Cut site preference allows influenza A virus PA-X to discriminate between host and viral mRNAs

Lea Gaucherand, Amrita Iyer, Isabel Gilabert, Chris Rycroft, Marta Gaglia

HHS Public Access

Author manuscript

Nat Microbiol. Author manuscript; available in PMC 2024 January 01.

Published in final edited form as:

Nat Microbiol. 2023 July ; 8(7): 1304–1317. doi:10.1038/s41564-023-01409-8.

Cut site preference allows influenza A virus PA-X to discriminate between host and viral mRNAs

Lea Gaucherand[1,2,7], **Amrita Iyer**[2], **Isabel Gilabert**[2,3], **Chris H. Rycroft**[4,5,6], **Marta M. Gaglia**[1,2,8]

[1]Program in Molecular Microbiology, Tufts University Graduate School of Biomedical Sciences, Boston, MA, United States.

[2]Department of Molecular Biology and Microbiology, Tufts University School of Medicine, Boston, MA, United States.

[3]Faculty of Experimental Sciences, Universidad Francisco de Vitoria, Madrid, Spain.

[4]John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, United States.

[5]Computational Research Division, Lawrence Berkeley National Laboratory, Berkeley, CA, United States.

[6]Department of Mathematics, University of Wisconsin - Madison, Madison, WI, United States

[7]Present address: Architecture et Réactivité de l'ARN, Institut de Biologie Moléculaire et Cellulaire du CNRS, Université de Strasbourg, Strasbourg, France

[8]Institute for Molecular Virology and Department of Medical Microbiology and Immunology, University of Wisconsin – Madison, Madison WI, United States

Abstract

Many viruses block host gene expression to take over the infected cell. This process, termed host shutoff, is thought to promote viral replication by preventing antiviral responses and redirecting cellular resources to viral processes. Several viruses from divergent families accomplish host shutoff through RNA degradation by endoribonucleases. However, viruses also need to ensure expression of their own genes. The influenza A virus endoribonuclease PA-X solves this problem by sparing viral mRNAs and some host RNAs necessary for viral replication. To understand how

Corresponding author: Marta M. Gaglia, marta.gaglia@wisc.edu.

Author Contributions Statement

Conceptualization, L.G. and M.M.G.; Methodology, L.G., C.H.R. and M.M.G.; Investigation, L.G., A.I. and I.G.; Writing – Original Draft, L.G. and M.M.G.; Writing – Review & Editing, A.I., I.G. and C.H.R.; Funding Acquisition, M.M.G.; Supervision, M.M.G.

Competing Interests Statement

The authors declare no competing interests.

Code availability

All python scripts used, including the PyDegradome 2.0 code can be found in our laboratory's GitHub page[71] (https://github.com/mgaglia81/PyDegradome). The workflow explaining the use of the PyDegradome 2.0 code can be found in the methods section, under "PyDegradome and other python analyses". The method of Reed and Muench[44] was used to calculate viral titers via the Bloom lab Python script at https://github.com/jbloom/reedmuenchcalculator[45]. The RNA structure prediction software LinearFold[53] is now available at https://linearfold.eecs.oregonstate.edu/[72]. WebLogo[57] is made available by the Computational Genomics Research Group (University of California, Berkeley) at https://weblogo.berkeley.edu/logo.cgi[58].

PA-X distinguishes between RNAs, we characterized PA-X cut sites transcriptome-wide using 5’ rapid amplification of cDNA ends (5’ RACE) coupled to high-throughput sequencing. This analysis, along with RNA structure predictions and validation experiments using reporters, shows that PA-Xs from multiple influenza strains preferentially cleave RNAs at GCUG tetramers in hairpin loops. Importantly, GCUG tetramers are enriched in the human but not the influenza transcriptome. Moreover, optimal PA-X cut sites inserted in the influenza A virus genome are quickly selected against during viral replication in cells. This finding suggests that PA-X evolved these cleavage characteristics to preferentially target host over viral mRNAs, in a manner reminiscent of cellular self vs. non-self discrimination.

Influenza A virus drives widespread host RNA depletion through its endoribonuclease (endoRNase) PA-X[1–4]. Immune modulation by PA-X *in vivo* has been widely documented. Infection with a PA-X-deficient virus induces more inflammation in animal models, surprisingly often causing higher morbidity and mortality than wild-type infections[3,5–8]. However, few studies have investigated PA-X’s mechanism of action, especially PA-X RNA targeting preference and its contribution to PA-X function during infection.

PA-X is produced from segment 3 of the negative-sense segmented influenza A virus genome, which also encodes the polymerase acidic (PA) subunit of the influenza RNA-dependent RNA-polymerase (FluPol), through a +1 ribosomal frameshift after amino acid 191[3,9]. PA-X and PA thus share an N-terminal domain, which has a PD-D/E-XK nuclease fold and endoRNase activity[10,11]. PA uses it to “snatch” capped ends of host mRNAs to initiate viral mRNA transcription[10], whereas PA-X to do host shutoff[4]. PA-X activity is not indiscriminate, as it down-regulates host RNAs synthesized by RNA Polymerase II (RNAPII) but not RNAPI and III[12]. This specificity may stem from a connection between PA-X and cellular RNA splicing[4]. However, unspliced reporter RNAs can also be down-regulated to some extent. Additionally, unlike some viral endoRNases, PA-X does not consistently down-regulate viral mRNAs[12], suggesting it can distinguish between host and viral transcripts. Viral RNA levels are unaffected by PA-X in cells infected with the influenza A/PuertoRico/8/1934 (H1N1) virus (“PR8”) strain[12], although some change was reported with A/California/04/2009 (H1N1)[13]. Splicing alone does not explain the virus-host discrimination, as influenza mRNAs M and NS are spliced but spared by PA-X. Therefore, there are additional unknown components to PA-X selectivity.

Knowing the cut site preference is critical to understand which RNAs are directly targeted by RNases and how, but cut site preference remains poorly understood for many endoRNases. The complexity of cut site motifs often hampers their identification using classical *in vitro* approaches. High-throughput sequencing approaches have emerged to profile endoRNase cleavage products in cells, identifying complex preferred motifs for *Salmonella* RNase E[14] and for Kaposi’s sarcoma-associated herpes virus (KSHV) SOX[15].

We have previously used RNA steady state levels to report on PA-X activity, but this approach may be confounded by feedback effects on transcription[16] and the inherent instability of certain RNAs. Thus, to better understand PA-X targeting, we directly probed PA-X cut sites throughout the transcriptome using 5’ Rapid Amplification of cDNA Ends adapted to high-throughput sequencing (5’ RACE-seq) and a custom analysis pipeline,

PyDegradome[15]. We report that PA-X preferentially cleaves host RNAs at GCUG sequences within hairpin loops. Importantly, these preferred cleavage motifs are more abundant in host than viral mRNAs. Moreover, inserting a preferred PA-X cleavage sequence in the virus reduces viral fitness only in the presence of PA-X. These findings suggest that PA-X cut site preference contributes to its ability to broadly target the host transcriptome while sparing the virus, and serves as a "self/non-self" discrimination mechanism for influenza A virus.

Results

Identification of PA-X cut sites transcriptome-wide

To identify PA-X cut sites transcriptome-wide, we used 5' RACE-seq (Fig. 1A) and compared RNA fragments from cells infected with wild-type (WT) PR8, or PR8 engineered to lack PA-X[4] (PR8-PA(ΔX) or ΔX, Extended Data Fig. 1A), or mock infected. The three mutations in PR8-PA(ΔX) introduce a stop codon in the PA-X +1 reading frame and should reduce frameshifting, abolishing production of full-length PA-X[4]. They are silent in the PA 0 reading frame, and do not affect PA protein levels (Extended Data Fig. 1B), but reduce host shutoff compared to WT PR8[4]. Since the host exonuclease Xrn1 degrades RNA fragments following cleavage by PA-X[12], we used Xrn1 knock-out human lung A549 cells to enrich for PA-X fragments[17] (Extended Data Fig. 1C). All samples expressed similar levels of viral genes, and host shutoff was detected in the WT PR8 infected samples (Fig. 1B–C). RNA cleavage fragments, including PA-X-generated fragments, were captured by ligating an RNA adapter to free 5' phosphate, followed by preparation of 5' RACE-seq libraries (Fig. 1A).

Our protocol yielded a high level of background reads, presumably due to the lack of basal RNA degradation in Xrn1 knock-out cells. Libraries had similar numbers of reads (Supplementary Table 1) and ~75% of reads mapped to unique locations. The number of reads that started at the same nucleotide was highly correlated among samples, suggesting some common background RNA degradation (Extended Data Fig. 1D–E). However, select locations had higher numbers of reads in infected cells, some likely due to PA-X cleavage (Extended Data Fig. 1E). To correct for background degradation and identify PA-X cut sites, i.e. fragment locations enriched in WT PR8 compared to PR8-PA(ΔX) infected cells, we used PyDegradome[15] (see methods).

The stringency of cut site definition in PyDegradome depends on two user-defined parameters: confidence level (cl) and multiplicative factor (mf)[15]. Using loose parameters (cl=99%, mf=2) on 16–22 million reads per sample, we identified 592 PA-X cut sites in 517 genes that were shared across three biological replicates (Fig. 1D–E, left, Supplementary Table 2). These sites are likely reproducible PA-X target sites, as a control analysis looking for fragments enriched in PR8-PA(ΔX) over WT PR8 infected cells (Fig. 1D, right) only returned 37 sites. The low number of RNAs cut by PA-X was surprising compared to the >3,000 genes down-regulated by PA-X at steady state levels[4]. To increase sequencing depth, we analyzed the replicates together by concatenating the read files ("concatenated" approach, Fig. 1E). This is valid from a statistical perspective and provides more power, although we lose replicability information. We optimized the parameters by maximizing both the number of "concatenated" cut sites detected and the percentage of the 592 shared cut sites identified in the concatenated analysis (Extended Data Fig. 1F). Using cl=99.99%

and mf=2, we identified 1361 cut sites in 1309 genes in WT PR8 vs. PR8-PA(ΔX) infected cells (Supplementary Table 2), but only 345 cut sites in the control comparison PR8-PA(ΔX) vs. WT PR8.

We carried out several quality control analyses, detailed in Methods and **Supplementary Fig. 1**, that support the idea that our two approaches identified reliable PA-X cleavage site. Given these considerations, we further analyzed the 1361 cut sites from the concatenated analysis (cl=99.99%/mf=2 parameters on all the reads), which give a broader overview of PA-X cut site locations transcriptome-wide, and the 592 cut sites shared between the three replicates (cl=99%/mf=2 parameters on individual replicates), which have a more biologically stringent cutoff and provide information on the preferred characteristics of PA-X cut sites (Supplementary Table 2).

PA-X cleavage is driven by RNA sequences

We validated PA-X cut sites using classical 5' RACE (Fig. 2A). We tested 12 sites and saw that RACE PCRs amplified products of the correct sizes in WT PR8 but not mock or PR8-PA(ΔX)-infected samples for 11, including the 6 examples shown (Fig. 2B, Extended Data Fig. 2A), although WT and PR8-PA(ΔX) samples were similarly infected (Extended Data Fig. 2B).

To test whether RNA sequences drive PA-X cleavage, we introduced 99 base pairs (bp) surrounding PA-X cut sites into a spliced luciferase reporter that is down-regulated by PA-X[4,18] and transfected the reporters in human embryonic kidney (HEK) 293T cells in which Xrn1 was knocked down (ishXrn1) (Fig. 2C, constructs 3–4, Extended Data Fig. 2C). These constructs had similar expression and down-regulation by PA-X as a control luciferase reporter with no additional sequences (Fig. 2C, construct 1), but harbored a new PA-X cut site (Fig. 2D; Extended Data Fig. 2D–E). To note, the control luciferase reporter has an existing cut site, and we obtained similar results with a construct in which we duplicated the 99 bp sequence around this site (Fig. 2C–D, construct 2, Extended Data Fig. 2D–E). We also detected single fragments coming from the luciferase and luciferase + *STOML2* (stomatin like 2) reporters by northern blotting, which does not require PCR and does not carry the risk of selective fragment amplification (Fig. 2E).

Importantly, we detected no cleavage at these sites upon expression of the PA-X D108A catalytic mutant, a PA(fs) mutant that makes PA but not PA-X due to a decrease in frameshifting[4], or host shutoff endoRNases from other viruses (Extended Data Fig. 3A–B). Conversely, the same sequences elicited cleavage by PA-X from the influenza A/Perth/16/2009 (H3N2) ("Perth") and A/Tennessee/1–560/2009 (H1N1) ("H1N1pdm09" for 2009 pandemic H1N1) viruses, representative of currently circulating H3N2 and H1N1 influenza A strains, in transfected cells and during infection (Fig. 2F, Extended Data Fig. 3C–D). The PA-X proteins from PR8, Perth and H1N1pdm09 (Extended Data Fig. 3E) are a good representation of PA-X diversity across human influenza A strains, as they represent the two different subtypes of circulating seasonal human influenza A viruses and have different C terminus lengths (41 aa for H1N1pdm09, 61 aa for PR8/Perth) and half-lives[19]. These experiments indicate that 5' RACE-seq and PyDegradome identified true

PA-X cut sites transcriptome-wide, and that PA-X preferentially cleaves RNAs at discrete sites depending on the RNA sequence across multiple strains.

PA-X cleaves RNA at a preferred sequence and structure

Since our results suggested a sequence-directed cleavage, we examined sequence features specific to PA-X sites and not enriched in fragments identified in the control comparison. These included a depletion in adenosines and an enrichment in GCUG at the cut site, and a modest enrichment for guanidines downstream of it (Fig. 3A–B, Extended Data Fig. 4A–B). Interestingly, most PA-X cut sites contained GCUG or a similar tetramer, although the location of the cut within the tetramer varied (Fig. 3C–D, Extended Data Fig. 4C–D).

Mutating the cut site GCTG to TAGC in the luciferase reporters with the insertions from *BCAP31 (B cell receptor associated protein 31)/STOML2/TUBA1B (tubulin alpha 1b)/YKT6 (YKT6 v-SNARE homolog)* (Fig. 3E, construct 5) prevented efficient cleavage of the inserted sequence, without impacting reporter expression and down-regulation by PA-X (presumably due to PA-X cleavage at the original luciferase sequence; Fig. 3F, Extended Data Fig. 4E–F). Thus, in these sequences, GCUG is necessary for PA-X cleavage, suggesting it is necessary for a strong cut site. However, introduction of GCUG alone was not sufficient for efficient cleavage (Fig. 3E, G, construct 6). This is consistent with PA-X not cleaving at every GCUG, as exemplified by the luciferase reporter, which is cut by PA-X at a specific GCUG, but not at GCUGs 17 and 92 nucleotides downstream of it (Fig. 3H).

Since GCUG alone was not sufficient, we examined the minimal sequence driving PA-X cleavage, which varied depending on the gene (≤15 bp for *STOML2/BCAP31/SLC7A5* vs. >27 bp for *YTK6*, Fig. 4A–B, Extended Data Fig. 5A). Given the lack of additional enriched sequence motifs, we tested a role for RNA secondary structures. Computational predictions indicated that PA-X cut sites (but not control sites enriched in PR8-PA(ΔX) infected cells) are preferentially found in the loops of RNA hairpins (Fig. 4C). To test their importance for cleavage, we disrupted the predicted hairpins in the *YKT6* 51 bp and *STOML2* 15 bp insertion constructs by mutating three C or G bases to break strong G-C bonds in the stems, moving the GCUG cut site to a bulge or paired region (different sequence, different structure, Fig. 4D–E). We then repaired the structure by mutating the complementary G or C bases to recreate the G-C bonds (different sequence, same structure, Fig. 4D–E). Disrupting the hairpin prevented efficient PA-X cleavage within the *YKT6* and *STOML2* sequences, while repairing it restored cleavage in PA-X transfected cells (Fig. 4F–G) and during infection (Extended Data Fig. 5B). These results show that PA-X displays a previously unappreciated layer of RNA targeting preference, targeting RNAs at GCUG tetramers located within hairpin loops.

PA-X cleaves RNAs within exons

Since PA-X activity is linked to splicing[4], we investigated whether PA-X targets pre- or mature mRNAs. PA-X cut sites were almost exclusively found in exons, although 12–16% of reads mapped to introns and 18% of PR8-PA(ΔX) specific fragments from the concatenated analysis mapped to introns (Extended Data Fig. 6A–C). Moreover, our 5’ RACE validation experiments suggested that PA-X cleaves after splicing, as RACE primers

spanning exon-exon junctions only amplified spliced fragments (Extended Data Fig. 6D–F, dark purple arrows). However, this could be a PCR artifact (i.e. amplification of smaller/more abundant spliced fragments over longer/less abundant non-spliced ones). We thus repeated the 5' RACE PCRs using intron-binding reverse primers specific for pre-mRNA, but again failed to detect fragments of expected sizes and saw background fragments instead (Extended Data Fig. 6D–F, light purple arrows). We also cloned the previously validated *STOML2* and *YKT6* 99 bp cut site sequences into the luciferase reporter intron, which did not alter reporter expression or splicing (Extended Data Fig. 7A–B, construct 10). 5' RACE did not amplify any bands that were unique to constructs with insertions and could correspond to cuts within the intron (Extended Data Fig. 7C, top gel, dark green arrow primer in 7A). Like in other cases where no additional cut site was introduced (Fig. 2D, 3G, 4B, 4F–G), we detected the original luciferase cut site using a reverse primer positioned further downstream (Extended Data Fig. 7C, bottom gel, magenta arrow primer in 7A). Mutating this luciferase cut site (GCUG → UAGC) to reduce cleavage also did not reveal an intronic cut site (Extended Data Fig. 7A, D, construct 11). These results suggest that PA-X preferentially cuts mRNAs within exonic sequences, likely after splicing.

GCUG tetramers are more abundant in host transcriptomes

Finally, we wondered why PA-X has evolved these preferred cleavage characteristics, and if this leads to targeting of specific RNAs. Interestingly, GCUG is one of the most abundant tetramers in the human transcriptome, but is less frequent in the genome (Fig. 5A, Extended Data Fig. 8A). As may be expected, mRNAs cut by PA-X have more GCUGs (Fig. 5B), but whether the GCUG preference has any role in targeting specific groups of host mRNAs is unclear. For example, antiviral interferon stimulated gene mRNAs do not contain more GCUGs than other mRNAs (Fig. 5C). However, the GCUG preference may contribute to virus vs. host discrimination. Whereas GCUG is abundant and overrepresented in the human transcriptome, it is far less abundant and underrepresented relative to other tetramers in the mRNAs of the influenza PR8, H1N1pdm09 and Perth strains (Fig. 5A, Extended Data Fig. 8B–E). GCUG is also more frequent in the transcriptomes of common influenza hosts (pigs, chickens, geese and ducks) than of influenza A virus strains isolated from these animals (Fig. 5D). Moreover, the GCUG percentage is higher in the negative viral genomic RNA strand than in the positive viral mRNA strand for most segments (Extended Data Fig. 8B). Structure-wise, computational predictions suggested that GCUGs in influenza transcripts are enriched in paired structures, likely inaccessible to PA-X (Fig. 5E). The few GCUGs predicted in hairpin loops were not cut very efficiently if at all, as we detected only faint bands that mapped close to the PR8 GCUG tetramers and no bands for H1N1pdm09 (Extended Data Fig. 9A–B). These observations are consistent with PA-X not targeting influenza transcripts[12]. Moreover, they indicate that PA-X cleavage may be biased towards host mRNAs and *away* from influenza mRNAs due to the relative representation of GCUG in hairpin loops, which elicit robust and reproducible targeting.

To test this idea experimentally, we inserted the WT *STOML2* 51 bp cut site sequence inside the PR8 spliced M segment (segment 7) (Fig. 5F), or a *STOML2* sequence with GCUG mutated to UAGC to prevent efficient PA-X cleavage (Fig. 3E–F, Extended Data Fig. 4F). We rescued viruses with WT or mutated *STOML2* sequences in both WT PR8 and PR8-

PA(ΔX) backgrounds (Fig. 5G). However, further passaging of the virus led to loss of the inserted WT *STOML2* sequence when PA-X was present (i.e. in the WT PR8 background) but not in its absence (i.e. in the PR8-PA(ΔX) background). This was apparent in the cDNA from infected A549 cells (Fig. 5H) and in viruses from supernatants collected 48 hours post infection (Fig. 5I). While there was no obvious difference in growth between WT PR8 and PR8-PA(ΔX) with or without the two inserts (Extended Data Fig. 9C), the WT PR8/WT *STOML2* virus had almost completely reverted to a virus without the insertion (Fig. 5I). These results were replicated with three separate rescues of the *STOML2* recombinant viruses. These findings suggest that during viral propagation, viruses that have lost the PA-X cleavage site have a fitness advantage, i.e that a preferred PA-X cut site inside a spliced viral segment is detrimental for the virus. The fact that the sequence was consistently retained in the absence of PA-X (i.e. in the PR8-PA(ΔX) background) and/or when PA-X cleavage was abolished (i.e. mutated *STOML2*) confirms that PA-X cleavage, not the insertion itself, is the problem. Overall, these results are consistent with PA-X evolving preferred cleavage characteristics to target host over viral mRNAs.

Discussion

Our results showing a preference for PA-X cleavage of GCUG sequences in hairpin loops are consistent with an *in vitro* study showing that PA-X cuts single-stranded RNA more efficiently[20]. However, this study did not identify any sequence preference, likely because these characteristics can be missed *in vitro* without prior knowledge, highlighting the advantage of our technique. Interestingly, while PA and PA-X share the same RNase domain, studies suggest they have slightly different cut site preferences[20–24], probably due to their different C-terminal domains and constraints on cleavage location from FluPol binding to mRNA caps. We did not see cap-snatching fragments in our dataset, likely because mRNA decapping in the nucleus leads to RNA degradation by the Xrn2 RNase[25], which is still present in our cells. It would be interesting to use Xrn2 knock out/down cells to examine PA specificity.

A testimony to the importance of the PA-X sequence and structure preference is its conservation across multiple influenza A strains. To note, although PA-Xs from different strains have different host shutoff activity[26–29], differences in cleavage efficiency between strains may be obscured in our results by the 5’ RACE PCRs, which make this method sensitive but not very quantitative. Nonetheless, PA-X preferentially recognizes a ubiquitous motif, likely a strategy to target a broad range of host mRNAs. Supporting this idea, we reported a similar preference for a degenerate recognition motif for SOX from KSHV[15]. Additionally, not only PA-X and SOX, but also the human endoRNase MCPIP1/Regnase-1[30], prefer cutting RNA within hairpin loop structures. Targeting exposed loops may be an easy way for endoRNases to access single-stranded RNA. Studying the cut site characteristics of other viral and human endoRNases will be important to test this idea.

How the molecular activity of host shutoff factors is linked to immunomodulation and how it is regulated to prevent inhibition of viral gene expression are key outstanding questions in the field. While we saw no link between PA-X cleavage preference and immunomodulation, we found that the preference may contribute to PA-X ability to spare viral mRNAs from

degradation. We propose that the targeting preference biases the activity of PA-X towards cutting host mRNA more efficiently than viral mRNA, essentially acting as a self vs. non-self discrimination mechanism for influenza A virus. Self/non-self discrimination is common in cellular responses to nucleic acids, as the cell needs to distinguish its nucleic acids from those of invading pathogens to prevent inappropriate immune activation. While common discrimination features include RNA modification and double-stranded RNA structures, the presence of specific sequences guide viral RNA degradation by the ZAP protein[31] and activation of Toll-like receptor TLR9[32]. Our results suggest that influenza A virus is using a similar process in reverse, identifying cellular RNAs as non-self based on the abundance of specific sequences.

It is intriguing that we only found 500–1,400 cut sites but that more than half of all host RNAs are down-regulated upon PA-X expression[4]. Although our follow-up analyses suggest that we identified true PA-X target sites and revealed preferred sequence characteristics, we cannot conclude that these are the only RNAs directly cut by PA-X, because of technical limitations (see below). However, it is worth considering whether PA-X needs to cleave thousands of RNAs to trigger widespread gene down-regulation. Degradation of mRNAs for key signaling molecules and transcription factors, as well as feedback inhibition of transcription associated with viral RNase activity[16,33,34], may be major contributors to global host shutoff. A decrease in transcription has been observed during influenza infection, although it is partly driven by another influenza protein, non-structural protein 1[35,36]. Future studies will need to decouple RNA degradation from a decrease in transcription to determine their relative contribution to host shutoff. They will also need to define whether degradation of specific RNAs, the number of degraded RNAs and/or the speed of degradation are important. Inhibition of transcription could also dampen the induction of antiviral RNAs when PA-X is present[4], even though few antiviral RNAs are directly cut by PA-X according to our analysis.

With 5’ RACE-seq and our updated PyDegradome pipeline, we present a streamlined way to identify and characterize endoRNase cut sites throughout the transcriptome. This method can uncover cut site preference characteristics that are masked *in vitro* and directly identifies cleaved RNAs, instead of relying on changes in RNA steady state levels, which can come from both degradation and transcription repression. Our study highlights the importance of determining the cleavage preference of RNases to understand their functional role, as our methodology was crucial to reveal the potential for viral self vs. non-self discrimination mechanisms. We hope that others will adapt this method to study other RNases.

Limitations of our study

A major limitation of our study is the large number of potential PA-X cut sites that were only detected in a single replicate, and what this implies for PA-X specificity. One possibility is that these sites are not cut by PA-X, but represent variations in background RNA degradation between samples, a source of noise that PyDegradome cannot readily distinguish from the desired signal. Alternatively, these additional sites could be PA-X cleavage locations that cross the significance threshold only in some replicates due to low RNA levels. We are presently refining our pipeline with these technical limitations in mind.

Lastly, PA-X may cut these additional RNAs, but at lower efficiency because they are a poorer match to the preferred cut characteristics. Regardless, our follow up work supports the idea that the cleavage characteristics we identified predict robust cleavage by PA-X. All the potential explanations for the large number of additional unique sites are still consistent with the model that PA-X has a preference for certain sequences and structures, and that this preference biases the cleavage towards host/non-viral-self RNAs and thus protect the virus from detrimental self-degradation.

Methods

Correspondence and requests for materials should be addressed to Marta M. Gaglia.

Plasmids—pCR3.1-PA-X-myc, pCR3.1-PA-X-D108A-myc and pCR3.1-PA-X-H1N1pdm09 were previously generated as described in[12]. Briefly, they were generated by amplifying sequences from pHW-193 and pSJ560-Tennessee2009-PA, kind gifts from Dr. R. Webby (St Jude's Children Research Hospital, Memphis, TN, USA), and introducing single nucleotide deletions to mimic the frameshift and a point mutation for the D108A mutation. pCR3.1-PA(fs)-myc (all sequences from the PR8 strain) was a kind from Dr. Craig McCormick (Dalhousie University, Halifax, NS, Canada) and was generated as described in[37]. pCDEF3-SOX and pCDNA3.1-vhs were kind gifts of Dr. Britt Glaunsinger (University of California, Berkeley, CA, USA) and were previously described in[38,39]. The luciferase constructs with and without the β-globin intron were a kind gift from Dr. Gideon Dreyfuss[18] (University of Pennsylvania, Philadelphia, PA, USA). Gibson cloning using HiFi assembly mix (New England Biolabs, cat# E5520S) was used to make all other constructs, unless otherwise stated. PR8 pHW-PA(ΔX) plasmid was generated as previously described[4] from pHW-193 by introducing mutations to reduce the level of frameshift and a premature stop codon in the PA-X +1 frame. pCR3.1-PA-X-Perth-myc was generated by PCR amplifying pHW-Perth09-PA, a kind gift from Dr. S. Lakdawala[40] (Emory University School of Medicine, Atlanta, GA, USA) while removing one nucleotide at the frameshift sequence. Fragments were then introduced into the pCR3.1-PA-X-myc vector digested with SalI and MluI to excise PR8 PA-X. pHW-Perth-PA(ΔX) was generated from pHW-Perth09-PA by inserting the ΔX mutations into PCR primers and reintroducing the fragments into pHW-Perth09-PA digested with NheI and BamHI. The same strategy was used to generate pSJ560-TN/CA/7-PA(ΔX) from NheI digested pSJ560-TN/CA/7-PA, a kind gift from Dr. R. Webby (St Jude's Children Research Hospital, Memphis, TN, USA). 99 base pair DNA sequences from genes selected for validation were amplified from human RNA using SuperScript IV One-STEP RT-PCR (Thermo Fisher Scientific, cat# 12594025), or PCR amplified from human cDNA using Vent Polymerase (New England Biolabs, cat# M0254), then introduced into the luciferase with an intron construct at the EcoRI site. The GCTG cut site was mutated to TAGC using the QuickChange site directed mutagenesis kit (Agilent, cat# 200519). 51 base pair constructs were generated by PCR amplification of their respective 99 base pair constructs, then inserted into the luciferase EcoRI site by DNA ligation with T4 DNA ligase (New England Biolabs, cat# M0202). 27 base pair constructs were generated by annealing primers together and inserting them into the luciferase EcoRI site by DNA ligation with T4 DNA ligase. 15 base pair constructs were generated by introducing the new 15 base pair sequence into the reverse primer

and PCR amplifying the rest of the luciferase sequence between the HindIII and EcoRI sites. The hairpin loop structure mutants were generated by inserting targeted mutations into primers and amplifying the cut site sequences with these primers with overlapping sequences to assemble the inserts back to the HindIII/EcoRI digested luciferase vector. The 99 base pair validation sequences were introduced into the β-globin intron of the luciferase construct by PCR amplification of the upstream luciferase sequence from the HindIII site, the 99 base pair sequence, and the downstream luciferase sequence up to the EcoRI site, all with overlapping sequences to assemble the fragments. The WT *STOML2* 51 bp cut site sequence was introduced into pHW-197-PR8-M at the end of each coding sequence by PCR amplification of the upstream M sequence from the BamHI site, the *STOML2* 51 bp sequence, a repeat of the full packaging sequence, and the downstream plasmid sequence up to the BstEII site, all with overlapping sequences to assemble the fragments. The construct containing the mutant *STOML2* 51 bp cut site in was generated from the WT sequence containing construct by mutating the GCTG sequence at the cut site to TAGC using the QuickChange site directed mutagenesis kit (Agilent, cat# 200519). All primers used for cloning are listed in Supplementary Table 3.

Cell lines and transfections—Human embryonic kidney HEK293T cells and Madin-Darby Canine Kidney (MDCK) cells were commercially obtained (ATCC cat# CRL-3216 and CCL-34, respectively). Wild type (ATCC cat# CCL-185) and Xrn1 knock out human adenocarcinoma alveolar basal epithelial (A549) cells were a kind gift from Dr. Bernard Moss[17] (National Institute of Health, Bethesda, MD, USA). HEK293T and MDCK cells are female and A549 cells are male. All cells were maintained in Dulbecco's modified Eagle's medium (DMEM) high glucose (Gibco, cat# 11965118) supplemented with 10% fetal bovine serum (Hyclone) at 37 °C and 5% CO_2. HEK293T cells expressing inducible shRNA against Xrn1 (ishXrn1) were previously generated using pTRIPZ-shXrn1 (Thermoscientific, clone V2THS_89028/RHS4696–99704634, targeting sequence: TATGGTGAGATATACTATG)[15]. For northern blotting and 5' RACE validation experiments, HEK293T ishXrn1 cells were treated with 1 μg/ml doxycycline (Thermo Fisher Scientific, cat# BP26531) for 3–4 days to induce expression of the shRNA, plated on 6-well or 12-well plates, and transfected with 1000 or 800 ng/ml total DNA (including 62.5 or 50 ng/ml PA-X construct, respectively) using jetPRIME transfection reagent (Polyplus transfection, VWR, cat# 89129–924). Cells were harvested 24 hours after transfection for RNA extraction and purification, or infected overnight as described below and harvested the next day for Extended Data Fig. 5B.

Viruses and infections—Recombinant wild-type influenza A/Puerto Rico/8/1934 (H1N1) ("PR8"), A/Tennessee/1–560/2009 (H1N1) ("H1N1pdm09") and A/Perth/16/2009 (H3N2) ("Perth") viruses, their mutated counterparts PR8-PA(ΔX), H1N1pdm09-PA(ΔX) and Perth-PA(ΔX), and the mutant recombinant PR8 containing the WT or mutated *STOML2* 51 bp cut site in the M segment were generated using the 8-plasmid reverse genetic system[41] as previously described[4,42]. Mutations were introduced as described in the plasmids section above. Plasmids were gifts from Drs. Webby, Bloom and Lakdawala. Viral stocks were propagated in MDCK cells and infectious titers determined by plaque assays in MDCK cells using 1.2% Avicel overlays as previously described[43]. Briefly, confluent

MDCK cells were plated in triplicates, then infected with 10-fold serially diluted viral stocks in low volume for 1 hour. Cells were then washed twice with PBS and incubated for 4 days at 37°C in 5% CO_2 atmosphere with a 1.2% Avicel overlay. Finally, cells were fixed with 4% paraformaldehyde and stained with crystal violet to observe plaques. We realized that some of the WT PR8 and PR8-PA(ΔX) virus preparations were contaminated with mycoplasma, but also validated our 5' RACE-seq results with clean preparations of PR8, H1N1pdm09 and Perth viruses. While we lack a reliable antibody to measure PR8 PA-X, we have previously profiled the transcriptional changes due to these mutations and confirmed that PR8-PA(ΔX) causes reduced host shutoff compared to WT PR8[4]. Influenza infections were performed in DMEM supplemented with 0.5% low endotoxin bovine serum albumin (BSA, Sigma-Aldrich, cat# A1470), referred to as infection media. For 5' RACE, western blotting and qRT-PCR experiments, WT or Xrn1 knock out A549 cells or transfected HEK293T ishXrn1 cells were mock-infected or infected with WT, PA(ΔX) or viruses containing the *STOML2* sequences at a MOI of 1 in a low volume of media for 1 hour, then more infection media supplemented with 0.5 μg/ml TPCK-treated trypsin (Sigma-Aldrich, cat# T1426) was added and cells were incubated for 15 hours at 37 °C in 5% CO_2 atmosphere. Cells were then collected for RNA isolation or preparation of lysates for western blotting. For time course experiments, MDCK cells were mock-infected or infected with WT PR8, PR8-PA(ΔX) or viruses containing the *STOML2* sequences at a MOI of 0.05 in a low volume of media for 1 hour. The inoculum was then removed, cells were washed with PBS, and infection media supplemented with 0.5 μg/ml TPCK-treated trypsin was added. Supernatant and RNA were collected for the 1h time point, while the rest of the cells were incubated at 37 °C in 5% CO_2 atmosphere until collection of supernatant and RNA at 8h, 24h and 48h post infection. Viral titers were quantified by 50% tissue culture infectious dose (TCID50). Briefly, 10 μl of undiluted (for 1h and 8h) or of a 1:100 dilution (for 24h and 48h) of the viral supernatant was serially diluted 1:10 in a 96-well plate containing 90 μl of infection media. 25×10^3 MDCK cells were then added to each well. Plates were incubated for 4 days at 37 °C in 5% CO_2 atmosphere before being scored for cytopathic effects. The method of Reed and Muench[44] was used to calculate viral titers via the Bloom lab Python script at https://github.com/jbloom/reedmuenchcalculator[45]. To test whether the *STOML2* sequence was retained, viral genomic RNA was purified from virus preparations and infected cell supernatant using the QIAamp Viral RNA Mini kit (Qiagen, cat# 52904) following manufacturer's protocol. The RNA was then treated with RNase-Free DNase set (Qiagen, cat# 79254), purified and concentrated using the RNA Clean & Concentrator kit (Zymo Research, cat# R1015). A detailed protocol for this procedure is available on Protocols.io[46] (dx.doi.org/10.17504/protocols.io.36wgqj8wxvk5/v1). The RNA was then reverse transcribed to cDNA using iScript Supermix (Bio-Rad, cat# 170–8841), per manufacturer's protocol. Taq DNA polymerase (New England Biolabs, cat# M0273S) was used to amplify the region around the *STOML2* sequence. PCR products were visualized on a 2% agarose gel, and DNA from bands were sequenced to confirm that there was no mutation in the *STOML2* sequences.

Protein harvesting and western blotting—Cell lysates were prepared using radioimmunoprecipitation assay (RIPA) buffer (50 mM Tris-HCl, pH 7.4, 150 mM NaCl, 2 mM EDTA, 0.5% sodium deoxycholate, 0.1% SDS, 1% NP-40) supplemented with

50 μg/ml phenylmethylsulfonyl fluoride (PMSF; G-Biosciences, cat# 501035662) and cOmplete protease cocktail inhibitor (Roche, cat# 11873580001). 20–100 μg of protein was loaded on an SDS-PAGE gel (Bio-Rad, cat# 456–1084) and transferred onto PVDF membranes (EMD Millipore, cat# IPFL00010), then blocked with 5% milk in phosphate-buffered saline with 0.1% Tween 20 (PBST). Western blots were performed with mouse anti Xrn1 C-1 antibodies (Santa Cruz Biotechnology #sc-165985, 1:500), rabbit anti influenza A virus PA antibodies (GeneTex #125932, 1:1000), or rabbit anti β-tubulin 9F3 antibodies (Cell Signaling Technologies #2128, 1:1000) diluted in 0.5% milk PBTS. Goat anti-mouse and rabbit secondary antibodies were purchased from Southern Biotech (cat# 1031–05 and 4030–05 respectively) and used at 1:5,000 dilution. Western blots were imaged with a Syngene G:Box chemi-XX6 system (GeneSys software version 1.7.2.0).

RNA purification, cDNA generation and qRT-PCR—RNA was extracted and purified using either the RNeasy Plus mini kit (Qiagen, cat# 74134) for 5’ RACE-seq experiments, or the Quick-RNA miniprep kit (Zymo Research, cat# R1055) for all other experiments, following manufacturer’s protocol. RNA was then treated with Turbo DNase (Life Technologies, cat# AM2239) and extracted from the DNase reaction by adding phenol chloroform, centrifuging at 12,000 x g for 5 min, collecting the aqueous layer and precipitating RNA in ethanol for 1 hour at −20 °C. RNA was then pelleted, washed with 75% ethanol, then resuspended in RNase-free water. For quantitative real time PCR (qRT-PCR) experiments, resuspended RNA was reverse transcribed to cDNA using iScript Supermix (Bio-Rad, cat# 170–8841), per manufacturer’s protocol. qRT-PCR was performed using iTaq Universal SYBR Green Supermix (Bio-Rad, cat# 172–5125), on the Bio-Rad CFX Connect Real-Time System qPCR and analyzed with Bio-Rad CFX Manager 3.1 or CFX Maestro 2.0 programs. The primers used are listed in Supplementary Table 3. Northern blotting, 5’ RACE and 5’ RACE-seq specific RNA processing are detailed in the next sections.

Northern blotting—Northern blotting was performed using the NorthernMax kit (Invitrogen, cat# AM1940) solutions. After DNase treatment and phenol chloroform extraction, 5 μg of DNase-treated RNA was separated on a 1.2% agarose gel, then transferred by capillary blotting onto a Biodyne B nylon membrane (ThermoFisher Scientific, cat# 77016). Biotinylated DNA probes against the 3’ UTR of the luciferase reporters were generated by PCR amplification using Taq polymerase (New England Biolabs, cat# M0273S) in the presence of Biotin-16-dUTP (Sigma-Aldrich, cat# 11093070910), see primers in Supplementary Table 3. Northern blots were probed with these biotinylated DNA probes, then blocked with Intercept PBS Blocking Buffer (LI-COR Biosciences, cat# 927–70001) supplemented with 1% SDS. Finally, northern blots were incubated with IRDye 800CW Streptavidin (LI-COR Biosciences, 1:10,000, cat# 926–32230) diluted in Intercept PBS Blocking Buffer supplemented with 1% SDS, then imaged on a LI-COR Odyssey CLx imaging system (Image Studio software version 5.2).

5’ RACE—After DNase treatment and phenol chloroform extraction, the RACE adapter (see Supplementary Table 3) was ligated to 1–5 μg RNA using T4 RNA ligase (Invitrogen, cat# AM2141) for 1 hour at 37 °C or 2 hours at 25 °C. Ligated RNA was reverse transcribed

to cDNA using MMLV RT (Thermo Fisher Scientific, cat# 28025013) per manufacturer's protocol. Taq DNA polymerase (New England Biolabs, cat# M0273S) was used to amplify fragments of interest, if present, in two rounds of nested PCRs using forward primers annealing to the RACE adapter (RACE outer and RACE inner, Supplementary Table 3) and reverse primers annealing within the gene of interest (see Supplementary Table 3). Finally, PCR products were run on a 2% agarose gel containing HydraGreen safe DNA dye (ACTGene, cat# ACT-IDMG04) to visualize the amplified fragments and imaged with a Syngene G:Box chemi-XX6 system (GeneSys software version 1.7.2.0). DNA fragments from gel bands at the expected sizes were extracted and sequenced to confirm their identities. A detailed protocol is available on Protocols.io[47] (dx.doi.org/10.17504/protocols.io.261ge3mzwl47/v1).

5' RACE-seq library preparation and high-throughput sequencing—Our 5' RACE-seq library preparation protocol was inspired by previous PARE-Seq protocols with some changes[14,48,49]. After DNase treatment and phenol chloroform extraction, 3 μg RNA for each sample were mixed with 6 μl of ERCC ExFold RNA spike-in mixes diluted to 1:100 (Invitrogen). Next, ribosomal RNA was removed from each sample using Ribo-Zero Plus kit (Illumina, cat# 20037135). The rest of the RNA was ligated to the RACE-seq adapter (see Supplementary Table 3) using T4 RNA ligase (Invitrogen, cat# AM2141) for 2 hours at 25 °C. The RACE-seq adapter contains a unique molecular identifier (UMI) in order to identify and remove reads that are the result of PCR duplication during the library preparation. Ligated RNA was then reverse transcribed to cDNA with SuperScript III RT (Invitrogen, cat# 18080044) using random primers of different lengths that also include a known sequence adapter (long and short RT primers, Supplementary Table 3). Finally, reads were amplified in two rounds of PCR (six cycles each) to enrich for ligated fragments and add the Illumina adapters, which include library barcodes for multiplexing (PCR1 and PCR2 primers, Supplementary Table 3). Between the two PCR rounds, DNA fragments of 150–400 bp were selected using SPRIselect Reagent (Beckman Coulter, cat# B23317). A detailed protocol is available on Protocols.io[50] (dx.doi.org/10.17504/protocols.io.n2bvj847wgk5/v1). The quality of each library was evaluated using a Fragment Analyzer (Advanced Analytical Technologies, Inc.) at the Tufts University Core Facility - Genomics Core. High-throughput sequencing was carried out by the Tufts Genomics facility on a HiSeq 2500, obtaining single-end 50 nucleotide reads on three total lanes (Supplementary Table 1). Bcl2fastq (Illumina) v2.17.1.14 was used for basecalling and demultiplexing. Fastqc v0.11.9 was used to ensure the quality of the sequencing reads before proceeding with the analysis.

Reads preprocessing and alignment—After labeling each read with its own UMI, cutadapt v3.5[51] was used to trim Illumina adapters and UMI from each read sequence. Trimmed reads were then aligned to the human genome version GRCh38/hg38, the PR8 genome and the ERCC spike sequences using HISAT2 v2.2.1[52] with the following specific parameters: retain unique alignments only (-k 1 option), disallow softclipping (--no-softclip option), and select fr-secondstrand library type (--rna-strandness F option). Finally, reads with the same UMI mapping to the same genome region were collapsed into one, to eliminate read duplications that can arise from PCR amplification. Of note, only limited duplication was observed. Uniquely aligned reads were then used for the PyDegradome

pipeline (Supplementary Table 1). Reads that aligned to the wrong strand (based on their strandness) and contained insertions or deletions were discarded for the PyDegradome analysis.

PyDegradome and other python analyses—PyDegradome is a Python-based peak finding pipeline developed to identify ribonuclease cut sites, which we originally reported in[15] and modified here. Once all reads are aligned to the genome, PyDegradome records the position of the first nucleotide of every read, and counts the number of reads that map to the same position for each library. It then uses a Bayesian probability model to identify which positions have statistically significantly more reads mapping to them in a test sample *vs.* a control sample, i.e. a higher rate of fragment generation. The read counts from each dataset are used to build a prior distribution of the underlying rates of fragment generation at each site. This stringency of the identification significance can be modulated based on three parameters that are optimized by the user: multiplicative factor, confidence level and scanning window. The multiplicative factor determines by how much the read count at a specific nucleotide position in the test sample needs to exceed the count in the control sample. The confidence level sets the cutoff of statistical significance. The multiplicative factor and confidence level together define the threshold that the test sample needs to exceed to be considered significantly above background. The scanning window determines how many consecutive nucleotides need to have an average read count above the threshold value for a peak to be called. This parameter removes isolated high read counts that could skew the analysis. For all the analyses showed here, a window of 4 nucleotides was used. For a more detailed explanation of PyDegradome, see our previous PyDegradome publication[15]. For the current study, we updated our older version of PyDegradome to be compatible with the alignment tool HISAT2 and the human genome annotation GRCh38/hg38. The older version of PyDegradome also restricted the cut site analysis to exons, and here we have updated it to include introns in case PA-X cuts pre-mRNA. The PyDegradome 2.0 updated script can be found in our laboratory's GitHub page (https://github.com/mgaglia81/PyDegradome). The downstream motif generation and scoring analysis was not modified and is still available to download as supplementary material in our previous article[15] or from our GitHub page. RNA structures were predicted by the online software LinearFold[53] using the CONTRAfold v2.0 machine-learning model[54] (LinearFold C) or the Vienna RNAfold thermodynamic model[55,56] (LinearFold V). WebLogo[57] was used to visualize the enriched nucleotide bases at the cut site, available at https://weblogo.berkeley.edu/logo.cgi[58].

Additional quality control analyses for PyDegradome results—We were surprised by the large number of predicted cut sites (comparison: test = WT PR8 vs. control = PR8-PA(ΔX)) that were not shared between biological replicates, even though we noted that the same large number of sites was detected in the control comparison (comparison: test = PR8-PA(ΔX) vs. control = WT PR8) (Fig. 1D). Therefore, we did additional quality control analyses to test whether the sites that are not shared are "missed" PA-X cut sites or noise from technical limitations of the pipeline. First, we tested different analyses parameters to see how they influenced our site detection. Increasing the stringency of the analysis (higher cutoff cl and mf values) increased the percentage of shared sites (Extended Data Fig. 10A). This effect suggests that at least some of the non-shared sites are due to

noise or limitations in the pipeline's ability to distinguish signal from noise, and not low PA-X cleavage specificity. Second, we compared the read counts at the cut sites for sites identified in all three replicates, only in one or two, or by the concatenated approach. Sites identified in all three replicates and by concatenated approach had higher reads counts than the ones identified in only one or two replicates (Extended Data Fig. 10B). This characteristic suggests that we are only identifying the most robust cut sites, and that some of the non-shared sites may also be PA-X cut sites, but their signal is too low in some replicates to be flagged as above background. Because this result suggests the sequencing depth may be limiting, we tested how much we may benefit from additional sequencing depth. We reanalyzed subsets of our data with the optimal parameters for the concatenated approach, with or without concatenating the read files. The number of PA-X cut sites increased with increasing input read depth while the number of "control" fragments enriched in the PR8-PA(ΔX) infected cells did not (Extended Data Fig. 10C). However, the number of detected PA-X sites plateaued around 100 million reads, which is similar to the number of reads we used for our concatenated analysis (Extended Data Fig. 10C). This result suggests that we are detecting specific cut sites and do not need to further increase the sequencing depth.

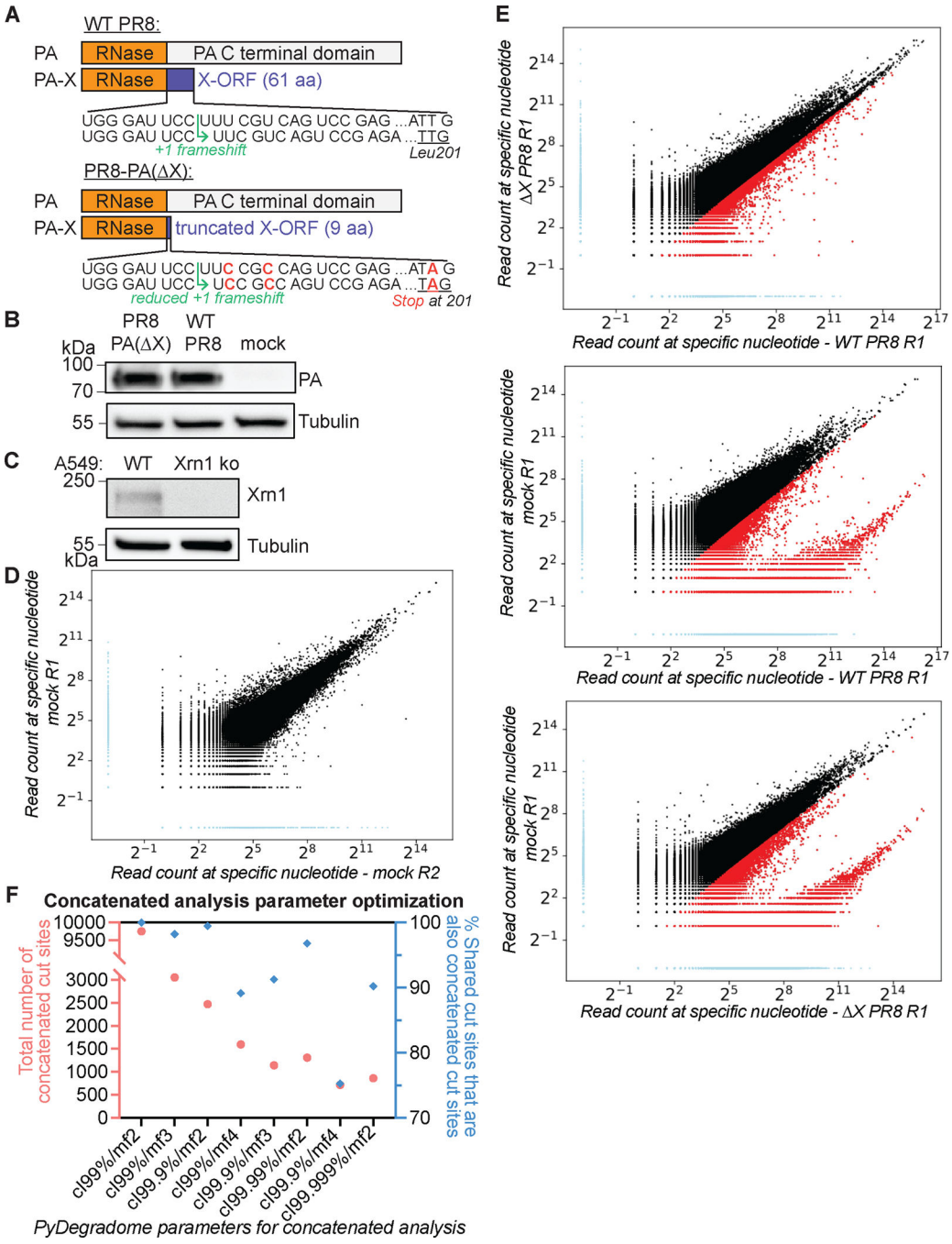
A separate concern is whether the sites we identified are truly cut by PA-X vs. another endoRNase. In particular, PA also can cut host mRNAs. However, our analyses suggest that the cut sites we identified in WT PR8-infected cells are unlikely coming from PA. First of all, PA is equally expressed in the two samples we are comparing, WT PR8 and PR8-PA(ΔX) infected cells (Extended Data Fig. 1B). This result is consistent with previous reports showing that if anything, PA levels are higher in the absence of PA-X[3,5,6,12,59,60]. Moreover, PA cuts RNAs at the very 5' end, but <0.2% of the identified cut sites mapped to the first 20 nucleotides of the transcripts (Extended Data Fig. 10D). These results indicate that PA is not responsible for the identified cut sites.

Analysis of GCUG representation in genomes and transcriptomes—To calculate the percentage of GCUG within transcriptomes, we counted the number of GCUG tetramers found in each RNA sequences and divided this number by the total possible number of tetramers in that RNA, i.e. length of RNA minus 3. To calculate the GCTG percentage in genomes, we counted the number of GCTG / GCUG tetramers in each chromosome / influenza segment and divided this number by the total number of tetramers in that chromosome / influenza segment, i.e. length of chromosome / segment minus 3. The python script used is available on our GitHub page. Transcriptome fasta files were downloaded from the Ensembl[61] ftp server available at https://ftp.ensembl.org/pub/current_fasta/. Swan goose: *Anser cygnoides* GooseV1.0, INSDC Assembly GCA_002166845.1, Jun 2017 by Poultry Science Institute. Pig: *Sus scrofa* Sscrofa11.1, INSDC Assembly GCA_000003025.6, Feb 2017 by the Swine Genome Sequencing Consortium (SGSC)[62,63]. Chicken: *Gallus gallus* bGalGal1.mat.broiler.GRCg7b, INSDC Assembly GCA_016699485.1, Jan 2021 by Vertebrate Genomes Project. Mallard duck: *Anas platyrhynchos* ASM874695v1, INSDC Assembly GCA_008746955.1, Sep 2019 by China Agricultural University. Influenza transcriptomes were downloaded from the NIAID Influenza Research Database (IRD)[64] (previously available at http://www.fludb.org). A/Puerto-Rico/8/34

(H1N1): Genbank CY033577-84, A/Tennessee/05/2009 (H1N1): Genbank GQ160528-31 GQ396729-30 GQ894895 GQ200205, A/Perth/16/2009 (H3N2): Genbank KJ609203-10, A/swine/Guangdong/2722/2011 (H1N1): Genbank KM027548-56[65], A/chicken/Pakistan/UDL-01/2008 (H9N2): Genbank CY038455-62[66], A/tree sparrow/Jiangsu/1/2008 (H5N1): Genbank GQ202207-14[67], A/Anhui/1/2005 (H5N1) HPAI: Genbank HM172438, HM172394, HM172342, HM172104, HM172254, HM172189, HM172159, HM172266[68].

Statistics & Reproducibility—Unless otherwise noted, data are plotted as mean ± standard deviation and represent three or more independent experiments. Statistical analyses were performed using Microsoft Excel (version 16.39) and GraphPad Prism (version 9). One-way or two-way analysis of variance (ANOVA) with Dunnett's multiple comparison test were used for multiple comparisons. Chi-Square Test for Goodness of Fit with degree of freedom 1 was used to compare distributions. Significance was defined as $p < 0.05$. Levels of significance are indicated as follows: * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$. No statistical method was used to predetermine sample size but our sample sizes are similar to those reported in previous publications[4,69]. No data were excluded from the analyses. The experiments were not randomized. Data collection and analysis were not performed blind to the conditions of the experiments. For reporter assays, data distribution was assumed to be normal but this was not formally tested.

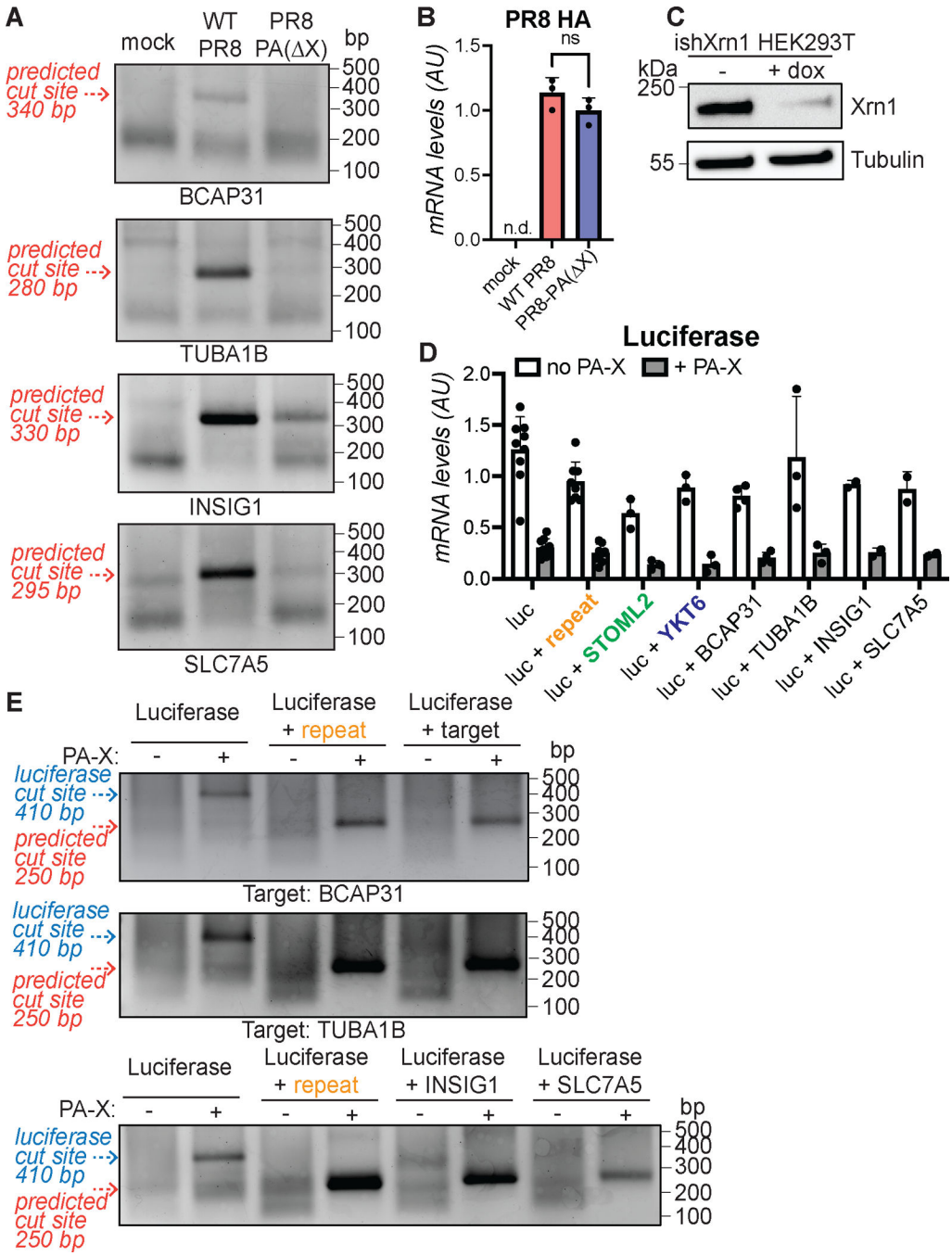
Extended Data



Extended Data Fig. 1. Characteristics of the system used to identify PA-X cut sites transcriptome-wide

(A) Strategy used to engineer a virus that lacks PA-X, PR8-PA(ΔX), compared to WT PR8. Adapted from4. (B) Protein lysates of Xrn1 knock out A549 cells infected with WT PR8 or PR8-PA(ΔX), or mock infected, were probed with antibodies against PR8 PA or b-tubulin as a loading control. Images are representative of 2 experiments. (C) Protein lysates of WT or Xrn1 knock out (ko) A549 cells were probed with antibodies against Xrn1 or b-tubulin as a

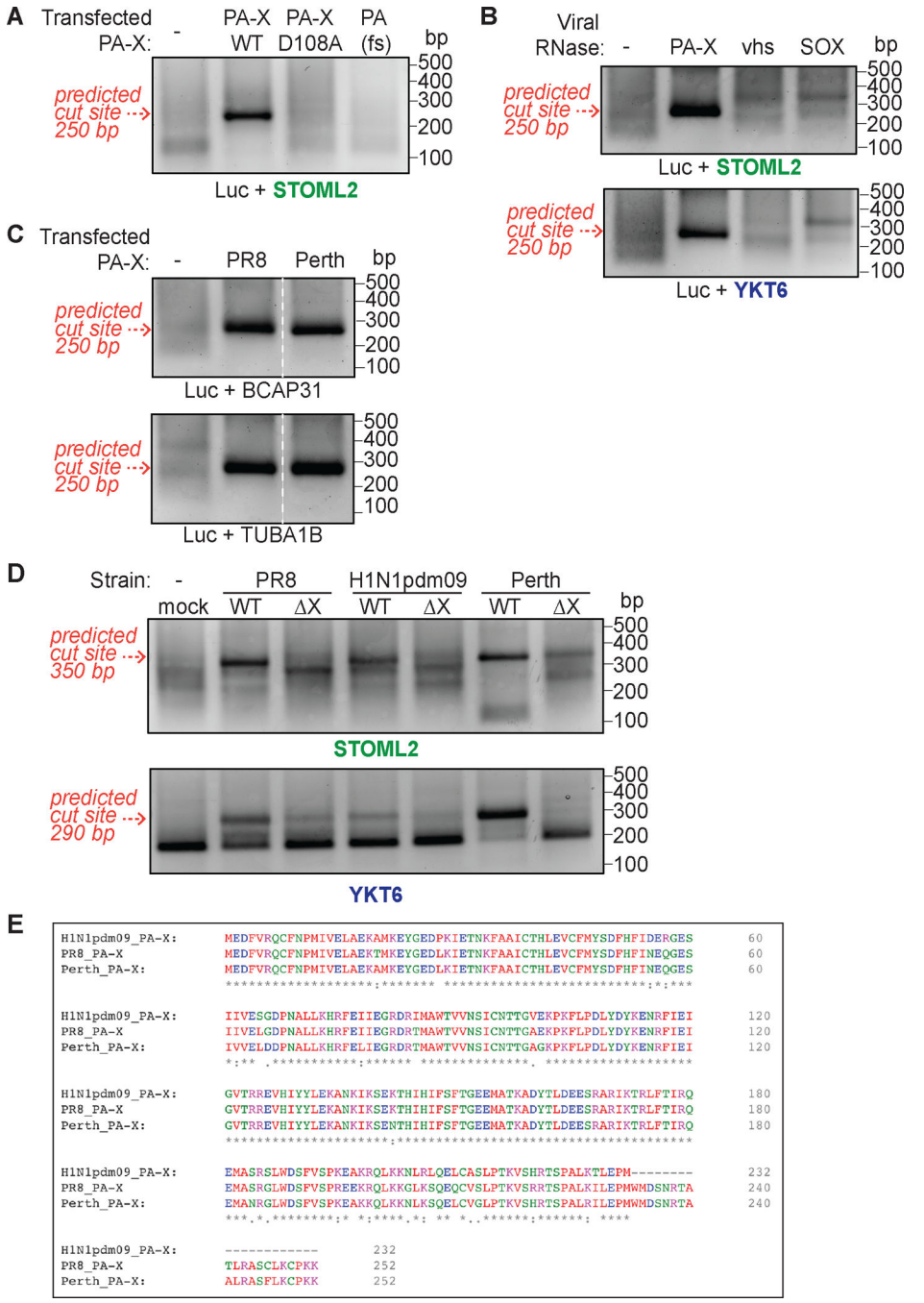
loading control to check for loss of Xrn1. Images are representative of 2 experiments. (D-E) Representation of individual chromosomal positions in the 5' RACE-seq data. For each sample, reads with their 5' end mapping to the same nucleotide were counted and plotted to compare different datasets: (D) mock samples from replicate 1 vs. replicate 2, (E) replicate 1 WT PR8 vs. PR8-PA(ΔX) (top), WT PR8 vs. mock infected (middle) and PR8-PA(ΔX) vs. mock infected (bottom). For each plot, light blue dots correspond to locations that are unique to one sample, while red and black dots correspond to locations that are common in the two samples. Red dots represent locations that have two-fold or more reads in the sample on the x axis vs. than in the sample on the y axis. Similar plots were obtained when comparing other replicates. (F) PyDegradome was run on concatenated read files comparing WT PR8 vs. PR8-PA(ΔX) infections using different parameters as indicated on the x axis (see Methods for parameter explanation). cl = confidence levels, mf = multiplicative factor. The window of analysis was 4 nucleotides in all cases. The total number of cut sites identified by PyDegradome (red circles, left y axis) and the percentage of shared cut sites also found as concatenated cut sites (blue diamonds, right y axis) were plotted for each set of parameters.



Extended Data Fig. 2. Further validation of PA-X cut sites identified by PyDegradome

(A-B) Xrn1 knock out A549 cells were infected with WT PR8 or PR8-PA(ΔX), or mock infected. (A) RNA was isolated, 5' RACE was performed using primers specific for *BCAP31*, *TUBA1B*, *INSIG1* or *SLC7A5*, and the PCR products were run on an agarose gel. Primers were positioned ~200–300 nucleotides downstream of the predicted cut sites. The predicted size of DNA bands coming from cut sites identified by PyDegradome are indicated by red dotted arrows. (B) PR8 HA RNA levels were quantified by qRT-PCR, normalized to 18S and plotted as mean ± standard deviation. AU, arbitrary units; n.d. not defined; ns, not

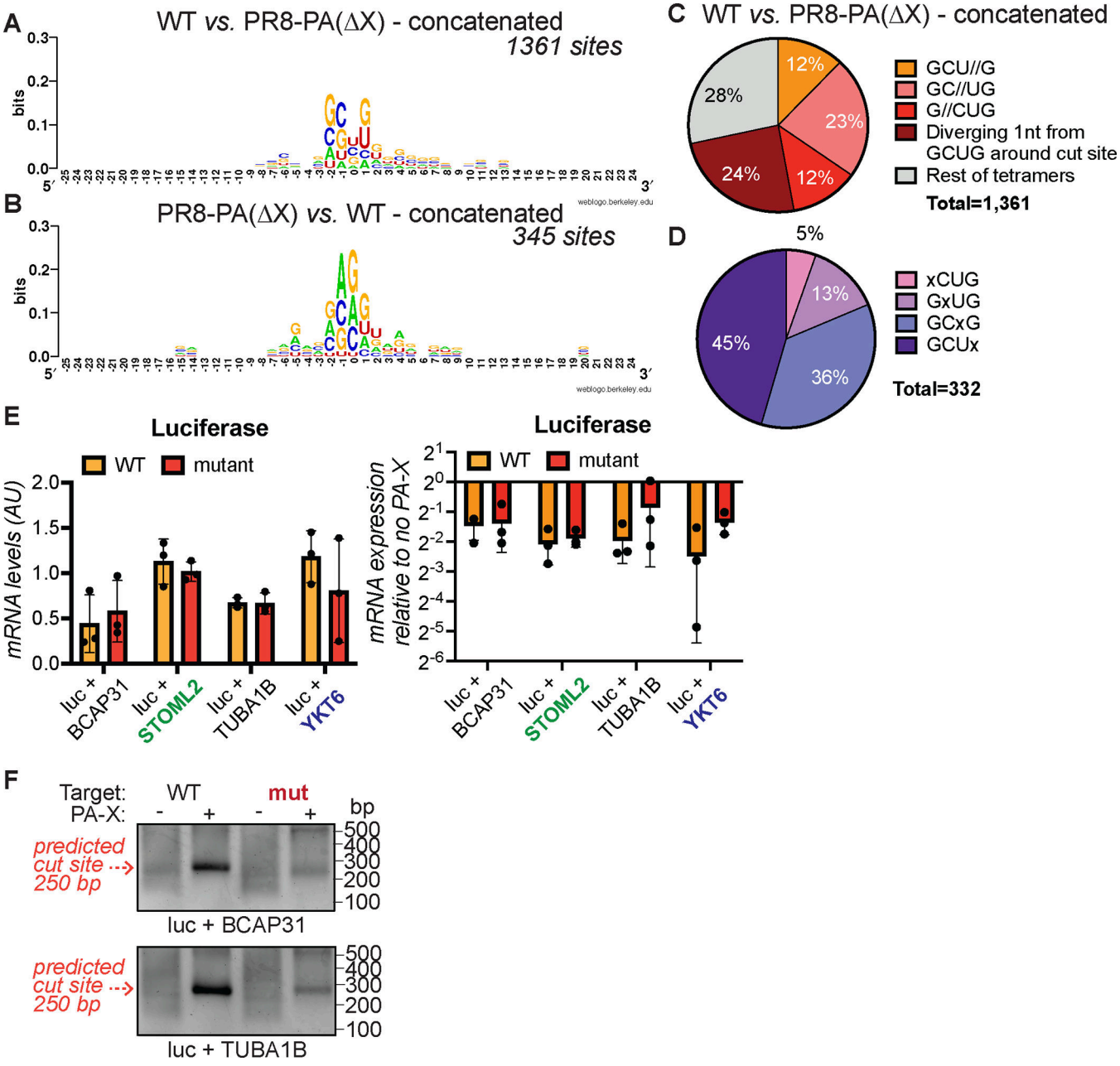
significant, One-way ANOVA with Dunnett's multiple comparison test. n = 3 independent experiments. (C) HEK293T ishXrn1 cells were treated with no drug or doxycycline for 3–4 days to induce shRNAs against Xrn1, then protein lysates were collected and probed with antibodies against Xrn1, or β-tubulin as a loading control to check for efficient knock down of Xrn1. Images are representative of 3 experiments. (D-E) HEK293T ishXrn1 cells were treated with doxycycline for 3–4 days to induce knock down of Xrn1, then transfected with luciferase reporters containing 99 bp insertions from the indicated genes, and where indicated, with PR8 PA-X. (D) RNA was extracted and luciferase mRNA levels were quantified by qRT-PCR, normalized to 18S and plotted as mean ± standard deviation. AU, arbitrary units. luc + *INSIG1*/*SLC7A5* n = 2, luc + *STOML2*/*YKT6*/*TUBA1B* n = 3, luc + *BCAP3*1 n = 4, luc + repeat n = 8, luc n = 9 independent experiments. (E) The RNA was also used to run 5' RACE. Expected sizes of DNA bands coming from cut sites in the introduced target sequences are indicated by red arrows, while the blue arrow in D indicates the size of the original luciferase cut site fragment. For all gels, the DNA bands were purified and sequenced to confirm their identities. Images are representative of 3 experiments or 2 experiments for luciferase + INSIG1/SLC7A5. The SLC7A5 sequence was only consistently cut when inside the luciferase reporter.



Extended Data Fig. 3. The cut sites identified by PyDegradome are specific to PA-X and conserved across multiple influenza strains

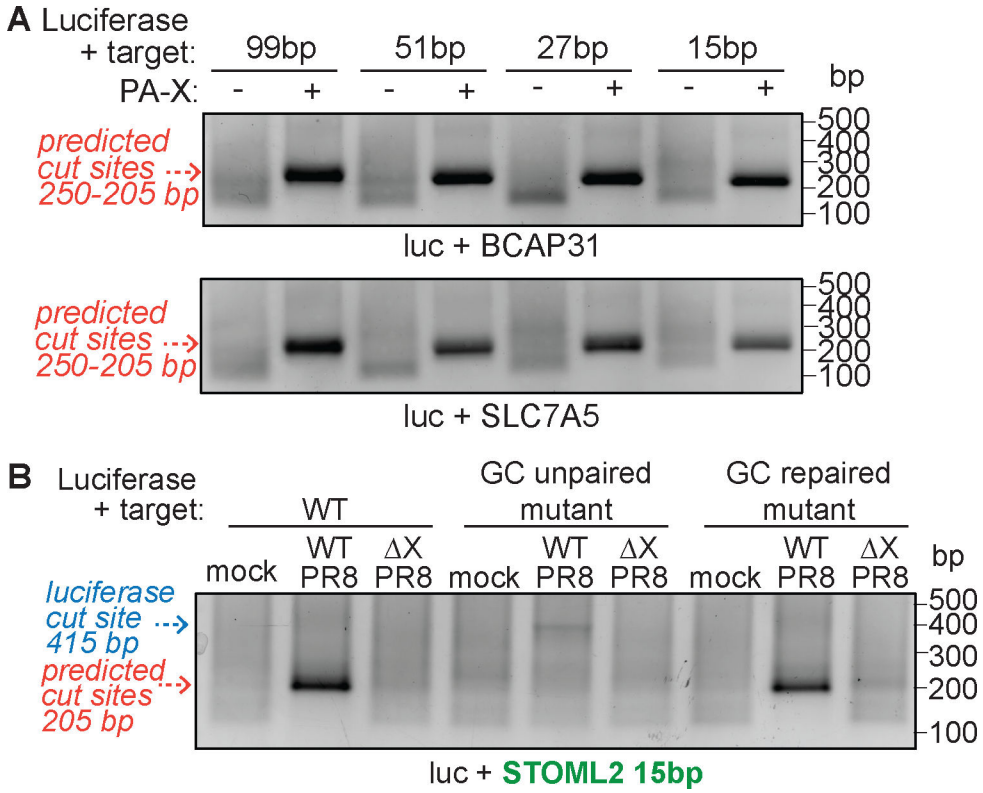
(A-C) HEK293T ishXrn1 cells were treated with doxycycline for 3–4 days to induce knock down of Xrn1, then transfected with luciferase reporters containing 99 bp insertions from the indicated genes, and where indicated, with WT PA-X from the PR8 strain (A-C), the catalytic mutant PR8 PA-X D108A (A), PR8 PA with a mutation to reduce frameshifting and prevent PA-X production (PA(fs)) (A), herpes simplex virus 1 (HSV-1) vhs or Kaposi's sarcoma-associated herpes virus (KSHV) SOX (B), or WT PA-X from the Perth influenza strain (C). RNA was extracted and used to run 5' RACE. Expected sizes of DNA bands

coming from cut sites in the introduced target sequences are indicated by the red dotted arrows. (D) Xrn1 knock out A549 cells were infected with WT PR8, Perth or pH1N1pdm09 or the corresponding PA(ΔX) mutants, or mock infected. 5' RACE was then performed using primers specific for *STOML2* or *YKT6* ~250–300 nt downstream of the predicted cut sites. The PCR products were run on an agarose gel. The predicted size of DNA bands coming from cut sites identified by PyDegradome are indicated by the red dotted arrows. For all gels, the DNA bands were purified and sequenced to confirm their identities, and images are representative of 3 experiments. (E) Protein alignment of PA-X from the three different influenza strains PR8, H1N1pdm09 and Perth, generated using Clustal Omega74.



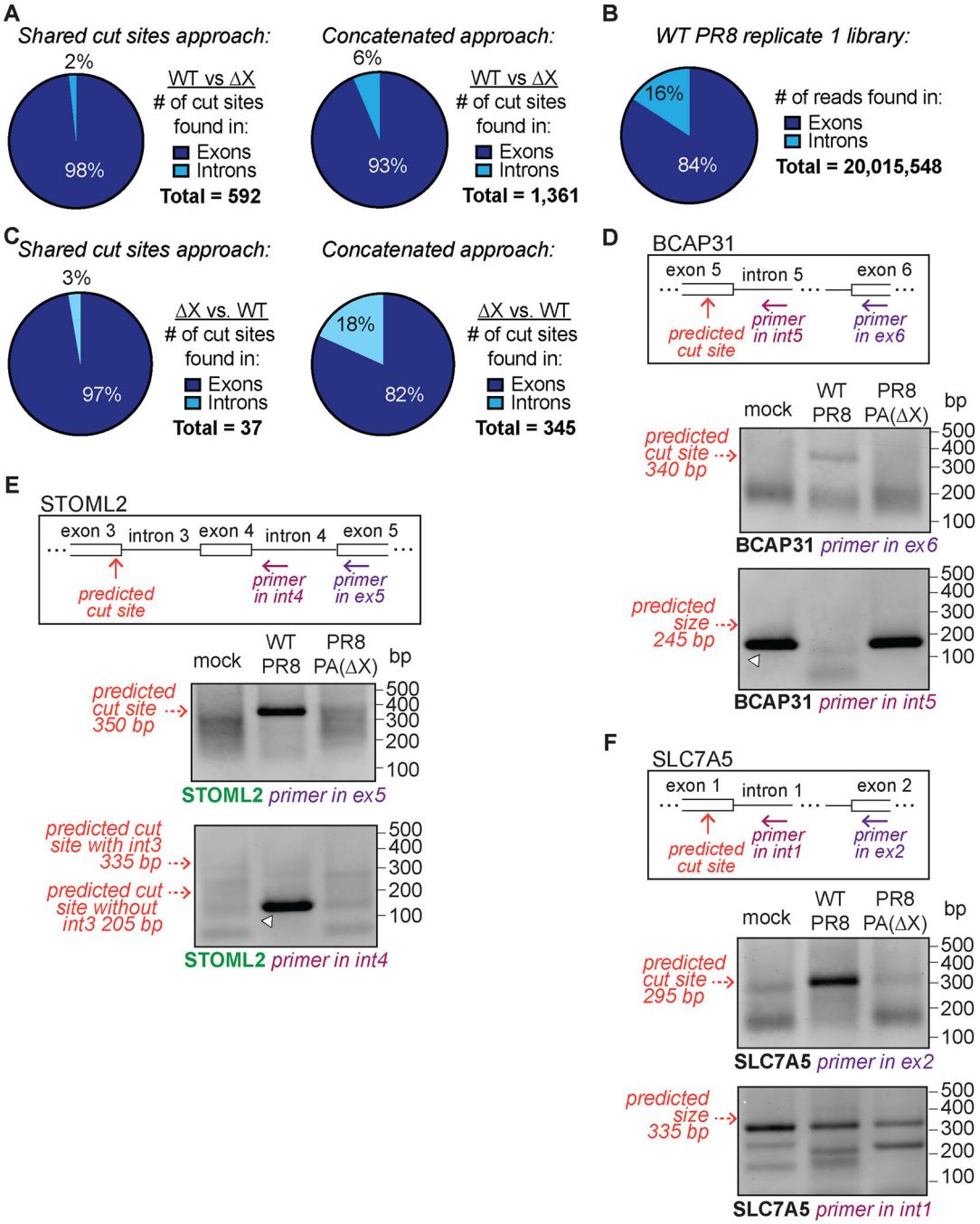
Extended Data Fig. 4. PA-X preferentially cleaves RNA at GCUG tetramers based on the concatenated cut site analysis

(A-B) WebLogo57 representation of base enrichment around PA-X cut sites (WT PR8 vs. PR8-PA(ΔX), A) or around control sites enriched in the PR8-PA(ΔX) sample (PR8-PA(ΔX) vs. WT PR8, B) for sites predicted by PyDegradome using the concatenated approach. (C) Percentage of PA-X cut sites containing GCUG or a tetramer with one nucleotide difference from GCUG, for sites identified by PyDegradome using the concatenated approach. // indicates the location of the cut, i.e. GCU//G indicates that PA-X cuts between the U and the G. (D) Further breakdown of the PA-X cut sites containing a tetramer with one nucleotide difference from GCUG around the cut site (marked by an x). (E-F) HEK293T ishXrn1cells were treated with doxycycline for 3–4 days to induce knock down of Xrn1, then transfected with luciferase reporters containing 99 bp insertions from the indicated genes, with or without PR8 PA-X. (E) RNA was extracted and luciferase mRNA levels were quantified by qRT-PCR, normalized to 18S and plotted as mean ± standard deviation. Left plot shows mRNA levels of each reporter in the absence of PA-X. Right plot shows mRNA levels of each reporter in the presence vs. absence of PA-X. AU, arbitrary units. n = 3 independent experiments. (F) The RNA was also used to run 5’ RACE. Expected sizes of DNA bands coming from cut sites in the introduced target sequences are indicated by the red dotted arrows. For both gels, DNA bands were purified and sequenced to confirm their identities, and images are representative of 3 experiments



Extended Data Fig. 5. PA-X preferentially cleaves RNA within a hairpin loop structure in transfected and infected cells

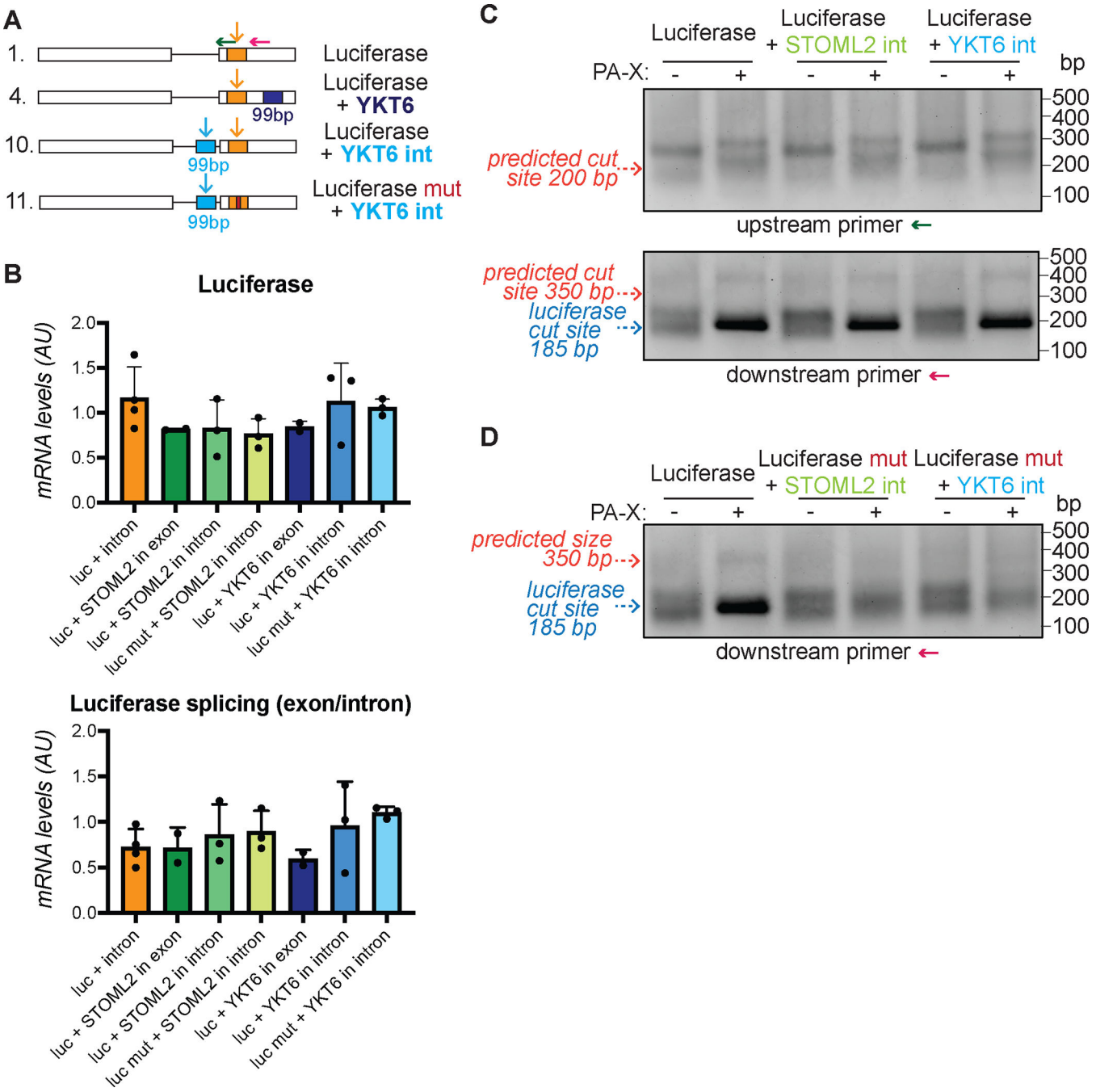
(A) HEK293T ishXrn1cells were treated with doxycycline for 3–4 days to induce knock down of Xrn1, then transfected with luciferase reporters containing insertions of the indicated lengths from the *BCAP31* and *SLC7A5* genes, with or without PR8 PA-X. RNA was extracted and used to run 5' RACE. Expected sizes of DNA bands coming from cut sites in the introduced target sequences are indicated by the red dotted arrows (~250 bp for 99 bp constructs, ~220 bp for 51 bp constructs, ~210 bp for 27 bp constructs and ~205 bp for the 15 bp constructs). (B) HEK293T ishXrn1cells were treated with doxycycline for 3–4 days to induce knock down of Xrn1, then transfected with luciferase reporters containing 15 bp insertions from the STOML2 gene with or without the indicated mutations. 24 hours post transfection, cells were infected with WT PR8 or PR8-PA(ΔX), or mock infected overnight. RNA was then extracted and used to run 5' RACE. Expected sizes of DNA bands coming from cut sites in the introduced target sequences are indicated by the red dotted arrow, while the blue arrow indicates the size of the original luciferase cut site fragment. For all gels, the DNA bands were purified and sequenced to confirm their identities, and images are representative of 3 experiments (B) or 2 experiments (A).



Extended Data Fig. 6. PA-X preferentially cleaves RNAs within exons

(A-C) Percentage of PA-X cut sites found within introns or exons for sites identified by PyDegradome using the shared cut sites approach (A and C, left) or the concatenated approach (A and C, right), compared to the percentage of reads found in exons vs. introns (B). The reads used in B are from WT PR8 replicate 1 as an example, but other samples have similar read distribution. The difference between reads and PA-X cut site distribution in introns vs. exons is statistically significant ($p < 0.001$), but the difference between reads and ΔX specific sites distribution in the concatenated analysis is not ($p > 0.2$), Chi-Square Test

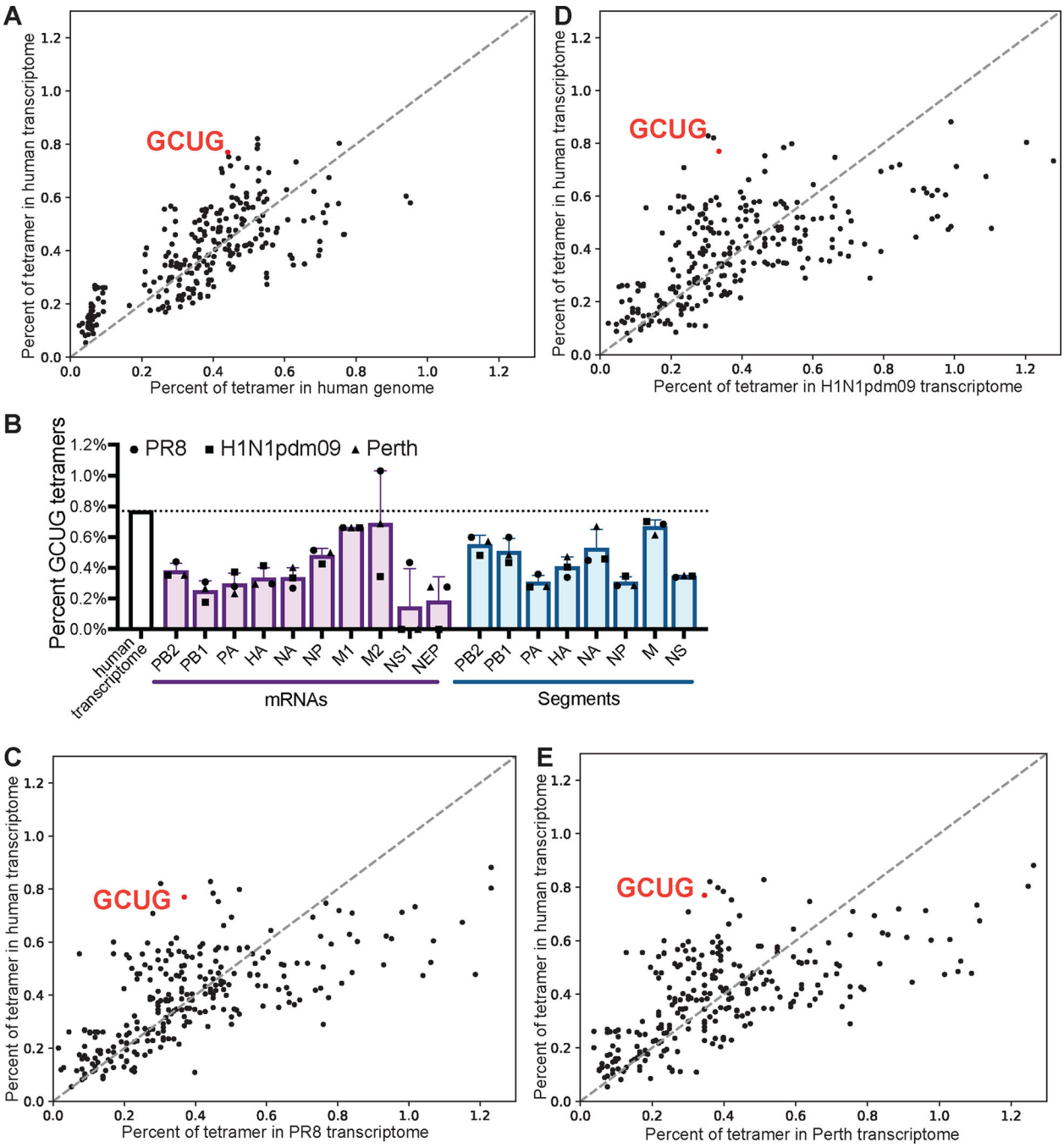
for Goodness of Fit with degree of freedom 1. While only 3% of PR8-PA(ΔX) fragments from the shared analysis mapped to introns ($0.025 < p < 0.05$ compared to reads distribution, Chi-Square Test for Goodness of Fit with degree of freedom 1), this may stem from the very low number of sites, 37, found with this method. (D-F) Xrn1 knock out A549 cells were infected with WT PR8 or PR8-PA(ΔX), or mock infected. 5'RACE was performed and PCR products were run on an agarose gel. The boxes show diagrams of the *BCAP31*, *STOML2* or *SLC7A5* genes and the positions of the reverse primers for 5' RACE. Top gels are the same gels as Fig. 2B and Extended Data Fig. 3A, and are included for comparison. Red dotted arrows indicate the size of fragments originating from previously validated cut sites. Bottom gels show products obtained using reverse primers in the intron (light purple primers), and red dotted arrows indicate the predicted size of PCR products that would appear if PA-X cleaved unspliced pre-mRNAs. White arrowhead indicates fragments that map to exon/intron junctions. New gel images are representative of 3 experiments.



Extended Data Fig. 7. PA-X only cleaves STOML2 and YKT6 preferred sequences if they are located within exons

(A) Diagram of the luciferase reporters tested. Green and magenta arrows indicate positions of 5' RACE PCR reverse primers, vertical arrows indicate location of predicted cut sites. (B-D) Testing of luciferase reporters with sequence insertions in the introns, as shown in diagram in A. HEK293T ishXrn1 cells were treated with doxycycline for 3–4 days to induce knock down of Xrn1, then transfected with luciferase reporters containing the 99 bp insertions in the indicated genes either in an exon or an intron, with or without PR8 PA-X. (B) RNA was extracted and luciferase mRNA levels were quantified by qRT-PCR,

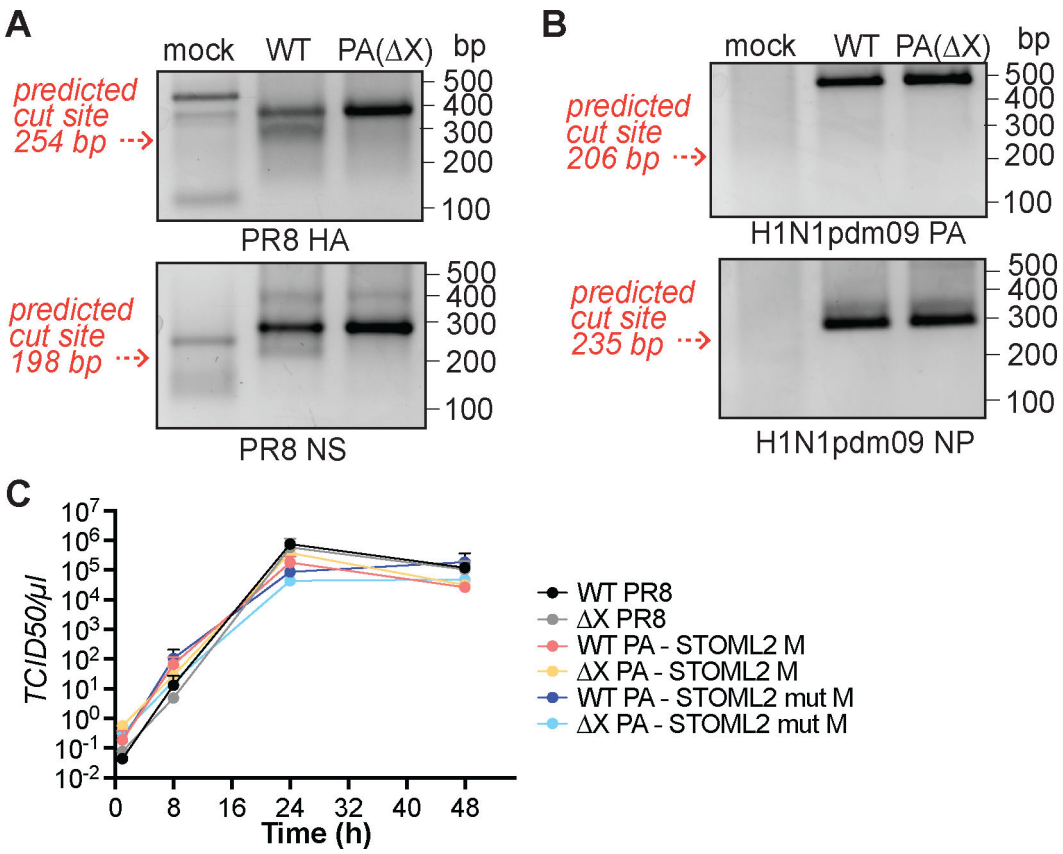
normalized to 18S and plotted as mean ± standard deviation. AU, arbitrary units; n = 3 except for luc + STOML2/YKT6 in exon n = 2, each independent experiments. The top graph shows levels of the processed mRNA (obtained using two primers that bind in exons). The bottom graph shows levels of the unspliced pre-mRNA (obtained using one primer in an exon and one primer in the intron). (C-D) RNA was also used to run 5' RACE. PCR products were separated on an agarose gel. The top gel in C and the gel in D represent products obtained using the green arrow primer from A, and the bottom gel in C using the magenta arrow primer from A. Blue dotted arrow indicates size of PCR products originating from the original luciferase cut site, red dotted arrow indicates the predicted sizes of PCR products that would originate from the sequence inserted in the introns. Gel images are representative of 3 experiments.



Extended Data Fig. 8. GCUG tetramers are enriched in the human transcriptome

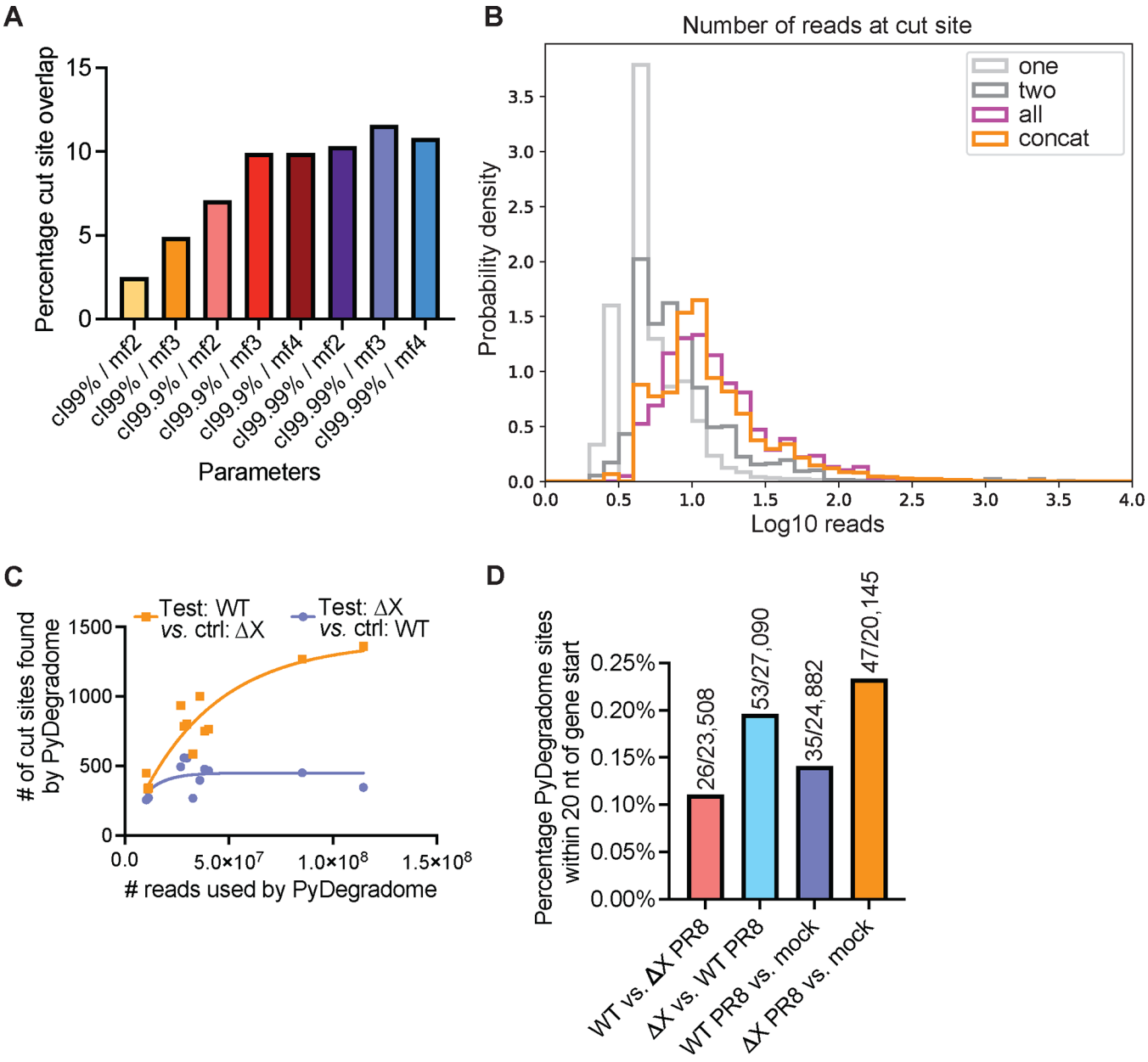
The percentage of tetramers in the human and viral genomes and transcriptomes was calculated by counting the number of each tetramer and dividing it by the total number of tetramers in the sequence (i.e. length of the sequence minus 3). Percentages for each tetramer are plotted to visualize which tetramers are more abundant in the human genome vs. transcriptome (A) and the human transcriptome vs. the transcriptome of 3 influenza A virus strains (C-E). Each dot represents a specific tetramer, with the red dot representing the GCUG tetramer. (B) The percentage of GCUG tetramers is plotted for each influenza

mRNA (i.e. the positive strand, purple), and for each influenza genomic RNA (i.e. negative strand, blue). Data are plotted as mean ± standard deviation, with each symbol representing a different influenza strain.



Extended Data Fig. 9. Viral mRNAs are not efficiently cleaved by PA-X

(A-B) Xrn1 knock out A549 cells were infected with WT PR8 or PR8-PA(ΔX) (A), or WT H1N1pdm09 or H1N1pdm09-PA(ΔX) (B), or mock infected. RNA was extracted to run 5' RACE using primers ~150–200 nt downstream of GCUG sites that are predicted to be in a hairpin loop in the indicated viral mRNAs. Red dotted arrows indicate the predicted sizes of PCR products that would originate from cleavage at these GCUG sites. Gel images are representative of 3 experiments. (C) MDCK cells were infected at MOI 0.05 with the indicated viruses (see Fig. 5F for details on virus construction). Supernatants were collected at 1, 8, 24 and 48 hours post infection and viral titers were quantified by TCID50. n = 2 independent experiments.



Extended Data Fig. 10. Analysis of PyDegradome output characteristics

(A) PyDegradome was run for each replicate on WT PR8 vs. PR8-PA(ΔX) infection samples using different parameters as indicated on the x axis (see Methods). cl = confidence levels, mf = multiplicative factor. The window of analysis was 4 nucleotides in all cases. The percentage of cut sites shared between all three replicates was plotted for each set of parameters. (B) Histogram of the number of reads at the cut site for PA-X sites (PyDegradome comparison: WT PR8 vs. PR8-PA(ΔX)) found in one (light grey), two (dark grey), or three replicates (“all”,magenta), or through the concatenated analysis (“concat”, orange). (C) Number of PA-X cut sites (orange squares) or RNA fragments enriched in the PR8-PA(ΔX) infected cells (purple circles) identified by PyDegradome relative to the number of input reads. (D) Percentage of sites identified by PyDegradome in at least one

replicate that are located within the first 20 nucleotides of a gene when comparing WT PR8 vs.PR8-PA(ΔX), PR8-PA(ΔX) vs. WT PR8, WT PR8 vs. mock and PR8-PA(ΔX) vs. mock.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Albert Tai and the personnel of the Tufts University Core Facility - Genomics Core for help with the sequencing. We thank Drs. Craig McCormick and Denys Khaperskyy (Dalhousie University, Halifax, NS, Canada) and members of their laboratories for their advice and input. We thank Drs. Craig McCormick (Dalhousie University, Halifax, NS, Canada), Britt Glaunsinger (University of California, Berkeley, CA, USA), Richard Webby (St Jude's Children Research Hospital, Memphis, TN, USA), Gideon Dreyfuss (University of Pennsylvania, Philadelphia, PA, USA), Jesse Bloom (Fred Hutchinson Cancer Center, Seattle, WA, USA) and Seema Lakdawala (Emory University School of Medicine, Atlanta, GA, USA) for constructs, Dr. Mariano Garcia Blanco (University of Virgina, Charlottesville, VA, USA) for providing the Northern Blot protocol, Dr. Bernard Moss (National Institute of Health, Bethesda, MD, USA) for cell lines, and Drs. John Coffin, Claire Moore, Karl Munger (Tufts University, Boston, MA, USA) and members of the Gaglia lab for critical reading of the manuscript. This work was supported by NIH grant R01 AI137358 (to M.M.G.). L.G. was supported by NIH F31 AI154587 and T32 GM007310. C.H.R. was partially supported by the Applied Mathematics Program of the U.S. DOE Office of Science Advanced Scientific Computing Research under contract number DE-AC02–05CH11231. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Data availability

All primary data (including uncropped gel and blot images) except for 5' RACE-seq data are publicly available as a Figshare collection, doi.org/10.6084/m9.figshare.c.6616663[70]. The raw 5' RACE-seq sequences and identified cut sites are deposited in the NCBI GEO database (GSE207253)

References

1. Gaucherand L & Gaglia MM The Role of Viral RNA Degrading Factors in Shutoff of Host Gene Expression. Annu Rev Virol 9, 213–238 (2022). [PubMed: 35671567]
2. Bercovich-Kinori A, Tai J, Gelbart IA, Shitrit A, Ben-Moshe S, Drori Y, Itzkovitz S, Mandelboim M & Stern-Ginossar N A systematic view on influenza induced host shutoff. eLife 5, e18311 (2016). [PubMed: 27525483]
3. Jagger BW, Wise HM, Kash JC, Walters KA, Wills NM, Xiao YL, Dunfee RL, Schwartzman LM, Ozinsky A, Bell GL, Dalton RM, Lo A, Efstathiou S, Atkins JF, Firth AE, Taubenberger JK & Digard P An overlapping protein-coding region in influenza A virus segment 3 modulates the host response. Science 337, 199–204 (2012) [PubMed: 22745253]
4. Gaucherand L, Porter BK, Levene RE, Price EL, Schmaling SK, Rycroft CH, Kevorkian Y, McCormick C, Khaperskyy DA & Gaglia MM The Influenza A Virus Endoribonuclease PA-X Usurps Host mRNA Processing Machinery to Limit Host Gene Expression. Cell Reports 27, 776–792.e7 (2019). [PubMed: 30995476]
5. Gao H, Sun Y, Hu J, Qi L, Wang J, Xiong X, Wang Y, He Q, Lin Y, Kong W, Seng L-G, Sun H, Pu J, Chang K-C, Liu X & Liu J The contribution of PA-X to the virulence of pandemic 2009 H1N1 and highly pathogenic H5N1 avian influenza viruses. Sci Rep 5, 8262 (2015). [PubMed: 25652161]
6. Gong XQ, Sun YF, Ruan BY, Liu XM, Wang Q, Yang HM, Wang SY, Zhang P, Wang XH, Shan TL, Tong W, Zhou YJ, Li GX, Zheng H, Tong GZ & Yu H PA-X protein decreases replication and pathogenicity of swine influenza virus in cultured cells and mouse models. Veterinary Microbiology 205, 66–70 (2017). [PubMed: 28622865]

7. Hayashi T, MacDonald LA & Takimoto T Influenza A Virus Protein PA-X Contributes to Viral Growth and Suppression of the Host Antiviral and Immune Responses. Journal of Virology (2015). doi:10.1128/JVI.00319-15
8. Hu J, Mo Y, Wang X, Gu M, Hu Z, Zhong L, Wu Q, Hao X, Hu S, Liu W, Liu H, Liu X & Liu X PA-X Decreases the Pathogenicity of Highly Pathogenic H5N1 Influenza A Virus in Avian Species by Inhibiting Virus Replication and Host Response. Journal of Virology 89, 6442–6452 (2015). [PubMed: 25855745]
9. Firth AE, Jagger BW, Wise HM, Nelson CC, Parsawar K, Wills NM, Napthine S, Taubenberger JK, Digard P & Atkins JF Ribosomal frameshifting used in influenza A virus expression occurs within the sequence UCC-UUU-CGU and is in the +1 direction. Open Biology 2, 1–7 (2012).
10. Dias A, Bouvier D, Crépin T, McCarthy AA, Hart DJ, Baudin F, Cusack S & Ruigrok RWH The cap-snatching endonuclease of influenza virus polymerase resides in the PA subunit. Nature 458, 914–918 (2009). [PubMed: 19194459]
11. Yuan P, Bartlam M, Lou Z, Chen S, Zhou J, He X, Lv Z, Ge R, Li X, Deng T, Fodor E, Rao Z & Liu Y Crystal structure of an avian influenza polymerase PA_N reveals an endonuclease active site. Nature 458, 909–913 (2009). [PubMed: 19194458]
12. Khaperskyy DA, Schmaling S, Larkins-Ford J, McCormick C & Gaglia MM Selective Degradation of Host RNA Polymerase II Transcripts by Influenza A Virus PA-X Host Shutoff Protein. PLoS Pathogens 12, (2016).
13. Chaimayo C, Dunagan M, Hayashi T, Santoso N & Takimoto T Specificity and functional interplay between influenza virus PA-X and NS1 shutoff activity. PLOS Pathogens 14, e1007465 (2018). [PubMed: 30496325]
14. Chao Y, Li L, Girodat D, Förstner KU, Said N, Corcoran C, Śmiga M, Papenfort K, Reinhardt R, Wieden H-J, Luisi BF & Vogel J In Vivo Cleavage Map Illuminates the Central Role of RNase E in Coding and Non-coding RNA Pathways. Molecular Cell 65, 39–51 (2017). [PubMed: 28061332]
15. Gaglia MM, Rycroft CH & Glaunsinger BA Transcriptome-Wide Cleavage Site Mapping on Cellular mRNAs Reveals Features Underlying Sequence-Specific Cleavage by the Viral Ribonuclease SOX. PLoS Pathogens 11, 1–25 (2015).
16. Abernathy E, Gilbertson S, Alla R & Glaunsinger B Viral nucleases induce an mRNA degradation-transcription feedback loop in mammalian cells. Cell Host and Microbe 18, 243–253 (2015). [PubMed: 26211836]
17. Liu R & Moss B Opposing Roles of Double-Stranded RNA Effector Pathways and Viral Defense Proteins Revealed with CRISPR-Cas9 Knockout Cell Lines and Vaccinia Virus Mutants. Journal of Virology 90, 7864–7879 (2016). [PubMed: 27334583]
18. Younis I, Berg M, Kaida D, Dittmar K, Wang C & Dreyfuss G Rapid-response splicing reporter screens identify differential regulators of constitutive and alternative splicing. Molecular and cellular biology 30, 1718–28 (2010). [PubMed: 20123975]
19. Levene RE, Shrestha SD & Gaglia MM The influenza A virus host shutoff factor PA-X is rapidly turned over in a strain-specific manner. J Virol 95, e02312–20 (2021). [PubMed: 33504608]
20. Bavagnoli L, Cucuzza S, Campanini G, Rovida F, Paolucci S, Baldanti F & Maga G The novel influenza A virus protein PA-X and its naturally deleted variant show different enzymatic properties in comparison to the viral endonuclease PA. Nucleic Acids Research 43, 9405–9417 (2015). [PubMed: 26384413]
21. Plotch SJ, Bouloy M, Ulmanen I & Krug RM A unique cap(m7GpppXm)-dependent influenza virion endonuclease cleaves capped RNAs to generate the primers that initiate viral RNA transcription. Cell 23, 847–858 (1981). [PubMed: 6261960]
22. Datta K, Wolkerstorfer A, Szolar OHJ, Cusack S & Klumpp K Characterization of PA-N terminal domain of Influenza A polymerase reveals sequence specific RNA cleavage. Nucl. Acids Res. 41, 8289–8299 (2013). [PubMed: 23847103]
23. Sikora D, Rocheleau L, Brown EG & Pelchat M Deep sequencing reveals the eight facets of the influenza A/HongKong/1/1968 (H3N2) virus cap-snatching process. Sci Rep 4, 6181 (2014). [PubMed: 25154590]

24. Gu W, Gallagher GR, Dai W, Liu P, Li R, Trombly MI, Gammon DB, Mello CC, Wang JP & Finberg RW Influenza A virus preferentially snatches noncoding RNA caps. RNA 21, 2067–2075 (2015). [PubMed: 26428694]
25. Brannan K, Kim H, Erickson B, Glover-Cutter K, Kim S, Fong N, Kiemele L, Hansen K, Davis R, Lykke-Andersen J & Bentley DL MRNA Decapping Factors and the Exonuclease Xrn2 Function in Widespread Premature Termination of RNA Polymerase II Transcription. Molecular Cell 46, 311–324 (2012). [PubMed: 22483619]
26. Wang X-H, Gong X-Q, Wen F, Ruan B-Y, Yu L-X, Liu X-M, Wang Q, Wang S-Y, Wang J, Zhang Y-F, Zhou Y-J, Shan T-L, Tong W, Zheng H, Kong N, Yu H & Tong G-Z The Role of PA-X C-Terminal 20 Residues of Classical Swine Influenza Virus in Its Replication and Pathogenicity. Veterinary Microbiology 251, 108916 (2020). [PubMed: 33197868]
27. Oishi K, Yamayoshi S & Kawaoka Y Identification of amino acid residues in influenza A virus PA-X that contribute to enhanced shutoff activity. Frontiers in Microbiology 10, 1–8 (2019). [PubMed: 30728808]
28. Feng KH, Sun M, Iketani S, Holmes EC & Parrish CR Comparing the functions of equine and canine influenza H3N8 virus PA-X proteins: Suppression of reporter gene expression and modulation of global host gene expression. Virology 496, 138–146 (2016). [PubMed: 27314620]
29. Nogales A, Villamayor L, Utrilla-Trigo S, Ortego J, Martinez-Sobrido L & DeDiego ML Natural Selection of H5N1 Avian Influenza A Viruses with Increased PA-X and NS1 Shutoff Activity. Viruses 2021, Vol. 13, Page 1760 13, 1760 (2021).
30. Wilamowski M, Gorecki A, Dziedzicka-Wasylewska M & Jura J Substrate specificity of human MCPIP1 endoribonuclease. Sci Rep 8, 7381 (2018). [PubMed: 29743536]
31. Takata MA, Gonçalves-Carneiro D, Zang TM, Soll SJ, York A, Blanco-Melo D & Bieniasz PD CG dinucleotide suppression enables antiviral defence targeting non-self RNA. Nature 550, 124–127 (2017). [PubMed: 28953888]
32. Pezda AC, Penn A, Barton GM & Coscoy L Suppression of TLR9 immunostimulatory motifs in the genome of a gammaherpesvirus. J Immunol 187, 887–896 (2011). [PubMed: 21666062]
33. Hartenian E & Glaunsinger BA Feedback to the central dogma: cytoplasmic mRNA decay and transcription are interdependent processes. Crit Rev Biochem Mol Biol 54, 385–398 (2019). [PubMed: 31656086]
34. Duncan-Lewis C, Hartenian E, King V & Glaunsinger BA Cytoplasmic mrna decay represses rna polymerase ii transcription during early apoptosis. eLife 10, (2021).
35. Zhao N, Sebastiano V, Moshkina N, Mena N, Hultquist J, Jimenez-Morales D, Ma Y, Rialdi A, Albrecht R, Fenouil R, Sánchez-Aparicio MT, Ayllon J, Ravisankar S, Haddad B, Ho JSY, Low D, Jin J, Yurchenko V, Prinjha RK, Tarakhovsky A, Squatrito M, Pinto D, Allette K, Byun M, Smith ML, Sebra R, Guccione E, Tumpey T, Krogan N, Greenbaum B, van Bakel H, García-Sastre A & Marazzi I Influenza virus infection causes global RNAPII termination defects. Nature Structural and Molecular Biology 25, 885–893 (2018).
36. Bauer DLV, Tellier M, Martínez-Alonso M, Nojima T, Proudfoot NJ, Murphy S & Fodor E Influenza Virus Mounts a Two-Pronged Attack on Host RNA Polymerase II Transcription. Cell Reports 23, 2119–2129.e3 (2018). [PubMed: 29768209]
37. Khaperskyy DA, Emara MM, Johnston BP, Anderson P, Hatchette TF & McCormick C Influenza A Virus Host Shutoff Disables Antiviral Stress-Induced Translation Arrest. PLOS Pathogens 10, e1004217 (2014). [PubMed: 25010204]
38. Glaunsinger B & Ganem D Lytic KSHV infection inhibits host gene expression by accelerating global mRNA turnover. Mol. Cell 13, 713–723 (2004). [PubMed: 15023341]
39. Jones FE, Smibert CA & Smiley JR Mutational analysis of the herpes simplex virus virion host shutoff protein: evidence that vhs functions in the absence of other viral proteins. J Virol 69, 4863–4871 (1995). [PubMed: 7609054]
40. Le Sage V, Kormuth KA, Nturibi E, Lee JM, Frizzell SA, Myerburg MM, Bloom JD & Lakdawala SS Cell-Culture Adaptation of H3N2 Influenza Virus Impacts Acid Stability and Reduces Airborne Transmission in Ferret Model. Viruses 13, 719 (2021). [PubMed: 33919124]

41. Hoffmann E, Neumann G, Kawaoka Y, Hobom G & Webster RG A DNA transfection system for generation of influenza A virus from eight plasmids. Proceedings of the National Academy of Sciences 97, 6108–6113 (2000).
42. Khaperskyy DA, Hatchette TF & McCormick C Influenza A virus inhibits cytoplasmic stress granule formation. The FASEB Journal 26, 1629–1639 (2012). [PubMed: 22202676]
43. Matrosovich M, Matrosovich T, Garten W & Klenk HD New low-viscosity overlay medium for viral plaque assays. Virology Journal 3, 1–7 (2006). [PubMed: 16390540]
44. REED LJ & MUENCH H A SIMPLE METHOD OF ESTIMATING FIFTY PER CENT ENDPOINTS12. American Journal of Epidemiology 27, 493–497 (1938).
45. Reed Muench Calculator. (2022). at <https://github.com/jbloomlab/reedmuenchcalculator>
46. Marta.Gaglia & lea.gaucherand. Purification of influenza A virus RNA from cell supernatant to check sequences. (2023). at <https://www.protocols.io/view/purification-of-influenza-a-virus-rna-from-cell-su-cthnwj5e>
47. Marta.Gaglia. 5’ RACE for RNA fragments with 5’ phosphate. (2023). at <https://www.protocols.io/view/5-39-race-for-rna-fragments-with-5-39-phosphate-cthawj2e>
48. German MA, Luo S, Schroth G, Meyers BC & Green PJ Construction of Parallel Analysis of RNA Ends (PARE) libraries for the study of cleaved miRNA targets and the RNA degradome. Nat Protoc 4, 356–362 (2009). [PubMed: 19247285]
49. Zhai J, Arikit S, Simon SA, Kingham BF & Meyers BC Rapid construction of parallel analysis of RNA end (PARE) libraries for Illumina sequencing. Methods 67, 84–90 (2014). [PubMed: 23810899]
50. Marta.Gaglia & lea.gaucherand. 5’ RACE-seq for RNA fragments with 5’ phosphate - library prep protocol. (2023). at <https://www.protocols.io/view/5-race-seq-for-rna-fragments-with-5-phosphate-libr-cthjwj4n>
51. Martin M Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17, 10–12 (2011).
52. Kim D, Paggi JM, Park C, Bennett C & Salzberg SL Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nature Biotechnology 37, 907–915 (2019).
53. Huang L, Zhang H, Deng D, Zhao K, Liu K, Hendrix DA & Mathews DH LinearFold: linear-time approximate RNA folding by 5’-to-3’ dynamic programming and beam search. Bioinformatics 35, i295–i304 (2019). [PubMed: 31510672]
54. Do CB, Woods DA & Batzoglou S CONTRAfold: RNA secondary structure prediction without physics-based models. Bioinformatics 22, e90–e98 (2006). [PubMed: 16873527]
55. Lorenz R, Bernhart SH, Höner zu Siederdissen C, Tafer H, Flamm C, Stadler PF & Hofacker IL ViennaRNA Package 2.0. Algorithms for Molecular Biology 6, 1–14 (2011). [PubMed: 21235792]
56. Mathews DH, Disney MD, Childs JL, Schroeder SJ, Zuker M & Turner DH Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. Proceedings of the National Academy of Sciences of the United States of America 101, 7287 (2004). [PubMed: 15123812]
57. Crooks GE, Hon G, Chandonia JM & Brenner SE WebLogo: A Sequence Logo Generator. Genome Research 14, 1188–1190 (2004). [PubMed: 15173120]
58. WebLogo - Create Sequence Logos. at <https://weblogo.berkeley.edu/logo.cgi>
59. Gao H, Xu G, Sun Y, Qi L, Wang J, Kong W, Sun H, Pu J, Chang K-C & Liu J PA-X is a virulence factor in avian H9N2 influenza virus. J. Gen. Virol. 96, 2587–2594 (2015). [PubMed: 26296365]
60. Rigby RE, Wise HM, Smith N, Digard P & Rehwinkel J PA-X antagonises MAVS-dependent accumulation of early type I interferon messenger RNAs during influenza A virus infection. Sci Rep 9, 7216 (2019). [PubMed: 31076606]
61. Cunningham F, Allen JE, Allen J, Alvarez-Jarreta J, Amode MR, Armean IM, Austine-Orimoloye O, Azov AG, Barnes I, Bennett R, Berry A, Bhai J, Bignell A, Billis K, Boddu S, Brooks L, Charkhchi M, Cummins C, Da Rin Fioretto L., Davidson C., Dodiya K., Donaldson S., El Houdaigui B., El Naboulsi T., Fatima R., Giron CG., Genez T., Martinez JG., Guijarro-Clarke C., Gymer A., Hardy M., Hollis Z., Hourlier T., Hunt T., Juettemann T., Kaikala V., Kay M., Lavidas I., Le T., Lemos D., Marugán JC., Mohanan S., Mushtaq A., Naven M., Ogeh DN., Parker A., Parton A., Perry M., Piližota I., Prosovetskaia I., Sakthivel MP., Salam AIA., Schmitt

BM., Schuilenburg H., Sheppard D., Pérez-Silva JG., Stark W., Steed E., Sutinen K., Sukumaran R., Sumathipala D., Suner M-M., Szpak M., Thormann A., Tricomi FF., Urbina-Gómez D., Veidenberg A., Walsh TA., Walts B., Willhoft N., Winterbottom A., Wass E., Chakiachvili M., Flint B., Frankish A., Giorgetti S., Haggerty L., Hunt SE., IIsley GR., Loveland JE., Martin FJ., Moore B., Mudge JM., Muffato M., Perry E., Ruffier M., Tate J., Thybert D., Trevanion SJ., Dyer S., Harrison PW., Howe KL., Yates AD., Zerbino DR. & Flicek P. Ensembl 2022. Nucleic Acids Research 50, D988–D995 (2022). [PubMed: 34791404]

62. Warr A, Affara N, Aken B, Beiki H, Bickhart DM, Billis K, Chow W, Eory L, Finlayson HA, Flicek P, Girón CG, Griffin DK, Hall R, Hannum G, Hourlier T, Howe K, Hume DA, Izuogu O, Kim K, Koren S, Liu H, Manchanda N, Martin FJ, Nonneman DJ, O'Connor RE, Phillippy AM, Rohrer GA, Rosen BD, Rund LA, Sargent CA, Schook LB, Schroeder SG, Schwartz AS, Skinner BM, Talbot R, Tseng E, Tuggle CK, Watson M, Smith TPL & Archibald AL An improved pig reference genome sequence to enable pig genetics and genomics research. Gigascience 9, giaa051 (2020). [PubMed: 32543654]
63. Li M, Chen L, Tian S, Lin Y, Tang Q, Zhou X, Li D, Yeung CKL, Che T, Jin L, Fu Y, Ma J, Wang X, Jiang A, Lan J, Pan Q, Liu Y, Luo Z, Guo Z, Liu H, Zhu L, Shuai S, Tang G, Zhao J, Jiang Y, Bai L, Zhang S, Mai M, Li C, Wang D, Gu Y, Wang G, Lu H, Li Y, Zhu H, Li Z, Li M, Gladyshev VN, Jiang Z, Zhao S, Wang J, Li R & Li X Comprehensive variation discovery and recovery of missing sequence in the pig genome using multiple de novo assemblies. Genome Res. 27, 865–874 (2017). [PubMed: 27646534]
64. Zhang Y, Aevermann BD, Anderson TK, Burke DF, Dauphin G, Gu Z, He S, Kumar S, Larsen CN, Lee AJ, Li X, Macken C, Mahaffey C, Pickett BE, Reardon B, Smith T, Stewart L, Suloway C, Sun G, Tong L, Vincent AL, Walters B, Zaremba S, Zhao H, Zhou L, Zmasek C, Klem EB & Scheuermann RH Influenza Research Database: An integrated bioinformatics resource for influenza virus research. Nucleic Acids Res 45, D466–D474 (2017). [PubMed: 27679478]
65. Liang H, Lam TT-Y, Fan X, Chen X, Zeng Y, Zhou J, Duan L, Tse M, Chan C-H, Li L, Leung T-Y, Yip C-H, Cheung C-L, Zhou B, Smith DK, Poon LL-M, Peiris M, Guan Y & Zhu H Expansion of genotypic diversity and establishment of 2009 H1N1 pandemic-origin internal genes in pigs in China. J Virol 88, 10864–10874 (2014). [PubMed: 25008935]
66. Iqbal M, Yaqub T, Reddy K & McCauley JW Novel genotypes of H9N2 influenza A viruses isolated from poultry in Pakistan containing NS genes similar to highly pathogenic H7N3 and H5N1 viruses. PLoS One 4, e5788 (2009). [PubMed: 19517011]
67. Liu Q, Ma J, Kou Z, Pu J, Lei F, Li T & Liu J Characterization of a highly pathogenic avian influenza H5N1 clade 2.3.4 virus isolated from a tree sparrow. Virus Res 147, 25–29 (2010). [PubMed: 19804800]
68. Li Y, Shi J, Zhong G, Deng G, Tian G, Ge J, Zeng X, Song J, Zhao D, Liu L, Jiang Y, Guan Y, Bu Z & Chen H Continued evolution of H5N1 influenza viruses in wild birds, domestic poultry, and humans in China from 2004 to 2009. J Virol 84, 8389–8397 (2010). [PubMed: 20538856]
69. Tran V, Ledwith MP, Thamamongood T, Higgins CA, Tripathi S, Chang MW, Benner C, García-Sastre A, Schwemmle M, Boon ACM, Diamond MS & Mehle A Influenza virus repurposes the antiviral protein IFIT2 to promote translation of viral mRNAs. Nat Microbiol 5, 1490–1503 (2020). [PubMed: 32839537]
70. Gaglia M Cut site preference allows influenza A virus PA-X to discriminate between host and viral mRNAs. (2023). doi:10.6084/m9.figshare.c.6616663
71. mgaglia81. PyDegradome. (2022). at <https://github.com/mgaglia81/PyDegradome>
72. LinearX Webserver (Beta). at <https://linearfold.eecs.oregonstate.edu/>
73. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD & Higgins DG Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular Systems Biology 7, 539 (2011). [PubMed: 21988835]

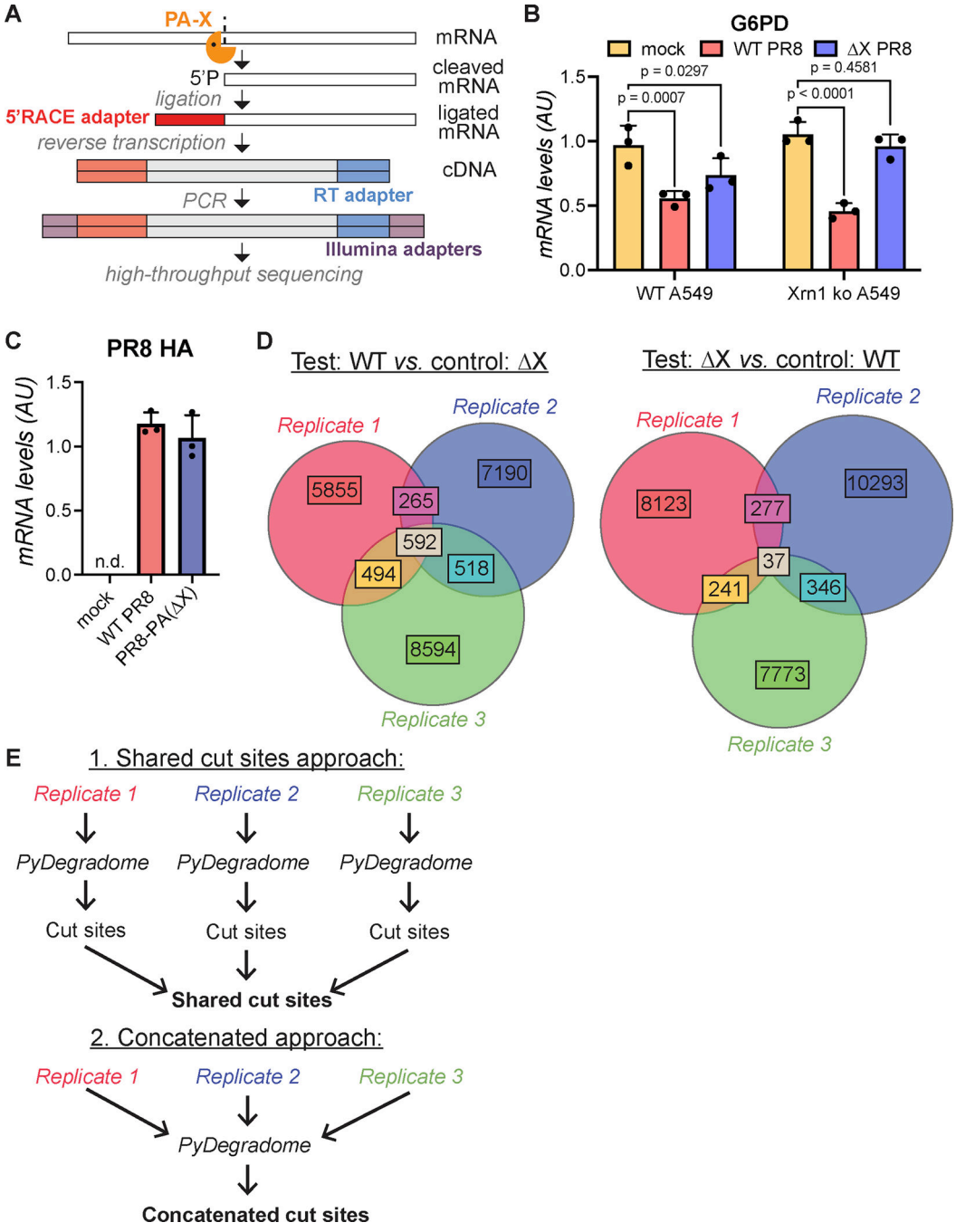


Figure 1: 5' RACE-seq with PyDegradome analysis identifies PA-X cut sites transcriptome-wide.

(A) Schematic diagram of 5' RACE-seq workflow. (B-C) Wild-type (B) or Xrn1 knock out (B-C) A549 cells were either mock infected or infected with WT PR8 or PR8-PA(ΔX) for 16 hours before RNA extraction. mRNA levels of influenza A virus HA and human G6PD, a PA-X target, were quantified by qRT-PCR, normalized by 18S rRNA levels, and plotted as mean ± standard deviation. AU, arbitrary units; n.d., not defined; Two-way ANOVA with Dunnett's multiple comparison test. n = 3 independent experiments. (D) Venn diagrams of the number of cut sites identified by PyDegradome for each replicate. Left diagram shows

WT PR8 specific fragments (test: WT PR8, control: PR8-PA(ΔX)), which are likely to be PA-X cut sites; right diagram shows PR8-PA(ΔX) specific fragments (test: PR8-PA(ΔX), control: WT PR8), which are used throughout this study as a negative control. (E) Flow chart of steps used to define PA-X cut sites for further analysis, either through the shared cut sites approach or the concatenated approach.

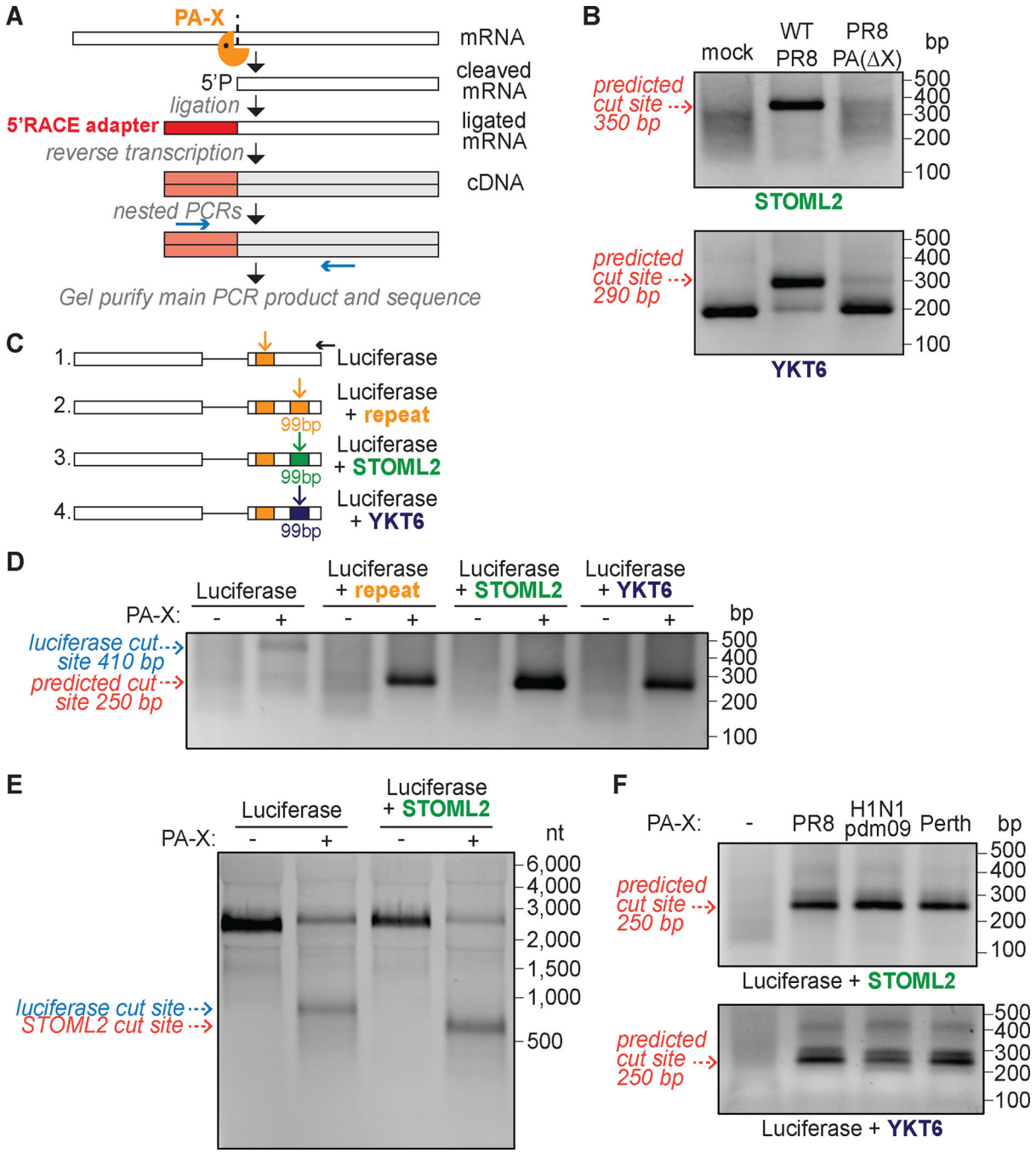


Figure 2: Sequences around the cut sites drive cleavage by PA-X.

(A) Schematic diagram of classical 5' RACE workflow. (B) Xrn1 knock out A549 cells were infected with WT or PR8-PA(ΔX), or mock infected for 16 hours before RNA extraction. 5'RACE was then performed using primers specific for *STOML2* or *YKT6*, ~250–300 nucleotides (nt) downstream of the predicted cut sites. The PCR products were run on an agarose gel. The predicted size of DNA bands coming from cut sites identified by PyDegradome are indicated by the red dotted arrows. Background bands can be observed with YKT6 primers at ~200 bp. These fragments map to TTG//AAC site in exon 7, 93 nt

downstream of GCTG cut site, and may represent a constitutive fragment of degradation. (C) Diagram of the luciferase reporters tested. The black horizontal arrow indicates the position of the 5' RACE PCR reverse primer, vertical arrows indicate the predicted position of the cut sites. (D-F) HEK293T ishXrn1 cells were treated with doxycycline for 3–4 days to induce knock down of Xrn1, then transfected with one of the luciferase reporters in (C), and where indicated, with PA-X from the PR8 (D-F), H1N1pdm09 or Perth influenza strains (F). RNA was extracted to run 5' RACE (D, F) or northern blotting (E). Expected sizes of DNA/RNA bands coming from cut sites in the introduced target sequences are indicated by the red dotted arrows, while the blue dotted arrow indicates the size of the original luciferase cut site fragment. For all DNA gels, the DNA bands were purified and sequenced to confirm their identities. All gel and northern blot images are representative of 3 experiments.

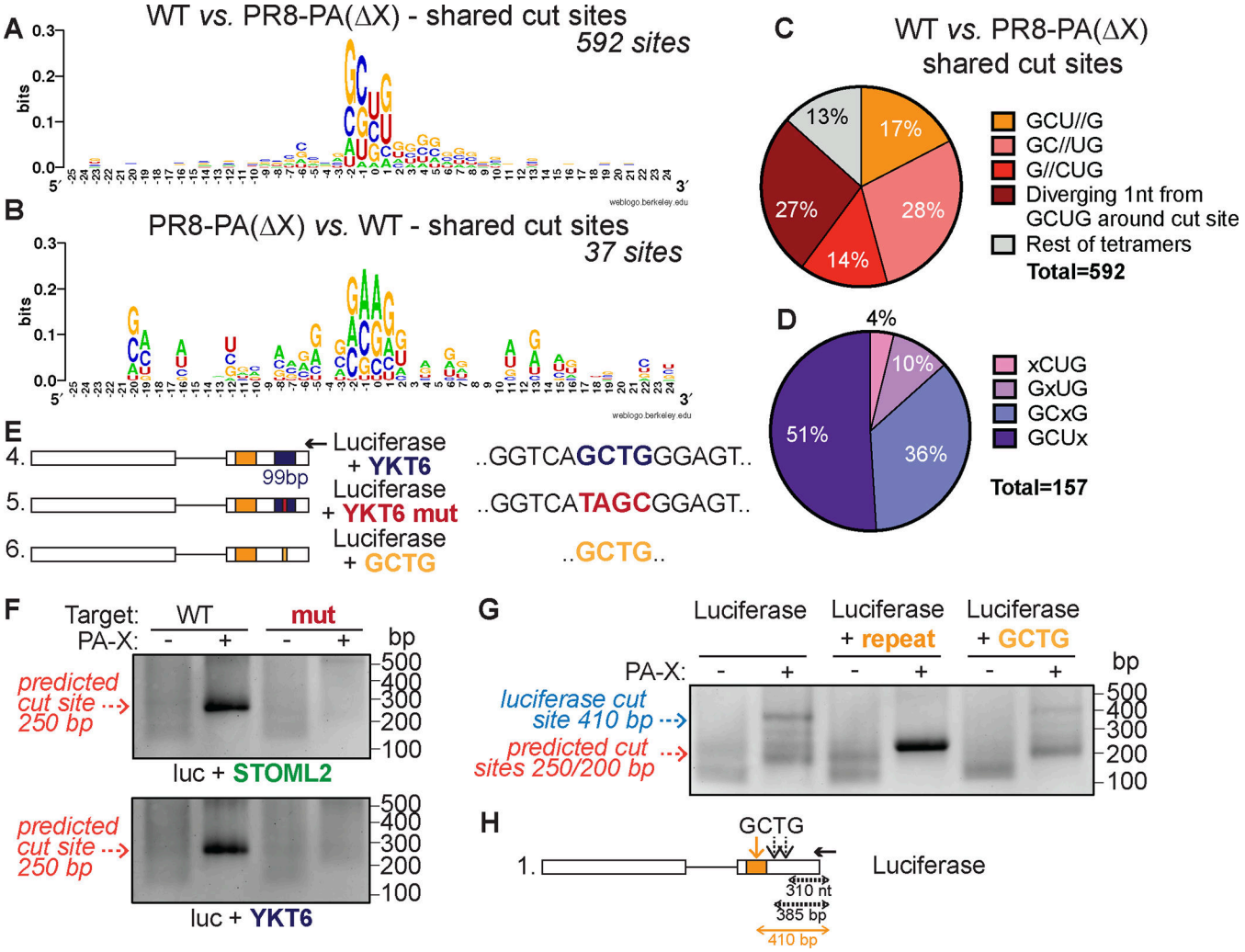


Figure 3: PA-X preferentially cleaves RNAs at GCUG tetramers.

(A-B) WebLogo[57] representation of base enrichment around PA-X cut sites (test: WT PR8 *vs.* control: PR8-PA(ΔX), A) or around control sites enriched in the PR8-PA(ΔX) sample (test: PR8-PA(ΔX) vs. control: WT PR8, B) for sites predicted by PyDegradome using the shared cut sites approach. (C) Percentage of PA-X cut sites containing GCUG or a tetramer with one nucleotide difference from GCUG, for sites identified by PyDegradome using the shared cut sites approach. // indicates the location of the cut, i.e. GCU//G indicates that PA-X cuts between U and G. (D) Further breakdown of PA-X cut sites containing a tetramer with one nucleotide difference from GCUG around the cut site (marked by an x). (E) Diagram of the luciferase reporters tested. The black arrow indicates the position of the 5' RACE PCR reverse primer. (F-G) HEK293T ishXrn1cells were treated with doxycycline for 3–4 days to induce knock down of Xrn1, then transfected with luciferase reporters containing 99 bp insertions from the indicated genes with or without the GCTG → TAGC mutation shown in E (F), a repeat of the luciferase cut site (G), or the GCTG tetramer alone (G), together with PR8 PA-X or an empty vector. RNA was extracted to run 5' RACE. Expected sizes of DNA bands coming from cut sites in the introduced target sequences are indicated by red dotted arrows (~250 bp for luciferase + *STOML2*, luciferase + *YKT6* and

luciferase + repeat, ~200 bp for luciferase + GCUG), while the dotted blue arrow indicates the size of the original luciferase cut site fragment. DNA bands were purified and sequenced to confirm their identities. Gel images are representative of 3 experiments. (H) Diagram of the positions of several GCTG tetramers in the luciferase reporter. The orange vertical arrow indicates the original luciferase cut site and the black vertical dotted arrows additional GCTG sites that are not clearly cleaved by PA-X. Black horizontal arrow indicates location of 5’ RACE reverse primer, and lengths underneath diagram indicate the expected length of 5’ RACE PCR products each cut site would generate.

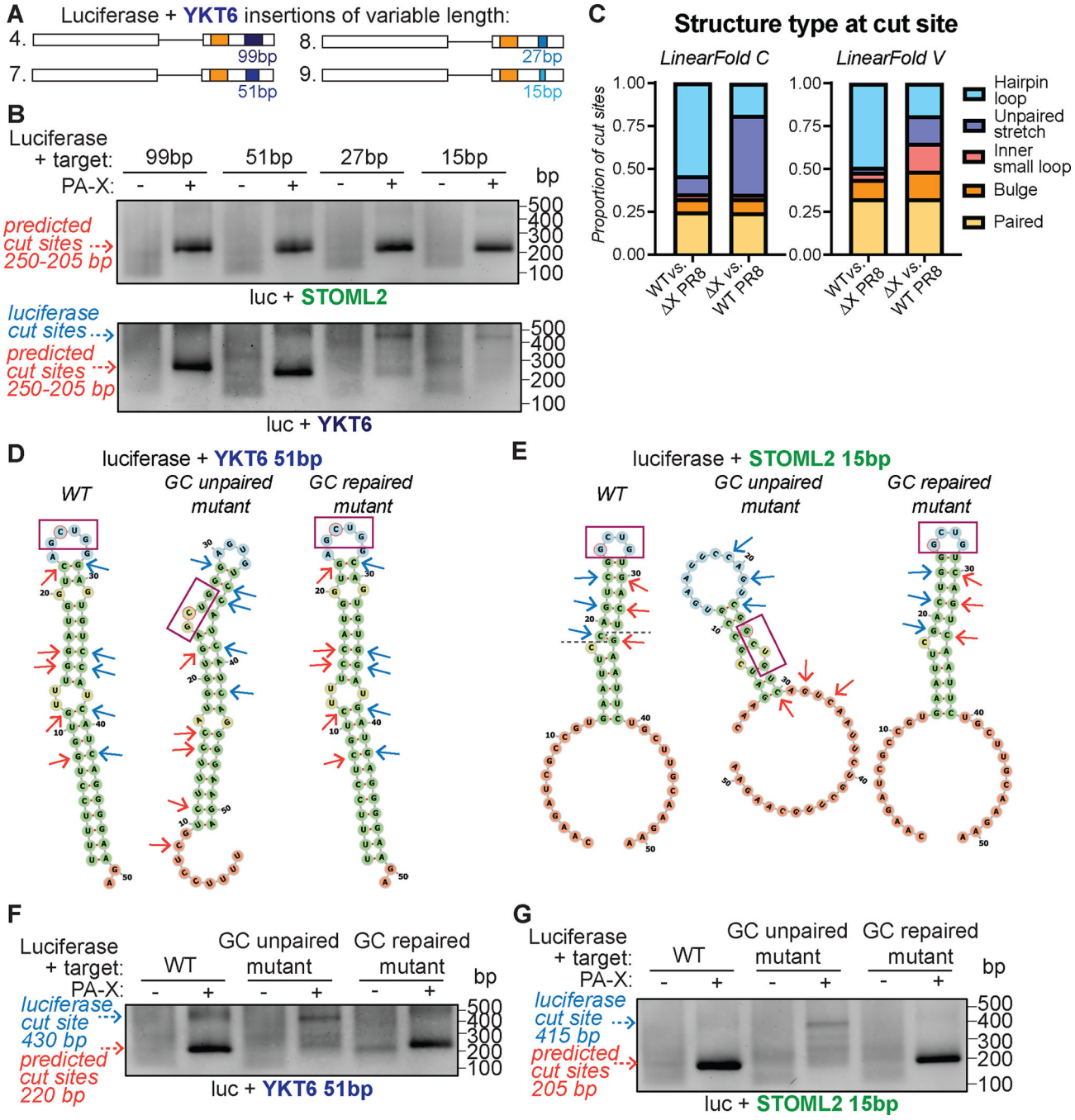


Figure 4: PA-X preferentially cleaves RNA within hairpin loop structures.

(A) Diagram of the luciferase reporters tested. (B, F, G) HEK293T ishXrn1cells were treated with doxycycline for 3–4 days to induce knock down of Xrn1, then transfected with luciferase reporters containing insertions of the indicated lengths from the *YKT6* or *STOML2* genes (B), with or without the mutations indicated in D-E (F-G), together with PR8 PA-X or an empty vector. RNA was extracted to run 5' RACE. Expected sizes of DNA bands coming from cut sites in the introduced target sequences are indicated by red dotted arrows (~250 bp for 99 bp constructs, ~220 bp for 51 bp constructs, ~210 bp

for 27 bp constructs and ~205 bp for the 15 bp constructs). Blue dotted arrows indicate the size of the original luciferase cut site fragments (~430 bp for 51 bp constructs, ~420 bp for 27 bp constructs and ~415 bp for 15 bp constructs). DNA bands were purified and sequenced to confirm their identities. Gel images representative of 3 experiments. (C) Predicted RNA secondary structures of the 99 bp sequence around PA-X cut sites (WT PR8 vs. PR8 PA(ΔX), n = 592) or around control sites enriched in the PR8-PA(ΔX) sample (PR8 PA(ΔX) vs. WT PR8, n =37) (shared cut sites approach). Structures were predicted with LinearFold[53], using the CONTRAfold v2.0 machine-learning model[54] (LinearFold C) or the Vienna RNAfold thermodynamic model[55,56] (LinearFold V). (D-E) Diagram of the LinearFold C predicted structures for the 51 bp *YKT6* sequence (D) or the 15 bp *STOML2* and surrounding luciferase sequences (E). Left = structures for the WT sequences, middle = structures for sequences with mutations in the nucleotides indicated by the red arrows (GC unpaired mutant), right = structures for sequences with mutations in the nucleotides indicated by both red and blue arrows (GC repaired mutant). The GCUG cut sites are indicated by the purple boxes. PA-X cuts these sequences after the “C” or “G” circled in red. In E, the dark grey dotted line on the WT diagram indicates where the *STOML2* sequence ends and the luciferase surrounding sequence begins.

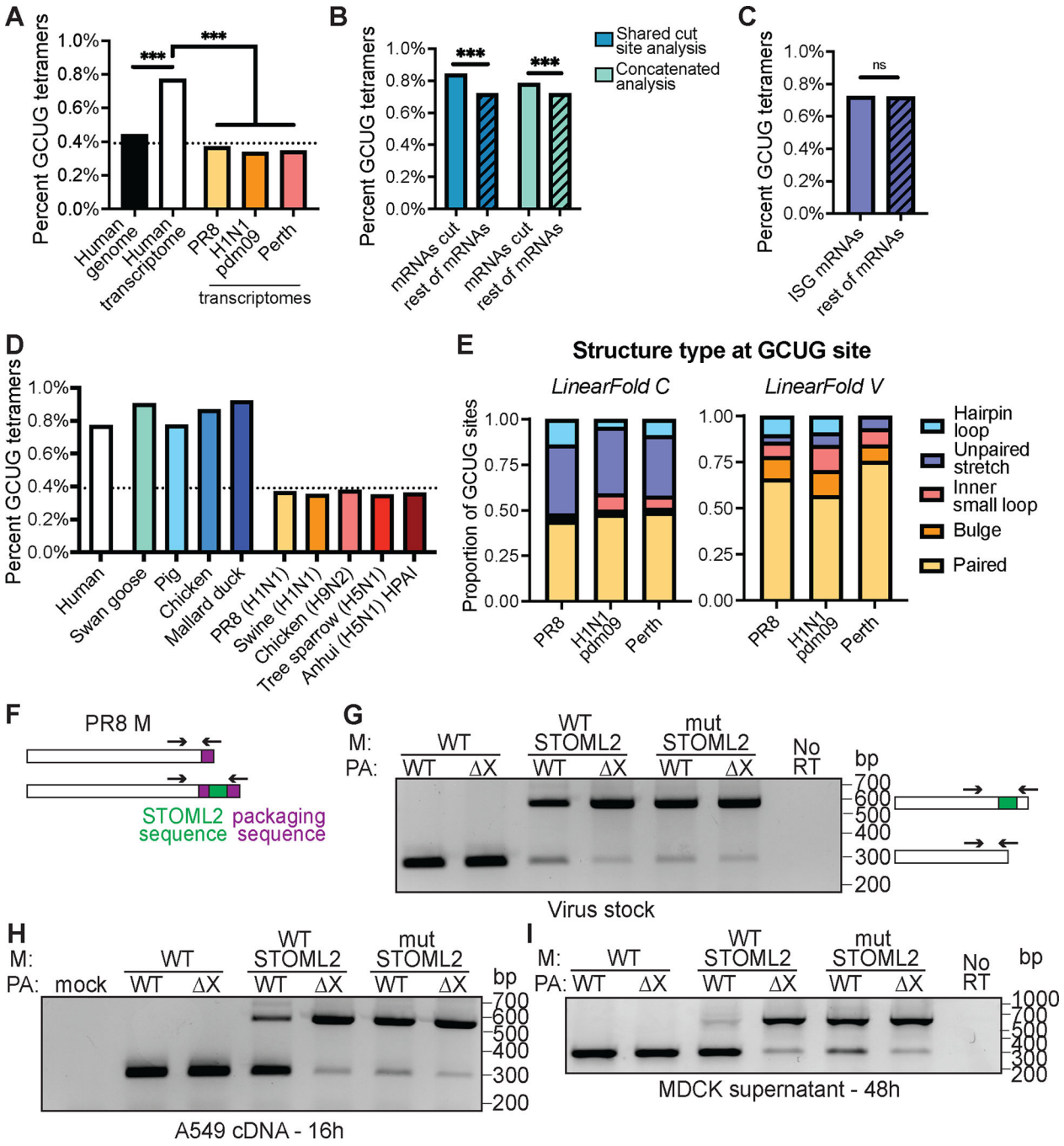


Figure 5: PA-X likely cleaves GCUG sequences to preferentially target host over viral mRNAs.

(A-D) Percentage of GCUG tetramers in indicated sequences calculated by counting the number of GCUG tetramers and dividing by the total number of tetramers in the sequence (i.e. length of sequence minus 3). ISG: interferon stimulated genes; ns = not significant, *** = $p < 0.001$, Chi-Square Test for Goodness of Fit with degree of freedom 1. Black dotted lines in (A) and (D) represent the average tetramer abundance expected by chance (i.e. 1/256). Transcriptomes in (D) are from *Anser cygnoides* (swan goose), *Sus scrofa* (pig), *Gallus gallus* (chicken), *Anas platyrhynchos* (duck), A/swine/Guangdong/2722/2011

(H1N1), A/chicken/Pakistan/UDL/01/2008 (H9N2), A/tree sparrow/Jiangsu/1/2008 (H5N1) and A/Anhui/1/2005 (H5N1) highly pathogenic avian influenza (HPAI). (E) Predicted RNA secondary structures of the 99 nt sequence around GCUG tetramers in the indicated influenza transcriptomes, using the CONTRAfold v2.0 machine-learning (LinearFold C) or the Vienna RNAfold thermodynamic (LinearFold V) model. (F) Diagram showing the location of the insertion of the *STOML2* 51 bp cut site sequence in the PR8 M segment after the M2 coding sequence (green rectangle). The full M segment packaging signal (purple rectangle) was positioned after the *STOML2* sequence to ensure the modified segment is packaged inside the virion, resulting in a duplication of the end of the M2 coding region. Arrows mark the position of primers to amplify sequences. The reverse primer only binds the full packaging sequence at the end of the engineered *M-STOML2* fusion segment. (G-I) Viral RNA was extracted from viral stocks (G), Xrn1 knock out A549 cells 16 hours post-infection (H), or 48-hour supernatants from infected MDCK cells in Extended Data Fig. 9C (H). Viruses analyzed: WT PR8 and PR8-PA(ΔX) strains with no insert, or PR8 with WT or mutant *STOML2* sequences inserted in the WT PR8 or PR8-PA(ΔX) background. Mutant *STOML2* sequences contain the GCUG → UAGC mutation preventing PA-X cleavage (Fig. 3). Viral RNA was reverse-transcribed and PCR-amplified using primers shown as black arrows in F-G to determine whether the *STOML2* sequences were retained. Gel images representative of 3 experiments with separate virus rescues.