

RESEARCH ARTICLE

The transcriptional gradient in negative-strand RNA viruses suggests a common RNA transcription mechanism

Connor R. King^{1,2,3*}, Casey-Tyler Berezin², Brian Munsky², Jean Peccoud²

1 LSI Genomics, Princeton University, Princeton, New Jersey, United States of America, **2** School of Biomedical and Chemical Engineering, Colorado State University, Fort Collins, Colorado, United States of America, **3** Cell and Molecular Biology Program, Colorado State University, Fort Collins, Colorado, United States of America

* connor.king@princeton.edu



OPEN ACCESS

Citation: King CR, Berezin C-T, Munsky B, Peccoud J (2026) The transcriptional gradient in negative-strand RNA viruses suggests a common RNA transcription mechanism. *PLoS Comput Biol* 22(6): e1014441. <https://doi.org/10.1371/journal.pcbi.1014441>

Editor: Eric C. Dykeman, University of York, UNITED KINGDOM OF GREAT BRITAIN AND NORTHERN IRELAND

Received: September 19, 2025

Accepted: June 12, 2026

Published: June 24, 2026

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pcbi.1014441>

Copyright: © 2026 King et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution,

Abstract

Nonsegmented negative-strand RNA viruses (NNSV) are a diverse class of medically relevant viruses which display a conserved attenuation gradient in the transcription of their genomes. This gradient has been traditionally explained by the Stop-Start model which attributes attenuation to polymerase behavior at gene junctions. In this article, we evaluate an alternative explanation where the gradient arises from polymerase dynamics during transcription. We introduce the RNA Polymerase Association Mechanism (RAM) model, a coarse-grained stochastic framework that describes transcription using two parameters related to polymerase processivity and the ability of the polymerase to backtrack. The RAM model accurately reproduces transcriptional gradients across diverse NNSVs as well as in gene-shuffled VSV variants. Additionally, the inferred polymerase processivity appears correlated to the length of the viral genomes suggesting a conserved constraint on transcription across these viruses. While the RAM model does not account for all known molecular features of NNSV transcription, it provides a parsimonious and predictive framework for relating genome architecture and transcription. These results support the view that, in tandem with the traditional junction-centric mechanisms governing transcription, nonspecific attenuation mechanisms contribute to the NNSV transcriptional gradient and warrant closer inspection in future studies which could lead to better rational genome design in viral studies and biomedical applications.

Author summary

Nonsegmented negative-strand RNA viruses (NNSVs), including Ebola virus, rabies virus, and vesicular stomatitis virus (VSV), share a common transcriptional gradient in which genes closer to the genome's 3' end are transcribed at higher levels than those farther away. The prevailing Start-Stop model attributes

and reproduction in any medium, provided the original author and source are credited.

Data availability statement: The code for this article can be found at https://github.com/Peccoud-Lab/NNSV_Transcription_Model_2024. The data used to fit the models can be found in Notebook1 and Notebook2 on GitHub. The mRNA expression data utilized comes from a variety of NNSVs including members of the Rhabdoviridae [8], Filoviridae [36], Pneumoviridae [37], and Paramyxoviridae [13,38]. This data was collected using a variety of methodologies including blotting-based methods and RNA-sequencing. When referring to the expression gradient, this is the normalized relative expression of mRNAs ordered from 3' to 5' along the genome. The data for sequence positions can be found in the same notebooks. The GenBank Accession Numbers of the viral genomic sequences used can be found in S3 Table.

Funding: This work was supported by the National Institutes of Health R01GM147816 to JP and the National Institute of General Medical Sciences Award T32GM132057 to CK. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

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

this gradient to polymerase behavior at gene junctions. The sequences located at these junctions control many essential aspects of transcription in these viruses such as initiation, termination, and post-transcriptional modifications (such as capping and polyadenylation). However, it is unclear whether the overarching transcriptional gradient in all NNSVs is the result of variation in the actions that occur at these junctions. Here, we propose a coarse-grained alternative framework that can recapitulate the observed gradient in these viruses without incorporating gene junction information. We propose the RNA-polymerase Association Mechanism (RAM) model, which describes transcription as a simple stochastic process where during transcription, the viral polymerase advances along the genome and either adds the next base to the end of the nascent transcript or dissociates from the nascent transcript resulting in premature termination of transcription. Our model requires only two parameters to reproduce transcriptional gradients across diverse NNSVs, including gene-shuffled VSV variants. We further show that the probability that the polymerase releases the transcript is correlated with genome length, suggesting a universal constraint on NNSV transcription. By shifting the focus from gene-junction sequences to polymerase dynamics, our work redefines how viral transcription is understood. These findings have broad implications for synthetic virology and vaccine design, providing a predictive framework for modifying viral gene expression, thereby enabling more precise engineering of viral vectors for therapeutic applications.

Introduction

Nonsegmented negative-strand RNA viruses (NNSVs), including Ebola (EBOV), rabies virus, and vesicular stomatitis virus (VSV), are a diverse class of viruses that are commonly used in biomedical applications such as vaccine vectors [1]. These viruses possess a single RNA strand encoding tightly packed genes (Fig 1) and their transcriptional output is linked to gene order. In the 1970s, UV inactivation studies in VSV suggested a single polymerase initiates at the 3' end and transcribes each gene sequentially [2–4]. Although early data showed similar protein levels [5], differences in protein size implied non-equimolar expression. This difference in relative expression was characterized at both the protein and transcript level to reveal a 3' to 5' gradient in expression, suggesting that the control was at the transcription level [6]. This gradient led researchers to suggest two hypotheses.

Iverson and Rose rationalized that since the transcripts of VSV had been shown to have roughly equal degradation rates [7], the difference in relative expression was due to actions in the genome either at specific locations (gene junctions) or non-specifically [8]. They designed probes to quantify the relative amounts of the 3' and 5' end of transcripts as they hypothesized that nonspecific attenuation would result in a gradient at the transcript level while specific attenuation would show equimolar levels of these fragments. This paper also established that at most gene junctions there is a 29–33% drop while the junction before the L polymerase causes a 80–90%

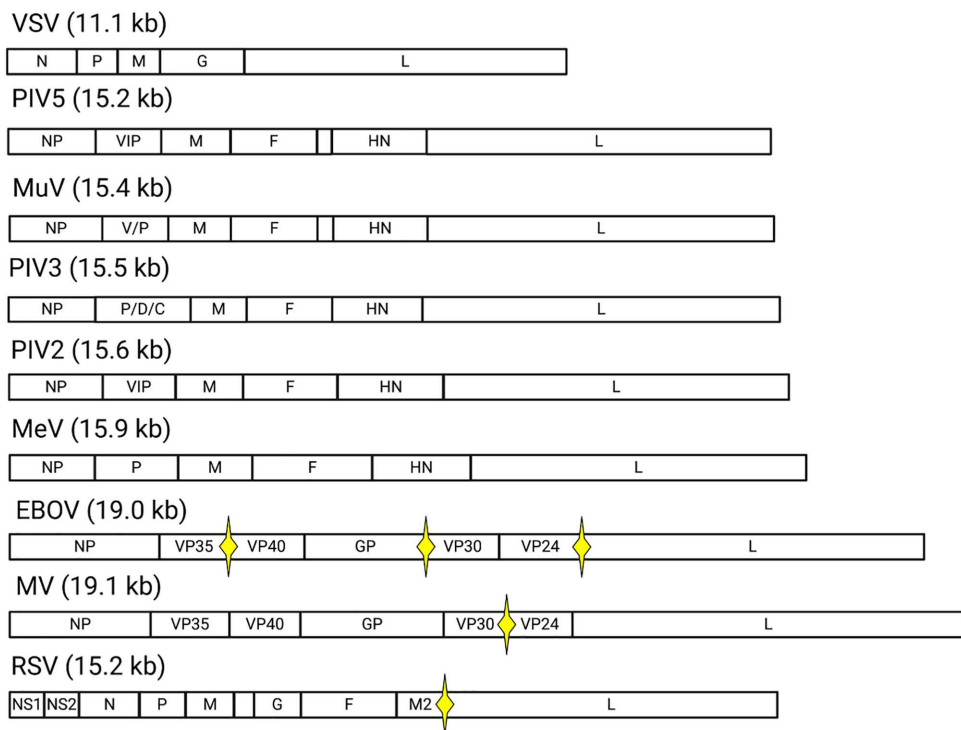


Fig 1. Genome Architecture of NNSVs. Graphical representation of the genomes of nonsegmented negative-strand RNA viruses (NNSVs) in 3' to 5' orientation with gene lengths to scale. Yellow stars indicate overlapping genes in the genome. The genes that lack labels are the regions containing the small hydrophobic (SH) gene. Common proteins present in most of these species are the N/NP = nucleocapsid, P = phosphoprotein, M = matrix protein, G/GP = glycoprotein, L = RNA-dependent RNA polymerase, F = fusion protein, and HN = hemagglutinin-neuraminidase protein.

<https://doi.org/10.1371/journal.pcbi.1014441.g001>

decrease in transcription. This hypothesis was soon dubbed the “Stop-Start Hypothesis” [9] and was extended to many other NNSVs, including EBOV [10,11], Marburg virus (MV) [10,12], parainfluenza virus type 2 (PIV2), parainfluenza virus type 3 (PIV3) [13,14], parainfluenza virus type 5 (PIV5) [13,15], mumps virus (MuV) [13,16], measles virus (MeV) [17], and respiratory syncytial virus (RSV) [18,19]. While this was a hot topic for many years, the debate over this phenomenon has quieted and to this day, the Stop-Start Hypothesis remains the dominant hypothesis [20].

Much of the literature has focused on dissecting what sequences at these gene junctions control the level of attenuation observed. Particular focus has been interested in the gene junction prior to the L gene as in general, we tend to see consistent levels of attenuation for all junctions at around 30% except for the L gene which exhibits about 80–90% attenuation. The cause of this has remained elusive, as in some viruses this attenuation can be strongly attributed to gene junctions while others it cannot. Indeed, in rabies virus we can see that the level of attenuation is strongly tied to the intergenic sequence located between the stop and start signal sequences and this has been used to generate variants of rabies virus that overexpress the L polymerase [21]. Additionally, RSV also has been shown to have termination depend strongly on the gene junction present, but intriguingly there is no effect from the very structurally diverse intergenic regions [18,22].

On the other hand, some viruses like EBOV do not offer such a clean explanation. For EBOV, attenuation at the L junction is unremarkable, while we do see an effect on the NP/VP35 and VP30/VP24 junctions [23]. Although, it should be noted that the inclusion of the 5' and 3' UTRs might have influenced these experiments as they may impact the stability of transcripts. Another intriguing study using PIV5 looked at attenuation from different gene junctions in the context of

expressing HN and L. Three different levels of attenuation were observed across the 6 genomes with the M-F, F-SH, and SH-HN junctions producing higher attenuation than the HN-L junction. Even more perplexing, these variations in attenuation rates were not observed when the genes in the minigenome were swapped for F and V/P [15].

These works suggest that there are several mechanisms that come together to orchestrate the gradient observed in literature. Indeed, some of these mechanisms that have come into focus have more to do with intrinsic polymerase properties and could be viewed as a source of nonspecific attenuation. The efficiency of 5' capping has gained more focus in recent years where it has been shown that the formation of the 5' cap during transcription is necessary for the proper elongation and termination of transcription [17,24–26]. When this process fails, it results in the production of truncated transcripts. Indeed such cap-less truncated transcripts had even been noted in RSV a decade earlier [27]. Something intriguing here is the observation of truncated transcripts, the absence of which was central to the initial formation of the stop-start model.

As such, there may be value in revisiting the study of NNSV transcription from the lens of nonspecific attenuation. Indeed, it has been shown that in measles virus the slope of the transcriptional gradient can be altered based on the affinity between the X domain of the P protein (which is part of the transcription complex) and the C terminal tail of the Nucleoprotein [28]. Revisiting these proposed mechanisms is also warranted as the template dissociation-based mechanism appears to be at odds with the structure of the polymerase-template complex. The L polymerase completely encircles the genome while moving down the genome [29]. So, the polymerase would need to undergo a dramatic rearrangement that is highly unlikely in order to dissociate from the template. Moreover, junction sequences are highly diverse, even within the same virus (Figs 2A, S1 Fig.). The initiation sequence itself is poorly conserved, with the minimal required motif being 3'-UYGnnnnnn-5' [30]. This raises the question of how this consistent gradient that is present across all NNSVs can be the direct result of such a diverse array of sequences.

In this article, we develop a model of transcription that attributes the amount of attenuation at each junction due to the distance the polymerase must travel from the 3' end of the genome to the gene's termination signal. First, we demonstrate that the mechanism is consistent with a wide variety of NNSV transcriptional gradient datasets. Then we show that the resulting predictions from a model fit to VSV transcriptional data is as consistent with the observed protein expression gradients in gene shuffled VSV variants as the Stop-Start model despite having fewer parameters (1 instead of 4). Additionally, there is even some weak evidence that the RAM model produces slightly better predictions than the Stop-Start model when considering the expression of genes in gene-shuffled VSV variants. Given that it is unlikely that the polymerase dissociates from the genome template [29], it is unclear why moving down the genome results in fewer transcripts being produced. It is possible that "losing" the nascent transcript results in a polymerase state that is unable to continue transcription until it restarts all over again at the 3' entry site. Indeed, there are extensive regulatory interactions that have been characterized between the nascent RNA and the RNA polymerase [31]. Since we attribute the amount of attenuation at each junction to the polymerase's ability to remain associated with the nascent transcript during transcription, we name this model the RNA Polymerase Association Mechanism (RAM) model.

Results

Modeling NNSV transcription as a walk

The RAM model consists of a single polymerase performing a walk beginning at the 3' end of the genome (Fig 3). In each step, the polymerase takes one of two actions. The polymerase either takes a single base step forward and continues productive transcription or it dissociates from the transcript and enters a nonproductive state where it can no longer initiate transcription. The probabilities associated with these actions will be referred to as p_{maintain} and p_{drop} , respectively. Since p_{drop} is the complement of p_{maintain} , these represent a single parameter in this model. The relative level of expression of most genes (p_{express}) can be calculated from the distance of the stop signal for the gene from the 3' end of the genome (n) using the equation:

$$p_{\text{express}} = (p_{\text{maintain}})^n$$

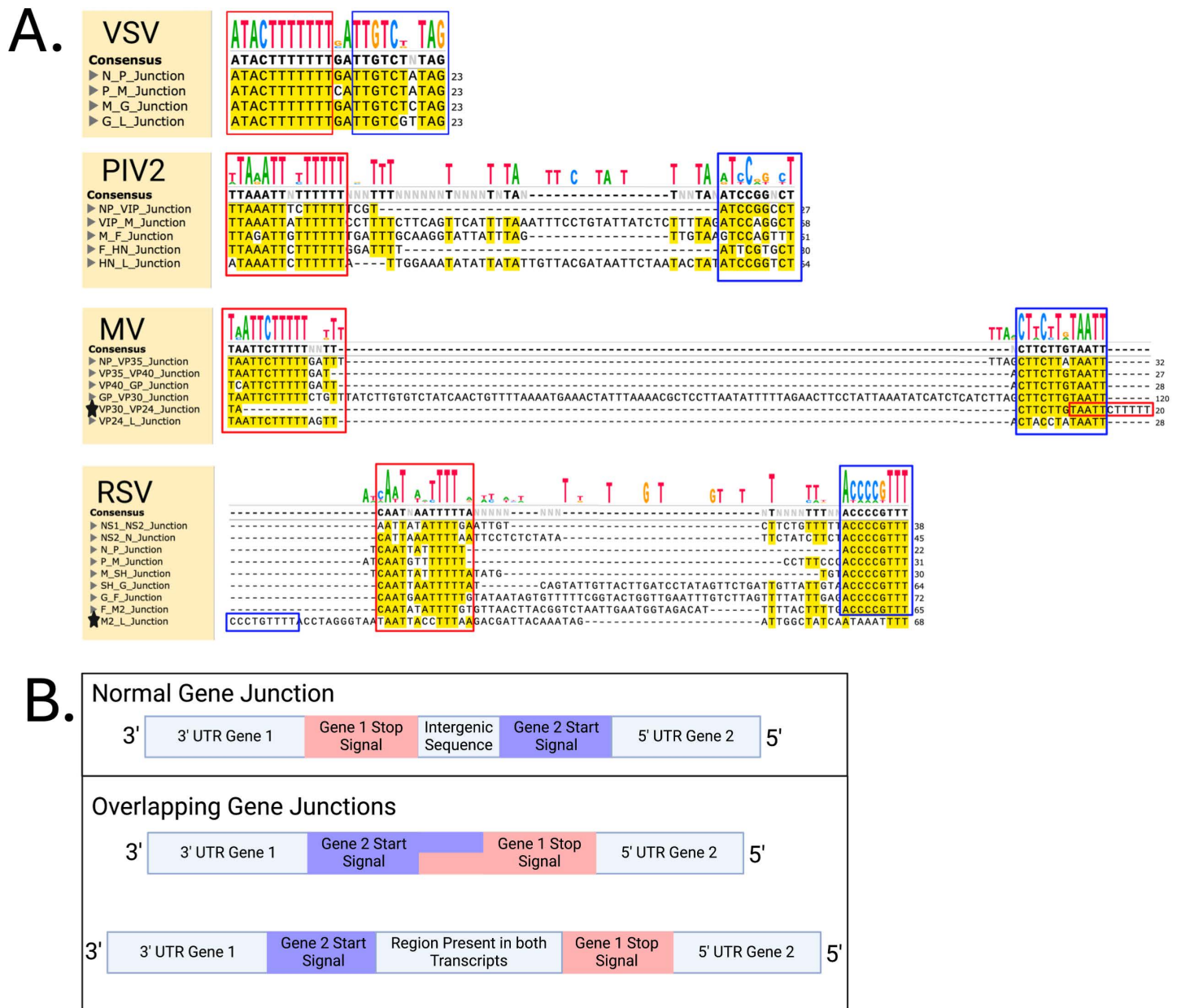


Fig 2. Gene Junction Diversity of NNSVs. A. Alignments of gene junctions for 4 NNSVs. Each junction has blue and red boxes highlighting the conserved transcriptional start and stop sequences respectively. Black stars to the left of the sequence indicate that the junction is an overlapping junction and so the locations of the start and stop sequences are inverted compared to the others. **B.** Cartoon diagram of the types of gene junctions observed in NNSVs and overlapping gene junctions. The UTRs in this figure refer to the untranslated regions of the transcripts that are produced.

<https://doi.org/10.1371/journal.pcbi.1014441.g002>

Where n is the number of bases that the termination sequence is from the 3' end of the genome. It is important to note that some NNSVs have overlapping genes, where the transcriptional start signal of one gene is located within the upstream gene (Fig 2). In cases where overlapping occurs and the transcription start site of a gene (Gene 2) is located within an upstream gene (Gene 1), then the transcription of Gene 2 only occurs if the polymerase backtracks back to the

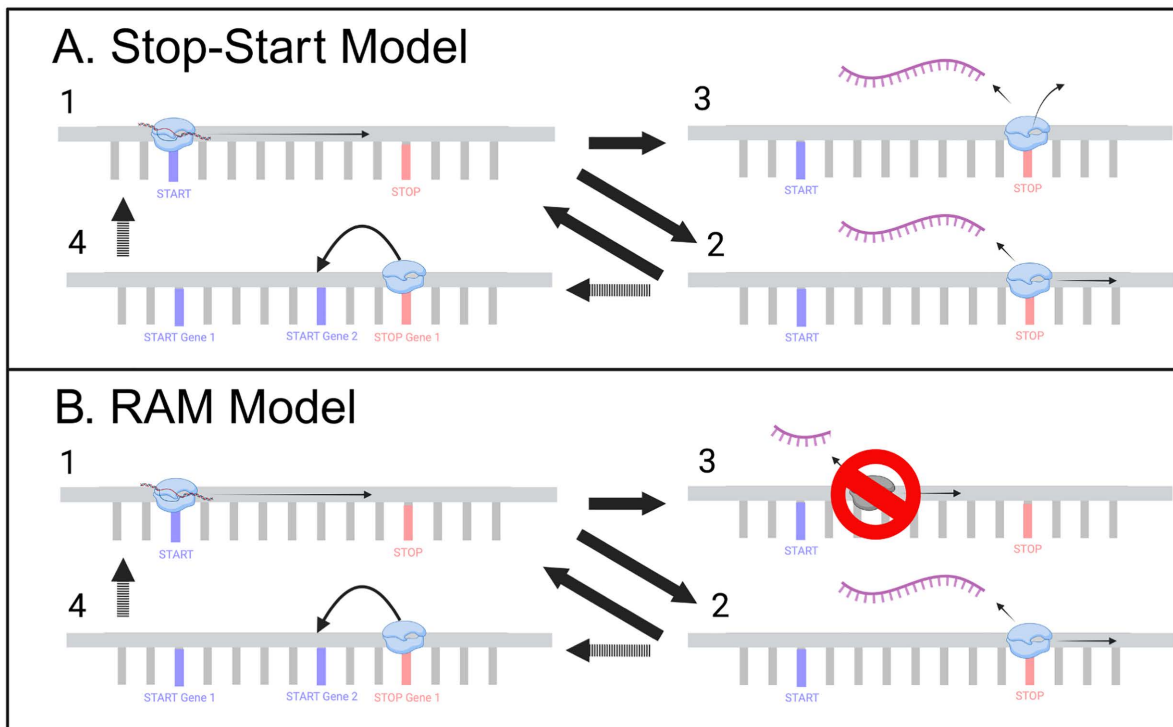


Fig 3. Diagram of Stop-Start and RAM Models. (A.) Stop-Start Model and (B.) RAM Model. In both models transcription begins at the 3' most end of the genome (state 1) where the polymerase recognizes the start signal and initiates transcription. The polymerase then proceeds down the genome to the termination signal (state 2) and releases the synthesized transcript. For standard gene structures, the polymerase reinitiates at the next downstream start signal (state 1). If there is an overlap, the polymerase must move backwards to state 4 before proceeding to state 1. The models differ in where the attenuation we observe comes from. In the RAM model, the polymerase slips and releases the nascent transcript, entering a state that can no longer recognize transcription start signals (state 3). In the Stop-Start model, the polymerase dissociates from the genome at the junction (state 3). Although this too may instead be due to some sort of conformational change into an inactive polymerase conformation at these junctions. These two models are not necessarily mutually exclusive. Both processes are likely to be occurring. Created in BioRender. King, C. (2026) <https://BioRender.com/z9eob9v>.

<https://doi.org/10.1371/journal.pcbi.1014441.g003>

start signal of Gene 2 after finishing transcription of Gene 1 [20,32]. We characterize this using one additional parameter called $p_{\text{backtrack}}$. In these cases p_{express} is calculated using this formula:

$$p_{\text{express}} = (p_{\text{maintain}})^n \times p_{\text{backtrack}}$$

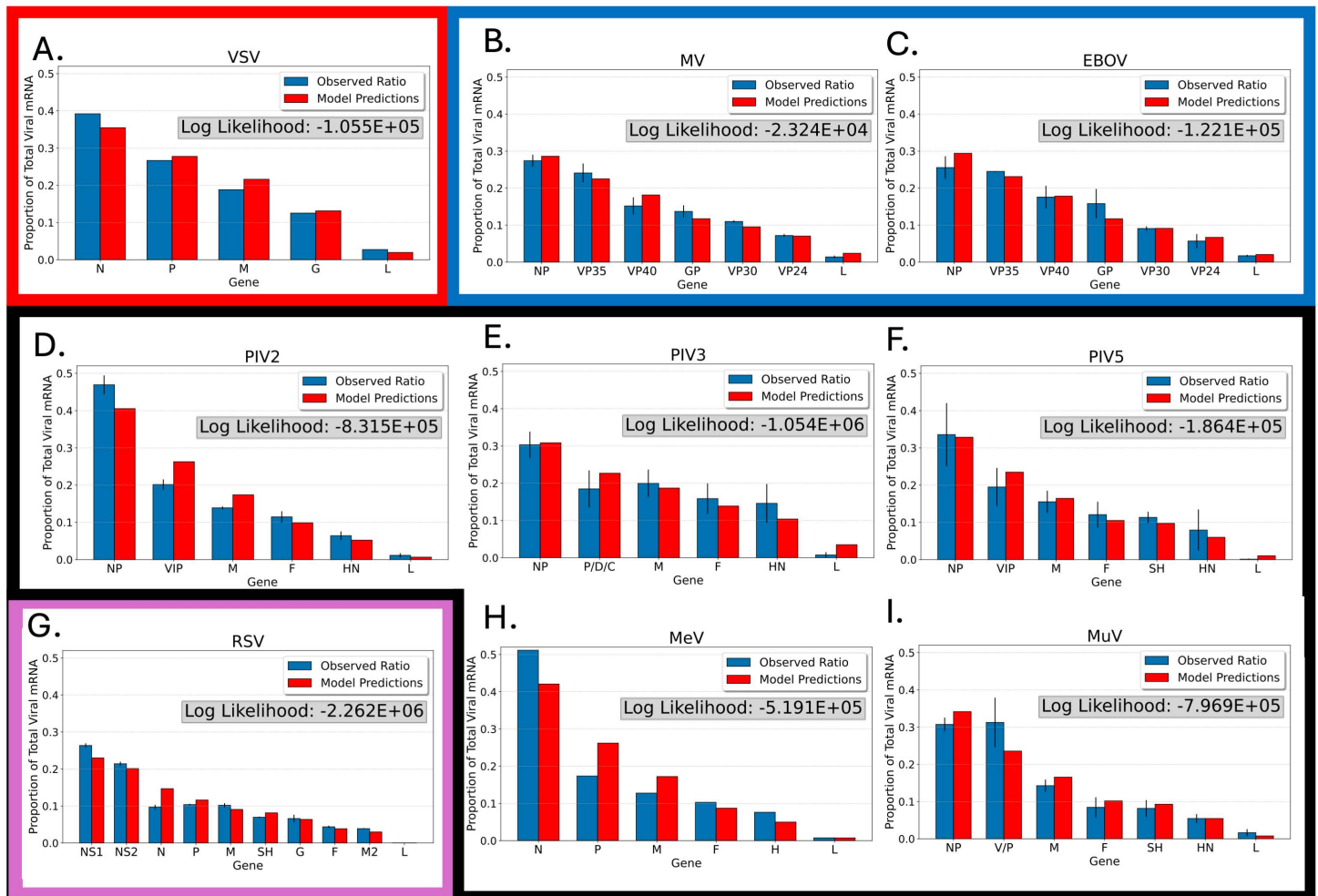
The parameters p_{maintain} , p_{drop} and $p_{\text{backtrack}}$ were determined by the fitting the model to maximize the likelihood of the data given the model under the assumption that their values were between 0 and 1, and posterior uncertainties in parameters were quantified using Markov Chain Monte Carlo (MCMC) algorithm (S2 Table) [33,34]. The data used in this article are very diverse, including sequencing data from Illumina sequencing and direct RNA sequencing as well as blotting-based methodologies. The code for fitting the models can be found in Notebook1 on the GitHub associated with this article.

The RAM model fits the transcriptional gradient of diverse NNSVs

The RAM model was fit to the observed gradients from 9 NNSVs of the *Pneumoviridae*, *Paramyxoviridae*, *Filoviridae*, and *Rhabdoviridae* families (Fig 4, S2 Table). For *Rhabdoviridae*, the model was fit to the original dataset used to describe the ratio of VSV expression in 1981 [8]. For *Filoviridae*, EBOV and MV were used [35]. For *Pneumoviridae*, RSV was used [36]. For *Paramyxoviridae*, data from MeV, MuV, PIV2, PIV3, and PIV5 were used [13,37]. It is important to note that the

Rhabdoviridae

Filoviridae



Pneumoviridae

Paramyxoviridae

Fig 4. The RAM Model Accurately Predicts Transcriptional Gradients Across NNSV Species. Model predictions of the transcriptional gradient of NNSVs using the RAM model (red bars) compared to published data (blue bars). These predictions are across predominant families of NNSVs: *Rhabdoviridae* (A), *Filoviridae* (B and C), and *Paramyxoviridae* (D-F, H, I), and *Pneumoviridae* (G). Black error bars are 4 standard deviations in length, or were excluded if standard deviations were not available for the data. The log likelihood is included for every model.

<https://doi.org/10.1371/journal.pcbi.1014441.g004>

published data associated with the transcriptional behavior of RSV are inconsistent. Some articles propose an entirely different gradient behavior of RSV with several inconsistencies when compared to the characteristic simple gradient [32,38]. In this article, the RSV gradient that most closely resembled the characteristic NNSV gradient was selected, as this is the one that is most consistent with the current understanding of NNSV transcription in literature.

The RAM model was able to achieve high-quality fits for all these viruses (Fig 4). The confidence in parameter p_{maintain} associated with each virus are well-determined (S2 Table). In general, the $p_{\text{backtrack}}$ parameter is well-determined in the viruses that have overlaps in their genomes while it appears underdetermined in those that do not (S2 Table). This is not surprising as models of viruses without overlaps never actually utilize this parameter. As such, in these scenarios, this

model can be reduced to a single parameter. In viruses that have overlapping genes, in order for the downstream gene of an overlapping junction to be transcribed the polymerase must backtrack to the transcriptional start signal upon completion of transcribing the upstream gene. This should lead to additional attenuation that comes from the parameter $p_{\text{backtrack}}$ in these viruses. However, the junctions in each virus appears to behave substantially different from one another with RSV demonstrating the largest amount of additional attenuation (94.15-94.33% more) attributed to the overlap while MV and EBOV had very little if any attenuation (1.6-20.05% more and 0-.065% more respectively). In RSV, it has been shown that part of this additional attenuation is due to premature stoppage of transcription at the internal stop signal [39]. It is possible that since the overlap in MV and EBOV are less dramatic than in RSV (Figs 2A, S1 Fig.) this effect does not occur in these viruses. Alternatively, it could be due to the larger distance the polymerase must travel.

Genome size is correlated to polymerase processivity

Since p_{drop} is the probability of the polymerase releasing the nascent transcript at any point, it can be viewed as a type of polymerase processivity. Intriguingly, there appears to be a negative correlation between the value of p_{drop} and the length of the genome (Fig 5). This correlation suggests that for every base increase in the genome, the probability of the polymerase releasing the transcript at every step down the genome decreases by about $1.79 \times 10^{-8} \pm 1.72 \times 10^{-8}$ (p -value = .044, t -test statistic = 2.45). This observation is reasonable as all polymerases must enter the 3' end of the genome and so the polymerase must stay active for longer stretches in order for genes at the 5' end of longer viral genomes to be expressed.

The RAM model accurately predicts gene expression in reorganized genomes

Historically, VSV has been studied using “gene-shuffled” variants where the position of genes is shifted to leverage the transcriptional gradient to predictably alter transcription rates [40,41]. Here, the RAM model and the Stop-Start model were both evaluated on their ability to predict how shuffling the position of VSV’s genes alters the observed expression ratio of the genes [41,42]. The Stop-Start model used here is model with 4 parameters, one for each junction. The amount of attenuation expected at each junction is essentially the amount of attenuation observed in the training dataset. The available data is all at the protein level, however, it is generally understood that the gradient in the expression of VSV genes at the transcript level also occurs at the protein level, with some potential to deviate due to other mechanisms that may be controlling protein expression [40–42].

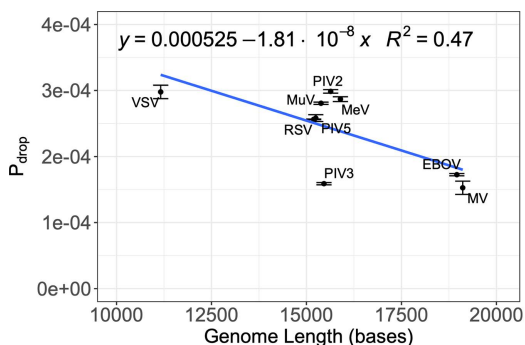


Fig 5. The Processivity of Viral Polymerases is Correlated to Genome Length. This figure shows a linear model showing how the probability of the polymerase releasing the nascent transcript at each step (P_{drop}) decreases as the length of the genome increases. The estimated slope ($-1.81 \times 10^{-8} \pm 1.70 \times 10^{-8}$) and intercept ($5.25 \times 10^{-4} \pm 2.72 \times 10^{-4}$) are both statistically significant ($t(7) = -2.50$, $p = .041$ and $t(7) = 4.58$, $p = .003$ respectively). The dots and error bars are the estimates and 95% confidence intervals respectively for the value of p_{drop} for each virus.

<https://doi.org/10.1371/journal.pcbi.1014441.g005>

Both models were first fit to the Iverson and Rose VSV transcriptional gradient data [8] using MCMC. Then, to evaluate the models parameters were drawn from the posterior distributions. Each of these draws was used to simulate the models and the average of these predictions was used to calculate the Mean Absolute Error between the predictions and observed expression levels for gene-shuffled variants from two datasets [41,42]. Both models do well predicting the relative expression levels (Fig 6A and B) and show a statistically significant correlation between the predicted expression levels. This is consistent with the previous statement that the expression gradient is carried on to the protein level and may occur at a nearly 1:1 level. However, there do appear to be some deviations between predicted and observed expression levels. It is possible that this is due to differing environmental factors, such as the degradation rate [43,44].

When looking at the error in these two models, while it is not statistically significant there is a very weak benefit from using the RAM model which has an MAE of 0.0501 as opposed to the Stop-Start model which has an MAE of 0.0545 (Fig 6C). It is likely that this is since the Stop-Start model is much more susceptible to overfitting since the number of variables is equal to the number of gene junctions (4) while the RAM model relies on only 1 parameter.

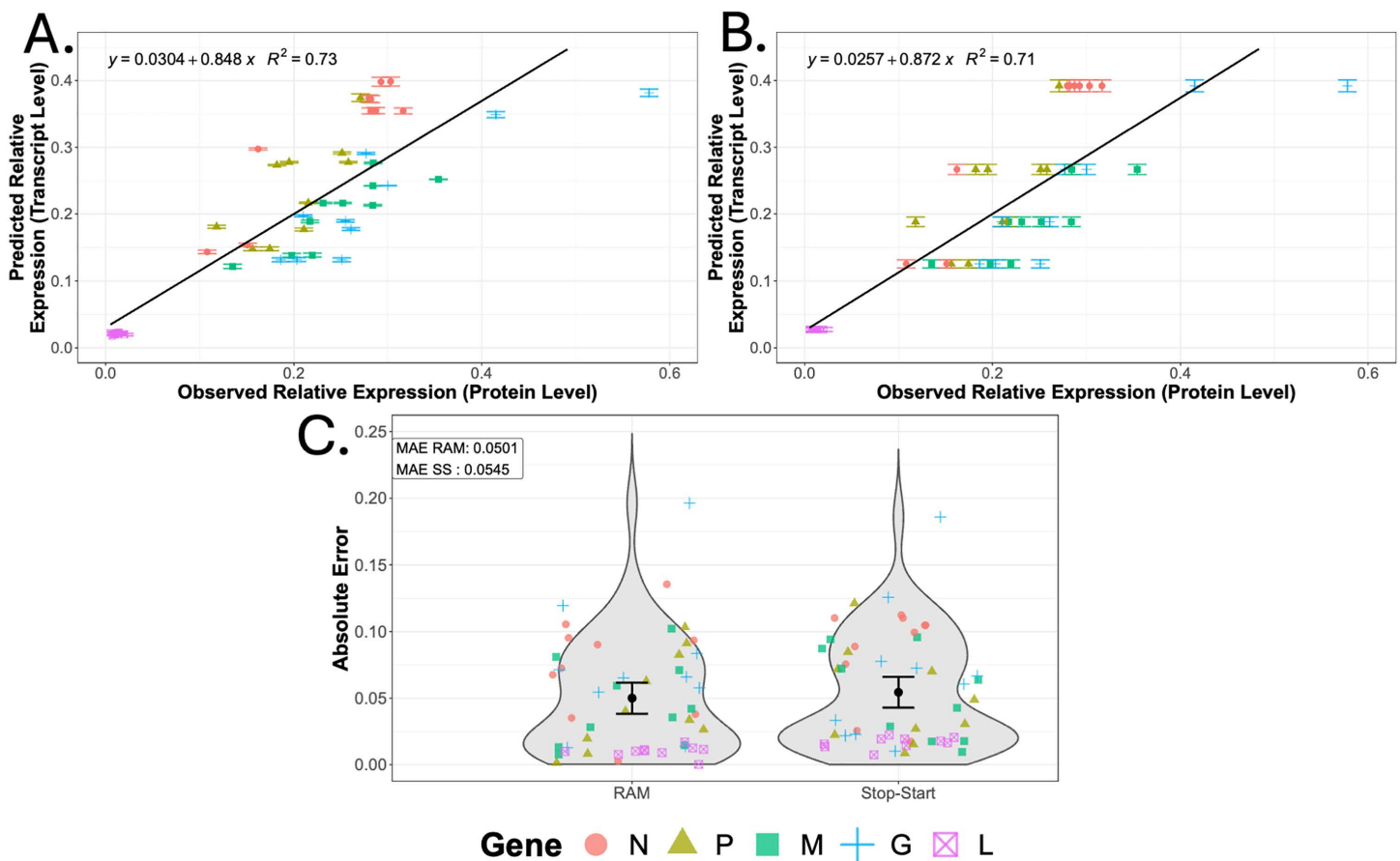


Fig 6. The RAM Model Accurately Predicts Gradients Across Gene-Shuffled VSV Variants. This figure shows the linear fits for the predicted relative mRNA transcript levels in different VSV gene-shuffled variants for the (A.) RAM Model and (B.) Stop-Start Model (y-axis) compared to the observed protein expression levels of experimentally constructed gene-shuffled VSV variants from 2 previous publications, Ball et al. 10.1128/jvi.73.6.4705-4712.1999 and Flanagan et al. 10.1128/jvi.74.17.7895-7902.2000 (x-axis). The slopes of both the RAM model (0.848 ± 0.152) and the Stop-Start model (0.872 ± 0.161) are statistically significant ($t(48) = 11.25$, $p = 4.67 \times 10^{-15}$ and $t(48) = 10.89$, $p = 1.46 \times 10^{-14}$). C. Shows the distribution of absolute errors between predicted and observed values between both models along with the Mean Absolute Error. The differences between these errors is not statistically significant ($t(49) = -1.74$, $p = 0.087$).

<https://doi.org/10.1371/journal.pcbi.1014441.g006>

Discussion

The model proposed here provides a simplified, quantitative approximation of the transcriptional gradient in NNSVs without explicitly incorporating junction-specific regulation. The RAM model can recapitulate the transcriptional gradient in a diverse array of viruses using only one (and at most two) parameters (Fig 4) providing a more parsimonious model as opposed to the Stop-Start model which has one parameter for each gene junction. This simplicity lends the model to be less susceptible to overfitting, the effects of which can be seen in the minor advantage the model has over the Stop-Start model when predicting the relative expression of genes in gene-shuffled VSV variants (Fig 6C).

The RAM model's robustness across diverse datasets ranging from traditional methods that use ³H-uridine to quantify viral RNAs [8] to next-generation sequencing data [35] further supports its broad applicability. The model presented in this article is effectively a successor to the non-specific attenuation model that had been previously discarded in the early 1980s [8]. By showing that a simple polymerase-based stochastic process can reproduce the transcriptional gradient across multiple NNSVs, our results suggest that these observed polymerase dynamics may represent a conserved underlying constraint in viral transcription. A result supported by the correlation between polymerase processivity and genome length (Fig 5).

Despite the benefits of the RAM model, there are limitations. The RAM model inherently does not capture all aspects of NNSV transcription. There are many cases where the gene expression gradient is strongly impacted by the gene junctions, such as rabies virus [21] and RSV [18,22]. These observations are more consistent with the Stop-Start model and highlight why the Stop-Start model is still a central model in NNSV research. Rather than replacing the Stop-Start model, the RAM model represents a complementary explanation that captures baseline attenuation and encourages the investigation of nonspecific mechanisms also known to be at work, such as premature termination due to failed 5' capping [17,24–26].

However, open questions remain. One of the major implications of the RAM model is the generation of truncated transcripts. Such transcripts have a complicated history in the NNSV transcriptions space. On one hand, Iverson and Rose used the lack of observed truncated transcripts as evidence for refuting a nonspecific attenuation mechanism [8]. On the other, truncated transcript have been observed in RSV [27] and when studying the role of 5' capping in [17,24–26]. It is unclear if these discrepancies are due to biological differences in the viruses or methodological differences in these studies. It also should be noted that in most studies the transcript levels observed are the steady-state levels. So, in some cases the absence of truncated transcripts could be due to reduced stability from a lack of 5' caps and/or polyadenylated tails [13,45]. Another alternative interpretation is that this is an artifact of the assumption that there is a uniform probability of attenuation across the genome. Rather, the dissociation of the polymerase from the transcript may be limited to specific stages in transcription. For example, it may be that the phenomenon associated with this is in fact the aforementioned 5' capping-dependent transcriptional termination mechanism. However, why the amount of attenuation would be correlated with the length of the gene is unclear.

To resolve these discrepancies, we advocate moving beyond minigenome assays toward large-scale variant libraries. While this was extremely challenging during these initial studies, it is now possible to leverage homology-based methods to generate large panels of variants [46,47]. Gene-shuffled and junction-swapped constructs, analyzed by qPCR or RNA-seq (in the absence of poly-A pulldowns), could reveal the individual contributions of gene order, junction sequences, and polymerase behavior. Such data would allow for a more rigorous understanding of NNSV transcription and help discern what mechanisms are at play. However, we should not neglect the fact that it is highly likely that both the specific and nonspecific mechanisms discussed in this article participate together in the regulation of transcription and therefore the emergence of a transcriptional gradient in NNSVs.

Overall, this work brings attention back to the potential for a per-base attenuation mechanism that may also be at work in the NNSV transcriptional gradient. At minimum, this model provides a useful null control for the assumed gradient for any NNSV with an uncharacterized gradient. At best, this brings back the idea that there should be some revisitation to a

per-base nonspecific attenuation mechanism that was discarded very early on in the study of NNSV transcription. Such per-base attenuation could be exceedingly useful in the biomedical community as it may allow for improved calibration of viral transcriptional profiles in therapies. For example, it has been a major point of focus to understand how the alteration of transcription rates influences the fitness of the VSV, primarily to attenuate it for medical applications [41,44,48]. Our model suggests new design strategies beyond gene shuffling (namely, gene overlaps and gene length modulation). Considering overlapping genes as binary variables yields 1,920 possible VSV variants. Incorporating gene-length changes, which might influence transcription rates and are feasible due to VSV's tolerance for large genome insertions [49–51], dramatically expands the design space. Together, gene order, overlap, and length could provide unprecedented control over viral transcription. This would enable not just understanding of viral biology but also the rational design of attenuated or customized NNSV variants with finely tuned expression profiles.

Materials and methods

Identification of gene junctions

Gene junctions were identified in reference genomes by using literature reported gene junctions as a guide (S1 Table) [15,22,23,37,52–56]. Junctions were then aligned in SnapGene using MUSCLE. Reference genomes, junctions, and alignments can be found in Folder4 on GitHub.

Parameter estimation

To estimate parameters, the log likelihood for observing the data based on the model-predicted ratios was maximized. The minimize function from SciPy was utilized. Each model was fit 10,000 times starting at a value between 0 and 1. The parameters chosen for downstream analyses were the ones associated with the largest log likelihood. Confidence in parameters was assessed using MCMC algorithm [33,34]. The code for fitting models and performing MCMC can be found in Notebook1 on GitHub. The code for then making predictions for gene shuffled variants can be found in Notebook2.

Linear regression

Linear regression was performed in R using the base lm function. The performance package was used to test model assumptions of normality, homogeneity of variance, and linearity. No violations were observed. The code for fitting the linear regression model connecting polymerase processivity and genome length can be found in the Markdown1 file on GitHub. The code for fitting the linear regression model connecting RAM model predicted expression and observed expression for gene shuffled variants can be found in the Markdown2 file on GitHub.

Supporting information

S1 Fig. Alignments of Gene Junctions for 9 NNSVs. Each junction has blue and red boxes highlighting the conserved transcriptional start and stop sequences respectively. Black stars to the left of the sequence indicate that the junction is an overlapping junction and so the locations of the start and stop sequences are inverted compared to the others.

(TIFF)

S1 Table. Literature Reported Gene Junction Consensus Sequences.

(PDF)

S2 Table. Best Fit Parameter Along with Confidence Statistics from MCMC for RAM Model.

(PDF)

S3 Table. GenBank Accession Numbers for Nonsegmented Negative-Stranded RNA Virus Genomes.

(PDF)

Author contributions

Conceptualization: Connor R. King.

Data curation: Connor R. King.

Formal analysis: Connor R. King.

Funding acquisition: Jean Peccoud.

Methodology: Connor R. King, Brian Munsky.

Project administration: Jean Peccoud.

Software: Connor R. King.

Supervision: Jean Peccoud.

Writing – original draft: Connor R. King, Casey-Tyler Berezin.

References

1. Bukreyev A, Skiadopoulos MH, Murphy BR, Collins PL. Nonsegmented negative-strand viruses as vaccine vectors. *J Virol.* 2006;80(21):10293–306. <https://doi.org/10.1128/JVI.00919-06> PMID: [17041210](https://pubmed.ncbi.nlm.nih.gov/17041210/)
2. Abraham G, Banerjee AK. Sequential transcription of the genes of vesicular stomatitis virus. *Proc Natl Acad Sci U S A.* 1976;73(5):1504–8. <https://doi.org/10.1073/pnas.73.5.1504> PMID: [179088](https://pubmed.ncbi.nlm.nih.gov/179088/)
3. Reddy DV, Schnitzlein WM, Reichmann ME. Use of defective interfering particle RNA probes in the determination of the order of in vitro transcription of vesicular stomatitis virus genes. *J Virol.* 1977;21(1):432–4. <https://doi.org/10.1128/JVI.21.1.432-434.1977> PMID: [189078](https://pubmed.ncbi.nlm.nih.gov/189078/)
4. Ball LA, White CN. Order of transcription of genes of vesicular stomatitis virus. *Proc Natl Acad Sci U S A.* 1976;73(2):442–6. <https://doi.org/10.1073/pnas.73.2.442> PMID: [174107](https://pubmed.ncbi.nlm.nih.gov/174107/)
5. Mudd JA, Summers DF. Protein synthesis in vesicular stomatitis virus-infected HeLa cells. *Virology.* 1970;42(2):328–40. [https://doi.org/10.1016/0042-6822\(70\)90277-1](https://doi.org/10.1016/0042-6822(70)90277-1) PMID: [4321299](https://pubmed.ncbi.nlm.nih.gov/4321299/)
6. Villarreal LP, Breindl M, Holland JJ. Determination of molar ratios of vesicular stomatitis virus induced RNA species in BHK21 cells. *Biochemistry.* 1976;15(8):1663–7. <https://doi.org/10.1021/bi00653a012> PMID: [178352](https://pubmed.ncbi.nlm.nih.gov/178352/)
7. Pennica D, Lynch KR, Cohen PS, Ennis HL. Decay of vesicular stomatitis virus mRNAs in vivo. *Virology.* 1979;94(2):484–7. [https://doi.org/10.1016/0042-6822\(79\)90480-x](https://doi.org/10.1016/0042-6822(79)90480-x) PMID: [222066](https://pubmed.ncbi.nlm.nih.gov/222066/)
8. Iverson LE, Rose JK. Localized attenuation and discontinuous synthesis during vesicular stomatitis virus transcription. *Cell.* 1981;23(2):477–84. [https://doi.org/10.1016/0092-8674\(81\)90143-4](https://doi.org/10.1016/0092-8674(81)90143-4) PMID: [6258804](https://pubmed.ncbi.nlm.nih.gov/6258804/)
9. Keene JD, Chien IM, Lazzarini RA. Vesicular stomatitis virus defective interfering particle containing a muted internal leader RNA gene. *Proc Natl Acad Sci U S A.* 1981;78(4):2090–4. <https://doi.org/10.1073/pnas.78.4.2090> PMID: [6264463](https://pubmed.ncbi.nlm.nih.gov/6264463/)
10. Mühlberger E. Filovirus Replication and Transcription. *Future Virol.* 2007;2: 205–15. <https://doi.org/10.2217/17460794.2.2.205>
11. Brauburger K, Boehmann Y, Krähling V, Mühlberger E. Transcriptional Regulation in Ebola Virus: Effects of Gene Border Structure and Regulatory Elements on Gene Expression and Polymerase Scanning Behavior. *J Virol.* 2015;90(4):1898–909. <https://doi.org/10.1128/JVI.02341-15> PMID: [26656691](https://pubmed.ncbi.nlm.nih.gov/26656691/)
12. Edwards MR, Vogel OA, Mori H, Davey RA, Basler CF. Marburg Virus VP30 Is Required for Transcription Initiation at the Glycoprotein Gene. *mBio.* 2022;13(5):e0224322. <https://doi.org/10.1128/mbio.02243-22> PMID: [35997284](https://pubmed.ncbi.nlm.nih.gov/35997284/)
13. Wignall-Fleming EB, Hughes DJ, Vattipally S, Modha S, Goodbourn S, Davison AJ, et al. Analysis of Paramyxovirus Transcription and Replication by High-Throughput Sequencing. *J Virol.* 2019;93(17):e00571-19. <https://doi.org/10.1128/JVI.00571-19> PMID: [31189700](https://pubmed.ncbi.nlm.nih.gov/31189700/)
14. Hoffman MA, Banerjee AK. Precise mapping of the replication and transcription promoters of human parainfluenza virus type 3. *Virology.* 2000;269(1):201–11. <https://doi.org/10.1006/viro.2000.0223> PMID: [10725212](https://pubmed.ncbi.nlm.nih.gov/10725212/)
15. He B, Lamb RA. Effect of inserting paramyxovirus simian virus 5 gene junctions at the HN/L gene junction: analysis of accumulation of mRNAs transcribed from rescued viable viruses. *J Virol.* 1999;73(8):6228–34. <https://doi.org/10.1128/JVI.73.8.6228-6234.1999> PMID: [10400712](https://pubmed.ncbi.nlm.nih.gov/10400712/)
16. Zengel J, Pickar A, Xu P, Lin A, He B. Roles of Phosphorylation of the Nucleocapsid Protein of Mumps Virus in Regulating Viral RNA Transcription and Replication. *J Virol.* 2015;89(14):7338–47. <https://doi.org/10.1128/JVI.00686-15> PMID: [25948749](https://pubmed.ncbi.nlm.nih.gov/25948749/)
17. Ogino M, Green TJ, Ogino T. GDP polyribonucleotidyltransferase domain of vesicular stomatitis virus polymerase regulates leader-promoter escape and polyadenylation-coupled termination during stop-start transcription. *PLoS Pathog.* 2022;18(2):e1010287. <https://doi.org/10.1371/journal.ppat.1010287> PMID: [35108335](https://pubmed.ncbi.nlm.nih.gov/35108335/)

18. Hardy RW, Harmon SB, Wertz GW. Diverse gene junctions of respiratory syncytial virus modulate the efficiency of transcription termination and respond differently to M2-mediated antitermination. *J Virol.* 1999;73(1):170–6. <https://doi.org/10.1128/JVI.73.1.170-176.1999> PMID: [9847319](https://pubmed.ncbi.nlm.nih.gov/9847319/)
19. Griffiths C, Drews SJ, Marchant DJ. Respiratory Syncytial Virus: Infection, Detection, and New Options for Prevention and Treatment. *Clin Microbiol Rev.* 2017;30(1):277–319. <https://doi.org/10.1128/CMR.00010-16> PMID: [27903593](https://pubmed.ncbi.nlm.nih.gov/27903593/)
20. Kleiner VA, Fearn R. How does the polymerase of non-segmented negative strand RNA viruses commit to transcription or genome replication?. *J Virol.* 2024;98(8):e0033224. <https://doi.org/10.1128/jvi.00332-24> PMID: [39078194](https://pubmed.ncbi.nlm.nih.gov/39078194/)
21. Finke S, Cox JH, Conzelmann KK. Differential transcription attenuation of rabies virus genes by intergenic regions: generation of recombinant viruses overexpressing the polymerase gene. *J Virol.* 2000;74(16):7261–9. <https://doi.org/10.1128/jvi.74.16.7261-7269.2000> PMID: [10906180](https://pubmed.ncbi.nlm.nih.gov/10906180/)
22. Kuo L, Fearn R, Collins PL. The structurally diverse intergenic regions of respiratory syncytial virus do not modulate sequential transcription by a dicistronic minigenome. *J Virol.* 1996;70(9):6143–50. <https://doi.org/10.1128/JVI.70.9.6143-6150.1996> PMID: [8709239](https://pubmed.ncbi.nlm.nih.gov/8709239/)
23. Neumann G, Watanabe S, Kawaoka Y. Characterization of Ebolavirus regulatory genomic regions. *Virus Res.* 2009;144(1–2):1–7. <https://doi.org/10.1016/j.virusres.2009.02.005> PMID: [19481829](https://pubmed.ncbi.nlm.nih.gov/19481829/)
24. Ogino T. Capping of vesicular stomatitis virus pre-mRNA is required for accurate selection of transcription stop-start sites and virus propagation. *Nucleic Acids Res.* 2014;42(19):12112–25. <https://doi.org/10.1093/nar/gku901> PMID: [25274740](https://pubmed.ncbi.nlm.nih.gov/25274740/)
25. Neubauer J, Ogino M, Green TJ, Ogino T. Signature motifs of GDP polyribonucleotidyltransferase, a non-segmented negative strand RNA viral mRNA capping enzyme, domain in the L protein are required for covalent enzyme-pRNA intermediate formation. *Nucleic Acids Res.* 2016;44(1):330–41. <https://doi.org/10.1093/nar/gkv1286> PMID: [26602696](https://pubmed.ncbi.nlm.nih.gov/26602696/)
26. Li J, Rahmeh A, Morelli M, Whelan SPJ. A conserved motif in region v of the large polymerase proteins of nonsegmented negative-sense RNA viruses that is essential for mRNA capping. *J Virol.* 2008;82(2):775–84. <https://doi.org/10.1128/JVI.02107-07> PMID: [18003731](https://pubmed.ncbi.nlm.nih.gov/18003731/)
27. Liuzzi M, Mason SW, Cartier M, Lawetz C, McCollum RS, Dansereau N, et al. Inhibitors of respiratory syncytial virus replication target cotranscriptional mRNA guanylation by viral RNA-dependent RNA polymerase. *J Virol.* 2005;79(20):13105–15. <https://doi.org/10.1128/JVI.79.20.13105-13115.2005> PMID: [16189012](https://pubmed.ncbi.nlm.nih.gov/16189012/)
28. Bloyet L-M, Brunel J, Dosnon M, Hamon V, Erales J, Gruet A, et al. Modulation of Re-initiation of Measles Virus Transcription at Intergenic Regions by PXD to NTAIL Binding Strength. *PLoS Pathog.* 2016;12(12):e1006058. <https://doi.org/10.1371/journal.ppat.1006058> PMID: [27936158](https://pubmed.ncbi.nlm.nih.gov/27936158/)
29. Sala FA, Ditter K, Dybkov O, Urlaub H, Hillen HS. Structural basis of Nipah virus RNA synthesis. *Nat Commun.* 2025;16(1):2261. <https://doi.org/10.1038/s41467-025-57219-5> PMID: [40050611](https://pubmed.ncbi.nlm.nih.gov/40050611/)
30. Stillman EA, Whitt MA. Mutational analyses of the intergenic dinucleotide and the transcriptional start sequence of vesicular stomatitis virus (VSV) define sequences required for efficient termination and initiation of VSV transcripts. *J Virol.* 1997;71(3):2127–37. <https://doi.org/10.1128/JVI.71.3.2127-2137.1997> PMID: [9032346](https://pubmed.ncbi.nlm.nih.gov/9032346/)
31. Zhang J, Landick R. A Two-Way Street: Regulatory Interplay between RNA Polymerase and Nascent RNA Structure. *Trends Biochem Sci.* 2016;41(4):293–310. <https://doi.org/10.1016/j.tibs.2015.12.009> PMID: [26822487](https://pubmed.ncbi.nlm.nih.gov/26822487/)
32. Piedra F-A, Henke D, Rajan A, Muzny DM, Doddapaneni H, Menon VK, et al. Modeling nonsegmented negative-strand RNA virus (NNSV) transcription with ejective polymerase collisions and biased diffusion. *Front Mol Biosci.* 2023;9:1095193. <https://doi.org/10.3389/fmolb.2022.1095193> PMID: [36699700](https://pubmed.ncbi.nlm.nih.gov/36699700/)
33. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics.* 1953;21(6):1087–92. <https://doi.org/10.1063/1.1699114>
34. Aguilera LU, Weber LM, Ron E, King CR, Öcal K, Poppinga A, et al. Methods in quantitative biology—from analysis of single-cell microscopy images to inference of predictive models for stochastic gene expression. *Phys Biol.* 2025;22(4):042001. <https://doi.org/10.1088/1478-3975/adda85> PMID: [40388970](https://pubmed.ncbi.nlm.nih.gov/40388970/)
35. Shabman RS, Jabado OJ, Mire CE, Stockwell TB, Edwards M, Mahajan M, et al. Deep sequencing identifies noncanonical editing of Ebola and Marburg virus RNAs in infected cells. *mBio.* 2014;5(6):e02011. <https://doi.org/10.1128/mBio.02011-14> PMID: [25370495](https://pubmed.ncbi.nlm.nih.gov/25370495/)
36. Donovan-Banfield I, Milligan R, Hall S, Gao T, Murphy E, Li J, et al. Direct RNA sequencing of respiratory syncytial virus infected human cells generates a detailed overview of RSV polycistronic mRNA and transcript abundance. *PLoS One.* 2022;17(11):e0276697. <https://doi.org/10.1371/journal.pone.0276697> PMID: [36355791](https://pubmed.ncbi.nlm.nih.gov/36355791/)
37. Cattaneo R, Rebmann G, Schmid A, Baczko K, ter Meulen V, Billeter MA. Altered transcription of a defective measles virus genome derived from a diseased human brain. *EMBO J.* 1987;6(3):681–8. <https://doi.org/10.1002/j.1460-2075.1987.tb04808.x> PMID: [3582370](https://pubmed.ncbi.nlm.nih.gov/3582370/)
38. Piedra F-A, Qiu X, Teng MN, Avadhanula V, Machado AA, Kim D-K, et al. Non-gradient and genotype-dependent patterns of RSV gene expression. *PLoS One.* 2020;15(1):e0227558. <https://doi.org/10.1371/journal.pone.0227558> PMID: [31923213](https://pubmed.ncbi.nlm.nih.gov/31923213/)
39. Fearn R, Collins PL. Model for polymerase access to the overlapped L gene of respiratory syncytial virus. *J Virol.* 1999;73(1):388–97. <https://doi.org/10.1128/JVI.73.1.388-397.1999> PMID: [9847343](https://pubmed.ncbi.nlm.nih.gov/9847343/)
40. Ball LA, Pringle CR, Flanagan B, Perepelitsa VP, Wertz GW. Phenotypic consequences of rearranging the P, M, and G genes of vesicular stomatitis virus. *J Virol.* 1999;73(6):4705–12. <https://doi.org/10.1128/JVI.73.6.4705-4712.1999> PMID: [10233930](https://pubmed.ncbi.nlm.nih.gov/10233930/)
41. King CR, Berezin C-T, Peccoud J. Stochastic model of vesicular stomatitis virus replication reveals mutational effects on virion production. *PLoS Comput Biol.* 2024;20(2):e1011373. <https://doi.org/10.1371/journal.pcbi.1011373> PMID: [38324583](https://pubmed.ncbi.nlm.nih.gov/38324583/)

42. Flanagan EB, Ball LA, Wertz GW. Moving the glycoprotein gene of vesicular stomatitis virus to promoter-proximal positions accelerates and enhances the protective immune response. *J Virol.* 2000;74(17):7895–902. <https://doi.org/10.1128/jvi.74.17.7895-7902.2000> PMID: [10933697](https://pubmed.ncbi.nlm.nih.gov/10933697/)
43. Tran NHN, Clayton AHA. Transfer function approach to understanding periodic forcing of signal transduction networks. *Phys Biol.* 2023;20(3):10.1088/1478-3975/acc300. <https://doi.org/10.1088/1478-3975/acc300> PMID: [36893467](https://pubmed.ncbi.nlm.nih.gov/36893467/)
44. Lim K-I, Lang T, Lam V, Yin J. Model-based design of growth-attenuated viruses. *PLoS Comput Biol.* 2006;2(9):e116. <https://doi.org/10.1371/journal.pcbi.0020116> PMID: [16948530](https://pubmed.ncbi.nlm.nih.gov/16948530/)
45. Barr JN, Tang X, Hinzman E, Shen R, Wertz GW. The VSV polymerase can initiate at mRNA start sites located either up or downstream of a transcription termination signal but size of the intervening intergenic region affects efficiency of initiation. *Virology.* 2008;374(2):361–70. <https://doi.org/10.1016/j.virol.2007.12.023> PMID: [18241907](https://pubmed.ncbi.nlm.nih.gov/18241907/)
46. Hernandez SI, Peccoud SJ, Berezin C-T, Peccoud J. Self-documenting plasmids. *Synthetic Biology.* 2024. <https://doi.org/10.1101/2024.10.29.620927>
47. Moles CM, Basu R, Weijmarshausen P, Ho B, Farhat M, Flaas T, et al. Leveraging Synthetic Virology for the Rapid Engineering of Vesicular Stomatitis Virus (VSV). *Viruses.* 2024;16(10):1641. <https://doi.org/10.3390/v16101641> PMID: [39459973](https://pubmed.ncbi.nlm.nih.gov/39459973/)
48. Lim K, Yin J. Computational fitness landscape for all gene-order permutations of an RNA virus. *PLoS Comput Biol.* 2009;5(2):e1000283. <https://doi.org/10.1371/journal.pcbi.1000283> PMID: [19197345](https://pubmed.ncbi.nlm.nih.gov/19197345/)
49. Lichty BD, Power AT, Stojdl DF, Bell JC. Vesicular stomatitis virus: re-inventing the bullet. *Trends Mol Med.* 2004;10(5):210–6. <https://doi.org/10.1016/j.molmed.2004.03.003> PMID: [15121047](https://pubmed.ncbi.nlm.nih.gov/15121047/)
50. Geisbert TW, Daddario-Dicaprio KM, Lewis MG, Geisbert JB, Grolla A, Leung A, et al. Vesicular stomatitis virus-based ebola vaccine is well-tolerated and protects immunocompromised nonhuman primates. *PLoS Pathog.* 2008;4(11):e1000225. <https://doi.org/10.1371/journal.ppat.1000225> PMID: [19043556](https://pubmed.ncbi.nlm.nih.gov/19043556/)
51. Overend C, Yuan L, Peccoud J. The synthetic futures of vesicular stomatitis virus. *Trends Biotechnol.* 2012;30(10):497–8. <https://doi.org/10.1016/j.tibtech.2012.06.002> PMID: [22789133](https://pubmed.ncbi.nlm.nih.gov/22789133/)
52. Rose JK. Complete intergenic and flanking gene sequences from the genome of vesicular stomatitis virus. *Cell.* 1980;19(2):415–21. [https://doi.org/10.1016/0092-8674\(80\)90515-2](https://doi.org/10.1016/0092-8674(80)90515-2) PMID: [6244108](https://pubmed.ncbi.nlm.nih.gov/6244108/)
53. Elango N, Varsanyi TM, Kövamees J, Norrby E. Molecular cloning and characterization of six genes, determination of gene order and intergenic sequences and leader sequence of mumps virus. *J Gen Virol.* 1988;69 (Pt 11):2893–900. <https://doi.org/10.1099/0022-1317-69-11-2893> PMID: [3183630](https://pubmed.ncbi.nlm.nih.gov/3183630/)
54. Kawano M, Okamoto K, Bando H, Kondo K, Tsurudome M, Komada H, et al. Characterizations of the human parainfluenza type 2 virus gene encoding the L protein and the intergenic sequences. *Nucleic Acids Res.* 1991;19(10):2739–46. <https://doi.org/10.1093/nar/19.10.2739> PMID: [1645865](https://pubmed.ncbi.nlm.nih.gov/1645865/)
55. Spriggs MK, Collins PL. Human parainfluenza virus type 3: messenger RNAs, polypeptide coding assignments, intergenic sequences, and genetic map. *J Virol.* 1986;59(3):646–54. <https://doi.org/10.1128/JVI.59.3.646-654.1986> PMID: [3016327](https://pubmed.ncbi.nlm.nih.gov/3016327/)
56. Feldmann H, Mühlberger E, Randolph A, Will C, Kiley MP, Sanchez A, et al. Marburg virus, a filovirus: messenger RNAs, gene order, and regulatory elements of the replication cycle. *Virus Res.* 1992;24(1):1–19. [https://doi.org/10.1016/0168-1702\(92\)90027-7](https://doi.org/10.1016/0168-1702(92)90027-7) PMID: [1626422](https://pubmed.ncbi.nlm.nih.gov/1626422/)