Deciphering the molecular mechanism of action of the influenza A virus endoribonuclease PA-X

A thesis submitted by

Léa Gaucherand

in partial fulfillment of the requirements for the degree of

PhD

in

Molecular Microbiology

Tufts University

Graduate School of Biomedical Sciences

February 2023

Advisor: Marta Gaglia, PhD

Abstract

An efficient strategy for viruses to take over infected cells and evade the antiviral response is to block host gene expression, a process termed host shutoff. Influenza A virus carries out host shutoff mainly through its endoribonuclease PA-X. As such, PA-X expression leads to the widespread depletion of host RNAs. Yet, PA-X is not indiscriminate, but specifically down-regulates RNAs transcribed by the host RNA polymerase (Pol) II but not Pol I and III. Moreover, viral mRNAs are spared by PA-X. How PA-X achieves this specificity is unclear, as little is known about its molecular mechanism of action. I report that PA-X selects its targets through two complementary mechanisms. First, PA-X uses the host mRNA processing machinery to access RNAs transcribed by Pol II. Through transcriptomic studies, we identified a link between PA-X and host mRNA splicing. Spliced RNAs are more down-regulated by PA-X than intronless RNAs, and the degree of down-regulation of an RNA correlates with its number of exons. Splicing also promotes PA-X targeting of reporter constructs. Second, PA-X cleaves RNA at a specific sequence and structure to discriminate between host and viral mRNAs. Transcriptome-wide identification of PA-X cut sites during influenza infection showed that PA-X preferentially cuts RNA at GCUG tetramers located within the loop of hairpin structures. Importantly, GCUG is one of the most abundant tetramers in the human transcriptome, while it is of average/low abundance in the influenza transcriptome. Moreover, the GCUGs found in the influenza transcriptome are generally found in paired structures and not hairpin loops. Both the splicing and cut site selectivity are conserved across PA-X from multiple influenza strains, suggesting I have uncovered a fundamental targeting mechanism. In addition, my results suggest that PA-X acts in a manner analogous to cellular self *vs.* non-self discrimination mechanisms, taking advantage of a motif that is overrepresented in the human transcriptome but not in the

influenza transcriptome to specifically target host mRNAs. These findings provide new insights into how influenza A virus regulates host gene expression and antiviral responses, and show how viral endoribonuclease cut site specificity can play a functional role to selectively target the host.

Dedication

I would like to dedicate this thesis to all the people that have made its existence possible, the mentors that have helped me throughout my personal and scientific journey. Allen, who planted in me the love of research. Adam, Tom and the Wight lab family, who gave me a chance to prove myself and learn the ropes. Anna, who infected me with her passion for viruses. Caroline and Simon, always here to give me the best research and career advice #frenchresearchers. Marta, a true mentor in all its forms (and hair colors). And most of all my parents, Alex, Luc and Valérie, ma grand-mère Nicole, and my wonderful husband Cody, who have given me everything. Thank you for your support and for always believing in me. I would not be who and where I am today without you all.

Acknowledgments

First and foremost, thank you to my advisor Marta, the best role model and PhD mentor anyone could have. Thank you for your support and guidance, for all the time spent reviewing documents and presentations, for listening and caring about your students, for always trying to be a better mentor and always pushing me to be the best version of myself. Another big thank you to all the Gaglia lab members, especially Râë, Rächêl, Ali and Machika my goofy lab partners in crime, and T8 and Matt my older lab brothers. You have made my PhD so much more fun and enjoyable. Thank you for all your feedback, love, secret notes, cheese parties, late night sample collections, support and encouragement, and for being the best labmates.

Thank you to my committee members, Drs. Claire Moore, Karl Munger and John Coffin, for all their insights, time and support, and for going through these 200 pages. I would also like to thank Dr. Mandy Muller for accepting to be my outside examiner and reviewing my thesis work.

This work could not have been done without collaborators and co-authors, so I would like to deeply thank them for all of their work. In particular, thank you to Drs. Denys Khaperskyy and Craig McCormick and their team for their work, feedback and encouragement, and for reviewing all my conference abstracts every year. Thank you to my undergraduate student Amrita for all her work and efforts, and for keeping me young. Thank you to Dr. Chris Rycroft for all his help and work on PyDegradome and power laws, and for being our occasional (overly qualified) “IT consultant”.

I would also like to thank everyone in the Micro Program for their supportive and collaborative spirit. The faculty’s investment in students has been really helpful throughout my PhD and is a great model to aspire to for the future. Thank you to everyone who runs the administrative side of things, we would not go very far without

you. I would also like to thank my amazing cohort and all my friends in the Micro program and beyond for their friendship and support, and for creating Café Disco.

Finally, I would like to thank my family, Cody's family and my friends outside of Tufts for keeping me grounded into the non-science world. Thank you to my parents for their unconditional love, for always making me believe that I could do anything, and for trusting and supporting me when I told them I wanted to follow my passion and take an unpaid position on the other side of the planet. Thank you most of all to Cody for being his silly self, always ready to make me laugh, always having my back and supporting me in so many ways throughout this whole experience.

Table of Contents

List of Tables

List of Figures

List of Copyrighted Materials Used

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. (2019) *Cell Reports* 27(3):776-792.e7. PMID: 30995476 PMCID: PMC6499400

Gaucherand L. & Gaglia MM. The Role of Viral RNA Degrading Factors in Shutoff of Host Gene Expression (2022) *Annu Rev Virol* 9(1):213-238. PMID: 35671567 PMCID: PMC9530000

List of Abbreviations

4A	X-ORF mutant with 4 positively charged residues replaced by alanine
4-SU	4-thiouridine
5’ RACE	5’ rapid amplification of cDNA ends
A549	Human adenocarcinoma alveolar basal epithelial cell line
aa	Amino acid
Ankrd17	Ankyrin repeat domain 17
ANP32	Acidic nuclear phosphoprotein 32
ARE	AU-rich sequence element
ASFV	African swine fever virus
BA	BirA*-myc
BCAP31	B cell receptor associated protein 31
BHV	Bovine herpesvirus
bp	Base pairs
BSA	Bovine serum albumin
CA/7	Influenza A/California/7/09 (H1N1) virus
CDC	Center for Disease Control
CFIm	Cleavage factor I
CFIIm	Cleavage factor II
cGAS	cyclic GMP-AMP synthase
ChIP	Chromatin immunoprecipitation
cl	Confidence level
CPSF	Cleavage and polyadenylation specificity factor
CRM1	Cellular chromosome region maintenance 1
cRNA	Complementary RNA
CstF	Cleavage stimulatory factor
CTD	Carboxy-terminal domain
ctrl	Control
DI	Defective interfering
DMEM	Dulbecco’s modified Eagle’s medium
DNase	Deoxyribonuclease
dox	Doxycycline
DPV	Duck plague virus
EBV	Epstein-Barr virus
eIF	Eukaryotic translation initiation factor
endoRNase	Endoribonuclease
EPC1	Enhancer of polycomb homolog 1
FluPol	Influenza RNA-dependent RNA polymerase
FPKM	Fragments per kilobase of transcript per million mapped reads
FT	Flow through
G6PD	Glucose-6-phosphate dehydrogenase
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
GC	Guanine-cytosine
GFP	Green fluorescent protein
GO	Gene ontology
H1N1pdm09	Influenza A/Tennessee/1-560/2009 (H1N1) virus
HA	Hemagglutinin
HEK	Human embryonic kidney
hnRNP	Heterogeneous nuclear ribonucleoprotein

hpi	Hours post infection
HR	Hammerhead ribozyme
HSV	Herpes simplex virus
IAV	Influenza A virus
IBV	Influenza B virus
IFN	Interferon
INSIG1	Insulin induced gene 1
iPA-X	Doxycycline-inducible PA-X
IPO7	Importin 7
IRES	Internal ribosome entry site
IRF	Interferon regulatory factor
ISG	Interferon stimulated gene
ish	Inducible small hairpin RNA
kb	Kilo base
ko	Knock out
KSHV	Kaposi's sarcoma-associated herpes virus
lnc RNA	Long non-coding RNA
luc	Luciferase
M1	Matrix protein
M2	Membrane ion channel
MDA5	Melanoma differentiation-associated gene 5
MDCK	Madin-Darby canine kidney
MERS-CoV	Middle Earth Respiratory Syndrome coronavirus
mf	Multiplicative factor
MHV68	Murine gammaherpesvirus 68
miRNA	Micro RNA
MOI	Multiplicity of infection
mRNA	Messenger RNA
mRNP	Messenger ribonucleoprotein
N term	PA-X ribonuclease domain
NA	Neuraminidase
Nat	N-terminal acetylase
NCL	Nucleolin
n.d.	Not detected/not defined
NEP	Nuclear export protein
NF-κB	Nuclear factor-kappa B
NLS	Nuclear localization signal
NP	Nucleoprotein
NPM1	Nucleophosmin
n.s.	Not significant
NS1	Non-structural protein 1 (for influenza viruses)
NS1-BP	NS1-binding protein
nsp1	Non-structural protein 1 (for coronaviruses)
nt	Nucleotide
NUDT21	Nudix hydrolase 21
NXF1	Nuclear RNA export factor 1
NXT1	Nuclear transport factor 2-related export protein 1
OE	Overexpressing
ORF	Open reading frame
PA	Polymerase acidic
PA(ΔX)	Influenza virus with frameshift and nonsense mutations in PA

PA(fs)	Influenza virus with frameshift mutation in PA
PA(X9)	Influenza virus with nonsense mutation in PA
PABP	Poly(A) binding protein
PAMP	Pathogen-associated molecular pattern
PA-N	PA-X ribonuclease domain
PAS	Poly(A) signal
PB1	Polymerase basic 1
PB2	Polymerase basic 2
PBS	Phosphate buffered saline
PEI	Polyethylenimine
PCR	Polymerase chain reaction
Perth	Influenza A/Perth/16/2009 (H3N2) virus
PKR	Protein kinase R
PMSF	Phenylmethylsulfonyl fluoride
Pol	RNA polymerase
PR8	Influenza A/PuertoRico/8/1934 (H1N1) virus
PRPF4	Pre-mRNA processing factor 4
PRR	Pattern recognition receptor
PrV	Pseudorabies virus
PUF60	Poly(U) binding splicing factor 60
RBM39	RNA binding motif protein 39
RBP	RNA binding protein
RdRp	RNA-dependent RNA polymerase
RIG-I	Retinoic acid-inducible gene-I
RIPA	Radioimmunoprecipitation assay buffer
RNA-seq	RNA sequencing
RNAP	RNA polymerase
RNase	Ribonuclease
rRNA	Ribosomal RNA
Rrp6	Ribosomal RNA-processing protein-6
RT-qPCR	Real-time quantitative polymerase chain reaction
SARS-CoV	Severe acute respiratory syndrome-related coronavirus
SDS	Sodium dodecyl sulfate
SG	Stress granule
siRNA	Small interfering RNA
SL	Stem loop
SLAM-Seq	Thiol(sh)-linked alkylation for the metabolic sequencing of RNA
SLC7A5	Solute carrier family 7 member 5
SMG6	SMG6 nonsense mediated mRNA decay factor
snRNP	Small nuclear ribonucleoprotein
STAT	Signal transducer and activator of transcription
STOML2	Stomatin-like protein 2
SOX	Shutoff and exonuclease
TAF7	TATA-box binding protein associated factor 7
TCID50	50% tissue culture infectious dose
TEAB	Triethylammonium bicarbonate
TERC	Telomerase RNA component
TFA	Trifluoroacetic acid
TGEV	Transmissible gastroenteritis virus
TLR	Toll-like receptor
TMPRSS2	Transmembrane protease serine S1 member 2

TREX	Transcription/Export complex
TTP	Tristetraprolin
TUBA1B	Tubulin alpha 1b
Udorn	Influenza A/Udorn/72 (H3N2) virus
UMI	Unique molecular identifier
UTR	Untranslated region
VacV	Vaccinia virus
vhs	Virion host shutoff
VP	Virion protein
vRNP	Viral ribonucleoprotein
VZV	Varicella zoster virus
W	Wash
WT	Wild-type
X41	41 amino acid X-ORF
X61	61 amino acid X-ORF
YKT6	YKT6 V-SNARE Homolog
ZAP	Zinc-finger antiviral protein

Chapter 1. Introduction

1.1. Influenza A virus

With hundreds of thousands of deaths every year worldwide, influenza viruses are a major public health threat. The Center for Disease Control (CDC) estimates that between 2010 and 2020, influenza virus has caused between 9 and 41 million illnesses, 140,000 and 710,000 hospitalizations, and 12,000 and 52,000 deaths annually in the United States alone (CDC, 2022a). This volume of infections also puts a high economic burden on countries. In 2007, it was estimated that the seasonal influenza epidemic cost the Unites States $10.4 billion in direct annual medical costs and $87.1 billion when also taking into account lost productivity from missed work days and lost lives (Molinari et al., 2007). In addition to seasonal influenza, there is a constant threat of influenza pandemic emergence from zoonotic infections. Previous examples include the famous so called 1918 “Spanish Influenza” that killed 2% of the world population (Johnson and Mueller, 2002), and the 2009 H1N1 pandemic that caused between 150,000 and 576,000 deaths worldwide, depending on the estimates (Dawood et al., 2012). While seasonal vaccines are available, immunity is short lived (Davis et al., 2020) and varies in efficiency (CDC, 2022b) so new vaccines have to be designed every year, and few vaccines are currently approved against zoonotic strains. A few therapeutic options are also available, but instances of drug resistance have been observed (Nguyen et al., 2012). There is thus an urgent need to develop new and improved vaccines and therapeutic solutions to fight this influenza threat.

An important aspect of influenza pathogenesis is that most of the morbidity and mortality of influenza do not come from the virus per se, but from our body’s immune reaction against the virus. Indeed, excessive inflammation in response to infection can lead to severe lung damage, which can in turn provoke acute respiratory distress

syndrome and death from respiratory failure and/or fatal bacterial pneumonia (Herold et al., 2015). To develop new therapeutic solutions, it is thus critical to better understand how immune responses and inflammation are modulated during infection and how the virus impacts these responses.

1.1.1. Classification

Influenza viruses are negative-sense, single-stranded, segmented RNA viruses from the *Orthomyxoviridae* family (Wright P, Neumann G, Kawaoka Y., 2013). This family is unusual amongst RNA viruses as they replicate their genome in the nucleus. Influenza viruses are enveloped viruses that are pleomorphic. They form spherical virions of 100 nm diameter in tissue culture but are usually found as filamentous elongated structures of over 300 nm in fresh clinical isolates (Dadonaite et al., 2016; Wright P, Neumann G, Kawaoka Y., 2013). There are four types of influenza viruses: A, B, C and D. Influenza C viruses can infect humans, dogs and pigs but usually only cause mild disease in young children (Sederdahl and Williams, 2020), while influenza D viruses primarily infect cattle with some spillover notably in swine, but overall remain a low threat to humans (Liu et al., 2020). In contrast, influenza A and B viruses can cause seasonal epidemics in humans, and thus constitute the biggest threats to humans. In addition, while influenza B viruses primarily replicate in humans, influenza A viruses constantly circulate in many animal hosts such as humans, birds, pigs, horses, seals and bats, and their main reservoirs are waterfowls (Medina and García-Sastre, 2011). This widespread circulation brings the additional worry that influenza A viruses can evolve rapidly outside of the human population, so they have the potential to cause severe disease from zoonotic infections, as well as pandemics. For this reason, my thesis work has focused on influenza A virus (IAV).

IAVs are further divided into subtypes, based on two glycoproteins in the virion envelope: hemagglutinin (HA) and neuraminidase (NA) (Wright P, Neumann G, Kawaoka Y., 2013). There are 18 HA subtypes (H1 through H18) and 11 NA subtypes (N1 through N11). While all the different combinations of HA and NA are technically possible, about 130 have been observed in nature, and only the two subtypes, H1N1 and H3N2, currently circulate in humans. Additionally, subtypes H17N10 and H18N11 differ from the other subtypes, as H17 and H18 do not use sialic acid as a receptor unlike all other HAs, and are only found in certain bat species (Ciminski and Schwemmle, 2021).

1.1.2. Influenza A virus proteins and life cycle

The IAV genome is comprised of 8 RNA segments, which encode for 10 core proteins summarized in **Table 1.1** (Wright P, Neumann G, Kawaoka Y., 2013). In addition, IAVs also encode 1-8 accessory proteins, depending on the strain, that are produced through alternative splicing or non-canonical mechanisms (Vasin et al., 2014; Yamayoshi et al., 2015) (**Table 1.2**).

Table 1.1: Summary of IAV core proteins.

Segment	Protein	Symbol	Function
1	Polymerase basic 2	PB2	Subunit of RdRp, cap recognition
2	Polymerase basic 1	PB1	Subunit of RdRp, elongation
3	Polymerase acidic	PA	Subunit of RdRp, cap snatching
4	Hemagglutinin	HA	Viral entry
5	Nucleocapsid protein	NP	Viral RNA coating
6	Neuraminidase	NA	Virion release
7	Matrix protein	M1	Viral egress, coat beneath membrane
	Membrane ion channel	M2	Genome unpacking (alternative splicing of M mRNA)
8	Non-structural protein 1	NS1	Immunomodulation, host shutoff
	Nuclear export protein	NEP	Nuclear export of vRNPs (alternative splicing of NS mRNA)

Table 1.2: Summary of IAV accessory proteins.

Segment	Protein	Mechanism	Function
1	PB2-S1	Alternative splicing	Immunomodulation, interferes with RdRp
2	PB1-F2	Leaky scanning	Immunomodulation, pro-apoptotic
	PB1 N40	Leaky scanning	Maintains PB1/PB1-F2 balance
3	PA-X	+1 frameshift	Host shutoff, immunomodulation
	PA-N155	Leaky scanning	Unknown
	PA-N182	Leaky scanning	Unknown
7	M42	Alternative splicing	Similar function as M2, help with segment evolution
8	NS3	Alternative splicing	Potential adaptation to mouse host

Influenza A virus entry into cells is mediated by the viral glycoprotein HA. HA trimers bind to sialic acids on the host cell surface, triggering endocytosis of the virus. The drop in pH as part of the normal endosome maturation then leads to a change in HA structural conformation, which allows for fusion of the viral membrane with the endosomal membrane, and subsequent release of the viral genome into the host cell cytoplasm (Sempere Borau and Stertz, 2021; Skehel and Wiley, 2000). Only HA proteins that have previously been cleaved by cellular proteases such as transmembrane protease serine S1 member 2 (TMPRSS2) can undergo conformational change and fusion (Bestle et al., 2021; Skehel and Wiley, 2000). NA also participates in viral entry by cleaving sialic acid on mucins, allowing the virion to escape non-productive binding (Cohen et al., 2013; Yang et al., 2014). Inside the virions, the viral genome segments are encapsidated by the nucleocapsid protein (NP) and capped by a trimer formed by the polymerase basic 1 (PB1), PB2 and polymerase acidic (PA) proteins, forming the viral ribonucleoproteins (vRNPs). NP protects the viral genome from damage and prevents its recognition by the host cell, and remains associated with the segments throughout the viral replication cycle. The PB1, PB2 and PA trimer forms the viral RNA-dependent RNA polymerase (RdRp), which carries out both replication of the viral RNAs and viral messenger RNA (mRNA) transcription (Wright P, Neumann G, Kawaoka Y., 2013). The vRNPs are

connected to the envelope by the matrix protein M1. To be released into the host cytoplasm, the vRNPs need to detach from M1 and the viral envelope, a step called uncoating. Low pH in the late endosome activates the ion channel M2, leading to an accumulation of H^+ and K^+ ions inside the virion and to the release of the vRNPs (Sempere Borau and Stertz, 2021).

Once uncoated, vRNPs are trafficked to the nucleus, where replication and transcription occur. Transcription of viral mRNAs is initiated by cap snatching, i.e. stealing of the host mRNA 5' cap by the RdRp. This process entails binding of the RdRp to the host RNA polymerase II carboxy-terminal domain, binding of PB2 to the 5' cap of nascent host RNAs, and cleavage of these RNAs 10-14 bases downstream of the 5' cap by PA (Fodor and Velthuis, 2020). The resulting capped RNA fragment is then used as a primer by the RdRp to initiate transcription of viral mRNAs. Transcription eventually ends by stuttering of the RdRp on a small stretch of uridine residues near the end of the vRNA template, leading to the addition of a poly(A) tail (Poon et al., 1999; Robertson et al., 1981). IAV segments 7 (encoding M1 and M2), 8 (encoding NS1 and NEP) and in some strains segment 1 (encoding PB2 and in some cases PB2-S1) are spliced by the host splicing machinery, with the help of additional host proteins such as NS1-binding protein (NS1-BP) and heterogeneous nuclear ribonucleoprotein K (hnRNPK) (Tsai et al., 2013). Since they are capped and possess a poly(A) tail, viral mRNAs are disguised as host mRNAs and can be exported to the cytoplasm and translated by the host machinery.

Once enough viral proteins have been synthesized, the RdRp is thought to switch to replicase activity. Replication of the viral genome occurs in two steps, and unlike transcription, is primer independent. In the first step, the RdRp copies viral RNAs into positive sense complementary RNAs (cRNAs) replicative intermediates that are immediately coated with NP and bound by a newly synthesized RdRp (Fodor and Velthuis, 2020). This cRNA encapsidation requires the new RdRp to dimerize with the

replicating RdRp. The host protein acidic nuclear phosphoprotein 32 (ANP32) is also required to stabilize the RdRp dimer and to promote NP coating (Carrique et al., 2020; Wang et al., 2022). In the second step, the RdRp dimer uses these cRNP intermediates as templates to produce negative sense viral RNAs, which are also immediately coated with NP and a new RdRp (Fodor and Velthuis, 2020).

The last step of the life cycle is viral assembly and budding, which primarily occurs at the apical plasma membrane. HA, NA and M2 are trafficked from the ER to the cell surface and incorporated into the plasma membrane (Nayak et al., 2004). Following replication, newly synthesized vRNPs are exported to the cytoplasm via a multiprotein complex that includes the host protein cellular chromosome region maintenance 1 (CRM1) and the viral proteins NEP and M1, and are trafficked to the plasma membrane for budding of the new virion (Giese et al., 2016). As M1 lies just beneath the lipid envelope and interacts with both vRNPs and HA/NA/M2, it is thought to serve as a bridge between the vRNPs and the cytoplasmic tails of the envelope glycoproteins (Nayak et al., 2009). Packaging of all 8 viral RNA segments is promoted by packaging signals within each segment sequence, as well as RNA-RNA interactions between segments (Giese et al., 2016). Packaging of fewer or wrong segments that contain polymerase replication errors can form defective interfering (DI) particles that compete with wild type (WT) viruses (Alnaji and Brooke, 2020). After completion of budding, NA removes the sialic acid at the cell surface that would otherwise be bound to HA, allowing for the release of the viral particle (Basak et al., 1985; Palese et al., 1974).

1.1.3. Infection and disease

Human IAV epidemics are usually seasonal, occurring predominantly during the colder months, although influenza activity can be detected all year round (Azziz Baumgartner et al., 2012). Human IAVs cause acute infections in non-

immunocompromised individuals (Wright P, Neumann G, Kawaoka Y., 2013) and primarily infect the respiratory tract. This tissue tropism in part comes from the fact that the cellular proteases that can cleave HA are mostly expressed in the respiratory tract, allowing for infection to occur there. In contrast, some highly pathogenic avian influenza strains can cause systemic infections in chickens because their HA can be cleaved more promiscuously by ubiquitous cellular proteases (Wright P, Neumann G, Kawaoka Y., 2013). Epithelial cells in the human respiratory tract predominantly possess sialic acids with a2,6-linkages on their surface, including in the nasal mucosa, paranasal sinuses, pharynx, trachea, and bronchi, whereas a2,3-linkages are found deeper within the lungs in some bronchiolar and alveolar cells (Shinya et al., 2006). For this reason, humans are predominantly infected by IAVs with a2,6 specificity (i.e. human virus isolates) and not a2,3 specificity (i.e. avian virus isolates), but can still in some cases be infected by avian IAVs (Couceiro et al., 1993). In addition, lung epithelial cells are polarized. IAVs use this characteristic to bud and release virions from the apical side of the cell, which leads to an accumulation of virions in the lumen of the respiratory tract and promotes transmission (Wright P, Neumann G, Kawaoka Y., 2013). Transmission occurs the most effectively through aerosols from the cough or sneeze of an infected individual, leading to direct person-to-person spread from breathing these aerosols (Uyeki et al., 2022).

Infection with human seasonal IAVs cause mild to severe respiratory symptoms that can be fatal, especially in the very young and old populations. In a healthy individual, viral replication usually peaks about 2 days after inoculation, and slowly declines the following days up to little shedding after 6 to 8 days (Wright P, Neumann G, Kawaoka Y., 2013). This is associated with a balanced activation of the innate immune response, which leads to lung inflammation, triggering of the adaptive immune response and tissue repair for efficient recovery. However, failure or dysregulation of the immune response in response to infection can lead to the over activation of pro-inflammatory cytokines and

chemokines, leading to acute lung injury, respiratory distress syndrome, sepsis and multiorgan failure (Uyeki et al., 2022). Aberrant coagulation and secondary bacterial infections are common complications that further contribute to viral replication and immune pathogenesis (Morris et al., 2017; Yang and Tang, 2016). In addition, increased disease severity is often associated with infection of the lower respiratory tract (Uyeki et al., 2022).

Some antiviral therapeutics are available and effective if taken early (i.e. during the viral replication phase). They include the NA inhibitors oseltamivir, zanamivir, laninamivir, and peramivir, and the cap snatching inhibitor baloxavir (Uyeki et al., 2022). Antivirals against the M2 channel also exist (amantadine and rimantadine), but they are of low efficiency since many circulating IAVs have already acquired resistance to them (Uyeki et al., 2022). Yearly vaccines also lower the risk of severe disease and death, although their effectiveness can vary (Flannery et al., 2020, 2019; Rolfes et al., 2019; Tenforde et al., 2021). The development of drug resistance and the need for yearly vaccines are due to antigenic drift, i.e. the gradual evolution of IAV over time from accumulation of mutations by the error prone RdRp. Unlike most other viruses, the segmented nature of IAVs also allows them to undergo antigenic shift, i.e. viral segment reassortment between two IAVs during a co-infection, leading to the introduction of a new or antigenically distinct HA or NA that was not previously circulating in humans (Wright P, Neumann G, Kawaoka Y., 2013). While antigenic drift causes seasonal epidemics, antigenic shift can lead to high infection rates and pandemics because of no prior immunity in the population.

In view of the major role that inflammation plays in IAV morbidity and mortality, it is critical to understand how inflammation and the immune response in general is modulated during infection. Importantly, IAV itself is able to regulate host gene expression during infection, and with it, the innate immune response that starts off this

immune cascade. The next section will look into more details at the mechanisms by which IAV can modulate host gene expression.

1.2. Modulation of host gene expression by influenza A virus

Any virus needs components of the cellular machinery to be able to replicate its genome and make progeny virions. It thus competes against the cell for its resources. It also has to protect itself from any attacks by the cell. To do this, many viruses have evolved strategies to limit host gene expression to a bare minimum, in a process called host shutoff. By globally decreasing host gene expression, cellular resources and ribosomes are now free to be used for viral replication and to make viral proteins. Moreover, the cell is no longer able to make cytokines and other antiviral RNAs and/or proteins to defend itself. To understand the different strategies used by IAV to modulate host gene expression and induce host shutoff, we first need to review the mechanism of host gene expression in human cells.

1.2.1. Normal life cycle of cellular mRNAs

Human cells possess three different RNA polymerases (RNAP or Pol I, II and III) with which they transcribe RNAs. All protein-coding mRNAs are synthesized by Pol II, although Pol II also transcribes some noncoding RNAs. Pol II transcription occurs in the nucleus (**Figure 1.1**), and begins with assembly of the Pol II subunits on a gene promoter, along with transcription factors that regulate escape from the promoter and productive elongation (Sainsbury et al., 2015). Elongation and termination are then regulated by a series of phosphorylation events and other post-translational modifications on the heptad repeats (consensus Tyr1–Ser2–Pro3–Thr4–Ser5–Pro6–Ser7) of the Pol II C-terminal domain (CTD) (Hsin and Manley, 2012). While the actual pattern of phosphorylation is complex, a general observation is that phosphorylation of

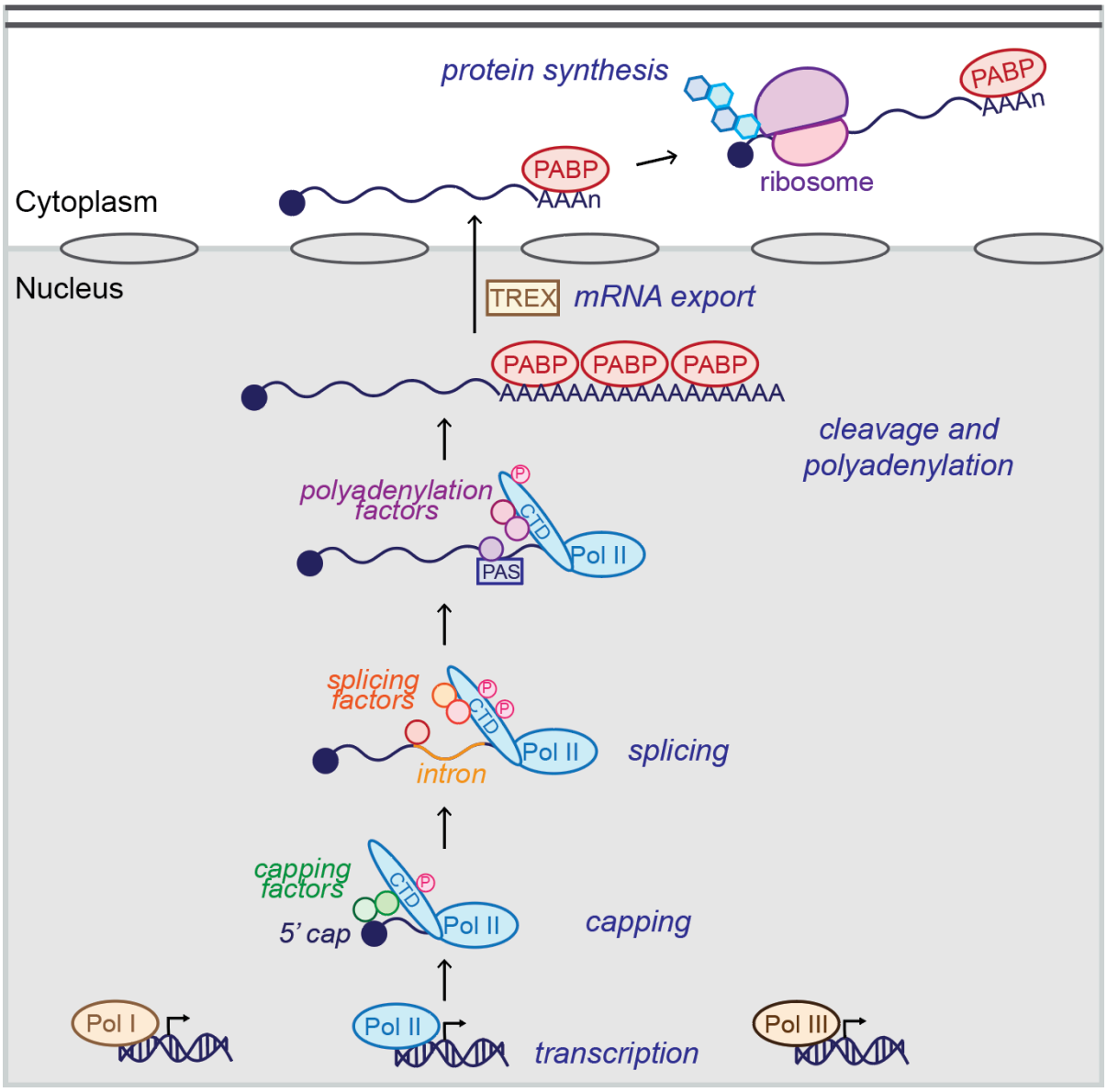

Figure 1.1: Life cycle of cellular mRNAs.

Of the three cellular RNA polymerases, Pol I, II and III, only Pol II transcribes mRNAs. mRNA processing occurs co-transcriptionally, and consists of the addition of a 5' methylguanosine cap, the splicing of introns (for most mRNAs), and cleavage and polyadenylation. These events are coordinated by the Pol II C-terminal domain (CTD), which binds to various splicing factors and polyadenylation factors over time. In turn, phosphorylation (P) of specific amino acids on the CTD regulates the progression of transcription. Cleavage and polyadenylation complexes recognize the poly(A) signal (PAS), leading to cleavage of the nascent RNA and addition of hundreds of adenosines that bind Poly(A) binding proteins (PABP). Finally, mRNAs are exported to the cytoplasm by the Transcription/Export (TREX) complex, or through other less common mechanisms. Once in the cytoplasm, mRNAs can be translated into proteins thanks to ribosomes.

Ser5 is more abundant during transcription initiation, while a shift to Ser2 phosphorylation occurs during elongation (Hsin and Manley, 2012). Multiple RNA processing steps are associated with Pol II elongation and termination and are coordinated by the Pol II CTD. First, a 5′ methylguanosine cap is added on the nascent RNA after synthesis of the first 20-30 nucleotides (nt) (Proudfoot et al., 2002). Second, most human nascent transcripts possess non-coding regions called introns (as opposed to coding regions that are called exons) that are removed in the mature mRNA, in a process termed splicing. Intron boundaries are defined by specific sequences, most notably the canonical GU and AG dinucleotides that mark the splice sites at the 5' and 3' ends of introns, respectively, and the branchpoint adenosine sequence 18–40 nt upstream of the 3′ splice site (Matera and Wang, 2014). These sequences are recognized by small nuclear ribonucleoproteins (snRNPs) that form the spliceosome and in a coordinated event lead to removal of the intron and ligation of the two exons (Matera and Wang, 2014). The Pol II CTD coordinates this process by inducing Pol II pausing and helping to recruit splicing factors for the nascent RNA to be spliced (Herzel et al., 2017; Hsin and Manley, 2012). Finally, the Pol II CTD also recruits and/or stabilizes termination and polyadenylation factors to the nascent RNA. Polyadenylation consists of endonucleolytic cleavage of the nascent RNA, followed by the addition of hundreds of non-templated adenosines that form the poly(A)tail. Four protein complexes largely direct these events: components of the cleavage and polyadenylation specificity factor (CPSF) bind a poly(A) signal (usually AAUAAA) and cleave the RNA, the cleavage stimulatory factor (CstF) recognizes and binds a U- or GU- rich motif downstream of the poly(A) signal, and the cleavage factor I and II (CFIm and CFIIm) complexes promote the selection of distal and proximal polyadenylation sites respectively through recognition of additional regulatory sequences (Dharmalingam et al., 2022). Nuclear poly(A) binding proteins (PABPs) bind the emerging poly(A) tail, and in turn stimulate the

polyadenylation process and define the poly(A) tail length in humans (Kühn et al., 2017). Once the mRNA is fully processed, it is exported through nuclear pore complexes to the cytoplasm to be translated into proteins by the ribosome. Most spliced mRNAs are transported by the splicing-dependent Transcription/Export (TREX) complex, although a subset of mRNAs use a pathway dependent on eukaryotic translation initiation factor 4E (eIF4E) instead. Intronless mRNAs also use a similar TREX pathway, through interactions with specific sequence motifs found on intronless mRNAs (Khan et al., 2022). Noncoding RNAs transcribed by Pol II are processed similarly to mRNAs, i.e. they are capped, tailed, and spliced if they contain introns. However, they do not associate with the translation machinery and in some cases stay in the nucleus. In contrast, noncoding RNAs transcribed by Pol I and III are processed by different machinery and in different ways, as the protein complexes responsible for capping, splicing and polyadenylation do not associate with Pol I and III.

Throughout the mRNA life cycle, constant surveillance is at work to ensure that every step of biogenesis is correctly carried out. Non-adenylated mRNAs, 3' unprocessed mRNAs and mRNAs with improper RNA binding protein loading are degraded in the nucleus by the ribosomal RNA-processing protein-6 (Rrp6p)-containing nuclear exosome (Houseley et al., 2006). Once in the cytoplasm, specific quality control pathways are triggered based on the type of defect. The nonsense-mediated decay pathway degrades mRNAs that contain premature termination codons (Lykke-Andersen and Jensen, 2015), while the non-stop decay pathway degrades mRNAs that lack termination codons (Frischmeyer et al., 2002; van Hoof et al., 2002), and the no-go decay pathway degrades mRNAs on which ribosomes have stalled (Doma and Parker, 2006). On top of these pathways, basal mRNA decay regulates the general level of mRNA in the cell. It begins with deadenylation of the mRNA by the major deadenylases Pan2-Pan3 and Ccr4-Not (Mugridge et al., 2018). PABPs are released as the poly(A) tail

is shortened, which destabilizes the mRNA (Yi et al., 2018). The deadenylation machinery also interacts with the core decapping complex, composed of the decapping enzyme Dcp2, the decapping activator Dcp1 and the scaffold protein Edc4, leading to the removal of the mRNA 5' cap (Mugridge et al., 2018). This removal in turn triggers degradation of the mRNA mostly by the 5'-3' exonuclease Xrn1, with some participation of the 3'-5' exonuclease Dis3L2 and the exosome (Łabno et al., 2016; Lubas et al., 2013). mRNA decay can also be further stimulated by the addition of uridines, or uridylation, at the 3' end of deadenylated transcripts by terminal uridylyl transferases (Łabno et al., 2016; Lim et al., 2014). Finally, certain sequences on the RNA can regulate the half-life of RNAs. The most well studied are AU-rich sequence elements (AREs), which can be bound by multiple RNA-binding proteins to either promote or block RNA degradation (Otsuka et al., 2019).

1.2.2. RdRp/NS1

Since all mRNAs are transcribed by Pol II, host shutoff mechanisms usually target Pol II transcribed RNAs, as degrading these RNAs also prevents protein expression (Gaucherand and Gaglia, 2022). To do so, host shutoff proteins highjack specific parts of the RNA processing and/or translation pathways. IAV carries out host shutoff through at least three proteins or protein complexes: PA-X, which is the focus of my thesis and I will review in the next section, NS1 and the RdRp complex (i.e. PB1, PB2 and PA).

Widespread degradation of host mRNAs is a classical way to induce host shutoff, and was reported to happen during IAV infection as early as 1982 (Inglis, 1982). Prior to the discovery of PA-X, host RNA degradation was thought to be a by-product of cap snatching by the RdRp. Indeed, host RNAs that have been cap snatched become susceptible to 5'-3' degradation by host exonucleases such as Xrn1 or Xrn2. While we now know that PA-X is the main driver of host mRNA degradation in IAV infected cells

(Desmet et al., 2013; Hayashi et al., 2016; Jagger et al., 2012), cap snatching still partially contributes to host shutoff by altering host RNA transcription. This is because the RdRp needs to bind to the Pol II CTD to carry out cap snatching of nascent RNAs (Engelhardt et al., 2005), and this association prevents Pol II elongation and likely decreases host transcription (Bauer et al., 2018; Chan et al., 2006). Specifically, the RdRp binds to the initiating Ser5 phosphorylated form of Pol II but not the elongating Ser2 phosphorylated form (Engelhardt et al., 2005; Martínez-Alonso et al., 2016). This finding suggests that cap snatching occurs early during transcription, which is in agreement with chromatin immunoprecipitation (ChIP) data showing that the RdRp is mainly associated with gene promoter regions (Chan et al., 2006). In addition to binding the CTD, IAV RdRp also interacts with multiple Pol II associated factors (Bradel-Tretheway et al., 2011; Naito et al., 2007; Zhang et al., 2010), as well as the nuclear exosome, which may promote the recruitment of the RdRp to gene promoter regions (Rialdi et al., 2017). As a result, Pol II is depleted from gene bodies (Bauer et al., 2018; Chan et al., 2006). A defect in transcription termination has also been noted, which is partially due to NS1 inhibiting host mRNA 3' end processing but also through an NS1-independent mechanism, probably from the RdRp (Bauer et al., 2018; Zhao et al., 2018). Finally, the RdRp of some virulent strains has also been involved in mediating Pol II degradation, although reports differ on whether the ubiquitin pathway is involved (Rodriguez et al., 2007; Vreede et al., 2010). One study also identified residues 504 of PB2 and 550 of PA as critical for Pol II degradation (Llompart et al., 2014).

In addition to the RdRp and PA-X, the NS1 protein has also been involved in host shutoff. NS1 is well-known to block the antiviral interferon (IFN) pathway, as IAVs with deleted NS1 only replicate in IFN-defective cells (García-Sastre et al., 1998; Krug, 2015). Multiple mechanisms have been described for how NS1 achieves this, although these mechanisms do not occur in all IAVs (Hale et al., 2008; Krug, 2015). Although NS1

can directly bind and regulate proteins in the IFN pathway, NS1 host shutoff through inhibition of host mRNA 3' end processing has also been implicated in the IFN evasion (Bauer et al., 2018; Heinz et al., 2018; Nacken et al., 2021; Zhao et al., 2018). NS1 proteins from multiple, but not all, IAV strains interact with CPSF30, a component of the CPSF complex, and in doing so prevent cleavage and polyadenylation of the nascent RNA and thus protein expression (Nemeroff et al., 1998). While there are exceptions, in general, residues at position 103 and 106 determine whether NS1 from a specific IAV strain binds efficiently to CPSF30 or not (Hale et al., 2008; Krug, 2015). Notably, the NS1 protein from the commonly used lab adapted influenza A/PuertoRico/8/1934 (H1N1) virus (PR8) strain and the original 2009 pandemic H1N1 strain that jumped to humans from swine are not thought to have this CPSF30 binding function (Hale et al., 2010; Hayman et al., 2006; Kochs et al., 2007). This host shutoff activity of NS1 nonetheless appears to be less global than PA-X host shutoff activity and more targeted towards IFN and the host immune response (Chaimayo et al., 2018; Noah et al., 2003). It is also difficult to detangle whether this host shutoff is specifically due to NS1 interacting with CPSF30 or another mechanism, especially since some mechanisms are strain specific. Indeed, other less characterized mechanisms of NS1 interference with host gene expression have been observed, such as interaction with nuclear PABP to reduce the elongation of the poly(A) tail (Chen et al., 1999), and interactions with the mRNA export receptor complex, nuclear RNA export factor 1-nuclear transport factor 2-related export protein 1 (NXF1-NXT1), to prevent export of host mRNAs to the cytoplasm (Satterly et al., 2007; Zhang et al., 2019). NS1 can also directly bind pre-mRNA intronic sequences, including the pre-mRNA of retinoic acid-inducible gene-I (RIG-I), the main cytoplasmic antiviral sensor of influenza virus infection, which slows down mRNA processing and host gene expression (Zhang et al., 2018). Finally, NS1 has been implicated in modulating host gene expression by promoting viral mRNA

translation (Hale et al., 2008). This modulation could be mediated by direct interaction with the ribosome, eukaryotic initiation factor 4 gamma 1 (eIF4G1) and PABP (Aragón et al., 2000; Arias-Mireles et al., 2018; Burgui et al., 2003; Cruz and Joseph, 2022; Panthu et al., 2017). However, this effect may be strain specific, as a transcriptomic study found no translational advantage of viral mRNAs over host mRNAs in the PR8 strain (Bercovich-Kinori et al., 2016).

1.2.3. PA-X

As previously stated, the main protein responsible for RNA degradation-dependent host shutoff in IAV is the endoribonuclease PA-X. PA-X was only discovered ten years ago, as it is non-canonically produced from +1 ribosomal frameshifting of segment 3 mRNA (Jagger et al., 2012) (**Figure 1.2**). Indeed, during translation, the ribosome stalls at a specific U-rich stretch followed by a rare codon (UCC UUU CGU C), a sequence that is fully conserved in 98% of IAV strains (Firth et al., 2012; Jagger et al., 2012). In a small fraction of translation runs, the ribosome resolves this stall through a +1 shift in the codon reading frame, generating PA-X's unique C-terminal region termed the X-ORF (Firth et al., 2012; Jagger et al., 2012) (**Figure 1.2**). Despite this non-canonical production mechanism, PA-X is encoded by all IAVs, attesting to its importance for the virus (Shi et al., 2012). Since the frameshifting event occurs after translation of amino acids (aa) 1-191 of PA, PA-X and PA share the same N-terminal RNase domain. This RNase domain, which is used for cap-snatching in PA (Decroly et al., 2012), belongs to the PD-D/E-X-K superfamily (Dias et al., 2009; Yuan et al., 2009). Cleavage by this family of nucleases leaves an RNA fragment with a free 5' phosphate that is then degraded by the host exoribonuclease Xrn1 (Khaperskyy et al., 2016).

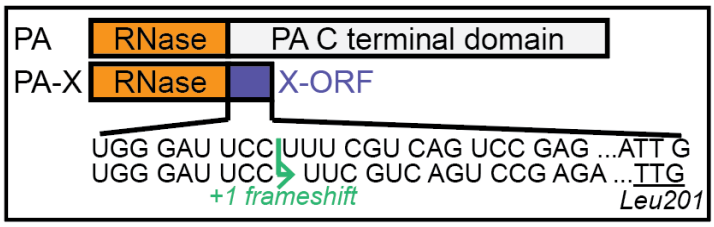

Figure 1.2: Diagram of the mechanism of PA-X production.

In a small fraction of translation runs of the PA mRNA, +1 ribosomal frameshifting occurs, leading to the production of the PA-X protein. PA-X thus has the same N-terminal ribonuclease (RNase) domain as PA but a unique C-terminal domain termed the X-ORF. Figure adapted with permission from Gaucherand L, Porter BK, Levene RE, Price EL, Schmaling SK, Rycroft CH, Kevorkian Y, McCormick C, Khaperskyy DA, Gaglia MM. 2019. *Cell Reports* 27:776-792.e7 and *Gaucherand L, Gaglia MM. 2022. The Role of Viral RNA Degrading Factors in Shutoff of Host Gene Expression. Annual Review of Virology 9:213–238*. Changes include change of X-ORF color and label.

Expression of PA-X leads to a global decrease in host gene expression, as observed by RT-qPCR analyses and metabolic labeling experiments (Jagger et al., 2012; Khaperskyy et al., 2016). Yet, more in depth analysis of down-regulated RNAs revealed some specificities for PA-X activity. Indeed, PA-X only down-regulates RNAs transcribed by Pol II, and not Pol I or III (Khaperskyy et al., 2016). Interestingly, this trait is shared by other viral RNases, as we will discuss in the next section [reviewed in (Gaucherand and Gaglia, 2022)]. Consistent with this finding, PA-X does not impact the levels of genomic viral RNA and viral mRNA, which are produced by the RdRp (Khaperskyy et al., 2016). Of note, some non-coding RNAs that are transcribed by Pol II are down-regulated by PA-X, suggesting that PA-X activity is linked to transcription and not translation (Khaperskyy et al., 2016). Conversely, PA-X does not down-regulate an mRNA that is translatable but has not gone through canonical 3' end processing (Khaperskyy et al., 2016). Consistent with this link to RNA biogenesis rather than translation, PA-X can be found in the nucleus, and various PA-X mutants that are fully cytoplasmic are not active (Hayashi et al., 2016; Khaperskyy et al., 2016). PA-X is thus thought to be active in the nucleus, although different studies disagree on whether PA-X is also able to degrade transfected mRNAs or T7 polymerase-synthesized RNAs that are restricted to the

cytoplasm (Hayashi et al., 2016; Khaperskyy et al., 2016). PA-X nuclear localization is mediated by the X-ORF, as the RNase domain alone is more cytoplasmic (Hayashi et al., 2016; Khaperskyy et al., 2016) and a chimeric GFP protein fused to the X-ORF is primarily found within the nucleus (Khaperskyy et al., 2016). In addition, point mutations in six basic amino acid residues (arginine or lysine) within the first 15 aa of the X-ORF lead to a primarily cytoplasmic localization of PA-X (Hayashi et al., 2016; Khaperskyy et al., 2016).

While the RNase domain alone (which I will refer to as “PA-N”) is still active *in vitro* (Bavagnoli et al., 2015; Datta et al., 2013; Dias et al., 2009; Yuan et al., 2009), the X-ORF is required for PA-X activity in cells, and mutations within it can abolish activity (Hayashi et al., 2016; Khaperskyy et al., 2016; Oishi et al., 2015, 2018a, 2019). Of note, fusing PA-N to a classical SV40 nuclear localization signal does not fully restore the host shutoff activity of PA-X, suggesting that the role of the X-ORF in PA-X activity goes beyond nuclear localization (Hayashi et al., 2016). Interestingly, the first 15 aa of the X-ORF are sufficient for host shutoff activity in transfected cells (Hayashi et al., 2016; Oishi et al., 2015). Yet, the sequence analysis of many IAV lineages suggests that there is evolutionary pressure to keep an X-ORF of 41 or 61 aa (depending on the IAV strain). Indeed, the sequence of PA/PA-X would allow for stop codons in the PA-X reading frame that are silent in the PA frame earlier in the X-ORF, but these have not been found in existing viral isolates (Shi et al., 2012). This discrepancy suggests that the rest of the X-ORF has an additional function, perhaps in regulation of PA-X activity. This would be in agreement with the recently identified ability of the X-ORF to regulate the stability of the PA-X protein, as PA-X is short lived and kept at low levels within the cell (Levene et al., 2021). Indeed, the half-life of PA-X ranges from 30 minutes to 3.5 hours depending on the IAV strain (Levene et al., 2021), compared to 9 hours for an average human protein (Chen et al., 2016). In particular, PA-X aa 220 appears important for PA-X

turnover in certain IAV strains (Levene et al., 2021). The X-ORF may also contribute to IAV host adaptation, as the different lengths of the X-ORF are found within distinct IAV lineages. Indeed, the shorter 41 aa X-ORF is only found in lineages that primarily infect dogs and pigs and in the 2009 pandemic H1N1 strain that jumped to humans from swine (Shi et al., 2012). A study of IAV swine isolates also analyzed the X-ORF truncation over time and found that having a 41 aa X-ORF increases viral replication and transmission in pigs (Xu et al., 2016). These extra 20 aa also lead to differences in activity in various cell types and *in vitro*, which could allow the X-ORF to modulate replication and pathogenicity (Bavagnoli et al., 2015; Feng et al., 2016; Gao et al., 2015a; Wang et al., 2020).

In addition to its regulation through the X-ORF, PA-X is also regulated through N-terminal acetylation, a co-translational modification that can influence protein subcellular localization, stability, and protein-protein interactions (Nguyen et al., 2018). The N-terminal acetylase (Nat) complex NatB is responsible for modifying PA-X, and modification by NatB only is required for proper activity of PA-X (Oishi et al., 2018b). There is also some evidence for a functional and regulatory interplay between PA-X and NS1, as both proteins seem to have coevolved in several IAV strains to limit their host shutoff activity in tandem and regulate pathogenicity and viral fitness *in vivo* (Chaimayo et al., 2018; Nogales et al., 2021, 2018, 2017).

Apart from its targeting specificity to RNAs transcribed by Pol II and its requirement for nuclear localization, nothing is really known about the mechanism of action of PA-X at the molecular level. It is also unknown how PA-X achieves the Pol II specificity. One hypothesis is that PA-X needs to interact with a host protein involved in some way in Pol II RNA transcription or processing to reach its target RNAs. While some studies have investigated the interactome of PA-X in cells, no follow up analysis was carried out to confirm and investigate the consequences of these interactions, except for the interactor

ankyrin repeat domain 17 (Ankrd17) that PA-X uses to dampen the immune response (Li et al., 2021, 2016). In addition, northern blot analyses suggested that cleavage of the RNA by PA-X occurs nonspecifically throughout the RNA, although this has not been investigated thoroughly (Khaperskyy et al., 2016). Therefore, more work is clearly needed to uncover the full molecular mechanism of action of PA-X.

1.2.4. Secondary consequences of PA-X activity

Direct degradation of RNAs by PA-X triggers secondary changes in the cell that can further intensify host shutoff (**Figure 1.3**). Widespread RNA degradation in cells leads to the redistribution of RNA-binding proteins. In particular, the previously mentioned PABP, which binds RNA poly(A) tails to stabilize RNAs and enhance translation, falls off RNAs during their degradation. Release from the RNA unmasks PABP's NLS, leading to the relocalization of PABP to the nucleus through interaction with the nuclear import machinery (Kumar et al., 2011). This in turn leads to mRNA hyperadenylation in the nucleus and subsequent inhibition of mRNA export to the cytoplasm. Host RNAs are thus trapped in the nucleus and host shutoff is enhanced (Khaperskyy et al., 2014; Kumar and Glaunsinger, 2010; Lee and Glaunsinger, 2009). PABP nuclear localization has been observed under many conditions linked to reduced RNA levels, including cellular stress such as heat shock, oxidative stress, transcriptional block or expression of viral host shutoff proteins (Kumar and Glaunsinger, 2010; Lee and Glaunsinger, 2009; Ma et al., 2009; Salaun et al., 2010). It is also a striking consequence of PA-X activity that can be directly observed by microscopy (Khaperskyy et al., 2016, 2014).

Virus-induced RNA decay also inhibits stress granule (SG) formation. SGs are cytoplasmic aggregates of mRNAs, translation machinery and mRNA-binding proteins that form upon stress-induced activation of kinases such as the double-stranded RNA-activated protein kinase R (PKR) (Anderson and Kedersha, 2008). In response to stress,

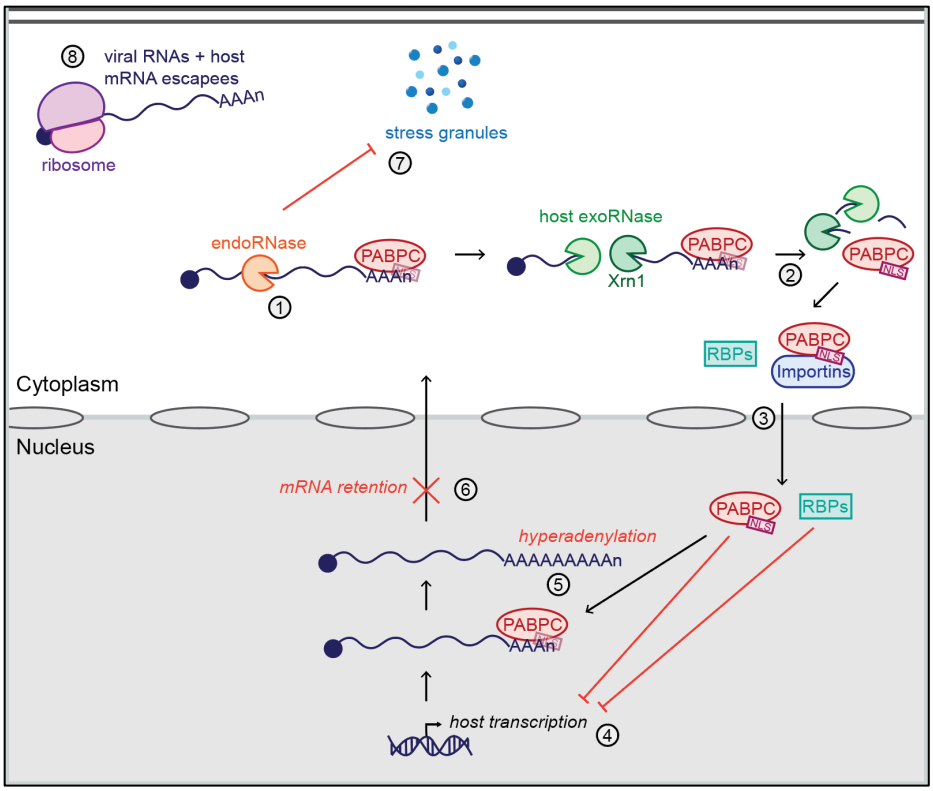

Figure 1.3: Secondary consequences of PA-X activity.
Poly(A)-binding protein (PABPC) usually binds the poly(A) tail of RNAs in the cytoplasm. However, upon cleavage of RNAs by PA-X or other endoribonucleases (endoRNases) (#1) followed by RNA fragment degradation by Xrn1 and other host exoribonucleases (exoRNases) (#2), PABPC falls off the RNA, which exposes its nuclear localization signal (NLS). PABPC and other RNA-binding proteins (RBPs) that are no longer RNA bound can now interact with nuclear import proteins and shuttle to the nucleus (#3). This leads to the inhibition of host transcription (#4). PABPC accumulation in the nucleus also leads to hyperadenylation (#6) and nuclear retention of mRNAs (#6). In parallel, PA-X activity prevents the formation of stress granules (SGs) (#7), allowing for the translation of viral mRNAs and the host mRNAs that are able to escape host shutoff (#8). Figure and legend adapted with permission from Gaucherand L, Gaglia MM. 2022. The Role of Viral RNA Degrading Factors in Shutoff of Host Gene Expression. Annual Review of Virology 9:213–238. Changes include removing elements not relevant for PA-X (viral replication in protective compartment and nsp1-specific mechanisms of nuclear retention) and modifying the numbering accordingly.

PKR phosphorylates the translation initiation factor eIF2α, which in turn inhibits translation (Anderson and Kedersha, 2008). SGs thus form during this stress response to protect host mRNAs and allow their translation to resume once stress is relieved (Anderson and Kedersha, 2008). In the context of a viral infection, SGs also serve an antiviral function by trapping viral mRNAs to prevent their translation (McCormick and

Khaperskyy, 2017). IAV actively prevents SG formation through PA-X RNase activity (Khaperskyy et al., 2014). Inhibition of SG formation is also a consequence of the activity of other viral RNases (Burgess and Mohr, 2018; Finnen et al., 2016). For example, the RNA degradation activity of the herpes simplex virus (HSV) virion host shutoff (vhs) protein limits activation of PKR by double-stranded RNA, which in turn prevents SG formation (Burgess and Mohr, 2018; Dauber et al., 2016). While the exact mechanism by which PA-X inhibits SG formation is unknown, it appears to be independent of eIF2α phosphorylation (Khaperskyy et al., 2014). It may thus just be due to the lack of cytoplasmic mRNAs from host shutoff to nucleate SGs.

Evidence recently emerged that global depletion of mRNAs in the cytoplasm can be sensed by the cell, triggering a feedback loop to inhibit host transcription (Abernathy et al., 2015). This has been described as a consequence of mRNA decay triggered by viral RNases and apoptosis, leading to a decrease in Pol II recruitment to promoters and in mRNA synthesis (Abernathy et al., 2015; Duncan-Lewis et al., 2021; Friedel et al., 2021; Hartenian et al., 2020). The exact mechanism is still unclear but seems to involve cellular mRNA decay factors such as Xrn1 and a change in the nuclear/cytoplasmic distribution of RNA-binding proteins that ultimately inhibits transcription (Duncan-Lewis et al., 2021; Gilbertson et al., 2018; Hartenian et al., 2020; Hartenian and Glaunsinger, 2019). It is thus likely that PA-X also plays a part in the host transcription inhibition following IAV infection discussed in section 1.2.2, although it has not yet been investigated.

Ultimately, PA-X-mediated RNA decay and secondary consequences of this decay define which mRNAs are translated during infection, referred to as the translatome. A few hours following IAV infection, the overwhelming majority of translated mRNAs becomes viral mRNAs (Bercovich-Kinori et al., 2016). This shift is likely a consequence of PA-X depleting host mRNAs but not targeting viral mRNAs (Khaperskyy et al., 2016),

as at least in the particular IAV strain used in this study, PR8, viral mRNAs are not translated more efficiently than host mRNAs (Bercovich-Kinori et al., 2016). Viral mRNAs also compose the majority of the translatome during infection with other viruses that deplete host mRNAs, such as HSV, Vaccinia virus (VacV) and severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (Dai et al., 2017; Rutkowski et al., 2015; Finkel et al., 2021). Conversely, some host RNAs can escape host shutoff and are still translated. These generally belong to categories of genes that are critical for the cell to function, and thus for viruses to replicate, such as genes involved in translation and oxidative phosphorylation (Bercovich-Kinori et al., 2016; Dai et al., 2017; Rao et al., 2021). Looking more specifically at PA-X activity, some RNAs that escape down-regulation by PA-X also escape down-regulation by the other viral RNases vhs and the Kaposi's sarcoma-associated herpes virus (KSHV) RNase SOX, but not by cellular RNases such as SMG6 (Muller et al., 2015; Muller and Glaunsinger, 2017; Rodriguez et al., 2019). This interesting finding suggests a potential common mechanism for protection from viral RNase activity. While the mechanism of protection has not been investigated for PA-X and vhs, specific mRNAs protected from SOX contain SOX resistance elements in their 3' UTRs that recruit RNA-binding proteins like AUF1, HuR, and nucleolin, forming a protective protein complex (Glaunsinger and Ganem, 2004a; Hutin et al., 2013; Muller et al., 2015; Muller and Glaunsinger, 2017; Rodriguez et al., 2019). N6-adenosine methylation in the SOX resistance element is also required for protection from degradation of at least some of these mRNAs (Macveigh-Fierro et al., 2022).

1.3. Immunomodulatory role of PA-X

The first study that identified PA-X also reported a striking phenotype *in vivo*: infection of a mouse model with an IAV strain that was engineered to lack PA-X caused

increased immune and inflammatory gene expression in the lungs of the animals compared to infection with the WT virus, even though both viruses replicated to the same extent (Jagger et al., 2012). Importantly, this heightened immune and inflammatory response was not protective, and instead lead to increased morbidity and mortality in this mouse model (Jagger et al., 2012). Since then, a similar phenotype has been observed across multiple IAV strains and animal models. These include mice, chickens, ducks and pigs infected with H1N1 IAV strains from the 1918 and 2009 pandemics, highly pathogenic avian H5N1 IAV strains and classical swine H1N1 and H5N1 IAV strains (Gao et al., 2015b; Gong et al., 2017; Hayashi et al., 2015; Hu et al., 2016, 2015; Jagger et al., 2012). Of note, decreased embryonic lethality was observed in chicken eggs infected with PA-X deficient avian H5N1, H7N1 and H9N2 IAV strains (Clements et al., 2021; Hussain et al., 2019), while similar or decreased virulence was observed in mice infected with PA-X deficient H9N2 and H7N9 avian IAV strains and PR8 . While these data show that lack of PA-X can have a different phenotype depending on the strain, it still generally implicates PA-X in the modulation of the host immune response.

Antiviral immune responses are initiated when infected cells sense the presence of the virus. All human cells express proteins termed pattern recognition receptors (PRRs) that can detect the presence of pathogen-associated molecular patterns (PAMPs) such as viral nucleic acids (Wu and Chen, 2014). Activation of PRRs triggers a signaling cascade that ultimately result in transcription of antiviral genes to fight the infection and of cytokines to alert neighboring cells of the infection (Wu and Chen, 2014). The major and best studied innate response pathway against viruses is the type I and III IFN pathway. Sensing of viral RNA through RNA sensors such as Toll-like receptor (TLR) 3, RIG-I or melanoma differentiation-associated gene 5 (MDA5), or of viral DNA by sensors like TLR9 or cyclic GMP-AMP synthase (cGAS), leads to phosphorylation and nuclear

import of the transcription factors interferon regulatory factor (IRF) 3 and 7 and nuclear factor-kappa B (NF-κB) (Lazear et al., 2019; Santoro et al., 2003). These transcription factors induce the transcription and secretion of the type I/III IFNs and other pro-inflammatory cytokines. Secreted IFNs in turn signal in an autocrine and paracrine fashion to activate kinases that phosphorylate and activate the transcription factors signal transducer and activator of transcription (STAT) 1 and 2. Finally, STAT1 and STAT2 induce expression of the IFN-stimulated genes (ISGs), which directly and indirectly fight the virus, setting up a potent antiviral state (Lazear et al., 2019).

By inducing host shutoff, PA-X modulates the expression of IFNs and other proteins involved in this pathway. Multiple studies have shown down-regulation of mRNAs for type I IFNs, pro-inflammatory cytokines, ISGs and other antiviral genes by PA-X from multiple human and avian IAV strains (Chaimayo et al., 2018; Galvin and Husain, 2019; Hayashi et al., 2015; Hu et al., 2015; Jagger et al., 2012; Narkpuk et al., 2018; Rigby et al., 2019). PA-X also restricts NF-κB p65 nuclear translocation and activity, leading to changes in NF-κB target gene expression, although the exact mechanism was not investigated (Hu et al., 2020). Interestingly, PA-X can also dampen the activation of type I IFN independently of its RNA degradation activity. PA-X achieves this immunomodulation by interacting with Ankrd17, a protein that activates RIG-I to induce IFN-β expression (Li et al., 2021; Wang et al., 2012). This overall dampening of the immune response at the molecular level likely explains the phenotypes observed *in vivo*. Few studies have investigated the direct link between the two, although many have observed increased secretion of pro-inflammatory cytokines and chemokines in the lungs of animal models infected with PA-X deficient IAV strains compared to WT (Dunagan et al., 2021; Gao et al., 2015b; Hayashi et al., 2015; Hu et al., 2016, 2015; Qin et al., 2022a, 2022b). One study showed that mice infected with a PA-X deficient 2009 pandemic H1N1 IAV strain have increased recruitment of neutrophils and lymphocytes

to the lungs compared to WT (Dunagan et al., 2021), while another found that PA-X from a low pathogenicity avian H9N2 IAV strain limits the recruitment of mucosal dendritic cells to the nasal submucosa of mice by reducing chemokine levels (Qin et al., 2022a). A third recent study similarly showed that PR8 PA-X limits the recruitment of dendritic cells but also reduces the ability of PR8 IAV to infect dendritic cells in the nasal mucosa of mice, limiting their migration and maturation (Qin et al., 2022b).

1.4. Other viral proteins that induce host shutoff through widespread RNA degradation

Aside from IAV, a few other virus families also induce host shutoff through RNA degradation. Some α-herpesviruses and γ-herpesviruses encode host shutoff RNases, while β-coronaviruses induce an unknown host RNase to trigger host shutoff, and some poxviruses and other large DNA viruses encode host shutoff decapping enzymes (Gaucherand and Gaglia, 2022). Although the viruses from these families are very different, they have all evolved to use RNA decay to shut down host gene expression. This convergent evolution is a testament to the efficacy of RNA degradation as a means to take over the host cell. In addition, there are common themes in the host shutoff mechanism of these divergent factors. First, they all specifically target RNAs transcribed by Pol II but not Pol I or III (Gaglia et al., 2012; Khaperskyy et al., 2016; Parrish et al., 2009, 2007; Parrish and Moss, 2007). Which subset of Pol II RNAs are targeted, as well as whether viral mRNAs are targeted, depend on the exact molecular mechanism of action of each of the viral RNA degradation factors. Second, as a result of endoribonucleolytic cleavage or decapping, an RNA with a 5’ monophosphate is generated, which is then fully degraded by the cellular exoribonuclease Xrn1 in the 5’-3’ direction (Covarrubias et al., 2011; Gaglia et al., 2012; Kamitani et al., 2009; Parrish et

al., 2009, 2007; Parrish and Moss, 2007). Understanding the mechanism of action of these other viral RNA degradation factors at the molecular level could thus help us draw hypotheses about the molecular mechanism of action of PA-X.

1.4.1. α-herpesviruses and vhs

Herpesviruses are large enveloped double-stranded DNA viruses that can infect an extremely wide range of animals, and include 8 different viruses that can infect humans. They can be classified into three subfamilies, α, β and γ. The α-herpesvirus subfamily includes the human viruses Herpes Simplex viruses (HSV) 1 and 2, which cause oral and genital herpes, and varicella zoster virus (VZV), which causes chickenpox and shingles. The HSV *UL41* gene encodes the virion host shutoff (vhs) protein, which is the most studied viral host shutoff RNase to date. vhs is an endoribonuclease of the FEN1 nuclease family (Doherty et al., 1996; Elgadi et al., 1999; Everly et al., 2002). Homologs of vhs in a few other α-herpesviruses can also induce host shutoff, although they have not been as extensively studied. This is the case for the swine pathogen pseudorabies virus (PrV) (Elgadi et al., 1999; Lin et al., 2010), the cattle pathogen bovine herpesvirus 1 (BHV) (Hinkley et al., 2000), the avian pathogen duck plague virus (DPV) (He et al., 2021), and the macaque pathogen monkey B virus (Black et al., 2014). While VZV also encodes a UL41 homolog, this protein does not induce host shutoff (Desloges et al., 2005). Of note, the other herpesvirus subfamilies β and γ do not encode UL41 homologs.

To reach its main target RNAs, HSV vhs binds to the eukaryotic translation initiation factors (eIFs) eIF4H, eIF4AII, and the cap-binding complex eIF4F (Feng et al., 2005; Page and Read, 2010) (**Figure 1.4**). Specifically, binding to eIF4H is necessary for vhs to degrade RNAs (Feng et al., 2001) and knock down of eIF4H abolishes vhs activity

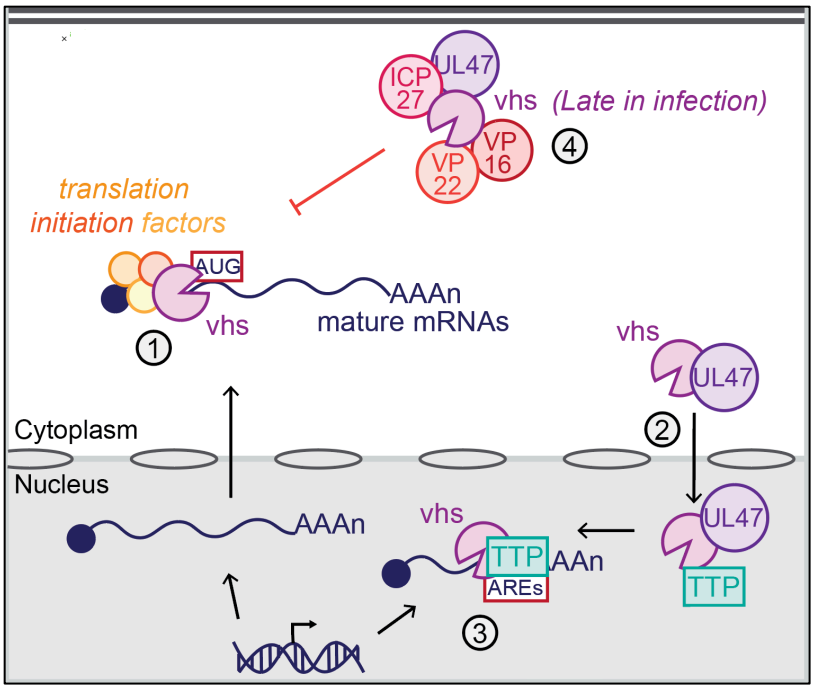

Figure 1.4: Current model of the molecular mechanism of action of the α-Herpesvirus vhs protein.

To access its target RNAs, vhs binds translation initiation factors in the cytoplasm, then cleaves capped RNAs or RNAs containing IRES at least at a position close to the translation start site (#1). The viral protein pUL47 also shuttles vhs to the nucleus (#2), where it binds the host protein TTP to cleave short-lived AREs containing mRNAs (#3). Later in infection, vhs activity is inhibited through interactions with the viral proteins VP16, VP22, pUL47 and ICP27 to allow translation of late proteins (#4). Figure and legend adapted with permission from Gaucherand L, Gaglia MM. 2022. The Role of Viral RNA Degrading Factors in Shutoff of Host Gene Expression. Annual Review of Virology 9:213–238. Changes include cropping out elements of the figure that do not focus on vhs.

(Sarma et al., 2008). Because these translation initiation factors bind capped mRNAs, or mRNAs containing internal ribosome entry sites (IRESs), vhs binds to the 5' end of host mRNAs right before they are translated. As a result, vhs specifically targets mature mRNAs in the cytoplasm, including viral mRNAs (Kwong and Frenkel, 1987; Oroskar and Read, 1989, 1987). In contrast, pre-mRNAs in the nucleus or uncapped noncoding RNAs transcribed by Pol I and III are not targeted (Gaglia et al., 2012). This specificity to fully processed mRNAs was also shown for its homolog pUL41 in DPV, as a transfected GFP reporter was only targeted by pUL41 when expressed from a Pol II promoter but

not a Pol I or III promoter (He et al., 2021). On top of targeting mature mRNAs in the cytoplasm, HSV vhs also targets specific mRNAs in the nucleus, presumably before they are exported to the cytoplasm (Shu et al., 2015, 2013). These mRNAs contain AU-rich elements (AREs) in their 3' untranslated region (UTR). As such, they are short-lived mRNAs that are induced by sensors of innate immunity. To target these mRNAs, vhs shuttles to the nucleus by interacting with the viral protein pUL47 (also known as VP13/14) (Shu et al., 2013). Once in the nucleus, vhs binds the host protein tristetraprolin (TTP), which recruits vhs to ARE-containing mRNAs (Shu et al., 2015) (**Figure 1.4**).

Because of its recruitment at the 5' end of transcripts by translation factors, HSV vhs is thought to cleave RNAs within their 5' UTR. This has been confirmed using cell lysates containing *in vitro*-translated vhs, where vhs cleaves RNA preferentially near the translation initiation site (Shiflett and Read, 2013). Moreover, the same study showed that these RNAs were preferentially cut if a 5' cap or an IRES were present, and mutations that altered the proximal AUG site abolished cleavage at nearby sites (Shiflett and Read, 2013). Similar results were obtained in an *in vitro* study of PrV vhs where RNAs were cut right downstream of an IRES, albeit a slightly different RNA degradation profile than HSV vhs (Liu et al., 2016). Unlike HSV vhs, PrV vhs was additionally able to target ribosomal RNA (rRNA) *in vitro* (Liu et al., 2015). These results suggest that vhs cleavage is tightly coupled to translation, potentially through the recognition of a specific RNA structure at the translation initiation site. However, northern blot analysis of cells co-transfected with HSV-1 vhs and green fluorescent protein (GFP) and DsRed showed a different degradation pattern, with multiple GFP and DsRed mRNA degradation fragments observed instead of just one cleavage site near the cap or IRES (Gaglia et al., 2012). vhs may thus also cut mRNAs at internal locations in cells compared to *in vitro*. Moreover, the RNA fragments were only observed in Xrn1 knock down cells, suggesting

that they are normally degraded by Xrn1 in cells after cleavage by vhs (Gaglia et al., 2012). In the case of stress response mRNAs, vhs cleaves these RNAs in the 3' UTR near their AREs (Esclatine et al., 2004b).

vhs is packaged in the α-herpesvirus virion tegument, i.e. the collection of proteins found between the viral capsid and the viral envelope. As a result, vhs is delivered into the cell and is active early in infection (Fenwick and McMenamin, 1984; Kwong and Frenkel, 1987). However, as infection progresses, HSV needs to regulate the activity of vhs to prevent any detrimental effect on viral replication. To do so, the viral proteins virion protein (VP) 16 (UL48), VP22 (UL49), UL47, and infected cell protein ICP27 bind to vhs and block its activity (Lam et al., 1996; Shu et al., 2013; Taddeo et al., 2010, 2007). This late inhibition of vhs activity allows HSV to switch from early to late viral gene expression, as vhs thus primarily degrades early viral but not late mRNAs, allowing to free the translation machinery for the translation of late viral mRNA (Dauber et al., 2014; Kwong and Frenkel, 1987; Oroskar and Read, 1989).

1.4.2. KSHV/MHV68/EBV and SOX/muSOX/BGLF5

γ-herpesviruses are well known to induce cancers in immunocompromised individuals. This subfamily includes the two human viruses Kaposi's sarcoma-associated herpesvirus (KSHV) and Epstein-Barr virus (EBV), as well as several animal viruses. The γ-herpesvirus murine herpesvirus 68 (MHV68) is commonly used to model infection in small animal models. As mentioned in the previous section, γ-herpesviruses do not encode a UL41/vhs homolog. Instead, the γ-herpesvirus homolog of the herpesviral alkaline exonuclease is responsible for widespread RNA degradation and host shutoff (Covarrubias et al., 2009; Glaunsinger and Ganem, 2004b; Rowe et al., 2007). This protein is encoded by *ORF37* and termed shutoff and exonuclease (SOX) in KSHV, muSOX in MHV68 or BGLF5 in EBV. Like PA-X, it belongs to the PD-D/E-X-K

superfamily of nucleases (Buisson et al., 2009; Dahlroth et al., 2009). The name alkaline exonuclease comes from a second function of the protein as a deoxyribonuclease (DNase) during genome replication. This protein is shared by all herpesviruses, encoded by the *UL12* gene in HSV and *UL98* in the b-herpesvirus human cytomegalovirus. Only in γ-herpesviruses has this protein evolved the additional function of endoribonuclease and host shutoff factor (Glaunsinger and Ganem, 2004b). This additional function could be due to the nuclear and cytoplasmic localization of SOX/muSOX/BGLF5, whereas their α- and β-herpesvirus homologs are only nuclear (Covarrubias et al., 2009). The mechanism of action of the γ-herpesvirus RNases SOX, muSOX and BGLF5 is still not well characterized at the molecular level (**Figure 1.5**). Like vhs, these RNases preferentially target mature mRNAs transcribed by Pol II that are capped, tailed and ready to be translated in the cytoplasm, but do not target Pol I and III transcripts (Covarrubias et al., 2011; Gaglia et al., 2012). Consistent with this finding, SOX co-sediments with the 40S ribosomal subunit in ribosomal profiling experiments (Covarrubias et al., 2011). Yet, active translation per se in not necessary for targeting, as the RNases still down-regulate translation-competent mRNAs bearing a 5' hairpin that prevents ribosome scanning after cap binding (Gaglia et al., 2012). Other than these findings, how the SOX/muSOX/BGLF5 proteins are able to target Pol II transcripts and whether they need to interact with a specific cellular factor to do so is still unknown. The RNases are also thought to target viral mRNAs for degradation, as in the absence of host shutoff, some viral proteins accumulate to higher levels, which alters the protein composition of the virion (Abernathy et al., 2014; Feederle et al., 2009).

While it is still unclear for vhs, SOX clearly cleaves mRNAs at a specific sequence and structure (Gaglia et al., 2015; Mendez et al., 2018) (**Figure 1.5**). This specificity has been observed both transcriptome-wide in cells overexpressing SOX and using *in vitro* assays (Gaglia et al., 2015; Mendez et al., 2018). However, the recognition and

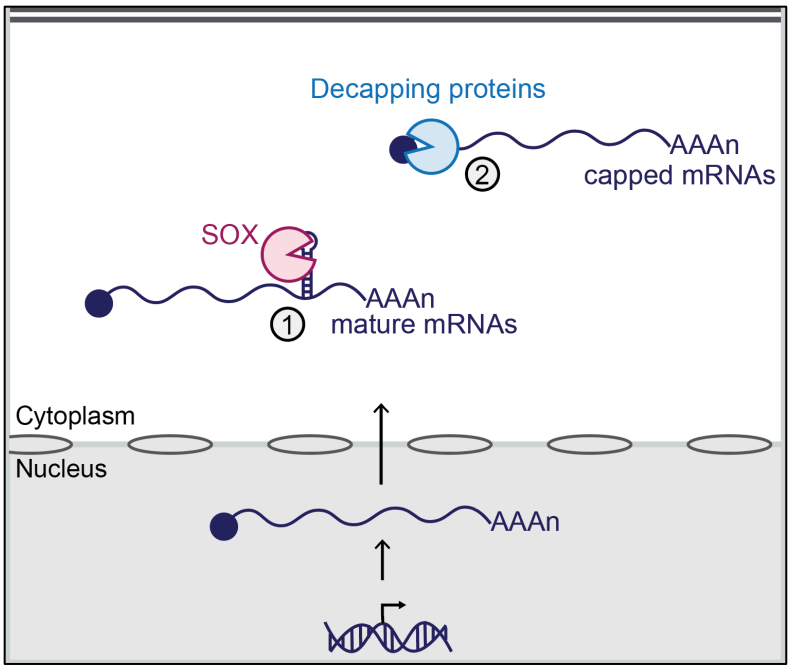

Figure 1.5: Current model of the molecular mechanism of action of the γ-Herpesvirus SOX protein and the large DNA viruses decapping proteins.

SOX recognizes and cleaves a specific sequence and structure within cytoplasmic mRNAs associated with translation complexes (#1). The viral decapping proteins ASFV-DP and L357 bind RNA to locate its 5' cap, while D9 and D10 bind to both the RNA and the 5' cap to locate the cap and cleave it (#2). Figure and legend adapted with permission from Gaucherand L, Gaglia MM. 2022. The Role of Viral RNA Degrading Factors in Shutoff of Host Gene Expression. Annual Review of Virology 9:213–238. Changes include cropping out elements of the figure that do not focus on SOX or decapping enzymes, and changing the numbering accordingly.

cleavage motif itself is complex and degenerate. Recognition of the target sequence requires a hairpin structure with a stretch of three adenosine in the loop (Gaglia et al., 2015; Mendez et al., 2018). SOX binds directly to the adenosine stretch and cleaves a few nt downstream of this location (Gaglia et al., 2015; Mendez et al., 2018). Northern blot analysis of cleaved transcripts in cells overexpressing EBV BGLF5 suggests this RNase may also cut RNAs in a sequence-specific manner, although this has not been thoroughly investigated (Gaglia et al., 2012). Finally, just like other viral RNA degradation factors, the host RNase Xrn1 degrades fragments generated by SOX/muSOX/BGLF5 (Gaglia et al., 2012).

1.4.3. Large DNA viruses and viral decapping enzymes

Aside from encoding viral RNases, another way to induce RNA degradation-mediated host shutoff is to use decapping enzymes. This is the case of a few large DNA viruses, the best characterized being vaccinia virus (VacV). VacV is a linear double stranded DNA virus of the poxviridae family that replicates in the cytoplasm. VacV is famous for being used as a vaccine to eradicate variola virus, the poxvirus responsible for the deadly disease smallpox (Fenner, 1980). Since then, VacV-based vectors have been used to design a wide range of vaccines against other viruses, such as human immunodeficiency virus, Rabies virus and Zika virus (Kieny et al., 1984; Pérez et al., 2018; Rerks-Ngarm et al., 2009; Zhang et al., 2021). VacV is also still extensively used as a model to study poxviruses. VacV actually encodes two decapping enzymes, called D9 and D10 (Parrish et al., 2007; Parrish and Moss, 2007). These enzymes remove the 5' cap from mRNAs, which triggers broad degradation of host mRNAs by Xrn1 in the cytoplasm, and with it host shutoff (Burgess and Mohr, 2015; Dai et al., 2017). It is still unclear why VacV encodes two proteins with similar functions, although several findings suggest they each play distinct roles for viral replication. First, these enzymes are expressed at different times during viral replication: whereas D9 is expressed early in infection, D10 is only expressed after DNA replication has occurred (Lee-Chen and Niles, 1988; Parrish and Moss, 2006). Second, mutant viruses that do not express D10 are more defective than viruses lacking D9 (Parrish and Moss, 2006). Third, D9 and D10 have different biochemical characteristics, as they bind RNAs and 5' caps with different affinities *in vitro* (Parrish and Moss, 2007). Consistent with these differences, D9 and D10 do not target exactly the same host mRNAs in cells, although D10 down-regulates most host transcripts (Ly et al., 2022).

Although less well-characterized, two other large DNA viruses, African swine fever virus (ASFV) and mimivirus, also encode decapping enzymes. ASFV is part of the

Asfaviridiae family and can infect different species of swine, including domestic pigs and wild boar (Costard et al., 2013). This virus encodes the decapping enzyme ASFV-DP, denoted as the g5R protein for the Malawi strain and D250R for the Ba71V strain (Cartwright et al., 2002; Parrish et al., 2009). Mimivirus is a giant virus from the *Mimiviridae* family that infects *Acanthamoeba* species (Raoult et al., 2004). It encodes the decapping enzyme L375 (Kago and Parrish, 2021). Interestingly, despite mimivirus and ASFV being very different viruses, their decapping enzymes share 21% amino acid identity (Kago and Parrish, 2021). While it has not been investigated for L375, ASFV-DP is expressed early and accumulates in the endoplasmic reticulum throughout infection (Quintas et al., 2017).

Similar to cellular decapping enzymes and the yeast decapping enzyme Dcp2, large DNA virus-encoded decapping enzymes contain a Nudix hydrolase domain (Cartwright et al., 2002; Kago and Parrish, 2021; Shors et al., 1999), with the highly conserved Nudix sequence signature motif (previously called mutT motif) GX5EX5[UA]XREX2EEXGU, where U is a hydrophobic amino acid and X can be any amino acid (Bessman et al., 1996; Koonin, 1993). This domain gives them intrinsic decapping activity to convert m^7GpppNm-capped RNAs into an uncapped 5' mono-phopsphorylated RNA product with the release of m^7GDP (Kago and Parrish, 2021; Parrish et al., 2009, 2007; Parrish and Moss, 2007).

To locate and cleave the 5' cap of their target RNAs, large DNA virus-encoded decapping enzymes first need to bind the RNA backbone (**Figure 1.5**). The requirement for this interaction was inferred from the finding that the addition of uncapped RNAs inhibits decapping activity *in vitro* (Kago and Parrish, 2021; Parrish et al., 2009, 2007; Parrish and Moss, 2007). The viral decapping proteins are also unable to cleave free methylated cap analogues *in vitro* that are not attached to an RNA moiety (Kago and Parrish, 2021; Parrish et al., 2009, 2007; Parrish and Moss, 2007). In addition, the

decapping activity of D9 and D10 is inhibited *in vitro* by the addition of free methylated cap derivatives, suggesting that these proteins also interact with the 5' cap structure (Parrish et al., 2007; Parrish and Moss, 2007). This was not the case for ASFV-DP, suggesting this protein does not bind the 5' cap (Parrish et al., 2009). Overall, it is still unclear whether these viral decapping enzymes also need to interact with host proteins to reach their target RNAs in cells. It is also unclear whether and how the decapping activity is limited to mRNAs, as even noncoding RNAs transcribed by Pol II are capped. While ASFV-DP was found to interact with the ribosomal protein RPL23a, the specific role of this interaction has not been investigated (Quintas et al., 2017). Similarly, a transcriptomic study found that spliced RNAs are more down-regulated by D10 than intronless RNAs (Ly et al., 2022). This finding suggests that D10 may interact with host proteins involved in mRNA processing, but this has not been directly examined. Finally, these viral decapping enzymes do not seem to discriminate between host and viral mRNAs. Indeed, the N-terminal domain of ASFV-DP interacts with both host and viral mRNAs in cells and during infection, and overexpression of ASFV-DP down-regulates both host and viral mRNA levels, suggesting this enzyme is also active on viral mRNAs (Quintas et al., 2017). Viral mRNAs also appear slightly targeted by D10 (Liu et al., 2014), although RNA sequencing data point to a weaker effect on viral than host mRNAs, in particular for late viral transcripts (Ly et al., 2022).

1.4.4. Coronaviruses and nsp1

Unlike the previously mentioned viruses, coronaviruses do not encode a host shutoff RNase or decapping enzyme. Instead, they are thought to induce widespread RNA degradation and host shutoff by activating a yet to be identified host RNase. Coronaviruses are large enveloped single-stranded positive-sense RNA viruses. They can cause respiratory or enteric infections in a wide variety of avian and mammalian

hosts. There are four genera of coronaviruses: α, β, γ and δ. Human coronaviruses include the common cold α-coronaviruses 229E and NL63 and β-coronaviruses HKU1 and OC43, as well as three β-coronaviruses that can cause more severe respiratory infections and pandemics, Middle East Respiratory Syndrome Coronavirus (MERS-CoV), and the two Severe Acute Respiratory Syndrome Coronaviruses (SARS-CoV and SARS-CoV-2). α- and β-coronaviruses, which encompass all coronaviruses known to infect humans, encode the host shutoff protein non-structural protein 1 (nsp1). While α- and β-coronavirus nsp1 lack sequence similarities, they retain similar biological functions, highlighting the importance of nsp1 for infection of mammalian hosts (Connor and Roper, 2007). Indeed, nsp1 proteins from human coronaviruses 229E, NL63, MERS-CoV, SARS-CoV and SARS-CoV-2, and from the swine transmissible gastroenteritis virus (TGEV) inhibit host translation, albeit with some mechanistic differences (Narayanan et al., 2015; Yuan et al., 2021). Importantly, in addition to inhibiting translation, all but TGEV nsp1 also induce degradation of host mRNAs (Burke et al., 2021; Huang et al., 2011a, 2011b; Kamitani et al., 2009, 2006; Lokugamage et al., 2015; Narayanan et al., 2008; Terada et al., 2017; Wang et al., 2010). However, nsp1 shares no sequence or structure similarities with known RNases, and requires the presence of translation extracts to cleave RNA *in vitro*, suggesting that nsp1 does not have intrinsic RNase activity (Almeida et al., 2007; Mendez et al., 2021). It is thus thought that nsp1 activates a host RNase instead to trigger mRNA degradation.

SARS-CoV and SARS-CoV-2 nsp1 bind to the ribosome to access their target mRNAs and induce their degradation (Huang et al., 2011a; Kamitani et al., 2009; Mendez et al., 2021; Narayanan et al., 2008) (**Figure 1.6**). As a result, nsp1 specifically targets capped translating mRNAs that are transcribed by RNAPII (Gaglia et al., 2012; Lokugamage et al., 2015). nsp1 can also target translating mRNAs containing

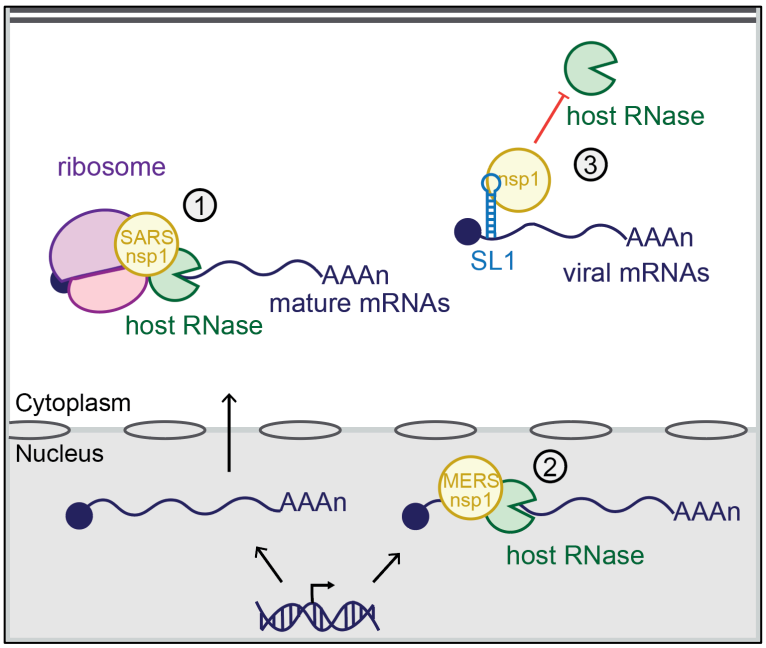

Figure 1.6: Current model of the molecular mechanism of action of the coronavirus nsp1 protein.

SARS-CoV-1 and 2 nsp1 bind to the 40S ribosome subunit to target capped or IRES-containing mRNAs that are actively translated (#1), inducing RNA degradation by an unknown host RNase in the cytoplasm. MERS-CoV nsp1 also induces RNA degradation, but within the nucleus and without binding to the ribosome (#2). SARS and MERS-CoV nsp1 bind to the SL1 structure within the 5' untranslated region of viral mRNAs to protect them from degradation (#3). Figure and legend adapted with permission from Gaucherand L, Gaglia MM. 2022. The Role of Viral RNA Degrading Factors in Shutoff of Host Gene Expression. Annual Review of Virology 9:213–238. Changes include cropping out elements of the figure that do not focus on nsp1, and changing the numbering accordingly.

picornavirus type I and II IRES elements (Huang et al., 2011a; Kamitani et al., 2009; Lokugamage et al., 2015). Because nsp1 needs to bind the ribosome to induce RNA degradation, one model is that nsp1 activates a host RNase part of an RNA quality control pathway linked to translation to trigger mRNA degradation. While the identity of this RNase is still unknown, it is not RNase L (Burke et al., 2021). As suggested in a recent review, the human homolog of the newly identified yeast endonuclease Cue2 may be a good candidate, as it cleaves mRNAs after translation stalls (Rodriguez et al., 2021). *In vitro* assays using cellular extracts show that SARS-CoV nsp1 induces

endonucleolytic RNA cleavage within 30 nt of the ribosome loading region, whether this occurs via 5' cap or IRES element (Huang et al., 2011a). Similar results were obtained in cells overexpressing SARS-CoV nsp1, followed by degradation of the RNA fragment by Xrn1 (Gaglia et al., 2012). The *in vitro* study did not identify any clear sequence preference for cleavage, although this has not been investigated in cells (Huang et al., 2011a). On top of inducing mRNA cleavage, binding of nsp1 to the ribosome also inhibits host mRNA translation, by altering ribosome function and complex formation (Setaro and Gaglia, 2021). Different mechanism of translation inhibition have been described, including the insertion of the C-terminal domain of SARS-CoV-2 nsp1 inside the ribosomal mRNA channel to interfere with host mRNA binding (Schubert et al., 2020; Yuan et al., 2020). Interestingly, the two functions of nsp1 are separable, as a SARS-CoV nsp1 R124A/K125A mutant was isolated that still blocks translation but does not induce RNA decay (Lokugamage et al., 2012). Consistent with this idea, SARS-CoV nsp1 inhibits translation but does not induce degradation of mRNAs containing type III and IV IRES elements (Kamitani et al., 2009).

Importantly, viral mRNAs are able to evade SARS-CoV and SARS-CoV-2 nsp1-induced RNA cleavage due to the presence of a stem loop (termed SL1) structure inside their 5' UTR (Banerjee et al., 2020; Huang et al., 2011a; Mendez et al., 2021; Rao et al., 2021; Schubert et al., 2020; Tidu et al., 2021). This SL1 protective structure is part of the leader sequence, which all genomic and subgenomic viral mRNAs have (Kim et al., 2020; Sola et al., 2015). Interestingly, nsp1 itself may confer the protection by binding to SL1 (Tanaka et al., 2012; Vankadari et al., 2020) (**Figure 1.6**). The mechanism of protection is still unclear, as different mechanisms have been proposed. One study suggested that binding of nsp1 to viral mRNA triggers remodeling of the ribosome complex to enable translation (Mendez et al., 2021). Another study has proposed instead that the viral leader protective sequence can remodel the structure of nsp1,

triggering its dissociation from the 40S ribosome subunit to allow for viral mRNA translation (Banerjee et al., 2020). However, a third study found that SARS-CoV-2 nsp1 remains bound to the ribosome during viral mRNA translation (Tidu et al., 2021).

Like the SARS-CoVs nsp1, MERS-CoV nsp1 also inhibits host translation and induces degradation of host mRNA transcribed by Pol II, and these functions can be separated by the analogous R146A/K147A mutations (Lokugamage et al., 2015). Moreover, MERS-CoV nsp1 also binds to SL1 to protect and promote translation of its viral mRNAs (Terada et al., 2017). However, it does not bind the 40S ribosome stably (Lokugamage et al., 2015). In addition, SARS-CoV nsp1 is exclusively found in the cytoplasm, whereas MERS-CoV nsp1 is located in both nucleus and cytoplasm (Lokugamage et al., 2015). As a result, MERS-CoV nsp1 induces selective degradation of mRNAs that are transcribed in the nucleus and exported to the cytoplasm, but not of cytoplasmic-restricted mRNAs that are introduced exogenously in the cytoplasm or virus-like mRNAs synthesized in the cytoplasm using the Rift Valley fever virus cytoplasmic transcription machinery (Lokugamage et al., 2015). Overall, these observations suggest MERS-CoV nsp1 uses a different mechanism to reach its target RNAs, which has not been identified yet (**Figure 1.6**).

1.4.5. Implications for the molecular mechanism of action of PA-X

As previously mentioned, the mechanism of action of PA-X is still largely unknown at the molecular level. However, the diverse mechanisms of action of other viral RNA degradation factors can give us a solid base to infer some hypotheses as to the general mechanism of PA-X activity (**Figure 1.7**). The first step likely involves hijacking a cellular process to locate Pol II-transcribed RNAs. For example, HSV vhs binds translation initiation factors to reach Pol II transcribed mRNAs ready to be translated, while VacV D9 and D10 bind an RNA and its 5’ cap to ensure targeting of Pol II transcribed RNAs

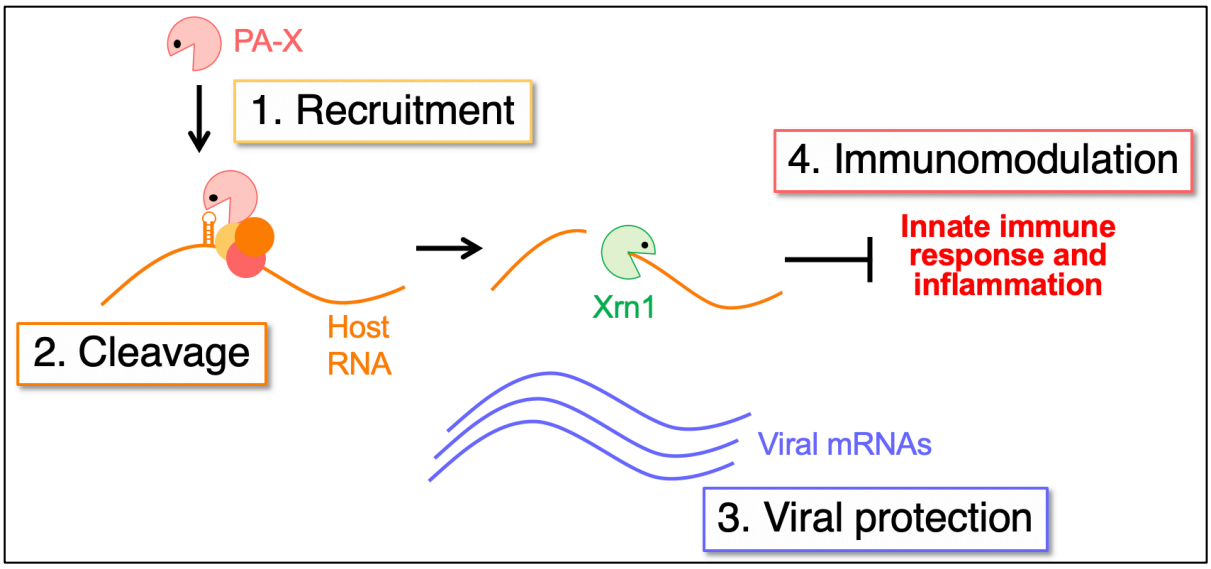

Figure 1.7: Working model of the general steps of the mechanism of action of PA-X to investigate.

Based on the molecular mechanism of action of other viral RNA degradation host shutoff factors, I hypothesize that PA-X specificity and activity involves 1. its recruitment to Pol II transcribed RNAs through interactions with specific RNA-binding proteins, 2. the recognition of a specific sequence and structure to cleave the RNA, 3. a specific intrinsic mechanism to protect viral mRNAs from degradation, and 4. the overall down-regulation of the host innate immune response and inflammation.

(Feng et al., 2005; Page and Read, 2010; Parrish et al., 2007; Parrish and Moss, 2007). Second, KSHV SOX, which is part of the PD-D/E-X-K nuclease superfamily (Buisson et al., 2009; Dahlroth et al., 2009), cuts mRNAs at a specific sequence and structure (Gaglia et al., 2015). Since PA-X also holds a PD-D/E-X-K nuclease family fold (Dias et al., 2009; Yuan et al., 2009), its cleavage may be sequence and structure specific as well. Third, unlike the herpesvirus RNases, PA-X does not seem to impact viral replication (Khaperskyy et al., 2016). Since coronavirus nsp1 protects its own viral mRNAs from degradation (Tanaka et al., 2012; Terada et al., 2017; Vankadari et al., 2020), PA-X may also have an intrinsic mechanism to avoid the targeting of viral mRNAs. Finally, infection with a PA-X deficient virus often leads to striking pro-

inflammatory phenotypes *in vivo* (Gao et al., 2015b; Gong et al., 2017; Hayashi et al., 2015; Hu et al., 2016, 2015; Jagger et al., 2012). The mechanism of PA-X target selection may thus allow for specific down-regulation of the innate immune and inflammatory response.

To uncover the exact mechanisms involved in these different steps, I chose an RNA centric approach. I hypothesized that identifying which RNAs are down-regulated and/or cleaved by PA-X, and which are not, would reveal the mechanism of selection of these RNAs. This thesis describes work focused on uncovering the mechanism of action of PA-X at the molecular level, and the possible implications of RNA targeting specificity on viral replication and the modulation of the innate immune response and inflammation. Chapter 2 is a first manuscript presenting a mechanism by which PA-X is recruited to Pol II transcripts (Gaucherand et al., 2019). Chapter 3 is a second manuscript highlighting the sequence and structure preference of PA-X cleavage, and how this specificity spares viral mRNAs. I will then discuss these findings and their implications in Chapter 4.

Chapter 2. The Influenza A Virus Endoribonuclease PA-X Usurps Host mRNA Processing Machinery to Limit Host Gene Expression

Gaucherand L, Porter BK, Levene RE, Price EL, Schmaling SK, Rycroft CH, Kevorkian Y, McCormick C, Khaperskyy DA, Gaglia MM. 2019. *Cell Reports* 27:776-792.e7.
Reprinted here with permission of publisher.

2.1. Introduction

Despite their small genomes, influenza A viruses (IAVs) dedicate multiple proteins to the suppression of host gene expression, or “host shutoff”, which limits host antiviral responses. One of the IAV host shutoff proteins is the endoribonuclease PA-X, which selectively degrades host RNAs (Jagger et al., 2012; Khaperskyy et al., 2016) and limits innate immune responses *in vivo.* PA-X-deficient viruses induce stronger innate immune and inflammatory responses in mice, chickens and pigs (Gao et al., 2015b; Gong et al., 2017; Hayashi et al., 2015; Hu et al., 2016, 2015; Jagger et al., 2012; Xu et al., 2017). In some IAV strains, the immune-evasion activity of PA-X reduces inflammation-induced pathology, thereby protecting the host and reducing mortality (Gao et al., 2015b; Gong et al., 2017; Hu et al., 2016, 2015; Jagger et al., 2012). While the role of PA-X in immune evasion is well established, its molecular mechanism of action remains poorly understood.

PA-X is produced by ribosomal frameshifting during translation of the PA mRNA (Firth et al., 2012; Jagger et al., 2012). The frameshift generates a protein with the PA amino-terminal ribonuclease (RNase) domain fused to a unique carboxy-terminal domain known as the X-ORF. The X-ORF is required for PA-X function (Hayashi et al., 2016; Khaperskyy et al., 2016; Oishi et al., 2015). Despite this non-canonical production mechanism, PA-X is encoded by all IAV strains (Shi et al., 2012). We previously reported that PA-X selectively degrades RNAs transcribed by host RNA polymerase II (Pol II), but not other polymerases (Khaperskyy et al., 2016). Importantly, this characteristic leads to the protection of viral RNAs created by the viral RNA-dependent RNA polymerase (RdRp) (Khaperskyy et al., 2016). However, the mechanism for PA-X targeting of Pol II transcripts is not known.

Other viruses encode host shutoff RNases that selectivity target Pol II transcripts, including alphaherpesviral vhs proteins, gammaherpesviral SOX/BGLF5 proteins and

the severe acute respiratory syndrome-related coronavirus (SARS-CoV) non-structural protein 1 (nsp1) (Covarrubias et al., 2011, 2009; Elgadi et al., 1999; Gaglia et al., 2012; Glaunsinger and Ganem, 2004b; Kamitani et al., 2006; Rowe et al., 2007). However, SARS nsp1 and the herpesviral host shutoff proteins operate in the cytoplasm and only degrade transcripts that are bound by components of the protein synthesis machinery (Covarrubias et al., 2011; Doepker et al., 2004; Feng et al., 2005, 2001; Gaglia et al., 2012; Kamitani et al., 2009). By contrast, PA-X accumulates in the nucleus, and the protein synthesis machinery has no role in RNA targeting and degradation (Hayashi et al., 2016; Khaperskyy et al., 2016). Our previous analysis of select transcripts suggests that not all Pol II transcripts are equally susceptible to PA-X degradation (Khaperskyy et al., 2016), similar to reports for other viral host shutoff RNases (Esclatine et al., 2004a; Glaunsinger and Ganem, 2004a). In agreement with this, a recent study of the host transcriptome in IAV-infected cells showed that certain functional classes of RNAs were spared from shutoff, although no specific link to PA-X activity was established (Bercovich-Kinori et al., 2016). By contrast, in the context of studying the relative contribution of IAV PA-X and NS1 proteins to host shutoff, Toru Takimoto's group recently reported that host mRNAs targeted by PA-X do not clearly belong to specific functional classes, whereas there is functional specificity among NS1 targets (Chaimayo et al., 2018). Collectively, these findings suggest that PA-X may have a unique mechanism to selectively target host RNAs in the nucleus, perhaps in conjunction with RNA processing and the assembly of functional messenger ribonucleoprotein (mRNP) complexes.

Here, we report transcriptome-wide analysis of PA-X targets in human lung A549 cells, in both *de novo* infection and ectopic expression models. This analysis revealed that PA-X susceptibility was tightly linked to Pol II transcript splicing. Moreover, we identified host proteins involved in mRNA processing that associated with the C-terminal

X-ORF, suggesting that PA-X target selection may involve physical interactions with components of the host mRNA processing machinery.

2.2. Results

2.2.1. PA-X causes global changes in RNA levels during infection

To determine the scope of PA-X specificity for host Pol II transcripts, we profiled RNA levels in cells infected with wild-type (wt) and PA-X-deficient IAV. To generate PA-X deficient mutants in the well-characterized strain A/PuertoRico/8/1934 H1N1 (PR8), we introduced two mutations in the frameshifting site and a nonsense mutation in PA-X, L201Stop, that truncated the X-ORF after 9 amino acids (aa); we dubbed this virus PA(ΔX) (**Figure 2.1A**). These mutations were designed to be silent in the PA ORF. We previously used a strain with only the frameshifting mutations, IAV PA(fs) (**Figure 2.2A**) (Khaperskyy et al., 2016), but created IAV PA(ΔX) to ensure that any residual frameshifting would produce a non-functional PA-X. We confirmed that the 9-aa truncated PR8 PA-X was largely inactive, as it lost the ability to degrade a β-globin reporter, whereas an X-ORF truncation to 15-aa retained activity (**Figure 2.1B**). The β-globin reporter includes the two introns of the native β-globin gene, and expresses an mRNA that is spliced. The results in **Figure 2.1B** recapitulate previous findings using truncations in PA-X variants from other strains (Hayashi et al., 2016; Oishi et al., 2015). We also generated a virus with only the L201Stop mutation and dubbed it ‘X9’ (**Figure 2.2A**). We chose to use the PR8 strain because it lacks two other known IAV host shutoff mechanisms: its NS1 protein does not block host mRNA processing (Das et al., 2008; Salvatore et al., 2002), and its RdRp does not trigger Pol II degradation (Rodriguez et al., 2009).

Using high-throughput RNA sequencing, we found that infection of cells with wt IAV caused a dramatic global decrease in transcript levels compared to mock-infected cells

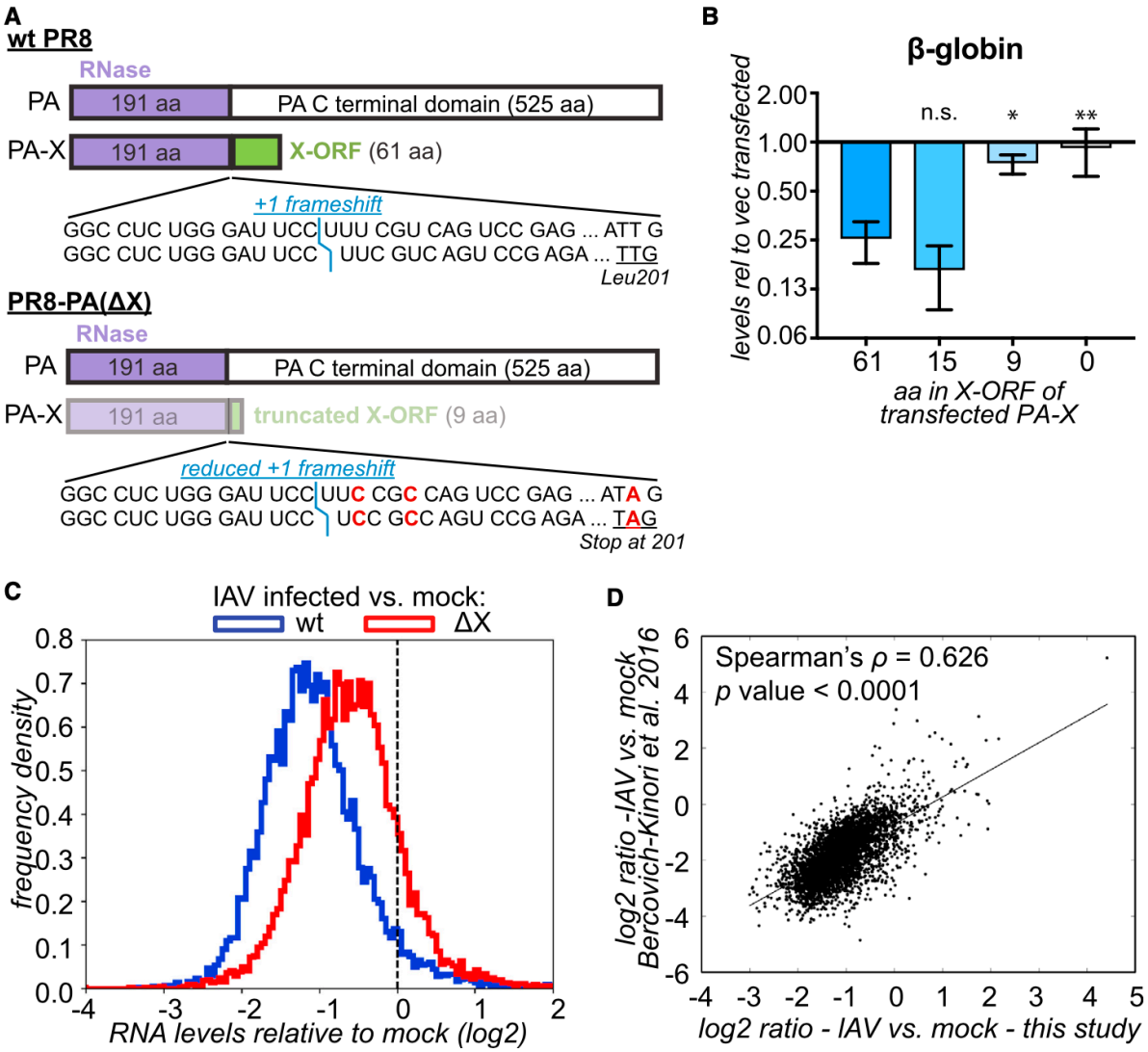

Figure 2.1: PA-X down-regulates most cellular RNAs and is a major contributor to host shutoff during influenza A virus infection.

*(A) Diagram of mutations in the PR8 PA(ΔX) virus. Less intense colors = lower levels of PA-X. Blue = position of the frameshift. Red = mutated nucleotides in the frameshifting sequence and at PA-X codon 201. (B) HEK293T cells were transfected for 24 h with a β-globin reporter and wt PR8 PA-X ("61") or variants with the C-terminal X-ORF truncated after the indicated number of amino acids (aa). Levels of β-globin in PA-X transfected cells were measured by RT-qPCR and are plotted relative to vector transfected cells, after normalization to cellular 18S rRNA. Values represent mean ± standard deviation. N = 3. ns, *, ** = p > 0.05, < 0.05, < 0.01, respectively, ANOVA followed by Dunnett's multiple comparison test* vs. *wt PA-X (61-aa). (C) RNA-seq was carried out on RNA collected 15 h after infection from A549 cells infected with wt PR8 or PR8 PA(ΔX). The ratio between levels in IAV-infected* vs. *mock-infected cells was computed for each RNA and the distribution of the ratios (log2) is plotted as a frequency histogram. The two populations are significantly different (p < 0.001) based on the Kolmogorov-Smirnoff test. The dashed line indicated a ratio of 1 (no change). N ≥ 2. (D) The ratio in RNA levels in wt IAV infected cells* vs. *mock-infected cells in our study (15 h post infection, MOI = 1) is*

plotted against the results from Bercovich-Kinori et al. *2016 (Bercovich-Kinori et al., 2016) (8 h post infection, MOI = 5).*

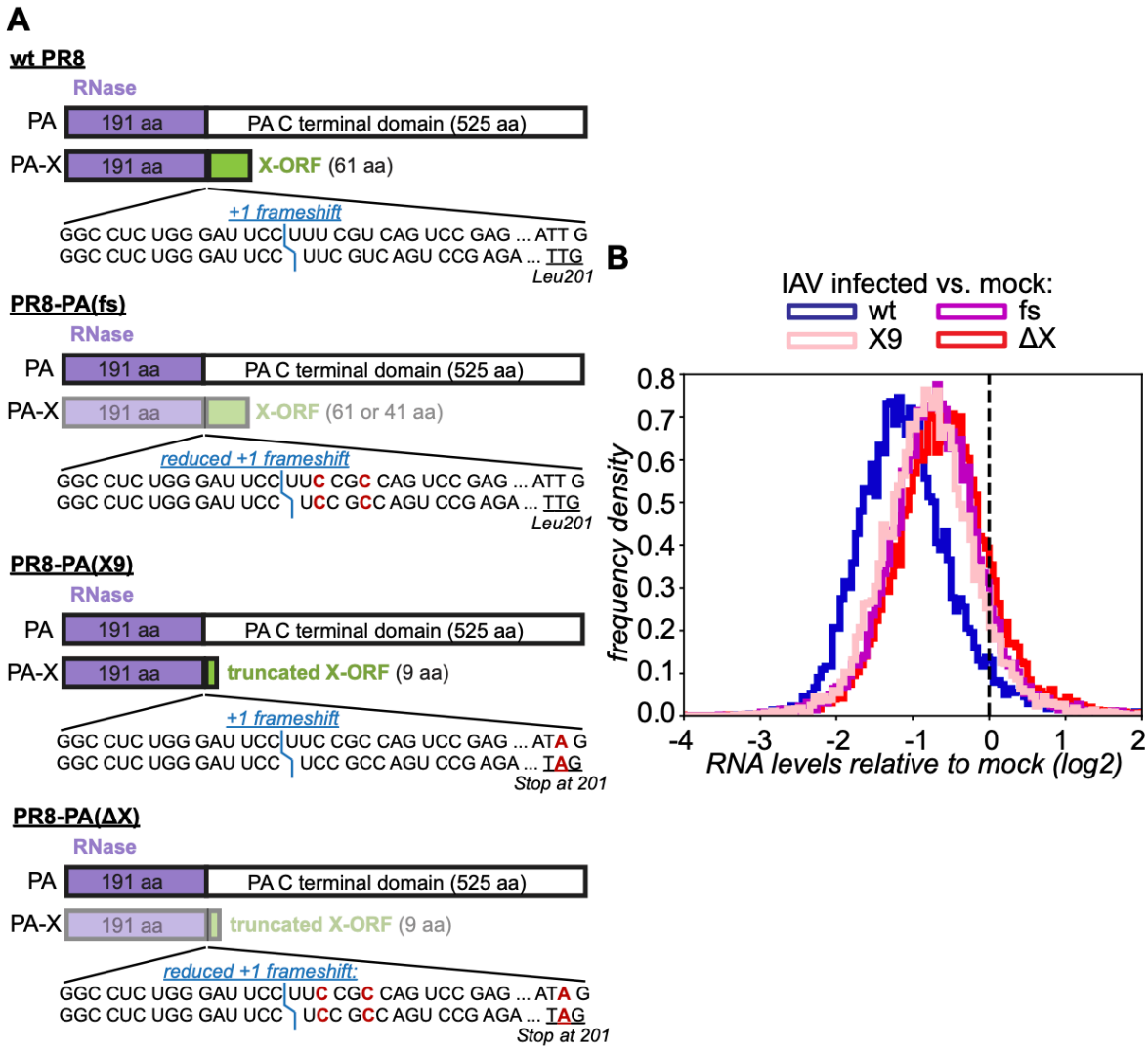

Figure 2.2: Reduction in frameshifting and truncation of the X-ORF have similar effects on host shutoff during infection.
(A) Diagrams of mutations in the PR8 PA(fs), PA(X9), PA(ΔX) viruses. The less intense colors represent lower levels of the PA-X protein. The position of the frameshift is marked in blue. The mutated nucleotides in the frameshifting sequence and at PA-X codon 201 are marked in red. (B) RNA was collected from infected A549 cells 15 hrs after infection with wt PR8 or the PR8 PA-X mutant viruses (PA(fs), PA(X9), PA(ΔX)), and the levels of all RNAs were measured by RNA-seq. The ratio between the levels in IAV-infected vs. mock-infected cells was computed for each RNA and the distribution of the ratios is plotted as a frequency histogram (The dataset plotted here for infections with wt and PA(ΔX) viruses are the same ones used in Figure 1C). n ≥ 2.

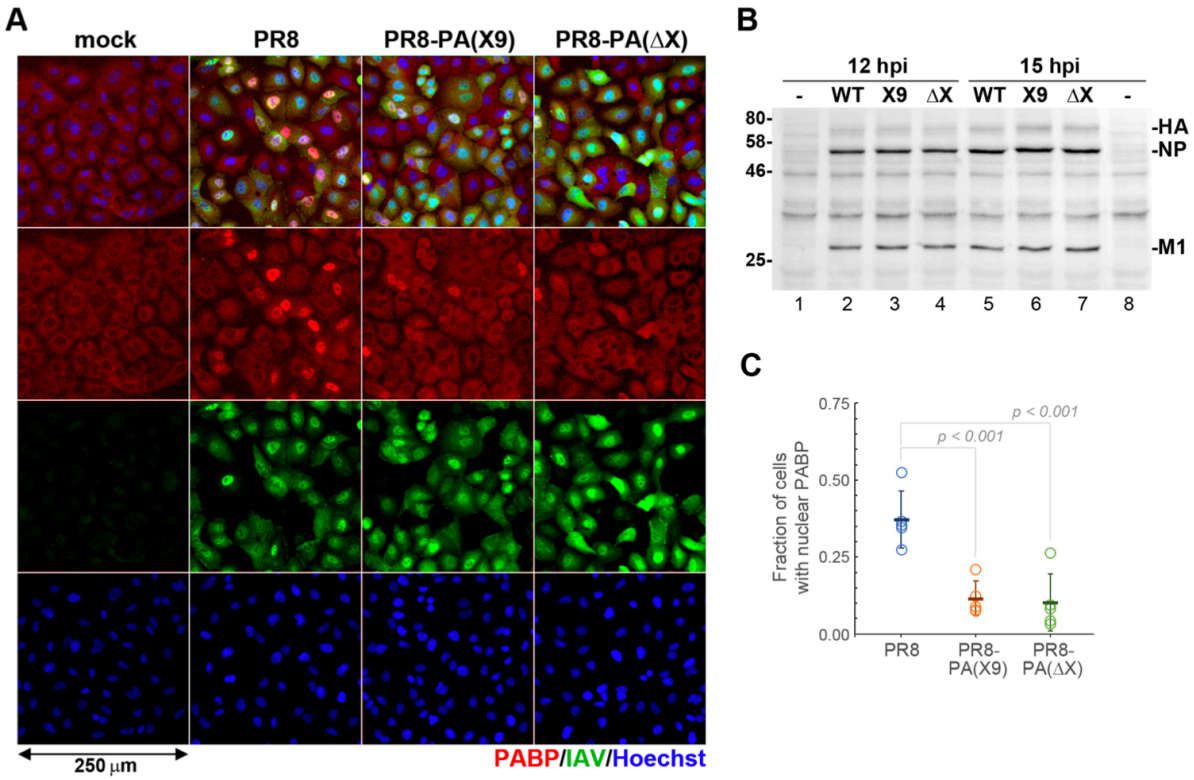

Figure 2.3: Infection rates are similar, but nuclear accumulation of PABP is reduced in cells infected with PA-X mutant viruses.
A549 cells were infected with PR8 wt, PR8-PA(X9), or PR8-PA(ΔX) at MOI = 1. (A) Immunofluorescent staining of virus-infect- ed cells at 15 h post infection (hpi) using antibodies to IAV proteins (green) and PABP1 (red). Nuclei are visualized with Hoechst dye (blue). In samples that display PABP nuclear localization, the PABP signal appears higher because it is concentrated in a smaller volume. We previously found that total levels of PABP1 do not change with PR8 infection (Khaperskyy et al., 2014)*. Scale bar = 250 μm, indicated as an arrow below the figure in the lower left corner. (B) Western blot analysis of whole cell lysates collected at 12 and 15 hpi showing viral protein accumulation in infected cells. Viral HA, NP, and M1 proteins were detected using the same polyclonal anti- IAV antibody used in A. (C) Nuclear accumulation of PABP at 15 hpi was quantified from A. The fraction of infected cells with nuclear PABP staining is plotted. Each open circle represents an independent infection experiments, in which > 200 cells from at least 3 random fields of view were counted. Horizontal lines show average values and error bars correspond to standard deviations (n = 5). p values were determined using unpaired Student's t test.*

(**Figure 2.1C, Data S2.1**). However, a small fraction of transcripts escaped shutoff (right tail end of distribution, **Figure 2.1C**). By contrast, shutoff was substantially attenuated in IAV PA(ΔX)-infected cells. Interestingly, cells infected with strains carrying either the PA(X9) or PA(fs) mutations also displayed attenuated host shutoff, and the defect was

similar in all three mutants (**Figure 2.2B**). This demonstrates that X-ORF truncation disrupts PA-X function during infection, as predicted from ectopic PA-X expression studies (Hayashi et al., 2016; Khaperskyy et al., 2016; Oishi et al., 2015). Importantly, infection rates by wt and mutant (**Figures 2.3A, 2.3B**). We also measured nuclear accumulation of cytoplasmic poly(A) binding protein (PABP), a well-described consequence of host shutoff (Khaperskyy et al., 2014; Kumar et al., 2011; Kumar and Glaunsinger, 2010; Lee and Glaunsinger, 2009). Infection with PA(X9) and PA(ΔX) viruses resulted in significantly lower rates of PABP nuclear accumulation compared to wt, confirming impairment of host shutoff (**Figures 2.3A, 2.3C**). Lastly, our RNA-seq results strongly correlated with those from a previous transcriptome profile of IAV PR8 infected cells (Bercovich-Kinori et al., 2016), despite differences in multiplicity of infection (MOI) and time course of analysis (**Figure 2.1D**). Collectively, these data demonstrate that PA-X controls the levels of the majority of host RNAs during infection.

2.2.2. PA-X causes global down-regulation of host RNAs in an ectopic expression model

To simplify our system, we also examined changes in RNA levels after ectopic PA-X expression. We used a doxycycline-inducible PA-X expression system, 'iPA-X' cells (Khaperskyy et al., 2016), to induce expression of wt PA-X or the catalytically-inactive D108A mutant in A549 cells. Because iPA-X cells were clonally selected, we analyzed two independently-generated cell lines for each variant. As expected from previous results with targeted RT-qPCR and metabolic labeling (Hayashi et al., 2015; Jagger et al., 2012; Khaperskyy et al., 2016), wt PA-X robustly down-regulated steady-state transcript levels (**Figure 2.4A, Data S2.2**). The degree of host shutoff correlated with the levels of PA-X (percentage of total reads mapping to PA-X: wt #1= 0.005%-0.006%,

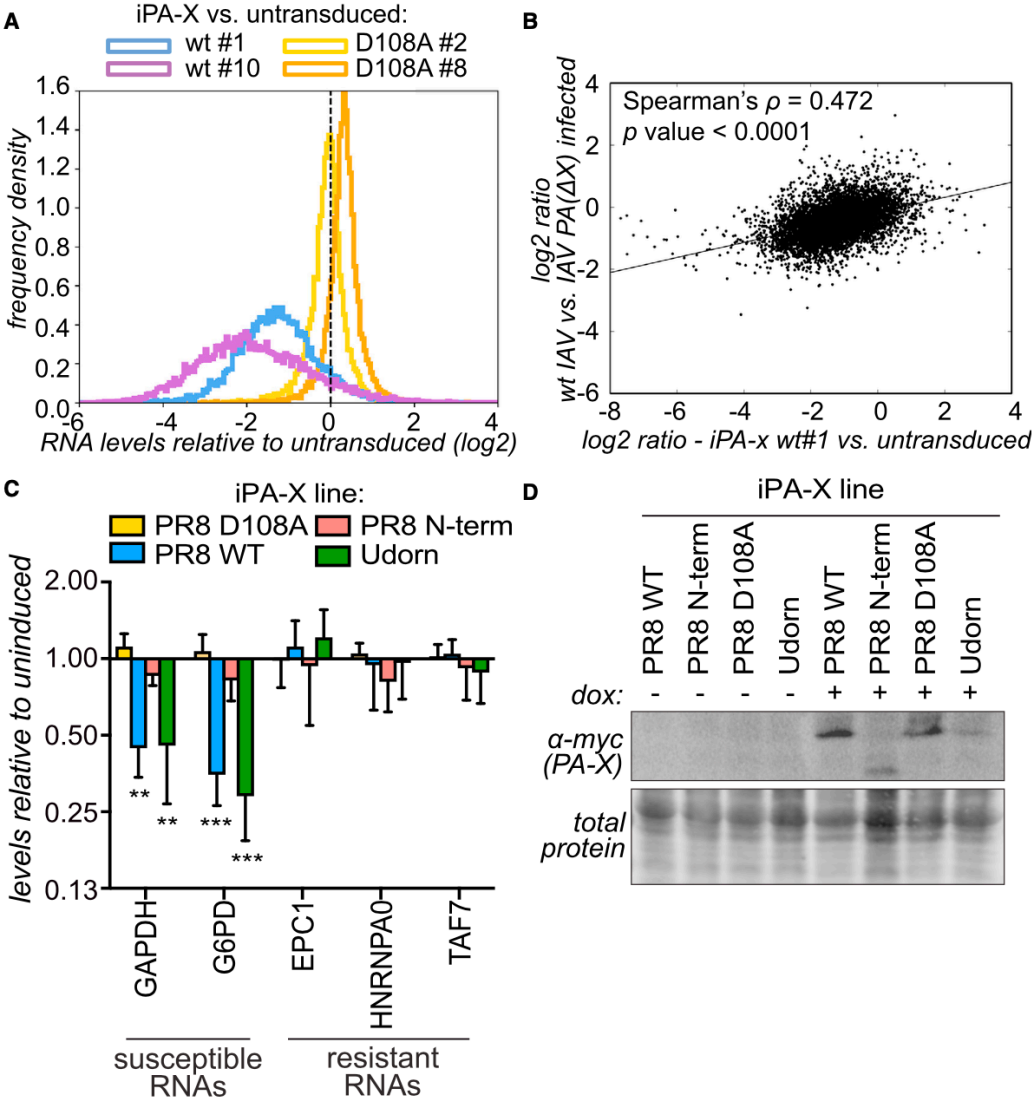

Figure 2.4: PA-X down-regulates most cellular RNAs in the absence of other viral proteins.

*RNA and protein samples were collected from control (untransduced) A549 cells, or A549 cells expressing doxycycline-inducible PR8 PA-X (wt), PR8 PA-X catalytic mutant (D108A), PR8 PA-X N terminal endonuclease domain (aa 1-191, "N term") or Udorn PA-X 18 h after addition of doxycycline. (A) RNA-seq was carried out on cells expressing wt or mutant PR8 PA-X (two clonal lines for each). The ratio between the levels in PA-X-expressing vs. control cells was computed for each RNA and the distribution of the ratios is plotted as a frequency histogram. The dashed line indicates a ratio of 1 (no change). N ≥ 2. (B) The PA-X-dependent changes in RNA levels in infected cells (ratio in PR8-PA(ΔX) vs. wt PR8) are plotted against changes in cells expressing PR8 PA-X vs. control cells. (C) Levels of several endogenous mRNAs were measured by RT-qPCR in cells expressing the indicated PA-X variants. After normalization to 18S, mRNA levels are plotted relative to uninduced cells. Values represent mean ± standard deviation. N ≥ 3. **, ***: p < 0.01, 0.001, ANOVA followed by Dunnett's multiple comparison test vs. PR8 D108A. (D) A representative western blot using anti-myc antibodies to detect myc-tagged PA-X and a total protein stain as loading control (blot section from 25 kDa to 35*

kDa) shows successful induction of PA-X in each cell line (corresponds to one of the experiments shown in panel C).

wt #10= 0.023%-0.026%), and was dependent on RNase activity, because expression of the PA-X catalytic mutant had no effect (**Figure 2.4A**). Importantly, a substantial minority of transcripts were unaffected by PA-X expression (**Figure 2.4A**, right tail end of distributions). Furthermore, we observed a highly significant correlation between the PA-X-dependent down-regulation of RNAs in the ectopic PA-X expression system and in virus-infected cells (**Figure 2.4B**). This indicates that PA-X largely targets the same RNAs in the absence of other viral proteins, and that the ectopic expression model accurately reflects the contribution of PA-X to host shutoff during infection. To further validate these findings, we selected representative RNAs, choosing RNAs that were strongly down-regulated (GAPDH, G6PD) or largely unaffected (HNRNPA0, TAF7, EPC1) in both *de novo* infection and ectopic expression models. We then validated the change in RNA levels in iPA-X cells by RT-qPCR. The RT-qPCR results agreed with the RNAseq data in terms of the selective effects on the tested transcripts (**Figure 2.4C**). By contrast, expression of the PR8 PA-X D108A mutant or the PA-X RNase domain (aa 1-191, “N term”) did not affect the level of any of the tested transcripts (**Figure 2.4C**). Expression of PA-X or the RNase domain was confirmed by western blotting (**Figure 2.4D**). Because the RNase domain alone is active *in vitro*, this result confirms that it is specifically the activity of PA-X and not overexpression of any active RNase that controls RNA levels in the iPA-X cells. Moreover, the mRNA levels were similarly affected by expression of PA-X from the A/Udorn/72 H3N2 (Udorn) strain, suggesting that target selection by PA-X is conserved among virus strains (**Figure 2.4C**). Collectively, these data demonstrate that PA-X broadly targets RNA for degradation, while a subset of RNAs remains unaffected.

2.2.3. Specific functional classes of host RNAs are differentially sensitive to PA-X

Although most RNAs were down-regulated by PA-X, the levels of ~25% of RNAs remained largely unchanged (**Figures 2.1C, 2.4A**). To identify resistant RNAs, we used *k*-means clustering to group RNAs with similar patterns of regulation (Gasch and Eisen, 2002). Clustering was carried out based on the relative RNA levels in PA-X-overexpressing *vs.* control cells or IAV *vs.* mock-infected cells in our eight datasets (**Figures 2.1C, 2.2B, 2.4A**). Two sets of RNAs comprising 55% of the RNAs that were detected in all conditions were identified as true PA-X targets (**Figure 2.5A, Data S2.1a, S2.2a**). They were down-regulated in a PA-X-dependent manner both during infection and by PA-X ectopic expression. The first set was completely PA-X-specific, as their levels were largely unchanged in IAV PA(ΔX) infected cells (**Figure 2.5A**, left panel). The RNAs in the second set were PA-X-sensitive but were partially down-regulated by other mechanisms during IAV PA(ΔX) infection (**Figure 2.5A**, right panel). By contrast, 28% of the RNAs were PA-X resistant and were not down-regulated by infection or PA-X expression (**Figure 2.5B, Data S2.1b, S2.2b**). In addition, the *k*-means algorithm identified a group of RNAs that were down-regulated during infection by a PA-X-independent mechanism, and yet were PA-X-sensitive in PA-X-expressing cells (**Figure 2.5C, Data S2.1c, S2.2c**). The levels of these transcripts may be substantially decreased by other regulatory mechanisms during infection, such that their targeting by PA-X is masked. Interestingly, based on gene ontology (GO) term analysis, the host shutoff-resistant RNAs were significantly enriched for genes involved in transcription and translation, including ribosomal RNA processing, ribosomal proteins, and membrane protein synthesis (**Figure 2.5D**). This result is consistent with the IAV requirement for host biosynthetic machinery and observations by Bercovich-Kinori *et al.* (Bercovich-Kinori et al., 2016). Collectively, these results suggest that while PA-X can target many RNAs, it retains some specificity for functional classes of RNAs.

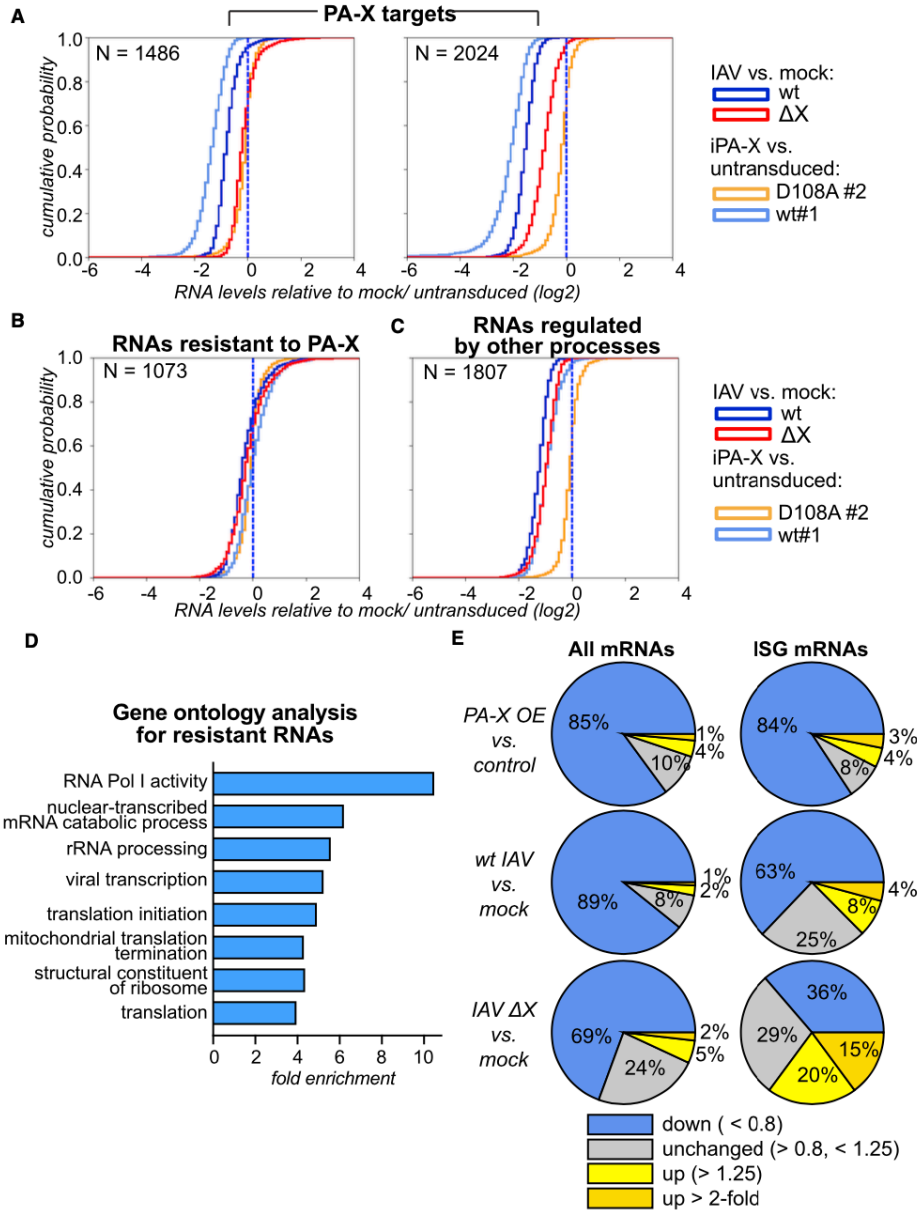

Figure 2.5: k-means clustering reveals differentially regulated groups of RNAs. *(A-C) Cluster3 was used to divide cellular RNAs in four clusters based on the pattern of fold changes in iPA-X cells (PA-X wt or D108A catalytic mutant vs. control) and cells infected with IAV (wt or PA-X-deficient IAV vs. mock). Cumulative probability histograms of fold changes for each of the classes are plotted: A: two groups of PA-X targets; B: PA-X resistant RNAs; C: potential PA-X targets that are regulated by other processes during infection. All datasets collected were used for clustering, but only select datasets are plotted for simplicity. (D) DAVID was used to identify overrepresented gene ontology (GO) terms for biological processes and molecular functions among PA-X resistant RNAs. Fold enrichment is plotted for GO terms that had corrected p values < 0.01. (E) Pie-charts showing the percentage of genes that are up- and down-regulated in infected and iPA-X cells. Left: all RNAs detected in RNAseq (wt vs. mock: N = 8573, PR8-PA(ΔX) vs. mock: N = 8848, PA-X overexpressing (OE) vs. control: N = 8554); Right: interferon-stimulated genes (ISGs; wt vs. mock: N = 167, PR8-PA(ΔX) vs. mock: N = 168, PA-X overexpressing (OE) vs. control: N = 159).*

In addition to this unbiased analysis, we examined how PA-X expression affected the levels of interferon-stimulated genes (ISGs), which are induced in infected cells and function in antiviral defense (Schoggins et al., 2011). We observed that although ISGs were induced during IAV infection, as shown by their higher expression compared to all detected RNAs, their levels were even higher in the absence of PA-X (**Figure 2.5E**). While the activity of PA-X is clearly not limited to ISGs, these data indicate that PA-X can contribute to dampening the cell-intrinsic response to infection.

2.2.4. PA-X strongly and preferentially down-regulates spliced Pol II transcripts

We previously showed that PA-X selectively degrades RNA transcribed by Pol II and spares Pol I and Pol III transcripts (Khaperskyy et al., 2016). Although all cellular transcripts are modified post-synthesis, only Pol II transcripts can be spliced. The process of RNA splicing is mechanistically linked to transcription, as recruitment of the spliceosome is mediated by the C-terminal domain of the large subunit of Pol II (Gu et al., 2013). However, a subset of Pol II transcripts naturally lack introns. When we analyzed spliced *vs*. intronless RNAs separately, we found that PA-X down-regulated spliced RNAs more than intronless RNAs (**Figure 2.6A**). Our targeted validation also showed that two intronless mRNAs, TAF7 and HNRNPA0, were not down-regulated by PA-X (**Figure 2.4C**), and, as mentioned above, the β-globin reporter used in **Figure 2.1B** encodes a spliced mRNA. Moreover, during infection down-regulation of spliced RNAs was clearly dependent on PA-X, whereas most of the down-regulation of intronless RNAs was PA-X-independent (**Figure 2.6B**). For these analyses, we only included intronless RNAs longer than 300 nt, excluding small non-coding RNAs and ensuring that the length distribution was similar between spliced and intronless RNAs. We also analyzed how the number of exons affected RNA down-regulation, as the

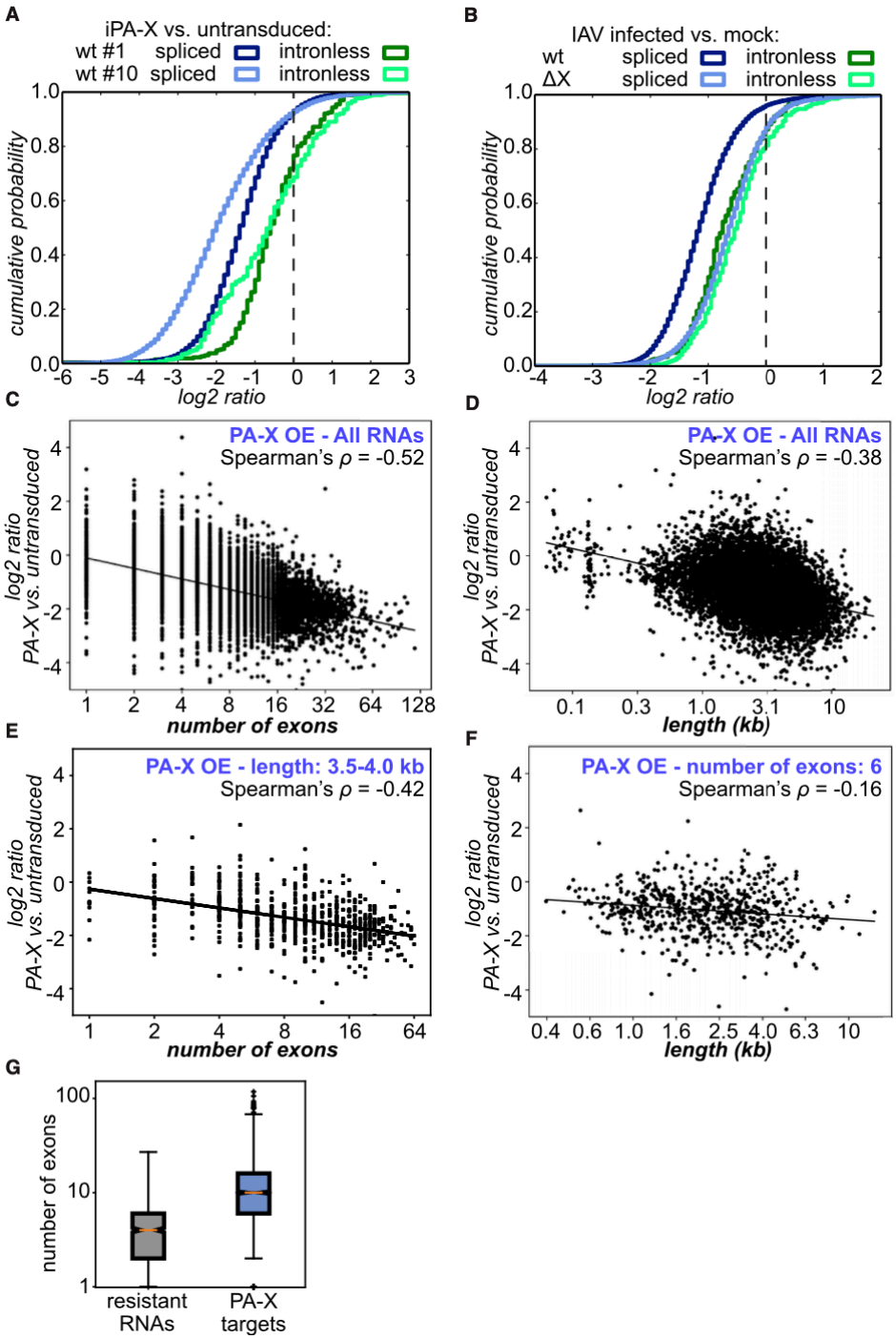

Figure 2.6: RNAs that are not spliced are less sensitive to regulation by PA-X. *(A-B) RNAseq results from Figures 2.1C, 2.4A are plotted separately for spliced and intronless RNA as a cumulative distribution histogram. A: cells overexpressing wt PA-X, clones #1 and #10; B: cells infected with wt PR8 vs. PR8-PA(ΔX). (C-F) Relative RNA levels in PA-X overexpressing ("PA-X OE", clone #1) vs. control cells are plotted against the number of exons (C, E, log2 scale) or transcript length in kb (D, F, log10 scale). C, D = all RNAs, E = RNAs with 6 exons, F = RNAs 3.5-4.0 kb in length. All correlations are statistically significant ($p < 0.001$, Spearman's test). (G) Number of exons for RNAs*

identified in the clustering analysis (Figure 2.5) is plotted. The two groups of PA-X targets (Figure 2.5A) are plotted together. p < 0.01, Kolmogorov-Smirnoff test.

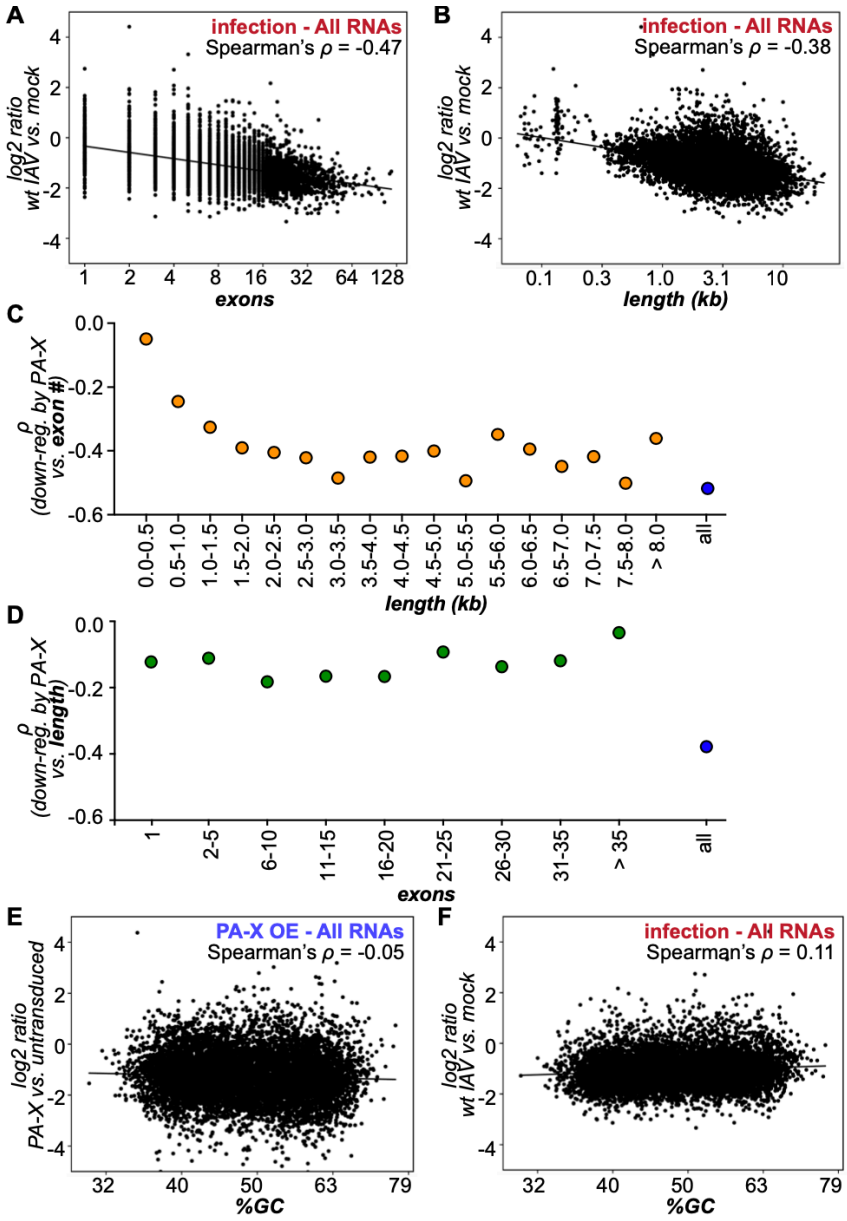

Figure 2.7: Degradation by PA-X correlates with number of exons and mRNA length, but not GC content of the mRNA.
(A-B) The relative RNA levels in wt IAV infected vs. mock infected cells are plotted against the number of exons (A, log2 scale) or transcript length in kb (B, log10 scale). (C-D) The Spearman's correlation coefficient (ρ) between the relative RNA levels in PA-X- overexpressing vs. control cells and the number of exons (C) or the RNA length (D) is plotted for groups of RNAs of the indicated length/exon number interval. Each data point represents between 99-2626 RNAs. The ρ for the entire populations are plotted in blue for comparison. (E-F) The relative RNA levels in PA-X overexpressing vs. control cells (E) or wt IAV infected vs. mock infected cells (F) are plotted against the percentage of GC of the transcript.

number of splice sites varies dramatically among spliced RNAs. Interestingly, there was a significant negative correlation between the number of exons in a transcript and its steady-state levels in PA-X-expressing and infected cells (PA-X-expressing cells: Spearman's ρ = -0.52, **Figure 2.6C**; IAV-infected cells: ρ = -0.47, **Figure 2.7A**). This result suggests that RNAs with more exons are more susceptible to PA-X degradation. However, the number of exons in an RNA is often proportional to RNA length. A prior study reported a relationship between IAV host shutoff and transcript length (Bercovich-Kinori et al., 2016) and there was also a correlation between degradation and RNA length in our data (ρ = -0.38 for both PA-X-expressing, **Figure 2.6D**, and IAV-infected cells, **Figure 2.7B**). To determine whether exon number or transcript length was important, we examined RNAs of similar length or with a specific number of exons. We still found a robust negative correlation between relative RNA levels in the presence of PA-X and exon number among RNAs of similar length (length = 3.5-4.0 kb, ρ = -0.42, N = 674, **Figure 2.6E**). Similar correlations were also seen for RNAs of other lengths (**Figure 2.7C**). By contrast, there was only a small correlation between degradation and RNA length among RNAs with the same number of exons (number of exons = 6, ρ = -0.16, N = 642, **Figure 2.6F**). Again, similar correlations were seen for other exon numbers (**Figure 2.7D**). We also tested another key characteristic of RNAs, GC content, and found no correlation with PA-X activity in our dataset (**Figures 2.7E, 2.7F**). The results from the clustering and GO analysis (**Figures 2.5A-D**) suggested that PA-X differentially regulates RNAs from specific functional groups, whereas **Figures 2.6A-F** suggested a difference based on the structure of the nascent transcript. Interestingly, there was a connection between the structural and functional specificity. RNAs classified as resistant by *k*-means clustering (shown in **Figure 2.5B**) had fewer exons than those classified as PA-X targets (shown in **Figure 2.5A**) (**Figure 2.6G**). Collectively, these

results suggest that targeting by PA-X is connected to RNA splicing, and that this preference has consequences for the selection of functionally relevant targets.

2.2.5. Splice sites confer susceptibility to PA-X

Endogenous RNAs with different numbers of exons also have different sequences, lengths and post-transcriptional modifications. To investigate the effect of splicing in a more controlled system, we examined the same RNA in both spliced and intronless forms. As a model transcript we used the mRNA for interferon λ2 (IFN-λ2), a type III IFN that contributes to IAV immune responses (Jewell et al., 2010). In the RNAseq, IFN-λ2 transcripts were only detectable in IAV-infected cells and were down-regulated by PA-X (**Figure 2.8A**). We cloned the IFN-λ2 cDNA and the full IFN-λ2 genomic sequence containing introns into plasmid expression vectors (**Figure 2.9A**), and co-transfected them into HEK293T cells with PR8 PA-X. In these transfection experiments, detection of PA-X protein is hindered by auto-cleavage of the PA-X Pol II transcript. As an alternative control to ensure comparable PA-X activity between transfections, we also measured the RNA levels of a co-transfected intron-containing luciferase reporter (Younis et al., 2010); luciferase down-regulation thus serves as a positive control for PA-X activity (**Figures 2.9B-D, 2.9G**). We confirmed that IFN-λ2 mRNA was expressed and exported to the cytoplasm at similar levels irrespective of the construct used (**Figures 2.8B, 2.8C**). We also checked that the five introns were properly spliced by PCR analysis (**Figures 2.9E-F, 2.8D**). As expected, we found that PA-X down-regulated the IFN-λ2 mRNA expressed from the full genomic region (**Figure 2.9B**). However, the levels of the same IFN-λ2 mRNA expressed from an intronless cDNA construct were only minimally reduced (**Figure 2.9B**). In both conditions, the control luciferase reporter was down-regulated by PA-X, confirming that the activity of PA-X was similar in all the samples. PA-X proteins from the 2009 pandemic H1N1 strains (such as A/California/7/09 (H1N1) (CA/7) and

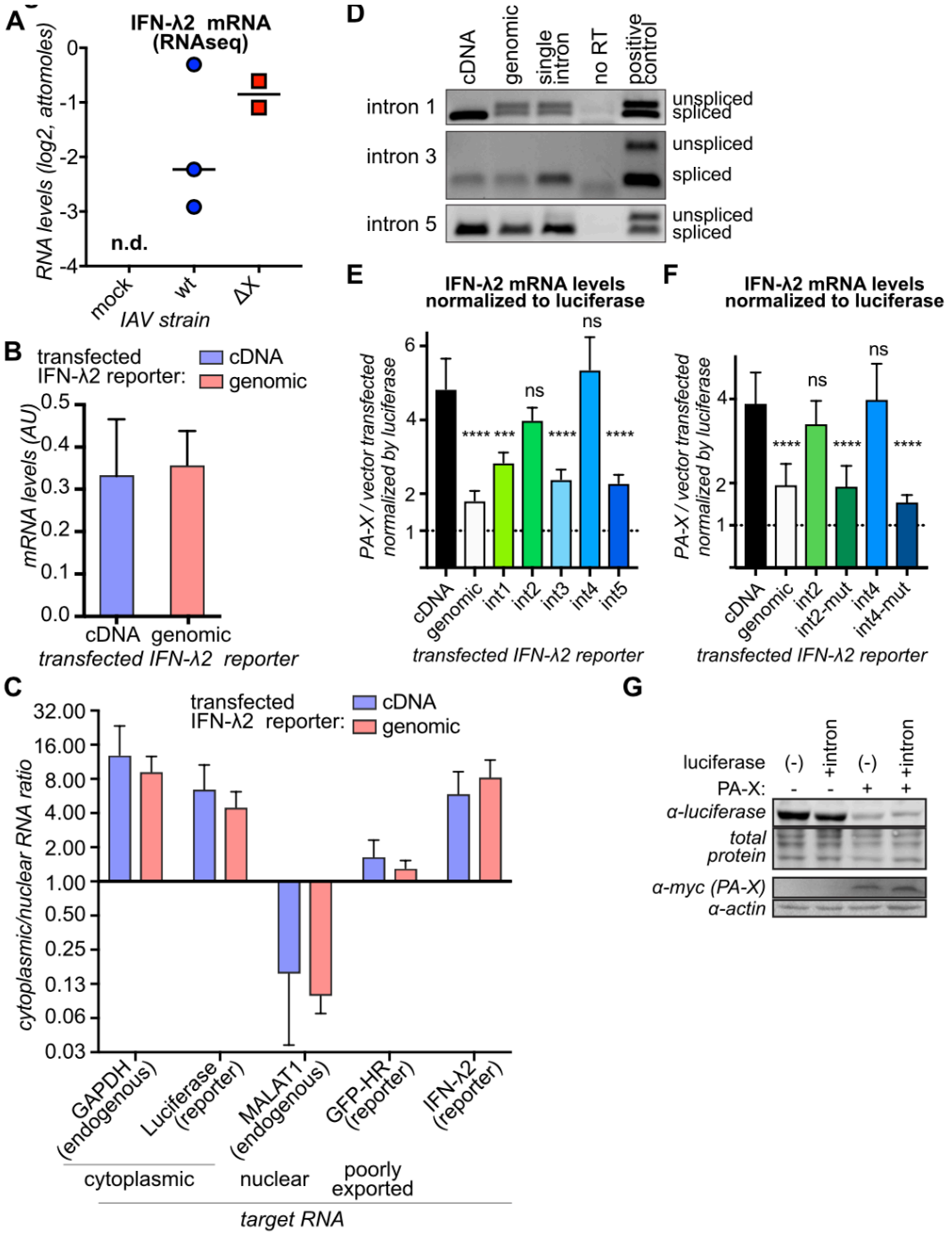

Figure 2.8: IFN-λ2 mRNA is expressed and exported at similar levels from both a genomic and a cDNA reporter constructs.
(A) The attomoles of IFN-λ2 mRNA in cells infected with wt IAV or IAV PA(ΔX) are plotted. Each point represents a biological replicate, lines represent the median value. The IFN-λ2 mRNA was undetectable in mock-infected cells (n.d. = not detected). n ≥ 2 (B) HEK293T cells were transfected with reporters expressing IFN-λ2 mRNA from cDNA or the genomic locus. RNA samples were collected 24 h after transfection and analyzed by RT-qPCR. The levels of IFN-λ2 in the absence of PA-X are plotted relative to 18S cellular rRNA. n = 4. (C) HEK293T cells were transfected with reporters expressing a luciferase mRNA, a GFP mRNA ending in a hammerhead ribozyme (GFP-HR) and IFN-λ2 mRNAs from cDNA or the genomic locus. Cells were fractionated into nuclear and cytoplasmic fraction, and the ratio between cytoplasmic and nuclear RNA levels was

calculated after normalization to 18S rRNA. The cytoplasmic/nuclear ratio for several control transcripts and the test IFN-λ2 mRNAs are plotted. As cytoplasmic controls, we included two well-exported mRNAs, the endogenous GAPDH mRNA and transfected luciferase mRNA. As nuclear control, we tested the nuclear- retained endogenous MALAT1 non-coding RNA. In addition we also included a poorly exported transfected mRNA, GFP-HR. In this construct, the 3' UTR of the original GFP vector was substituted with a self-cleaving hammerhead ribozyme (HR). While this allows processing of the RNA, it results in a non-polyadenylated RNA that is not well-exported (Lee and Glaunsinger, 2009)*. n = 4. (D) The cDNAs from vector-transfected cells from Figure 2.9C were PCR amplified across the different introns to confirm splicing. The amplified PCR products are shown (gel image is representative of four experiments). A 1:1 mix of the constructs for IFN-λ2 cDNA and genomic serves as a control to check that both products could be simultaneously amplified. (E-F) Alternative representation of the data plotted in Figure 2.9D and 2.9G, respectively, in which IFN-λ2 mRNA levels were normalized to luciferase mRNA levels (instead of 18S rRNA), and then plotted as relative levels in PA-X expressing cells vs. vector transfected cells. n ≥ 4. ****: p < 0.0001, ns: p > 0.05, ANOVA followed by Dunnett's test, p values relative to cDNA construct- transfected cells. (G) Cells were transfected with an intronless ("(-)") or an intron-containing ("+ intron") luciferase reporter and PR8 PA-X. Protein lysates were probed with antibodies against luciferase and against the myc tag to detect PA-X. For the luciferase blot, an image of the total protein stain around 40-55 kDa is shown as loading control. For the PA-X blot, actin was used as a loading control. The images are representative and correspond to one of the experiments included in Figure 2.9H. For panels B-F, bars represent mean ± standard deviation.*

A/Tennessee/1-560/2009) and the Udorn H3N2 strain also preferentially degraded the spliced mRNAs (**Figure 2.9C**). These results confirm the prediction that splicing is important for PA-X targeting. In addition, we tested the down-regulation of IFN-λ2 mRNAs expressed from chimeric constructs that contained only one of the five introns from the original genomic sequence, to determine whether a single splicing event was sufficient to restore PA-X targeting. Indeed, addition of a single intron increased susceptibility to PA-X (**Figure 2.9D**). Down-regulation of the luciferase reporter indicated that these changes were not due to varying PA-X activity in the samples. Interestingly, we found that while addition of introns 1, 3, or 5 alone restored PA-X susceptibility, introns 2 or 4 had little effect. The difference between the introns was also apparent when we normalized IFN-λ2 to luciferase mRNA levels (**Figure 2.8E**). Intron 4 was not spliced efficiently in the absence of other introns (**Figure 2.9E**), but intron 2 was still efficiently spliced (**Figure 2.9F**). Therefore, splicing efficiency alone does not explain the

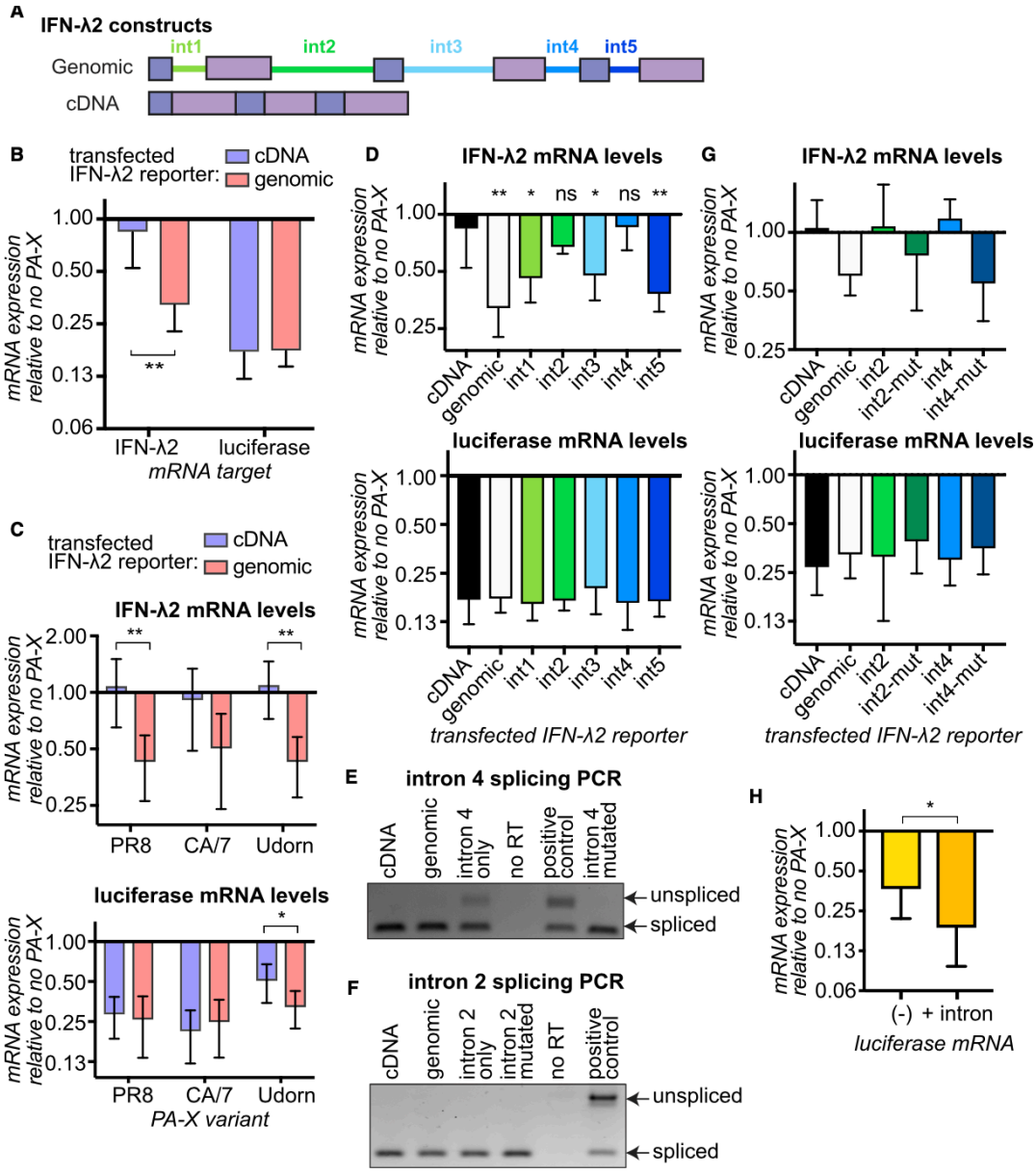

Figure 2.9: Addition of introns and splicing events promotes degradation by PA-X. *(A) Diagram of IFN-λ2 constructs. Int = intron. (B-G) HEK293T cells were transfected for 24 hrs with reporters expressing a luciferase control mRNA and IFN-λ2 mRNA from cDNA, the genomic locus, cDNA with one of the five IFN-λ2 introns added back, or cDNA with IFN-λ2 introns 2 or 4 carrying mutations that restore a canonical 5' splice site sequence. Cells were also transfected with PA-X (PR8 variant in B, D and G, PR8, CA/7 and Udorn variants in C) or vector. Levels of luciferase and IFN-λ2 mRNAs were measured by RT-qPCR and plotted as relative levels in PA-X expressing vs. vector-transfected cells, after normalization to 18S rRNA. The down-regulation of a spliced luciferase mRNA serves as a control to ensure similar PA-X activity across samples. In E and F, cDNA from vector-transfected cells was PCR amplified across the indicated introns to test splicing. Amplified PCR products are shown (image is representative of four experiments). A 1:1 mix of the IFN-λ2 cDNA and genomic constructs was included*

*to check that unspliced and spliced products could be simultaneously amplified. (H) Cells were transfected with an intronless (-) or an intron-containing (+ intron) luciferase reporter and PR8 PA-X. Luciferase RNA levels were measured by RT-qPCR, normalized by 18S rRNA and are plotted relative to vector transfected cells. For all panels, values represent mean ± standard deviation, N ≥ 4. *,**: p < 0.05, 0.01; B, C: ANOVA followed by Tukey's pairwise test; D: ANOVA followed by Dunnett's test, p values relative to cDNA construct; H: Student's t test.*

differential effects of the introns. More in-depth examination of the IFN-λ2 sequence revealed that the 5' splice sites for introns 1, 3, and 5 match the consensus 5' splice site sequence, AG|GT (where | marks the splice site). By contrast, the 5' splice sites for introns 2 and 4 are an imperfect match (TA|GT and GT|GT, respectively). 5' splice-site quality scores calculated using the MaxEntScan::score5ss program were also higher for introns 1,3 and 5 (8.8-10.5) than intron 2 and 4 (5.8) (Yeo and Burge, 2004). When we mutated these two 5' splice sites to match the consensus sequence AG|GT and increase the splice site quality score (mutated intron 2,4 = 10.7), we found that the mutated introns 2 and 4 could restore PA-X susceptibility (**Figure 2.9G, 2.8F** normalized to luciferase mRNA levels), and intron 4 splicing efficiency (**Figure 2.9E**). This result further strengthens the link between splicing and PA-X susceptibility.

Despite these results linking mRNA splicing and PA-X degradation, one unresolved issue is that we and others have previously used intronless reporters to study PA-X that appeared to be efficiently degraded. To resolve this issue, we compared intronless and spliced luciferase reporters (Younis et al., 2010). The spliced reporter, which we used as a control in **Figures 2.9B-G**, contains a portion of the β-globin intron (Younis et al., 2010). Interestingly, PA-X had a more robust effect on the spliced mRNA, although it could also down-regulate intronless luciferase mRNA (**Figure 2.9H, 2.8G**). These results suggest that addition of a single splicing event further promotes degradation by PA-X. Reporter constructs are selected for their robust expression; it is possible that certain intronless reporters also associate with cellular factors involved in PA-X targeting.

Nevertheless, our findings recommend the use of intron-containing reporters for future cell-based studies of PA-X.

2.2.6. The X-ORF mediates interaction with proteins involved in RNA metabolism

As shown in previous studies (Hayashi et al., 2016; Khaperskyy et al., 2016; Oishi et al., 2015) and the PR8 PA(X9) RNAseq results (**Figure 2.2B**), the C-terminal X-ORF is required for PA-X activity. We hypothesized that the X-ORF interacts with cellular proteins that mediate the association of PA-X with target mRNAs, especially in light of our results connecting PA-X targeting with splicing (**Figures 2.6, 2.9**) and mRNA 3'-end processing (Khaperskyy et al., 2016). To identify cellular X-ORF-interacting proteins we used BioID, a proteomic technique that relies on non-specific proximity biotinylation of lysine residues by a modified *E. coli* biotin ligase, BirA* (Roux et al., 2012). Since there are two major classes of PA-X isoforms that differ in X-ORF length (Shi et al., 2012), we fused BirA* to X-ORFs representative of each class: the 61-aa PR8 X-ORF (X61) and the 41-aa CA/7 X-ORF (X41) (**Figure 2.10A**). As negative controls, we used BirA* alone and BirA* fused to a mutated PR8 X-ORF in which four positively charged residues were replaced by alanine (X61(4A)). These mutations prevent nuclear localization of a GFP-X-ORF fusion and disrupt mRNA degradation by PA-X (Khaperskyy et al., 2016). As expected from our previous studies (Khaperskyy et al., 2016), fusion to the wt X-ORFs, but not X61(4A), led to accumulation of BirA* in the nucleus (**Figure 2.11A**). Moreover, all BirA* fusions efficiently biotinylated many cellular proteins (**Figure 2.11B**).

To identify cellular proteins that bind both the X61 and X41 X-ORFs, we affinity-purified biotinylated proteins from HEK293T cells expressing BirA*-X-ORF fusion proteins and prepared them for quantitative mass spectrometry using reductive dimethylation. Reductive dimethylation exploits formaldehyde variants with different molecular weights due to substituted carbon-13 or deuterium atoms to label captured

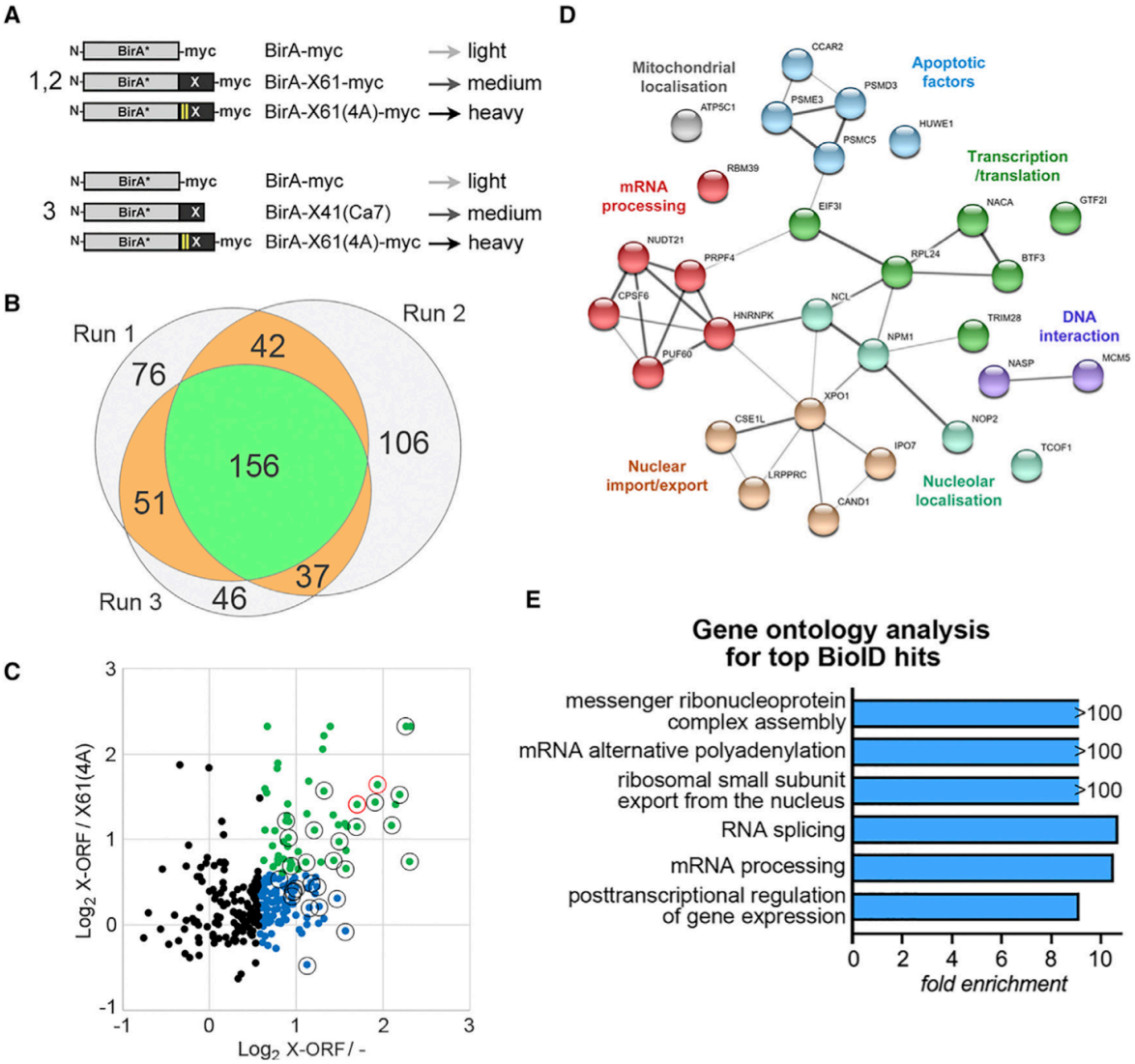

Figure 2.10: The X-ORF interactome is enriched for proteins involved in mRNA processing.

(A) Schematic diagram of X-ORF-BirA fusion baits used in BioID mass spectrometry experiment. Numbers indicate independent runs employing each construct set. Light, medium, and heavy = light, medium, or heavy isotope tags. (B) Overlap between proteins identified by mass spectrometry by ≥2 unique peptides in three BioID runs. (C) Average relative abundance of 286 proteins identified in at least two BioID experiments, plotted as log2 ratio of medium vs. light (x axis, X-ORF/-) and medium vs. heavy (y axis, X-ORF/X61(4A). Green dots = proteins with >1.5-fold enrichment over both negative controls, blue dots = proteins with >1.5-fold enrichment over BirA*-myc alone. Black and red open circles = high-confidence hits (>2.0-fold over BirA*-myc in ≥2 experiments or >1.5-fold over BirA*-myc and BirA*-X61(4A)-myc in 3 experiments). Red open circles = nucleolin (NCL) and nucleophosmin (NPM1), which were enriched > 2.0-fold over both negative controls in all three experiments. (D) STRING protein-protein interaction network of high-confidence hits. Apparent nodes were differentially colored (only one annotation per protein is shown for simplicity). (E) Gene ontology (GO) enrichment analysis of X-ORF BioID hits (black and red circles in C). All enriched functional classes*

are presented (excluding parental subclasses for each term). Note that the >100-fold enriched functional classes contain only 2 proteins each.

proteins with stable isotope tags and allow for quantitative comparisons between samples (Hsu et al., 2003). We performed the experiment three times, twice with BirA*-X61 and once with BirA*-X41, comparing them in each case to the BirA* alone and BirA*-X61(4A) controls (**Figure 2.10A**). 156 candidate X-ORF interacting proteins were represented by at least two unique peptides (**Table S2.4**) and were present in all runs (**Figure 2.10B**). Among these, we selected 29 high-confidence interacting proteins with higher relative peptide abundance in the test (BirA*-X61 or X41) *vs.* control (BirA*-X61(4A) or BirA* alone) conditions (>2-fold higher than controls in at least two experiments or >1.5-fold in all three experiments, **Figure 2.10D**, **Table S2.4**). **Figure 2.10C** depicts the relative peptide abundance of proteins in X61 samples compared to the X61(4A) control or BirA* alone control, with black and red open circles identifying the high confidence hits. Red circles indicate two proteins (nucleolin (NCL), nucleophosmin (NPM1)) that were enriched >2-fold compared to controls across all three runs. Because NCL and NPM1 are abundant proteins that traffic in and out of nucleoli, these interactions may explain the apparent nucleolar accumulation of biotinylated proteins (**Figures 2.11A, 2.11C**), even though neither the BirA-X-ORF fusion (**Figure 2.11A**) nor the full length PA-X (Khaperskyy et al., 2016, 2014) accumulate in nucleoli. STRING protein-protein interaction network analysis of the hits revealed several physical and functional interaction nodes, including protein trafficking, transcription/translation and mRNA processing (**Figure 2.10D**). Similarly, GO term analysis revealed a strong association with mRNA processing, RNA splicing and mRNA metabolic process functions among the high-confidence hits (**Figure 2.10E**). These data show that the PA-X C-terminal X-ORF is physically recruited to protein complexes involved in nuclear Pol II RNA processing (commonly referred to as mRNA processing), which likely explains

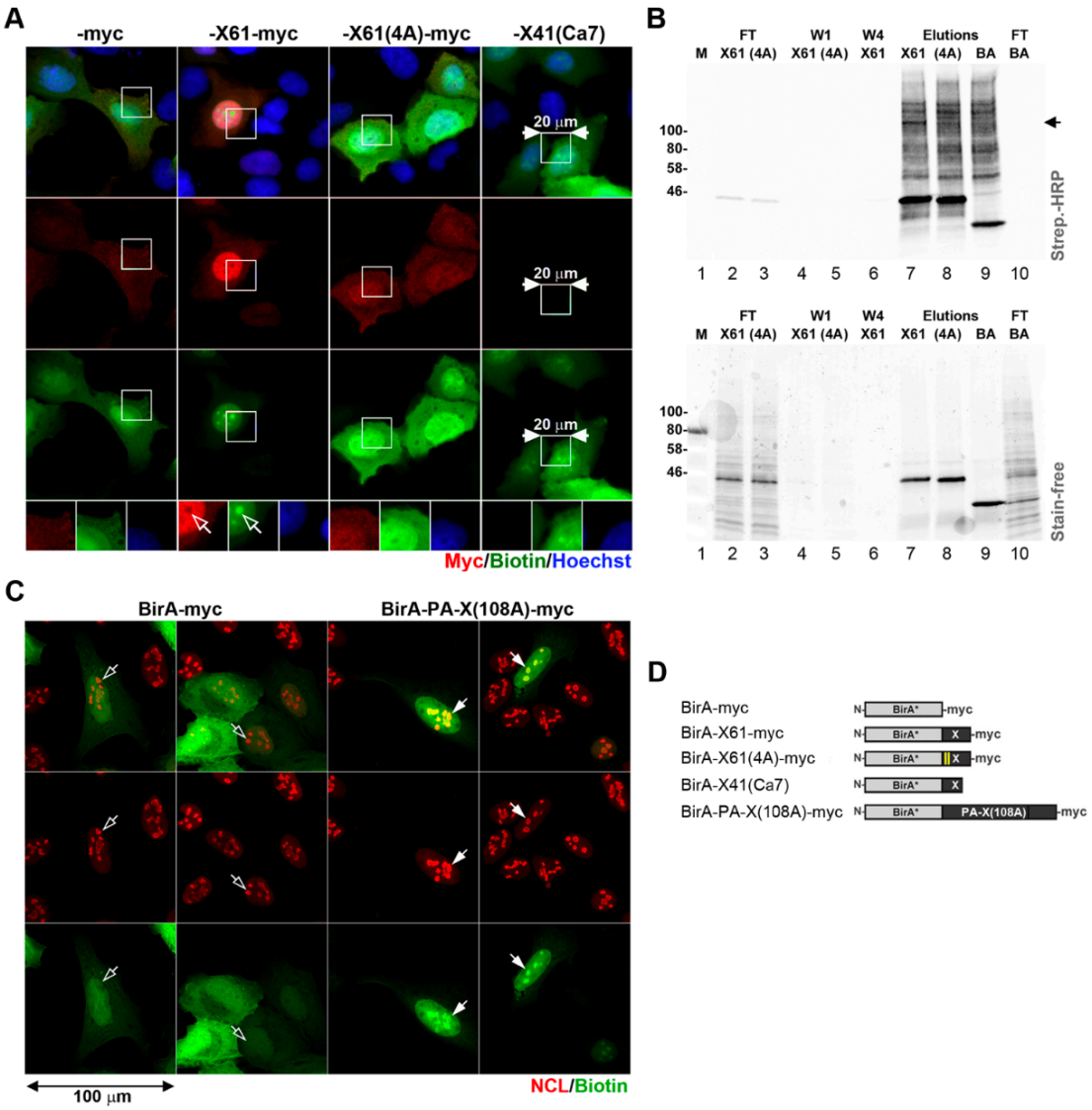

Figure 2.11: BirA*-X-ORF mediates biotinylation of cellular proteins.
(A) Immunofluorescent staining of biotinylated proteins in HEK293A cells transfected with BirA-myc (-myc), BirA*-X61-myc (-X61-myc), BirA*-X61(4A)-myc (-X61(4A)-myc), or BirA*-X41(Ca7) expression constructs. The localization of myc-tagged fusion proteins was visualized with anti-myc antibody (red), biotinylated proteins were stained with Streptavidin-Al- exa-Fluor-488 conjugate (Biotin, green), and nuclei were stained with Hoechst dye (blue). Open arrows indicate strong nucleolar biotin staining in BirA*-X61-myc expressing cells. (B) Flow through (FT), wash 1 (W1), wash 4 (W4), and elution fractions of neutravidin agarose pulldown of proteins biotinylated using BirA*-X61-myc (X61), BirA*-X61(4A)-myc (4A), or BirA*-myc (BA) were analyzed by western blot using Streptavidin-HRP conjugate (top panel). Total protein in the same fractions was detected using Bio-Rad Stain-free chemiluminescent reagent (bottom panel). Arrow indicates position of a major band corresponding to the size of nucleolin (apparent molecular weight ~110 kDa) that is enriched in BirA*-X61-myc biotinylated protein mix compared to proteins biotinylated by the other two baits. (C) Immunofluorescent staining of biotinylated proteins in cells transfected with BirA*-myc or the BirA* fused to full-length catalytically inactive PA-X mutant, BirA*-PA-X(108A)-myc. Biotinylated proteins were*

stained with Streptavidin-Alexa-Fluor-488 conjugate (Biotin, green), and nucleoli were stained with anti-nucleolin antibody (NCL, red). Filled arrows indicate strong biotinylation of nucleolar proteins, open arrows highlight nucleoli that are not preferentially labelled in control BirA-myc expressing cells. (D) Schematic diagram of BirA* fusion constructs used in panels B and C. Scale bar = 100 μm, indicated as an arrow below panel C in the lower left corner. Scale bar for magnified images = 20 μm, as indicated in panel A.*

the preferential degradation of RNAs that have undergone co/post-transcriptional processing.

2.2.7. The CFIm complex may regulate PA-X activity

The BioID screen identified several proteins involved in RNA splicing (RBM39, PUF60, PRPF4) and/or polyadenylation (NUDT21/CPSF5/CFIm25, CPSF6/CFIm68) as X-ORF-interacting proteins. We conducted co-immunoprecipitation experiments to validate these interactions using nuclear extracts derived from a HEK293T iPA-X cell line that produces high-levels of a myc-tagged PA-X (Khaperskyy et al., 2016). We recapitulated the interaction between full-length PA-X and endogenous NUDT21, suggesting this is a stable interaction that can survive affinity isolation procedures (**Figure 2.12A, 2.13A**). NUDT21 and CPSF6 assemble into a functional heterotetrameric CFIm complex (Kim et al., 2010) that enhances polyadenylation and guides polyadenylation site choice (Zhu et al., 2018). In addition, the CFIm complex is present in the spliceosome, and has been proposed to link splicing to polyadenylation during RNA processing (Rappsilber et al., 2002; Zhou et al., 2002). To test whether the interaction with the CFIm complex was required for PA-X activity, we used siRNAs to deplete NUDT21 and CPSF6 alone or in combination (**Figure 2.13B, 2.13F**). Partial silencing of CFIm proteins reduced PA-X down-regulation of IFN-λ2 upon co-transfection in HEK293T cells (**Figure 2.12B**). Moreover, in A549 cells silencing of the CFIm complex reduced PABP nuclear localization during IAV PR8 infection (**Figure 2.12C, 2.12E**), a hallmark of PA-X dependent host shutoff (**Figure 2.3**, (Khaperskyy et al., 2014)).

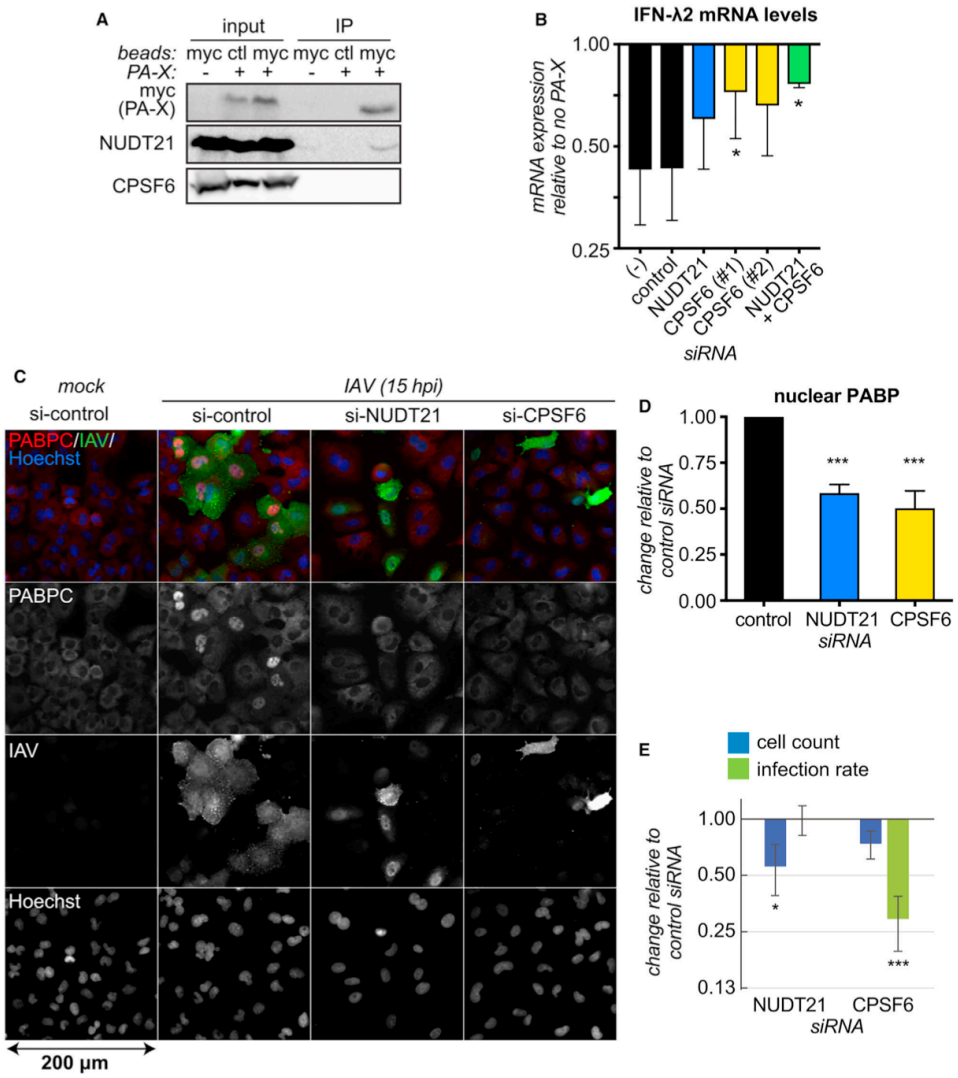

Figure 2.12: The CFIm complex is involved in PA-X activity.

*(A) Proteins were extracted from nuclei of uninduced or doxycycline-treated HEK293T cells expressing inducible wt PR8 PA-X, and incubated with myc-trap beads to immunoprecipitate PA-X-myc (myc) or control beads (ctrl). Input and IP samples were resolved by SDS-PAGE and analyzed by western blotting for PA-X-myc, NUDT21 and CPSF6. Image is representative of 3 independent experiments. (B) NUDT21 and CPSF6 were knocked down by siRNA, separately or in combination, in HEK293T cells. For NUDT21, siRNA #2 was used (see methods). For CPSF6, siRNA #1 was used for knockdown in combination with NUDT21. Cells were then transfected with a reporter expressing IFN-λ2 mRNA from the genomic locus, with and without wt PR8 PA-X. The levels of IFN-λ2 mRNA and 18S rRNA were measured by RT-qPCR. Expression of IFN-λ2 mRNA is plotted relative to vector-transfected cells, after normalization to 18S rRNA. (C-E) NUDT21 and CPSF6 were knocked down by siRNA in A549 cells, using a mixture of two siRNAs. Cells were then infected with WT PR8 IAV for 15 hrs. Infection rates were assessed by staining for IAV proteins, and host shutoff by staining for nuclear PABP. C = representative immunofluorescence images. D, E = change in fraction of infected cells with nuclear PABP or total cell counts and infected cells, respectively, relative to control siRNA. For all panels, bars are mean ± standard deviation, N ≥ 3. *,****

: p < 0.05, 0.001, ANOVA followed by Dunnett's multiple comparison test vs. control siRNA.

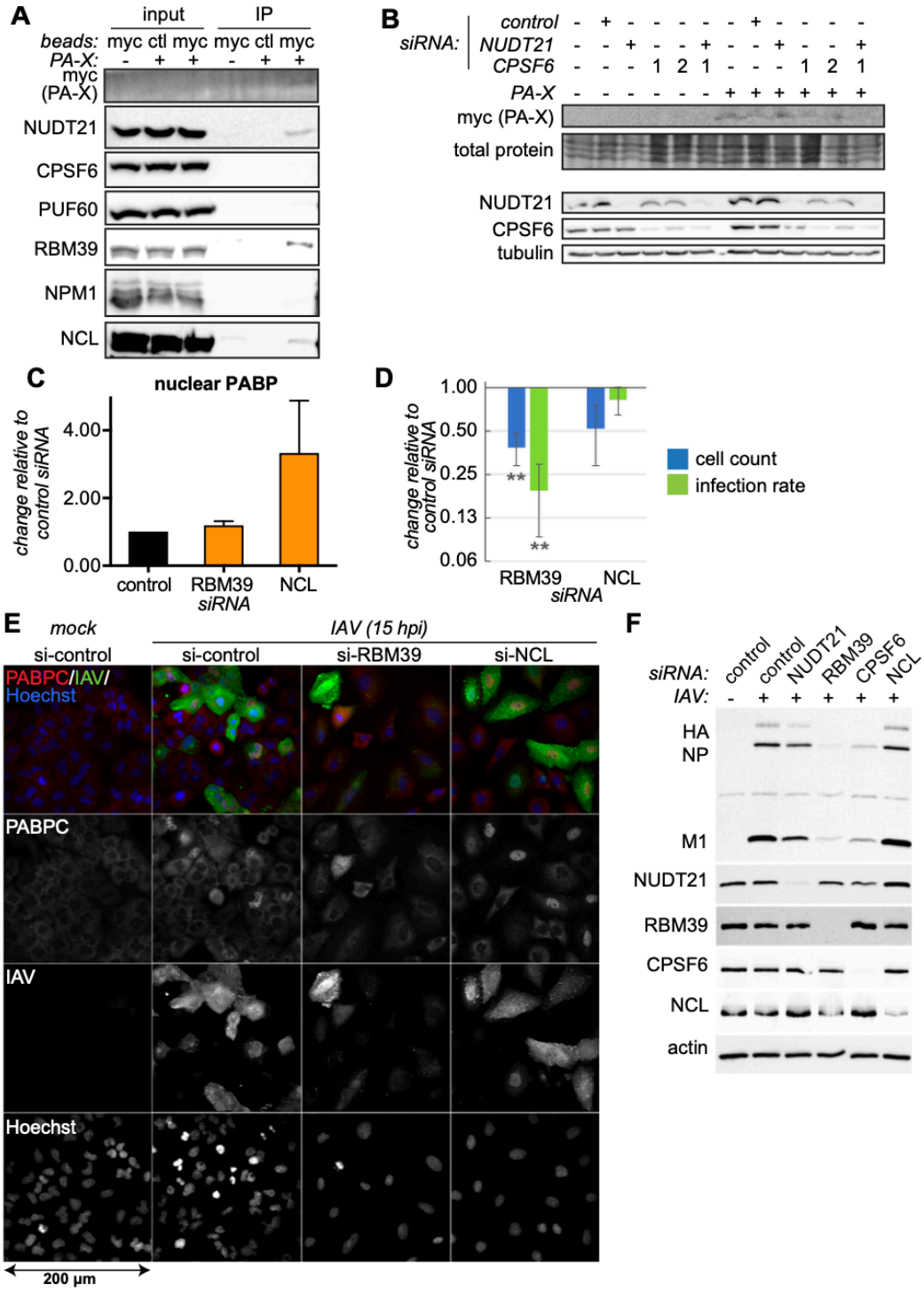

Figure 2.13: Validation and functional study of BioID hits involved in mRNA processing.

(A) Protein lysates were harvested from the nucleus of uninduced or doxy- cycline-treated HEK293T iPA-X cells, and incubated with myc-trap beads to immunoprecipitate

*PA-X-myc ("myc") or control beads ("ctrl"). Input and IP samples were resolved by SDS-PAGE and analyzed by western blotting for PA-X-myc, NUDT21, CPSF6, nucleolin (NCL), nucleophos- min (NPM1), RBM39, PUF60. A blot representative of 2-3 independent experiments is shown. (B) NUDT21 and CPSF6 were knocked down by siRNA, separately or in combination, in HEK293T cells. Protein lysates were probed with antibodies against the myc tag to visualize PA-X, and against NUDT21 and CPSF6. Tubulin serves as a loading control for the CPSF6 and NUDT21 blot, while total protein stain (shown for the portion of the blot between ~40 kDa and 55 kDa) serves as a loading control for the myc blot. The blots are representative of 3 or more replicates, and correspond to one of the experiments included in Figure 7B. (C-F) RBM39 and NCL were knocked down by siRNA in A549 cells, using a combination of two siRNAs in each condition. Cells were then infected with WT PR8 IAV for 15 hrs. Infection rates were assessed by staining for IAV proteins, and host shutoff by staining for nuclear PABP. (C) and (D) show the change in the fraction of infected cells with nuclear PABP, and in total cell counts and infected cells, respec- tively, relative to control siRNA. Bars are mean ± standard deviation, n ≥ 2. ** : p < 0.01, ANOVA followed by Dunnett's multiple comparison test vs. control siRNA (E) shows representative immunofluorescence images. Scale bar = 200 μm, indicated as an arrow in the lower left corner. (F) shows representative western blots confirming efficient knock-down, and showing accumulation of IAV proteins under knock-down conditions.*

Importantly, PABP relocalization can be quantified on a per-cell basis, which allowed us to control for unrelated effects of the knockdowns on cell viability and infection rates. In fact, we found that NUDT21 silencing reduced cell viability, and CPSF6 silencing dramatically reduced infection rates (**Figure 2.12D**). We also tested two additional potential interaction partners, NCL and RBM39, in the PABP relocalization assay (**Figure 2.13C, 2.13E**). While NCL silencing had little effect on cell viability, infection rates and host shutoff, RBM39 silencing reduced cell viability and infection rates (**Figure 2.13C-F**). We conclude that NCL is unlikely to have a role in PA-X-mediated host shutoff, whereas the effects of RBM39 silencing on cell physiology are too severe to assess its role in PA-X host shutoff. Nonetheless, collectively these data suggest that the RNA processing and spliceosome-associated CFIm complex is required for at least some of the activity of PA-X in cells.

2.3. Discussion

A thorough understanding of the molecular mechanism of action of PA-X is required to determine how it selectively degrades host RNAs and contributes to the control of the innate immune response. In this study, we discovered a key aspect of the PA-X mechanism of action: its selectivity for transcripts that have undergone splicing. This coupling to RNA processing sets PA-X apart from other viral host shutoff RNases that target mRNAs in the cytoplasm in association with translation (Covarrubias et al., 2009; Doepker et al., 2004; Feng et al., 2005, 2001; Gaglia et al., 2012; Kamitani et al., 2009). Our transcriptomic results show that, as expected, PA-X down-regulates many host RNAs, both on its own and in the context of infection. However, some RNAs are less susceptible to PA-X activity, and a key characteristic of these resistant RNAs is that they are intronless or have fewer introns. Moreover, the carboxy-terminal X-ORF of PA-X interacts with many proteins involved in cellular RNA metabolism. We propose a model whereby PA-X associates with a discrete set of RNA metabolism proteins that allows selective targeting of RNAs during transcription or early processing. In this model, RNAs that are not canonically processed, including viral RNAs, are spared.

Our transcriptomic study confirms that the PA-X-dependent down-regulation of host protein production (Hayashi et al., 2015; Jagger et al., 2012) is due to a reduction in RNA levels, and defines PA-X-dependent and PA-X-independent components of RNA down-regulation during infection. The PA-X-independent component is likely due to a recently described generalized reduction in cellular transcription (Bauer et al., 2018; Heinz et al., 2018; Zhao et al., 2018), because other known modalities of IAV host shutoff are not active in the PR8 strain (Das et al., 2008; Rodriguez et al., 2009; Salvatore et al., 2002). The mechanism of reduced host transcription in IAV infected cells remains a matter for debate (Bauer et al., 2018; Heinz et al., 2018; Zhao et al., 2018). However, it is most likely PA-X-independent, because it also occurs during

infection with influenza B viruses (Bauer et al., 2018), which do not encode PA-X (Shi et al., 2012). Our clustering analysis also revealed that some functional classes of RNAs are spared from PA-X degradation, including mRNAs for proteins involved in translation, which agrees with previous results from Bercovich-Kinori *et al.* (**Figure 2.5D**, (Bercovich-Kinori et al., 2016)). Our new results suggest that the small number of exons of these mRNAs, particularly RNAs for ribosomal proteins, may explain this phenomenon.

Another general conclusion of our RNAseq analysis is that PA-X with a 9-aa truncated C-terminal X-ORF is essentially non-functional in the context of infection. The shutoff impairment of IAV PA(X9) is very similar to that of the PA(fs) and PA(ΔX) viruses (**Fig. 2.2B**), which presumably have reduced PA-X production. This finding validates the results of multiple studies using ectopic PA-X expression models that concluded that at least 15 aa of the X-ORF is required for full RNA degrading activity in cells (Hayashi et al., 2016; Khaperskyy et al., 2016; Oishi et al., 2015). Similarly, a 1918 H1N1 chimeric virus with a stop codon after 15 aa had an intermediate host shutoff phenotype between IAV wt and PA(fs) (Jagger et al., 2012). The finding that truncating the X-ORF is sufficient to block PA-X activity in the virus is important because single point mutations in the X-ORF sequence are less disruptive than frameshifting mutations. Thus, viruses carrying X-ORF mutations could be better tools for *in vivo* studies of PA-X function and IAV pathogenesis.

The key unexpected finding from our study is the link between PA-X and splicing. All other viral host shutoff RNases appear to act at some stage of mRNP loading into the translation apparatus. For example, RNA targeting by alphaherpesvirus vhs is linked to physical interactions with translation initiation factors (Doepker et al., 2004; Feng et al., 2005, 2001) and SARS CoV nsp1 only degrades RNAs that are actively translated (Gaglia et al., 2012; Kamitani et al., 2009). Thus, to our knowledge, there is no other described instance of a host shutoff RNase using splicing as a targeting mechanism. In

fact, splicing was reported to protect mRNAs from cleavage by vhs (Sadek and Read, 2016). The connection between splicing and PA-X degradation is evident from the reduced effect of PA-X on intronless mRNAs (**Figures 2.6, 2.9**), the negative correlation between exon number and degree of degradation by PA-X (**Figures 2.6, 2.7**), and the fact that small changes in the 5' splice site can affect susceptibility to degradation by PA-X (**Figure 2.9G, 2.8F**). These findings begin to shed light on the specificity of PA-X for Pol II transcripts (Khaperskyy et al., 2016). In cellular transcription, the splicing machinery associates with RNAs through interactions with Pol II, and thus only Pol II transcripts are normally spliced (Gu et al., 2013). Protein-protein interactions with splicing factors may thus bring PA-X to its Pol II targets. This idea is corroborated by our proteomic analysis that shows that the PA-X X-ORF interacts with several splicing regulators (PUF60, RBM39, PRPF4) and spliceosome-associated polyadenylation proteins (the CFIm complex proteins CPSF5/NUDT21 and CPSF6) (**Figure 2.10**). Furthermore, PA-X activity is in part dependent on the CFIm complex (**Figure 2.12**). We speculate that more exons provide more chances for PA-X to be brought to the RNA by these factors, resulting into more efficient turnover of RNAs with more splice sites. Since these proteins do not regulate the processing of all mRNAs in the cell to the same extent, PA-X interactions with these proteins could provide an additional mechanism for target discrimination. Further studies will be needed to determine the exact role of these factors.

A targeting strategy based on splicing offers a major benefit to the virus, because it provides the ability to easily discriminate between host and viral mRNAs. Viral mRNAs are synthesized by the RdRp and most of them are not spliced, which renders them "invisible" to PA-X. That said, our published results suggest that even the two viral mRNAs that are spliced (NEP and M2) are PA-X resistant (Khaperskyy et al., 2016). However, splicing of viral mRNAs is a fundamentally different process, since the splicing

machinery needs to be recruited to the RNAs separately from Pol II (Dubois et al., 2014). It is possible that viral mRNA splicing does not require the CFIm complex or other PA-X-binding partners, because they are auxiliary components of the host RNA processing machinery. The PA-X splicing-based targeting strategy is more efficient at virus *vs.* host discrimination than the translation-based targeting strategy used by herpesviral host shutoff RNases, which leads to degradation of viral and host mRNAs alike (Abernathy et al., 2014). This is likely because viral translation relies on the same machinery as host translation. While herpesviruses can compensate for the degradation of their own RNAs, this self-sacrifice may not work for a virus like IAV that has a shorter replication cycle and a small genome with a limited gene expression program.

Our BioID results suggest that the preference for spliced RNAs may be linked to protein-protein interactions between the PA-X X-ORF and cellular factors. The X-ORF is required for PA-X nuclear localization (Hayashi et al., 2016; Khaperskyy et al., 2016; Oishi et al., 2015). However, enforced nuclear localization of the PA-X RNase domain alone does not fully rescue activity (Hayashi et al., 2016), suggesting that the X-ORF has additional functions. Indeed, we identified many X-ORF-interacting proteins with various roles in RNA metabolism in addition to a nuclear import protein (importin 7). By examining the X-ORF in isolation, we likely excluded indirect interactions via RNA binding of the RNase domain, as well as interactions that are important for PA rather than PA-X function. Among our hits, two nucleolar proteins, nucleolin (NCL) and nucleophosmin (NPM1), were also reported to interact with H5N1 PA-X (Li et al., 2016). The fact that biotinylated proteins accumulated in the nucleoli also supports the idea that nucleolin and nucleophosmin, which traffic to nucleoli, come in contact with the BirA*-fused X-ORF and full-length PA-X (**Figure 2.11C**). It is unclear whether the high nucleolar biotinylation is of functional importance, since silencing nucleolin had little effect on PA-X mediated host shutoff in infected cells (**Figure 2.13C-F**) and PA-X is not

localized specifically to this compartment (**Figure 2.11A**, (Khaperskyy et al., 2016, 2014)). Interestingly, nucleolin has been reported to protect specific RNAs from degradation by viral host shutoff RNases, including PA-X (Muller et al., 2015; Muller and Glaunsinger, 2017). Therefore, the PA-X-nucleolin interaction may reflect a different role of nucleolin in regulating RNA homeostasis during infection. By contrast, we found evidence for the involvement of the CFIm complex, NUDT21 and CPSF6, in host shutoff during infection (**Figure 2.12D, 2.12E**) and PA-X ectopic expression (**Figure 2.12B**). The findings in infected cells must be interpreted with caution because CPSF6 silencing markedly inhibited viral infection (**Figure 2.12D, 2.12F, 2.13F**). It is unclear whether the reduction in infection rates is connected to PA-X function. While the CFIm complex is more commonly studied for its roles in alternative polyadenylation and mRNA 3' processing (Hardy and Norbury, 2016), multiple studies have shown that NUDT21 and CPSF6 are found in purified spliceosome complexes (Rappsilber et al., 2002; Zhou et al., 2002). This finding has led to the idea that they may also have a role in coordination of splicing and 3' processing (Martinson, 2011). Interestingly, in a previous study we reported that canonical 3' end processing may also be linked to PA-X targeting (Khaperskyy et al., 2016); the CFIm complex may also explain this connection. Little is known about the function of the CFIm complex, so studying its contribution to PA-X activity and/or IAV infection may advance understanding of its normal physiological role. Roles for the other candidate PA-X-interacting proteins remain to be explored; such studies may be hindered if these proteins play PA-X-independent roles in the viral replication cycle or in maintaining general cell viability during infection. For example, we found that silencing of RBM39, an alternative splicing regulator, has profound negative effects on cell survival and infection rates (**Figure 2.13D-F**). We also wonder whether the association of PA-X with cellular proteins could compromise their normal function. We did not find dramatic changes in host splicing in our dataset (not shown), but others

have reported increased intron retention in cells infected with a PR8 chimeric virus bearing a 1918 NS1 protein (Zhao et al., 2018). This discrepancy could be due to the way we set up our sequencing pipeline, or the NS1 variant present in the virus. Because these changes were attributed to NS1 activity, we will have to be cautious in our assessment of PA-X-dependent and PA-X-independent effects in our system.

It is interesting that both PA-X and the well-known influenza host shutoff factor NS1 interact with nuclear mRNA processing machinery to control gene expression (Nemeroff et al., 1998). Many NS1 variants (but not PR8 NS1) cause host shutoff by binding and inhibiting a component of the 3'-end RNA processing machinery, CPSF30 (Das et al., 2008; Nemeroff et al., 1998). While this convergence could allow the two proteins to coordinate an attack on the host, studies of engineered and naturally-evolved viruses suggest that NS1 and PA-X activities are anti-correlated to prevent cytotoxicity. For example, the original 2009 pandemic H1N1 NS1 does not bind CPSF30, nor does it reduce host gene expression (Hale et al., 2010), but the more human-adapted NS1 from currently circulating pandemic H1N1 strains does (Clark et al., 2017; Nogales et al., 2018). Interestingly, these H1N1strains have also accumulated mutations that reduce PA-X activity, suggesting that having two highly active host shutoff proteins may impair viral fitness (Nogales et al., 2018). A recent study by the Takimoto lab comparing NS1 and PA-X targeting in a 2009 pandemic strain suggests that NS1 and PA-X have overlapping but not identical targets (Chaimayo et al., 2018). NS1 is more clearly directed at down-regulation of the innate immune response, whereas PA-X has a broader targeting range (Chaimayo et al., 2018). In general, host mRNA processing may be a hub of regulation for influenza because viral mRNAs are generally not processed by host machinery. Also, interactions with host processing do not directly compromise Pol II activity, which is required for viral replication (Lamb and Choppin, 1977).

Collectively, our results affirm the importance of PA-X for the viral replication cycle, as they show that in the absence of PA-X, the ability of the virus to regulate host gene expression is severely reduced. Moreover, we have uncovered a unique mechanism of host RNA targeting that can allow PA-X to distinguish not only between host and viral targets, but also among cellular targets. For example, the intronless mRNA TAF7, which we examined in **Figures 2.4B, 2.4C**, is a component of Pol II pre-initiation complexes. Therefore, PA-X selectivity could have repercussions for the viral replication cycle. Through further elucidation of PA-X mechanism of action, we will gain important insights into how host shutoff allows the virus to usurp host biosynthetic machinery and expand our knowledge of the link between PA-X and IAV pathogenesis.

2.4. Acknowledgments

We thank Albert Tai and the personnel of the Tufts University Core Facility - Genomics Core for help with the RNAseq. We thank Alejandro Cohen of the Dalhousie Proteomics and Mass Spectrometry Core Facility for support of BioID proteomics analysis. We thank Andrew Mehle and Richard Webby for constructs, Claire Moore and Andrew Bohm for suggestions and feedback, and members of the Gaglia and McCormick labs for critical reading of the manuscript. This work was supported by a Natalie V. Zucker award from Tufts University to MMG, National Institutes of Health grant R01 AI137358 to MMG, and Canada Institutes for Health Research grant MOP-136817 to CM. LG was supported by National Institutes of Health training grant T32 GM007310. REL and YK were supported by National Institutes of Health training grant T32 AI007422. CHR was partially supported by the Applied Mathematics Program of the U.S. DOE Office of Advanced Scientific Computing Research under contract number DE-AC02-05CH11231.

2.5. Declaration of interests

The authors declare no competing interests.

2.6. Methods

2.6.1. Contact for reagent sharing

Further information and request for resources and reagents should be directed to and will be fulfilled by the lead contact, Marta M. Gaglia (Marta.Gaglia@tufts.edu).

2.6.2. Cell lines

Human embryonic kidney cells HEK293A (Thermo Fisher) and HEK293T (ATCC), and human adenocarcinoma alveolar basal epithelial (A549, ATCC) cells were commercially obtained. HEK293A and HEK293T are female and A549 are male. All cell lines and derivatives were cultured in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum (Hyclone) at 37°C and 5% CO_2. HEK293T_iPA-X-PR8, A549-iPA-X_PR8 and A549-iPA-X-D108A_PR8 were previously described (Khaperskyy et al., 2016). A549-iPA-X_Udorn and iPA-X_PR8_Nterm were generated by transducing A549 cells with lentiviruses containing pTRIPZ-PA-X_Udorn-myc and pTRIPZ-PA-X_PR8-Nterm-myc.

2.6.3. Plasmids

pCR3.1-PA-X-myc (with PR8 PA-X) (Khaperskyy et al., 2014), pCR3.1-PA-N191 (PA-X 0 aa) (Khaperskyy et al., 2016), pCR3.1-PA-X_TN/CA/7-myc (Khaperskyy et al., 2016), pd2GFP-HR (Lee and Glaunsinger, 2009), pCDNA3.1-β-globin ((Covarrubias et al., 2011), subcloned from the pcTet2-βwt plasmid (Singh et al., 2008)) were previously described. The luciferase constructs with and without the intron were a kind gift from

Gideon Dreyfuss (Younis et al., 2010). pHW-PA(X9) and pHW-PA(ΔX) were generated from pHW-193 (kind gift from R. Webby) and pHW-PA(fs) vectors (Khaperskyy et al., 2016), respectively, using Phusion site-directed PCR mutagenesis to introduce the TAG stop codon in +1 ORF (synonymous ATT to ATA substitution at PA Ile-201 codon, TTG to TAG substitution at PA-X Leu-201). pCR3.1-PA-X_9aa-myc and pCR3.1-PA-X_15aa-myc were generated from the pCR3.1-PA-X-myc constructs by amplifying the truncated coding region and inserting it into the SalI-MluI sites of a pCR3.1-C-terminal-myc backbone. pCR3.1-PA-X_Udorn-myc was generated by PCR amplifying the 5' portion of the segment 3 RNA from a PolI-Udorn construct (kind gift from A. Mehle), adding a single nucleotide deletion to shift the frame of the X-ORF, and inserting into the SalI-MluI sites of a pCR3.1-C-terminal-myc backbone. pTRIPZ-PA-X-Nterm-myc and pTRIPZ_PA-X_Udorn-myc were generated by PCR amplifying PA-N191 from pCR3.1-PA-N191-myc and PA-X-Udorn-myc from pCR3.1-PA-X_Udorn-myc, respectively, and inserting these sequences into the backbone of pTRIPZ-RFP_SV40_3'UTR (Khaperskyy et al., 2016) after RFP excision with AgeI and ClaI . pCMV-IFN-λ2 cDNA, genomic and single intron constructs were generated by PCR amplifying the full human IFN-λ2 cDNA, the genomic locus or combinations of fragments of the two, and inserting into the pd2eGFP-N1 construct (Clontech) after the GFP was excised using NheI and NotI. The 5' splice site of intron 2 or 4 was then mutated from TTA|GT and TGT|GT, respectively, to CAG|GT within the single intron constructs to generate the intron 2 and 4 mutant constructs. Gibson cloning using HiFi assembly mix (New England Biolabs) was used to make all of these constructs, unless otherwise stated. The expression vector for the biotin ligase from *E. coli* with the R118G mutation BirA* (pcDNA3.1-myc-BioID2-MCS) (Roux et al., 2012) was obtained from Addgene (#74223) and the BirA* ORF was amplified by PCR and inserted into the pCR3.1-myc vector (Khaperskyy et al., 2012) between KpnI and EcoRI sites to generate pCR3.1-BirA*-myc. The PCR-amplified X-ORF sequences from

pCR3.1-PA-X-myc and pCR3.1-PA-X(4A)-myc (Khaperskyy et al., 2016) were inserted in frame with BirA* ORF using EcoRI and MluI to generate pCR3.1-BirA*-X61-myc and pCR3.1-BirA*-X61(4A)-myc, respectively. X-ORF coding sequence from A/California/7/2009 (H1N1) strain was amplified from pHW-C3 vector (Slaine et al., 2018) and inserted in frame with BirA* ORF using EcoRI and XhoI to generate pCR3.1-BirA*-X41(CA/7) vector.

2.6.4. Cell lines, lentiviral transduction and transfections

HEK293A, HEK293T, A549 cells and derivatives were cultured in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum (Hyclone) at 37°C and 5% CO_2. HEK293T_iPA-X-PR8, A549-iPA-X_PR8 and A549-iPA-X-D108A_PR8 were previously described (Khaperskyy et al., 2016). A549-iPA-X_Udorn and iPA-X_PR8_Nterm were generated by transducing A549 cells (ATCC) with lentiviruses containing pTRIPZ-PA-X_Udorn-myc and pTRIPZ-PA-X_PR8-Nterm-myc. Lentiviral packaging was carried out using the packaging plasmids psPAX2 and pMD2 (Addgene #12260, #12259). For experiments using iPA-X cells, cells were treated with 0.2 µg/ml doxycycline for 18 hrs to induce PA-X expression prior to RNA or protein sample collection. For the RNAseq experiments, untransduced A549 cells were also treated with doxycycline to serve as the control. For experiments using IFN-λ2 constructs and the β-globin reporter, HEK293T cells were plated in 24-well or 6-well plates (for fractionation experiments) and transfected with 800 ng/ml total DNA (including 50 ng/ml PA-X construct) using polyethylenimine (PEI). Cells were collected 24 h later for fractionation and/or RNA extraction and purification, and cell lysates for western blot. For siRNA transfections, HEK293T cells were plated in 6-well plates while transfecting 15 nM siRNA (ThermoFisher Scientific) per well using Lipofectamine RNAiMAX reagents (ThermoFisher Scientific). Cells were transfected two days later with 800 ng/ml total

DNA (including 50 ng/ml PA-X construct) using PEI, and collected for RNA and protein 24 hrs later.

2.6.5. Viruses and infections

Wild-type influenza A virus A/Puerto Rico/8/1934 H1N1 (PR8) and the mutant recombinant viruses PR8 PA(X9), PR8 PA(fs) and PR8 PA(ΔX) were generated using the 8-plasmid reverse genetic system (Hoffmann et al., 2000) as previously described (Khaperskyy et al., 2012). Viral stocks were produced in MDCK cells and infectious titers determined by plaque assays in MDCK cells using 1.2% Avicel overlays as described in Matrosovich *et al*. (Matrosovich et al., 2006). A549 cell monolayers were mock-infected or infected with the wild-type or mutant viruses at MOI = 1 for 1 h at 37°C. Then monolayers were washed briefly with PBS, fresh infection media (0.5% BSA in DMEM supplemented with 20 μM L-glutamine) was added and cells incubated at 37°C in 5% CO_2 atmosphere for 12 or 15 h prior to RNA isolation or preparation of lysates for western blotting. For immunofluorescence microscopy analysis cells grown on glass coverslips were infected as described above and fixed at 15 h post-infection using 4% paraformaldehyde in PBS.

2.6.6. Preparation of cell lysates containing biotinylated proteins

HEK293T cells grown on 10-cm dishes were washed briefly were transfected with BirA* fusion protein expression constructs using PEI. 6 hours post transfection media was changed to 10% FBS DMEM supplemented with 50 μM biotin (Sigma). 24 hours post-transfection (18 h post-biotin addition) cells were washed and collected in ice cold PBS, and centrifuged at 250 x *g* for 5 minutes at 4°C. Cell pellets were resuspended in 500 μl RIPA buffer (50 mM Tris-HCl pH 7.4, 150 mM NaCl, 1% Igepal, 0.5% sodium

deoxycholate, 0.1% SDS) with protease inhibitor cocktail (P8340, Sigma) and lysed at 4°C for 1 hour with gentle agitation, followed by passing through a 21-gauge needle. Lysates were cleared by centrifugation at 4°C for 20 min at 20,000 x *g*.

2.6.7. Neutravidin pull-down

60 µl of 50% slurry of High Capacity Neutravidin Agarose Beads (Thermo) was used for each 500 µl of clarified whole cell lysate. Beads were equilibrated in RIPA buffer by washing three times for 10 mins at 4°C. In one of the BirA*-X61 experimental runs, 1 µl of 500x RNase A (100 µg, Qiagen) was added to each sample to remove non-specific interactors. The lysate was then incubated for 5 minutes at room temperature before loading onto the beads. Untreated samples were loaded directly onto the beads post washing. 1 mg of protein sample was loaded to beads in 1.5 mL Eppendorf tubes, which were then placed on a rotator overnight at 4°C, and collect with centrifugation at 400 x g for 1 minute at 4°C. Beads were washed with RIPA buffer three times, followed by three washes with TAP buffer (50 mM HEPES-KOH pH 8.0, 100 mM KCl and 10% glycerol).

2.6.8. Mass spectrometry sample preparation

Beads were resuspended in 50 mM triethylammonium bicarbonate (TEAB) buffer (Sigma). 24 mM DTT and 32 mM IAcNH2 were added sequentially. Beads were then incubated for 30 minutes at 37°C, washed with 50 mM TEAB and centrifuged for 1 min at 400 x *g* before resuspending in 50 mM TEAB. On-bead trypsin (Pierce Trypsin protease, MS-Grade; Thermo Scientific) digest was performed with 1 µg trypsin in 50 mM TEAB buffer, shaken overnight at 37°C. Samples were acidified with 1 µl trifluoroacetic acid (TFA) and 3 µl Formic acid until a pH lower than 3 was achieved. Trypsinized peptides were collected by puncturing a hole in the bottom of the 1.5 mL

Eppendorf tube using a 30-gauge needle, placing it in a 2 mL Eppendorf tube and spinning it at 2000 rpm for 1 minute at room temperature. Beads were washed with 50 mM TEAB prior to desalting. Samples were desalted with Oasis/SepPak Desalting columns, eluted sequentially in 1 mL 50% ACN/0.1% TFA and 500 µl 70% ACN/0.1% TFA. Combined eluted samples were dried in a Thermo SPDIIIV speed vacuum centrifuge and frozen at -20°C.

2.6.9. Reductive dimethylation and quantitative mass spectrometry

Quantitative mass spectrometry analysis via reductive dimethylation enabled measurements of relative abundance of biotinylated proteins in each experimental condition (Hsu et al., 2003). In reductive dimethylation, formaldehyde molecules with different combinations of stable hydrogen and carbon isotopes are conjugated to peptide samples. Dried protein samples were resuspended by sonication for 15 minutes in 50 mM TEAB. BirA*, BirA*-XORF (X61 or X41) and BirA*-X61(4A) samples were labeled with light, medium and heavy isotopes respectively. 8 µl formaldehyde (Sigma) were added to the light, 15 µl D2-formaldehyde (Cambridge Isotope Laboratories Inc.) to the medium and 15 µl D2-C13-formaldehyde (Aldrich) to the heavy samples. Reactions were incubated for 5 minutes at room temperature. Once incubation was completed, 0.51 M $NaCNBH_3$ (sodium cyanoborohydride; Fluka) was added to the light and medium samples, while 0.51 M $NaCNBD_3$ (sodium cyanoborodeuteride; Aldrich) was added to the heavy reaction, to label terminal amines. All three reactions were incubated for 1 hour at room temperature before being combined into a single tube at a 1:1:1 ratio. The combined sample was acidified, desalted and dried as described above. Samples were resuspended in 3% ACN/0.1% formic acid and sonicated for 15 min to prepare for mass spectrometry. Mass spectrometry and peptide identification was performed at Dalhousie Proteomics CORE Facility by Dr. Alejandro Cohen (https://medicine.dal.ca/research-dal-

med/facilities/proteomics.html). Proteome Discoverer software (Thermo) was used for protein identification. Functional Protein Association Network analysis was conducted on 29 selected protein hits using online STRING version 10.5 (https://string-db.org/) (Szklarczyk et al., 2017).

2.6.10. Cell fractionation for RNA analysis

Fractionation was performed as described previously (Gagnon et al., 2014) with minor modifications. Briefly, cells were collected 24 h after transfection, centrifuged at 500 x *g* for 5 min at 4°C, then washed with PBS and counted. Equal numbers of cells were aliquoted into two tubes, one for the whole cell lysates collection and one for fractionation. Cells were pelleted again, and lysed on ice for 10 mins in 250 µl ice-cold hypotonic lysis buffer (10 mM Tris pH 7.5, 10 mM NaCl, 3 mM $MgCl_2$, 0.3% (vol/vol) NP-40, 10% (vol/vol) glycerol in nuclease-free water) supplemented with 100 U RNasin (Promega). For the whole cell lysate, Trizol (Life Technologies) was added directly to the lysate to extract RNA. For nuclear/cytoplasmic fraction, the lysate was centrifugated at 1000 x *g* for 3 min at 4°C to pellet membrane and nuclei. The supernatant was collected as the cytoplasmic fraction and Trizol was added to it to extract RNA. Finally, the nuclear pellet was washed 3 times with 1 ml hypotonic lysis buffer and collected by centrifugation at 200 x *g* for 2 mins at 4°C, then lysed directly in Trizol to extract nuclear RNA.

2.6.11. RNA purification, cDNA preparation and qPCR

For fractionation experiments, RNA was purified using Trizol. 1 ml Trizol, 2 µl glycogen and 200 µl chloroform (Fisher Scientific) were added to each fraction. Samples were then centrifuged at 16,000 x *g* for 15 min at 4°C, and the aqueous layer was

collected. RNA was precipitated by addition of 700 µl isopropanol and incubation for 10 min at room temperature, followed by centrifugation at 16,000 x *g*, 4°C for 20 min. The pellet was washed with 75% ethanol, and resuspended in RNase-free water. For other transfection experiments, RNA was extracted from cells and purified using the Quick-RNA miniprep kit (Zymo Research), following manufacturer's protocol. In all cases, the RNA was treated with Turbo DNase (Life Technologies), then reverse transcribed using iScript supermix (Bio-Rad) per manufacturer's protocol. In the fractionation experiment, the same cell equivalents of total, nuclear and cytoplasmic fraction were used for these steps. qPCR was performed using iTaq Universal SYBR Green supermix (Bio-Rad), on the Bio-Rad CFX Connect Real-Time System qPCR and analyzed with Bio-Rad CFX Manager 3.1 program. The primers used are listed in **Table 2.1**.

2.6.12. Intron splicing verification

Proper splicing of each intron for all IFN-λ2 construct was verified by PCR amplification across each splice sites using primers listed in **Table 2.1**. PCR products were run on a 2% agarose gel containing HydraGreen safe DNA dye (ACTGene) and imaged with a Syngene G:Box Chemi XT4 gel doc system.

2.6.13. Co-immunoprecipitation from nuclear lysates

HEK293T_iPA-X-PR8 cells were plated in 10-cm dishes. Cells were then treated with 1 µg/mL doxycycline for 18 hours to induce PA-X expression. Fractionation of nuclear lysates was performed as described previously (Dadi et al., 2013) with minor modifications. Briefly, cells were washed with ice-cold Dulbecco's PBS and centrifuged at 1,200 rpm for 6 minutes at 4°C to collect the cells. Cell pellets were lysed in 50 µL of a sucrose-based lysis buffer (0.32 M sucrose, 3 mM $CaCl_2$, 2 mM MgOAc, 0.1 mM EDTA,

Table 2.1: Primers used for qPCR and splicing PCR.

PRIMER	SOURCE	IDENTIFIER
IFN-λ2/3 Forward: AGTTCCGGGCCTGTATCCAG	Ank et al., 2006	N/A
IFN-λ2/3 Reverse: GAGCCGGTACAGCCAATGGT	Ank et al., 2006	N/A
18S rRNA Forward: GTAACCCGTTGAACCCCATT	Abernathy et al., 2015	N/A
18S rRNA Reverse: CCATCCAATCGGTAGTAGCG	Abernathy et al., 2015	N/A
Luciferase Forward: ATCGAGGTGGACATTACCTACG	Khaperskyy et al., 2016	N/A
Luciferase Reverse: CGCTCGTTGTAGATGTCGTTAG	Khaperskyy et al., 2016	N/A
GAPDH Forward: AGCCACATCGCTCAGACAC	Arias et al., 2014	N/A
GAPDH Reverse: TGGAAGATGGTGATGGGATT	Arias et al., 2014	N/A
G6PD Forward: TGAGCCAGATAGGCTGGAA	Hu et al., 2013	N/A
G6PD Reverse: TAACGCAGGCGATGTTGTC	Hu et al., 2013	N/A
TAF7 Forward: TAGCACTGATCCTAAAGCAAGC	Wang et al., 2012	PrimerBank ID: 14717406c3
TAF7 Reverse: GGCAGAGTAATTCCGTGGTTC	Wang et al., 2012	PrimerBank ID: 14717406c3
EPC1 Forward: AAGCAGCTCATTCACATACAGC	Feng et al., 2013	N/A
EPC1 Reverse: ACTTCATCTTCAGAATCCAAATCA	Feng et al., 2013	N/A
HNRNPA0 Forward: GCGACCTGATCGAGCACTTC	Wang et al., 2012	PrimerBank ID: 52426775c2
HNRNPA0 Reverse: CGCGTCGTGATTCTGGAAATAC	Wang et al., 2012	PrimerBank ID: 52426775c2
β-globin Forward: GTGGATCCTGAGAACTTCAGGCT	Morgado et al., 2012	N/A
β-globin Reverse: CAGCACACAGACCAGCACGT	Morgado et al., 2012	N/A

IFN-λ2 intron 1 Forward: AGCTGGTTTAGTGAACCGTCAGATCCGCTAGC	This paper	N/A
IFN-λ2 intron 1 Reverse: GCATCCGGGAGAGCCCCGTGGAGCCTGG	This paper	N/A
IFN-λ2 intron 2 Forward: CCAGGCTCCACGGGGCTCTCCCGGATGC	This paper	N/A
IFN-λ2 intron 2 Reverse: CTCAGGTCCCAGGTCCTGGGGAAGAGG	This paper	N/A
IFN-λ2 intron 3 Forward: CCTCTTCCCCAGGACCTGGGACCTGAG	This paper	N/A
IFN-λ2 intron 3 Reverse: CTGGTCCAAGACGTCCACCAGGGCTGG	This paper	N/A
IFN-λ2 intron 4 Forward: CCAGCCCTGGTGGACGTCTTGGACCAG	This paper	N/A
IFN-λ2 intron 4 Reverse: GGAGGCGGCCCCGGGTCCTGGGCCCTGC	This paper	N/A
IFN-λ2 intron 5 Forward: GCAGGGCCCAGGACCCGGGGCCGCCTCC	This paper	N/A
IFN-λ2 intron 5 Reverse: GCTGATTATGATCTAGAGTCGCGGCCGCATA GCGACTGGGTGGCAATAAATTAAGAC	This paper	N/A

10 mM DTT, 0.5 mM PMSF). An additional 50 µL of sucrose lysis buffer with 0.5% (vol/vol) NP40 was added to the lysate. The lysate was centrifuged for 10 minutes at 1,100 x *g* at 4°C. The supernatant was collected as the cytoplasmic fraction, whereas the pellet was washed in 100 µL of sucrose lysis buffer, and centrifuged for 10 minutes at 2,000 rpm at 4°C. The supernatant was discarded and the pellet was lysed in 100 µL of Soluble Nuclear Lysis Buffer (50 mM HEPES pH 7.8, 50 mM KCl, 300 mM NaCl, 0.1 mM EDTA, 10% Glycerol, 1 mM DTT, 0.1 mM PMSF, cOmplete protease inhibitors (Roche)) and incubated on a rotating rotor at 4°C for 45 minutes. Samples were centrifuged at 10,000 x *g* for 3 minutes at 4°C. 20 µL of the supernatant was collected as the input sample, and the rest was added to myc-trap magnetic agarose beads

(ChromoTek), or control magnetic agarose beads (ChromoTek) and incubated on a rotating rotor for 45 minutes at 4°C. Beads were collected on a magnetic rack and washed 6 times with wash buffer (50 mM Tris HCl pH 7.4, 150 mM NaCl, 1 mM EDTA, 0.05% NP40, cOmplete protease inhibitors (Roche)). 25 µL of Laemmli sample buffer was added to the beads, and samples were incubated for 10 minutes at 95°C prior to SDS-PAGE and western blotting.

2.6.14. Western blotting and immunofluorescence

Western blotting and immunofluorescence were carried out as previously described (Khaperskyy et al., 2016, 2014) using the antibodies listed in the STAR methods table. For detection of biotinylated proteins HRP-conjugated streptavidin (Cell Signaling) and Alexa-Fluor-488-conjugated streptavidin (Molecular Probes) were used.

2.6.15. RNA-seq

For RNAseq analysis of PA-X-expressing cells, A549 and A549 iPA-X cells were induced with 0.2 µg/ml doxycycline for 18 hours. For RNAseq analysis of infected cells, A549 cells were infected with PR8 wt, PA(fs), PA(X9), and PA(ΔX) viruses for 15 hours. RNA lysates were collected and purified using the RNeasy kit (Qiagen). 650-750 ng of RNA were mixed with ERCC ExFold RNA spike-in mix (0.65 µl or 0.75 µl, respectively, Invitrogen/ThermoFisher) prior to the start of library preparation. The spike-in controls were included to better normalize the final RNA levels. This particular kit contains two mixes that can be used for the control *vs.* test samples and that have known fold change differences. This information can be used to re-calibrate the samples. Prior to library preparation, the levels of select human mRNAs were tested by RT-qPCR, to confirm that PA-X overexpression/IAV infection had the expected effect. Libraries were prepared

using the TruSeq Stranded Total RNA Library Prep Kit with Ribo-Zero Human (Illumina) following the manufacturer's protocol. Library preparation 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. Single-end 50 nucleotide reads were obtained with a multiplexing strategies, using a total of four total lanes. For the replicates, the spike mixes for the samples and the barcoded primers were switched, in order to control for potential biases.

2.6.16. Read alignment and bioinformatic analysis

Reads were aligned with Tophat2 v2.1.1 (Kim et al., 2013) to hg19, IAV PR8 and the ERCC spike sequences. Default settings were used, except --library-type fr-firststrand, and a gtf of the hg19 annotation was provided as reference. **Table 2.2** summarizes the results of the alignment. FPKMs were computed using Cufflinks v2.2.1 (Trapnell et al., 2012). Default settings were used, except --library-type fr-firststrand and –u, and a gtf of the hg19 annotation was provided as reference. Only previously annotated RNAs with a level of >1 FPKM in all the control samples (A549 + dox or mock-infected A549) were used in further analyses. Because most annotated RNAs are Pol II transcripts, the downstream analysis includes predominantly Pol II transcripts. The FPKMs for the RNAs were converted to attomoles of RNA based on the known concentration of the spike-in controls. **Table 2.3** shows the high correlation in measured RNA levels between replicate samples. The absolute RNA levels in the replicates were averaged, and the relative RNA levels (RNA ratio) in infected *vs.* mock infected or PA-X-expressing *vs.* control cells was computed. All downstream analysis was carried out on the relative levels (ratio) in log2 scale. **Data S2.1 and S2.2** and subtables summarize the results of the RNAseq analyses. The hg19 annotation was used to derive the number of exons,

Table 2.2: Summary of RNAseq read data.

Number of total and aligned reads per sample.

replicate	sample		total	aligned	% aligned
1	IAV infection	mock	19667828	18123484	92.1%
1	IAV infection	IAV wt	26682554	24882935	93.3%
1	IAV infection	IAV PA(fs)	32310728	29248702	90.5%
2	IAV infection	mock	49002230	46981383	95.9%
2	IAV infection	IAV wt	61459700	59845236	97.4%
2	IAV infection	IAV PA(ΔX)	58099857	55889949	96.2%
2	IAV infection	IAV PA(X9)	60267367	58017799	96.3%
3	IAV infection	mock	50657894	48070697	94.9%
3	IAV infection	IAV wt	60876293	58944335	96.8%
3	IAV infection	IAV PA(fs)	59310604	57530179	97.0%
3	IAV infection	IAV PA(X9)	55085364	53229027	96.6%
4	IAV infection	mock	50453374	48618713	96.4%
4	IAV infection	IAV PA(ΔX)	59601781	57407566	96.3%
1	PA-X overexpression	A549 + dox	23541334	22689807	96.4%
1	PA-X overexpression	iPA-X D108A #2 + dox	26175509	24858959	95.0%
1	PA-X overexpression	iPA-X D108A #8 + dox	19188386	18183207	94.8%
1	PA-X overexpression	iPA-X wt # 1 + dox	21037487	19583024	93.1%
1	PA-X overexpression	iPA-X wt #10 + dox	17534270	15392606	87.8%
2	PA-X overexpression	A549 + dox	54385249	52298884	96.2%
2	PA-X overexpression	iPA-X D108A #2 + dox	50816019	49198442	96.8%
2	PA-X overexpression	iPA-X D108A #8 + dox	52964604	51451854	97.1%
2	PA-X overexpression	iPA-X wt # 1 + dox	41843644	39077421	93.4%
2	PA-X overexpression	iPA-X wt #10 + dox	55490125	52909892	95.4%
3	PA-X overexpression	A549 + dox	62518863	61112687	97.8%
3	PA-X overexpression	iPA-X D108A #8 + dox	68994621	67360890	97.6%
3	PA-X overexpression	iPA-X wt # 1 + dox	58113723	55911429	96.2%

length of transcripts and GC content for the analyses in **Figures 2.6, 2.7**. For analysis of intronless RNAs, only RNAs that were longer than 300 nt were used, to enrich for mRNAs and long non-coding RNAs and exclude small non-coding RNAs that are transcribed by Pol III or are produced through processing of longer transcripts, like microRNAs and small nuclear and nucleolar RNAs. The length distribution of the remaining intronless RNAs was similar to that of the spliced RNAs. All RNAs were used for the analysis of PA-X down-regulation *vs.* length, exon number, and GC content. *k*-means clustering analysis was carried out using the Cluster 3.0 program (http://bonsai.hgc.jp/~mdehoon/software/cluster/software.htm). The clustering was done only on RNAs that were detected in all samples tested (6,391 RNAs), because all eight

Table 2.3: Correlation between RNA levels measurements in RNAseq replicates.

List of correlation coefficients from pairwise comparison of replicates (Spearman's ρ).

sample		sample1-ID	sample2-ID	Spearman's ρ	P value
IAV Infection	mock	827_1	829_1	0.9738	< 0.0001
IAV Infection	mock	827_1	755_6	0.9735	< 0.0001
IAV Infection	mock	827_1	829_5	0.9804	< 0.0001
IAV Infection	mock	829_1	755_6	0.9684	< 0.0001
IAV Infection	mock	829_1	829_5	0.9723	< 0.0001
IAV Infection	mock	755_6	829_5	0.9698	< 0.0001
IAV Infection	IAV wt	829_2	755_7	0.9469	< 0.0001
IAV Infection	IAV wt	829_2	827_2	0.9686	< 0.0001
IAV Infection	IAV wt	755_7	827_2	0.9567	< 0.0001
IAV Infection	IAV PA(fs)	829_3	755_8	0.9544	< 0.0001
IAV Infection	IAV PA(X9)	827_4	829_4	0.9616	< 0.0001
IAV Infection	IAV PA(ΔX)	827_3	829_6	0.9593	< 0.0001
PA-X overexpression	A549 + dox	827_12	755_1	0.9658	< 0.0001
PA-X overexpression	A549 + dox	827_12	829_10	0.9771	< 0.0001
PA-X overexpression	A549 + dox	755_1	829_10	0.9733	< 0.0001
PA-X overexpression	iPA-X wt # 1 + dox	829_13	755_4	0.9703	< 0.0001
PA-X overexpression	iPA-X wt # 1 + dox	829_13	827_15	0.9770	< 0.0001
PA-X overexpression	iPA-X wt # 1 + dox	755_4	827_15	0.9688	< 0.0001
PA-X overexpression	iPA-X wt #10 + dox	755_5	827_16	0.9496	< 0.0001
PA-X overexpression	iPA-X D108A #8 + dox	829_12	827_14	0.9643	< 0.0001
PA-X overexpression	iPA-X D108A #8 + dox	829_12	755_3	0.9613	< 0.0001
PA-X overexpression	iPA-X D108A #8 + dox	827_14	755_3	0.9726	< 0.0001
PA-X overexpression	iPA-X D108A #2 + dox	827_13	755_2	0.9708	< 0.0001

datasets were used to generate the clusters, even though only the mRNA levels for select samples are plotted in **Figures 2.5A-C** for clarity. The classified RNAs listed in **Data S2.1a,b,c** and **S2.2a,b,c** represent these 6,391 RNAs. Because in *k*-means clustering the number of clusters is user-defined, clustering was attempted with three, four, and five clusters. Four clusters were chosen, because they provided more granularity. For example, they identified a group of RNAs that were only PA-X-dependent in the ectopic expression system. Initializing the program with more than four clusters led to separation of PA-X targets in multiple groups different only by the extent

of down-regulation, but did not identify other patterns of gene expression. Gene ontology (GO) term analysis was carried out on the DAVID server (Huang et al., 2009a, 2009b). The 5' splice site quality score was computed using the MaxEntScan::score5ss program using the maximum entropy model (Yeo and Burge, 2004). Other analyses were done using custom scripts in Python2.7. For the ISG analysis (**Figure 2.5E**) the list of ISG tested by Schoggins *et al.* was used (Schoggins et al., 2011).

2.6.17. Quantification and statistical analysis

For **Figures 2.1B, 2.4C, 2.8, 2.9, 2.12, 2.13**, statistical analysis and plotting were done in GraphPad Prism v7.0d software using the test recommended by the software and indicated in the figure legends. Generally, ANOVA followed by a corrected pairwise test (Tukey's or Dunnett's) was used when more than two samples were analyzed and Student's *t* test when two samples were compared. For **Figures 2.1C-D, 2.4A-B, 2.6**, statistical analysis and plotting were done using Python2.7 and the NumPy, SciPy and matplotlib libraries. The Kolmogorov-Smirnoff test was used to compare populations and Spearman's correlation coefficient to analyze the relationship between certain variables and RNA down-regulation. Statistical tests used for each panel are noted in Figure legends and/or Figures.

2.6.18. Data and software availability

RNAseq data is available on the GEO database, entry GSE120183.

2.7. Supplemental information

Data S2.1 (.zip file) (related to Figures 2.1, 2.4):

Includes four tab-delimited .txt tables reporting fold-changes from replicate averages and standard deviations:

DataS2.1all: Relative RNA levels in IAV-infected cells

Summary of relative RNA levels in IAV *vs.* mock infected cells for all four IAV strains used.

DataS2.1a,b,c: Relative RNA levels in IAV-infected cells classified by cluster

Summary of relative RNA levels in IAV-infected *vs.* mock infected cells for all four IAV strains used separated based on their classification by clustering analysis (**Figure 2.5A-C**). S3a = PA-X targets (**Figure 2.5A**). S3b = PA-X resistant RNAs (**Figure 2.5B**). S3c = RNAs regulated by other processes during infection (**Figure 2.5C**).

DataS2.2 (.zip file) (related to Figures 2.1, 2.4):

Includes four tab-delimited .txt tables reporting fold-changes from replicate averages and standard deviations*:*

DataS2.2all: Relative RNA levels in PA-X-expressing cells

Summary of relative RNA levels in PA-X-expressing *vs.* control cells for all four iPA-X cell lines used.

DataS2.2a,b,c: Relative RNA levels in PA-X-expressing cells classified by cluster

Summary of relative RNA levels in PA-X-expressing *vs.* control cells for all four iPA-X cell lines used separated based on their classification by clustering analysis (**Figure 2.5A-C**). S4a = PA-X targets (**Figure 2.5A**). S4b = PA-X resistant RNAs (**Figure 2.5B**). S4c = RNAs that are regulated by other processes during infection (**Figure 2.5C**).

Table S2.4 (.xlsx file) (related to Figure 2.10): Summary of BioID results.

List of proteins identified in each experimental run and of the 29 high-confidence hits

2.8. Author contribution

RNAseq experiments and analyses were performed by Marta Gaglia (Figures 2.1, 2.2, 2.4A-B, 2.5, 2.6, 2.7, 2.8A). BioID screen and analyses, and microscopy experiments were carried out by Brittany Porter, Emma Price, Craig McCormick and Denys Khaperskyy (Figures 2.3, 2.10, 2.11, 2.12C-E, 2.13C-F). The co-IP experiment depicted in Figures 2.12A and 2.13A was carried out by Rachel Levene. The rest of the experimental work (Figures 2.4C-D, 2.8B-G, 2.9, 2.12B, 2.13B) was carried out by Léa Gaucherand.

Chapter 3. The cut site specificity of the influenza A virus endoribonuclease PA-X allows it to discriminate between host and viral mRNAs

Gaucherand L, Iyer A, Gilabert I, Rycroft CH, Gaglia MM. To be resubmitted to *Nature Microbiology*.

3.1. Introduction

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 cells from mounting an antiviral response and redirecting cellular resources to viral processes. Several viruses from divergent families accomplish host shutoff through RNA degradation by endoribonucleases (endoRNases), including the important human pathogen influenza A virus, a negative-sense single-stranded RNA virus with a segmented genome (Bercovich-Kinori et al., 2016; Gaucherand and Gaglia, 2022; Jagger et al., 2012). This convergent evolution highlights the benefit of using endoRNases to induce host shutoff. In influenza A virus, the virus-encoded endoRNase PA-X drives widespread depletion of host RNAs (Gaucherand et al., 2019; Jagger et al., 2012). PA-X activity has critical consequences for the host, as infection with a PA-X-deficient virus induces more inflammation in animal models, often causing higher morbidity and mortality than wild-type infections (Gao et al., 2015b; Gong et al., 2017; Hayashi et al., 2015; Hu et al., 2015; Jagger et al., 2012). While immune modulation by PA-X *in vivo* has been widely documented, few studies have investigated its molecular mechanism of action. Moreover, the RNA targeting specificity of PA-X and its contribution to PA-X function during infection are still incompletely understood.

PA-X is produced from segment 3 of influenza A virus, which also encodes the polymerase acidic (PA) subunit of the influenza RNA-dependent RNA-polymerase (FluPol). PA-X is produced when a +1 ribosomal frameshift occurs after addition of amino acid 191 (Firth et al., 2012; Jagger et al., 2012). PA-X thus shares its N-terminal domain with PA, which has a PD-D/E-XK nuclease fold and endoRNase activity (Dias et al., 2009; Yuan et al., 2009). PA uses this domain to “snatch” capped 5’ ends of host mRNAs to initiate viral mRNA transcription (Dias et al., 2009), whereas PA-X uses it for host shutoff (Gaucherand et al., 2019), and *in vitro* studies suggest that the two proteins

have different substrate preference (Bavagnoli et al., 2015). We previously found that PA-X activity is not indiscriminate, as it down-regulates host RNAs synthesized by RNA Polymerase II (RNAPII) but not RNAPI and III (Khaperskyy et al., 2016). Interestingly, this feature is shared by all characterized viral endoRNases and degradation factors that drive host shutoff (Gaglia et al., 2012; Gaucherand and Gaglia, 2022). In influenza A virus, this specificity may be explained by a connection between PA-X and cellular RNA splicing, which may govern how PA-X reaches its target RNAs (Gaucherand et al., 2019). However, unspliced reporter RNAs can also be degraded to some extent, suggesting that there may also be other determinants. Additionally, unlike some other viral endoRNases, PA-X does not appear to consistently down-regulate viral mRNAs (Khaperskyy et al., 2016), suggesting it can distinguish between host and viral transcripts. Indeed, viral RNA levels were unaffected by PA-X in cells infected with the influenza A/PuertoRico/8/1934 (H1N1) virus strain (Khaperskyy et al., 2016). However, some change was reported in the context of infections with the A/California/04/2009 (H1N1) strain (Chaimayo et al., 2018). Moreover, splicing alone may not explain the virus-host discrimination, as the mRNAs from the influenza A virus M and NS segments are spliced by the host machinery but spared by PA-X. Moreover, not all spliced RNAs are down-regulated by PA-X, while some intronless RNAs are. These observations indicate that there are additional unknown components to PA-X selectivity.

While the role of cofactors in RNase specificity is well-established, cut site specificity remains poorly understood for many endoRNases. Yet sequence and/or structure preferences have been identified when it has been examined. For example, phage T4 RegB cleaves mRNA inside GGAG sequences in specific structures (Lebars et al., 2001; Saïda et al., 2003), and fungal α–sarcin cleaves 28S ribosomal RNA at GAGA sequences within loops (Endo et al., 1990). However, the complexity of cut site motifs often hampers their identification using classical *in vitro* approaches. Next generation

sequencing has recently been used to tackle this challenge by profiling endoRNase cleavage products in cells. For example, it was used to identify the complex preferred motif of *Salmonella* endoRNase E (Chao et al., 2017), and we have employed it to identify the cut site specificity of the Kaposi's sarcoma-associated herpes virus (KSHV) endoRNase SOX (Gaglia et al., 2015). Knowing the cut site specificity is critical to understand which RNAs are directly targeted by RNases and how.

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 (Abernathy et al., 2015) and the inherent instability of certain RNAs. To bypass these confounds and better understand PA-X targeting, we directly probed PA-X cut sites throughout the transcriptome using 5' Rapid Amplification of cDNA Ends (5' RACE) adapted to high-throughput sequencing and our PyDegradome pipeline (Gaglia et al., 2015). We report that PA-X preferentially cleaves host RNAs at GCUG sequences within the loop of hairpin structures, and likely after splicing. Importantly, these preferred cleavage motifs are more abundant in host than viral mRNAs. Moreover, inserting a preferred cleavage sequence in a viral segment is detrimental to viral growth only in the presence of PA-X. These findings suggest that PA-X cut site specificity 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.

3.2. Results

3.2.1. 5' RACE-seq and PyDegradome identify PA-X cut sites transcriptome-wide

In preliminary 5' RACE experiments, we found that PA-X cuts reporter mRNAs at a specific location. We thus hypothesized that PA-X cleaves host RNAs at distinct places in the transcriptome, contrary to what we previously thought (Khaperskyy et al., 2016). To test this hypothesis at a transcriptome-wide level, we used 5' RACE coupled to high-

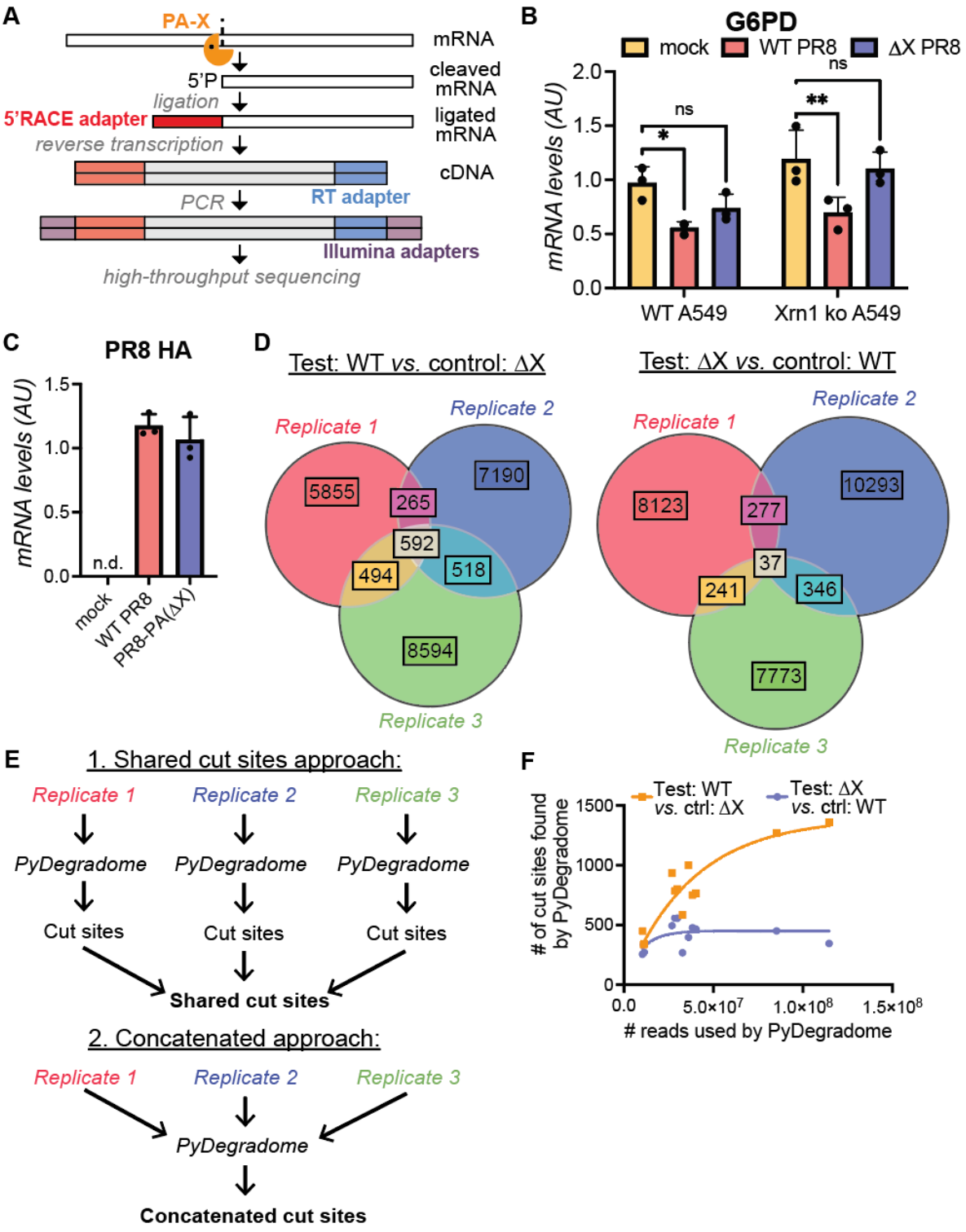

Figure 3.1: 5' RACE-seq with PyDegradome analysis identifies PA-X cut sites transcriptome-wide.

(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, infected with WT PR8 or with 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; ns, not significant; *, p < 0.05; **, p < 0.01; Two-way ANOVA with Dunnett's multiple comparison test. n = 3 (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(∆X)), which are likely to be PA-X cut sites; right diagram shows PR8-PA(∆X) specific fragments (test: PR8-PA(∆X), 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. (F) Number of PA-X cut sites (orange squares) or RNA fragments enriched in the PR8-PA(∆X) infected cells (purple circles) identified by PyDegradome relative to the number of input reads.*

throughput sequencing (5' RACE-seq, **Figure 3.1A**) and compared RNA fragments from mock infected cells to cells infected with the wild-type (WT) influenza A/PuertoRico/8/1934 (H1N1) virus strain (henceforth "PR8"), or PR8 engineered to lack PA-X (referred to as PR8-PA(∆X) or ∆X, **Figure 3.2A** (Gaucherand et al., 2019)). The mutations in PR8-PA(∆X) introduce a stop codon in the PA-X +1 reading frame and are also expected to reduce the amount of frameshifting, thus abolishing production of full-length PA-X (Gaucherand et al., 2019). They are silent in the PA 0 reading frame, so PA protein levels are not affected by these mutations (**Figure 3.2B**). Since the host exonuclease Xrn1 degrades RNA fragments following cleavage by PA-X (Khaperskyy et al., 2016), we used Xrn1 knock-out human lung A549 cells to enrich for PA-X fragments and improve identification of the location of the PA-X cut (Liu and Moss, 2016) (**Figure 3.2C**). We confirmed by qRT-PCR that the samples expressed similar levels of viral genes and that host shutoff was detected in the WT PR8 infected samples, using the PA-X target G6PD mRNA as a readout (Khaperskyy et al., 2016) in both WT and Xrn1 knock-out A549 cells (**Figure 3.1B-C**). For 5' RACE-seq, we ligated an RNA adapter to the 5' phosphate at the end of cleaved RNAs, which include PA-X cleaved RNAs, and prepared sequencing libraries (**Figure 3.1A**). After aligning reads to the human genome, the junction between the 5' RACE adapter and the human sequence represents the position where PA-X cut the RNA.

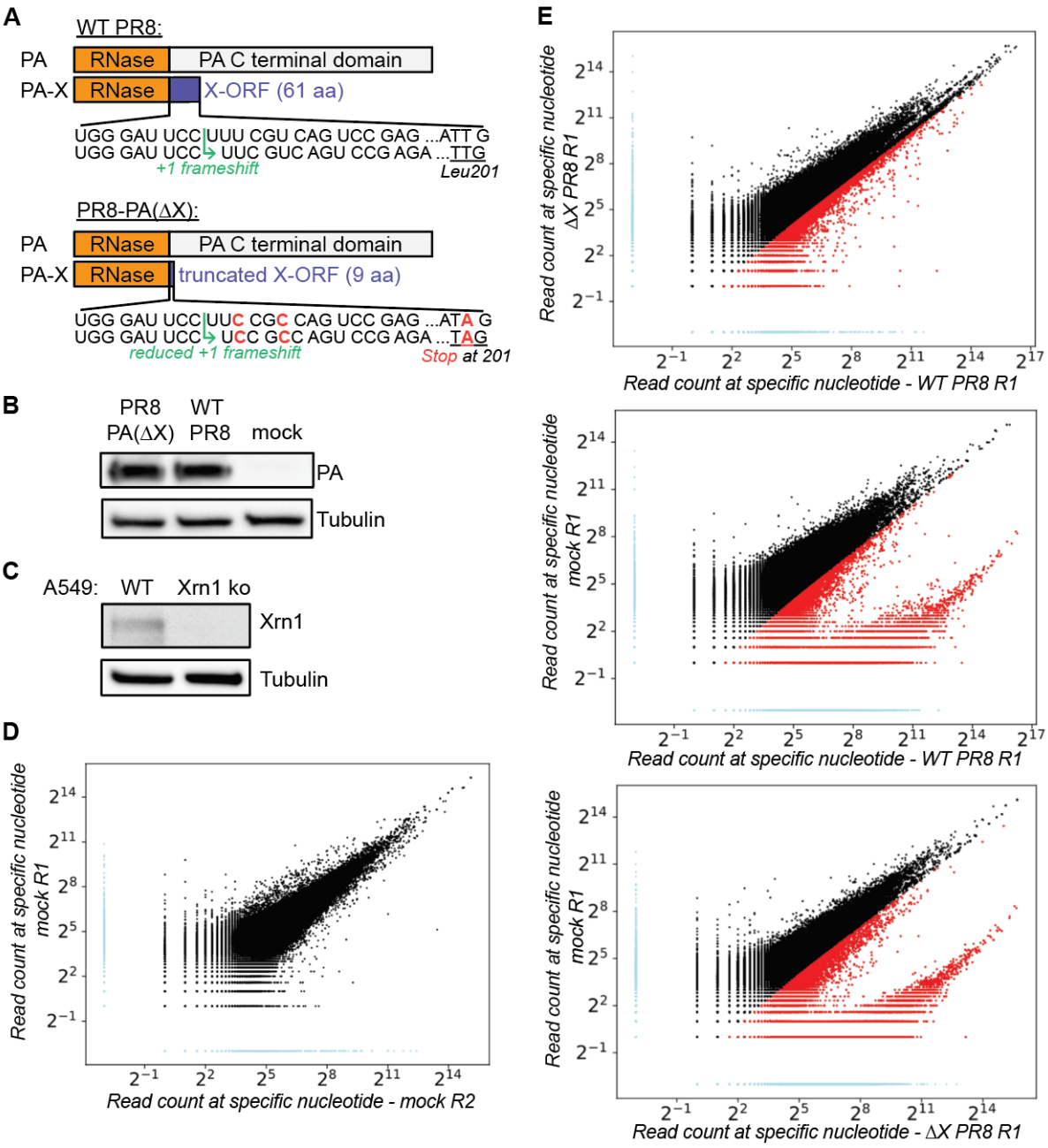

Figure 3.2: Characteristics of the system used to identify PA-X cut sites transcriptome-wide.

(A) Strategy used to engineer a virus that lacks PA-X, PR8-PA(ΔX), compared to WT PR8. Adapted from (Gaucherand et al., 2019). (B) Protein lysates of Xrn1 knock out (ko) A549 cells infected with WT PR8 or PR8-PA(ΔX), or mock infected, were probed with antibodies against PR8 PA, PR8 PA-X or b-tubulin as a loading control. Images are representative of 2 experiments. (C) Protein lysates of WT or Xrn1 ko A549 cells were probed with antibodies against Xrn1, or b-tubulin as a loading control, to check for efficient knock out 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(∆X) (top), WT PR8 vs. mock infected (middle) and PR8-PA(∆X) 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.

Our protocol yielded a high level of background reads, presumably due to the lack of basal RNA degradation in Xrn1 knock-out cells. All libraries had a similar number of reads (**Table 3.1**), and ~75% of reads mapped to locations that were not shared across samples (**Figure 3.2D-E**, light blue datapoints). Additionally, the number of reads that started at the same nucleotide across multiple samples was highly correlated (**Figure 3.2D-E**, black and red datapoints), suggesting common background RNA degradation. However, some locations had higher numbers of reads in infected cells (**Figure 3.2E**, red datapoints). These fragments could be due to PA-X cleavage, as well as other infection-related processes, such as cap-snatching and cellular stress responses. To correct for background degradation and identify locations that were enriched in WT PR8 compared to PR8-PA(∆X) infected cells, we used PyDegradome (Gaglia et al., 2015). This pipeline assumes the first read nucleotide after the RNA adapter is the 5' end of the RNA fragment and counts how many reads within each dataset map to the same 5'

Table 3.1: Summary of RACE-seq read data.

Number of total reads sequenced, reads that aligned exactly one time to hg38/ERCC spike ins/PR8, and reads used in PyDegradome analysis for each sample.

Replicate	Sample	Total	Uniquely aligned	% uniquely aligned	Used as PyDegradome input
1	Mock	49956233	29851634	59.8%	26375238
1	WT PR8	59832882	30563534	51.1%	18055280
1	PR8-PA(∆X)	61091210	32019504	52.4%	22210910
2	Mock	46603110	27547835	59.1%	23708011
2	WT PR8	50920984	28938148	56.8%	16459354
2	PR8-PA(∆X)	62507002	32229277	51.6%	21897094
3	Mock	57788531	37176976	64.3%	31313003
3	WT PR8	60000926	32354246	53.9%	17821338
3	PR8-PA(∆X)	53011926	27782595	52.4%	18229661

position. These counts are then compared between a test and a control sample, and a Bayesian probability model is used to determine which locations are significantly higher in the test sample. Here, we updated PyDegradome to identify cut sites in both mature mRNAs and unspliced pre-mRNAs, because PA-X accumulates in the nucleus (Hayashi et al., 2016; Khaperskyy et al., 2016) and its activity is linked to splicing (Gaucherand et al., 2019) (https://github.com/mgaglia81/PyDegradome).

The stringency of cut site definition can be adjusted based on two user-defined parameters: confidence level (cl) and multiplicative factor (mf) (see methods) (Gaglia et al., 2015). We used loose parameters (cl = 99%, mf = 2) on 16-22 million uniquely mapped reads per sample to find as many cut sites as possible, and identified 592 PA-X cut sites in 517 different genes that were shared across three biological replicates (**Figure 3.1D**, left, **Supplementary Table S3.1**). We will refer to this analysis as the "shared" approach (**Figure 3.1E).** These sites are likely true PA-X target sites, as a control analysis looking for fragments enriched in the PR8-PA(ΔX) infected cells (i.e. running PyDegradome with PR8-PA(ΔX) infected cells as the test sample and WT PR8 infected cells as the control sample; **Figure 3.1D**, right) only returned 37 sites. We considered whether the other non-shared sites we identified could also be PA-X sites, indicating a low specificity for this protein. However, we found that if we increased the stringency of the analysis by increasing the cutoff cl and mf values, we increased the percentage overlap (**Figure 3.3A**). This result suggests that at least some of the additional sites are due to either noise or limitations in the pipeline's ability to distinguish signal from noise. We chose to continue with the loose parameters to capture more sites. However, we were still surprised by the relatively low number of RNAs cut by PA-X compared to the >3,000 genes down-regulated by PA-X at steady state levels (Gaucherand et al., 2019). We thus wondered if this low number stemmed from limited sequencing depth. To increase the number of reads used as input for the PyDegradome

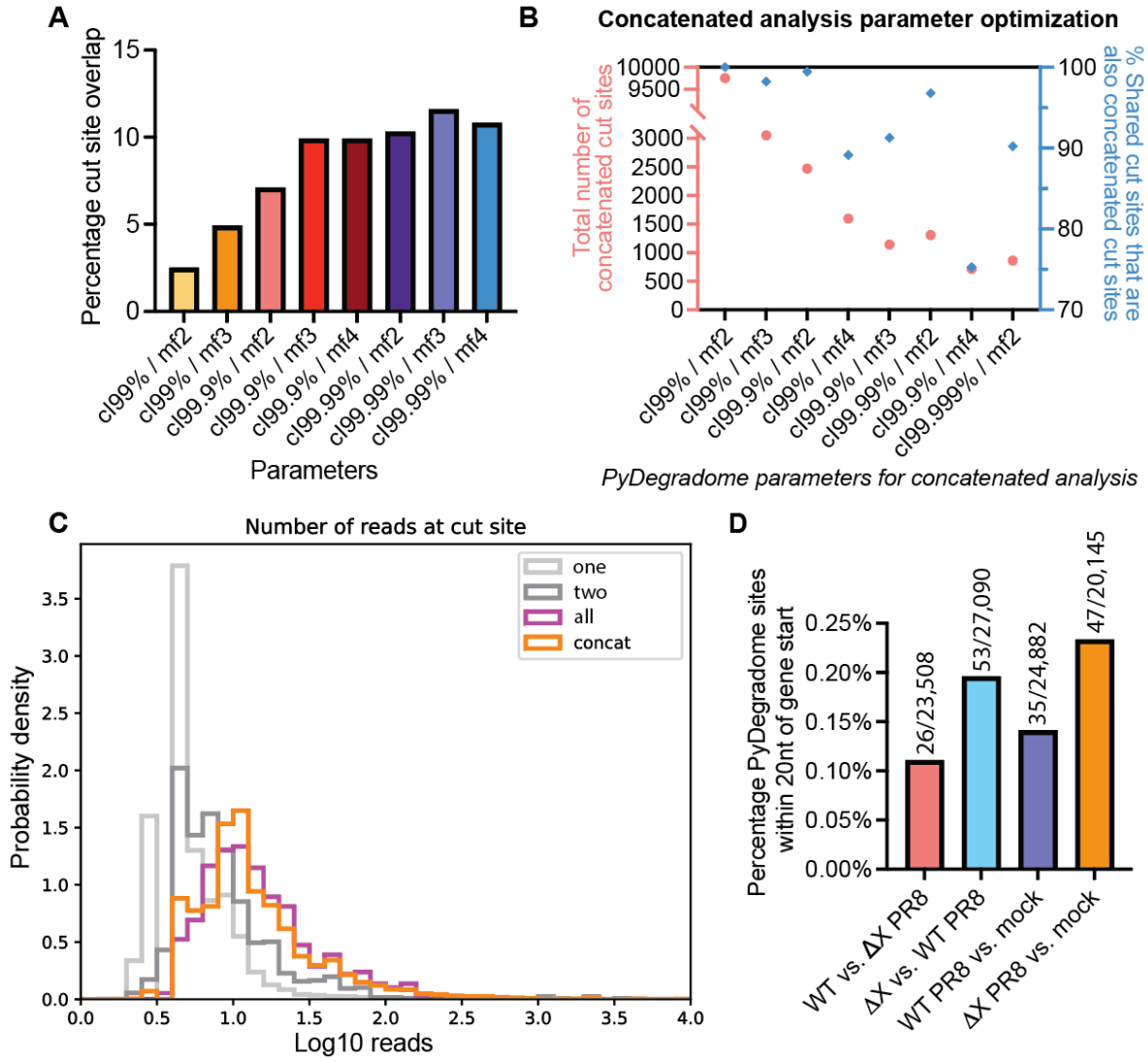

Figure 3.3: Optimization of parameters for the identification of PA-X cut sites transcriptome-wide and analysis of output characteristics.
(A-B) PyDegradome was run for each replicate (A) or on concatenated samples (B) on WT vs. PR8-PA(ΔX) samples using different parameters (see Methods) as indicated on the x axis. cl = confidence levels, mf = multiplicative factor. The window of analysis was 4 nucleotides in all cases. In A, the percentage of cut sites shared between each replicate was plotted for each set of parameters. In B, 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. (C) Histogram of the number of reads at the cut site for sites identified by PyDegradome when analyzing WT vs. PR8-PA(ΔX) samples in one (light grey), two (dark grey), or all three replicates ("all", magenta) or through the concatenated analysis ("concat", orange). (D) Percentage of sites identified by PyDegradome that are located within the first 20 nucleotides of the 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.

pipeline, we analyzed the replicates together by concatenating the read files ("concatenated" approach, **Figure 3.1E**). While we lose information on the replicability of the cut sites, this method is equally valid from a statistical perspective and has more power. We optimized the parameters by maximizing both the total number of concatenated cut sites detected (**Figure 3.3B**, red circles) and the percentage of the 592 shared cut sites identified in the concatenated analysis (**Figure 3.3B**, blue diamonds). Using cl = 99.99% and mf = 2, we identified 1361 cut sites in 1309 genes in WT vs. PR8-PA(ΔX) infected cells (**Supplementary Table S3.1**), but only 345 cut sites in the control comparison PR8-PA(ΔX) vs. WT PR8.

As mentioned before, it is possible that the limited numbers of sites we identified was due to the intensity of the signal and the signal-to-noise ratio, i.e., in practical terms, the number of reads mapping to the cut site in test vs. control samples. To test this possibility, we plotted the distribution of the read count at the cut site for sites that were identified by one, two or three individual replicates, or by the concatenated approach. The cut sites identified in three replicates or by the concatenated approach had higher reads counts at the cut site, and the ones identified in one replicate only had lower counts (**Figure 3.3C**). This analysis suggests we may be only identifying the most robust cut sites and that at least some of the non-shared sites may also be real, but that the signal is too low in some replicates to distinguish from random degradation also present in the absence of PA-X. Considering this result, to determine how much we may benefit from additional sequencing depth, we reanalyzed subsets of our data with the optimal parameters for the concatenated approach and plotted the number of cut sites identified by PyDegradome against the number of reads used in the analysis. Increasing the number of reads by re-sequencing our libraries and/or concatenating the read files led to an increase in the number of PA-X cut sites identified (**Figure 3.1F**, orange squares). However, the number of cut sites plateaued around 100 million reads. In contrast, the

number of fragments enriched in the PR8-PA(ΔX) infected cells did not increase with the number of reads (**Figure 3.1F**, purple circles). This suggests that we can use the current analysis without further increasing the sequencing depth, and that we are detecting specific cut sites.

Because the endonuclease domain is shared by PA, we considered the possibility that some of the sites we identified are PA-generated rather than PA-X-generated. However, we detected similar levels of PA protein in WT PR8 and PR8-PA(ΔX) infected cells, consistent with the literature that suggests that PA protein levels are the same or even higher in cells infected with PA-X-deficient viruses (Gao et al., 2015c, 2015b; Gong et al., 2017; Jagger et al., 2012; Khaperskyy et al., 2016; Rigby et al., 2019). Our PyDegradome pipeline comparing WT and PR8-PA(ΔX) samples would thus remove PA cut sites. Moreover, we found that very few (<0.5%) of the sites identified by PyDegradome map to the first 20 nucleotides of the transcripts, where PA cuts in the context of FluPol (**Figure 3.3D**). In addition, this fraction is slightly higher among sites enriched in the PR8-PA(ΔX)-infected cells vs. mock or WT PR8-infected cells (**Figure 3.3D).** This analysis suggests that the cut sites we identified in WT PR8-infected cells are unlikely to be generated by PA.

Given these considerations, we proceeded to further analyze the 1361 cut sites from the concatenated analysis (cl=99.99%/mf=2 parameters on all the reads), which give us a better overview of the PA-X cut site locations transcriptome-wide, as well as 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 can inform us on the preferred characteristics of PA-X cut sites (**Supplementary Table S3.1**).

3.2.2. PA-X cleavage is driven by RNA sequences

To confirm that PyDegradome identified true PA-X cut sites, we performed classical 5' RACE (**Figure 3.4A**) for 12 sites on RNA from Xrn1 knock-out A549 cells infected with WT or PR8-PA(∆X), or mock infected, and validated 11, including the 6 examples shown in **Figures 3.4B** and **3.5A**. For this assay, we concluded that the RNA was cut at the predicted location if the RACE PCR amplified a product of the correct size (i.e. the distance between the gene-specific primer and the primer in the ligated adaptor, **Figure 3.4A,** blue arrows) only in the WT PR8 infected samples. PCR products were also sequenced to confirm the cut site identification. Of note, additional PCR bands were apparent, likely due to basal RNA degradation in Xrn1 knock out cells, but were not specific to the WT PR8 infected cells. We also confirmed by RT-qPCR that WT and PR8-PA(∆X) samples were similarly infected (**Figure 3.5B**). To test whether the sequences around the predicted cut sites were sufficient to drive PA-X cleavage, we introduced stretches of 99 base pairs (bp) surrounding PA-X cut sites into a luciferase reporter that contains an intron and is down-regulated by PA-X (Gaucherand et al., 2019; Younis et al., 2010) (**Figure 3.4C**, constructs 3 and 4). We had previously found an "endogenous" cut site in this mRNA (**Figure 3.4C**, construct 1, orange rectangle). We transfected this reporter with or without PR8 PA-X into human embryonic kidney (HEK) 293T cells expressing inducible shRNA against Xrn1 (ishXrn1), which were pre-treated with doxycycline to knock down Xrn1 (**Figure 3.5C**). As negative and positive controls, we used the original luciferase reporter with no inserted sequence (**Figure 3.4C**, construct 1) and a construct in which we had duplicated the 99 bp sequence around the original luciferase cut site (**Figure 3.4C**, construct 2). Introducing extra sequences into the luciferase reporter did not interfere with its expression and its down-regulation by PA-X (**Figure 3.5D**), but added a new PA-X cut site (red dotted arrows, **Figure 3.4D**, **Figure 3.5E**). For the negative control, we detected fragments from the original cut site

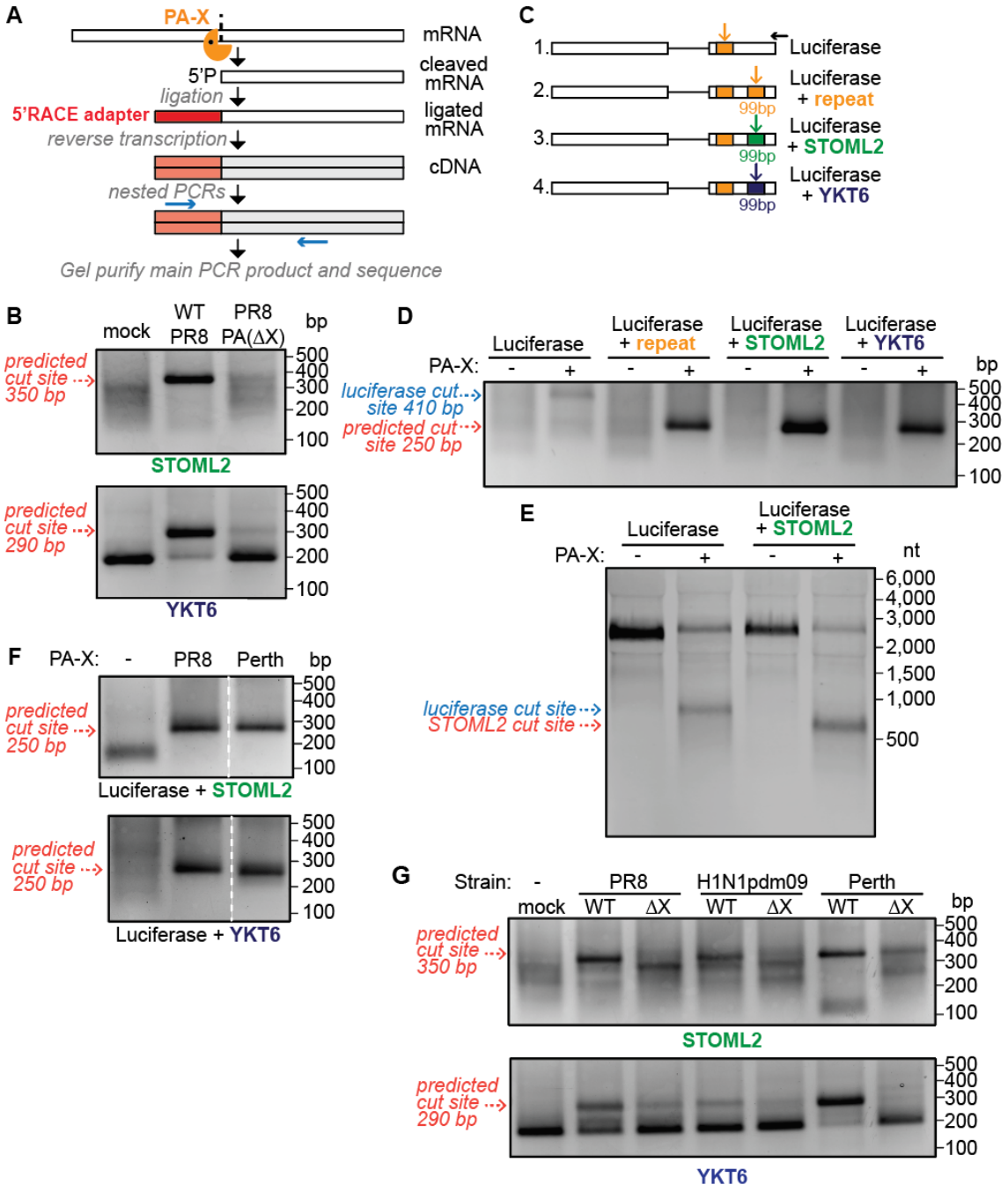

Figure 3.4: Sequences around the cut sites drive cleavage by PA-X.
(A) Schematic diagram of 5' RACE workflow. (B) Xrn1 knock out A549 cells were infected with WT or PR8-PA(ΔX), or mock infected for 16 hours before RNA extraction. 5'RACE was then performed using primers specific for STOML2 or YKT6, ~250-300 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 ~200bp. 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) or Perth influenza strain (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, and images are representative of 3 experiments.

instead (blue dotted arrows, **Figure 3.4D**, **Figure 3.5E).** We also confirmed these results using a second experimental approach, northern blotting (**Figure 3.4E**). Importantly, while RACE could in principle selectively amplify some fragments, northern blotting should allow us to see all the fragments generated. Therefore, the results in **Figure 3.4E** suggest that the there is only one major cut site in the luciferase and luciferase + STOML2 reporters, supporting our model that PA-X cleaves RNAs at discrete sites. These results further validate our cut site identification pipeline, and suggest that the sequences around the cut sites are sufficient to drive cleavage by PA-X.

Importantly, the sequences we identified are specific to PA-X activity, as no cleavage is detected upon expression of the PA-X D108A catalytic mutant or a PA(fs) mutant that makes PA but not PA-X due to a decrease in frameshifting event (Gaucherand et al., 2019) (**Figure 3.6A**). Moreover, we did not detect the same fragments upon expression of other host shutoff endoRNases (**Figure 3.6B**). Conversely, these sequences elicited cleavage by PA-X from the influenza A/Perth/16/2009 (H3N2) (henceforth "Perth") virus, representative of currently circulating H3N2 influenza A strains (**Figure 3.4F**, **Figure 3.6C**). We also saw conserved cleavage specificity during infection with the same Perth virus, as well as influenza A/Tennessee/1-560/2009 (H1N1) (henceforth "H1N1pdm09" for 2009 pandemic H1N1) virus, representative of currently circulating H1N1 strains (**Figure 3.4G**). The PA-X proteins from these three influenza A strains are highly

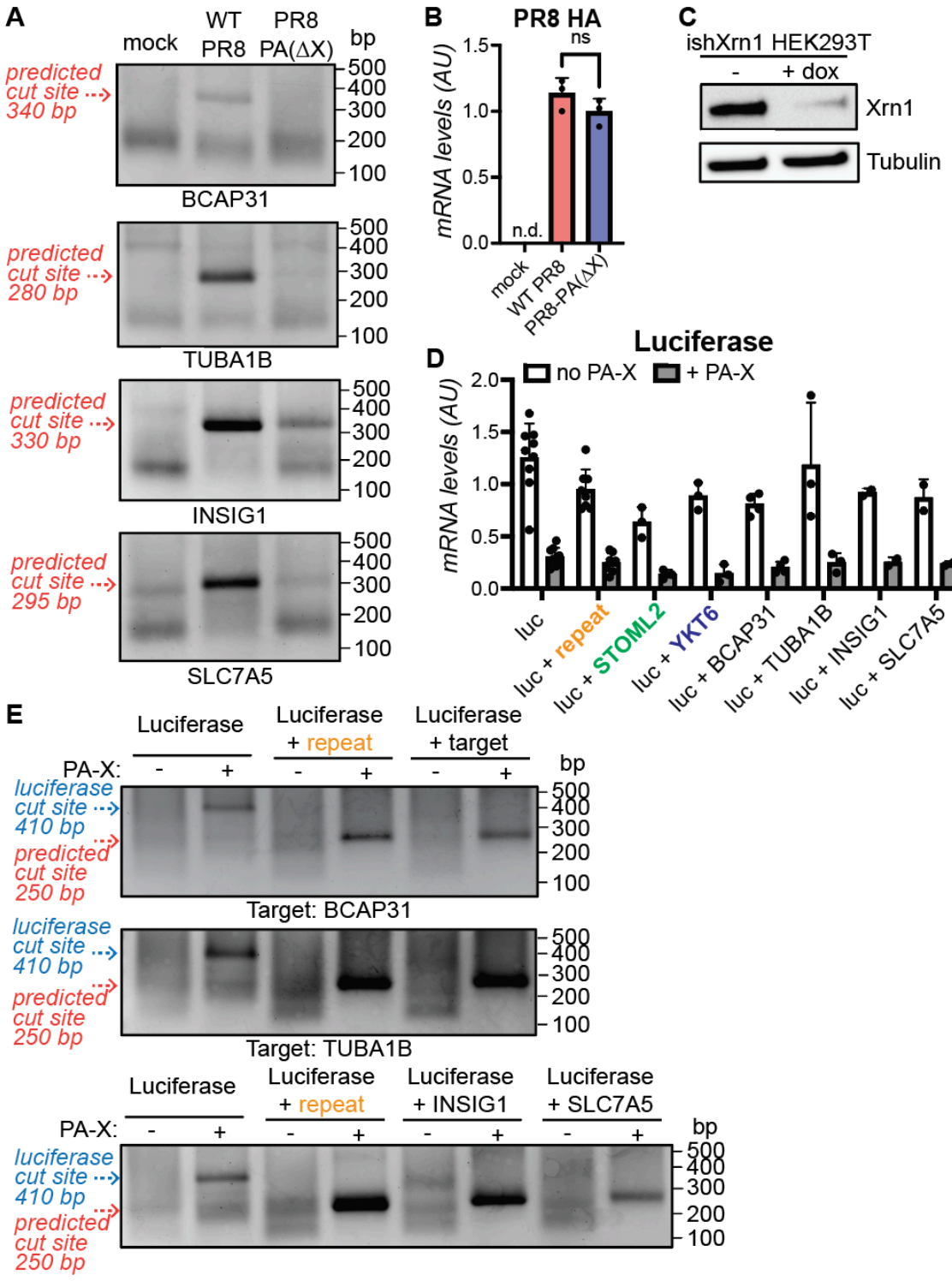

Figure 3.5: Further validation of PA-X cut sites identified by PyDegradome. *(A-B) Xrn1 knock out A549 cells were infected with WT or PR8-PA(ΔX), or mock infected. 5' RACE was then performed using primers specific for BCAP31, TUBA1B, INSIG1 or SLC7A5, positioned ~ 200-300 nucleotides downstream of the predicted cut sites. (A) 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. (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. (C) HEK293T ishXrn1 cells were treated with no drug or doxycycline for 3-4 days, then protein lysates were collected and probed with antibodies against Xrn1, or b-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 the indicated luciferase reporters, and where indicated, with PR8 PA-X. (D) RNA was extracted and levels were quantified by qRT-PCR, normalized to 18S and plotted as mean ± standard deviation. AU, arbitrary units. n ≥ 2. (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 the 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, and images are representative of 3 experiments or 2 experiments for the luciferase + INSIG1 and + SLC7A5 reporters.

conserved (**Figure 3.6D**), but this is largely driven by the conservation of the N-terminal domain that is shared with PA. Nonetheless, they are a good representation of the existing diversity of PA-X proteins across the influenza A strains that infect humans. They represent three different subtypes of influenza A viruses, and have different X-ORF lengths (41 aa for H1N1pdm09 *vs.* 61 aa for PR8 and Perth) and half-lives (Levene et al., 2021). These experiments confirm that 5' RACE-seq and PyDegradome identified bona fide PA-X cut sites transcriptome-wide and indicate that PA-X cleavage across multiple strains is sequence specific.

3.2.3. PA-X preferentially cleaves RNAs at a specific sequence and structure

Because the sequences surrounding PA-X cut sites were sufficient to trigger cleavage, we wondered whether conserved RNA features guided cleavage. Indeed, we saw a depletion in adenosines and an enrichment in GCUG at the cut site, and a modestly enriched guanidine stretch downstream of the cut site (**Figure 3.7A** 592 sites from shared analysis, **Figure 3.8A** 1361 cut sites from concatenated analysis). These features were specific to PA-X, as we found different patterns when analyzing fragments specific to PR8-PA(∆X) infected cells, our control comparison (**Figure 3.7B**, **Figure**

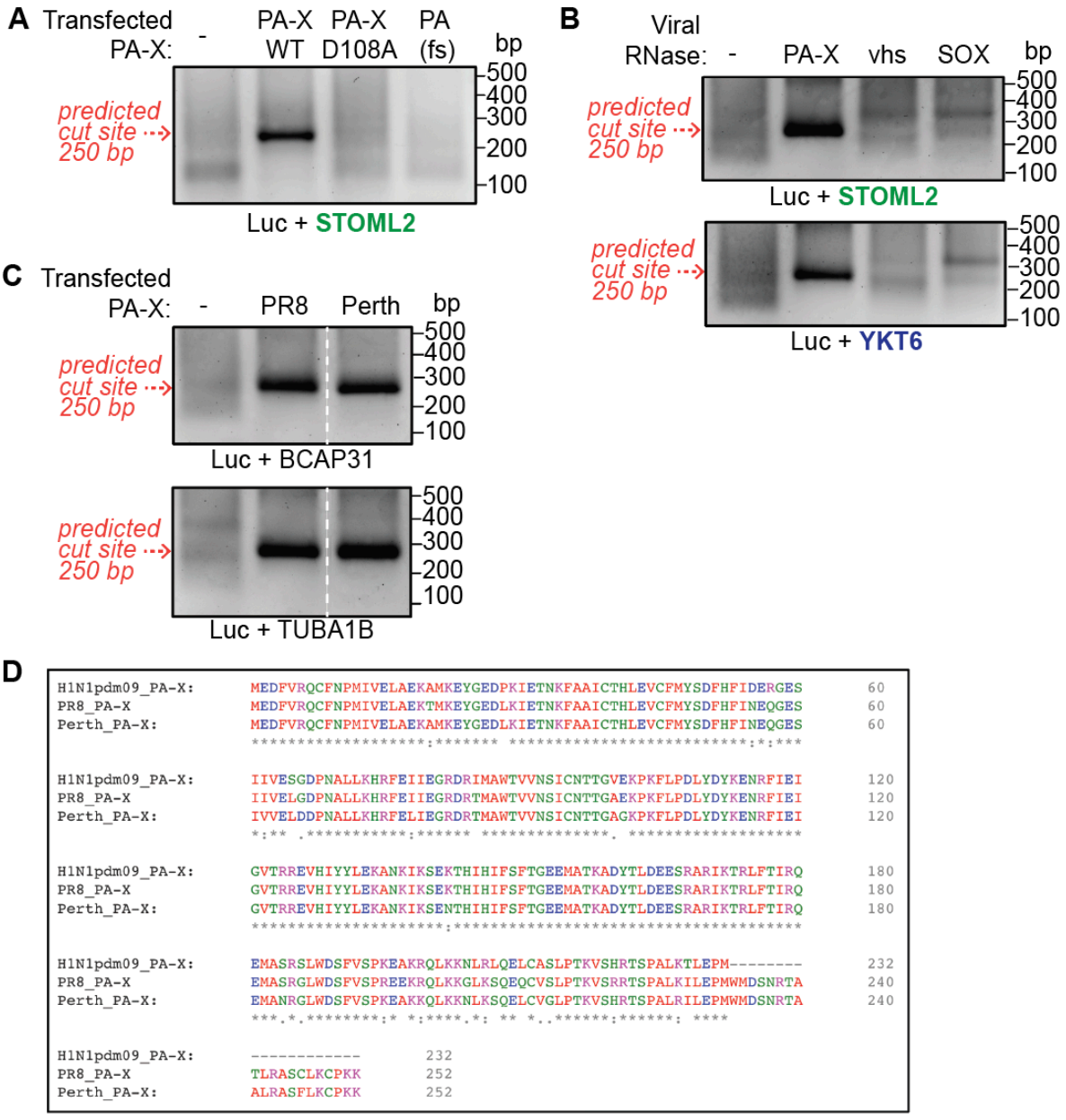

Figure 3.6: The cut sites identified by PyDegradome and specific to PA-X and conserved across multiple influenza strains.

(A-C) HEK293T ishXrn1 cells were treated with doxycycline for 3-4 days to induce knock down of Xrn1, then transfected with the indicated luciferase reporters, 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 (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 strains (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 or PR8-PA(ΔX), WT H1N1pdm09 or H1N1pdm09-PA(ΔX), WT Perth or Perth-PA(ΔX), 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 Omega (Sievers et al., 2011).

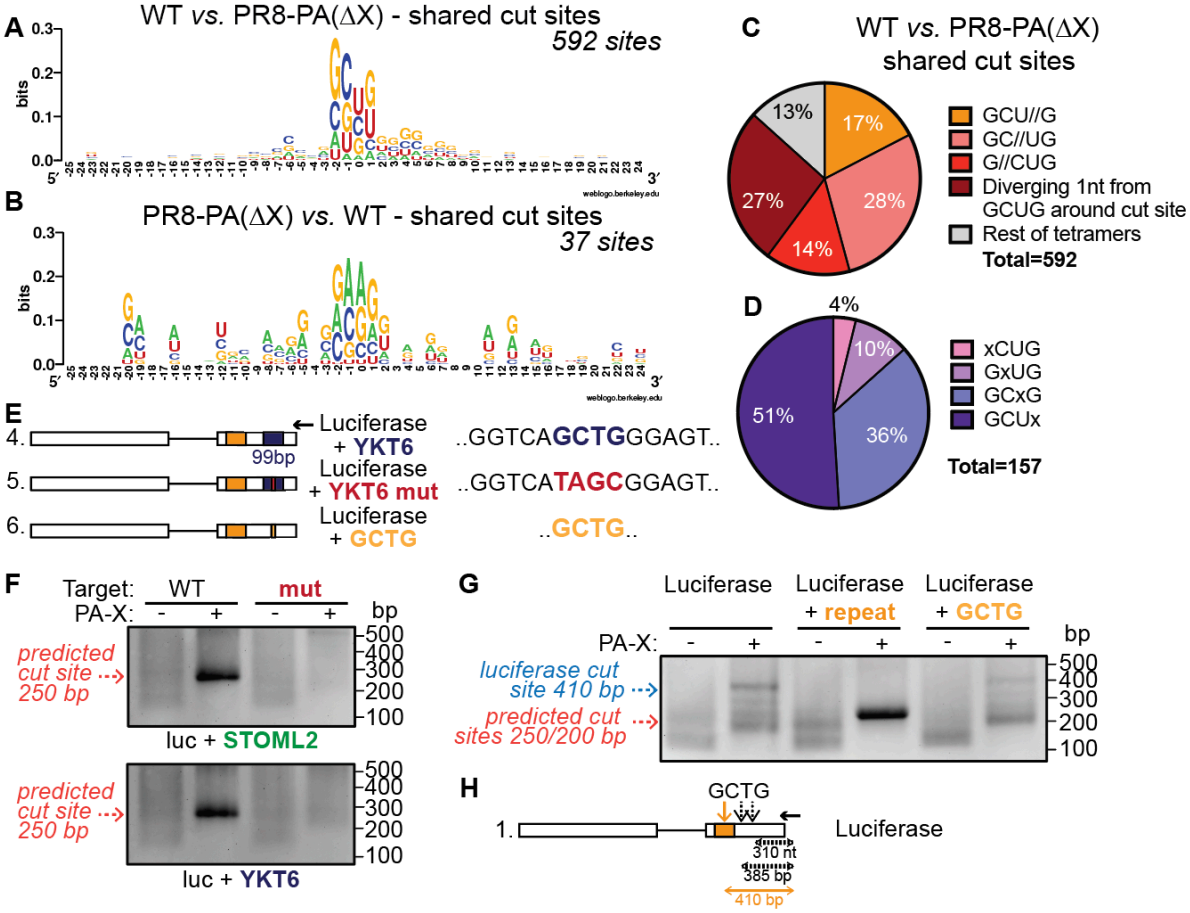

Figure 3.7: PA-X preferentially cleaves RNAs at GCUG tetramers.
(A-B) WebLogo (Crooks et al., 2004) representation of base enrichment around PA-X cut sites (test: WT PR8 vs. control: PR8-PA(ΔX), A) or around control sites enriched in the PR8-PA(ΔX) sample (test: PR8-PA(ΔX) vs. control: WT PR8, B) for 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 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) Diagram of the luciferase reporters tested. The black arrow indicates the position of the 5' RACE PCR reverse primer used. (F-G) HEK293T ishXrn1cells were treated with doxycycline for 3-4 days to induce knock down of Xrn1, then transfected with the indicated luciferase reporters, with or without PR8 PA-X. RNA was extracted 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 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.

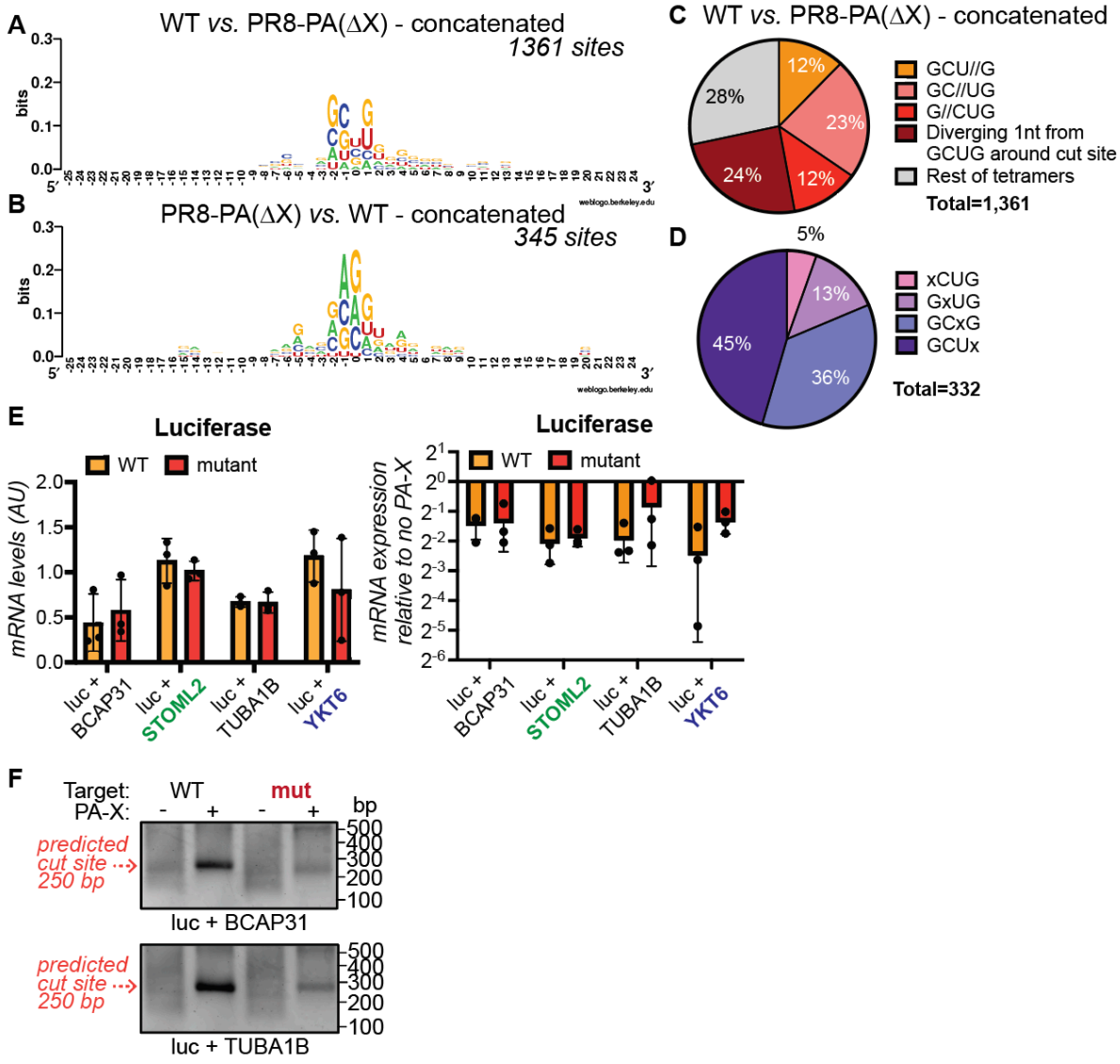

Figure 3.8: PA-X preferentially cleaves RNA at GCUG tetramers within a hairpin loop structure based on the concatenated cut site analysis and in infected cells. *(A-B) WebLogo (Crooks et al., 2004) representation of base enrichment around PA-X cut sites (WT vs. PR8-PA(ΔX), A) or around control sites enriched in the PR8-PA(ΔX) sample (PR8-PA(ΔX) vs. WT, 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-G) HEK293T ishXrn1cells were treated with doxycycline for 3-4 days to induce knock down of Xrn1, then transfected with the indicated luciferase reporters, with or without PR8 PA-X. (E) RNA was extracted and 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 of PA-X normalized to the levels in the absence of PA-X. AU, arbitrary units. n = 3. (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.

3.8B). Interestingly, most PA-X cut sites contained GCUG or a tetramer with one nucleotide difference from GCUG, but the location of the cut within the tetramer varied (**Figure 3.7C**, **Figure 3.8C**). Moreover, the first two nucleotides in the GCUG tetramer were the most important for PA-X cleavage (**Figure 3.7D**, **Figure 3.8D**).

To determine the importance of the GCUG tetramer, we mutated the cut site GCTG to TAGC in the luciferase reporters with the 99 bp insertions from BCAP31, STOML2, TUBA1B and YKT6. (**Figure 3.7E**, construct 5). This mutation did not impact expression of the reporters or their down-regulation by PA-X **(Figure 3.8E)** but prevented efficient cleavage by PA-X (**Figure 3.7F**, **Figure 3.8F**). However, introduction of an isolated GCTG (**Figure 3.7E**, construct 6) did not result in efficient cleavage (**Figure 3.7G**). These results suggest that the GCUG tetramer is necessary but not sufficient for PA-X cleavage, consistent with PA-X not cleaving at every GCUG. For example, there are GCUG tetramers 17 and 92 nucleotides downstream of the original cut site in the luciferase mRNA (**Figure 3.7H**, black vertical dotted arrows), but we do not see evidence of robust cleavage by PA-X at these locations.

Since GCUG alone was not sufficient, we investigated how much of the surrounding sequence was needed to drive PA-X cleavage (**Figure 3.9A**). The minimal sequence required for PA-X cleavage varied depending on the gene tested: whereas inserting 15

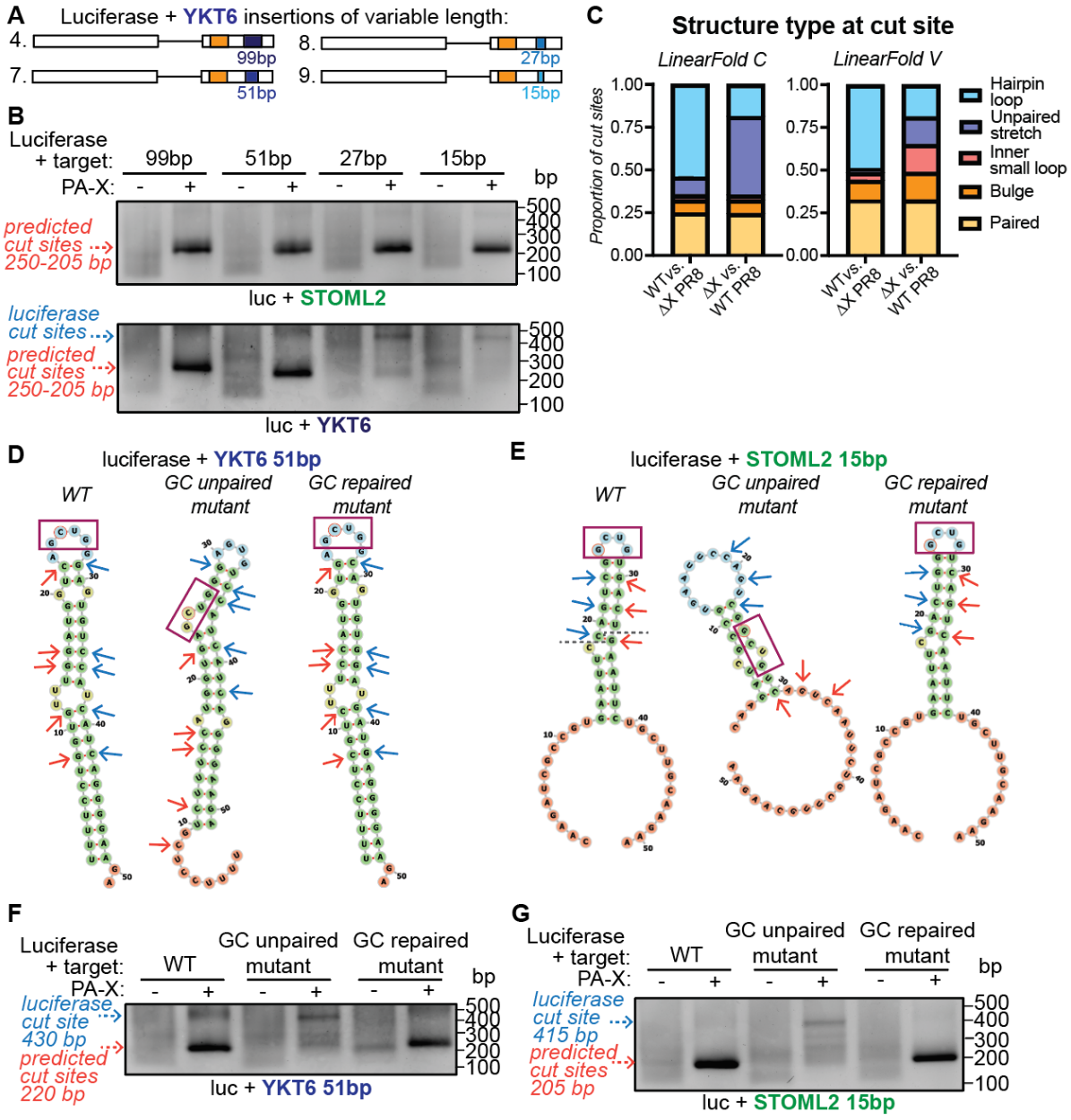

Figure 3.9: PA-X preferentially cleaves RNA within hairpin loop structures.
(A) Diagram of the luciferase reporters tested. (B, F, G) HEK293T ishXrn1cells were treated with doxycycline for 3-4 days to induce knock down of Xrn1, then transfected with the indicated different luciferase reporters, with or without PR8 PA-X. RNA was extracted 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). 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 vs. ΔX, n = 592) or around control sites enriched in the PR8-PA(ΔX) sample (ΔX vs. WT, n =37) (shared cut sites

approach). Structures were predicted using the CONTRAfold v2.0 machine-learning model (LinearFold C) or the Vienna RNAfold thermodynamic model (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 the 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.

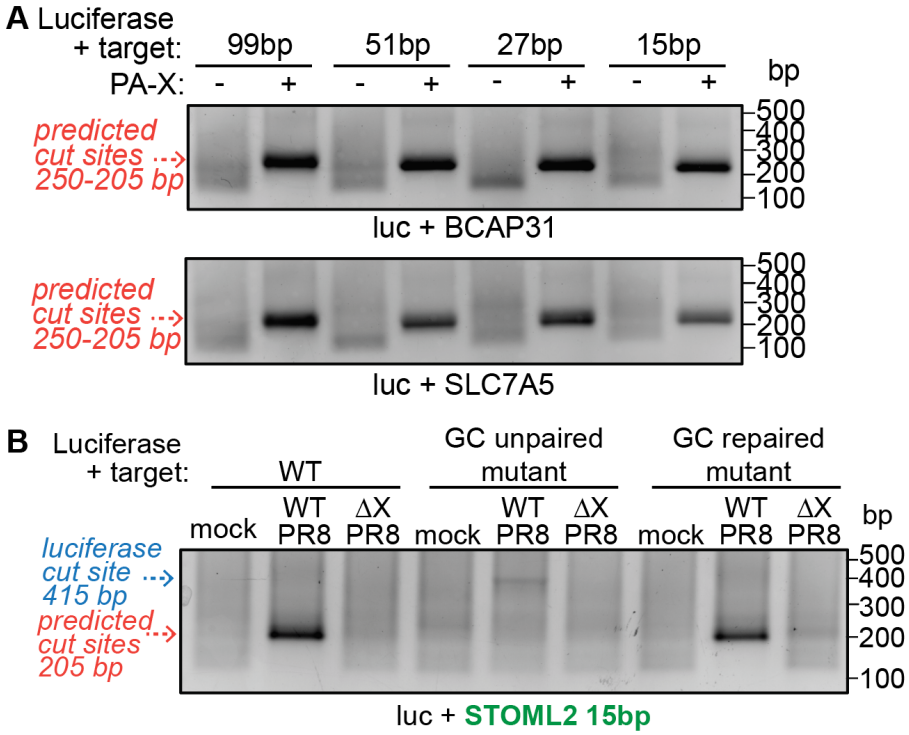

Figure 3.10: PA-X preferentially cleaves RNA within a hairpin loop structure in infected cells.

(A) HEK293T ishXrn1cells were treated with doxycycline for 3-4 days to induce knock down of Xrn1, then transfected with the indicated different luciferase reporters, 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 the indicated luciferase reporters. 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).

bp was sufficient for PA-X cleavage of the STOML2, BCAP31 and SLC7A5 cut sites, more than 27 bp were required for the YKT6 sequence (**Figure 3.9B**, **Figure 3.10A**). Since there was no clear sequence enrichment aside from GCUG (**Figure 3.7A**, **Figure 3.8A**), we hypothesized that an RNA secondary structure was also required for cleavage. We predicted the RNA structure around the shared PA-X cut sites using LinearFold (Huang et al., 2019), which provides structure predictions using either the CONTRAfold v2.0 machine-learning model (Do et al., 2006) (LinearFold C) or the Vienna RNAfold thermodynamic model (Lorenz et al., 2011; Mathews et al., 2004) (LinearFold V). According to both models, PA-X cut sites (but not control sites enriched in PR8-PA(ΔX) infected cells) were preferentially found in the loops of RNA hairpin structures (**Figure 3.9C**). To test whether the hairpin was required for cleavage, we manipulated the structures of the YKT6 51 bp and the STOML2 15 bp insertion constructs (**Figure 3.9A**). To disrupt the predicted hairpin structures, we mutated select C or G bases to break strong G-C bonds in the stems (**Figure 3.9D-E**, *WT*, red arrows). These mutations led to very different predicted structures, with the GCUG cut site in bulge or paired regions instead of in the loop (**Figure 3.9D-E**, *GC unpaired mutant*). As a control, we additionally mutated the complementary G or C bases to recreate the G-C bonds and repair the structures (**Figure 3.9D-E**, *GC repaired mutant*, blue arrows). Consistent with PA-X cut site structure predictions, disrupting the hairpin prevented efficient PA-X cleavage within the YKT6 and STOML2 sequences. The reporter was instead cleaved at the original luciferase cut site location (**Figure 3.9F-G**, blue dotted arrow). Conversely, repairing the hairpin structure restored efficient cleavage by PA-X within the inserted fragment despite the mutated surrounding sequence (**Figure 3.9F-G**, red dotted arrow). Selective cleavage of GCUG sequences in a hairpin loop was also seen using the same strategy in infected cells, confirming the importance of a hairpin

structure for PA-X cleavage during infection (**Figure 3.10B**). Collectively these experiments show that PA-X preferentially cleaves RNAs at GCUG tetramers located within hairpin loops and reveal a new layer of RNA targeting specificity by PA-X.

3.2.4. PA-X preferentially cleaves RNAs within exons

Since PA-X activity is linked to splicing (Gaucherand et al., 2019), we wondered whether PA-X targets pre-mRNAs or mature mRNAs. Interestingly, cut sites were almost exclusively found within exons (**Figure 3.11A**) despite 12-16% of reads mapping to introns (**Figure 3.11B**, $p < 0.001$, Chi-Square Test for Goodness of Fit with degree of freedom 1). In contrast, 18% of PR8-PA(ΔX) specific fragments from the concatenated analysis mapped to introns (**Figure 3.12A**, bottom). While only 3% of PR8-PA(ΔX) fragments from the shared analysis mapped to introns, this may stem from the very low number of sites (37) found with this method (**Figure 3.12A**, top). Nonetheless, these results suggest that PA-X preferentially cleaves RNA within exons, although this may occur before or after splicing.

Our 5' RACE validation experiments (**Figure 3.4B**, **Figure 3.5A**) suggested that PA-X cleaves after splicing, as we used primers that spanned exon-exon junctions and only saw PCR amplification of spliced fragments (**Figure 3.11C**, **Figure 3.12B**, dark purple arrows). However, because PCR favors the amplification of smaller fragments, we could have preferentially amplified smaller spliced fragments over longer non-spliced ones. We thus repeated the 5' RACE PCRs using reverse primers in the intron 3' of the cut sites, which should amplify cleaved unspliced pre-mRNAs (**Figure 3.11C**, **Figure 3.12B**, light purple primers). However, we did not detect fragments of expected sizes specifically in WT PR8 infected cells, but instead saw background fragments, including some that mapped to the exon/intron junction (**Figure 3.11C**, **Figure 3.12B**, white arrowheads). This result suggests that PA-X cleaves RNAs after splicing of the neighboring introns.

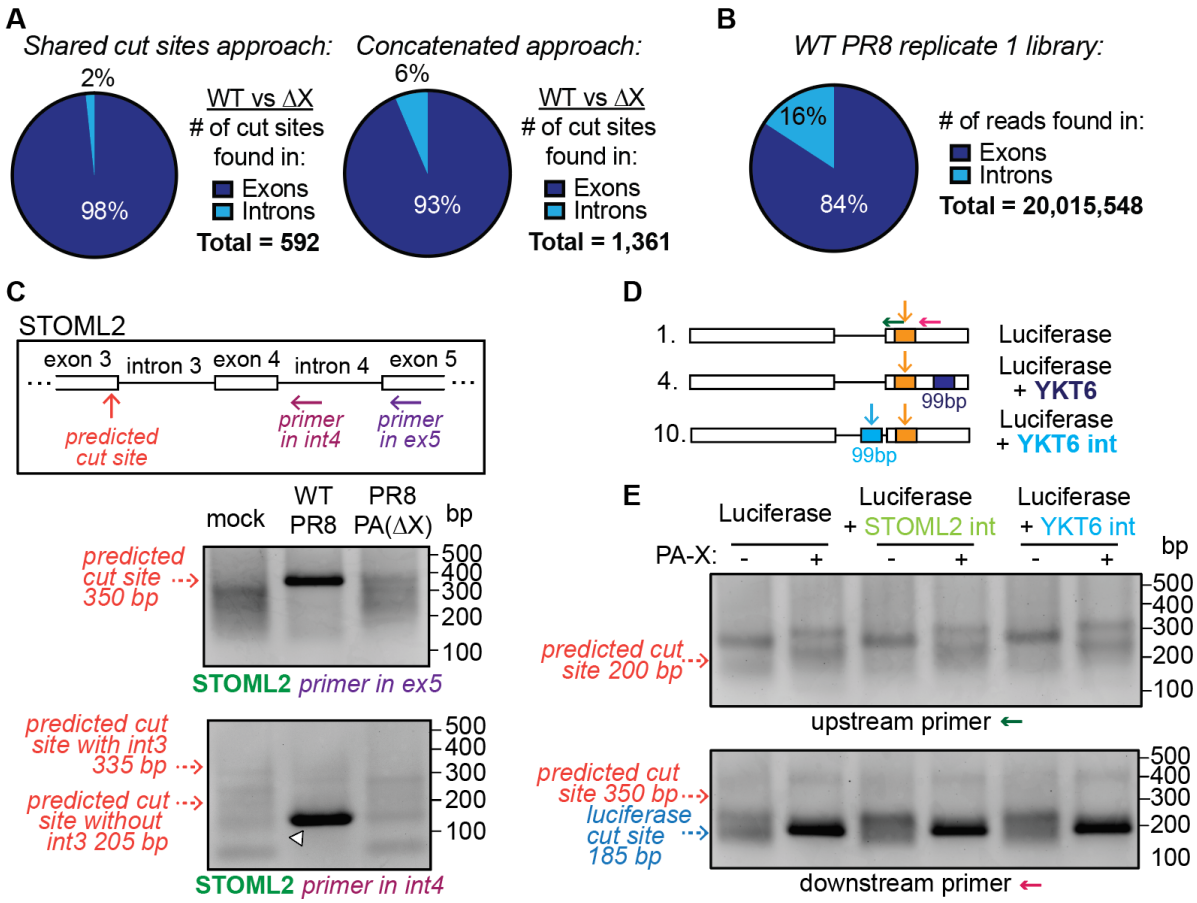

Figure 3.11: PA-X preferentially cleaves RNAs within exons.

(A-B) Percentage of PA-X cut sites found within introns or exons, for sites identified by PyDegradome using the shared cut sites approach (A, left) or the concatenated approach (A, 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. (C) Box: diagram of the STOML2 gene and positions of the reverse primers for 5' RACE. Xrn1 knock out A549 cells were infected with WT or PR8-PA(ΔX), or mock infected. 5'RACE was performed using the STOML2 primers indicated in the diagram, and the PCR products were run on an agarose gel. The top gel is the same gel as Fig. 2B for comparison, i.e. using a reverse primer in the exon (dark purple primer), and the red dotted arrows indicate size of fragment originating from previously validated cut site. The bottom gel shows products obtained using a reverse primer in the intron (light purple primer), and red dotted arrows indicate the predicted size of PCR products that would appear if PA-X cleaved unspliced pre-mRNAs. White arrowhead indicates fragment mapped to exon/intron 4 junction. Gel images are representative of 3 experiments. (D) Diagram of the luciferase reporters tested. Dark green and magenta arrows indicate positions of 5' RACE PCR reverse primers, vertical arrows indicate location of predicted cut sites. (E) HEK293T ishXrn1 cells were treated with doxycycline for 3-4 days to induce knock down of Xrn1, then transfected with the indicated different luciferase reporters, with or without PR8 PA-X. RNA was extracted and used to run 5' RACE. PCR products were separated on an agarose gel. The top gel represents products obtained using dark green arrow primer from (D) and the bottom gel using magenta arrow primer from (D). Blue dotted arrow indicates size of PCR products originating from the original luciferase cut site, red

dotted arrows indicate the predicted sizes of PCR products that would originate from the sequence inserted in the introns. Gel images representative of 3 experiments.

As an additional test, we cloned the previously validated STOML2 and YKT6 99 bp cut site sequences inside the intron of the luciferase reporter (**Figure 3.11D**, construct 10). Introducing the sequence did not significantly affect mRNA expression or splicing (**Figure 3.12C**). However, 5' RACE did not amplify bands that would correspond to cuts within the intron (**Figure 3.11E**, top gel, see dark green arrow in **Figure 3.11D** for position of the primer). The closest band at ~250 bp was also present in the luciferase construct without the STOML2 or YKT6 cut site sequences. Like in other cases where no additional cut site was introduced (**Figures 3.7G**, **3.9B**, **3.9F-G**), we could still detect the original luciferase cut site using a reverse primer positioned further downstream (**Figure 3.11E**, bottom gel, see magenta arrow in **Figure 3.11D** for the position of the primer). We also mutated the original luciferase cut site GCUG sequence to UAGC to reduce cleavage at this site (**Figure 3.12D**, construct 11), then ran 5' RACE using the primer indicated by the magenta arrow. However, we still saw no PCR product matching a cut site in the luciferase intron (**Figure 3.12E**). These results suggest that the STOML2 and YKT6 99 bp sequences are sufficient for PA-X cleavage only when located within an exon. Overall, these experiments are consistent with PA-X cutting mRNAs within exonic sequences after splicing of at least the neighboring intron.

3.2.5. GCUG tetramers are more abundant in the human than influenza transcriptome

Finally, we wondered why PA-X has evolved these preferred cleavage characteristics, and if this could lead to targeting of specific RNAs. Interestingly, GCUG is one of the most abundant tetramers in the human transcriptome (**Figure 3.13A**) and is found at twice the frequency in the transcriptome compared to the human genome,

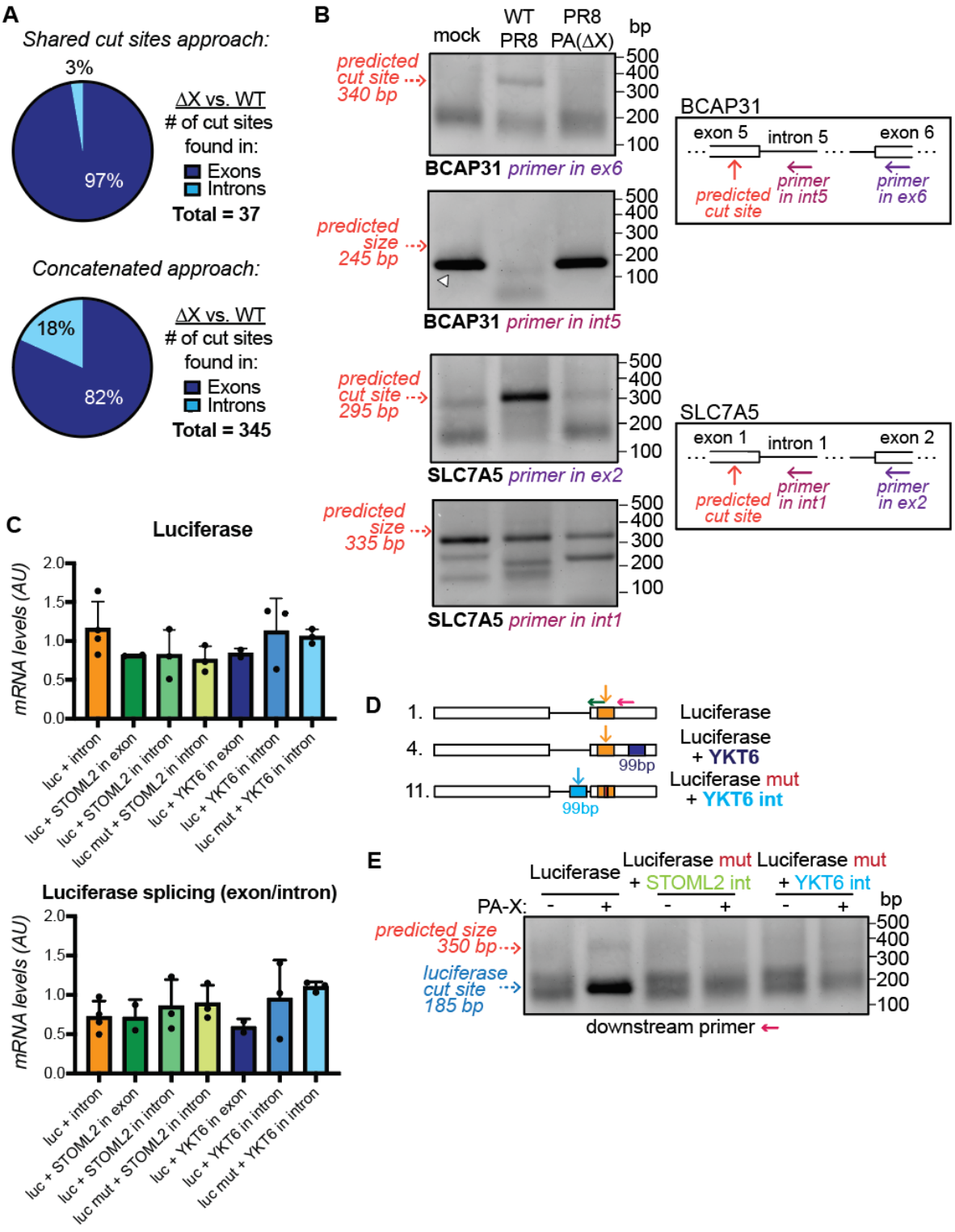

Figure 3.12: PA-X preferentially cleaves RNAs within exons.
(A) Percentage of PR8-PA(ΔX) specific fragments found within introns or exons, for sites identified by PyDegradome using the shared cut sites approach (top) or the concatenated approach (bottom). (B) Boxes: diagrams of the BCAP31 and SLC7A5

genes and positions of the reverse primers for 5' RACE. Xrn1 knock out A549 cells were infected with WT or PR8-PA(∆X), or mock infected. 5' RACE was performed using the BCAP31 and SLC7A5 primers indicated in the diagram, and the PCR products were run on an agarose gel. Top gels are the same gels as Extended Data Fig. 3A for comparison, i.e. using reverse primers in exons (dark purple primers), and 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 fragment mapped to exon/intron junction. Gel images are representative of 3 experiments. (D) Diagram of the luciferase reporters tested. Dark green and magenta arrows indicate positions of 5' RACE PCR reverse primers, vertical arrows indicate location of predicted cut sites. (C, E) HEK293T ishXrn1 cells were treated with doxycycline for 3-4 days to induce knock down of Xrn1, then transfected with the indicated luciferase reporters, with or without PR8 PA-X. (C) 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 ≥ 2. Top gel is using both primers in exons, bottom gel is using one primer in an exon and one primer in the intron. (E) RNA was also used to run 5' RACE. PCR products were obtained using the magenta arrow primer from (D), then were separated on an agarose gel. Blue dotted arrow indicates size of PCR products originating from the original luciferase cut site, red dotted arrow indicates the predicted size of PCR products that would originate from the sequence inserted in the introns. Gel images representative of 3 experiments.

which includes the intronic sequences (**Figure 3.14A**). Given that mRNAs cut by PA-X do tend to have more GCUG tetramers (**Figure 3.14B**), PA-X may have evolved these cleavage characteristics to preferentially target the host transcriptome.

It is unclear if GCUG specificity has any role in targeting specific groups of host mRNAs. For example, the mRNAs for the antiviral interferon stimulated genes (ISGs) do not contain more GCUG tetramers than other mRNAs (**Figure 3.14C**). However, GCUG specificity may contribute to virus vs. host discrimination. Indeed, whereas GCUG is abundant and overrepresented in the human transcriptome, it is far less abundant in the mRNAs of three influenza A virus strains, PR8 H1N1, H1N1pdm09 and Perth H3N2 (**Figure 3.14A, Figure 3.13B**) and is underrepresented relative to other tetramers (**Figure 3.13C-E**). Interestingly, this difference in GCUG levels between cellular and viral transcriptome is conserved across common influenza hosts such as pigs, chickens and ducks, and representative influenza A virus strains that were isolated from these animals

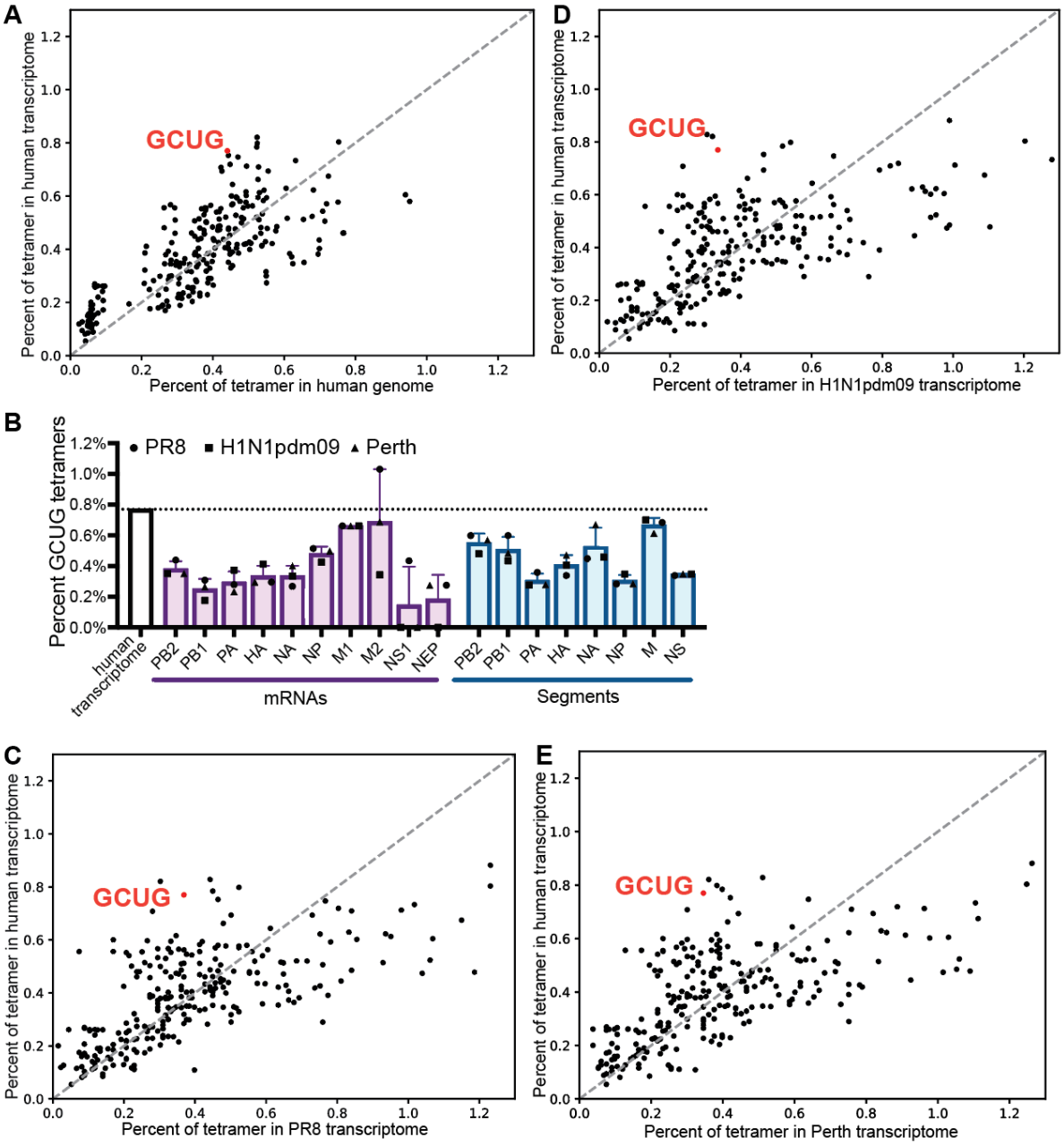

Figure 3.13: GCUG tetramers are enriched in the human transcriptome.
(A, C-E) The percentage of all possible tetramers in the human genome, human transcriptome or viral transcriptomes was calculated by counting the number of each tetramer and dividing it by the total number of tetramers in the sequence (i.e. length of the sequence minus 3). These percentages were then plotted to visualize which tetramers were most abundant in one genome/transcriptome compared to another. Each dot represents a specific tetramer, with the red dot representing the GCUG tetramer. (B) The percentage of GCUG tetramers was calculated for each influenza mRNA (i.e. the positive strand, purple), and for each influenza genomic RNA (i.e. negative strand, blue). Each symbol represents a different influenza strain.

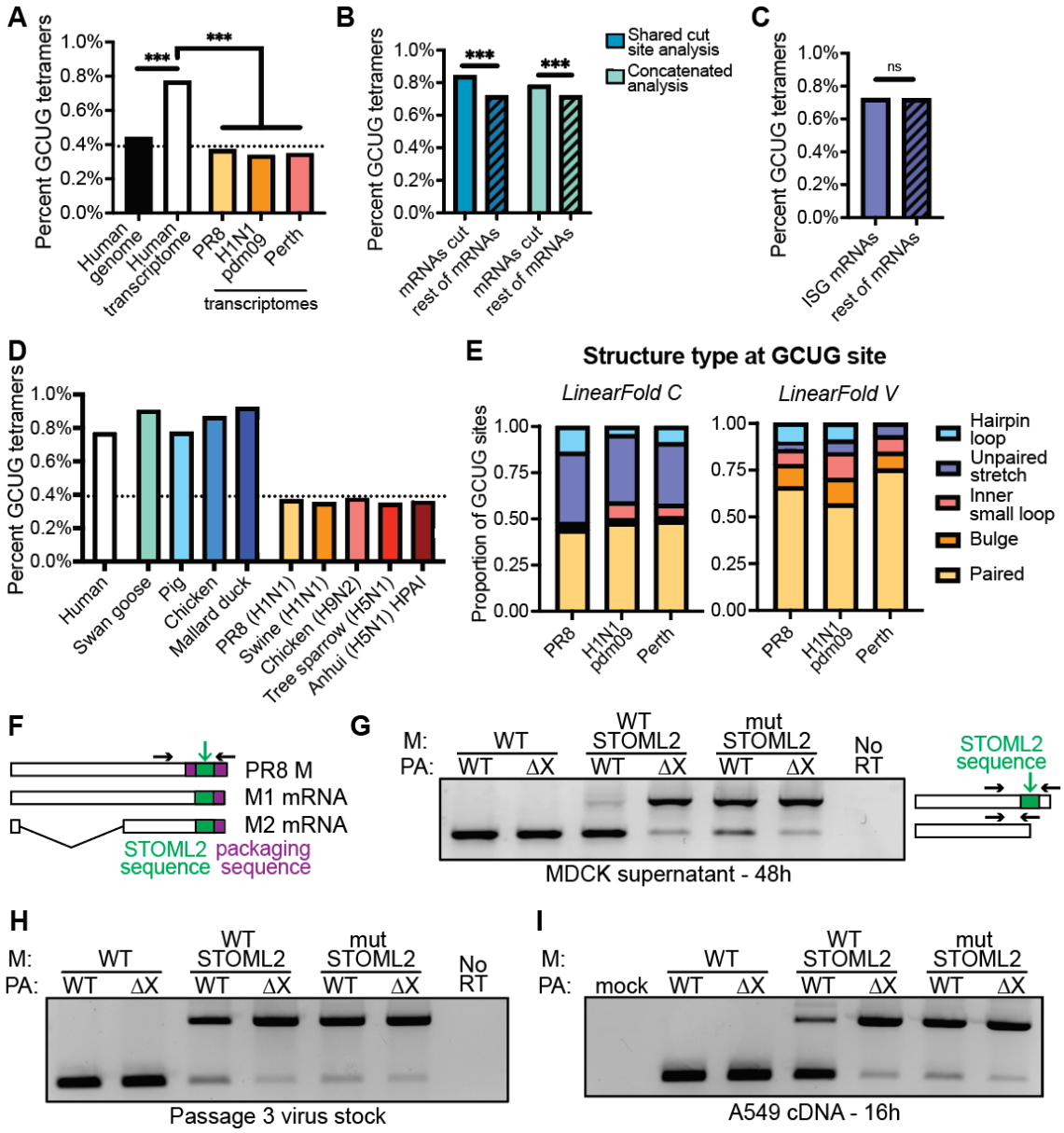

Figure 3.14: PA-X likely cleaves GCUG sequences to preferentially target host over viral mRNAs.

*(A-D) The percentage of GCUG tetramers in each indicated sequence was calculated by counting the number of GCUG tetramers and dividing it by the total number of tetramers in the sequence (i.e. length of the 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. The black dotted lines in (A) and (D) represent the average tetramer abundance (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. Structures were*

predicted using the CONTRAfold v2.0 machine-learning model (LinearFold C) or the Vienna RNAfold thermodynamic model (LinearFold V). (F) Diagram showing the location of the insertion of the STOML2 51 bp cut site sequence (green rectangle) inside the PR8 M segment right at the end of the M2 coding sequence. The M segment packaging signal (purple rectangle) was repeated after the STOML2 sequence to make sure the modified segment is packaged inside the virion. (G-I) Viral RNA was extracted from our viral stocks (G), 48-hour supernatants from MDCK cells from Figure 3.15C (H), or Xrn1 ko A549 cells 16 hours post-infection for viruses harboring no insert, inserted WT STOML2 or mutant STOML2 sequences, either in the WT PR8 or PR8-PA(∆X) background. The RNA was then reverse-transcribed to cDNA and PCR amplified using primers located on either side of the STOML2 sequence (black arrows in (F)) to visualize on an agarose gel whether the STOML2 sequence was retained. Gel images are representative of 2 experiments.

(**Figure 3.14D**). The differential GCUG representation is consistent with our finding that PA-X activity does not down-regulate influenza transcripts (Khaperskyy et al., 2016), and points to GCUG specificity as a way to distinguish host vs. viral mRNA. In support of this idea, we also found that the percentage of GCUG is higher in the negative RNA strand (i.e. the viral genomic RNA) than in the positive RNA strand (i.e. the viral mRNA) for almost all segments (**Figure 3.13B**). Additionally, LinearFold structure predictions suggested that GCUG tetramers in influenza transcripts are enriched in paired structures that may not be accessible to PA-X rather than hairpin loops (**Figure 3.14E**). Since a few GCUG tetramers in the viral transcriptome were predicted to be within hairpin loops, we tested whether these sequences were cut by PA-X. While we were able to detect faint bands that mapped close to the GCUG tetramers in the case of PR8 (**Figure 3.15A**), we were not able to detect cut sites for H1N1pdm09 (**Figure 3.15B**). Overall, these results suggest that PA-X may have evolved to target GCUG sequences in hairpin loops to avoid degrading influenza mRNAs while preferentially targeting the host transcriptome.

To test this idea experimentally, we inserted the STOML2 51 bp cut site sequence inside the PR8 M segment (**Figure 3.14F**). We picked a segment that is spliced because we previously found that PA-X activity is linked to splicing (Gaucherand et al., 2019). As a control, we inserted the STOML2 51 bp sequence with GCUG mutated to UAGC, as

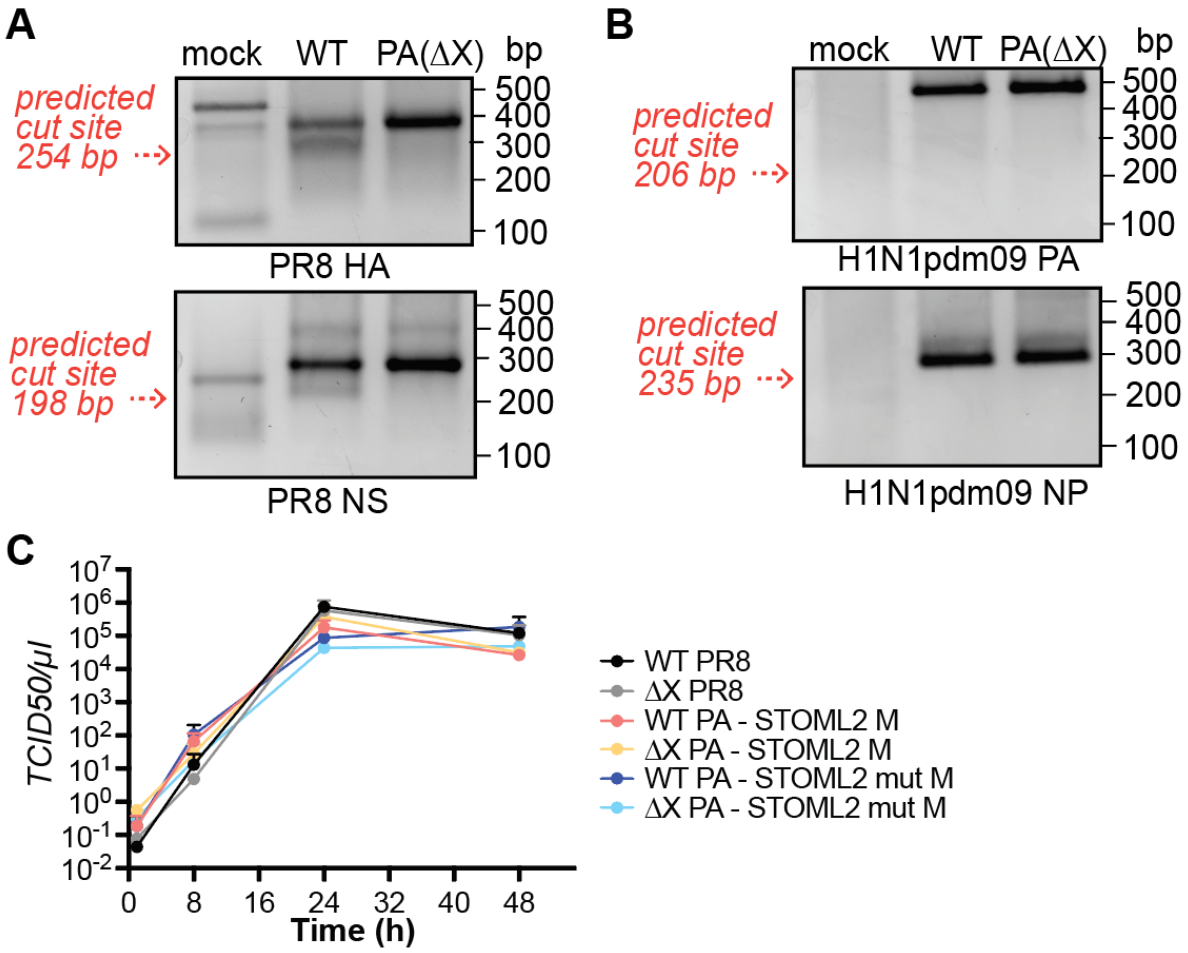

Figure 3.15: Viral mRNAs are not efficiently cleaved by PA-X.
(A-B) Xrn1 ko A549 cells were infected with WT or PR8-PA(ΔX) (A), or WT or H1N1pdm09-PA(ΔX) (B), or mock infected. RNA was extracted to run 5' RACE using primers ~150-200 nt downstream of GCUG sites that are predicted to be inside a hairpin loop within 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 representative of 3 experiments. (C) MDCK cells were infected at MOI 0.05 with the indicated viruses. Supernatants were collected at 1, 8, 24 and 48 hours post infection and viral titers were quantified by TCID50. n = 2.

this mutation prevented efficient PA-X cleavage (**Figure 3.7E-F, Figure 3.8F**). While we were able to recover recombinant virus, we noticed during the virus rescuing procedure that the titers were lower for viruses containing the WT STOML2 cut site when PA-X was present. We were thus surprised when we did not see any difference in viral growth compared to WT PR8 (**Figure 3.15C**). However, when we purified the viral RNA present

in the supernatant 48 hours post infection, we noticed that the STOML2 sequence was missing when PA-X was present (i.e. in the WT PR8 background) (**Figure 3.14G**). In contrast, the inserted STOML2 sequence was retained in the absence of PA-X (i.e. in the PR8-PA(ΔX) background) (**Figure 3.14G**). So was the mutated STOML2 sequence that is not cleaved by PA-X, in both WT PR8 and PR8-PA(ΔX) backgrounds (**Figure 3.14G**). Infecting A549 cells overnight also led to loss of the STOML2 sequence insertion (**Figure 3.14H**), even though the STOML2 sequence was present in the viral inoculum (**Figure 3.14I**). These results suggest that during viral propagation, viruses that have lost the cleavage site have an advantage for growth, but also that this is only true when PA-X is present. To confirm that this was not due to an error during the virus making process, we rescued the STOML2 recombinant viruses from plasmid two more times. We still saw that the inserted STOML2 sequence was selectively lost in the presence of PA-X, suggesting that having a preferred PA-X cut site inside a spliced viral segment is detrimental for the virus. Overall, these results are consistent with PA-X evolving specific cleavage characteristics to preferentially target host over viral mRNAs.

3.3. Discussion

Studies of viral host shutoff endoRNases have mostly relied on RNA steady-state levels as a proxy for RNA degradation. However, widespread RNA degradation can have downstream effects on other aspects of RNA metabolism (Gaucherand and Gaglia, 2022). Steady-state level analysis may thus preclude a clear understanding of how these enzymes target RNAs and which RNAs they degrade. Therefore, we examined RNA degradation directly by identifying PA-X cut sites transcriptome-wide and analyzing their characteristics in infected cells. 5’ RACE-seq analysis coupled with a custom pipeline allowed us to reveal several important characteristics of PA-X cleavage. We found that PA-X preferentially cuts RNAs at GCUG or similar tetramers located within hairpin loops.

This sequence/structure preference is conserved across PA-X from multiple influenza strains. Additionally, PA-X preferentially cuts RNAs within exons, likely after splicing. These preferred cleavage characteristics are more frequent in human transcripts than in intergenic regions or introns, and most importantly, than in influenza mRNAs. Moreover, PA-X prevents the growth of a recombinant virus carrying a preferred PA-X target sequence. Collectively these results suggest that PA-X cleavage specificity is tailored to induce host shutoff across a broad range of host RNA. They also suggest that influenza A virus has evolved a self vs. non-self discrimination mechanism, similar to cellular immune mechanisms that distinguish between host and viral nucleic acids.

A previous *in vitro* study of PA-X cleavage activity did not identify sequence specificity (Bavagnoli et al., 2015), likely because these characteristics can be missed *in vitro* without prior knowledge, highlighting the advantage of our technique. The study did find that PA-X preferentially cuts single-stranded RNA, likely explaining the preference for single-stranded loops. Interestingly, while PA and PA-X share the same RNase domain, they seem to have different cut site specificity (Bavagnoli et al., 2015), probably due to their different C-terminal domains. Studies that investigated which host RNAs are cap-snatched by PA have found little sequence specificity (Datta et al., 2013; Gu et al., 2015; Plotch et al., 1981; Sikora et al., 2017, 2014), suggesting that the location of PA cleavage is likely governed by the distance between the cap binding site in the polymerase basic 2 (PB2) subunit of FluPol and the PA active site. 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, which is still present in our cells (Brannan et al., 2012). It would thus be interesting to repeat the experiments in Xrn2 knock out/down cells, and also test whether Xrn2 has an effect on PA-X fragments in the nucleus.

A testimony to the importance of the sequence and structure preference of PA-X is that it is conserved across multiple influenza A strains (**Figure 3.4F, Figure 3.6C-D**). This is particularly interesting as PA-Xs from different strains have differences in host shutoff activity (Feng et al., 2016; Nogales et al., 2021; Oishi et al., 2019; Wang et al., 2020). However, the 5' RACE PCRs make this method very sensitive but not very quantitative. Therefore, it is difficult to say whether there are quantitative differences in efficiency of cleavage between strains. Nonetheless, the specific sequence and structure that is preferred by PA-X (**Figures 3.7, 3.9**) is abundantly found in host mRNAs. Recognizing a ubiquitous motif may thus be a strategy to target a broad range of host mRNAs. Supporting this idea, we reported a similar specificity for a degenerate recognition motif for another viral RNase, SOX from KSHV (Gaglia et al., 2015). Additionally, both PA-X and SOX, as well as the human endoRNase MCPIP1/Regnase-1 (Wilamowski et al., 2018), prefer cutting RNA within hairpin loop structures. Targeting exposed loops may be an easy way for endoRNases to access and cleave single-stranded RNA. Studying the cut site characteristics of other viral and human endoRNases will be important to test these ideas.

Two key questions in host shutoff are how the molecular activity of host shutoff factors is linked to immunomodulation and how it is regulated to prevent inhibition of viral gene expression. While we did not uncover a relationship between PA-X cleavage specificity and immunomodulation, we found that it may contribute to PA-X ability to spare viral mRNAs from degradation. Our analysis indicates that influenza A virus mRNAs contain few GCUG tetramers, mostly located in paired regions (**Figure 3.14**). This result is exciting as it provides an additional mechanism for virus vs. host discrimination by PA-X. Interestingly, most viral genomic RNAs have a higher GCUG percentage than their complementary mRNAs. This result may indicate that there is less pressure to decrease the GCUG percentage in the influenza genome, perhaps because

the genomic segments are associated with NP, which could protect them from cleavage (Lee et al., 2017). Conversely, GCUG is one of the most abundant tetramers in the human transcriptome (**Figure 3.13**). PA-X could have evolved this optimal target sequence to destroy as many host transcripts as possible without also targeting anything unnecessary. Consistent with this idea, inserting a sequence targeted by PA-X in one of the viral segments is detrimental to the virus only in the presence of PA-X (**Figure 3.14**). Therefore, we propose that this targeting specificity may act as a self vs. non-self discrimination mechanism for influenza A virus. Self/non-self discrimination is common in immune responses, as the cell needs to discriminate its own nucleic acids from those of invading pathogens to prevent inappropriate activation of immune and inflammatory signals. Common discrimination features include RNA modification, double-stranded RNA structures and DNA methylation. However, sequences can also be important. For example, the ZAP protein targets viral RNAs for degradation based on the presence of CG dinucleotides (Takata et al., 2017) and the Toll-like receptor TLR9 has preferred target sequences (Pezda et al., 2011). Our results suggest that influenza A virus is using a similar process in reverse, identifying cellular RNAs as non-self due to the abundance of specific sequences.

The fact that we only found 500-1,400 cut sites through 5' RACE-seq but that most host RNAs are down-regulated upon PA-X expression (Gaucherand et al., 2019) is intriguing. Of course, technical aspects of the method could limit our ability to capture more sites, even though our follow-up analyses suggest the ones we did find where true PA-X cut sites and could reveal preferred sequence characteristics. Nonetheless, we cannot conclude that these are the only RNAs directly cut by PA-X. Indeed, we detected many sites that appeared to not be shared between replicates, which tended to have lower read counts than the PA-X cut sites that were shared by all replicates or that were detected by the concatenated approach (**Figure 3.3C).** There are several possibilities as

to what may cause this result. The sites may not be PA-X cut sites at all, and the pipeline may be picking them up due to limitation in distinguishing signal from noise. Alternatively, PA-X may cut different RNAs with varying efficiency depending on how close sequences match the preferred cut characteristics. Thus, these additional sites could be cut by PA-X at lower efficiency. Lastly, we did find that the transcripts that harbor PA-X cut sites are expressed at higher levels, suggesting these additional sites may also be real PA-X cleavage locations that rise to the level of significance only in some replicates due to low gene expression. However, if we are indeed capturing the majority of PA-X cut sites in a cell, this suggests that PA-X may not need to cleave thousands of RNAs to trigger widespread down-regulation of RNAs. Indeed, rapid cellular RNA degradation triggered by expression of other viral RNases, infection with the murine gammaherpesvirus 68 (MHV68) and apoptosis causes a secondary inhibition of transcription (Abernathy et al., 2015; Duncan-Lewis et al., 2021; Hartenian and Glaunsinger, 2019). It is thus unclear how much of the widespread depletion in host RNAs during influenza infection and PA-X expression is due to RNA degradation *vs.* decrease in transcription. A decrease in transcription has been observed during influenza infection, although it may be partly driven by another influenza protein, non-structural protein 1 (Bauer et al., 2018; Zhao et al., 2018). Future studies will need to decouple RNA degradation from transcription to study this question. It would also be interesting to determine if specific RNAs need to be degraded by PA-X to induce host shutoff, or if the number of degraded RNAs and/or the speed of degradation determines transcription inhibition. Finally, inhibition of transcription could also dampen the induction of antiviral RNAs when PA-X is present (Gaucherand et al., 2019), even though few antiviral RNAs are directly cut by PA-X according to our analysis.

With 5' RACE-seq and our updated PyDegradome pipeline (**Figure 3.1**), we present a streamlined way to identify and characterize endoRNase cut sites throughout the

transcriptome that can be easily adapted to study other RNases. This method can uncover cut site preference characteristics that are masked *in vitro*, because RNase activity *in vitro* can be pushed to the limit by providing sufficient substrate or because cofactors are missing. This method directly identifies cleaved RNAs, instead of relying on RNA steady state levels, which can come from both degradation and transcription repression. Using this method, we have identified a new level of specificity in the mechanism of action of PA-X, adding to our understanding of how PA-X selects RNAs for degradation and distinguishes between host and viral mRNAs. This knowledge brings us one step closer to understanding how PA-X drives host shutoff, and more generally how PA-X is able to modulate inflammation and influenza pathogenesis. It also highlights the importance of determining the cleavage specificity of RNases to understand their functional role, and reveals the potential for viral self vs. non-self discrimination mechanisms.

3.4. 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 and members of their laboratories for their advice and input. We thank Drs. Richard Webby, Gideon Dreyfuss, Jesse Bloom and Seema Lakdawala for constructs, Dr. Mariano Garcia Blanco for providing the Northern Blot protocol, Dr. Bernard Moss for cell lines, and Drs. John Coffin, Claire Moore, Karl Munger 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.

3.5. Declaration of interests

The authors declare no competing interests.

3.6. Methods

3.6.1. Plasmids

pCR3.1-PA-X-myc, pCR3.1-PA-X-D108A-myc and pCR3.1-PA(fs)-myc (from PR8) were previously described (Khaperskyy et al., 2014). PR8 pHW-PA(ΔX) plasmid was generated as previously described (Gaucherand et al., 2019) from pHW-193, a kind gift from Dr. R. Webby (St Jude's Children Research Hospital, Memphis TN). pCR3.1-PA-X-Perth-myc was generated by PCR amplifying pHW-Perth09-PA, a kind gift from Dr. S. Lakdawala (Le Sage et al., 2021) 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 ligating each fragment to 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). pCDEF3-SOX and pCDNA3.1-vhs were previously described (Glaunsinger and Ganem, 2004b; Jones et al., 1995). The luciferase constructs with and without the β-globin intron were a kind gift from Dr. Gideon Dreyfuss (Younis et al., 2010). 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), or PCR amplified from human cDNA using Vent Polymerase (New England Biolabs), 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). 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). 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). Gibson cloning using HiFi assembly mix (New England Biolabs) was used to make all of these constructs, unless otherwise stated. All primers used for cloning are listed in Supplementary Table 3.2.

3.6.2. Cell lines and transfections

Human embryonic kidney HEK293T cells and Madin-Darby Canine Kidney (MDCK) cells were commercially obtained (ATCC). Wild type and Xrn1 knock out human

adenocarcinoma alveolar basal epithelial (A549) cells were a kind gift from Dr. Bernard Moss (Liu and Moss, 2016). 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) supplemented with 10% fetal bovine serum (Hyclone) at 37 °C and 5% CO_2. HEK293T inducible shXrn1 (ishXrn1) cells were previously described (Gaglia et al., 2015). For northern blotting and 5' RACE validation experiments, HEK293T ishXrn1 cells were treated with 1 µg/ml doxycycline (Thermo Fisher Scientific) 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). Cells were harvested 24 hours after transfection for RNA extraction and purification, or infected overnight as described below and harvested the next day for **Figure 3.10B**.

3.6.3. Viruses and infections

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, as well as their mutant recombinant virus counterparts PR8-PA(ΔX), H1N1pdm09-PA(ΔX) and Perth-PA(ΔX), and the mutant recombinant viruses containing the WT or mutated STOML2 51 bp cut site in the M or NA segment were generated using the 8-plasmid reverse genetic system (Hoffmann et al., 2000) as previously described (Gaucherand et al., 2019; Khaperskyy et al., 2012). 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 (Matrosovich et al., 2006). 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. Influenza infections were

performed in DMEM supplemented with 0.5% low endotoxin bovine serum albumin (BSA, Sigma-Aldrich), 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) 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 90μl of infection media. 25 x 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 (REED and MUENCH, 1938) was used to calculate viral titers via the Bloom lab Python script at https://github.com/jbloom/reedmuenchcalculator. 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) following manufacturer's protocol. The RNA was then treated with RNase-Free DNase set (Qiagen), purified and concentrated using the RNA Clean & Concentrator kit (Zymo Research), and reverse

transcribed to cDNA using iScript Supermix (Bio-Rad), per manufacturer's protocol. Taq DNA polymerase (New England Biolabs) 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.

3.6.4. 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) and cOmplete protease cocktail inhibitor (Roche). 20-100 µg of protein was loaded on an SDS-PAGE gel (Bio-Rad) and transferred onto PVDF membranes (EMD Millipore), 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), rabbit anti influenza A virus PA-X antibodies (Biorbyt #184340, 1:1000) or rabbit anti *b*-tubulin 9F3 antibodies (Cell Signaling Technologies #2128, 1:1000) diluted in 0.5% milk PBTS. Secondary antibodies were purchased from Southern Biotech 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).

3.6.5. RNA purification, cDNA generation and qRT-PCR

RNA was extracted and purified using either the RNeasy Plus mini kit (Qiagen) for 5' RACE-seq experiments, or the Quick-RNA miniprep kit (Zymo Research) for all other experiments, following manufacturer's protocol. RNA was then treated with Turbo DNase

(Life Technologies) 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), per manufacturer's protocol. qRT-PCR was performed using iTaq Universal SYBR Green Supermix (Bio-Rad), 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 3.2. Northern blotting, 5' RACE and 5' RACE-seq specific RNA processing are detailed in the next sections.

3.6.6. Northern blotting

Northern blotting was performed using the NorthernMax kit (Invitrogen) 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). Biotinylated DNA probes against the 3' UTR of the luciferase reporters were generated by PCR amplification using Taq polymerase (New England Biolabs) in the presence of Biotin-16-dUTP (Sigma-Aldrich), see primers in Supplementary Table 3.2. Northern blots were probed with these biotinylated DNA probes, then blocked with Intercept PBS Blocking Buffer (LI-COR Biosciences) supplemented with 1 % SDS. Finally, northern blots were incubated with IRDye 800CW Streptavidin (LI-COR Biosciences, 1:10,000) 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).

3.6.7. 5' RACE

After DNase treatment and phenol chloroform extraction, the RACE adapter (see Supplementary Table 3.2) was ligated to 1-5 µg RNA using T4 RNA ligase (Invitrogen) 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) per manufacturer's protocol. Taq DNA polymerase (New England Biolabs) 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 3.2) and reverse primers annealing within the gene of interest (see Supplementary Table 3.2). Finally, PCR products were run on a 2% agarose gel containing HydraGreen safe DNA dye (ACTGene) 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.

3.6.8. 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 (Chao et al., 2017; German et al., 2009; Zhai et al., 2014). 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). The rest of the RNA was ligated to the RACE-seq adapter (see Supplementary Table 3.2) using T4 RNA ligase (Invitrogen) 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)

using random primers of different lengths that also include a known sequence adapter (long and short RT primers, Supplementary Table 3.2). 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 3.2). Between the two PCR rounds, DNA fragments of 150-400 bp were selected using SPRIselect Reagent (Beckman Coulter). 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). The raw sequences and identified cut sites are deposited on the NCBI GEO database (GSE207253, reviewer token: idatiuqodnyxtev).

3.6.9. Reads preprocessing and alignment

After labeling each read with its own UMI, cutadapt v3.5 (Martin, 2011) 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 (Kim et al., 2019) 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.

3.6.10. PyDegradome and other python analyses

PyDegradome is a Python-based peak finding pipeline developed to identify ribonuclease cut sites, which we originally reported (Gaglia et al., 2015) and modified here. Once all reads are aligned to the genome, the position of the first nucleotide of every read is recorded, and the number of reads that map to the same position is counted for each read of each library. PyDegradome 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. This significance can be made more or less stringent based on three parameters that can be 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. All the data showed here use a window of 4 nucleotides. For a more detailed explanation of PyDegradome, see our previous PyDegradome publication (Gaglia et al., 2015). In this publication, we updated our older version of PyDegradome to be usable with the alignment tool HISAT2 and the human genome GRCh38/hg38. Our older version of PyDegradome also restricted the cut site analysis to exons, and we have updated it to include introns in case PA-X cut pre-mRNA. The PyDegradome 2.0 updated script can be found in our laboratory's GitHub page (https://github.com/mgaglia81/PyDegradome). The down-stream motif generation and scoring analysis was not modified and is still available to download as supplementary material in our previous article (Gaglia et al., 2015) or our

GitHub page. 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.

3.7. Supplemental information

Supplementary Table 3.2 – Summary of PA-X cut sites

Supplementary Table 3.3 – Primers and RNA adapters.

3.8. Author contribution

Amrita Iyer constructed plasmids and carried out replicates of the 5' RACE experiments pictured in Figures 3.9B and 3.10A, and helped me process the structure data in Figure 3.9C. Isabel Gilabert constructed the Luciferase + repeat plasmid depicted in Figure 3.4C and performed original 5' RACE experiments on it, not shown here. Marta Gaglia performed the analysis to plot Figure 3.3C. Marta Gaglia and Chris Rycroft wrote the original PyDegradome pipeline and helped me optimize it for PA-X. The rest of the experiments and analyses were carried out by Léa Gaucherand.

Chapter 4. Discussion

4.1. Main conclusions of work

Many studies using animal models have highlighted the critical role of PA-X for virulence and pathogenesis *in vivo*. Despite its importance, the molecular mechanism of action of PA-X was largely unknown when I started my thesis work in May 2018. We knew that PA-X is an endoribonuclease that induces host shutoff (Jagger et al., 2012), that nuclear localization and the first 15 aa of the X-ORF are required for activity (Hayashi et al., 2016; Khaperskyy et al., 2016; Oishi et al., 2015), and that PA-X preferentially degrades host RNAs transcribed by Pol II but not viral mRNAs (Khaperskyy et al., 2016). My thesis work has focused on identifying which RNAs are specifically targeted by PA-X and what their characteristics are, in order to uncover the mechanism of action of PA-X at the molecular level. In view of the impact of PA-X activity *in vivo*, studying PA-X could allow us to better understand how viral factors can influence inflammation and disease outcome.

4.1.1. PA-X activity is linked to splicing

Based on the mechanism of action of other viral RNA degradation factors, we hypothesized that PA-X binds specific proteins to reach its target RNAs. We took two complementary approaches to identify these proteins: a transcriptomic approach to identify RNAs down-regulated by PA-X, and a proteomics approach led by our collaborators to identify proteins interacting with the X-ORF (Gaucherand et al., 2019). The idea behind our transcriptomic approach was to identify specific features common to RNAs down-regulated by PA-X that would hint at processes important for PA-X activity. Our transcriptomic analysis uncovered a link between PA-X activity and splicing (**Figures 2.6, 2.7**), which we validated experimentally using reporters (**Figures 2.8, 2.9**).

The main targets of PA-X are thus spliced RNAs, suggesting that PA-X may be preferentially recruited to spliced RNAs, potentially by interacting with a protein involved in splicing. Since only Pol II transcribed RNAs are spliced, this hypothesis would be consistent with the previous finding that PA-X preferentially targets RNAs transcribed by Pol II .

One intriguing finding was that the only IFN-λ2 reporters that were down-regulated by PA-X were the ones that had an intron with the 5' splice site consensus sequence (**Figures 2.8E-F, 2.9D-G**). It suggested that splicing of only certain introns can trigger targeting by PA-X. This could either come from increased splicing efficiency associated with the consensus splice site sequence, or from the use of different proteins that recognize a specific splice site to carry out splicing of the intron. Since these results were published, I have continued to study this question using multiple approaches (see appendix 5.1 for a detailed description). While experiments with our IFN-λ2 reporter point to splicing efficiency and the 5' splice site sequence as being key for PA-X targeting, these results were not as clear cut using two additional reporters. Moreover, mutating the intron sequence of some IFN-λ2 reporters to be recognized by the minor spliceosome instead of the major spliceosome, which use different small nuclear RNAs and proteins to splice RNAs, did not interfere with PA-X targeting (see appendix 5.1). Overall, these experiments suggest that while PA-X activity is linked to splicing, the 5' splice site sequence is not the only determinant for PA-X targeting, and the molecular mechanism of how PA-X targets RNAs is more complex than just binding to a spliceosome protein.

If not a spliceosome protein, PA-X may still interact with a protein involved in mRNA processing that would be indirectly linked to splicing. Our collaborators carried out a proteomic screen to look for interactors of the X-ORF (**Figure 2.10**). My colleague

Rachel Levene was able to validate by co-immunoprecipitation the interaction between PA-X and nudix hydrolase 21 (NUDT21), a member of the CFIm complex (Kim et al., 2010) (**Figure 2.12A**). While NUDT21 is mostly known for its role in polyadenylation site selection (Zhu et al., 2018), the CFIm complex is detected in proteomic analyses of the spliceosome, and has been proposed to link splicing to polyadenylation (Rappsilber et al., 2002; Zhou et al., 2002). In addition, a recent study showed that NUDT21 is involved in splicing of MAT2A mRNA, independently of poly(A) site choice (Scarborough et al., 2021). Binding to NUDT21 may thus allow PA-X to preferentially reach spliced RNAs. Consistent with this idea, I found that knocking down the CFIm complex decreases PA-X activity (**Figure 2.12B**). However, additional experiments should be carried out to more thoroughly test the hypothesis that PA-X is recruited to the RNA through NUDT21. For example, recruitment of PA-X to the RNA could be directly studied by RNA-immunoprecipitation, which I was unfortunately unable to optimize. Sophisticated microscopy experiments could also be carried out to confirm the interaction of PA-X and NUDT21 on the RNA. Moreover, one caveat of the proteomics strategy was that both negative controls were localized to the cytoplasm, while the wild-type X-ORF localizes to the nucleus. This difference in localization likely led to a number of false positive hits. On the other hand, some parts of the nucleus such as chromatin can be difficult to probe without the use of DNase or other ways to break down DNA, which was not used in the proteomics experiment. Any potential PA-X interactor in the chromatin fraction may thus not have been detected.

4.1.2. The RNAs down-regulated by PA-X may not all be cleaved by PA-X

Since host shutoff by other viral RNases can trigger transcription inhibition (Abernathy et al., 2015), one alternative hypothesis for the link between PA-X activity and splicing is that host shutoff preferentially inhibits transcription of spliced RNAs.

Indeed, down-regulation by PA-X could theoretically come from direct degradation of the RNA but also from a decrease in transcription. To avoid this caveat, we used 5' RACE-seq along with our PyDegradome pipeline to directly study cleavage by PA-X in cells instead of measuring RNA steady state levels. Depending on the PyDegradome analysis used (**Figure 3.2D**), we either found 592 or 1361 PA-X cut sites, which corresponded to 517 or 1309 different genes, respectively. Yet, in both cases, the number of distinct RNAs cleaved by PA-X is much smaller than the > 3,000 genes down-regulated by PA-X that we observed by RNA sequencing (Gaucherand et al., 2019). This discrepancy could stem from a limitation of our analysis, as our results suggest that very deep sequencing coverage may be required to reliably identify PA-X cut sites over the background of basal RNA degradation, ideally deeper than what we used. On the other hand, our analyses suggest that the number of cut sites identified by PyDegradome plateaus at less than 1500 cut sites with increased input reads (**Figure 3.2E**), which implies that we already have enough sequencing depth to identify most PA-X cut sites. There could nonetheless be other limitations. For example, I have observed that cut sites are predominantly found in RNAs that are the most abundant, as assessed by RNA sequencing of the same samples in parallel of 5' RACE-seq (**Figure 4.1A-B**). This bias could come from the inefficiency of the adapter 5' ligation during our library protocol, or from our PyDegradome analysis not being able to pick up low level RNAs cut over the background. We are currently addressing the latter by incorporating a normalization step in our PyDegradome pipeline, which normalizes the number of reads of a cut RNA by the total level of the same RNA. We will test whether this improves cut site identification. Alternatively, PA-X could naturally preferentially target abundant RNAs, which would make sense from a physical perspective. The discrepancy between the number of PA-X cut sites identified by PyDegradome and the number of RNAs down-regulated by PA-X could also stem from the biology of PA-X. Indeed, since widespread RNA degradation

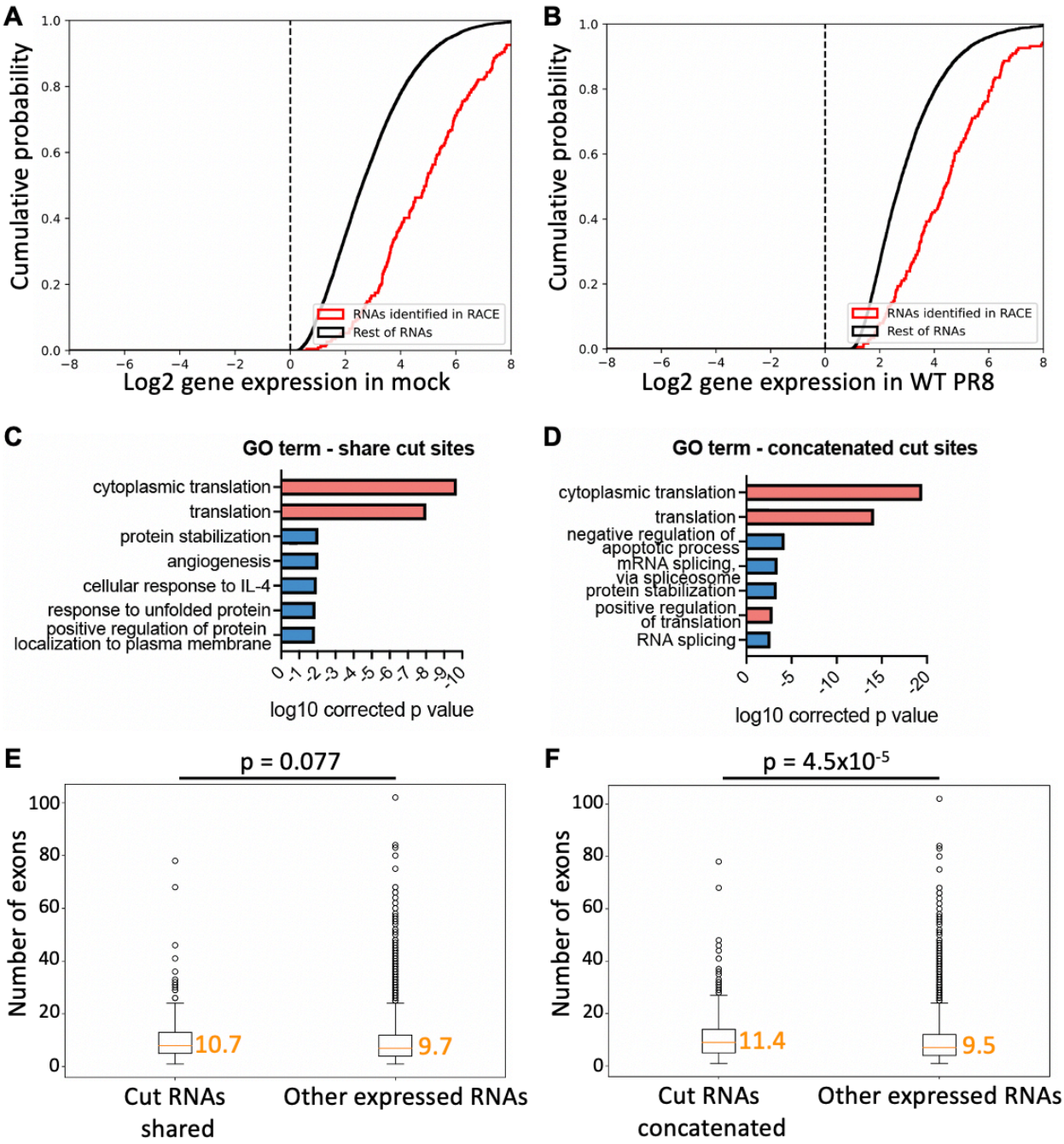

Figure 4.1: Characterization of PA-X cut sites.

(A-B) Cumulative distribution histograms of RNAseq levels averaged from three mock samples (A) or WT PR8 samples (B), plotted separately for RNAs that have PA-X cut sites according to PyDegradome shared analysis in red, or the rest of RNAs in black. (C-D) DAVID was used to identify overrepresented Gene Ontology (GO) terms for biological processes among RNAs cut by PA-X identified by PyDegradome using the shared (C) or concatenated (D) analysis. The Benjamini corrected p value is plotted for the 7 most significant GO terms. (E-F) Box plot representing the number of exons for RNAs cut by PA-X identified by PyDegradome using the shared (E) or concatenated (F) analysis that were also detected by RNAseq, compared to the other RNAs detected by RNAseq. p values obtained using Kolmogorov-Smirnoff test.

triggers secondary consequences that exacerbate host shutoff (see section 1.2.4), PA-X may only need to cleave a few abundant RNAs to trigger transcription inhibition of more genes and global down-regulation of most RNAs. In addition, many genes can regulate each other or are similarly regulated. An already cited example is the induction of hundreds of ISGs upon sensing of type I and III IFNs. In this example, cleavage of any RNA along the IFN sensing pathway could lead to a down-regulation of all ISGs, even though these ISGs are not directly cut by PA-X.

There are also additional potential confounds. GO term analysis of the 517 or 1309 distinct RNAs cut by PA-X identified translation as the most enriched biological process (**Figure 4.1C-D**). This enrichment is intriguing, as we previously found that RNAs resistant to down-regulation by PA-X were enriched in GO terms such as translation, translation initiation and structural constituent of ribosome (**Figure 2.5D**). One explanation could be that these RNAs are the only abundant RNAs left in the cell for PA-X to cleave at the time of collection, as PA-X would have down-regulated the other RNAs, either through cleavage or secondary transcription inhibition. On the other hand, these experiments are performed in Xrn1 knock out cells, which could also have an effect on the transcription inhibition feedback mechanism. Indeed, silencing of Xrn1 prevented transcription inhibition in cells overexpressing muSOX (Abernathy et al., 2015). Unfortunately, this mechanism has not been investigated in the context of IAV infection or PA-X overexpression. It is thus difficult to predict the effect of Xrn1 knock down in our system. Moreover, in some contexts such as during MHV68 infection, silencing of Xrn1 is not enough to prevent transcription inhibition (Hartenian et al., 2020). Studying the effect of PA-X on secondary transcription inhibition and/or carrying out 5' RACE-seq at earlier time points to see if different cut RNAs are detected over time based on their abundance will help resolve these possibilities. I have attempted to probe RNA degradation and transcription repression separately using thiol(sh)-linked alkylation

for the metabolic sequencing of RNA (SLAM-Seq) during infection with WT or PA-X deficient IAV (Fasching et al., 2022; Herzog et al., 2017). This method involves labeling of nascent RNAs with 4-thiouridine (4-SU) then high-throughput sequencing to measure the abundance of both labeled and unlabeled RNAs over time. While I was able to confirm incorporation of 4-SU over time, I obtained very inconsistent PA-X activity at different time points across my three replicates, which prevented me from drawing any conclusion. This could be due to various technical errors, as well as variable cellular stress from prolonged 4-SU treatment, especially at the 12h time point. It will be important to try this experiment again or carry out similar types of experiments that allow to directly look at degradation of RNAs by PA-X over time and potential transcription inhibition separately.

Interestingly, RNAs cut by PA-X have a slightly higher number of exons than the rest of expressed RNAs for both shared cut sites and concatenated cut sites, suggesting that splicing may be directly involved in PA-X cleavage activity (**Figure 4.1E-F**). However, this difference in exon number is small. In contrast, RNAs down-regulated by PA-X have a much higher number of exons than PA-X resistant RNAs (**Figure 2.6G**), suggesting that spliced RNAs may additionally be preferentially down-regulated by PA-X as an indirect consequence of PA-X activity. Alternatively, there may be a smaller difference in exon numbers when looking at cut RNAs because of a lack of statistical power from the low number of distinct RNAs cleaved by PA-X. Our potential experimental bias towards detecting abundant RNAs may also skew the results. Nonetheless, this result hints at a potential link between splicing and PA-X direct degrading activity and should be more thoroughly investigated.

4.1.3. PA-X cleaves RNA at a preferred sequence and structure to distinguish between host and viral mRNAs

Although I may not have identified all PA-X cut sites transcriptome-wide, I have identified enough high confidence sites to infer the preferred characteristics of PA-X cleavage. Both the transcriptomic data and the experimental follow-up indicate that PA-X cleavage is sequence specific, and that PA-X preferentially cleaves GCUG or a similar tetramer. However, this GCUG also needs to be located in a loop within a hairpin structure. This finding makes sense, as GCUG tetramers are ubiquitous but PA-X does not cut RNA at every GCUG sequence. Moreover, since the loop is single-stranded, this finding is consistent with previous *in vitro* results showing that PA-X preferentially cleaves single-stranded over double-stranded RNA (Bavagnoli et al., 2015). Interestingly, the actual location of cleavage within the GCUG sequence seem to vary based on the RNA being cut (**Figures 3.7C, 3.8C**). It is unclear whether this variation is an artefact of our method or a true characteristic of PA-X. It could potentially be influenced by the location of GCUG relative to the hairpin loop.

I also found that the cut site specificity of PA-X is conserved across multiple IAV strains (**Figures 3.4F-G, 3.6C**). This finding makes sense biochemically, since the RNase domain of PA-X is very well conserved across IAV strains (**Figure 3.6D**). Yet, studies have reported differences in the level of down-regulation by PA-X based on the IAV strain, with the pandemic H1N1 strain usually having stronger host shutoff activity (Desmet et al., 2013; Hayashi et al., 2016; Nogales et al., 2018). These differences in activity may come from differences in PA-X stability in cells (Levene et al., 2021), as well as potential differences in PA-X interacting proteins, or an unknown function of PA-X in certain IAV strains that has not been identified yet. Nonetheless, the fact that different IAV strains evolved differences in PA-X activity while keeping the same PA-X cut site specificity is a testament to the efficiency of this cut site specificity.

It is interesting that PA and PA-X share the same RNase domain but do not seem to have the exact same cleavage specificity. Indeed, PA preferentially cleaves RNA for cap-snatching after a guanine residue within a 5′-GC-3′ motif (Datta et al., 2013). Of note, the two nucleotides least likely to vary in the PA-X cut site motif are the first GC dinucleotides (**Figure 3.7D**), so PA and PA-X may still recognize similar RNA sequences. However, the PA cleavage preference is largely governed by binding of the RdRp to the 5' cap through PB2, which restricts RNA cleavage within 50 angstroms of the cap, or 10 to 14 nucleotides (Reich et al., 2014). Cleavage is thus sequence dependent if a GC dinucleotide is present within this length range, or in general if the sequence of the resulting cap-snatched primer is compatible with the 3' end of the viral RNA segment to be transcribed (Geerts-Dimitriadou et al., 2011a, 2011b; Sikora et al., 2015; te Velthuis and Oymans, 2018). In contrast, PA-X is not restricted to the 5' end, and it is still unclear if PA-X is recruited to a specific location on the RNA through interactions with RNA processing proteins. Of note, the PA-X cut sites identified by PyDegradome can be found throughout each exon, with no clear bias towards a specific RNA landmark such as splice sites (**Figure 4.2A**). However, almost all identified cut sites were located within exons and not introns (**Figures 3.11A, 3.12A**). Even inserting what appears to be a strongly favored PA-X cut site sequence within the intron of a reporter did not lead to cleavage in the intron (**Figures 3.11D-E, 3.12D-E**). Since PA-X activity is linked to splicing, which occurs co-transcriptionally for most spliced introns (Gaucherand et al., 2019), these data are consistent with splicing occurring very rapidly and suggest that PA-X cleaves RNA after it has been spliced. Alternatively, in addition to cleaving RNAs after splicing has occurred, PA-X may cut RNAs within both introns and exons during transcription, but the resulting fragments may be degraded by Xrn2, the host exoribonuclease responsible for the 5' to 3' degradation of nascent uncapped RNAs (Nagarajan et al., 2013). Since we did not silence Xrn2, we may have missed these

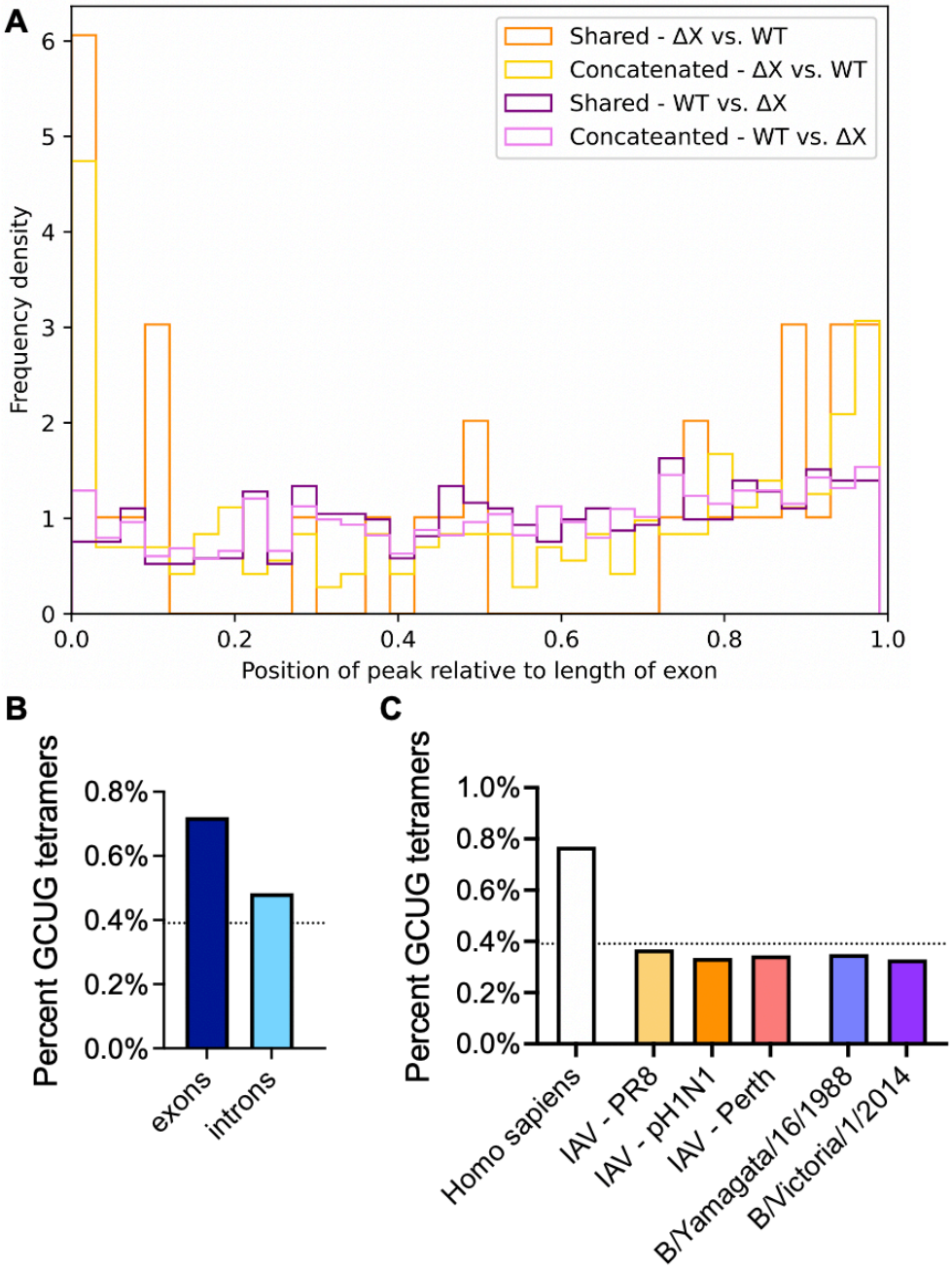

Figure 4.2: Abundance of GCUG sites mostly directs PA-X cleavage.
(A) Distribution along exons of PA-X cut sites (WT vs. ΔX) or control sites enriched in the ΔX samples (ΔX vs. WT) as identified by PyDegradome using the shared approach (purple or orange lines) or concatenated approach (pink or yellow lines). (B-C) Percentages of GCUG tetramers found within human exons compared to introns (B), or within the human transcriptome compared to the transcriptomes of the IAV strains PR8, pH1N1 and Perth, and the two common influenza B strains B/Yamagata/16/1988 and B/Victoria/1/2014 (C).

fragments. It would be important to carry out northern blot or 5' RACE experiments in Xrn1 *vs.* Xrn2 knock down/out cells to test this idea.

Another possible explanation for why most PA-X cut sites are in exons is that the preferred cleavage characteristics of PA-X may bias towards cleavage of exons over introns. Indeed, GCUG tetramers are twice as abundant in exons than in introns (**Figure 4.2B**). PA-X may thus still cut the RNA co-transcriptionally but preferentially recognize GCUG sites within exons. Furthermore, this could be a way for PA-X to ensure that the actual RNA will be degraded and not risk cleaving "decoy" introns after they are spliced out, which would have little influence on host gene expression. In fact, GCUG is one of the most abundant tetramers in the human transcriptome (i.e. in human mature mRNAs and noncoding RNAs), while it is of average abundance in the human genome, which also includes untranscribed and regulatory regions (**Figures 3.13A, 3.14A**). This preference could thus ensure that on top of not cleaving unnecessary RNAs, PA-X preferentially cleaves RNAs that code for proteins. Importantly, GCUG tetramers are of average percentage or slightly underrepresented in the transcriptomes of multiple human IAV strains, and the GCUG sequences that are present in these strains are mostly predicted to be in paired regions and not hairpin loop structures (**Figures 3.13C-E, 3.14A, 3.14E**). This exciting finding suggests that PA-X may have evolved its cleavage specificity to avoid targeting viral mRNAs while preferentially targeting human mRNAs. The fact that viruses with the STOML2 cut site sequence in the spliced M segment can only grow to high titers in the absence of PA-X also supports this idea (**Figure 3.14F-I**). Interestingly, high viral titers can still be obtained when the same PA-X cut site sequence is inserted into the NA segment, which is not spliced (**Figure 4.3A**). Of note, lower levels of NA but not HA viral mRNA were recovered after inserting the STOML2 sequence in the NA segment (**Figure 4.3B-C**). However, the same low levels were obtained in the absence of PA-X and when inserting a mutated STOML2 sequence

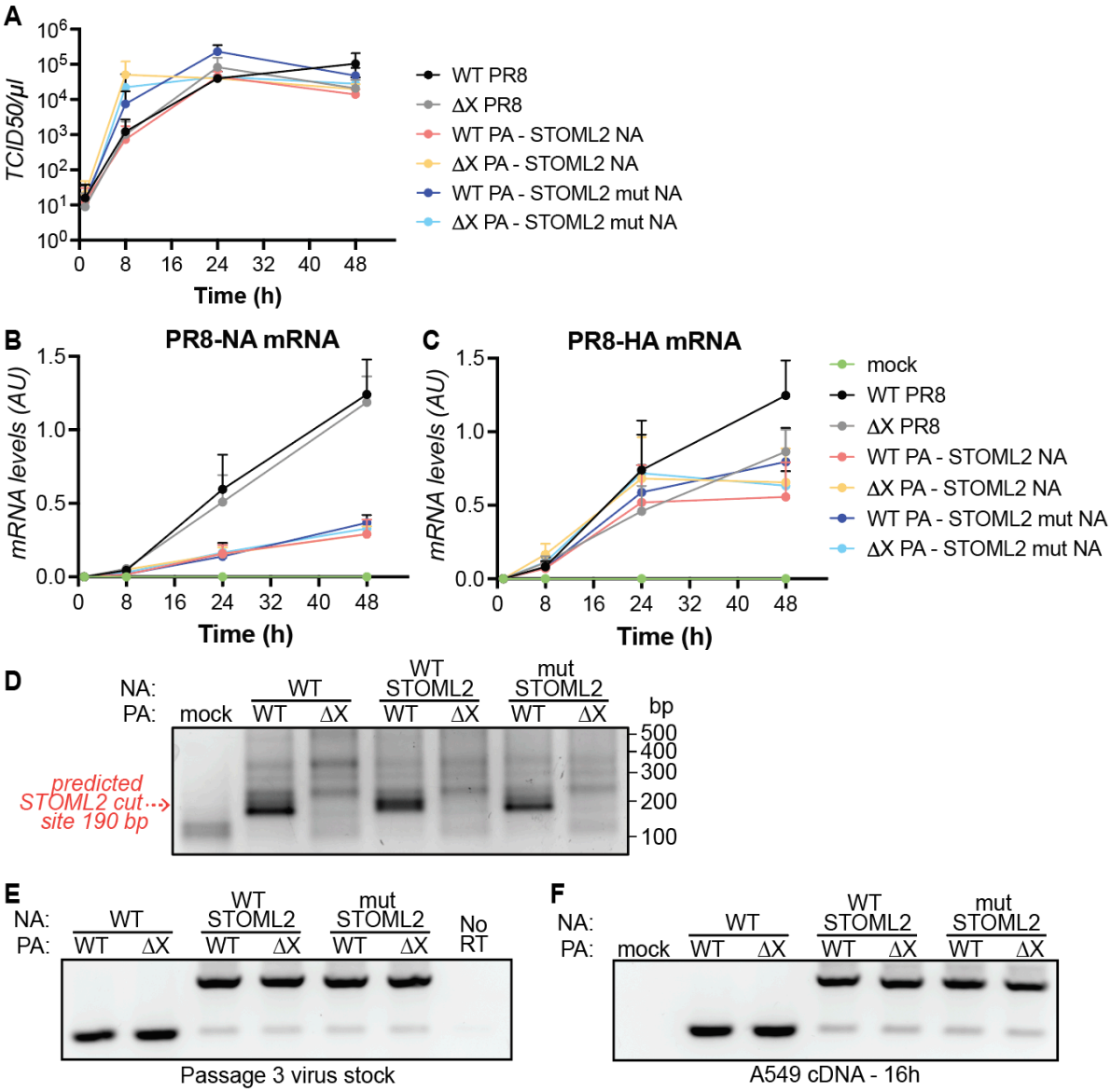

Figure 4.3: Inserting a PA-X cut site sequence inside the NA segment does not affect viral replication.

(A-C) MDCK cells were infected at MOI 0.05 with the indicated viruses. Supernatants and RNA were collected at 1, 8, 24 and 48 hours post infection. (A) viral titers were quantified by TCID50. n = 3. (B-C) PR8 NA (B) and HA (C) mRNA levels were quantified by RT-qPCR, normalized by 18S rRNA levels, and plotted as mean ± standard deviation. AU, arbitrary units. n = 3. (D) Xrn1 knock out A549 cells were infected with the indicated viruses, or mock infected for 16 hours before RNA extraction. 5'RACE was then performed using primers specific for PR8 NA, ~150 nt downstream of the predicted cut site inside the STOML2 sequence. The PCR products were run on an agarose gel. The predicted size of DNA bands coming from a cut site inside the STOML2 sequence is indicated by the red dotted arrow. DNA bands were purified and sequenced to confirm their identities, and image is representative of 3 experiments. (E) Viral RNA was extracted from our viral stocks, reverse-transcribed to cDNA and PCR amplified using primers located on either side of the STOML2 sequence to visualize on an agarose gel

whether the STOML2 sequence was retained. n = 1. (F) cDNA from (D) was also PCR amplified using the same primers to visualize whether the STOML2 sequence was retained. Gel image is representative of 3 experiments.

that is not cut by PA-X, suggesting that this defect comes from inserting any sequence and not a specific sequence cleaved by PA-X. Inserting the STOML2 sequence lead to a faint cut site inside the STOML2 sequence by 5' RACE (**Figure 4.3D**). However, there was also a faint cut site at a GCUG sequence in the NA segment in the absence of the STOML2 sequence or when inserting the mutated STOML2 sequence (**Figure 4.3D**). While the band appeared faint, the two rounds of nested PCRs make our 5' RACE method not very quantitative, so we cannot really conclude about the abundance of this cut site. Nonetheless, cleavage of the STOML2 sequence was not detrimental enough for the recombinant virus to grow, as there was no enrichment of the WT virus over time as was seen when inserting the STOML2 sequence in the M segment (**Figure 4.3E-F**). This result suggests that PA-X does not cleave NA mRNA efficiently, which is consistent with our model that spliced mRNAs are preferentially cleaved over intronless ones, since as far as we know NA mRNA is not spliced. Targeting spliced RNAs thus adds a second layer of protection for viral mRNAs, as most of them are not spliced, like NA mRNA and viral genomic RNAs. Alternatively, PA-X may already cleave NA mRNA as part of the normal viral life cycle, so adding a second cut site does not influence the amount of NA being cut. This is consistent with the faint band detected by 5' RACE (**Figure 4.3D**). It would be interesting to distinguish these two possibilities by determining the actual amount of NA mRNA being cleaved by PA-X, for example using northern blotting.

This GCUG enrichment in the host transcriptome is conserved across common IAV hosts, such as chickens, pigs and ducks (**Figure 3.14D**). The reason for this GCUG enrichment in host transcripts is unknown, and is an interesting observation to follow up on. Moreover, GCUG tetramers are also slightly underrepresented in the transcriptome

of influenza B virus (IBV) (**Figure 4.2C**). Yet, influenza B viruses do not encode a PA-X protein, suggesting that the low/average percentage of GCUG tetramers could stem from a different pressure than PA-X that would be advantageous for all influenza viruses. This evolutionary pressure could be the same that pushes GCUG tetramers to be enriched in the transcriptomes of humans, birds and swine. PA-X could then have evolved to use this difference to its advantage, as a way to distinguish between host and viral mRNA. This distinction is reminiscent of the host self *vs.* non-self discrimination that is commonly found in immune responses. Analogous to PA-X using the high percentage of GCUG in the human transcriptome to its advantage, the host antiviral protein zinc-finger antiviral protein (ZAP) uses the low percentage of CG dinucleotide in the human genome to identify non-self CG rich viral RNA and prevent viral replication (Takata et al., 2017).

Based on the results from my two manuscripts and the points raised in the discussion, my current model of PA-X mechanism of action is that PA-X is recruited to Pol II transcribed spliced RNAs, potentially through interactions with the RNA processing machinery. Once there, PA-X recognizes and cleaves the RNA within a GCUG sequence inside the loop of a hairpin structure, if present. Viral mRNAs mostly lack these characteristics, as they are intronless (with two exceptions) and have low GCUG content. Therefore they are spared from cleavage by PA-X. As a result of this activity, many RNAs are down-regulated, either through direct cleavage or potentially through transcription inhibition, leading to host shutoff and a general blocking of the host innate immune response and inflammation. This working model is summarized in **Figure 4.4**.

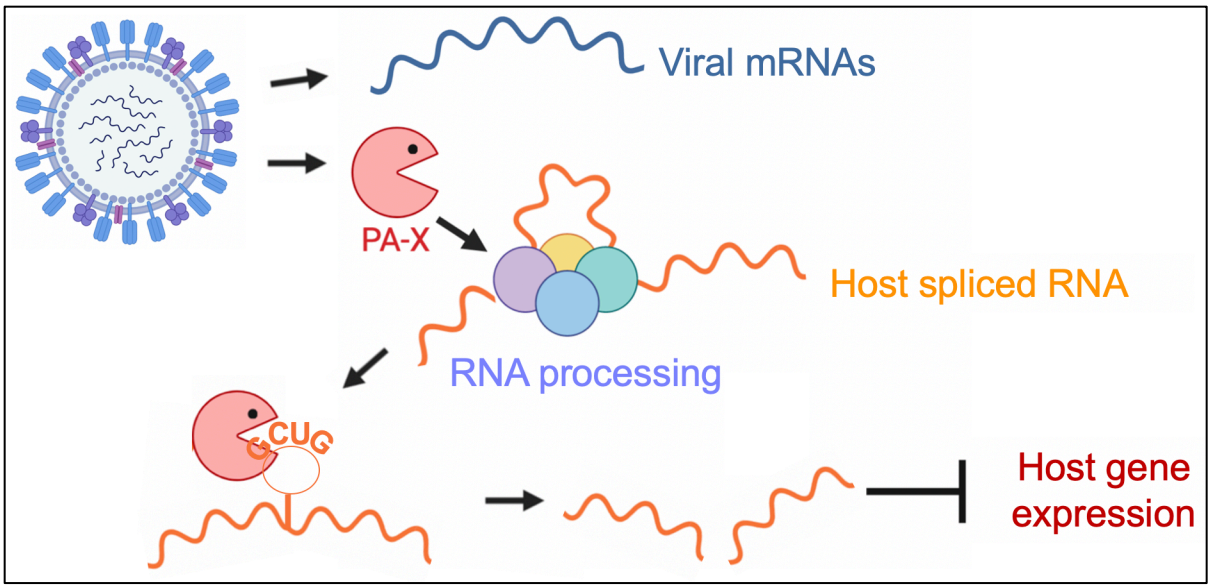

Figure 4.4: Current model of PA-X molecular mechanism of action.
Based on my thesis work, my current working model RNA targeting by PA-X at the molecular level is that PA-X is preferentially recruited to spliced RNAs through interactions with the RNA processing machinery. Once at the RNA, PA-X can recognize and cleave GCUG or similar sequences within loops of hairpin structures, leading to the degradation specific host RNAs, while viral mRNAs are spared from degradation because they mostly lack these cut site characteristics. This activity overall leads to host shutoff. Diagram made using biorender.

4.2. Implications of using the mRNA processing machinery as a targeting strategy

All viral RNA degradation-dependent host shutoff factors, i.e. the viral RNases PA-X, vhs, SOX, muSOX and BGLF5, the host RNase inducing protein nsp1 and the viral decapping enzymes D9, D10, ASFV-DP and L375, specifically target RNAs transcribed by Pol II (Gaucherand and Gaglia, 2022). This ensures the degradation of host mRNAs to prevent protein expression and thus induce host shutoff. Therefore, all these viral proteins need a mechanism to find the correct RNAs. Our results indicate that PA-X may solve this problem by binding to a protein involved in mRNA processing that will guide it to Pol II transcribed RNAs. This mechanism provides an efficient strategy to specifically target coding RNAs to induce host shutoff.

Additional advantages can come from hijacking the mRNA processing machinery. Targeting RNAs during processing could allow PA-X to degrade RNAs early in their life cycle. Degrading RNAs early could in turn provide a way to free up more resources for the virus than if the RNAs were degraded in the cytoplasm, which is the case for other viral RNases such as vhs and SOX. For example, degrading host mRNAs in the nucleus could allow to reallocate RNA nuclear export proteins to export viral mRNAs instead. In addition, many RNAs are naturally cleaved in the nucleus as part of their normal processing, for example for polyadenylation and splicing. Some faulty mRNAs are also cleaved and degraded during transcription as part of quality check pathways. Cleavage by PA-X may thus not be detected as aberrant right away if it occurs in an environment where RNA cleavage is common.

In addition to directing PA-X to its target RNAs, binding to an mRNA processing machinery protein could modulate host gene expression by interfering with the normal activity of that protein. Similar mechanisms have been uncovered for other viral proteins. For example, as mentioned previously, IAV NS1 binding to the CPSF complex prevents cleavage and polyadenylation of nascent RNAs, which in turn decreases protein expression (Nemeroff et al., 1998). Interfering with mRNA processing could thus exacerbate host shutoff by PA-X. Moreover, since mRNA processing is co-transcriptional and many proteins involved in mRNA processing bind to the Pol II CTD, PA-X could be in complex with Pol II and interfere with Pol II elongation. Consistent with a co-transcriptional activity of PA-X, some PA-X can be found within the chromatin fraction (**Figure 4.5A**). Moreover, in one preliminary experiment, I observed that PA-X co-immunoprecipitated with the Ser2P elongating/terminating form of Pol II, although the interaction could only be detected in the presence of DNase, which liberates chromatin-associated proteins (**Figure 4.5B**), and not RNase (**Figure 4.5C**). However, I was not able to detect an interaction in two following replicates, nor through co-

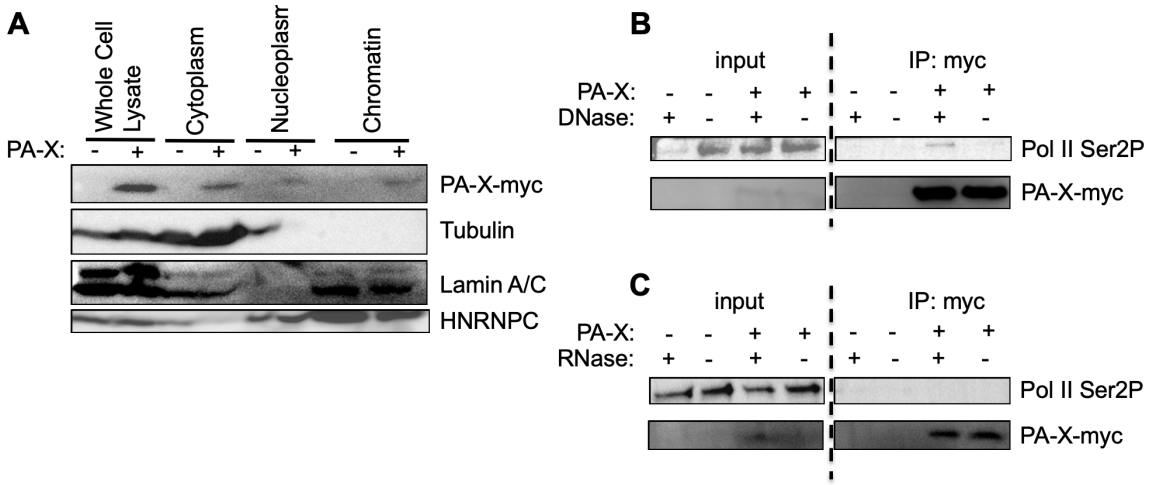

Figure 4.5: PA-X may be in complex with elongating Pol II.
(A) PA-X-myc expressing human embryonic kidney (HEK)293T cells were fractionated then run on a western blot to observe localization of PA-X. Tubulin was used as a cytoplasmic marker, Lamin A/C as a chromatin marker, and HNRNPC as a general nuclear marker. (B-C) PA-X-myc was immunoprecipitated from the nuclei of PA-X-myc expressing HEK293T cells, with or without DNase (B) or RNase (C) treatment. The elongating/termination isoform of Pol II was detected using Ser2P antibodies.

immunoprecipitation of the Ser2P or Ser5P active forms of Pol II. I also attempted to perform ChIP to test whether PA-X was associated with transcription, but the results were inconclusive. It is thus still unclear whether PA-X binds Pol II and could directly interfere with transcription.

One consequence of binding an mRNA processing factor, and potentially Pol II, is that this interaction may recruit PA-X to nascent RNAs. In such a system, RNAs that are induced upon infection, such as antiviral genes, would be the ones preferentially targeted by PA-X. Consistent with this idea, in our previous RNA sequencing analysis we found that PA-X preferentially down-regulated genes involved in the type I IFN pathway (**Figure 4.6**). However, these genes were not enriched in the analysis of RNAs cleaved by PA-X (**Figure 4.1C-D**). Moreover, RNA half-life combined with fractionation experiments suggest that PA-X does not degrade RNAs in the chromatin fraction (see appendix 5.2 for more details). These results argue against PA-X degrading nascent RNAs during transcription, although it is possible that RNAs are degraded so rapidly in

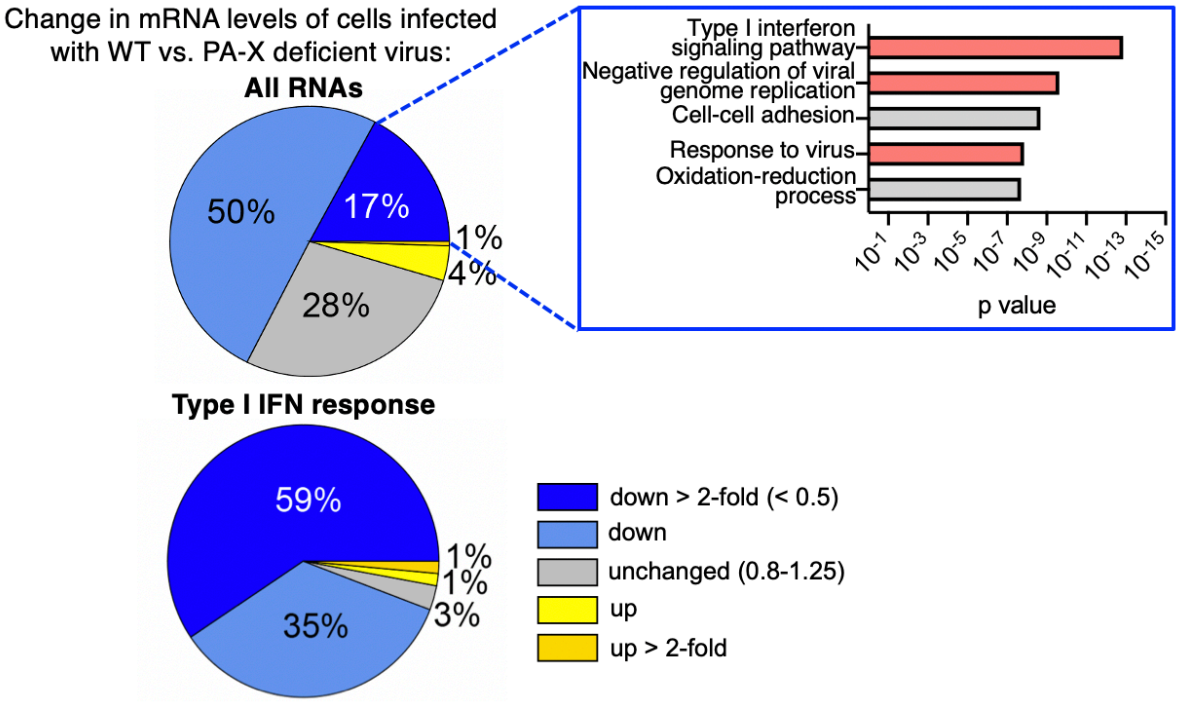

Figure 4.6: PA-X globally reduces host gene expression but preferentially targets the type I IFN pathway.
RNA sequencing from (Gaucherand et al., 2019) re-analyzed to compare cells infected with IAV PR8 WT vs. ΔX. Top: data for all RNAs; bottom: RNAs in the MSigDB hallmark type I IFN response gene set (Liberzon et al., 2015). The underlying distribution for type I IFN is significantly different from all RNAs based on Kolgorov-Smirnoff test ($p < 0.0001$). Inset: GO term analysis of RNAs down-regulated more than 2-fold by PA-X. p value computed by the DAVID program with Benjamini correction (Huang et al., 2009a, 2009b). Antiviral immunity pathways are shown in red to highlight their enrichment.

the chromatin fraction that I was not able to detect any change. Moreover, one could still imagine a model where PA-X is recruited to these RNAs after they are processed or exported to the cytoplasm. This model would still be consistent with fully cytoplasmic PA-X mutants being inactive and fully cytoplasmic RNAs not being targeted by PA-X (Hayashi et al., 2016; Khaperskyy et al., 2016; Oishi et al., 2015), as PA-X would still need to travel to the nucleus to find its target RNAs. Moreover, I have always been puzzled by the fact that microscopy images show PA-X predominantly in the nucleus, whereas fractionation experiments by our group and others show similar amounts of PA-X in the cytoplasm and the nucleus (**Figure 4.5A**) (Hayashi et al., 2016). A mechanism of action that involves shuttling may explain this discrepancy, although it could also stem

from limitations of the two techniques. Alternatively, since splicing is also important for export of mRNAs (Khan et al., 2022), PA-X could be recruited to mRNAs by a protein involved in linking splicing to mRNA export, and cleave RNAs as they are exported to the cytoplasm. PA-X could thus be located at the nuclear pore and appear nuclear by microscopy but could leak out of the nucleus during the fractionation process. Both hypotheses would be interesting to test to really pinpoint the molecular mechanism of action of PA-X.

4.3. Implications of cleaving RNA with a sequence and structure specificity

The determination of where an RNase cleaves RNA is based on where it binds the RNA and/or on the sequence and structure of the RNA at the cut site. This is true for host shutoff viral RNases. For example, HSV vhs is thought to cleave RNA near the translation initiation site because of its binding to translation initiation factors (Shiflett and Read, 2013), or near AREs when recruited to ARE-containing mRNAs (Esclatine et al., 2004a). Conversely, KSHV SOX recognizes a specific motif and cleaves RNA at a specific sequence and structure (Gaglia et al., 2015; Mendez et al., 2018). Similarly, PA in complex with PB1 and PB2 binds the 5’ cap, which restricts RNA cleavage 10-14 bp downstream of the 5’ cap (Fodor and Velthuis, 2020), while PA-X preferentially cleaves at the specific GCUG sequence when this sequence is found within the loop of a hairpin structure. Similar examples can be found within cellular RNases. For example, the cellular RNase DICER involved in micro (mi)RNA and small interfering (si)RNA generation binds to specific structures in the RNA, leading to a specific cleavage site (MacRae et al., 2007; Nguyen et al., 2022). The fungal α–sarcin cleaves 28S ribosomal RNA at GAGA sequences within loops (Endo et al., 1990). Cutting an RNA based on its specific sequence and structure *vs.* where the RNase binds the RNA could have different targeting implications. For example, using a degenerate sequence and structure

specificity means a wide range of RNAs can be targeted. In contrast, having a combination of binding to a specific protein or location on the RNA and cleaving at a specific sequence and structure could bring about more specificity to ensure that only useful RNAs are cut. This level of specificity is likely beneficial for a virus to reach a balance between cleaving as many RNAs as possible without wasting precious resources on cleaving RNAs that do not need to be cleaved.

It is interesting that the KSHV RNase SOX, which has a PD-D/E-XK nuclease fold like PA-X, also cuts RNA at a specific sequence within the loop of a hairpin structure (Gaglia et al., 2015). This specificity is not just a characteristic of PD-D/E-XK nucleases, as the human endoribonuclease MCPIP1/Regnase-1 also cleaves RNAs within a hairpin loop structure but its RNase activity is due to a PIN domain (Wilamowski et al., 2018). Cutting RNA within exposed loops may thus be a common easy way for RNases to access single-stranded RNA. Accordingly, other host shutoff RNases may also cut RNA at a specific sequence within hairpin loop structures. To study this question, I have carried out 5' RACE-seq on cells expressing HSV-1 vhs, SARS-CoV nsp1, SARS-CoV-2 nsp1, or RFP as a negative control. My results are still preliminary, but they argue for a specific sequence at the cleavage site for each protein (see appendix 5.3 for more details). More work is needed to validate these cut sites and investigate whether the cleavages are truly sequence specific and/or whether a structure is also involved. Nonetheless, it is interesting to think that all these RNases have evolved to cleave RNA at a specific sequence, but not the same sequence. Analyzing these specific sequences in each context could inform us on the biological consequences of having this specificity.

4.4. Future Work

As discussed throughout this chapter, many questions are still unanswered about the molecular mechanism of action of PA-X, and how it influences infection and disease. I will summarize the most important ones here and highlight their implications.

4.4.1. Does PA-X activity directly and/or indirectly inhibit host transcription?

Since both RNA degradation and transcription may be altered during host shutoff, measuring steady state levels may not appropriately capture the dynamics of direct RNA degradation by PA-X. However, one should first test whether transcription is indeed inhibited following PA-X overexpression or as a consequence of PA-X activity during infection. If so, conclusions about the degradation of specific RNAs by PA-X should be drawn only from assays that can directly probe RNA degradation, or separate RNA degradation from transcription. Here we have used 5' RACE-seq to directly identify PA-X cut sites throughout the transcriptome, but other techniques may also be appropriate to directly analyze RNA fragments. Other options can be to block transcription and follow RNA degradation over time, or to label nascent RNAs to differentiate them from pre-existing RNAs that are being degraded. Conversely, general transcription inhibition could be investigated directly through ChIP-seq experiments or other high throughput transcription monitoring methods. If transcription is indeed inhibited upon PA-X expression, it would also be interesting to test whether this is solely stemming from the previously described secondary consequence of host shutoff or if PA-X additionally inhibits transcription through interactions with the mRNA processing machinery. Moreover, different mRNAs could be affected by one process of transcription inhibition or the other, or both.

4.4.2. How does the mechanism of action of PA-X allow for immunomodulation?

In view of the striking *in vivo* phenotype of PA-X deficient viruses, PA-X activity is able to somehow regulate inflammation and the innate immune response. Yet, how PA-X achieves this specificity is currently unknown. While antiviral genes such as the type I IFN pathway are particularly down-regulated by PA-X (**Figure 4.6**), I did not find an enrichment for antiviral RNAs being cleaved by PA-X (**Figure 4.1C-D**). Moreover, antiviral genes do not have higher numbers of exons compared to other genes. Therefore, there is so far no obvious connection between the immune response selectivity of PA-X and its mechanism of RNA targeting. One possibility is that immunomodulation could be a secondary effect of PA-X. For example, if transcription inhibition indeed occurs following PA-X expression, this would prevent transcription of antiviral genes, even though their RNAs are not directly cleaved by PA-X. PA-X could also degrade mRNAs involved in the transcription of antiviral genes instead of the antiviral mRNA directly. For example, PA-X is known to down-regulate the histone deacetylase HDAC4, which controls the expression of a few ISGs (Galvin and Husain, 2019). Another possibility is that PA-X could directly inhibit the transcription of certain genes, including antiviral genes, for example by binding and preventing the activity of an mRNA processing protein. Alternatively, PA-X could use a yet unknown mechanism to specifically target the IFN pathway, similar to its previously reported mechanism of NF-κb inhibition (Hu et al., 2020) or through an RNase activity independent mechanism such as the suppression of Ankrd17-mediated immune responses (Li et al., 2021). More work is needed to identify the mechanism by which PA-X preferentially down-regulates antiviral genes. Moreover, it is still unclear how PA-X down-regulation of specific antiviral genes at the molecular level leads to a broadly immunomodulatory phenotype *in vivo*, since the host immune response against IAV also involves many uninfected immune cells.

One important question that also remains is why infection with a PA-X deficient virus leads to increased immunopathology and death in animal models for some IAV strains but not for others, especially since sequence analysis indicates that there is a strong evolutionary pressure to retain PA-X in all IAV strains (Shi et al., 2012). It would be interesting to compare which RNAs are preferentially targeted in each case and see if this correlates with the immune phenotype observed. One possible model is that different IAV strains use different strategies to increase transmission based on their PA-X activity, either by increasing symptoms or preventing early death of the host. While very few studies have investigated the effect of PA-X on IAV transmission, there is evidence that an optimized PA-X sequence can increase transmission. For example, truncation of the X-ORF to 41 aa increases viral transmission in pigs (Xu et al., 2016), and a specific R195K amino acid change in the PA-X protein of some avian IAV strains appears to increase transmission to mammals (Sun et al., 2020). However, no study has tested whether infection with a PA-X deficient IAV affects transmission.

4.4.3. Where PA-X is active in the cell?

PA-X is thought to cleave RNAs in the nucleus, as PA-X mutants that are fully cytoplasmic are not active and fractionation experiments show down-regulation of RNAs in the nuclear fraction (Hayashi et al., 2016; Khaperskyy et al., 2016; Oishi et al., 2015). However, some PA-X activity has been reported on transfected RNAs, which only enter the cytoplasm (Hayashi et al., 2016). Moreover, I was not able to detect direct PA-X activity in the nucleoplasm and chromatin fractions (see appendix 5.2). One limitation of all these experiments is that they were done in cells overexpressing PA-X (and with different levels of overexpression between studies), but not confirmed in infected cells where levels of PA-X may be lower. I also previously mentioned the discrepancy between microscopy images that place PA-X primarily in the nucleus, and fractionation

experiments that show equal distribution of PA-X in the nucleus and cytoplasm (**Figure 4.5A**) (Hayashi et al., 2016; Khaperskyy et al., 2016). It would thus be important to more thoroughly investigate the location of PA-X inside the cell, and where RNA cleavage occurs. High resolution microscopy could be used along RNA FISH to pinpoint the precise location of PA-X in relation to target RNAs. Another possibility would be to carry out a version of a proximity ligation assay with one probe attached to an antibody recognizing PA-X and the second probe attached to a DNA oligonucleotide that hybridizes to a target RNA (Zhang et al., 2016). This method would allow to detect transient interactions between PA-X and RNA and localize them within the cell. Determining where PA-X is active within the cell may also provide information on the role of splicing in PA-X activity. Finally, better controlled BioID interactome studies could be used to identify PA-X interactors that are critical for activity or that explain the localization of PA-X within the cell.

4.4.4. What is the cleavage specificity of other viral RNases?

Our transcriptome-wide study of PA-X cut sites has shown that the cut site sequence and structure specificity of an RNase can play a critical functional role, allowing the preferential targeting of host RNAs for PA-X. Moreover, our PyDegradome pipeline has proven an easy and effective method to identify this cleavage specificity. It will thus be important to continue identifying the cleavage specificity of other viral and host RNases. My preliminary data suggests PyDegradome was able to identify specific vhs and nsp1 cut sites throughout the transcriptome (see appendix 5.3). These sites will still need to be experimentally validated to confirm they are true cut sites. In addition, the requirements for the specific sequences enriched at the cut sites will also need to be tested using reporters, as well as potential structural requirements. Finally, these results will be critical to settle whether vhs only cleaves near the translations start site, as was

shown *in vitro* (Shiflett and Read, 2013), or if vhs also cleaves at other locations inside the RNA, as was suggested in cells expressing vhs (Gaglia et al., 2012). Finally, since the PyDegradome analysis was performed in an overexpression system, it will also be important to validate our results in infected cells and in a more biologically relevant cell type where the viral protein levels may be lower or differentially expressed over time as a normal part of the viral life cycle.

Prior to these studies, little was known about the mechanism of action of PA-X at the molecular level. Overall, my work has uncovered a two-step selection process by which PA-X effectively targets host mRNAs to induce host shutoff. Our finding that PA-X uses its cleavage specificity to distinguish between host and viral mRNAs was particularly surprising and exciting, as it shows that IAV can discriminate self *vs.* non-self similar to host cells. Future work should focus on completing this mechanism to determine how PA-X is physically recruited to its target RNAs, where the cleavage activity occurs, and what consequences ensue in terms of transcription inhibition and how this mechanism is linked to the striking immunomodulatory role of PA-X *in vivo*.

4.5. Author contribution

Rachel Levene carried out the co-immunoprecipitation experiments pictured in Figure 4.4B-C that I used to stain for Pol II. Marta Gaglia collected and analyzed the RNAseq data used to plot Figure 4.5. All other experiments and analyses were conducted by Lea Gaucherand.

Chapter 5. Appendix

5.1. PA-X activity is not linked to the consensus splice site sequence

Through transcriptomic studies, we identified a link between PA-X activity and splicing. In addition, our experimental validation data using IFN-λ2 as a reporter hinted at the 5' splice site being important for PA-X targeting. Indeed, while an intronless IFN-λ2 reporter was not efficiently down-regulated by PA-X, the addition of introns 1, 3 or 5 was sufficient to rescue down-regulation of the mRNA by PA-X (**Figures 2.8E, 2.9D**). These introns have a 5' splice site that matches the consensus sequence AG|GT. In contrast, the addition of introns 2 or 4, whose 5' splice site does not match the consensus, did not lead to mRNA down-regulation by PA-X unless their 5' splice site sequence was mutated to AG|GT (**Figures 2.8F, 2.9G**). Since the 5' splice site sequence dictates the binding efficiency of some splicing factors, these results could indicate a mechanistic link between how PA-X targets RNAs and these specific splicing factors. However, having a consensus 5' splice site also increases splicing efficiency, which could help targeting by PA-X by other means. I thus wanted to investigate further this link to the splice site sequence.

To test the importance of the 5' splice site, I first tested the effect of mutating the 5' splice site sequence of intron 3 away from the consensus sequence. As expected, I found that the mutation slightly decreases the targeting of IFN-λ2 single intron construct mRNA by PA-X (**Figure 5.1A**). Of note, this mutation concurrently decreased the splicing efficiency of intron 3, as assessed by PCR across the intron (**Figure 5.1B**). Second, I tested whether the 3' splice site sequence could also have an effect on PA-X activity. The rationale behind this experiment is that if targeting is solely dependent on the 5' splice site sequence, mutating the 3' splice site should not interfere with targeting, although it may decrease splicing efficiency. I used the MaxEnt splicing score algorithm

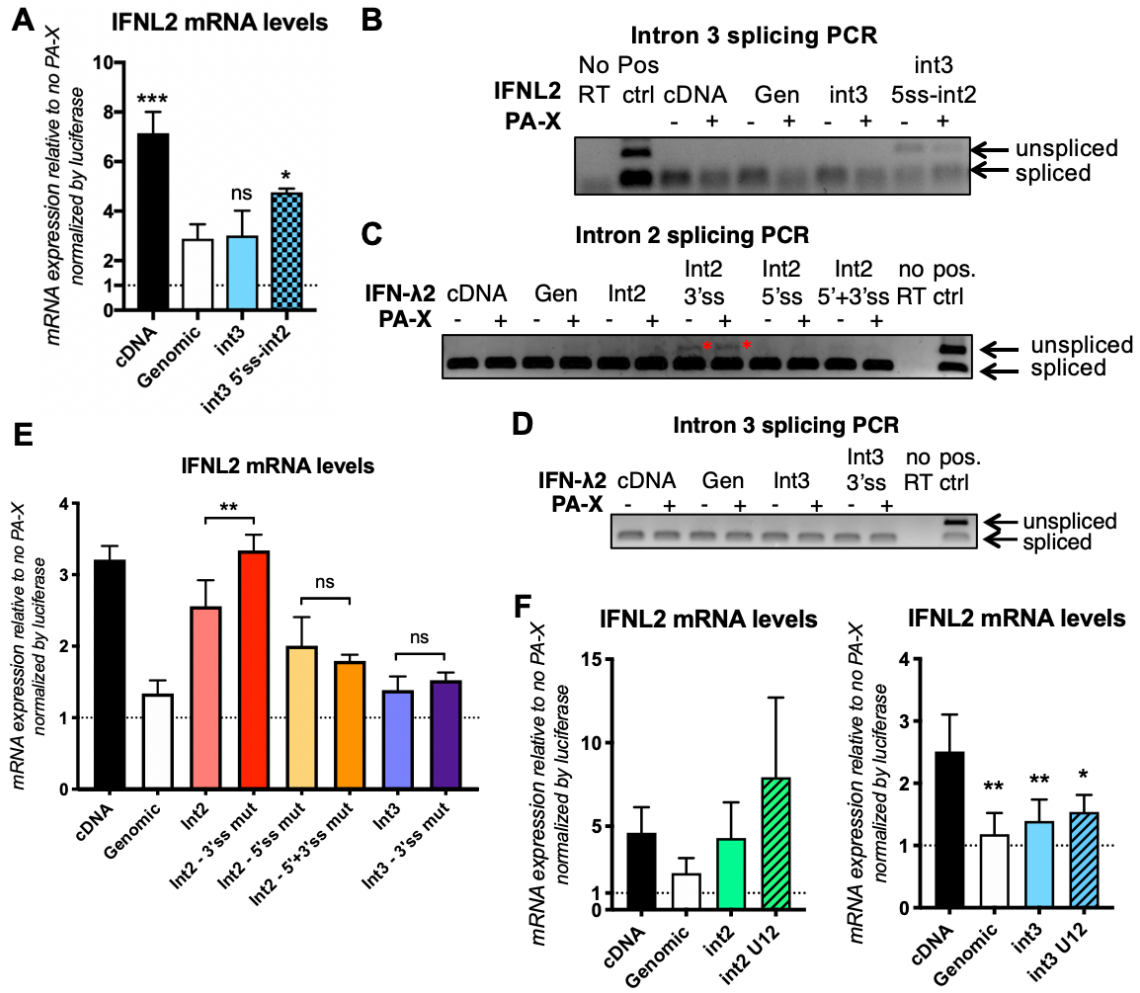

Figure 5.1: PA-X activity may be linked to splicing efficiency and not 5' splice site sequence.

*(A) mRNA levels of reporters expressing IFN-λ2 from the cDNA, genomic locus, or constructs with intron 3 with its original 5' splice site sequence or mutated to the intron 2 5' splice site sequence, normalized to luciferase expression in HEK293T cells transfected for 24hrs with PA-X. (B-D) Splicing PCR across intron 2 (C) or intron 3 (B, D) run on an agarose gel to assess splicing efficiency of each IFN-λ2 construct. (E) mRNA levels of reporters expressing IFN-λ2 from the cDNA, genomic locus, or constructs with introns 2 or 3 with and without mutations in the 5' splice site (to consensus sequence) and 3' splice site (to non-consensus sequence), normalized to luciferase expression in HEK293T cells transfected for 24hrs with PA-X. (F) mRNA levels of reporters expressing IFN-λ2 from the cDNA, genomic locus, or constructs with introns 2 (left) or 3 (right) with and without mutations in their splicing sequence to be recognized by the minor spliceosome (U12), normalized to luciferase expression in HEK293T cells transfected for 24hrs with PA-X. *, p<0.05; **, p<0.01; ns, not significant; One-way ANOVA with Dunnett's multiple comparison test. n = 3.*

(Yeo and Burge, 2004) to design mutations in the 3' splice site of 3 constructs (intron 2 only, intron 2 with the 5' splice site consensus sequence, and intron 3 only) that lower

the splice site quality score of the target introns (**Table 5.1**), which should be indicative of splicing efficiency. Down-regulation by PA-X was unchanged for most of these constructs (**Figure 5.1E**). The exception was mutating the 3' splice site of the single intron 2 construct, which lead to less down-regulation of the RNA by PA-X (**Figure 5.1E**). However, these changes are consistent with PA-X activity being linked to splicing efficiency, rather than splice site sequences. Indeed, only mutating the 3' splice site of intron 2 lead to a noticeable decrease in splicing efficiency (**Figure 5.1C**, red stars), whereas the 3' splice site mutations did not seem to affect the splicing efficiency of the other constructs (**Figure 5.1C-D**). Finally, I tested the need for a specific 5' splice site interacting protein by fully mutating the 5', branchpoint and 3' splice site sequences of introns 2 and 3 so that they would now be recognized and spliced by the minor spliceosome, as opposed to the major spliceosome. Introns recognized by the minor spliceosome represent less than 0.5% of introns in the human genome and are called U12-type introns because they can use different small nuclear RNAs and proteins for splice site recognition (Turunen et al., 2013). The U12 mutations lead to a decrease in splicing efficiency for both introns but did not significantly affect targeting of the IFN-λ2 single intron construct mRNAs by PA-X (**Figure 5.1F**). This was also supported by a preliminary analysis of our RNAseq data, in which I did not observe any difference in PA-X targeting between genes that contain introns that are spliced by the major *vs.* minor spliceosome. While I do not have the confirmation that these mRNAs are indeed spliced by the minor spliceosome, these preliminary experiments go against the need for a specific splicing factor unique to the major spliceosome to recognize the consensus 5' splice site sequence and drive targeting by PA-X.

To make sure that these results were not artifacts of using a specific gene like IFN-λ2, I repeated these experiments with two additional model transcripts, TMEM141 and PSME1, that are also targeted by PA-X based on our RNAseq results. I picked these

genes because they have 10 and 4 introns respectively, yet the size of their genomic locus is less than 3 kb so I was able to clone the entire locus in an expression plasmid. PSME1 is also an ISG, so it is relevant for IAV infection. In both cases, the spliced transcript was more down-regulated by PA-X than the intronless transcript, confirming our original finding that PA-X activity is linked to splicing (**Figure 5.2A, D**). Of note, the PSME1 cDNA construct was already efficiently targeted by PA-X. However, no PSME1 intron addition was able to rescue the level of PA-X targeting observed with the PSME1 genomic construct (**Figure 5.2E**). This result was probably because none of the introns (apart from intron 6) are spliced efficiently in the single intron constructs (**Figure 5.2F**), even though introns 5 and 7 possess the consensus 5' splice site sequence (**Table 5.2**). As for TMEM141, there was no correlation between consensus 5' splice site sequence, splicing efficiency and down-regulation by PA-X (**Table 5.2**, **Figure 5.2B-C**). Indeed, for the only introns 2 and 4 that have the consensus 5' splice site sequence, the single intron 2 construct was not spliced efficiently but efficiently down-regulated by PA-X, while the single intron 4 construct was spliced efficiently but not down-regulated by PA-X.

All these results are summarized in **Tables 5.1** and **5.2**. Overall, these experiments suggest that while PA-X activity is linked to splicing, the 5' splice site sequence is not the only determinant for PA-X targeting, and the molecular mechanism of how PA-X targets RNAs appears more complex than just binding to a spliceosome protein or having a high splicing efficiency.

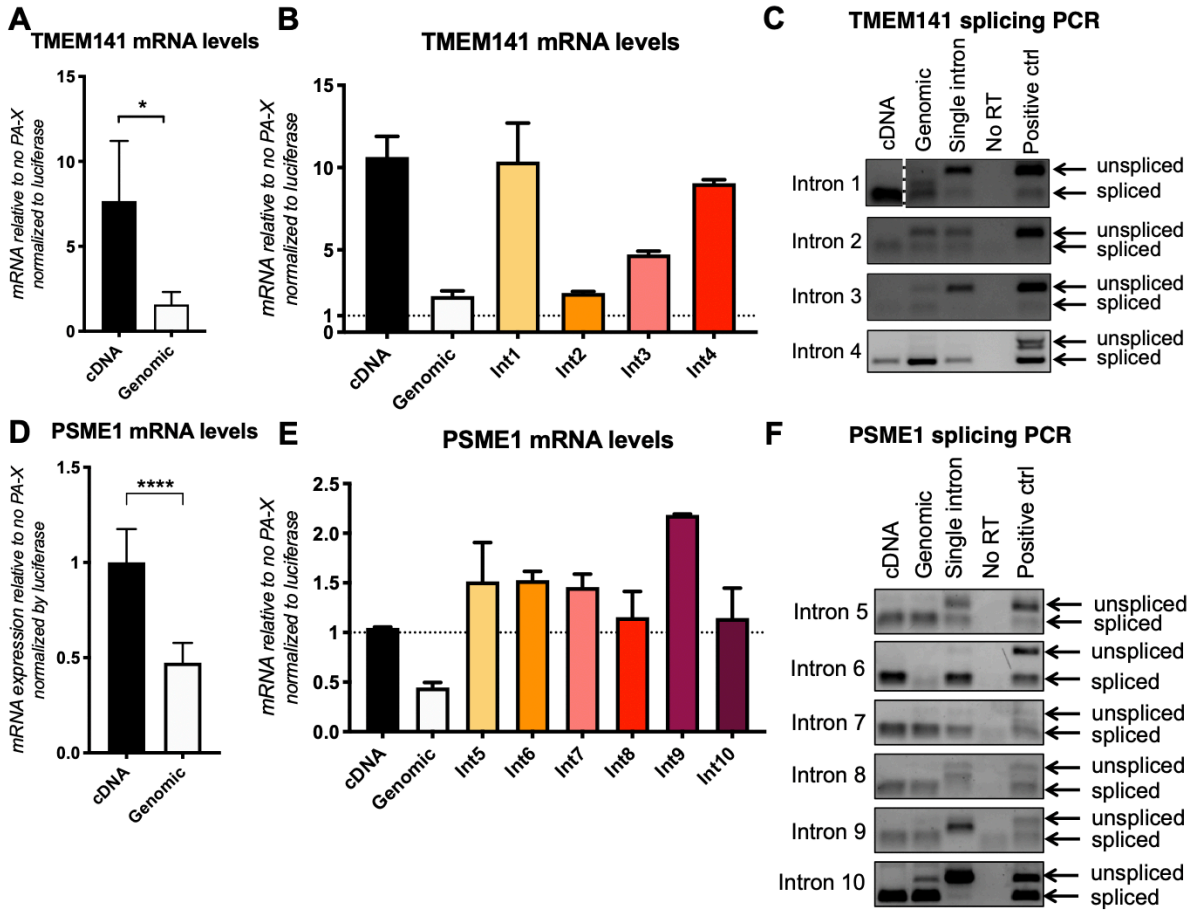

Figure 5.2: The 5' splice site sequence and splicing efficiency are not the only determinant for PA-X targeting.

*(A-B) TMEM141 mRNA levels of reporters expressing TMEM141 from the cDNA or genomic locus (n = 6) (A), or constructs with single introns (n = 2) (B), normalized to luciferase expression in HEK293T cells transfected for 24hrs with PA-X. (D-E) PSME1 mRNA levels of reporters expressing PSME1 from the cDNA or genomic locus (n = 4) (D), or constructs with single introns (n = 2) (E), normalized to luciferase expression in HEK293T cells transfected for 24hrs with PA-X. *, p<0.05; ****, p<0.0001; Unpaired T test. (C, F) Splicing efficiency for each intron as assessed by splicing PCR across introns then run on an agarose gel. n = 2.*

Table 5.1: Summary of IFN- λ2 constructs and results.

Constructs	5' ss sequence	5' Splicing score*	3' ss sequence	3' Splicing score*	Splicing *vs* Gen (gel)	PA-X targeting
PSME1 Int5	AAG|GT	8.4	AG|GTC	9.0	Worse	Less than cDNA
PSME1 Int6	CTG|GT	8.7	AG|GTC	10.1	Worse	Less than cDNA
PSME1 Int7	CAG|GT	9.2	AG|GAG	11.8	Same	Less than cDNA
PSME1 Int8	TAA|GT	6.4	AG|GTA	10.5	Worse	Same as cDNA
PSME1 Int9	GTG|GT	7.2	AG|GGT	4.1	Worse	Less than cDNA
PSME1 Int10	TAT|GT	2.4	AG|GCT	8.6	Worse	Same as cDNA
TMEM141 Int1	CCG|GT	8.2	AG|GGA	7.1	Worse	No
TMEM141 Int2	CAG|GT	9.9	AG|GCA	5.3	Same	Yes
TMEM141 Int3	TGG|GT	3.8	AG|TTG	8.4	Worse	Yes
TMEM141 Int4	CAG|GT	9.2	AG|ATC	12.4	Same	No

*As assessed by the MaxEnt splicing score program (Yeo and Burge, 2004).

Table 5.2: Summary of PSME1 and TMEM141 constructs and results.

IFN-λ2 construct	5' ss sequence	5'Splicing score*	3' ss sequence	3'Splicing score*	Splicing *vs* Genomic (gel)	Splicing *vs* Genomic (qPCR)	PA-X targeting
Int1	TAG|GT	8.8	AG|ACA	7.5	Same	*Not tested*	Yes
Int2	TTA|GT	5.8	AG|GAA	9.2	Same	Worse	No
Int2 5' mut	**CAG**|GT	**10.7**	AG|GAA	9.2	Same	Better	Yes
Int2 3' mut	TTA|GT	5.8	AG|**C**AA	**6.5**	Worse	Worse	No
Int2 5'+3' mut	**CAG**|GT	**10.7**	AG|**C**AA	**6.5**	Same	Same	Yes
Int2 U12	**AAG|AT**		**AC|GTT**		*Not tested*	Worse	No
Int3	CAG|GT	9.2	AG|GTG	11.3	Same	Same	Yes
Int3 5'mut-int2	**TTA**|GT	**5.8**	AG|GTG	11.3	Worse	Worse	Less
Int3 3' mut	CAG|GT	9.2	AG|**CAA**	**7.7**	Same	Same	Yes
Int3 U12	**AAG|AT**		**AC|GTT**		*Not tested*	Worse	Yes
Int4	TGT|GT	5.8	AG|ATC	12.3	Worse	Worse	No
Int4 5' mut	**CAG**|GT	**10.7**	AG|ATC	12.3	Same	Worse	Yes
Int5	AAG|GT	10.5	AG|GAG	11.2	Same	*Not tested*	Yes

*As assessed by the MaxEnt splicing score program (Yeo and Burge, 2004).

5.2. PA-X does not appear to degrade RNAs in the chromatin fraction

Microscopy images show that PA-X is primarily found within the nucleus (Khaperskyy et al., 2016). While fractionation experiments show PA-X both in the nucleus and the cytoplasm, PA-X mutants that are primarily cytoplasmic have decreased or no activity (Hayashi et al., 2016; Khaperskyy et al., 2016; Oishi et al., 2018a, 2015). Moreover, PA-

X expression leads to down-regulation of mRNAs in the nucleus (Hayashi et al., 2016; Khaperskyy et al., 2016). These results strongly suggest a nuclear activity of PA-X. More specifically, our data indicate that PA-X activity is linked to splicing, which is mostly co-transcriptional, and that PA-X interacts with the CFIm complex, which is involved in splicing and cleavage and polyadenylation (Gaucherand et al., 2019). In addition, my preliminary data indicate that some PA-X can be found within the chromatin fraction (**Figure 4.5A**) and may interact with elongating/terminating Pol II (**Figure 4.5B**). I thus hypothesized that PA-X was active in the chromatin fraction and actively degrading nascent RNAs during transcription.

To study this question, I induced PA-X expression from an integrated doxycycline-inducible transcript in human embryonic kidney (HEK)293T cells for 4 hours. I then blocked transcription with actinomycin D and followed the degradation of mRNAs over time in the cytoplasmic, nucleoplasmic and chromatin fraction (**Figure 5.3A**). I confirmed efficient fractionation using GAPDH mRNA as a cytoplasmic marker, MALAT1 long non-coding (lnc) RNA as a nuclear marker and GAPDH pre-mRNA as a chromatin marker (**Figure 5.3B**). As expected, the mRNA levels of the PA-X target G6PD decreased faster in the presence than in the absence of PA-X in the total fraction, while PA-X did not affect the PA-X resistant intronless TERC lncRNA (**Figure 5.3C**, total). The same was true in the cytoplasmic fraction (**Figure 5.3C**, cytoplasm). Degradation in the nucleoplasm was hard to assess due to nuclear retention of mRNAs, a known secondary consequence of PA-X activity (Khaperskyy et al., 2014). Indeed, increased mRNA levels are seen in the nucleoplasm in the presence of PA-X for both G6PD and TERC RNA (**Figure 5.3C**, nucleoplasm). However, there was no difference in RNA half-life in the presence of PA-X in the chromatin fraction (**Figure 5.3C**, chromatin). To check if this was because PA-X was active on pre-mRNA instead of mature mRNA, I used one

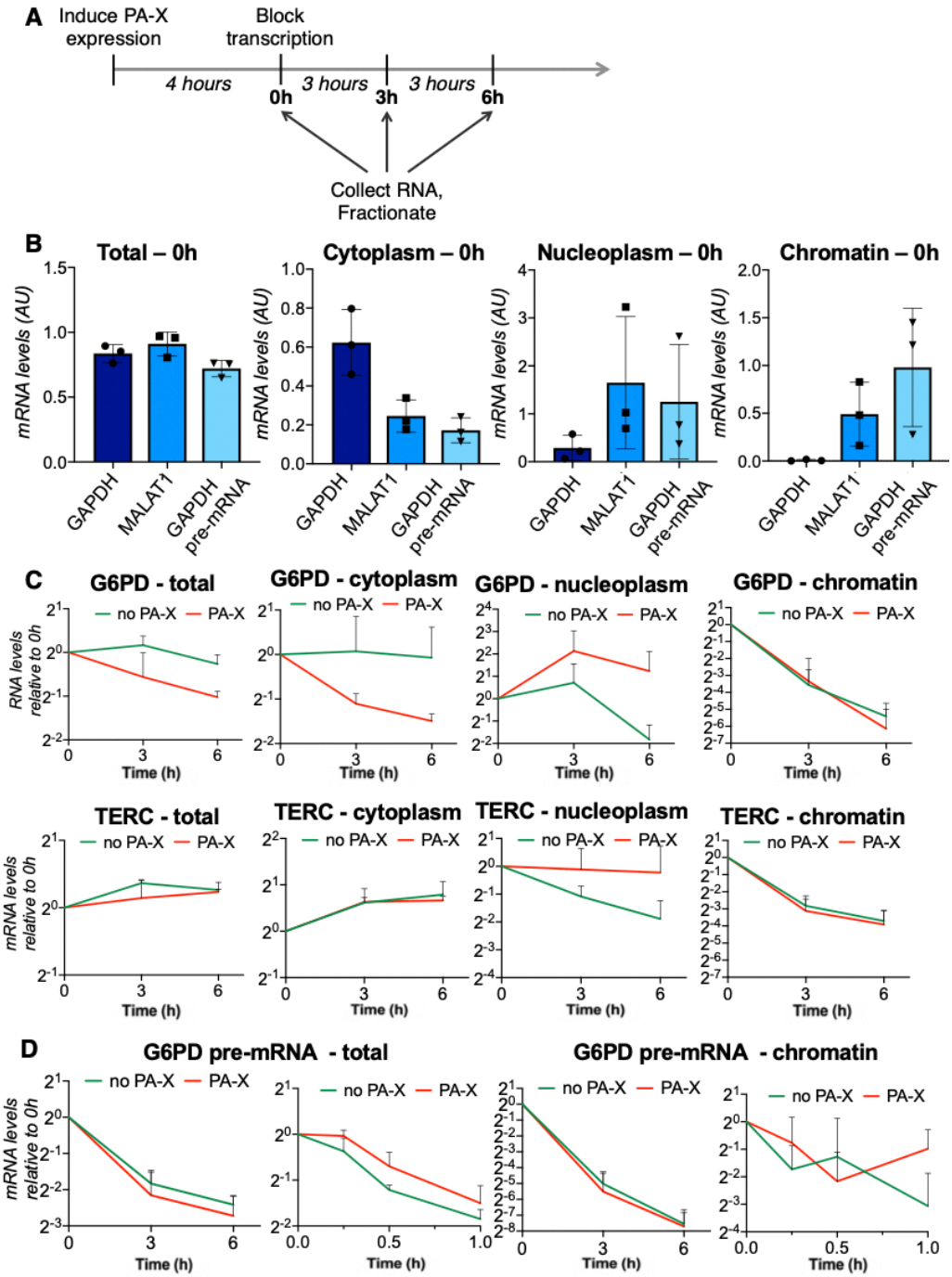

Figure 5.3: PA-X does not appear to degrade RNAs in the chromatin fraction.
(A) Experimental setup: PA-X inducible HEK293T cells were treated with no or low levels of doxycycline for 4h to induce PA-X expression, then treated with actinomycin D to block transcription. Cells were collected at 0h, 3h and 6h after addition of actinomycin D, then fractionated. RNA was extracted and levels were measured by RT-qPCR. (B) Marker RNA levels for each fraction at 0h without PA-X to control for efficient fractionation: GAPDH for cytoplasm, MALAT1 for nucleoplasm and GAPDH pre-mRNA (with one primer in an intron and one primer in an exon) for chromatin. (C-D) RNA levels over time in each fraction for PA-X targets G6PD (C), G6PD pre-mRNA (C) or PA-X

resistant TERC (D), normalized to levels without PA-X. n = 3 except for 1h time course n = 2.

intronic and one exonic primer to measure G6PD pre-mRNA levels. However, there was also no difference when measuring G6PD pre-mRNA half-lives in both the total and chromatin fractions, even at earlier time points after addition of actinomycin D (**Figure 5.3D**).

Overall, these results suggest that PA-X does not directly degrade RNAs in the chromatin fraction but appears active in the cytoplasmic fraction. However, there are caveats to this interpretation, especially since chromatin-associated RNA is already very short-lived. Nuclear retention of RNAs could be partly responsible for the decrease in cytoplasmic RNAs observed upon PA-X expression. The previously observed nuclear down-regulation of RNAs could also come from a block in transcription, either as a secondary consequence of RNA depletion in the cytoplasm or through a yet unknown additional effect of PA-X on transcription. Alternatively, these experiments could point to a model where PA-X is loaded on the RNA in the nucleus, but is only active in the cytoplasm. One important question would also be to know how fast RNAs are degraded by host exonucleases after PA-X cleavage, as this could influence our fractionation results.

5.3. Preliminary cut site preference results for vhs and nsp1

Our results indicate that the two viral host shutoff RNases KSHV SOX and IAV PA-X cleave RNAs at a preferred sequence and structure (Gaglia et al., 2015). In addition, cut site specificity can play an important functional role, such as to distinguish between host and viral mRNAs in the case of PA-X. I thus hypothesized that other viral RNases would also cut RNAs at a preferred sequence and/or structure. To test this, I selected the HSV-1 RNase vhs, as well as the SARS-CoV and SARS-CoV-2 nsp1 proteins, which are

thought to induce host shutoff through an unknown host RNase. I overexpressed either of these viral proteins, or RFP as negative control, in HEK293T cells for 24 hours, then collected RNA and carried out 5' RACE-seq to identify cut sites transcriptome wide.

I first confirmed that the viral proteins were active by monitoring mRNA and protein levels of a co-transfected GFP reporter. GFP mRNA was indeed down-regulated upon expression of vhs and nsp1 compared to the control RFP expressing cells (**Figure 5.4A**). As a result, GFP protein levels were also reduced upon expression of vhs and nsp1 (**Figure 5.4B**). Of note, in the case of nsp1, the observed decrease in protein levels likely comes from both mRNA degradation and translation inhibition (Narayanan et al., 2015; Yuan et al., 2021). While host shutoff by SARS-CoV-2 nsp1 appeared stronger than by SARS-CoV nsp1, this could be due to a lower level of SARS-CoV nsp1 expression (**Figure 5.4B**). I then used our PyDegradome pipeline to identify cut sites transcriptome-wide. Of note, the results depicted here were obtained using a faster version of PyDegradome that only looks for cut sites in the transcriptome, since RNA cleavage upon expression of vhs and nsp1 primarily occurs in the cytoplasm. Similar results were obtained using the version of PyDegradome used in Chapter 3 that looks for cut sites in both exons and introns. Using the shared cut sites approach (**Figure 3.2D**), PyDegradome identified 1800, 648 and 1818 high confidence cut sites following vhs, SARS-CoV nsp1 and SARS-CoV-2 expression, respectively, that were shared amongst three biological replicates (**Figure 5.4C**). In contrast, a lot fewer RFP specific fragments were shared amongst the three biological replicates, increasing our confidence that PyDegradome identified true cut sites (**Figure 5.4D**). Interestingly, the two different nsp1 samples had a high percentage of overlap in their cut sites, while there was almost no overlap between nsp1 and vhs cut sites (**Figure 5.4E**). This result is consistent with both nsp1 proteins activating the same host RNase to induce host shutoff, and again increases our confidence that we have identified true cut sites. As a

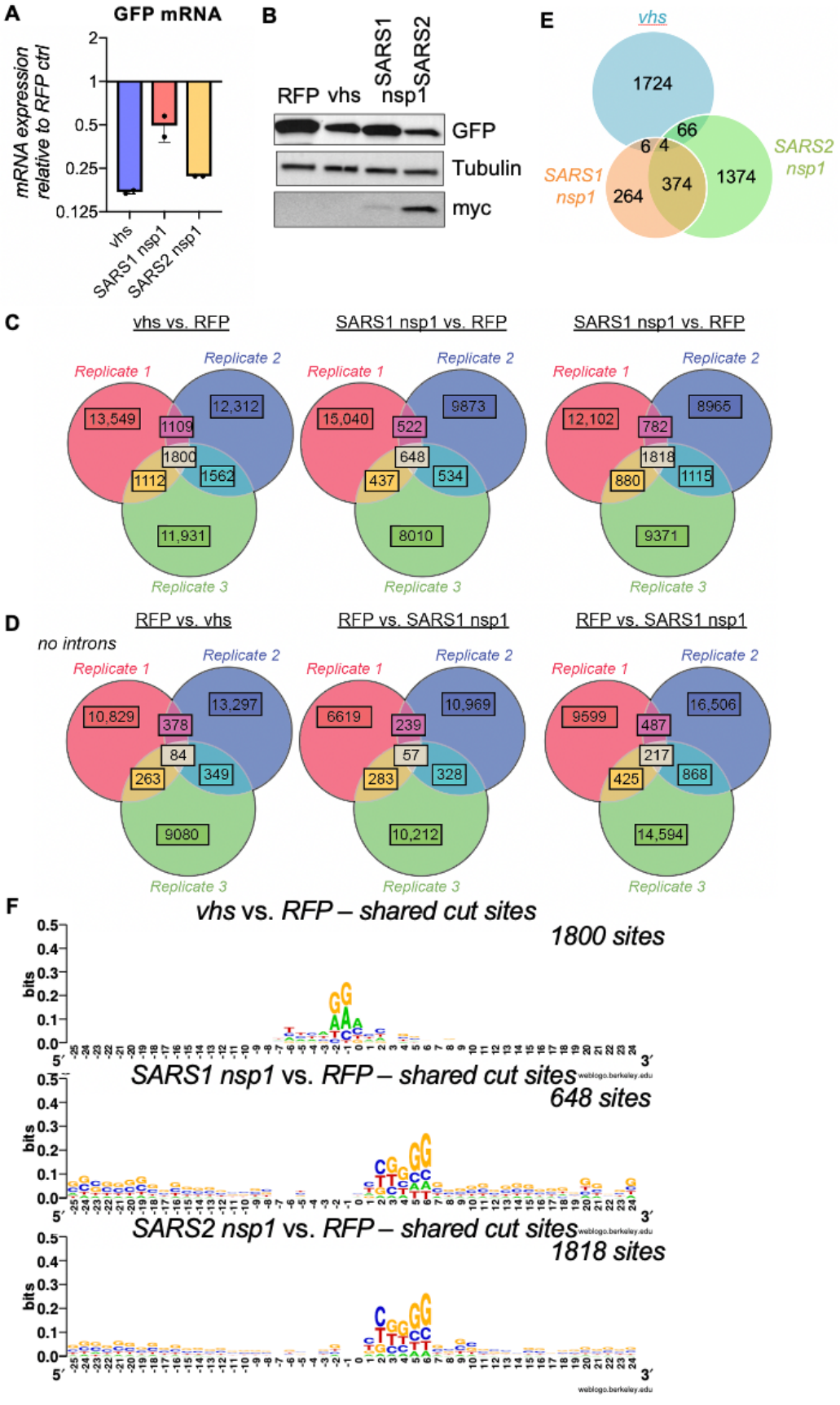
A
GFP mRNA
mRNA expression relative to RFP ctrl
2
1
0.5
0.25
0.125
vhs
SARS1 nsp1
SARS2 nsp1
B
RFP vhs
SARS1 nsp1
SARS2
GFP
Tubulin
myc
E
vhs
1724
66
6 4
SARS1 nsp1
264
374
1374
SARS2 nsp1
C
vhs vs. RFP
SARS1 nsp1 vs. RFP
SARS1 nsp1 vs. RFP
Replicate 1
Replicate 2
Replicate 3
13,549
1109
12,312
1800
1112
1562
11,931
15,040
522
9873
648
437
534
8010
12,102
782
8965
1818
880
1115
9371
D
RFP vs. vhs
RFP vs. SARS1 nsp1
RFP vs. SARS1 nsp1
no introns
10,829
378
13,297
84
263
349
9080
6619
239
10,969
57
283
328
10,212
9599
487
16,506
217
425
868
14,594
F
vhs vs. RFP – shared cut sites
1800 sites
SARS1 nsp1 vs. RFP – shared cut sites
648 sites
SARS2 nsp1 vs. RFP – shared cut sites
1818 sites
bits
5′
3′
weblogo.berkeley.edu

Figure 5.4: Preliminary studies suggest that vhs and nsp1 expression lead to sequence specific cut sites.
Inducible shXrn1 HEK293T cells were treated with 0 or 1µg/ml doxycycline for 3-4 days to knock down Xrn1, then transfected with the HSV-1 RNase vhs, the SARS-CoV or SARS-CoV-2 nsp1 protein that activates an unknown host RNase, or RFP as control. GFP was also co-transfected to monitor host shutoff activity. (A) In cells not treated with doxycycline, GFP mRNA levels were measured by qRT-PCR, normalized by 18S and plotted as fold change compared to RFP control. n = 2 (B) In cells with Xrn1 knock down, GFP protein levels were measured by western blot. Anti myc antibodies were used to confirm myc-tagged nsp1 expression. Image representative of 2 experiments. (C-D) Venn diagrams of the number of cut sites identified by PyDegradome for each replicate. (C) Left diagram shows vhs specific fragments, middle diagram SARS-CoV nsp1 specific fragments and right diagram SARS-CoV-2 nsp1 specific fragments. (D) As controls, RFP specific fragments were also identified as compared to vhs (left), SARS-CoV nsp1 (middle) and SARS-CoV-2 nsp1 (right). (E) Venn diagram showing the number of cut sites shared between vhs, SARS-CoV nsp1 and SARS-CoV-2 nsp1. (F) WebLogo (Crooks et al., 2004) representation of base enrichment around vhs (top), SARS-CoV nsp1 (middle) or SARS-CoV-2 nsp1 (bottom) for shared cut sites predicted by PyDegradome.

consequence, similar bases were enriched around the cut sites for the two nsp1 proteins, with a “CGGGG” motif right downstream of the cleavage site (**Figure 5.4F**, middle and bottom). Long GC-rich stretches were also observed on either side of this motif, which could be indicative of a base-pairing structure (**Figure 5.4F**, middle and bottom). In contrast, the bases “GA” or “GGA” were enriched right before the cleavage site in vhs shared cut sites (**Figure 5.4F**, top).

While still preliminary, these results argue for at least a sequence preference at the cut site for both vhs and the host RNase induced by nsp1. Future work will need to validate these results experimentally, as well as investigate the need for a specific structure and the potential functional importance of these preferred cut site characteristics.

5.4. Author contribution

All experiments and analyses in this chapter were conducted by Léa Gaucherand.

Chapter 6. Bibliography

Abernathy E, Clyde K, Yeasmin R, Krug LT, Burlingame A, Coscoy L, Glaunsinger B. 2014. Gammaherpesviral gene expression and virion composition are broadly controlled by accelerated mRNA degradation. *PLoS Pathog* **10**:e1003882. doi:10.1371/journal.ppat.1003882

Abernathy E, Gilbertson S, Alla R, Glaunsinger B. 2015. Viral Nucleases Induce an mRNA Degradation-Transcription Feedback Loop in Mammalian Cells. *Cell Host & Microbe* **18**:243–253. doi:10.1016/j.chom.2015.06.019

Almeida MS, Johnson MA, Herrmann T, Geralt M, Wüthrich K. 2007. Novel beta-barrel fold in the nuclear magnetic resonance structure of the replicase nonstructural protein 1 from the severe acute respiratory syndrome coronavirus. *J Virol* **81**:3151–3161. doi:10.1128/JVI.01939-06

Alnaji FG, Brooke CB. 2020. Influenza virus DI particles: Defective interfering or delightfully interesting? *PLOS Pathogens* **16**:e1008436. doi:10.1371/journal.ppat.1008436

Anderson P, Kedersha N. 2008. Stress granules: the Tao of RNA triage. *Trends in Biochemical Sciences* **33**:141–150. doi:10.1016/j.tibs.2007.12.003

Aragón T, de la Luna S, Novoa I, Carrasco L, Ortín J, Nieto A. 2000. Eukaryotic Translation Initiation Factor 4GI Is a Cellular Target for NS1 Protein, a Translational Activator of Influenza Virus. *Molecular and Cellular Biology* **20**:6259–6268. doi:10.1128/MCB.20.17.6259-6268.2000

Arias-Mireles BH, de Rozieres CM, Ly K, Joseph S. 2018. RNA Modulates the Interaction between Influenza A Virus NS1 and Human PABP1. *Biochemistry* **57**:3590–3598. doi:10.1021/acs.biochem.8b00218

Azziz Baumgartner E, Dao CN, Nasreen S, Bhuiyan MU, Mah-E-Muneer S, Mamun AA, Sharker MAY, Zaman RU, Cheng P-Y, Klimov AI, Widdowson M-A, Uyeki TM, Luby SP, Mounts A, Bresee J. 2012. Seasonality, Timing, and Climate Drivers of Influenza Activity Worldwide. *The Journal of Infectious Diseases* **206**:838–846. doi:10.1093/infdis/jis467

Banerjee AK, Blanco MR, Bruce EA, Honson DD, Chen LM, Chow A, Bhat P, Ollikainen N, Quinodoz SA, Loney C, Thai J, Miller ZD, Lin AE, Schmidt MM, Stewart DG, Goldfarb D, Lorenzo GD, Rihn SJ, Voorhees RM, Botten JW, Majumdar D, Guttman M. 2020. SARS-CoV-2 Disrupts Splicing, Translation, and Protein Trafficking to Suppress Host Defenses. *Cell* **183**:1325-1339.e21. doi:10.1016/j.cell.2020.10.004

Basak S, Tomana M, Compans RW. 1985. Sialic acid is incorporated into influenza hemagglutinin glycoproteins in the absence of viral neuraminidase. *Virus Research* **2**:61–68. doi:10.1016/0168-1702(85)90060-7

Bauer DLV, Tellier M, Martínez-Alonso M, Nojima T, Proudfoot NJ, Murphy S, Fodor E. 2018. Influenza Virus Mounts a Two-Pronged Attack on Host RNA Polymerase II Transcription. *Cell Reports* **23**:2119-2129.e3. doi:10.1016/j.celrep.2018.04.047

Bavagnoli L, Cucuzza S, Campanini G, Rovida F, Paolucci S, Baldanti F, Maga G. 2015. 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. doi:10.1093/nar/gkv926

Bercovich-Kinori A, Tai J, Gelbart IA, Shitrit A, Ben-Moshe S, Drori Y, Itzkovitz S, Mandelboim M, Stern-Ginossar N. 2016. A systematic view on influenza induced host shutoff. *eLife* **5**:e18311. doi:10.7554/eLife.18311

Bessman MJ, Frick DN, O'Handley SF. 1996. The MutT proteins or "Nudix" hydrolases, a family of versatile, widely distributed, "housecleaning" enzymes. *J Biol Chem* **271**:25059–25062. doi:10.1074/jbc.271.41.25059

Bestle D, Limburg H, Kruhl D, Harbig A, Stein DA, Moulton H, Matrosovich M, Abdelwhab EM, Stech J, Böttcher-Friebertshäuser E. 2021. Hemagglutinins of Avian Influenza Viruses Are Proteolytically Activated by TMPRSS2 in Human and Murine Airway Cells. *J Virol* **95**:e0090621. doi:10.1128/JVI.00906-21

Black D, Ritchey J, Payton M, Eberle R. 2014. Role of the virion host shutoff protein in neurovirulence of monkey B virus (Macacine herpesvirus 1). *Virol Sin* **29**:274–283. doi:10.1007/s12250-014-3495-x

Bradel-Tretheway BG, Mattiacio JL, Krasnoselsky A, Stevenson C, Purdy D, Dewhurst S, Katzel MG. 2011. Comprehensive proteomic analysis of influenza virus polymerase complex reveals a novel association with mitochondrial proteins and RNA polymerase accessory factors. *Journal of Virology* **85**:8569–8581. doi:10.1128/JVI.00496-11

Brannan K, Kim H, Erickson B, Glover-Cutter K, Kim S, Fong N, Kiemele L, Hansen K, Davis R, Lykke-Andersen J, Bentley DL. 2012. MRNA Decapping Factors and the Exonuclease Xrn2 Function in Widespread Premature Termination of RNA Polymerase II Transcription. *Molecular Cell* **46**:311–324. doi:10.1016/J.MOLCEL.2012.03.006/ATTACHMENT/C830929C-3EC3-4913-9183-4F88722FC95B/MMC2.XLS

Buisson M, Géoui T, Flot D, Tarbouriech N, Ressing ME, Wiertz EJ, Burmeister WP. 2009. A bridge crosses the active-site canyon of the Epstein-Barr virus nuclease with DNase and RNase activities. *J Mol Biol* **391**:717–728. doi:10.1016/j.jmb.2009.06.034

Burgess HM, Mohr I. 2018. Defining the Role of Stress Granules in Innate Immune Suppression by the Herpes Simplex Virus 1 Endoribonuclease VHS. *J Virol* **92**:e00829-18. doi:10.1128/JVI.00829-18

Burgess HM, Mohr I. 2015. Cellular 5'-3' mRNA exonuclease Xrn1 controls double-stranded RNA accumulation and anti-viral responses. *Cell Host Microbe* **17**:332–344. doi:10.1016/j.chom.2015.02.003

Burgui I, Aragón T, Ortín J, Nieto A. 2003. PABP1 and eIF4GI associate with influenza virus NS1 protein in viral mRNA translation initiation complexes. *J Gen Virol* **84**:3263–3274. doi:10.1099/vir.0.19487-0

Burke JM, St Clair LA, Perera R, Parker R. 2021. SARS-CoV-2 infection triggers widespread host mRNA decay leading to an mRNA export block. *RNA* **27**:1318–1329. doi:10.1261/rna.078923.121

Carrique L, Fan H, Walker AP, Keown JR, Sharps J, Staller E, Barclay WS, Fodor E, Grimes JM. 2020. Host ANP32A mediates the assembly of the influenza virus replicase. *Nature* **587**:638–643. doi:10.1038/s41586-020-2927-z

Cartwright JL, Safrany ST, Dixon LK, Darzynkiewicz E, Stepinski J, Burke R, McLennan AG. 2002. The g5R (D250) gene of African swine fever virus encodes a Nudix hydrolase that preferentially degrades diphosphoinositol polyphosphates. *J Virol* **76**:1415–1421. doi:10.1128/jvi.76.3.1415-1421.2002

CDC. 2022a. Burden of Influenza. Available at: https://www.cdc.gov/flu/about/burden/index.html. *Centers for Disease Control and Prevention*. https://www.cdc.gov/flu/about/burden/index.html

CDC. 2022b. CDC Seasonal Flu Vaccine Effectiveness Studies. Available at: https://www.cdc.gov/flu/vaccines-work/effectiveness-studies.htm. https://www.cdc.gov/flu/vaccines-work/effectiveness-studies.htm

Chaimayo C, Dunagan M, Hayashi T, Santoso N, Takimoto T. 2018. Specificity and functional interplay between influenza virus PA-X and NS1 shutoff activity. *PLOS Pathogens* **14**:e1007465. doi:10.1371/journal.ppat.1007465

Chan AY, Vreede FT, Smith M, Engelhardt OG, Fodor E. 2006. Influenza virus inhibits RNA polymerase II elongation. *Virology* **351**:210–217. doi:10.1016/j.virol.2006.03.005

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. 2017. In Vivo Cleavage Map Illuminates the Central Role of RNase E in Coding and Non-coding RNA Pathways. *Molecular Cell* **65**:39–51. doi:10.1016/j.molcel.2016.11.002

Chen W, Smeekens JM, Wu R. 2016. Systematic study of the dynamics and half-lives of newly synthesized proteins in human cells. *Chem Sci* **7**:1393–1400. doi:10.1039/C5SC03826J

Chen Z, Li Y, Krug RM. 1999. Influenza A virus NS1 protein targetspoly(A)-binding protein II of the cellular 3′-end processing machinery. *The EMBO Journal* **18**:2273–2283. doi:10.1093/emboj/18.8.2273

Ciminski K, Schwemmle M. 2021. Bat-Borne Influenza A Viruses: An Awakening. *Cold Spring Harb Perspect Med* **11**:a038612. doi:10.1101/cshperspect.a038612

Clark AM, Nogales A, Martinez-Sobrido L, Topham DJ, DeDiego ML. 2017. Functional Evolution of Influenza Virus NS1 Protein in Currently Circulating Human 2009 Pandemic H1N1 Viruses. *Journal of Virology* **91**:e00721-17. doi:10.1128/JVI.00721-17

Clements AL, Peacock TP, Sealy JE, Lee HM, Hussain S, Sadeyen J-R, Shelton H, Digard P, Iqbal M. 2021. PA-X is an avian virulence factor in H9N2 avian influenza virus. *J Gen Virol* **102**. doi:10.1099/jgv.0.001531

Cohen M, Zhang X-Q, Senaati HP, Chen H-W, Varki NM, Schooley RT, Gagneux P. 2013. Influenza A penetrates host mucus by cleaving sialic acids with neuraminidase. *Virol J* **10**:321. doi:10.1186/1743-422X-10-321

Connor RF, Roper RL. 2007. Unique SARS-CoV protein nsp1: bioinformatics, biochemistry and potential effects on virulence. *Trends in Microbiology* **15**:51–53. doi:10.1016/j.tim.2006.12.005

Costard S, Mur L, Lubroth J, Sanchez-Vizcaino JM, Pfeiffer DU. 2013. Epidemiology of African swine fever virus. *Virus Res* **173**:191–197. doi:10.1016/j.virusres.2012.10.030

Couceiro JNSS, Paulson JC, Baum LG. 1993. Influenza virus strains selectively recognize sialyloligosaccharides on human respiratory epithelium; the role of the host cell in selection of hemagglutinin receptor specificity. *Virus Research* **29**:155–165. doi:10.1016/0168-1702(93)90056-S

Covarrubias S, Gaglia MM, Kumar GR, Wong W, Jackson AO, Glaunsinger BA. 2011. Coordinated Destruction of Cellular Messages in Translation Complexes by the Gammaherpesvirus Host Shutoff Factor and the Mammalian Exonuclease Xrn1. *PLoS Pathogens* **7**:e1002339. doi:10.1371/journal.ppat.1002339

Covarrubias S, Richner JM, Clyde K, Lee YJ, Glaunsinger BA. 2009. Host shutoff is a conserved phenotype of gammaherpesvirus infection and is orchestrated exclusively from the cytoplasm. *J Virol* **83**:9554–9566. doi:10.1128/JVI.01051-09

Crooks GE, Hon G, Chandonia J-M, Brenner SE. 2004. WebLogo: a sequence logo generator. *Genome Res* **14**:1188–1190. doi:10.1101/gr.849004

Cruz A, Joseph S. 2022. Interaction of the Influenza A Virus NS1 Protein with the 5'-m7G-mRNA·eIF4E·eIF4G1 Complex. *Biochemistry* **61**:1485–1494. doi:10.1021/acs.biochem.2c00019

Dadi S, Payet-Bornet D, Ferrier P. 2013. ImmunoPrecipitation of Nuclear Protein with Antibody Affinity Columns. *Bio-protocol* **3**:e319–e319.
Dadonaite B, Vijayakrishnan S, Fodor E, Bhella D, Hutchinson EC. 2016. Filamentous influenza viruses. *J Gen Virol* **97**:1755–1764. doi:10.1099/jgv.0.000535
Dahlroth S-L, Gurmu D, Schmitzberger F, Engman H, Haas J, Erlandsen H, Nordlund P. 2009. Crystal structure of the shutoff and exonuclease protein from the oncogenic Kaposi's sarcoma-associated herpesvirus. *FEBS J* **276**:6636–6645. doi:10.1111/j.1742-4658.2009.07374.x
Dai A, Cao S, Dhungel P, Luan Y, Liu Y, Xie Z, Yang Z. 2017. Ribosome Profiling Reveals Translational Upregulation of Cellular Oxidative Phosphorylation mRNAs during Vaccinia Virus-Induced Host Shutoff. *J Virol* **91**:e01858-16. doi:10.1128/JVI.01858-16
Das K, Ma L-C, Xiao R, Radvansky B, Aramini J, Zhao L, Marklund J, Kuo R-L, Twu KY, Arnold E, Krug RM, Montelione GT. 2008. Structural basis for suppression of a host antiviral response by influenza A virus. *Proc Natl Acad Sci U S A* **105**:13093–13098. doi:10.1073/pnas.0805213105
Datta K, Wolkerstorfer A, Szolar OHJ, Cusack S, Klumpp K. 2013. Characterization of PA-N terminal domain of Influenza A polymerase reveals sequence specific RNA cleavage. *Nucl Acids Res* **41**:8289–8299. doi:10.1093/nar/gkt603
Dauber B, Poon D, Dos Santos T, Duguay BA, Mehta N, Saffran HA, Smiley JR. 2016. The Herpes Simplex Virus Virion Host Shutoff Protein Enhances Translation of Viral True Late mRNAs Independently of Suppressing Protein Kinase R and Stress Granule Formation. *J Virol* **90**:6049–6057. doi:10.1128/JVI.03180-15
Dauber B, Saffran HA, Smiley JR. 2014. The Herpes Simplex Virus 1 Virion Host Shutoff Protein Enhances Translation of Viral Late mRNAs by Preventing mRNA Overload. *J Virol* **88**:9624–9632. doi:10.1128/JVI.01350-14
Davis CW, Jackson KJL, McCausland MM, Darce J, Chang C, Linderman SL, Chennareddy C, Gerkin R, Brown SJ, Wrammert J, Mehta AK, Cheung WC, Boyd SD, Waller EK, Ahmed R. 2020. Influenza vaccine–induced human bone marrow plasma cells decline within a year after vaccination. *Science* **370**:237–241. doi:10.1126/science.aaz8432
Dawood FS, Iuliano AD, Reed C, Meltzer MI, Shay DK, Cheng P-Y, Bandaranayake D, Breiman RF, Brooks WA, Buchy P, Feikin DR, Fowler KB, Gordon A, Hien NT, Horby P, Huang QS, Katz MA, Krishnan A, Lal R, Montgomery JM, Mølbak K, Pebody R, Presanis AM, Razuri H, Steens A, Tinoco YO, Wallinga J, Yu H, Vong S, Bresee J, Widdowson M-A. 2012. Estimated global mortality associated with the first 12 months of 2009 pandemic influenza A H1N1 virus circulation: a modelling study. *The Lancet Infectious Diseases* **12**:687–695. doi:10.1016/S1473-3099(12)70121-4
Decroly E, Ferron F, Lescar J, Canard B. 2012. Conventional and unconventional mechanisms for capping viral mRNA. *Nat Rev Microbiol* **10**:51–65. doi:10.1038/nrmicro2675
Desloges N, Rahaus M, Wolff MH. 2005. The varicella–zoster virus–mediated delayed host shutoff: open reading frame 17 has no major function, whereas immediate–early 63 protein represses heterologous gene expression. *Microbes and Infection* **7**:1519–1529. doi:10.1016/j.micinf.2005.05.010
Desmet EA, Bussey KA, Stone R, Takimoto T. 2013. Identification of the N-Terminal Domain of the Influenza Virus PA Responsible for the Suppression of Host Protein Synthesis. *Journal of Virology* **87**:3108–3118. doi:10.1128/JVI.02826-12
Dharmalingam P, Mahalingam R, Yalamanchili HK, Weng T, Karmouty-Quintana H, Guha A, A Thandavarayan R. 2022. Emerging roles of alternative cleavage and

polyadenylation (APA) in human disease. *J Cell Physiol* **237**:149–160. doi:10.1002/jcp.30549
Dias A, Bouvier D, Crépin T, McCarthy AA, Hart DJ, Baudin F, Cusack S, Ruigrok RWH. 2009. The cap-snatching endonuclease of influenza virus polymerase resides in the PA subunit. *Nature* **458**:914–918. doi:10.1038/nature07745
Do CB, Woods DA, Batzoglou S. 2006. CONTRAfold: RNA secondary structure prediction without physics-based models. *Bioinformatics* **22**:e90–e98. doi:10.1093/BIOINFORMATICS/BTL246
Doepker RC, Hsu W-L, Saffran HA, Smiley JR. 2004. Herpes simplex virus virion host shutoff protein is stimulated by translation initiation factors eIF4B and eIF4H. *J Virol* **78**:4684–4699.
Doherty AJ, Serpell LC, Ponting CP. 1996. The Helix-Hairpin-Helix DNA-Binding Motif: A Structural Basis for Non-Sequence-Specific Recognition of DNA. *Nucleic Acids Research* **24**:2488–2497. doi:10.1093/nar/24.13.2488
Doma MK, Parker R. 2006. Endonucleolytic cleavage of eukaryotic mRNAs with stalls in translation elongation. *Nature* **440**:561–564. doi:10.1038/nature04530
Dubois J, Terrier O, Rosa-Calatrava M. 2014. Influenza viruses and mRNA splicing: doing more with less. *mBio* **5**:e00070-00014. doi:10.1128/mBio.00070-14
Dunagan MM, Hardy K, Takimoto T. 2021. Impact of Influenza A Virus Shutoff Proteins on Host Immune Responses. *Vaccines* **9**:629. doi:10.3390/vaccines9060629
Duncan-Lewis C, Hartenian E, King V, Glaunsinger BA. 2021. Cytoplasmic mRNA decay represses RNA polymerase II transcription during early apoptosis. *eLife* **10**:e58342. doi:10.7554/eLife.58342
Elgadi MM, Hayes CE, Smiley JR. 1999. The Herpes Simplex Virus vhs Protein Induces Endoribonucleolytic Cleavage of Target RNAs in Cell Extracts. *J Virol* **73**:7153–7164.
Endo Y, Glück A, Chan YL, Tsurugi K, Wool IG. 1990. RNA-protein interaction. An analysis with RNA oligonucleotides of the recognition by alpha-sarcin of a ribosomal domain critical for function. *J Biol Chem* **265**:2216–2222.
Engelhardt OG, Smith M, Fodor E. 2005. Association of the Influenza A Virus RNA-Dependent RNA Polymerase with Cellular RNA Polymerase II. *J Virol* **79**:5812–5818. doi:10.1128/JVI.79.9.5812-5818.2005
Esclatine A, Taddeo B, Evans L, Roizman B. 2004a. The herpes simplex virus 1 UL41 gene-dependent destabilization of cellular RNAs is selective and may be sequence-specific. *Proc Natl Acad Sci USA* **101**:3603–3608. doi:10.1073/pnas.0400354101
Esclatine A, Taddeo B, Roizman B. 2004b. The UL41 protein of herpes simplex virus mediates selective stabilization or degradation of cellular mRNAs. *Proc Natl Acad Sci USA* **101**:18165–18170. doi:10.1073/pnas.0408272102
Everly DN, Feng P, Mian IS, Read GS. 2002. mRNA Degradation by the Virion Host Shutoff (Vhs) Protein of Herpes Simplex Virus: Genetic and Biochemical Evidence that Vhs Is a Nuclease. *J Virol* **76**:8560–8571. doi:10.1128/JVI.76.17.8560-8571.2002
Fasching N, Petržílek J, Popitsch N, Herzog VA, Ameres SL. 2022. Transcriptome-Wide Profiling of RNA Stability. *Methods in molecular biology (Clifton, NJ)* **2404**:311–330. doi:10.1007/978-1-0716-1851-6_17
Feederle R, Bannert H, Lips H, Müller-Lantzsch N, Delecluse H-J. 2009. The Epstein-Barr Virus Alkaline Exonuclease BGLF5 Serves Pleiotropic Functions in Virus Replication. *J Virol* **83**:4952–4962. doi:10.1128/JVI.00170-09
Feng KH, Sun M, Iketani S, Holmes EC, Parrish CR. 2016. 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. doi:10.1016/j.virol.2016.06.001
Feng P, Everly DN, Read GS. 2005. mRNA decay during herpes simplex virus (HSV) infections: protein-protein interactions involving the HSV virion host shutoff protein and translation factors eIF4H and eIF4A. *J Virol* **79**:9651–9664. doi:10.1128/JVI.79.15.9651-9664.2005
Feng P, Everly DN, Read GS. 2001. mRNA decay during herpesvirus infections: interaction between a putative viral nuclease and a cellular translation factor. *J Virol* **75**:10272–10280. doi:10.1128/JVI.75.21.10272-10280.2001
Fenner F. 1980. The Global Eradication of Smallpox. *Medical Journal of Australia* **1**:455–456. doi:10.5694/j.1326-5377.1980.tb135034.x
Fenwick ML, McMenamin MM. 1984. Early Virion-associated Suppression of Cellular Protein Synthesis by Herpes Simplex Virus is Accompanied by Inactivation of mRNA. *Journal of General Virology* **65**:1225–1228. doi:10.1099/0022-1317-65-7-1225
Finkel Y, Gluck A, Nachshon A, Winkler R, Fisher T, Rozman B, Mizrahi O, Lubelsky Y, Zuckerman B, Slobodin B, Yahalom-Ronen Y, Tamir H, Ulitsky I, Israely T, Paran N, Schwartz M, Stern-Ginossar N. 2021. SARS-CoV-2 uses a multipronged strategy to impede host protein synthesis. *Nature* **594**:240–245. doi:10.1038/s41586-021-03610-3
Finnen RL, Zhu M, Li J, Romo D, Banfield BW. 2016. Herpes Simplex Virus 2 Virion Host Shutoff Endoribonuclease Activity Is Required To Disrupt Stress Granule Formation. *J Virol* **90**:7943–7955. doi:10.1128/JVI.00947-16
Firth AE, Jagger BW, Wise HM, Nelson CC, Parsawar K, Wills NM, Napthine S, Taubenberger JK, Digard P, Atkins JF. 2012. Ribosomal frameshifting used in influenza A virus expression occurs within the sequence UCC_UUU_CGU and is in the +1 direction. *Open Biol* **2**:120109. doi:10.1098/rsob.120109
Flannery B, Chung JR, Monto AS, Martin ET, Belongia EA, McLean HQ, Gaglani M, Murthy K, Zimmerman RK, Nowalk MP, Jackson ML, Jackson LA, Rolfes MA, Spencer S, Fry AM, US Flu VE Investigators. 2019. Influenza Vaccine Effectiveness in the United States During the 2016-2017 Season. *Clin Infect Dis* **68**:1798–1806. doi:10.1093/cid/ciy775
Flannery B, Kondor RJG, Chung JR, Gaglani M, Reis M, Zimmerman RK, Nowalk MP, Jackson ML, Jackson LA, Monto AS, Martin ET, Belongia EA, McLean HQ, Kim SS, Blanton L, Kniss K, Budd AP, Brammer L, Stark TJ, Barnes JR, Wentworth DE, Fry AM, Patel M. 2020. Spread of Antigenically Drifted Influenza A(H3N2) Viruses and Vaccine Effectiveness in the United States During the 2018-2019 Season. *J Infect Dis* **221**:8–15. doi:10.1093/infdis/jiz543
Fodor E, Velthuis AJW te. 2020. Structure and Function of the Influenza Virus Transcription and Replication Machinery. *Cold Spring Harb Perspect Med* **10**:a038398. doi:10.1101/cshperspect.a038398
Friedel CC, Whisnant AW, Djakovic L, Rutkowski AJ, Friedl M-S, Kluge M, Williamson JC, Sai S, Vidal RO, Sauer S, Hennig T, Grothey A, Milić A, Prusty BK, Lehner PJ, Matheson NJ, Erhard F, Dölken L. 2021. Dissecting Herpes Simplex Virus 1-Induced Host Shutoff at the RNA Level. *J Virol* **95**. doi:10.1128/JVI.01399-20
Frischmeyer PA, van Hoof A, O'Donnell K, Guerrerio AL, Parker R, Dietz HC. 2002. An mRNA Surveillance Mechanism That Eliminates Transcripts Lacking Termination Codons. *Science* **295**:2258–2261. doi:10.1126/science.1067338
Gaglia MM, Covarrubias S, Wong W, Glaunsinger BA. 2012. A Common Strategy for Host RNA Degradation by Divergent Viruses. *Journal of Virology* **86**:9527–9530. doi:10.1128/JVI.01230-12

Gaglia MM, Rycroft CH, Glaunsinger BA. 2015. Transcriptome-Wide Cleavage Site Mapping on Cellular mRNAs Reveals Features Underlying Sequence-Specific Cleavage by the Viral Ribonuclease SOX. *PLoS Pathog* **11**:e1005305. doi:10.1371/journal.ppat.1005305

Gagnon KT, Li L, Janowski BA, Corey DR. 2014. Analysis of nuclear RNA interference in human cells by subcellular fractionation and Argonaute loading. *Nat Protoc* **9**:2045–2060. doi:10.1038/nprot.2014.135

Galvin HD, Husain M. 2019. Influenza A virus-induced host caspase and viral PA-X antagonize the antiviral host factor, histone deacetylase 4. *J Biol Chem* **294**:20207–20221. doi:10.1074/jbc.RA119.010650

Gao H, Sun H, Hu J, Qi L, Wang J, Xiong X, Wang Y, He Q, Lin Y, Kong W, Seng L-G, Pu J, Chang K-C, Liu X, Liu J, Sun Y. 2015a. Twenty amino acids at the C-terminus of PA-X are associated with increased influenza A virus replication and pathogenicity. *Journal of General Virology* **96**:2036–2049. doi:10.1099/vir.0.000143

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. 2015b. The contribution of PA-X to the virulence of pandemic 2009 H1N1 and highly pathogenic H5N1 avian influenza viruses. *Sci Rep* **5**:8262. doi:10.1038/srep08262

Gao H, Xu G, Sun Y, Qi L, Wang J, Kong W, Sun H, Pu J, Chang K-C, Liu J. 2015c. PA-X is a virulence factor in avian H9N2 influenza virus. *J Gen Virol* **96**:2587–2594. doi:10.1099/jgv.0.000232

García-Sastre A, Egorov A, Matassov D, Brandt S, Levy DE, Durbin JE, Palese P, Muster T. 1998. Influenza A Virus Lacking the NS1 Gene Replicates in Interferon-Deficient Systems. *Virology* **252**:324–330. doi:10.1006/viro.1998.9508

Gasch AP, Eisen MB. 2002. Exploring the conditional coregulation of yeast gene expression through fuzzy k-means clustering. *Genome Biol* **3**:RESEARCH0059. doi:10.1186/gb-2002-3-11-research0059

Gaucherand L, Gaglia MM. 2022. The Role of Viral RNA Degrading Factors in Shutoff of Host Gene Expression. *Annual Review of Virology* **9**:213–238. doi:10.1146/annurev-virology-100120-012345

Gaucherand L, Porter BK, Levene RE, Price EL, Schmaling SK, Rycroft CH, Kevorkian Y, McCormick C, Khaperskyy DA, Gaglia MM. 2019. The Influenza A Virus Endoribonuclease PA-X Usurps Host mRNA Processing Machinery to Limit Host Gene Expression. *Cell Reports* **27**:776-792.e7. doi:10.1016/j.celrep.2019.03.063

Geerts-Dimitriadou C, Goldbach R, Kormelink R. 2011a. Preferential use of RNA leader sequences during influenza A transcription initiation in vivo. *Virology* **409**:27–32. doi:10.1016/j.virol.2010.09.006

Geerts-Dimitriadou C, Zwart MP, Goldbach R, Kormelink R. 2011b. Base-pairing promotes leader selection to prime in vitro influenza genome transcription. *Virology* **409**:17–26. doi:10.1016/j.virol.2010.09.003

German MA, Luo S, Schroth G, Meyers BC, Green PJ. 2009. 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. doi:10.1038/nprot.2009.8

Giese S, Bolte H, Schwemmle M. 2016. The Feat of Packaging Eight Unique Genome Segments. *Viruses* **8**:165. doi:10.3390/v8060165

Gilbertson S, Federspiel JD, Hartenian E, Cristea IM, Glaunsinger B. 2018. Changes in mRNA abundance drive shuttling of RNA binding proteins, linking cytoplasmic RNA degradation to transcription. *eLife* **7**:e37663. doi:10.7554/eLife.37663

Glaunsinger B, Ganem D. 2004a. Highly Selective Escape from KSHV-mediated Host mRNA Shutoff and Its Implications for Viral Pathogenesis. *J Exp Med* **200**:391–398. doi:10.1084/jem.20031881

Glaunsinger B, Ganem D. 2004b. Lytic KSHV infection inhibits host gene expression by accelerating global mRNA turnover. *Mol Cell* **13**:713–723.

Gong X-Q, Sun Y-F, Ruan B-Y, Liu X-M, Wang Q, Yang H-M, Wang S-Y, Zhang P, Wang X-H, Shan T-L, Tong W, Zhou Y-J, Li G-X, Zheng H, Tong G-Z, Yu H. 2017. PA-X protein decreases replication and pathogenicity of swine influenza virus in cultured cells and mouse models. *Veterinary Microbiology* **205**:66–70. doi:10.1016/j.vetmic.2017.05.004

Gu B, Eick D, Bensaude O. 2013. CTD serine-2 plays a critical role in splicing and termination factor recruitment to RNA polymerase II in vivo. *Nucleic Acids Res* **41**:1591–1603. doi:10.1093/nar/gks1327

Gu W, Gallagher GR, Dai W, Liu P, Li R, Trombly MI, Gammon DB, Mello CC, Wang JP, Finberg RW. 2015. Influenza A virus preferentially snatches noncoding RNA caps. *RNA* **21**:2067–2075. doi:10.1261/rna.054221.115.)

Hale BG, Randall RE, Ortín J, Jackson D. 2008. The multifunctional NS1 protein of influenza A viruses. *Journal of General Virology* **89**:2359–2376. doi:10.1099/vir.0.2008/004606-0

Hale BG, Steel J, Medina RA, Manicassamy B, Ye J, Hickman D, Hai R, Schmolke M, Lowen AC, Perez DR, García-Sastre A. 2010. Inefficient Control of Host Gene Expression by the 2009 Pandemic H1N1 Influenza A Virus NS1 Protein. *J Virol* **84**:6909–6922. doi:10.1128/JVI.00081-10

Hardy JG, Norbury CJ. 2016. Cleavage factor Im (CFIm) as a regulator of alternative polyadenylation. *Biochem Soc Trans* **44**:1051–1057. doi:10.1042/BST20160078

Hartenian E, Gilbertson S, Federspiel JD, Cristea IM, Glaunsinger BA. 2020. RNA decay during gammaherpesvirus infection reduces RNA polymerase II occupancy of host promoters but spares viral promoters. *PLoS Pathog* **16**:e1008269. doi:10.1371/journal.ppat.1008269

Hartenian E, Glaunsinger BA. 2019. Feedback to the central dogma: cytoplasmic mRNA decay and transcription are interdependent processes. *Critical Reviews in Biochemistry and Molecular Biology* **54**:385–398. doi:10.1080/10409238.2019.1679083

Hayashi T, Chaimayo C, McGuinness J, Takimoto T. 2016. Critical Role of the PA-X C-Terminal Domain of Influenza A Virus in Its Subcellular Localization and Shutoff Activity. *J Virol* **90**:7131–7141. doi:10.1128/JVI.00954-16

Hayashi T, MacDonald LA, Takimoto T. 2015. Influenza A Virus Protein PA-X Contributes to Viral Growth and Suppression of the Host Antiviral and Immune Responses. *J Virol* **89**:6442–6452. doi:10.1128/JVI.00319-15

Hayman A, Comely S, Lackenby A, Murphy S, McCauley J, Goodbourn S, Barclay W. 2006. Variation in the ability of human influenza A viruses to induce and inhibit the IFN-β pathway. *Virology* **347**:52–64. doi:10.1016/j.virol.2005.11.024

He T, Wang M, Cheng A, Yang Q, Jia R, Wu Y, Huang J, Tian B, Liu M, Chen S, Zhao X-X, Zhu D, Zhang S, Ou X, Mao S, Gao Q, Sun D. 2021. DPV UL41 gene encoding protein induces host shutoff activity and affects viral replication. *Veterinary Microbiology* **255**:108979. doi:10.1016/j.vetmic.2021.108979

Heinz S, Texari L, Hayes MGB, Urbanowski M, Chang MW, Givarkes N, Rialdi A, White KM, Albrecht RA, Pache L, Marazzi I, García-Sastre A, Shaw ML, Benner C. 2018. Transcription Elongation Can Affect Genome 3D Structure. *Cell* **174**:1522-1536.e22. doi:10.1016/j.cell.2018.07.047

Herold S, Becker C, Ridge KM, Budinger GRS. 2015. Influenza virus-induced lung injury: pathogenesis and implications for treatment. *European Respiratory Journal* **45**:1463–1478. doi:10.1183/09031936.00186214

Herzel L, Ottoz DSM, Alpert T, Neugebauer KM. 2017. Splicing and transcription touch base: co-transcriptional spliceosome assembly and function. *Nat Rev Mol Cell Biol* **18**:637–650. doi:10.1038/nrm.2017.63

Herzog VA, Reichholf B, Neumann T, Rescheneder P, Bhat P, Burkard TR, Wlotzka W, Von Haeseler A, Zuber J, Ameres SL. 2017. Thiol-linked alkylation of RNA to assess expression dynamics. *Nature Methods* **14**:1198–1204. doi:10.1038/nmeth.4435

Hinkley S, Ambagala APN, Jones CJ, Srikumaran S. 2000. A vhs-like activity of bovine herpesvirus-1. *Arch Virol* **145**:2027–2046. doi:10.1007/s007050070038

Hoffmann E, Neumann G, Kawaoka Y, Hobom G, Webster RG. 2000. A DNA transfection system for generation of influenza A virus from eight plasmids. *Proc Natl Acad Sci U S A* **97**:6108–6113. doi:10.1073/pnas.100133697

Houseley J, LaCava J, Tollervey D. 2006. RNA-quality control by the exosome. *Nat Rev Mol Cell Biol* **7**:529–539. doi:10.1038/nrm1964

Hsin J-P, Manley JL. 2012. The RNA polymerase II CTD coordinates transcription and RNA processing. *Genes Dev* **26**:2119–2137. doi:10.1101/gad.200303.112

Hsu J-L, Huang S-Y, Chow N-H, Chen S-H. 2003. Stable-isotope dimethyl labeling for quantitative proteomics. *Anal Chem* **75**:6843–6852. doi:10.1021/ac0348625

Hu J, Kong M, Cui Z, Gao Z, Ma C, Hu Z, Jiao X, Liu X. 2020. PA-X protein of H5N1 avian influenza virus inhibits NF-kappaB activity, a potential mechanism for PA-X counteracting the host innate immune responses. *Veterinary Microbiology* **250**:108838. doi:10.1016/j.vetmic.2020.108838

Hu J, Mo Y, Gao Z, Wang X, Gu M, Liang Y, Cheng X, Hu S, Liu W, Liu H, Chen S, Liu Xiaowen, Peng D, Liu Xiufan. 2016. PA-X-associated early alleviation of the acute lung injury contributes to the attenuation of a highly pathogenic H5N1 avian influenza virus in mice. *Med Microbiol Immunol* **205**:381–395. doi:10.1007/s00430-016-0461-2

Hu J, Mo Y, Wang X, Gu M, Hu Z, Zhong L, Wu Q, Hao X, Hu S, Liu W, Liu H, Liu Xiaowen, Liu Xiufan. 2015. PA-X decreases the pathogenicity of highly pathogenic H5N1 influenza A virus in avian species by inhibiting virus replication and host response. *J Virol* **89**:4126–4142. doi:10.1128/JVI.02132-14

Huang C, Lokugamage KG, Rozovics JM, Narayanan K, Semler BL, Makino S. 2011a. SARS coronavirus nsp1 protein induces template-dependent endonucleolytic cleavage of mRNAs: viral mRNAs are resistant to nsp1-induced RNA cleavage. *PLoS Pathog* **7**:e1002433. doi:10.1371/journal.ppat.1002433

Huang C, Lokugamage KG, Rozovics JM, Narayanan K, Semler BL, Makino S. 2011b. Alphacoronavirus transmissible gastroenteritis virus nsp1 protein suppresses protein translation in mammalian cells and in cell-free HeLa cell extracts but not in rabbit reticulocyte lysate. *J Virol* **85**:638–643. doi:10.1128/JVI.01806-10

Huang DW, Sherman BT, Lempicki RA. 2009a. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* **37**:1–13. doi:10.1093/nar/gkn923

Huang DW, Sherman BT, Lempicki RA. 2009b. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**:44–57. doi:10.1038/nprot.2008.211

Huang L, Zhang H, Deng D, Zhao K, Liu K, Hendrix DA, Mathews DH. 2019. LinearFold: linear-time approximate RNA folding by 5'-to-3' dynamic programming and beam search. *Bioinformatics* **35**:i295–i304. doi:10.1093/bioinformatics/btz375

Hussain S, Turnbull ML, Wise HM, Jagger BW, Beard PM, Kovacikova K, Taubenberger JK, Vervelde L, Engelhardt OG, Digard P. 2019. Mutation of Influenza A Virus PA-X Decreases Pathogenicity in Chicken Embryos and Can Increase the Yield of Reassortant Candidate Vaccine Viruses. *Journal of Virology* **93**:e01551-18. doi:10.1128/JVI.01551-18

Hutin S, Lee Y, Glaunsinger BA. 2013. An RNA Element in Human Interleukin 6 Confers Escape from Degradation by the Gammaherpesvirus SOX Protein. *J Virol* **87**:4672–4682. doi:10.1128/JVI.00159-13

Inglis SC. 1982. Inhibition of Host Protein Synthesis and Degradation of Cellular mRNAs During Infection by Influenza and Herpes Simplex Virus. *Mol Cell Biol* **2**:1644–1648. doi:10.1128/MCB.2.12.1644

Jagger BW, Wise HM, Kash JC, Walters K-A, Wills NM, Xiao Y-L, Dunfee RL, Schwartzman LM, Ozinsky A, Bell GL, Dalton RM, Lo A, Efstathiou S, Atkins JF, Firth AE, Taubenberger JK, Digard P. 2012. An Overlapping Protein-Coding Region in Influenza A Virus Segment 3 Modulates the Host Response. *Science* **337**:199–204. doi:10.1126/science.1222213

Jewell NA, Cline T, Mertz SE, Smirnov SV, Flaño E, Schindler C, Grieves JL, Durbin RK, Kotenko SV, Durbin JE. 2010. Lambda interferon is the predominant interferon induced by influenza A virus infection in vivo. *J Virol* **84**:11515–11522. doi:10.1128/JVI.01703-09

Johnson NPAS, Mueller J. 2002. Updating the Accounts: Global Mortality of the 1918-1920 “Spanish” Influenza Pandemic. *Bulletin of the History of Medicine* **76**:105–115. doi:10.1353/bhm.2002.0022

Jones FE, Smibert CA, Smiley JR. 1995. 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.

Kago G, Parrish S. 2021. The Mimivirus L375 Nudix enzyme hydrolyzes the 5’ mRNA cap. *PLOS ONE* **16**:e0245820. doi:10.1371/journal.pone.0245820

Kamitani W, Huang C, Narayanan K, Lokugamage KG, Makino S. 2009. A two-pronged strategy to suppress host protein synthesis by SARS coronavirus Nsp1 protein. *Nat Struct Mol Biol* **16**:1134–1140. doi:10.1038/nsmb.1680

Kamitani W, Narayanan K, Huang C, Lokugamage K, Ikegami T, Ito N, Kubo H, Makino S. 2006. Severe acute respiratory syndrome coronavirus nsp1 protein suppresses host gene expression by promoting host mRNA degradation. *Proc Natl Acad Sci USA* **103**:12885–12890. doi:10.1073/pnas.0603144103

Khan M, Hou S, Chen M, Lei H. 2022. Mechanisms of RNA export and nuclear retention. *Wiley Interdiscip Rev RNA* e1755. doi:10.1002/wrna.1755

Khaperskyy DA, Emara MM, Johnston BP, Anderson P, Hatchette TF, McCormick C. 2014. Influenza a virus host shutoff disables antiviral stress-induced translation arrest. *PLoS Pathog* **10**:e1004217. doi:10.1371/journal.ppat.1004217

Khaperskyy DA, Hatchette TF, McCormick C. 2012. Influenza A virus inhibits cytoplasmic stress granule formation. *FASEB J* **26**:1629–1639. doi:10.1096/fj.11-196915

Khaperskyy DA, Schmaling S, Larkins-Ford J, McCormick C, Gaglia MM. 2016. Selective Degradation of Host RNA Polymerase II Transcripts by Influenza A Virus PA-X Host Shutoff Protein. *PLoS Pathog* **12**:e1005427. doi:10.1371/journal.ppat.1005427

Kieny MP, Lathe R, Drillien R, Spehner D, Skory S, Schmitt D, Wiktor T, Koprowski H, Lecocq JP. 1984. Expression of rabies virus glycoprotein from a recombinant vaccinia virus. *Nature* **312**:163–166. doi:10.1038/312163a0

Kim D, Lee J-Y, Yang J-S, Kim JW, Kim VN, Chang H. 2020. The Architecture of SARS-CoV-2 Transcriptome. *Cell* **181**:914-921.e10. doi:10.1016/j.cell.2020.04.011
Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology* **37**:907–915. doi:10.1038/s41587-019-0201-4
Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology* **14**:R36. doi:10.1186/gb-2013-14-4-r36
Kim S, Yamamoto J, Chen Y, Aida M, Wada T, Handa H, Yamaguchi Y. 2010. Evidence that cleavage factor Im is a heterotetrameric protein complex controlling alternative polyadenylation. *Genes Cells* **15**:1003–1013. doi:10.1111/j.1365-2443.2010.01436.x
Kochs G, García-Sastre A, Martínez-Sobrido L. 2007. Multiple Anti-Interferon Actions of the Influenza A Virus NS1 Protein. *J Virol* **81**:7011–7021. doi:10.1128/JVI.02581-06
Koonin EV. 1993. A highly conserved sequence motif defining the family of MutT-related proteins from eubacteria, eukaryotes and viruses. *Nucleic Acids Res* **21**:4847.
Krug RM. 2015. Functions of the influenza A virus NS1 protein in antiviral defense. *Current Opinion in Virology* **12**:1–6. doi:10.1016/j.coviro.2015.01.007
Kühn U, Buschmann J, Wahle E. 2017. The nuclear poly(A) binding protein of mammals, but not of fission yeast, participates in mRNA polyadenylation. *RNA* **23**:473–482. doi:10.1261/rna.057026.116
Kumar GR, Glaunsinger BA. 2010. Nuclear Import of Cytoplasmic Poly(A) Binding Protein Restricts Gene Expression via Hyperadenylation and Nuclear Retention of mRNA. *Mol Cell Biol* **30**:4996–5008. doi:10.1128/MCB.00600-10
Kumar GR, Shum L, Glaunsinger BA. 2011. Importin α-Mediated Nuclear Import of Cytoplasmic Poly(A) Binding Protein Occurs as a Direct Consequence of Cytoplasmic mRNA Depletion. *Mol Cell Biol* **31**:3113–3125. doi:10.1128/MCB.05402-11
Kwong AD, Frenkel N. 1987. Herpes simplex virus-infected cells contain a function(s) that destabilizes both host and viral mRNAs. *Proc Natl Acad Sci USA* **84**:1926–1930.
Łabno A, Tomecki R, Dziembowski A. 2016. Cytoplasmic RNA decay pathways - Enzymes and mechanisms. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1863**:3125–3147. doi:10.1016/j.bbamcr.2016.09.023
Lam Q, Smibert CA, Koop KE, Lavery C, Capone JP, Weinheimer SP, Smiley JR. 1996. Herpes simplex virus VP16 rescues viral mRNA from destruction by the virion host shutoff function. *EMBO J* **15**:2575–2581.
Lamb RA, Choppin PW. 1977. Synthesis of influenza virus polypeptides in cells resistant to alpha-amanitin: evidence for the involvement of cellular RNA polymerase II in virus replication. *Journal of Virology* **23**:816–819. doi:10.1128/jvi.23.3.816-819.1977
Lazear HM, Schoggins JW, Diamond MS. 2019. Shared and Distinct Functions of Type I and Type III Interferons. *Immunity* **50**:907–923. doi:10.1016/j.immuni.2019.03.025
Le Sage V, Kormuth KA, Nturibi E, Lee JM, Frizzell SA, Myerburg MM, Bloom JD, Lakdawala SS. 2021. Cell-Culture Adaptation of H3N2 Influenza Virus Impacts Acid Stability and Reduces Airborne Transmission in Ferret Model. *Viruses* **13**:719. doi:10.3390/v13050719
Lebars I, Hu R-M, Lallemand J-Y, Uzan M, Bontems F. 2001. Role of the Substrate Conformation and of the S1 Protein in the Cleavage Efficiency of the T4

Endoribonuclease RegB*. *Journal of Biological Chemistry* **276**:13264–13272. doi:10.1074/jbc.M010680200
Lee N, Le Sage V, Nanni AV, Snyder DJ, Cooper VS, Lakdawala SS. 2017. Genome-wide analysis of influenza viral RNA and nucleoprotein association. *Nucleic Acids Res* **45**:8968–8977. doi:10.1093/nar/gkx584
Lee YJ, Glaunsinger BA. 2009. Aberrant herpesvirus-induced polyadenylation correlates with cellular messenger RNA destruction. *PLoS Biol* **7**:e1000107. doi:10.1371/journal.pbio.1000107
Lee-Chen GJ, Niles EG. 1988. Transcription and translation mapping of the 13 genes in the vaccinia virus HindIII D fragment. *Virology* **163**:52–63. doi:10.1016/0042-6822(88)90233-4
Levene RE, Shrestha SD, Gaglia MM. 2021. The influenza A virus host shutoff factor PA-X is rapidly turned over in a strain-specific manner. *J Virol* **95**:e02312-20. doi:10.1128/JVI.02312-20
Li M, Qi W, Chang Q, Chen R, Zhen D, Liao M, Wen J, Deng Y. 2021. Influenza A virus protein PA-X suppresses host Ankrd17-mediated immune responses. *Microbiology and Immunology* **65**:48–59. doi:https://doi.org/10.1111/1348-0421.12863
Li Q, Yuan X, Wang Q, Chang G, Wang F, Liu R, Zheng M, Chen G, Wen J, Zhao G. 2016. Interactomic landscape of PA-X-chicken protein complexes of H5N1 influenza A virus. *Journal of Proteomics* **148**:20–25. doi:10.1016/j.jprot.2016.07.009
Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. 2015. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* **1**:417–425. doi:10.1016/j.cels.2015.12.004
Lim J, Ha M, Chang H, Kwon SC, Simanshu DK, Patel DJ, Kim VN. 2014. Uridylation by TUT4 and TUT7 Marks mRNA for Degradation. *Cell* **159**:1365–1376. doi:10.1016/j.cell.2014.10.055
Lin H-W, Hsu W-L, Chang Y-Y, Jan M-S, Wong M-L, Chang T-J. 2010. Role of the UL41 protein of pseudorabies virus in host shutoff, pathogenesis and induction of TNF-α expression. *J Vet Med Sci* **72**:1179–1187. doi:10.1292/jvms.10-0059
Liu R, Moss B. 2016. 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. doi:10.1128/jvi.00869-16
Liu R, Sheng Z, Huang C, Wang D, Li F. 2020. Influenza D virus. *Current Opinion in Virology*, Preventive and therapeutic vaccines * Viral Immunology **44**:154–161. doi:10.1016/j.coviro.2020.08.004
Liu S-W, Wyatt LS, Orandle MS, Minai M, Moss B. 2014. The D10 Decapping Enzyme of Vaccinia Virus Contributes to Decay of Cellular and Viral mRNAs and to Virulence in Mice. *J Virol* **88**:202–211. doi:10.1128/JVI.02426-13
Liu Y-F, Tsai P-Y, Chulakasian S, Lin F-Y, Hsu W-L. 2016. The pseudorabies virus vhs protein cleaves RNA containing an IRES sequence. *The FEBS Journal* **283**:899–911. doi:10.1111/febs.13642
Liu Y-F, Tsai P-Y, Lin F-Y, Lin K-H, Chang T-J, Lin H-W, Chulakasian S, Hsu W-L. 2015. Roles of nucleic acid substrates and cofactors in the vhs protein activity of pseudorabies virus. *Vet Res* **46**:1–14. doi:10.1186/s13567-015-0284-y
Llompart CM, Nieto A, Rodriguez-Frandsen A. 2014. Specific Residues of PB2 and PA Influenza Virus Polymerase Subunits Confer the Ability for RNA Polymerase II Degradation and Virus Pathogenicity in Mice. *J Virol* **88**:3455–3463. doi:10.1128/JVI.02263-13

Lokugamage KG, Narayanan K, Huang C, Makino S. 2012. Severe Acute Respiratory Syndrome Coronavirus Protein nsp1 Is a Novel Eukaryotic Translation Inhibitor That Represses Multiple Steps of Translation Initiation. *Journal of Virology* **86**:13598–13608. doi:10.1128/JVI.01958-12

Lokugamage KG, Narayanan K, Nakagawa K, Terasaki K, Ramirez SI, Tseng C-TK, Makino S. 2015. Middle East Respiratory Syndrome Coronavirus nsp1 Inhibits Host Gene Expression by Selectively Targeting mRNAs Transcribed in the Nucleus while Sparing mRNAs of Cytoplasmic Origin. *Journal of Virology* **89**:10970–10981. doi:10.1128/JVI.01352-15

Lorenz R, Bernhart SH, Höner zu Siederdissen C, Tafer H, Flamm C, Stadler PF, Hofacker IL. 2011. ViennaRNA Package 2.0. *Algorithms for Molecular Biology* **6**:1–14. doi:10.1186/1748-7188-6-26/TABLES/2

Lubas M, Damgaard CK, Tomecki R, Cysewski D, Jensen TH, Dziembowski A. 2013. Exonuclease hDIS3L2 specifies an exosome-independent 3′-5′ degradation pathway of human cytoplasmic mRNA. *EMBO J* **32**:1855–1868. doi:10.1038/emboj.2013.135

Ly M, Burgess HM, Shah SB, Mohr I, Glaunsinger BA. 2022. Vaccinia virus D10 has broad decapping activity that is regulated by mRNA splicing. *PLoS Pathog* **18**:e1010099. doi:10.1371/journal.ppat.1010099

Lykke-Andersen S, Jensen TH. 2015. Nonsense-mediated mRNA decay: an intricate machinery that shapes transcriptomes. *Nat Rev Mol Cell Biol* **16**:665–677. doi:10.1038/nrm4063

Ma S, Bhattacharjee RB, Bag J. 2009. Expression of poly(A)-binding protein is upregulated during recovery from heat shock in HeLa cells. *The FEBS Journal* **276**:552–570. doi:10.1111/j.1742-4658.2008.06803.x

MacRae IJ, Zhou K, Doudna JA. 2007. Structural determinants of RNA recognition and cleavage by Dicer. *Nat Struct Mol Biol* **14**:934–940. doi:10.1038/nsmb1293

Macveigh-Fierro D, Cicerchia A, Cadorette A, Sharma V, Muller M. 2022. The m6A reader YTHDC2 is essential for escape from KSHV SOX-induced RNA decay. *Proc Natl Acad Sci U S A* **119**:e2116662119. doi:10.1073/pnas.2116662119

Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**:10–12. doi:10.14806/EJ.17.1.200

Martínez-Alonso M, Hengrung N, Fodor E. 2016. RNA-Free and Ribonucleoprotein-Associated Influenza Virus Polymerases Directly Bind the Serine-5-Phosphorylated Carboxyl-Terminal Domain of Host RNA Polymerase II. *J Virol* **90**:6014–6021. doi:10.1128/JVI.00494-16

Martinson HG. 2011. An active role for splicing in 3′-end formation. *WIREs RNA* **2**:459–470. doi:10.1002/wrna.68

Matera AG, Wang Z. 2014. A day in the life of the spliceosome. *Nat Rev Mol Cell Biol* **15**:108–121. doi:10.1038/nrm3742

Mathews DH, Disney MD, Childs JL, Schroeder SJ, Zuker M, Turner DH. 2004. 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. doi:10.1073/PNAS.0401799101

Matrosovich M, Matrosovich T, Garten W, Klenk H-D. 2006. New low-viscosity overlay medium for viral plaque assays. *Virology Journal* **3**:63. doi:10.1186/1743-422X-3-63

McCormick C, Khaperskyy DA. 2017. Translation inhibition and stress granules in the antiviral immune response. *Nat Rev Immunol* **17**:647–660. doi:10.1038/nri.2017.63

Medina RA, García-Sastre A. 2011. Influenza A viruses: new research developments. *Nat Rev Microbiol* **9**:590–603. doi:10.1038/nrmicro2613

Mendez AS, Ly M, González-Sánchez AM, Hartenian E, Ingolia NT, Cate JH, Glaunsinger BA. 2021. The N-terminal domain of SARS-CoV-2 nsp1 plays key roles in suppression of cellular gene expression and preservation of viral gene expression. *Cell Reports* 109841. doi:10.1016/j.celrep.2021.109841

Mendez AS, Vogt C, Bohne J, Glaunsinger BA. 2018. Site specific target binding controls RNA cleavage efficiency by the Kaposi's sarcoma-associated herpesvirus endonuclease SOX. *Nucleic Acids Research* **46**:11968–11979. doi:10.1093/nar/gky932

Molinari N-AM, Ortega-Sanchez IR, Messonnier ML, Thompson WW, Wortley PM, Weintraub E, Bridges CB. 2007. The annual impact of seasonal influenza in the US: Measuring disease burden and costs. *Vaccine* **25**:5086–5096. doi:10.1016/j.vaccine.2007.03.046

Morris DE, Cleary DW, Clarke SC. 2017. Secondary Bacterial Infections Associated with Influenza Pandemics. *Frontiers in Microbiology* **8**.

Mugridge JS, Coller J, Gross JD. 2018. Structural and molecular mechanisms for the control of eukaryotic 5′–3′ mRNA decay. *Nat Struct Mol Biol* **25**:1077–1085. doi:10.1038/s41594-018-0164-z

Muller M, Glaunsinger BA. 2017. Nuclease escape elements protect messenger RNA against cleavage by multiple viral endonucleases. *PLoS Pathog* **13**:e1006593. doi:10.1371/journal.ppat.1006593

Muller M, Hutin S, Marigold O, Li KH, Burlingame A, Glaunsinger BA. 2015. A Ribonucleoprotein Complex Protects the Interleukin-6 mRNA from Degradation by Distinct Herpesviral Endonucleases. *PLoS Pathog* **11**:e1004899. doi:10.1371/journal.ppat.1004899

Nacken W, Schreiber A, Masemann D, Ludwig S. 2021. The Effector Domain of the Influenza A Virus Nonstructural Protein NS1 Triggers Host Shutoff by Mediating Inhibition and Global Deregulation of Host Transcription When Associated with Specific Structures in the Nucleus. *mBio* **12**:e02196-21. doi:10.1128/mBio.02196-21

Nagarajan VK, Jones CI, Newbury SF, Green PJ. 2013. XRN 5′→3′ exoribonucleases: Structure, mechanisms and functions. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, RNA Decay Mechanisms **1829**:590–603. doi:10.1016/j.bbagrm.2013.03.005

Naito T, Kiyasu Y, Sugiyama K, Kimura A, Nakano R, Matsukage A, Nagata K. 2007. An influenza virus replicon system in yeast identified Tat-SF1 as a stimulatory host factor for viral RNA synthesis. *Proceedings of the National Academy of Sciences* **104**:18235–18240. doi:10.1073/pnas.0705856104

Narayanan K, Huang C, Lokugamage K, Kamitani W, Ikegami T, Tseng C-TK, Makino S. 2008. Severe Acute Respiratory Syndrome Coronavirus nsp1 Suppresses Host Gene Expression, Including That of Type I Interferon, in Infected Cells. *Journal of Virology* **82**:4471–4479. doi:10.1128/JVI.02472-07

Narayanan K, Ramirez SI, Lokugamage KG, Makino S. 2015. Coronavirus nonstructural protein 1: Common and distinct functions in the regulation of host and viral gene expression. *Virus Research*, Nidoviruses II **202**:89–100. doi:10.1016/j.virusres.2014.11.019

Narkpuk J, Jongkaewwattana A, Teeravechyan S. 2018. The avian influenza virus PA segment mediates strain-specific antagonism of BST-2/tetherin. *Virology* **525**:161–169. doi:10.1016/j.virol.2018.09.016

Nayak DP, Balogun RA, Yamada H, Zhou ZH, Barman S. 2009. Influenza virus morphogenesis and budding. *Virus Research*, Virus Research - 25th Anniversary Issue **143**:147–161. doi:10.1016/j.virusres.2009.05.010

Nayak DP, Hui EK-W, Barman S. 2004. Assembly and budding of influenza virus. *Virus Research*, Mechanisms of Enveloped Virus Release **106**:147–165. doi:10.1016/j.virusres.2004.08.012

Nemeroff ME, Barabino SM, Li Y, Keller W, Krug RM. 1998. Influenza virus NS1 protein interacts with the cellular 30 kDa subunit of CPSF and inhibits 3'end formation of cellular pre-mRNAs. *Mol Cell* **1**:991–1000.

Nguyen HT, Fry AM, Gubareva LV. 2012. Neuraminidase inhibitor resistance in influenza viruses and laboratory testing methods. *Antivir Ther* **17**:159–173. doi:10.3851/IMP2067

Nguyen KT, Mun S-H, Lee C-S, Hwang C-S. 2018. Control of protein degradation by N-terminal acetylation and the N-end rule pathway. *Exp Mol Med* **50**:1–8. doi:10.1038/s12276-018-0097-y

Nguyen TD, Trinh TA, Bao S, Nguyen TA. 2022. Secondary structure RNA elements control the cleavage activity of DICER. *Nat Commun* **13**:2138. doi:10.1038/s41467-022-29822-3

Noah DL, Twu KY, Krug RM. 2003. Cellular antiviral responses against influenza A virus are countered at the posttranscriptional level by the viral NS1A protein via its binding to a cellular protein required for the 3' end processing of cellular pre-mRNAS. *Virology* **307**:386–395.

Nogales A, Martinez-Sobrido L, Chiem K, Topham DJ, DeDiego ML. 2018. Functional Evolution of the 2009 Pandemic H1N1 Influenza Virus NS1 and PA in Humans. *Journal of Virology* **92**. doi:10.1128/JVI.01206-18

Nogales A, Rodriguez L, DeDiego ML, Topham DJ, Martínez-Sobrido L. 2017. Interplay of PA-X and NS1 Proteins in Replication and Pathogenesis of a Temperature-Sensitive 2009 Pandemic H1N1 Influenza A Virus. *J Virol* **91**:e00720-17. doi:10.1128/JVI.00720-17

Nogales A, Villamayor L, Utrilla-Trigo S, Ortego J, Martinez-Sobrido L, DeDiego ML. 2021. Natural Selection of H5N1 Avian Influenza A Viruses with Increased PA-X and NS1 Shutoff Activity. *Viruses* **13**:1760. doi:10.3390/v13091760

Oishi K, Yamayoshi S, Kawaoka Y. 2019. Identification of Amino Acid Residues in Influenza A Virus PA-X That Contribute to Enhanced Shutoff Activity. *Front Microbiol* **10**. doi:10.3389/fmicb.2019.00432

Oishi K, Yamayoshi S, Kawaoka Y. 2018a. Identification of novel amino acid residues of influenza virus PA-X that are important for PA-X shutoff activity by using yeast. *Virology* **516**:71–75. doi:10.1016/j.virol.2018.01.004

Oishi K, Yamayoshi S, Kawaoka Y. 2015. Mapping of a Region of the PA-X Protein of Influenza A Virus That Is Important for Its Shutoff Activity. *J Virol* **89**:8661–8665. doi:10.1128/JVI.01132-15

Oishi K, Yamayoshi S, Kozuka-Hata H, Oyama M, Kawaoka Y. 2018b. N-Terminal Acetylation by NatB Is Required for the Shutoff Activity of Influenza A Virus PA-X. *Cell Reports* **24**:851–860. doi:10.1016/j.celrep.2018.06.078

Oroskar AA, Read GS. 1989. Control of mRNA stability by the virion host shutoff function of herpes simplex virus. *J Virol* **63**:1897–1906.

Oroskar AA, Read GS. 1987. A mutant of herpes simplex virus type 1 exhibits increased stability of immediate-early (alpha) mRNAs. *J Virol* **61**:604–606.

Otsuka H, Fukao A, Funakami Y, Duncan KE, Fujiwara T. 2019. Emerging Evidence of Translational Control by AU-Rich Element-Binding Proteins. *Frontiers in Genetics* **10**.

Page HG, Read GS. 2010. The Virion Host Shutoff Endonuclease (UL41) of Herpes Simplex Virus Interacts with the Cellular Cap-Binding Complex eIF4F. *J Virol* **84**:6886–6890. doi:10.1128/JVI.00166-10

Palese P, Tobita K, Ueda M, Compans RW. 1974. Characterization of temperature sensitive influenza virus mutants defective in neuraminidase. *Virology* **61**:397–410. doi:10.1016/0042-6822(74)90276-1

Panthu B, Terrier O, Carron C, Traversier A, Corbin A, Balvay L, Lina B, Rosa-Calatrava M, Ohlmann T. 2017. The NS1 Protein from Influenza Virus Stimulates Translation Initiation by Enhancing Ribosome Recruitment to mRNAs. *Journal of Molecular Biology*, RNA processing and genome integrity **429**:3334–3352. doi:10.1016/j.jmb.2017.04.007

Parrish S, Hurchalla M, Liu S-W, Moss B. 2009. The African swine fever virus g5R protein possesses mRNA decapping activity. *Virology* **393**:177–182. doi:10.1016/j.virol.2009.07.026

Parrish S, Moss B. 2007. Characterization of a second vaccinia virus mRNA-decapping enzyme conserved in poxviruses. *J Virol* **81**:12973–12978. doi:10.1128/JVI.01668-07

Parrish S, Moss B. 2006. Characterization of a Vaccinia Virus Mutant with a Deletion of the D10R Gene Encoding a Putative Negative Regulator of Gene Expression. *J Virol* **80**:553–561. doi:10.1128/JVI.80.2.553-561.2006

Parrish S, Resch W, Moss B. 2007. Vaccinia virus D10 protein has mRNA decapping activity, providing a mechanism for control of host and viral gene expression. *Proc Natl Acad Sci U S A* **104**:2139–2144. doi:10.1073/pnas.0611685104

Pérez P, Q Marín M, Lázaro-Frías A, Jiménez de Oya N, Blázquez A-B, Escribano-Romero E, S Sorzano CÓ, Ortego J, Saiz J-C, Esteban M, Martín-Acebes MA, García-Arriaza J. 2018. A Vaccine Based on a Modified Vaccinia Virus Ankara Vector Expressing Zika Virus Structural Proteins Controls Zika Virus Replication in Mice. *Sci Rep* **8**:17385. doi:10.1038/s41598-018-35724-6

Pezda AC, Penn A, Barton GM, Coscoy L. 2011. Suppression of TLR9 immunostimulatory motifs in the genome of a gammaherpesvirus. *J Immunol* **187**:887–896. doi:10.4049/jimmunol.1003737

Plotch SJ, Bouloy M, Ulmanen I, Krug RM. 1981. A unique cap(m7GpppXm)-dependent influenza virion endonuclease cleaves capped RNAs to generate the primers that initiate viral RNA transcription. *Cell* **23**:847–858. doi:10.1016/0092-8674(81)90449-9

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

Proudfoot NJ, Furger A, Dye MJ. 2002. Integrating mRNA Processing with Transcription. *Cell* **108**:501–512. doi:10.1016/S0092-8674(02)00617-7

Qin T, Chen Y, Huangfu D, Yin Yinyan, Miao X, Yin Yuncong, Chen S, Peng D, Liu X. 2022a. PA-X protein assists H9N2 subtype avian influenza virus in escaping immune response of mucosal dendritic cells. *Transbound Emerg Dis*. doi:10.1111/tbed.14665

Qin T, Chen Y, Huangfu D, Yin Yinyan, Miao X, Yin Yuncong, Chen S, Peng D, Liu X. 2022b. PA-X protein of H1N1 subtype influenza virus disables the nasal mucosal dendritic cells for strengthening virulence. *Virulence* **13**:1928–1942. doi:10.1080/21505594.2022.2139474

Quintas A, Pérez-Núñez D, Sánchez EG, Nogal ML, Hentze MW, Castelló A, Revilla Y. 2017. Characterization of the African Swine Fever Virus Decapping Enzyme during Infection. *J Virol* **91**:e00990-17. doi:10.1128/JVI.00990-17

Rao S, Hoskins I, Tonn T, Garcia PD, Ozadam H, Cenik ES, Cenik C. 2021. Genes with 5′ terminal oligopyrimidine tracts preferentially escape global suppression of translation by the SARS-CoV-2 Nsp1 protein. *RNA* **27**:1025–1045. doi:10.1261/rna.078661.120

Raoult D, Audic S, Robert C, Abergel C, Renesto P, Ogata H, La Scola B, Suzan M, Claverie J-M. 2004. The 1.2-megabase genome sequence of Mimivirus. *Science* **306**:1344–1350. doi:10.1126/science.1101485

Rappsilber J, Ryder U, Lamond AI, Mann M. 2002. Large-scale proteomic analysis of the human spliceosome. *Genome Res* **12**:1231–1245. doi:10.1101/gr.473902

REED LJ, MUENCH H. 1938. A SIMPLE METHOD OF ESTIMATING FIFTY PER CENT ENDPOINTS12. *American Journal of Epidemiology* **27**:493–497. doi:10.1093/oxfordjournals.aje.a118408

Reich S, Guilligay D, Pflug A, Malet H, Berger I, Crépin T, Hart D, Lunardi T, Nanao M, Ruigrok RWH, Cusack S. 2014. Structural insight into cap-snatching and RNA synthesis by influenza polymerase. *Nature* **516**:361–366. doi:10.1038/nature14009

Rerks-Ngarm S, Pitisuttithum P, Nitayaphan S, Kaewkungwal J, Chiu J, Paris R, Premsri N, Namwat C, de Souza M, Adams E, Benenson M, Gurunathan S, Tartaglia J, McNeil JG, Francis DP, Stablein D, Birx DL, Chunsuttiwat S, Khamboonruang C, Thongcharoen P, Robb ML, Michael NL, Kunasol P, Kim JH. 2009. Vaccination with ALVAC and AIDSVAX to Prevent HIV-1 Infection in Thailand. *New England Journal of Medicine* **361**:2209–2220. doi:10.1056/NEJMoa0908492

Rialdi A, Hultquist J, Jimenez-Morales D, Peralta Z, Campisi L, Fenouil R, Moshkina N, Wang ZZ, Laffleur B, Kaake RM, McGregor MJ, Haas K, Pefanis E, Albrecht RA, Pache L, Chanda S, Jen J, Ochando J, Byun M, Basu U, García-Sastre A, Krogan N, van Bakel H, Marazzi I. 2017. The RNA Exosome Syncs IAV-RNAPII Transcription to Promote Viral Ribogenesis and Infectivity. *Cell* **169**:679-692.e14. doi:10.1016/j.cell.2017.04.021

Rigby RE, Wise HM, Smith N, Digard P, Rehwinkel J. 2019. PA-X antagonises MAVS-dependent accumulation of early type I interferon messenger RNAs during influenza A virus infection. *Sci Rep* **9**:7216. doi:10.1038/s41598-019-43632-6

Robertson JS, Schubert M, Lazzarini RA. 1981. Polyadenylation sites for influenza virus mRNA. *Journal of Virology* **38**:157–163. doi:10.1128/jvi.38.1.157-163.1981

Rodriguez A, Pérez-González A, Hossain MJ, Chen L-M, Rolling T, Pérez-Breña P, Donis R, Kochs G, Nieto A. 2009. Attenuated Strains of Influenza A Viruses Do Not Induce Degradation of RNA Polymerase II. *Journal of Virology* **83**:11166–11174. doi:10.1128/JVI.01439-09

Rodriguez A, Pérez-González A, Nieto A. 2007. Influenza virus infection causes specific degradation of the largest subunit of cellular RNA polymerase II. *J Virol* **81**:5315–5324. doi:10.1128/JVI.02129-06

Rodriguez W, Macveigh-Fierro D, Miles J, Muller M. 2021. Fated for decay: RNA elements targeted by viral endonucleases. *Seminars in Cell & Developmental Biology* **111**:119–125. doi:10.1016/j.semcdb.2020.05.010

Rodriguez W, Srivastav K, Muller M. 2019. C19ORF66 Broadly Escapes Virus-Induced Endonuclease Cleavage and Restricts Kaposi's Sarcoma-Associated Herpesvirus. *J Virol* **93**:e00373-19. doi:10.1128/JVI.00373-19

Rolfes MA, Flannery B, Chung JR, O'Halloran A, Garg S, Belongia EA, Gaglani M, Zimmerman RK, Jackson ML, Monto AS, Alden NB, Anderson E, Bennett NM, Billing L, Eckel S, Kirley PD, Lynfield R, Monroe ML, Spencer M, Spina N, Talbot HK, Thomas A, Torres SM, Yousey-Hindes K, Singleton JA, Patel M, Reed C, Fry AM, US Influenza Vaccine Effectiveness (Flu VE) Network, the Influenza

Hospitalization Surveillance Network, and the Assessment Branch, Immunization Services Division, Centers for Disease Control and Prevention. 2019. Effects of Influenza Vaccination in the United States During the 2017-2018 Influenza Season. *Clin Infect Dis* **69**:1845–1853. doi:10.1093/cid/ciz075

Roux KJ, Kim DI, Raida M, Burke B. 2012. A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *J Cell Biol* **196**:801–810. doi:10.1083/jcb.201112098

Rowe M, Glaunsinger B, van Leeuwen D, Zuo J, Sweetman D, Ganem D, Middeldorp J, Wiertz EJHJ, Ressing ME. 2007. Host shutoff during productive Epstein-Barr virus infection is mediated by BGLF5 and may contribute to immune evasion. *Proc Natl Acad Sci USA* **104**:3366–3371. doi:10.1073/pnas.0611128104

Rutkowski AJ, Erhard F, L'Hernault A, Bonfert T, Schilhabel M, Crump C, Rosenstiel P, Efstathiou S, Zimmer R, Friedel CC, Dölken L. 2015. Widespread disruption of host transcription termination in HSV-1 infection. *Nat Commun* **6**:7126. doi:10.1038/ncomms8126

Sadek J, Read GS. 2016. The Splicing History of an mRNA Affects Its Level of Translation and Sensitivity to Cleavage by the Virion Host Shutoff Endonuclease during Herpes Simplex Virus Infections. *J Virol* **90**:10844–10856. doi:10.1128/JVI.01302-16

Saïda F, Uzan M, Bontems F. 2003. The phage T4 restriction endoribonuclease RegB: a cyclizing enzyme that requires two histidines to be fully active. *Nucleic Acids Research* **31**:2751–2758. doi:10.1093/nar/gkg377

Sainsbury S, Bernecky C, Cramer P. 2015. Structural basis of transcription initiation by RNA polymerase II. *Nat Rev Mol Cell Biol* **16**:129–143. doi:10.1038/nrm3952

Salaun C, MacDonald AI, Larralde O, Howard L, Lochtie K, Burgess HM, Brook M, Malik P, Gray NK, Graham SV. 2010. Poly(A)-binding protein 1 partially relocalizes to the nucleus during herpes simplex virus type 1 infection in an ICP27-independent manner and does not inhibit virus replication. *J Virol* **84**:8539–8548. doi:10.1128/JVI.00668-10

Salvatore M, Basler CF, Parisien J-P, Horvath CM, Bourmakina S, Zheng H, Muster T, Palese P, García-Sastre A. 2002. Effects of influenza A virus NS1 protein on protein expression: the NS1 protein enhances translation and is not required for shutoff of host protein synthesis. *J Virol* **76**:1206–1212. doi:10.1128/jvi.76.3.1206-1212.2002

Santoro MG, Rossi A, Amici C. 2003. NF-kappaB and virus infection: who controls whom. *EMBO J* **22**:2552–2560. doi:10.1093/emboj/cdg267

Sarma N, Agarwal D, Shiflett LA, Read GS. 2008. Small Interfering RNAs That Deplete the Cellular Translation Factor eIF4H Impede mRNA Degradation by the Virion Host Shutoff Protein of Herpes Simplex Virus. *J Virol* **82**:6600–6609. doi:10.1128/JVI.00137-08

Satterly N, Tsai P-L, van Deursen J, Nussenzveig DR, Wang Y, Faria PA, Levay A, Levy DE, Fontoura BMA. 2007. Influenza virus targets the mRNA export machinery and the nuclear pore complex. *Proc Natl Acad Sci U S A* **104**:1853–1858. doi:10.1073/pnas.0610977104

Scarborough AM, Flaherty JN, Hunter OV, Liu K, Kumar A, Xing C, Tu BP, Conrad NK. 2021. SAM homeostasis is regulated by CFIm-mediated splicing of MAT2A. *eLife* **10**:e64930. doi:10.7554/eLife.64930

Schoggins JW, Wilson SJ, Panis M, Murphy MY, Jones CT, Bieniasz P, Rice CM. 2011. A diverse range of gene products are effectors of the type I interferon antiviral response. *Nature* **472**:481–485. doi:10.1038/nature09907

Schubert K, Karousis ED, Jomaa A, Scaiola A, Echeverria B, Gurzeler L-A, Leibundgut M, Thiel V, Mühlemann O, Ban N. 2020. SARS-CoV-2 Nsp1 binds the ribosomal mRNA channel to inhibit translation. *Nature Structural & Molecular Biology* **27**:959–966. doi:10.1038/s41594-020-0511-8

Sederdahl BK, Williams JV. 2020. Epidemiology and Clinical Characteristics of Influenza C Virus. *Viruses* **12**:89. doi:10.3390/v12010089

Sempere Borau M, Stertz S. 2021. Entry of influenza A virus into host cells — recent progress and remaining challenges. *Current Opinion in Virology* **48**:23–29. doi:10.1016/j.coviro.2021.03.001

Setaro AC, Gaglia MM. 2021. All hands on deck: SARS-CoV-2 proteins that block early anti-viral interferon responses. *Current Research in Virological Science* **2**:100015. doi:10.1016/j.crviro.2021.100015

Shi M, Jagger BW, Wise HM, Digard P, Holmes EC, Taubenberger JK. 2012. Evolutionary Conservation of the PA-X Open Reading Frame in Segment 3 of Influenza A Virus. *Journal of Virology* **86**:12411–12413. doi:10.1128/JVI.01677-12

Shiflett LA, Read GS. 2013. mRNA Decay during Herpes Simplex Virus (HSV) Infections: Mutations That Affect Translation of an mRNA Influence the Sites at Which It Is Cleaved by the HSV Virion Host Shutoff (Vhs) Protein. *J Virol* **87**:94–109. doi:10.1128/JVI.01557-12

Shinya K, Ebina M, Yamada S, Ono M, Kasai N, Kawaoka Y. 2006. Avian flu: influenza virus receptors in the human airway. *Nature* **440**:435–436. doi:10.1038/440435a

Shors T, Keck JG, Moss B. 1999. Down regulation of gene expression by the vaccinia virus D10 protein. *J Virol* **73**:791–796. doi:10.1128/JVI.73.1.791-796.1999

Shu M, Taddeo B, Roizman B. 2015. Tristetraprolin Recruits the Herpes Simplex Virion Host Shutoff RNase to AU-Rich Elements in Stress Response mRNAs To Enable Their Cleavage. *J Virol* **89**:5643–5650. doi:10.1128/JVI.00091-15

Shu M, Taddeo B, Roizman B. 2013. The Nuclear-Cytoplasmic Shuttling of Virion Host Shutoff RNase Is Enabled by pUL47 and an Embedded Nuclear Export Signal and Defines the Sites of Degradation of AU-Rich and Stable Cellular mRNAs. *J Virol* **87**:13569–13578. doi:10.1128/JVI.02603-13

Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD, Higgins DG. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology* **7**:539. doi:10.1038/msb.2011.75

Sikora D, Rocheleau L, Brown EG, Pelchat M. 2017. Influenza A virus cap-snatches host RNAs based on their abundance early after infection. *Virology* **509**:167–177. doi:10.1016/j.virol.2017.06.020

Sikora D, Rocheleau L, Brown EG, Pelchat M. 2015. Deep sequencing reveals the eight facets of the influenza A/HongKong/1/1968 (H3N2) virus cap-snatching process. *Sci Rep* **4**:6181. doi:10.1038/srep06181

Sikora D, Rocheleau L, Brown EG, Pelchat M. 2014. Deep sequencing reveals the eight facets of the influenza A/HongKong/1/1968 (H3N2) virus cap-snatching process. *Scientific Reports* **4**. doi:10.1038/srep06181

Singh G, Rebbapragada I, Lykke-Andersen J. 2008. A Competition between Stimulators and Antagonists of Upf Complex Recruitment Governs Human Nonsense-Mediated mRNA Decay. *PLOS Biology* **6**:e111. doi:10.1371/journal.pbio.0060111

Skehel JJ, Wiley DC. 2000. Receptor Binding and Membrane Fusion in Virus Entry: The Influenza Hemagglutinin. *Annual Review of Biochemistry* **69**:531–569. doi:10.1146/annurev.biochem.69.1.531

Slaine PD, MacRae C, Kleer M, Lamoureux E, McAlpine S, Warhuus M, Comeau AM, McCormick C, Hatchette T, Khaperskyy DA. 2018. Adaptive Mutations in Influenza A/California/07/2009 Enhance Polymerase Activity and Infectious Virion Production. *Viruses* **10**:272. doi:10.3390/v10050272

Sola I, Almazán F, Zúñiga S, Enjuanes L. 2015. Continuous and Discontinuous RNA Synthesis in Coronaviruses. *Annu Rev Virol* **2**:265–288. doi:10.1146/annurev-virology-100114-055218

Sun Y, Hu Z, Zhang X, Chen M, Wang Z, Xu G, Bi Y, Tong Q, Wang M, Sun H, Pu J, Iqbal M, Liu J. 2020. An R195K Mutation in the PA-X Protein Increases the Virulence and Transmission of Influenza A Virus in Mammalian Hosts. *Journal of Virology* **94**:e01817-19. doi:10.1128/JVI.01817-19

Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, Santos A, Doncheva NT, Roth A, Bork P, Jensen LJ, von Mering C. 2017. The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. *Nucleic Acids Research* **45**:D362–D368. doi:10.1093/nar/gkw937

Taddeo B, Sciortino MT, Zhang W, Roizman B. 2007. Interaction of herpes simplex virus RNase with VP16 and VP22 is required for the accumulation of the protein but not for accumulation of mRNA. *PNAS* **104**:12163–12168. doi:10.1073/pnas.0705245104

Taddeo B, Zhang W, Roizman B. 2010. Role of Herpes Simplex Virus ICP27 in the Degradation of mRNA by Virion Host Shutoff RNase. *J Virol* **84**:10182–10190. doi:10.1128/JVI.00975-10

Takata MA, Gonçalves-Carneiro D, Zang TM, Soll SJ, York A, Blanco-Melo D, Bieniasz PD. 2017. CG dinucleotide suppression enables antiviral defence targeting non-self RNA. *Nature* **550**:124–127. doi:10.1038/nature24039

Tanaka T, Kamitani W, DeDiego ML, Enjuanes L, Matsuura Y. 2012. Severe Acute Respiratory Syndrome Coronavirus nsp1 Facilitates Efficient Propagation in Cells through a Specific Translational Shutoff of Host mRNA. *Journal of Virology* **86**:11128–11137. doi:10.1128/JVI.01700-12

te Velthuis AJW, Oymans J. 2018. Initiation, Elongation, and Realignment during Influenza Virus mRNA Synthesis. *Journal of Virology* **92**:e01775-17. doi:10.1128/JVI.01775-17

Tenforde MW, Kondor RJG, Chung JR, Zimmerman RK, Nowalk MP, Jackson ML, Jackson LA, Monto AS, Martin ET, Belongia EA, McLean HQ, Gaglani M, Rao A, Kim SS, Stark TJ, Barnes JR, Wentworth DE, Patel MM, Flannery B. 2021. Effect of Antigenic Drift on Influenza Vaccine Effectiveness in the United States-2019-2020. *Clin Infect Dis* **73**:e4244–e4250. doi:10.1093/cid/ciaa1884

Terada Y, Kawachi K, Matsuura Y, Kamitani W. 2017. MERS coronavirus nsp1 participates in an efficient propagation through a specific interaction with viral RNA. *Virology* **511**:95–105. doi:10.1016/j.virol.2017.08.026

Tidu A, Janvier A, Schaeffer L, Sosnowski P, Kuhn L, Hammann P, Westhof E, Eriani G, Martin F. 2021. The viral protein NSP1 acts as a ribosome gatekeeper for shutting down host translation and fostering SARS-CoV-2 translation. *RNA* **27**:253–264. doi:10.1261/rna.078121.120

Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* **7**:562–578. doi:10.1038/nprot.2012.016

Tsai P-L, Chiou N-T, Kuss S, García-Sastre A, Lynch KW, Fontoura BMA. 2013. Cellular RNA Binding Proteins NS1-BP and hnRNP K Regulate Influenza A Virus RNA Splicing. *PLOS Pathogens* **9**:e1003460. doi:10.1371/journal.ppat.1003460

Turunen JJ, Niemelä EH, Verma B, Frilander MJ. 2013. The significant other: splicing by the minor spliceosome: Splicing by the minor spliceosome. *WIREs RNA* **4**:61–76. doi:10.1002/wrna.1141

Uyeki TM, Hui DS, Zambon M, Wentworth DE, Monto AS. 2022. Influenza. *The Lancet* **400**:693–706. doi:10.1016/S0140-6736(22)00982-5

van Hoof A, Frischmeyer PA, Dietz HC, Parker R. 2002. Exosome-Mediated Recognition and Degradation of mRNAs Lacking a Termination Codon. *Science* **295**:2262–2264. doi:10.1126/science.1067272

Vankadari N, Jeyasankar NN, Lopes WJ. 2020. Structure of the SARS-CoV-2 Nsp1/5′-Untranslated Region Complex and Implications for Potential Therapeutic Targets, a Vaccine, and Virulence. *J Phys Chem Lett* **11**:9659–9668. doi:10.1021/acs.jpclett.0c02818

Vasin AV, Temkina OA, Egorov VV, Klotchenko SA, Plotnikova MA, Kiselev OI. 2014. Molecular mechanisms enhancing the proteome of influenza A viruses: An overview of recently discovered proteins. *Virus Research* **185**:53–63. doi:10.1016/j.virusres.2014.03.015

Vreede FT, Chan AY, Sharps J, Fodor E. 2010. Mechanisms and functional implications of the degradation of host RNA polymerase II in influenza virus infected cells. *Virology* **396**:125–134. doi:10.1016/j.virol.2009.10.003

Wang F, Sheppard CM, Mistry B, Staller E, Barclay WS, Grimes JM, Fodor E, Fan H. 2022. The C-terminal LCAR of host ANP32 proteins interacts with the influenza A virus nucleoprotein to promote the replication of the viral RNA genome. *Nucleic Acids Res* **50**:5713–5725. doi:10.1093/nar/gkac410

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. 2020. The role of PA-X C-terminal 20 residues of classical swine influenza virus in its replication and pathogenicity. *Veterinary Microbiology* **251**:108916. doi:10.1016/j.vetmic.2020.108916

Wang Y, Shi H, Rigolet P, Wu N, Zhu L, Xi X-G, Vabret A, Wang X, Wang T. 2010. Nsp1 proteins of group I and SARS coronaviruses share structural and functional similarities. *Infection, Genetics and Evolution* **10**:919–924. doi:10.1016/j.meegid.2010.05.014

Wang Y, Tong X, Li G, Li J, Deng M, Ye X. 2012. Ankrd17 positively regulates RIG-I-like receptor (RLR)-mediated immune signaling. *European Journal of Immunology* **42**:1304–1315. doi:10.1002/eji.201142125

Wilamowski M, Gorecki A, Dziedzicka-Wasylewska M, Jura J. 2018. Substrate specificity of human MCPIP1 endoribonuclease. *Sci Rep* **8**:7381. doi:10.1038/s41598-018-25765-2

Wright P, Neumann G, Kawaoka Y. 2013. Orthomyxoviruses.Field's Virology, 6th Ed. Lippincott Williams & Wilkins, Philadelphia, PA, United States. pp. 1186–1243.

Wu J, Chen ZJ. 2014. Innate immune sensing and signaling of cytosolic nucleic acids. *Annu Rev Immunol* **32**:461–488. doi:10.1146/annurev-immunol-032713-120156

Xu G, Zhang X, Liu Q, Bing G, Hu Z, Sun H, Xiong X, Jiang M, He Q, Wang Y, Pu J, Guo X, Yang H, Liu J, Sun Y. 2017. PA-X protein contributes to virulence of triple-reassortant H1N2 influenza virus by suppressing early immune responses in swine. *Virology* **508**:45–53. doi:10.1016/j.virol.2017.05.002

Xu G, Zhang X, Sun Y, Liu Q, Sun H, Xiong X, Jiang M, He Q, Wang Y, Pu J, Guo X, Yang H, Liu J. 2016. Truncation of C-terminal 20 amino acids in PA-X contributes to adaptation of swine influenza virus in pigs. *Sci Rep* **6**:21845. doi:10.1038/srep21845

Yamayoshi S, Watanabe M, Goto H, Kawaoka Y. 2015. Identification of a Novel Viral Protein Expressed from the PB2 Segment of Influenza A Virus. *Journal of Virology* **90**:444–456. doi:10.1128/JVI.02175-15

Yang X, Steukers L, Forier K, Xiong R, Braeckmans K, Van Reeth K, Nauwynck H. 2014. A beneficiary role for neuraminidase in influenza virus penetration through the respiratory mucus. *PLoS One* **9**:e110026. doi:10.1371/journal.pone.0110026

Yang Y, Tang H. 2016. Aberrant coagulation causes a hyper-inflammatory response in severe influenza pneumonia. *Cell Mol Immunol* **13**:432–442. doi:10.1038/cmi.2016.1

Yeo G, Burge CB. 2004. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J Comput Biol* **11**:377–394. doi:10.1089/1066527041410418

Yi H, Park J, Ha M, Lim J, Chang H, Kim VN. 2018. PABP Cooperates with the CCR4-NOT Complex to Promote mRNA Deadenylation and Block Precocious Decay. *Mol Cell* **70**:1081-1088.e5. doi:10.1016/j.molcel.2018.05.009

Younis I, Berg M, Kaida D, Dittmar K, Wang C, Dreyfuss G. 2010. Rapid-response splicing reporter screens identify differential regulators of constitutive and alternative splicing. *Mol Cell Biol* **30**:1718–1728. doi:10.1128/MCB.01301-09

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. 2009. Crystal structure of an avian influenza polymerase PA_N reveals an endonuclease active site. *Nature* **458**:909–913. doi:10.1038/nature07720

Yuan S, Balaji S, Lomakin IB, Xiong Y. 2021. Coronavirus Nsp1: Immune Response Suppression and Protein Expression Inhibition. *Frontiers in Microbiology* **12**:2683. doi:10.3389/fmicb.2021.752214

Yuan S, Peng L, Park JJ, Hu Y, Devarkar SC, Dong MB, Shen Q, Wu S, Chen S, Lomakin IB, Xiong Y. 2020. Nonstructural Protein 1 of SARS-CoV-2 Is a Potent Pathogenicity Factor Redirecting Host Protein Synthesis Machinery toward Viral RNA. *Molecular Cell* **80**:1055-1066.e6. doi:10.1016/j.molcel.2020.10.034

Zhai J, Arikit S, Simon SA, Kingham BF, Meyers BC. 2014. Rapid construction of parallel analysis of RNA end (PARE) libraries for Illumina sequencing. *Methods* **67**:84–90. doi:10.1016/j.ymeth.2013.06.025

Zhang J, Li G, Ye X. 2010. Cyclin T1/CDK9 interacts with influenza A virus polymerase and facilitates its association with cellular RNA polymerase II. *Journal of Virology* **84**:12619–12627. doi:10.1128/JVI.01696-10

Zhang K, Xie Y, Muñoz-Moreno R, Wang J, Zhang L, Esparza M, García-Sastre A, Fontoura BMA, Ren Y. 2019. Structural basis for influenza virus NS1 protein block of mRNA nuclear export. *Nat Microbiol* **4**:1671–1679. doi:10.1038/s41564-019-0482-x

Zhang L, Wang J, Muñoz-Moreno R, Kim M, Sakthivel R, Mo W, Shao D, Anantharaman A, García-Sastre A, Conrad NK, Fontoura BMA. 2018. Influenza Virus NS1 Protein-RNA Interactome Reveals Intron Targeting. *J Virol* **92**:e01634-18. doi:10.1128/JVI.01634-18

Zhang W, Xie M, Shu M-D, Steitz JA, DiMaio D. 2016. A proximity-dependent assay for specific RNA–protein interactions in intact cells. *RNA* **22**:1785–1792. doi:10.1261/rna.058248.116

Zhang Z, Dong L, Zhao C, Zheng P, Zhang X, Xu J. 2021. Vaccinia virus-based vector against infectious diseases and tumors. *Human Vaccines & Immunotherapeutics* **17**:1578–1585. doi:10.1080/21645515.2020.1840887

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. 2018. Influenza virus infection causes global RNAPII termination defects. *Nat Struct Mol Biol* **25**:885–893. doi:10.1038/s41594-018-0124-7

Zhou Z, Licklider LJ, Gygi SP, Reed R. 2002. Comprehensive proteomic analysis of the human spliceosome. *Nature* **419**:182–185. doi:10.1038/nature01031

Zhu Y, Wang X, Forouzmand E, Jeong J, Qiao F, Sowd GA, Engelman AN, Xie X, Hertel KJ, Shi Y. 2018. Molecular Mechanisms for CFIm-Mediated Regulation of mRNA Alternative Polyadenylation. *Mol Cell* **69**:62-74.e4. doi:10.1016/j.molcel.2017.11.031