state mean relative mRNA levels and standard deviation were calculated using the mean of each relevant time-point (24, 48, 72 hours post-infection). The mean of each time-point was calculated from two independent experiments, and the mean from each experiment was calculated from duplicate measurements as described.

Fig 4. Relative mRNA levels are comparable in different cell lines and in nasal wash and lung lavage samples from infected cotton rats. (a) Relative mRNA levels are comparable in different cell lines. Viral mRNA levels were measured from infected A549 (in yellow), Vero (in orange), and HEp-2 (in blue) cell lines (MOI = 0.01) at 24 hours post-infection (pi). Each bar depicts the mean mRNA # / total mRNA # of the indicated species and error bars show the standard deviation (n = 2). For each independent experiment, the mean was calculated from duplicate measurements and used in subsequent calculations. **(b)** Relative mRNA levels are comparable in lung lavage (LL) and nasal wash (NW) samples from infected cotton rats. Each bar depicts

the mean mRNA # / total mRNA # of the indicated species and error bars show the standard deviation calculated from duplicate measurements of the same sample. Results from LL samples collected 4 days pi are shown in blue (cotton rat A = light blue; cotton rat B = dark blue) and NW samples shown in green (cotton rat A = light green; cotton rat B = dark green).

Fig 5. Transcript stabilities do not account for non-gradient patterns, indicating that relative mRNA levels strongly reflect RSV gene expression. (a) Viral mRNAs decay after addition of GS-5734, a viral polymerase inhibitor. Viral mRNA levels were divided by RNase P mRNA levels to control for well-to-well variation in the amount of sample obtained, then normalized. Each dot represents the mean normalized mRNA # and error bars the standard deviation of two independent experiments (n=2). For each independent experiment, a mean was calculated from the means of two different samples; and each sample mean was obtained from duplicate measurements. **(b)** Decay constants obtained from exponential decay functions fit to each data set were used to calculate mRNA half-lives (RSV/A/GA1Tracy [pale blue]; RSV/A/ON/121301043A [dark blue]; RSV/B/GB1/18537 [light green]; RSV/B/BA/80171 [dark green]). **(c)** Transcript stabilities cannot account for non-gradient mRNA levels. Grey bars depict relative mRNA levels expected from an expression gradient resulting from a 20% decrease in transcription at every gene junction. Each dot depicts the mean expressed mRNA # / total expressed mRNA # estimated for the indicated mRNA species and virus isolate in HEp-2 cells (MOI = 0.01) at 24 hours post-infection (mRNA expressed = mRNA # observed * $e^{(\text{decay constant} * 24\ \text{hr})}$).

Fig 6. GS signal sequence is highly conserved but alignment with N protein (N-phase) is variable, revealing a potential source of genotype-specific and non-gradient gene expression patterns. (a) GS signals are highly conserved and show only a U to C substitution at position ten of the G gene GS signal. (Genomic, i.e., negative-strand, sequence displayed). **(b)** GS signals have variable N-phase. Diagrams show the estimated GS signal N-phase for each gene whose mRNA levels were measured (NS1 (cyan), NS2 (green), N (tawny), G (purple), & F (charcoal)) from the 4 virus isolates used.

Fig 7. Gene expression changes with GS signal N-phase in minigenomes encoding two luciferase enzymes. (a) A minigenome encoding two reporter genes, the second with variable GS signal N-phase. RSV minigenomes contained Renilla and Firefly luciferase genes separated by an N GE signal, SH/G intergenic (IG) sequence, and F GS signal of variable N-phase. **(b)** The F GS signal and neighboring sequences in minigenomes with variable GS signal N-phase. In *italics*: SH-G intergenic (IG) sequence; in **bold**: consensus GS signal sequence; underlined: F GS signal sequence block. (Genomic, i.e., negative-strand, sequence displayed). **(c)** Changing ratios of Firefly to Renilla luciferase activity indicate that GS signal N-phase can affect gene expression. Each bar represents the mean and error bars the standard deviation of measurements of four samples taken at 24 hr pi from a single experiment.

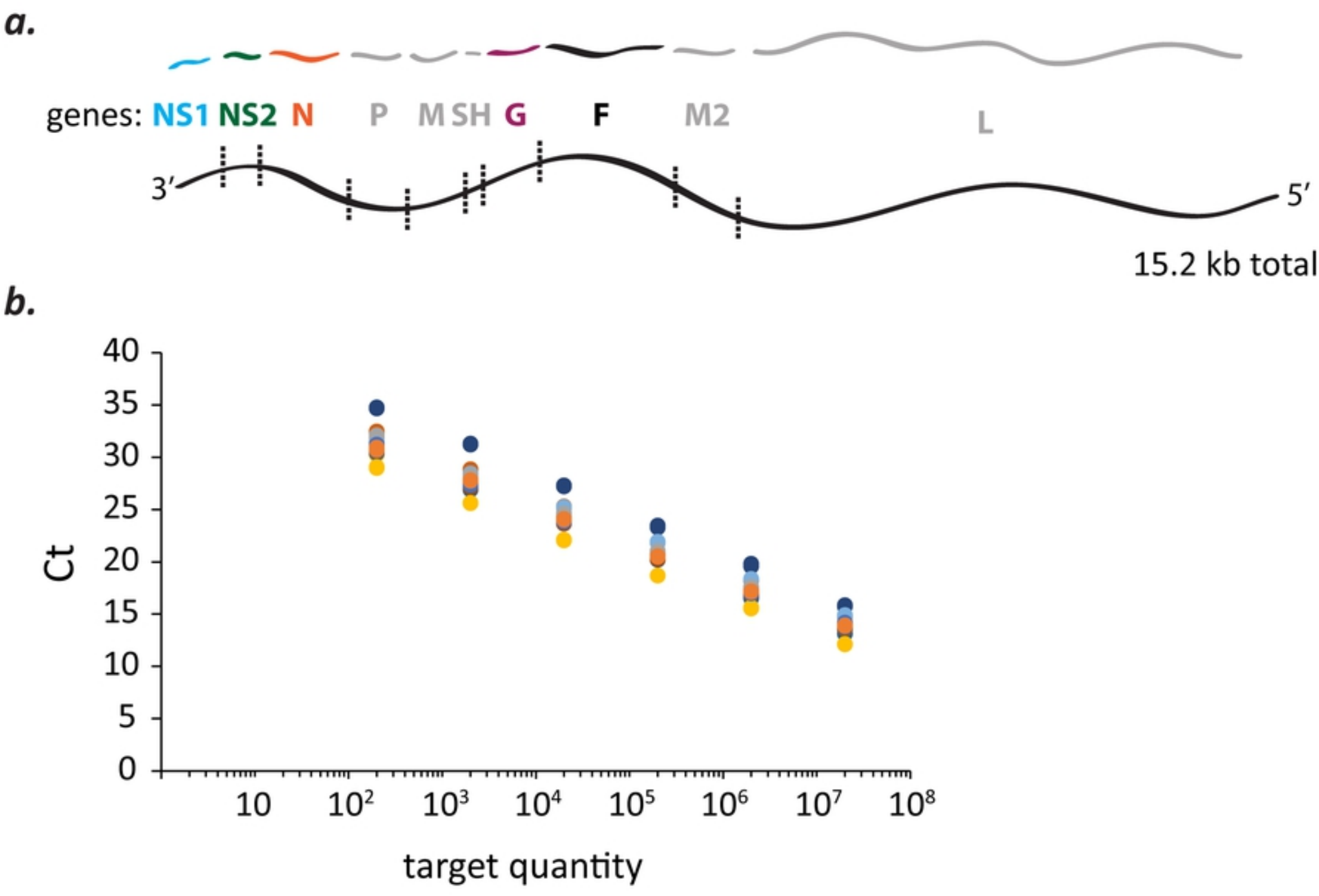

a.
genes: NS1 NS2 N P M SH G F M2 L
3′
5′
15.2 kb total
b.
Ct
40
35
30
25
20
15
10
5
0
10
10²
10³
10⁴
10⁵
10⁶
10⁷
10⁸
target quantity

Fig 1.

Fig 1

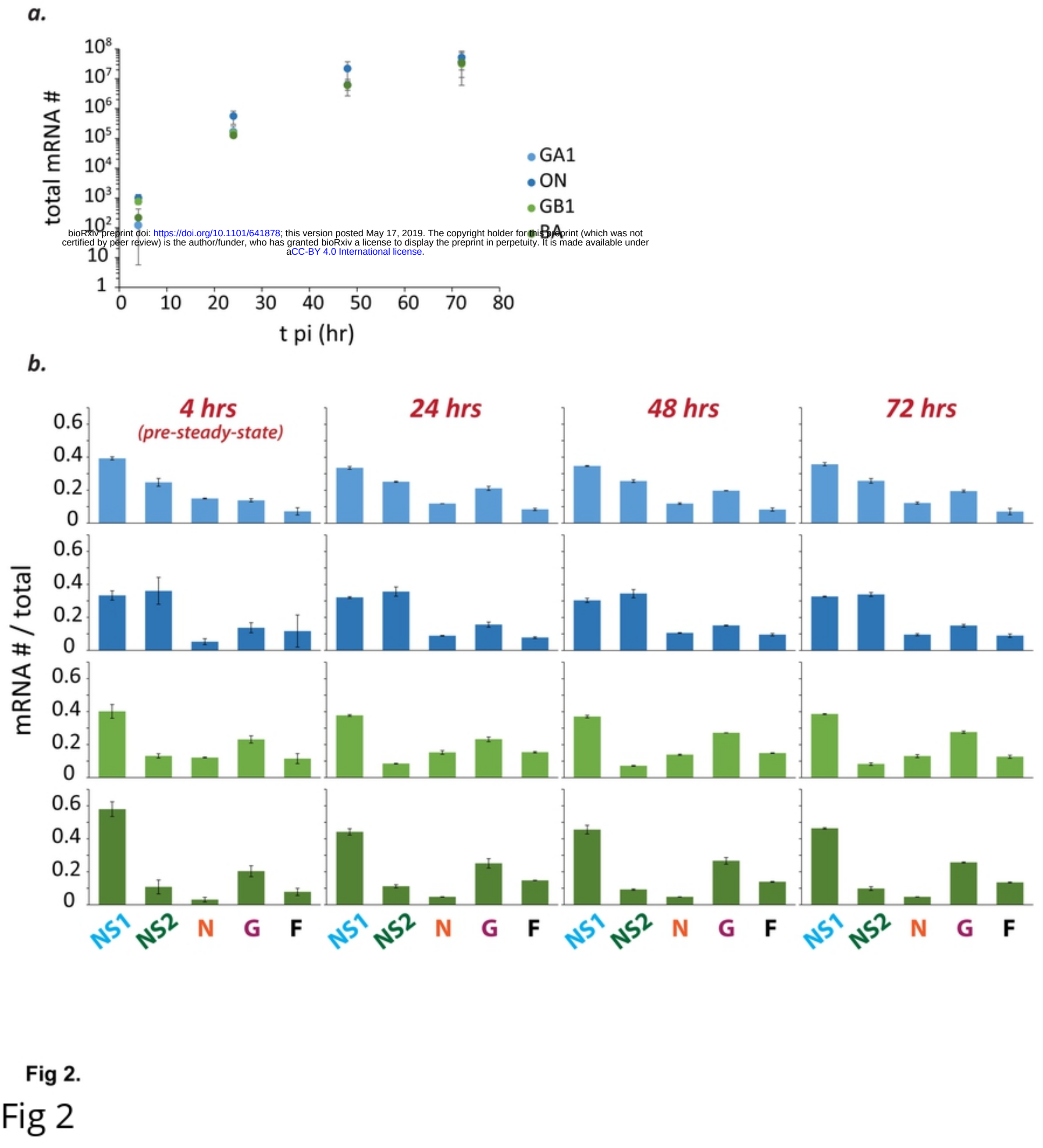


Fig 2.

Fig 2

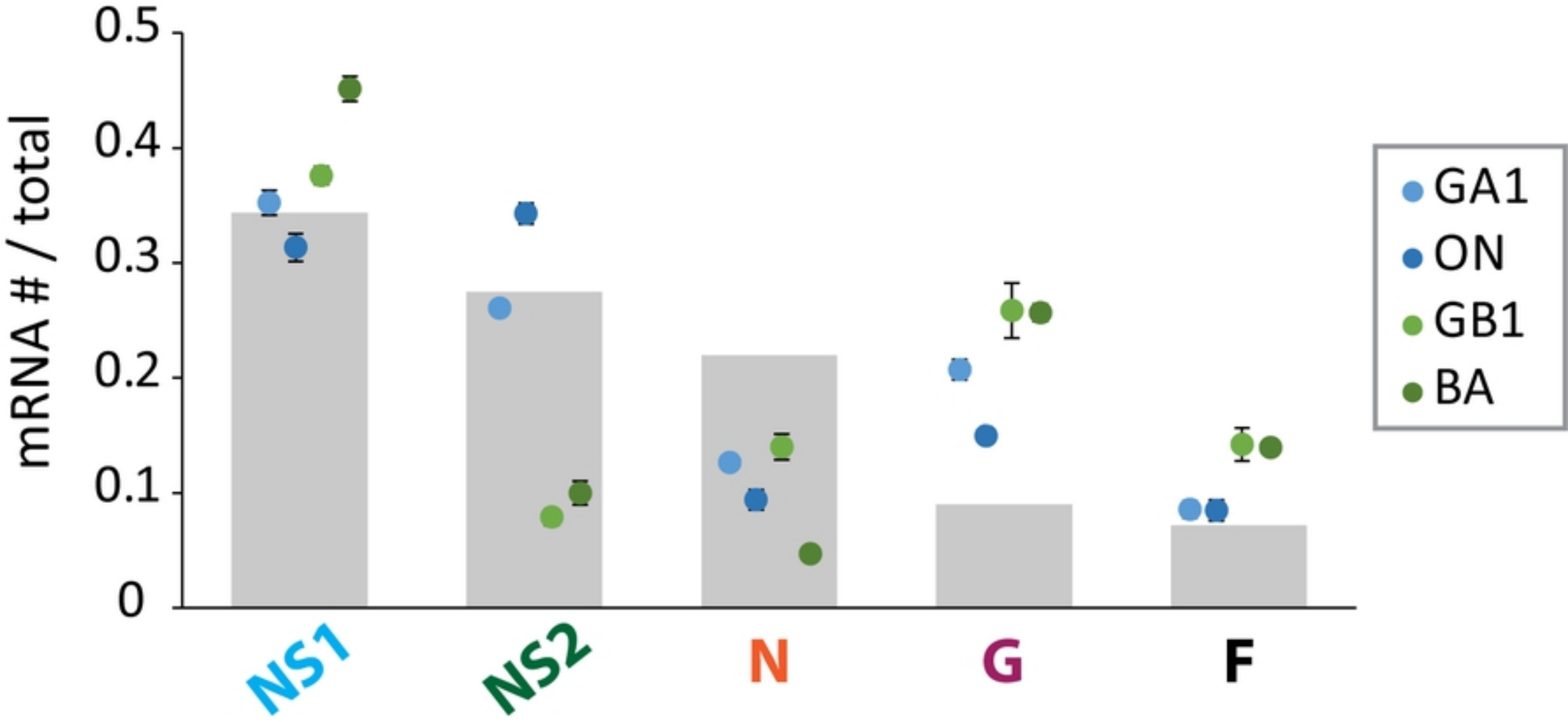


Fig 3.

Fig 3

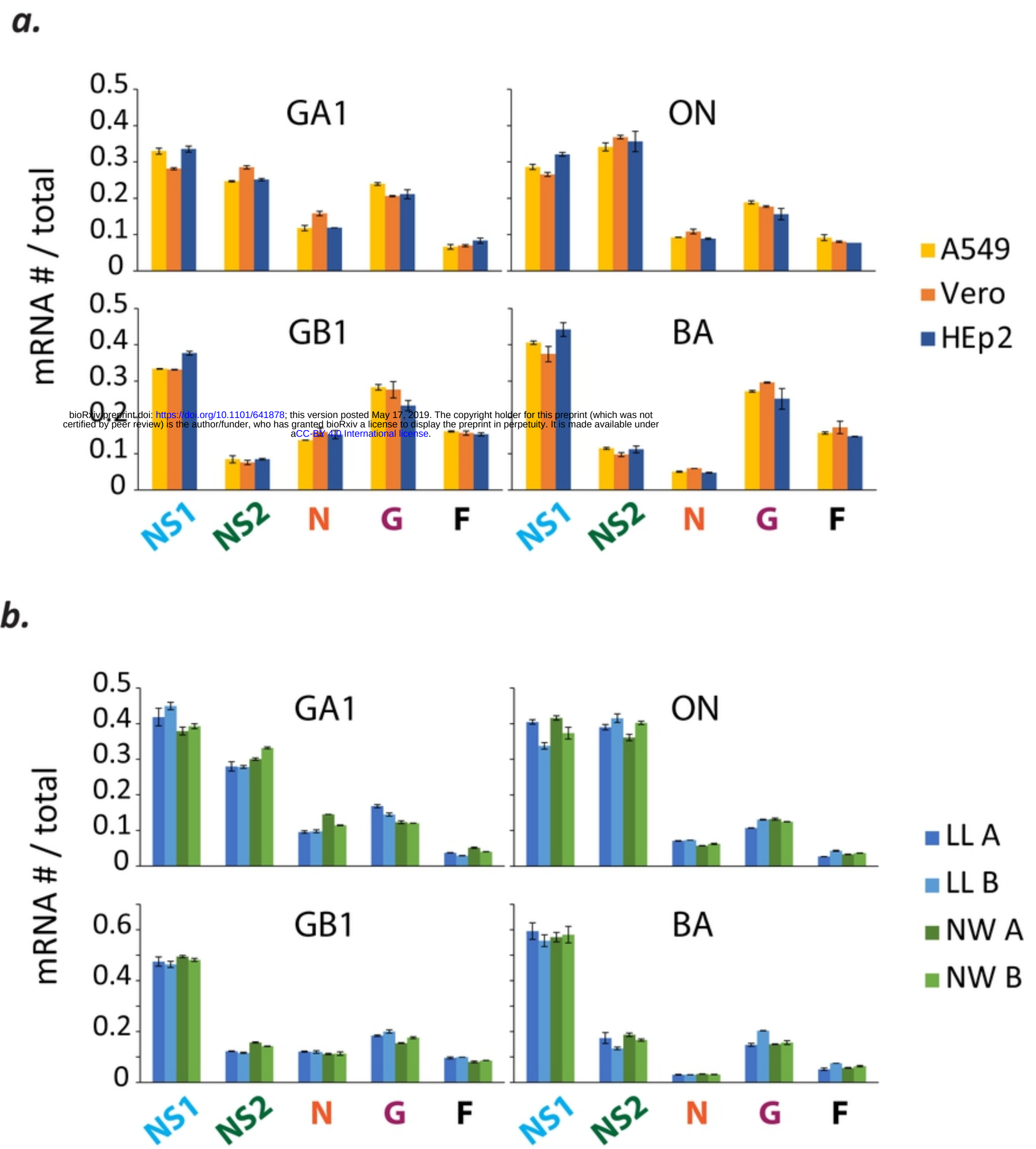


Fig 4.

Fig 4

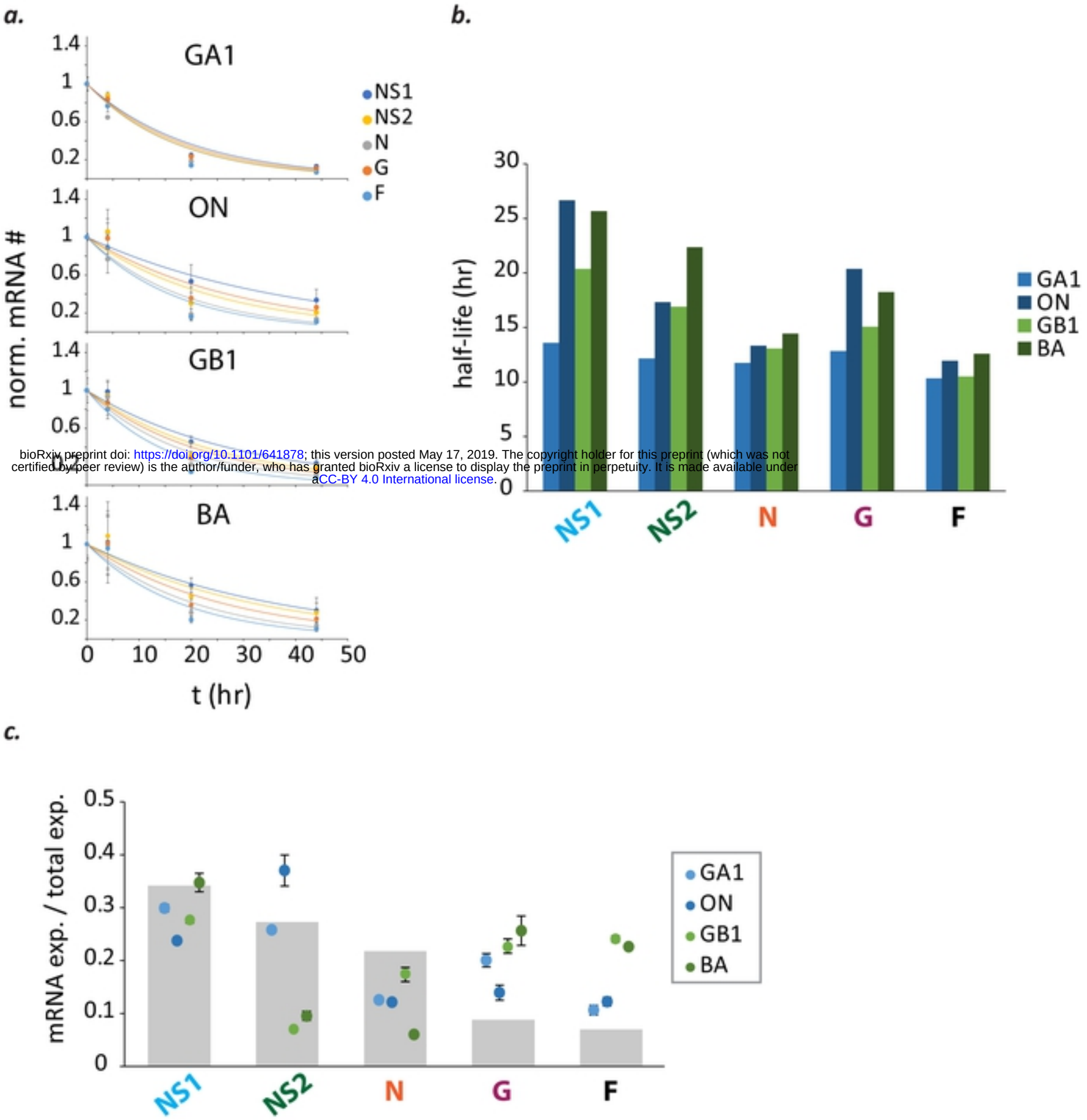


Fig 5.

Fig 5

a.

	NS1, NS2, N, F	G
GA1	CCCCGUUUA**U**	CCCCGUUUA**C**
ON	• • • • • • • • • •	• • • • • • • • • •
GB1	• • • • • • • • • •	• • • • • • • • • •
BA	• • • • • • • • • •	• • • • • • • • • •

b.

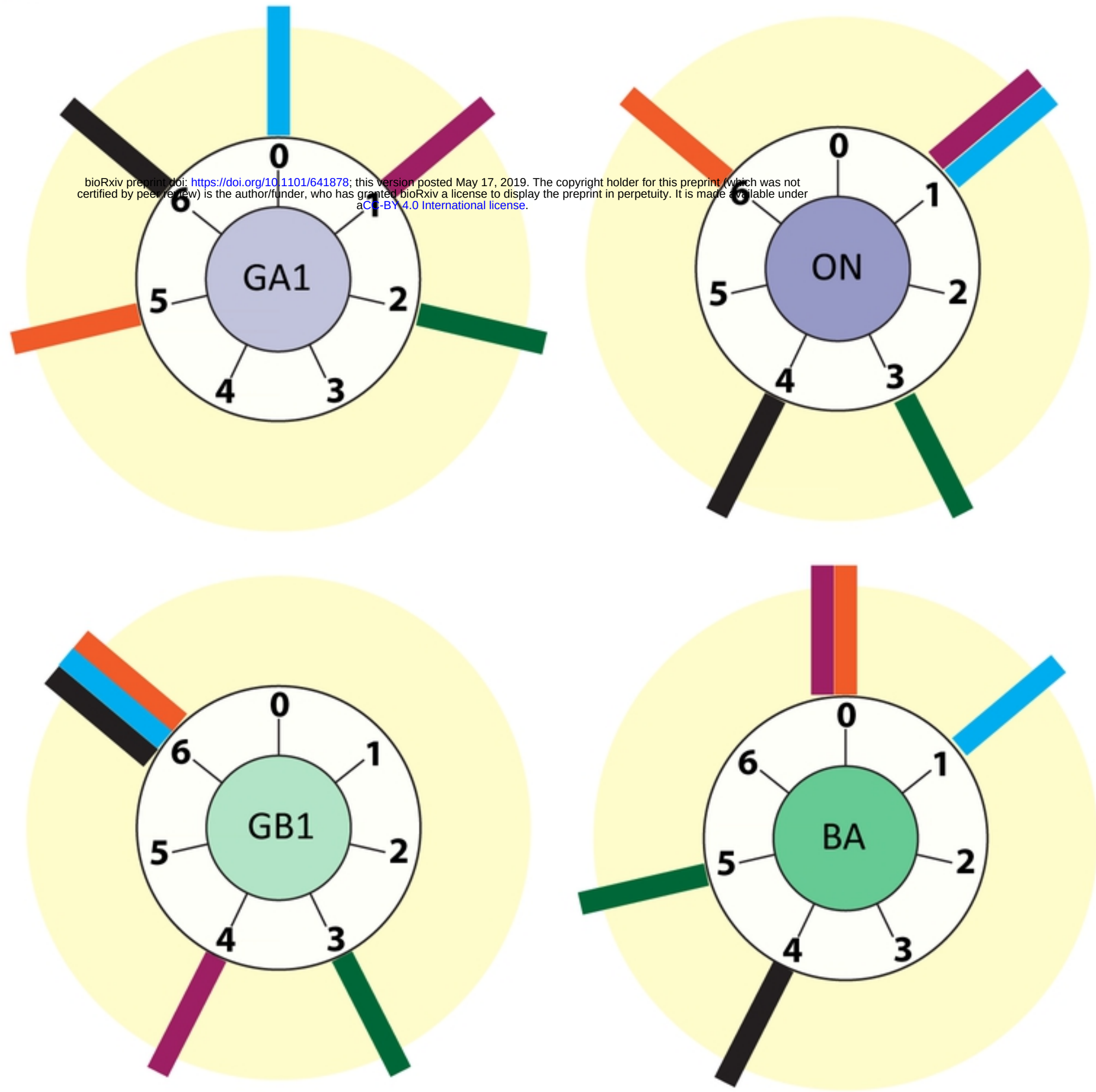
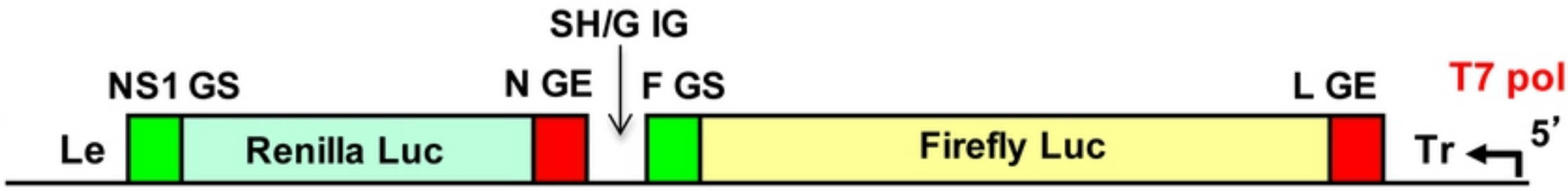


Fig 6.

Fig 6

a.



b.

NP0 *GUAUUGUUACUUGAUCCUAUAGUUCUGAUUGUUAUUGUA***ACCCCGUUUA***UUGU*

NP1 *GUAUUGUUACUUGAUCCUAUAGUUCUGAUUGUUAUUGU***ACCCCGUUUA***UUGU**A**

NP2 *GUAUUGUUACUUGAUCCUAUAGUUCUGAUUGUUAUUG***ACCCCGUUUA***UUGU**UA**

NP3 *GUAUUGUUACUUGAUCCUAUAGUUCUGAUUGUUAUU***ACCCCGUUUA***UUGU**GUA**

NP4 *GUAUUGUUACUUGAUCCUAUAGUUCUGAUUGUUAU***ACCCCGUUUA***UUGU**UGUA**

NP5 *GUAUUGUUACUUGAUCCUAUAGUUCUGAUUGUUA***ACCCCGUUUA***UUGU**UUGUA**

NP6 *GUAUUGUUACUUGAUCCUAUAGUUCUGAUUGUU***ACCCCGUUUA***UUGU**AUUGUA**

NP7 *GUAUUGUUACUUGAUCCUAUAGUUCUGAUUGU***ACCCCGUUUA***UUGU**UAUUGUA**

c.

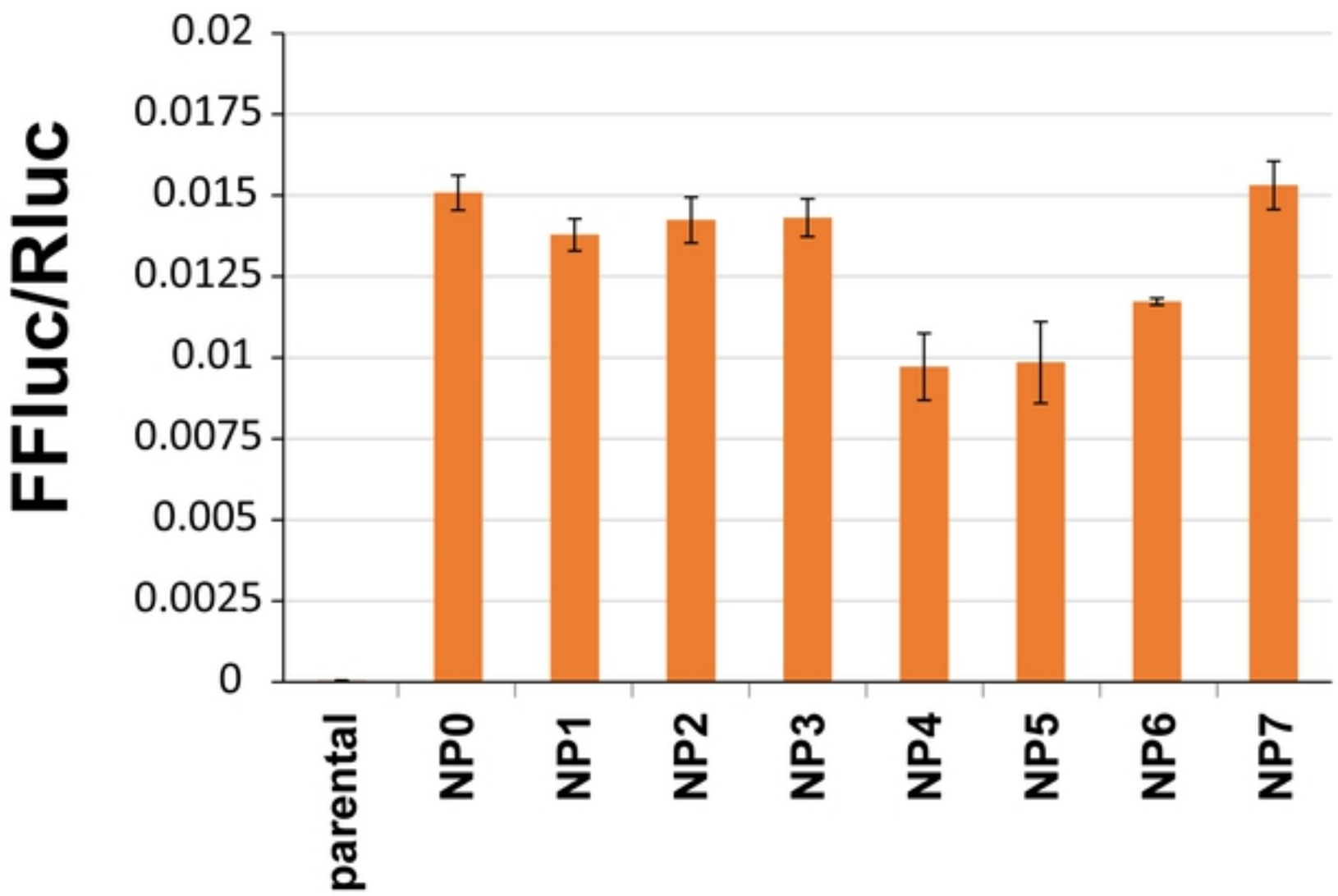


Fig 7.

Fig 7