Proteomic and functional studies of the influenza A virus PA-X protein

by

Brittany Porter

Submitted in partial fulfilment of the requirements
for the degree of Master of Science

at

Dalhousie University
Halifax, Nova Scotia
January 2018

For Ted, Kay, and Reinhilde

Table of Contents

List of Tables

List of Figures

Abstract

Viral RNA endonuclease activity is required for influenza A virus (IAV) replication. This activity resides in the polymerase acidic (PA) protein, which assembles into viral RNA-dependent RNA polymerase (RdRp) complexes and cleaves nascent host pre-mRNAs proximal to 5'-m^7G caps, creating primers for viral mRNA synthesis. A rare (+1) ribosomal frameshifting event during translation of the PA open reading frame (ORF) creates the polymerase acidic-X (PA-X) protein. PA-X retains the amino-terminal PA RNA endonuclease domain, but contains a novel short carboxy-terminus, dubbed the X-ORF. Accumulating evidence indicates PA-X is a host shutoff protein functioning in the nucleus, selectively cleaving RNAs transcribed by host RNA polymerase II (pol II) while sparing RNA pol I, III, and viral transcripts. The molecular mechanism for this specificity remains to be elucidated. I hypothesize that PA-X gains access to target RNAs by X-ORF-mediated interaction with host proteins. In this study, I used a proximity labeling proteomic method known as BioID to identify host proteins that interact with the X-ORF. In BioID, fusion of the bait protein to a promiscuous biotin ligase allows efficient biotinylation of lysine residues on nearby proteins. X-ORF baits subjected to BioID included a 61-amino acid variant from A/Puerto Rico/8/1934 (H1N1) and a truncated 41-amino acid variant from A/California/7/2009 (H1N1), as well as a mutant X-ORF lacking basic residues required for nuclear localization. Affinity-purified proteins were trypsinized, subjected to reductive dimethylation with stable isotope tags, and identified by mass spectrometry. Using quantitative analysis, 29 high-confidence candidate X-ORF-interacting proteins were identified. X-ORF interacting proteins were validated using a luciferase-based functional assay in cells where each candidate host gene was silenced by short-hairpin RNAs. Through this study, the cleavage factor I (CFI_m) complex proteins, cleavage and polyadenylation specificity factor subunit 5 (CPSF5) and CPSF6, were identified as required for PA-X function. The CFI_m complex is poorly characterized but is known to influence site selection for mRNA 3'-end cleavage and polyadenylation. My observations are concordant with the emerging model for PA-X host shutoff activity, which has been shown to require canonical mRNA 3'-end processing mechanisms.

List of Abbreviations Used

5'-m^{7}G	5' -7 methyl guanosine cap
ACN	Acetonitrile
ADP	Adenosine Diphosphate
AMP	Adenosine Monophoshphate
AP-MS	Affinity Purification - Mass Spectrometry
ARE	AU-Rich Elements
ATP	Adenosine Triphosphate
BirA	Biotin Ligase
BirA*	Promiscuous Biotin Ligase
BSA	Bovine Serum Albumin
CBC	Cap-Binding Complex
CMV	Cytomegalovirus
CFIm	Cleavage Factor Im
CPSF	Cleavage And Polyadenylation Specificity Factor
cRNA	Complementary Ribonucleic Acid
cRNP	Complementary Ribonucleoprotein
CstF	Cleavage Stimulation Factor
CTD	C-Terminal Domain
DMEM	Dulbecco's Modified Eagle's Medium
DNA	Deoxyribonucleic Acid
dsRBD	Double-Stranded RNA Binding Domain
dsRNA	Double-Stranded Ribonucleic Acid
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic Acid
EGTA	Ethylene-Bis(Oxyethylenenitriolo)Tetraacetic Acid
eIF	Eukaryotic Initiation Factor
FBS	Fetal Bovine Serum
FS	Frameshift
GADD45B	Growth Arrest And DNA Damage Inducible 45 β
GDP	Guanosine Diphosphate
GTP	Guanosine Triphosphate
HA	Haemagglutinin
HEK	Human Embryonic Kidney
HRP	Horseradish Peroxidase
HSV	Herpes Simplex Virus
IAV	Influenza A Virus
IBB	Importin-β Binding Domain
IF	Immunofluorescence
IFN	Interferon

IL-6	Interleukin-6
KSHV	Kaposi's sarcoma-associated herpes virus
LB	Lysogeny Broth
LC-MS/MS	Liquid Chromatography - Tandem Mass Spectrometry
mRNA	Messenger Ribonucleic Acid
NA	Neuraminidase
$NaCNBD_3$	Sodium cyanoborodeuteride
$NaCNBH_3$	Sodium cyanoborohydride
NEP	Nuclear Export Protein
NES	Nuclear Export Signal
NLS	Nuclear Localization Signal
NP	Nucleoprotein
NP-40	Nonyl Phenoxypolyethoxylethanol
NPC	Nuclear Pore Complex
ORF	Open Reading Frame
PA	Polymerase Acidic
PABP	Poly(A) Binding Protein
PAP	Poly(A) Polymerase
PAR-CLIP	Photoactivatable Ribonucleoside Enhanced Crosslinking and Immunoprecipitation
PB1	Polymerase Basic-1
PB2	Polymerase Basic-2
PBS	Phosphate Buffered Saline
PEI	Polyethylenimine
pH1N1	Pandemic H1N1
PIC	Protease Inhibitor Cocktail
PKR	Protein Kinase R
PTC	Peptidyl-Transferase Complex
Puro	Puromycin
PVDF	Polyvinylidene Difluroide
qPCR	Quantitative Polymerase Chain Reaction
RBD	RNA Binding Domain
RIPA	Radioimmunoprecipitation Assay Buffer
RLU	Relative Light Unit
RNA	Ribonucleic Acid
RNA Pol II	RNA Polymerase II
Rnase	Ribonuclease
RNP	Ribonucleoprotein
RT	Reverse Transcriptase
SDS-PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
shRNA	Short Hairpin RNA

snRNA	Small Nuclear Ribonucleic Acid
snRNPs	Small Nuclear Ribonucleoproteins
SOX	Shutoff And Exonuclease
ssRNA	Single-Stranded Ribonucleic Acid
Strep	Streptavidin
TBS	Tris-Buffered Saline
TBS-T	Tris-Buffered Saline + Tween 20
TFA	Trifluoroacetic acid
TGN	*Trans*-Golgi Network
TM	Transmembrane
TOPO2B	Topoisomerase II B
tRNA	Transfer Ribonucleic Acid
UTR	Untranslated Region
vhs	Virion Host Shutoff Protein
vRNA	Viral Ribonucleic Acid
vRNP	Viral Ribonucleoprotein
WT	Wild Type
Y2H	Yeast-2-Hybrid

Acknowledgements

First, I would like to thank my supervisor, Dr. Craig McCormick, for not only allowing me the opportunity to explore the world of molecular virology, but also for all the academic support over the course of my degree; pushing me to further my understanding in the field.

To my committee members Dr. John Rohde and Dr. Roy Duncan, thank you for all of your expertise and input into making this project what it turned out to be.

I would also like to thank the expertise and guidance I have received from training with Dr. Denys Khaperskyy. Thank you for taking the time to train me and allowing me to ask questions along the way.

Also, I would like to thank Dr. Alejandro Cohen of the proteomics CORE facility at Dalhousie University for taking the time to train me in labelling techniques used in this thesis, as well as conducting the data base searches for my mass spectrometry analysis.

To all my lab members, thank you for your support both academic and moral. To Team Flu, your input and ideas have helped shape this thesis. Special thanks to Ben Johnston, Carolyn Robinson and Eric Pringle, for all the advice and guidance you have given me in the course of my project. It would not have been as successful if it were not for you.

Thank you to Stephen Whitefield for training in the Cellular and Molecular Digital Imaging CORE Facility at Dalhousie University.

To the funding agencies who allowed the financial support for this project to be completed: NSERC, CHIR, Nova Scotia Research and Innovation Graduate Scholarship from Faculty of Graduate Studies at Dalhousie University, and Dalhousie University.

Last, but not least, to my friends and family, thank you for all the love and support through this journey of science. It means the world to me to have you in my corner.

Chapter 1: Introduction

Influenza A virus (IAV) nuclear replication relies on the endonuclease activity of the polymerase acidic (PA) component of the RNA dependent RNA polymerase (RdRp). PA cleaves host mRNA near the 5'-m^{7}G cap, which then serves as a primer for viral mRNA synthesis. IAV has multiple host shutoff mechanisms that limit the accumulation of host antiviral proteins, while giving viral mRNAs priority access to the host protein synthesis machinery. A new IAV host shutoff protein called PA-X was recently discovered; PA-X comprises the N-terminal endonuclease domain of PA and a unique C-terminus called the X-ORF. PA-X accumulates in the nucleus and degrades host RNA pol II transcripts, while sparing viral and RNA pol I and III transcripts. This introduction will review the IAV replication cycle, known IAV host shutoff mechanisms, and detail PA-X biosynthesis and known properties. This introduction also details a proximity labelling proteomic method known as BioID, which played a central role in the discovery of new candidate PA-X interacting proteins.

1.1 Influenza virus entry and replication

IAV is a member of the *Orthomyxoviridae* family of viruses. The segmented, single-stranded RNA (ssRNA) genome produces at least 14 proteins [1,2]. The virion is wrapped in a host-derived lipid envelope studded with three different viral proteins: haemagglutinin (HA), neuraminidase (NA), and the M2 ion channel [3]. The virion interior contains the abundant matrix 1 (M1) protein and eight viral

ribonucleoproteins (vRNPs), comprising a single viral genomic RNA segment bound to nucleoprotein (NP) and the tripartite viral RNA-dependent RNA polymerase (RdRp; PA (polymerase acidic protein), PB1 (polymerase basic protein 1), and PB2 (polymerase basic protein 2)) [3]. HA and NA are the most divergent IAV proteins; there are 18 known subtypes of HA and 11 known subtypes of NA, which are often found in different combinations [4]. IAV nomenclature is derived from these subtype groupings characterizing the ability of HA and NA bind specific antibodies [5]. Figure 1 illustrates the general features of IAV replication, which will be discussed in greater detail in the following sections.

1.1.1 IAV attachment and entry into host cells

IAV infection begins when trimeric HA glycoproteins bind sialic acid receptors on the surface of epithelial cells [6-9]. The mature active form of HA is a result of cleavage of the HA_0 polypeptide by host proteases resulting in the production of two disulphide-linked domains: the globular HA_1 and stem-shaped HA_2 [10-12]. The HA trimer interacts with sialic acid moieties which are conjugated to galactose sugars via α(2,3)- or α(2,6)-linkages. HA proteins from human IAV isolates typically bind to α(2,6)-linkages found in upper respiratory epithelial cells, whereas HA proteins from equine and avian IAV isolates bind to α(2,3)-linkages commonly found on epithelial cells of the bird intestines and the lower respiratory tract in humans [7,13]. These preferences create a barrier to transmission between host species. The abundance of both types of sialic acid linkages in pigs enables co-infection between divergent IAV isolates and emergence of re-assortant viruses with

new properties [14]. This process, known as 'antigenic shift', is a major source of new viruses with pandemic potential [14].

Once the virus binds to the sialic acid residues it is brought into the cell through a process formerly known as viropexis, more commonly known as receptor-mediated endocytosis [15,16]. The virus envelope must fuse with the host endosome membrane to allow entry of the viral genome into the host cell and for viral replication to begin [15].

1.1.2 The role of M2 in IAV uncoating

Fusion of the host endosome membrane with the viral envelope occurs once endosomal pH drops below approximately 5.2 [17]. In acidic conditions, the HA trimer undergoes conformational changes whereby the structure of HA_1 and HA_2 is rearranged to allow the globular head of the trimer to disassociate from the stem of HA [18]. This rearrangement exposes the highly conserved fusion peptide found in HA_2 [19]. HA fusion peptide penetrates the endosome membrane initiating a dramatic membrane deformation and close apposition of host and viral membranes to promote membrane mixing and pore formation [20].

The M2 ion channel plays an important role in the final stage of viral entry, known as "uncoating", in which the viral genome segments are released into the cytoplasm. Acidification occurs by transport of protons from the endosome through the M2 ion channel into the virion core causing the release of M1 proteins from vRNPs and exposing the nuclear localization signals (NLSs) that direct translocation of vRNPs to the nucleus [21,22].

The role of M2 in IAV replication was elucidated during investigations of the mechanism of action of the antiviral drug amantadine [23]. Hay *et al.* determined that in the presence of amantadine, early stages of viral replication were inhibited [23]. Furthermore, they determined it was the product of the M genes that were susceptible to the effect of the drug [23]. At the molecular level, amantadine was shown to block the M2 ion channel through interaction with hydrophobic residues in the channel [23,24]. Though it was an effective antiviral, global resistance to amantadine quickly spread among H3N2 viruses; prior to 2002 resistance occurred in 1-2% of H3N2 viruses and progressed to 90.5% resistance in the 2005-2006 season [25]. Amantadine resistance arises due to a single amino acid mutation in the M2 ion channel, with the most prevalent mutation being Ser31Asn [25].

1.1.3 IAV transcription and translation

The viral genomic RNA is segmented and packaged into vRNPs consisting of a strand of genomic RNA, an abundance of nucleoproteins (NP), and a single RdRp[1]. Acidification of the virion causes M1 to dissociate from the vRNPs, exposing nuclear localization signals that enable translocation to the nucleus (the site of viral transcription and genome replication) [1,22]. Genomic RNA is held in a helical shape with a highly conserved terminal panhandle structure [26,27]. The structure is maintained through the association of NP at a ratio of approximately 20 nucleotides RNA per subunit of NP [27]. This association is also what promotes the translocation of vRNA to the nucleus. NP contains two NLS motifs that are recognized by host import machinery and guide the vRNPs through the nuclear pore complex [28-30].

In the nucleus, the RdRp is responsible for synthesizing viral mRNA and genome replication [31]. The initial step of viral mRNA synthesis is known as "cap-snatching", where by PB2 binds the 5'-m^{7}G cap of a host mRNA aligning the mRNA into the endonuclease domain of PA [32,33]. The 5'-m^{7}G cap is cleaved at a fixed position 10-13 nucleotides downstream from the cap and acts as a primer for transcription [34]. Once cap-snatching occurs, the vRNA is translocated through the polymerase complex to the active site in PB1 aided by rearrangement of NP on the RNA [35]. Viral mRNAs are transcribed until the RdRp detects a polyuridine region [36]. Reiterative decoding of the polyuridine region by the RdRp produces a poly (A) tail on the viral mRNA [36].

The primary product of IAV segment 7 is the M1 protein, however mRNA splicing permits the production of the M2 ion channel [37]. A similar event occurs on genome segment 8. The primary unspliced product of segment 8 is the non-structural protein 1 (NS1) [38]. The splicing event allows for the production of non-structural protein 2 (NS2), sometimes referred to as the nuclear export protein (NEP) [38].

Once the mature viral mRNA is produced, it must be exported to the cytoplasm to be translated. IAV transcripts are exported via the Transcription-export (TREX) pathway [39]. More recently, viral mRNA export has been found to be more complex and utilizes different pathways in a cell dependent manner [40]. In A549 lung epithelial cells HA, NA and NP transcripts rely on the TREX pathway and association with NXF1 for mRNA export from the nucleus [40]. However, PA, PB1, and PB2 transcripts did not rely on the NXF1 pathway or the CRM1 pathway for

mRNA export, suggesting an alternative pathway may be required for export of certain viral mRNA transcripts [40]. Irrespective of export pathway, all viral mRNAs translocate to the cytoplasm to undergo translation.

IAV does not encode its own translation machinery and therefore must rely on translation complexes in the host cell cytoplasm to produce viral proteins [41]. The IAV NS1 protein enables the efficient production of viral proteins [42]. By recognizing and interacting with the 5'-UTRs of viral mRNAs, NS1 ensures specificity and an increased rate of translation for IAV proteins [42,43]. The 5'-UTR of certain viral mRNAs, such as M1 and NP, contain conserved sequences which stimulate translation through interaction with NS1 [43]. Other viral transcripts, such as NA mRNA, do not contain this same sequence and as a result are produced in less abundance [43]. However, it is the association between NS1, eIF4G, and PABP1 that aid in IAV translation initiation [44].

1.1.4 IAV genome replication, virion assembly and release

During viral replication, the RdRp transcribes an intermediate positive strand RNA (cRNA) in the complementary RNP (cRNP) complex [45]. The cRNA is then used as a template to create the negative-sense genomic strand of RNA for the virus [45]. The cRNP complex is made out of necessity, as the cRNA will be degraded unless freestanding NP and RdRps associate with the RNA [46]. Initiation of cRNP formation occurs due to a multitude of events, primarily the accumulation of viral mRNAs and the interaction of NP with the RdRp [47,48] . Upon cRNP formation, a small viral RNA is synthesized and loaded into PA, allowing genomic viral RNA to be

transcribed [49] . A trans-acting RdRp is also required to activate the RdRp on the cRNP to produce the genomic vRNA [45].

In late stages of viral infection, newly synthesized vRNA genomes in vRNPs are transported out of the nucleus using the CRM-1 nuclear export pathway [1]. M1 associates with the vRNPs in the nucleus followed by the recruitment of NEP (NS2); NEP acts as an adaptor between the M1-vRNP complex and nuclear pore complex (NPC) promoting export of the vRNP complex [50]. Upon exiting the nucleus, vRNP-M1 complexes localize to the perinuclear endosomal-recycling complex where they interact with Rab11 [51]. The vRNP complex then utilizes the microtubule network to traffic to the site of budding on the cellular plasma membrane [51]. At the plasma membrane, the vRNP-M1 complex associates with the M2 protein for incorporation of the vRNPs in the virion [52,53].

M1 associates with HA and NA upon their localization to the plasma membrane [57]. This interaction occurs once HA and NA are delivered to the plasma membrane [54,55]. The trans-membrane domains of HA and NA contain sequences that enable association with lipid rafts within the trans-Golgi Network (TGN), allowing for delivery to the apical plasma membrane of the cell [56-58].

The initial action in viral budding is the localization of HA to plasma membrane lipid rafts [59]. Acylation to the C-terminal domain of HA must occur in order to obtain the curvature and proper budding of the virion [60]. This process has also been shown to promote the formation of the M1 layer within the virion [60]. M2 recruitment to the virion further promotes the budding process. The cytoplasmic amphipathic helix binds cholesterol, which further alters the budding of

the virion [59,61]. Localization of M2 is predominately at the neck of the budding virion where the amphipathic helix controls scission [59]. NA cleavage of α(2,6)- or α(2,3)-linked sialic acid moieties is required for virion escape from the surface of the host cell [4]. This action is crucial for IAV infectivity. If NA does not cleave the sialic acid moieties the virus will remain tethered to the host cell and virions will not be released [62].

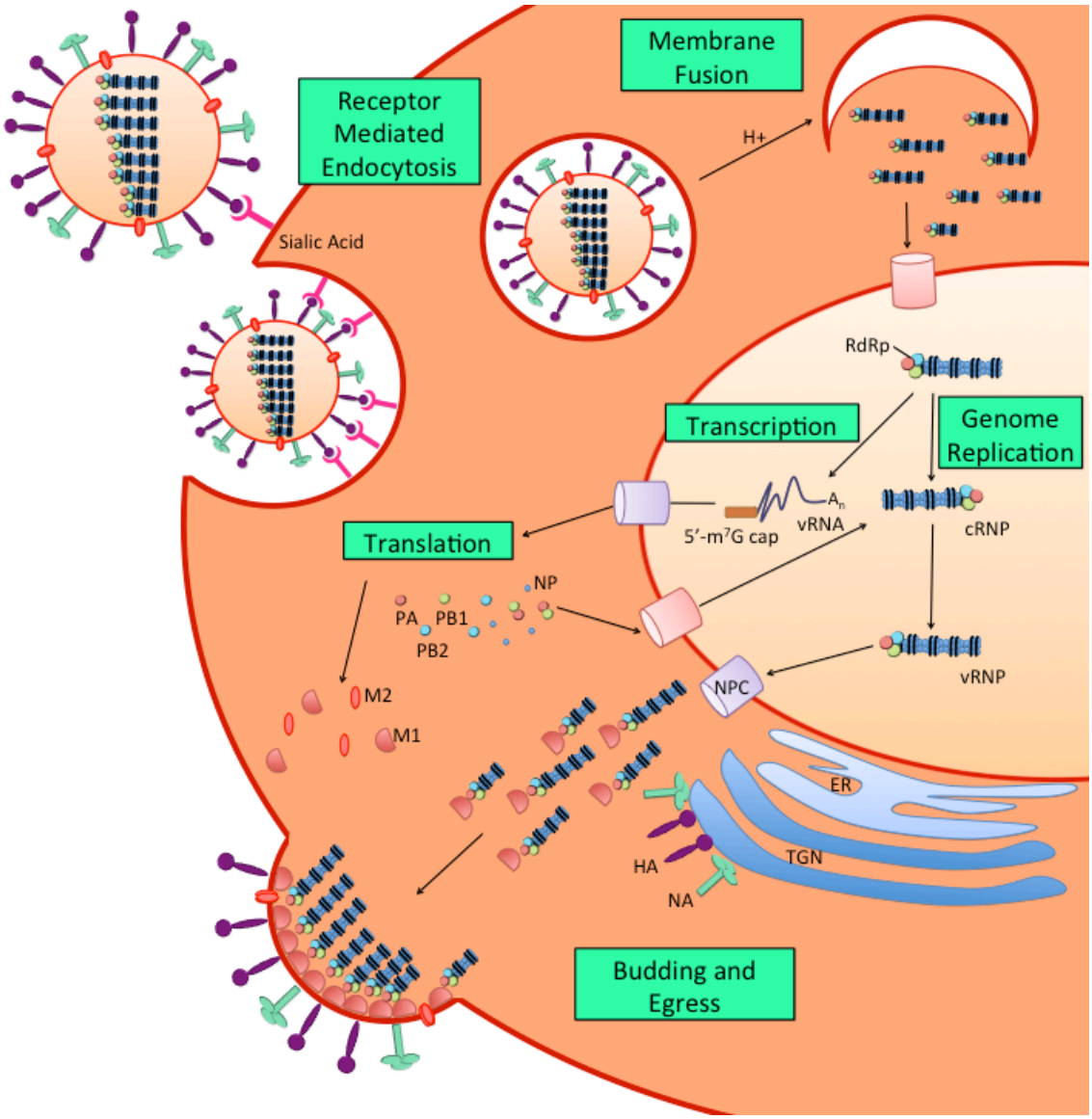


Figure 1: IAV replication cycle
Modified from [35] and [63]

1.2 Host shutoff mechanisms of IAV

Known strategies of IAV host shutoff include: the cap-snatching mechanism of RNA pol II transcripts, inhibition of splicing, 3'-end processing, and host mRNA export, and NS1 blocking PKR activation.

1.2.1 PA cap-snatching activity

The PA component of the RdRp cleaves the 5'-m^7G cap from cellular pre-mRNAs (illustrated in Figure 2) [35]. The PA active site is a type II endonuclease, similar to the common restriction endonuclease *Eco*RV [64]. PA requires the divalent cation, Mn^{2+}, to stabilize its structure in the active form [64]. The catalytic active site has been mapped to residues $P_{107}D_{108}X_{10}E_{119}K_{134}$ in the N-terminus of PA [65]. K134 has been identified as the catalytic residue in the active site, whereas D108 is the Mn^{2+} binding site; mutations to these residues individually results in loss of PA function [65,66]. Proximity of the viral RdRp to the C-terminal domain (CTD) of host RNA pol II enables the cap-snatching mechanism to occur [67]. The RdRp associates with the phosphorylated phosphoserine-5 on the CTD of RNA pol II at early stages of host transcription, coincident with the addition of the 5'-m^7G cap [67].

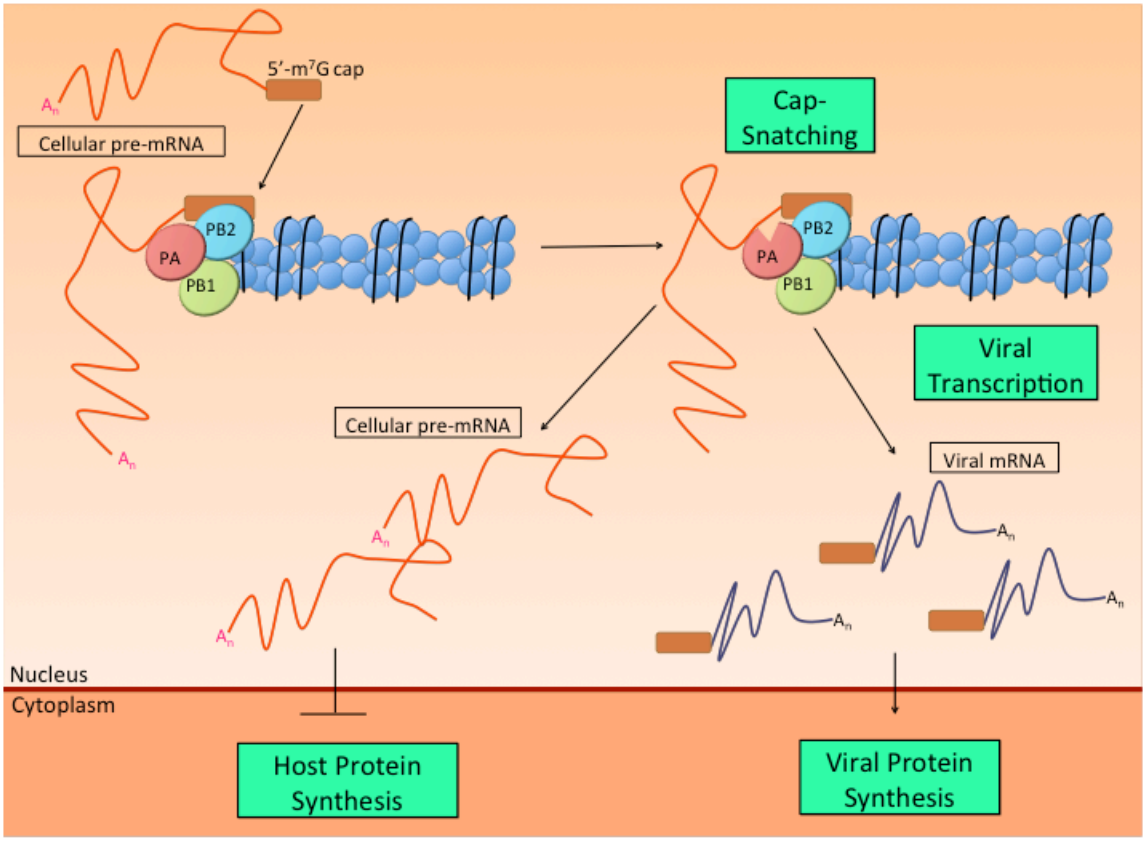


Figure 2: IAV PA removes 5'-m^{7}G cap from host mRNA transcripts to prime viral transcription

PB2 binds host pre-mRNA 5'-m^{7}G cap aligning the host RNA into the PA endonuclease domain. PA cleavage of mRNA releases a fragment that the RdRp can use to initiate transcription. Host pre-mRNAs are retained in the nucleus and degraded, whereas viral mRNAs are exported and translated into proteins. Modified from [68].

1.2.2 NS1 inhibits mRNA processing: splicing, 3'-end processing, and export

NS1 can be divided into two separate domains, the N-terminal RNA binding domain and the C-terminal effector domain [69]. NS1 has several implications in IAV host shutoff within the cell including blocking of the splicing machinery from editing host pre-mRNAs.

The spliceosome, which is composed of several snRNPs (small nuclear ribonucleoproteins), orchestrates mRNA splicing reactions [70]. In the first steps snRNPs U1 and U2 bind to the 5'- and 3'- splice sites respectively; upon binding, U1 cleaves the 5'-end of the intron allowing attack on the branch site, which lies upstream of the 3'-splice site [70,71]. The branch site is denoted by an adenine, this adenine is attacked by the 5'- end of the intron to form the lariat intermediate [70]. The snRNP U4-U6 complex mediates lariat formation [70]. The U2, U6 and U5 snRNPS then aid in intron 3'-end cleavage; after cleavage of the lariat, the two ends of the exon are then fused together [70,71].

NS1 has been shown to inhibit mRNA splicing [72]. NS1 directly binds to the U6 snRNP through two predicted stem loop structures at residues 38-43 and 88-95 on the U6 snRNA [73]. NS1 disrupts both U6-U2 and U6-U4 snRNA complex formation, thereby impeding the splicing reaction[73].

Normally, RNA pol II transcripts undergo 3'-end processing to produce a polyadenylated tail; utilizing the cleavage and polyadenylation specificity factor (CPSF) complex, cleavage stimulation factor (CstF), and the cleavage factor I and II (CFIm and CFIIm) complexes [74]. Other proteins that are involved in 3'-end processing include nuclear PABP and the C-terminus of RNA pol II [74]. The CPSF

complex (CPSF73, CPSF160 and CPSF30) binds to the AAUAAA sequence upstream of the cleavage site on the host pre-mRNAs where they cleave and signal for polyadenylation to occur [74,75].

NS1 binding to CPSF30 prevents CPSF30's ability to bind host pre-mRNA, therefore inhibiting cleavage and polyadenylation [75]. The NS1 effector domain also interacts with nuclear PABP [76]. Chen *et al.* determined that even in the presence of NS1-CPSF30 binding, mRNAs with shortened poly (A) tails of approximately 12 bases are formed [76]. These short poly (A) tails cannot be elongated due to the lack of PABP II and nuclear export is blocked [76]. Subsequently, CPSF30 and PABP II could bind to NS1 simultaneously at different sites, indicating the requirement of both CPSF30 and PABP II NS1 binding to prevent poly (A) tail formation [76].

The majority of cellular transcripts utilize the TREX pathway for export of mRNA from the nucleus to the cytoplasm; the CRM-1 export pathway handles a limited set of mRNAs, including those that encode cell-cycle regulating proteins [77]. IAV utilizes another level of regulation to inhibit the export of cellular mRNAs by disrupting the TREX pathway complex of NXF1, p15, Rae1 and E1B-AP5 [78]. NS1 binds the TREX pathway complex preventing NXF1 from acquiring cargo to carry to the NPC [77,78].

1.2.3 NS1 inhibition of PKR activation and translation arrest

NS1 selectively enhances the translation of viral mRNA through interactions in the 5'-UTR and translation initiation complex components [43,44]. However, NS1

can indirectly influence translation through its interactions with host protein kinase R (PKR).

In IAV infection the viral RNA panhandle structure of vRNP is able to stimulate PKR in the same manner as dsRNA [79]. To prevent activation of the integrated stress response, NS1 interacts with the stress response protein PKR [80]. PKR is a serine/threonine kinase that, in the presence of dsRNA, dimerizes and binds RNA with its N-terminal RNA binding domain (RBD) [81]. dsRNA binds to the two dsRBD in PKR in a sequence-independent manner [82]. PKR is found in the cytoplasm as both a monomer and dimer, however in the presence of dsRNA there is a shift towards its active dimeric form [83]. RNA binding promotes a conformational change in the PKR catalytic domain [83]. When dimerized, an autophosphorylation event occurs leading to the phosphorylation of the α subunit of eIF2 (eIF2α) [81]. The phosphorylation of eIF2α is detrimental to the translation of mRNA within the cell. eIF2 is a component of the eIF2-GTP- Met-tRNA$_i$ complex allowing initiation of translation [84]. During translation initiation, eIF2-GTP undergoes hydrolysis to eIF2-GDP and must be exchanged back to GTP form; this is done through the guanine nucleotide exchange factor (GEF) eIF2B [81]. Phosphorylation of eIF2α prevents the exchange of GDP to GTP by eIF2B and therefore inhibits translation initiation [85].

Prevention of eIF2α phosphorylation can be done though inhibition of PKR activation solely through interactions of the N-terminus of NS1 [79]. NS1 has the ability to bind both dsRNA and the interferon-inducible dsRNA dependent protein kinase activator A (PACT) [86]. The activating domain of PACT binds to the C-

terminus of PKR causing a conformational change, promoting activation and phosphorylation of eIF2α [86]. NS1 inhibits this activity and that of the dsRNA binding to PKR though binding to the N-terminal of PKR itself [86]. Binding of NS1 to PKR is done in a dsRNA independent manner [86].

In summary, NS1 has the ability to promote IAV translation by preventing eIF2α phosphorylation (and global translation inhibition) and inhibition of host mRNA processing events such as splicing, 3'-end processing, and nuclear export.

1.3 Newly discovered IAV host shutoff protein PA-X

1.3.1 Formation of PA-X, a secondary product of PA

It has long been known that segment 3 of the IAV genome encodes the un-spliced product PA [87]. More recently, a conserved sequence with no known structural motif was discovered within the PA open reading frame (ORF) between codons 200-250 [88]. No start codons or alternative translation initiation sites were evident in this sequence [88]. In 2012, Jagger *et al*. determined this sequence was an overlapping ORF within the PA segment, which was dubbed the X-ORF [87]. Deep sequencing analysis of several IAV PA sequences revealed a 9 codon-sliding window with synonymous mutations and a prominent site of conservation between amino acid residues 190-253 [87].

The mechanism of production of this ORF was determined to be due to a rare +1 ribosomal frameshift during translation [87]. Near the 5'-end of the X-ORF is a conserved sequence of 5'-UCC UUU CGU C-3' where slippage occurs when UUU is in the P site of the ribosome and CGU in the A site; CGU is a rare codon in both

mammals and birds [87]. UUU and UUC share the same anti-codon encoding for Phe, however UUU is much weaker at binding the anti-codon than that of UUC [89]. PA-X, a 29-kDa protein retaining the N-terminal RNase domain of PA, is the product of this ribosomal frameshift (Figure 3) [87,90]. Under normal circumstances the rate at which the frameshift occurs is between 1.2-2.5% [87]. This rate is controlled by the rare CGU codon in the A site, UCC in the E site and 5'-distal control elements [87]. PA-X is produced in all IAV strains, containing either 61 or 41 amino acid long X-ORFS.

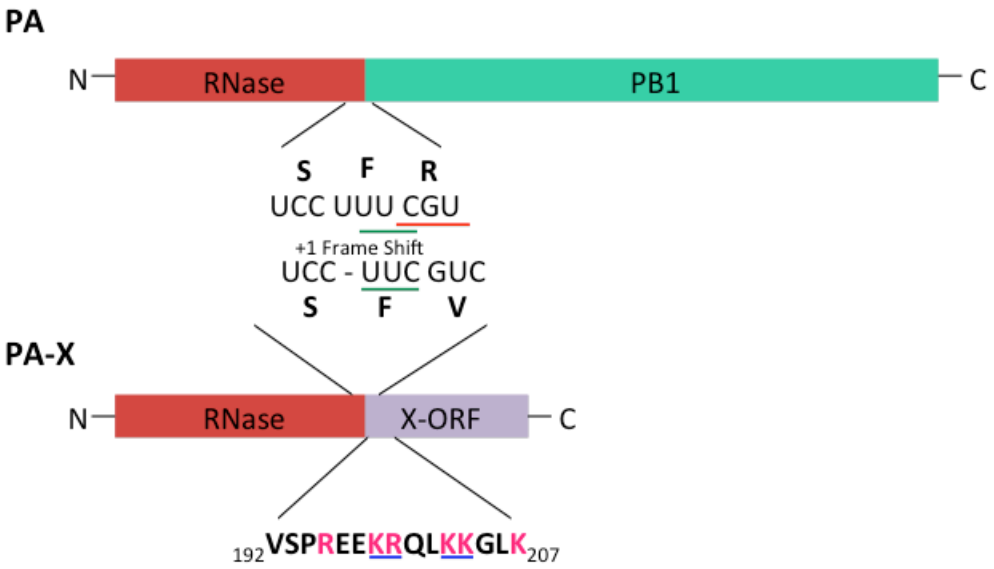


Figure 3: Novel host shutoff protein PA-X produced via a rare +1 ribosomal frameshift

RdRp component PA (upper panel) contains an N-terminal ribonuclease (RNase) domain and a long C-terminal PB1 binding domain. During translation a rare +1 ribosomal frameshift occurs producing a protein containing the first 191 amino acids of PA (the RNase domain) and a novel 61 or 41 amino acid C-terminus, deemed PA-X (bottom panel) [87] . The plus one ribosomal frameshift occurs due to a rare codon, CGU (underlined in red), and a wobble codon for Phe (underlined in green). Amino acids 192-207 of PA-X are crucial for PA-X host shutoff activity (pink) [91]. Blue underlining indicates the four basic amino acids that are required for nuclear import of PA-X. Modified from [92] and [93] .

The most common length of the X-ORF is 61 amino acids, however, a truncated X-ORF of 41 amino acids can occur [87]. Indeed, truncated forms of the X-ORF are commonly found in swine, equine, canine, and bat IAVs [90]. These truncated forms are due to a stop codon at residue 42 of the X-ORF [90]. Full length PA-X is found in all avian and human IAVs, including the 1918 H1N1 pandemic strain [90].

PA-X is found in IAVs, but completely absent from influenza B viruses [90]. Prior to 1985, all swine IAVs had the 61 amino acid X-ORF; but since then most swine viruses bore the truncated 41 amino acid X-ORF [94] . This truncated X-ORF has been associated with increased pathogenesis in animal models [94] .

1.3.2 PA-X role in virulence and pathogenesis

Mutations were made at the site of the +1 ribosomal frameshift preventing frameshifting from occurring to understand the effect PA-X had on the pathogenicity of the virus [95] . These changes included synonymous mutations from UCC UUU CGU to AGC UUC AGA – denoted as the frameshift (FS)-mutant; most notably is the mutation from the UUU to UUC, which is the preferred binding of the Phe anti-codon [89,95] . This mutation eliminates both the Phe anti-codon wobble and rare codon, so that only PA is produced.

Gao *et al*. used this FS-mutation and pH1N1 (pandemic H1N1; truncated X-ORF) and avian H5N1 (full length X-ORF) in their experiment to investigate the role of PA-X in IAV pathogenesis [95]. Diminished production of PA-X by the FS-virus led to an increase in viral titre, inflammatory responses and animal mortality rates [95] Interestingly, when Gao *et al*. examined the effects of the avian influenza virus H9N2

(full length X-ORF) they observed the opposite; loss of PA-X decreased viral titre, animal mortality rate, and immune response [96] . Further studies revealed that both *in vitro* and *in vivo* interferon- β (IFN-β) release is diminished in wild type IAV infections compared to infections with PA-X-deficient IAVs [97] . Regulation of immune response is not the only role for PA-X; it also regulates viral transcription rates by regulating the amount of PA produced [95,98] . This phenomenon is clearly indicated with the increase in PA during infections with PA-X deficient IAVs [95] .

1.3.3 The current model for PA-X mechanism of action

The function of PA-X was initially proposed to be similar to that of herpes simplex virus (HSV) virion host shutoff protein (vhs) [87]. Reviewed by Rivas *et al.*, vhs is released early in infection and targets both cellular and viral mRNAs for degradation [99]. Vhs degradation of viral mRNAs signals for HSV to switch from early to late gene expression [99] . mRNA cleavage by vhs is position specific; targeting active mRNAs bound to the cap-binding complex proximal to the 5'-m^7G cap and cytokine mRNAs containing AU-rich elements (AREs) in the 3'-UTR [99] . PA-X has also been compared to the SOX host shutoff endonuclease of Kaposi's sarcoma-associated herpesvirus (KSHV), which behaves differently from vhs in that cleavage of RNA is done in a sequence-specific manner [99] . This sequence is conserved among host and viral mRNAs and is a degenerate sequence in structure and residues [99,100] . Similar to vhs, SOX cleaves its targets in the cytoplasm of the host cell [99] .

PA-X retains the PA N-terminus endonuclease domain (Figure 3). Previous work from Khaperskyy *et al*. determined that PA-X reduces the levels of cytoplasmic RNAs containing poly (A) tails [93] . Another striking observation found by Khaperskyy *et al.* was the nuclear localization of PABP1 and poly (A) RNAs in the presence of PA-X [93] . Furthermore, Hayashi *et al*. indicated the degradation of host cellular mRNAs, such as actin, was conducted in IAV infections with PA-X [97] . Similar to the strain specificity for immune response in the presence of PA-X, the strength of the host shutoff mechanism is strain specific [101] .

The shared N-terminus of PA and PA-X is insufficient for PA-X activity [102] . This led to the discovery of the importance of the X-ORF for PA-X host shutoff activity [102] . Further examination revealed that the first 15 amino acids of the X-ORF are required for full PA-X host shutoff activity (Figure 3) [102] . Within the first 15 amino acids of the X-ORF are 6 conserved basic residues between human, avian and swine strains (Figure 3); basicity is maintained however, differences occur with the presence of Arg and Lys residues between strains [102] . When mutated to acidic residues, the activity of PA-X is inhibited [102] .

PA-X retains specificity for cleavage of RNAs retaining poly (A) tails [93,103] This selective degradation was further investigated by Khaperskyy *et al*., who reported that PA-X specifically targets RNA pol II driven mRNA and non-coding RNA, but not RNA pol I or III transcripts [92] . Moreover, IAV transcripts are protected from PA-X degradation [92] . PA-X initiates degradation of RNA pol II transcripts, however, it is host 5'-3' exonuclease Xrn1 that completes transcript degradation, similar to both vhs and SOX [92] .

To determine the requirement of the canonical 3'-end poly (A) tail in PA-X transcript targeting, Khaperskyy *et al*. utilized alternative 3'-end configurations on mRNA transcripts [92] . A decreased effect of PA-X on mRNAs that did not require canonical 3'-end processing machinery was observed, leading to the hypothesis that PA-X activity is modulated by 3'-end processing events [92] . Therefore, IAV transcripts are spared from PA-X degradation due to lack of canonical mRNA 3'-end processing mechanisms.

The X-ORF guides PA-X into the nucleus of the cell [92] . More specifically, it is the basic amino acids within the N-terminus of the X-ORF that target PA-X to the nucleus; when mutated to alanine residues PA-X and the X-ORF were retained in the cytoplasm [92] . Targeting to the nucleus only requires four of the six basic amino acid residues identified by Oishi *et al.*, (Figure 3) [92,102] . Nuclear localization of PA-X is a unique feature correlating to its host shutoff function.

1.4 Proximity proteomics

1.4.1 BioID: promiscuous biotin ligase

BioID is a proteomic method introduced by Roux *et al*. in 2012 [104] . It utilizes the 34-kDa biotin ligase (BirA) from *Escherichia coli,* which covalently attaches biotin to lysine residues [105,106] . Under normal conditions, BirA selectively adds biotin to a single specific lysine with an ε-amino group in the biotin carboxyl carrier protein (BCCP) subunit of acetyl-CoA carboxylase [107,108] . Labelling occurs through a two-part reaction as follows [106,109] :

i.) ATP + Biotin ⇌ Biotin-5'-AMP + PP_i
ii.) Biotin-5'-AMP + acceptor protein apo-domain → biotinylated holo-domain + AMP

A conserved region of BirA is subjected to mutations, one of which, R118G, results in an increased rate of dissociation of the intermediate product in the biotinylation reaction (biotin-5'-AMP) [110] . The free release of biotin-5'-AMP allows for the transfer of biotin to any freely available lysine residues [111,112] .

In 2012, Roux *et al.* demonstrated this promiscuous biotin ligase, referred to as BirA*, can be used to label proximal interacting proteins within a 10 nm radius [104,113] . Roux *et al.* further used the labelling ability of BirA* through fusion to lamin-A, where under biotin treatment, they were able to determine bait localization via streptavidin conjugated Alexa-Fluor. Furthermore, mass spectrometry identified biotinylated interacting proteins of lamin-A [104] . Figure 4 illustrates the interactions captured through the BioID system.

.

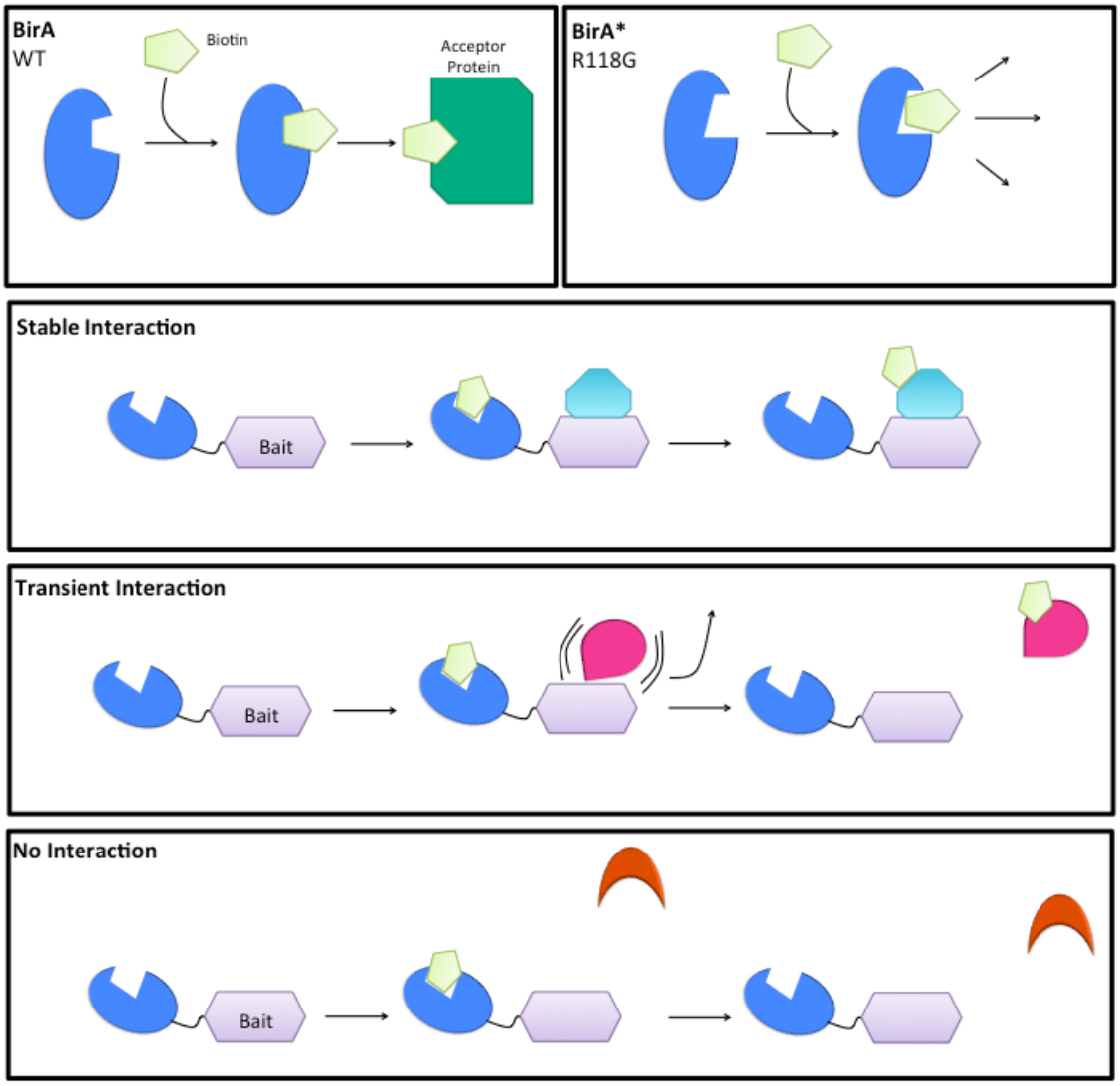


Figure 4: BioID uses BirA* R118G fused to bait protein to label stable and transient interacting proteins

The BioID method uses BirA from *E. coli* to biotinylate proteins within a 10 nm radius [113]. The top panels represent the differences between wild type BirA and R118G BirA*. The middle panels depict the types of interactions the BioID system can reveal: stable and transient. The bottom panel indicates a protein outside of the radius of BirA* labelling which is not labelled, even in the presence of biotin. Modified from [104] .

1.4.2 BioID applications

The BioID method has been used to investigate the cytoskeleton of trypanosomes [114] , define protein interactions during the duplication of centrioles [115] , and the nuclear pore complex [113] .

Anne-Claude Gingras is a leader in the BioID field. Her research has used BioID to label the interactomes of the Hippo pathway that governs organ size [116] , chromatin-associated complexes [117] , the centrosome-cillium interface and centrolsolar satellites [118] , F-box β-TrCP1 and β-TrCP2 E3 ligases [119] , Topoisomerase II-β [120] , RNA binding proteins [121] , and function of the autoimmune regulator AIRE in embryonic cells [122] .

1.5 Rationale, Hypothesis and Objective

Previous studies in the McCormick laboratory demonstrated the selectivity of PA-X in targeting RNA pol II derived transcripts, while sparing viral transcripts and products of host RNA pol I and III [95]. Moreover, we observed that PA-X, unlike other host shutoff proteins, acts in the nucleus, suggesting that host shutoff might be closely tied to the biogenesis of RNA pol II transcripts. PA-X nuclear localization and host shutoff activity was dependent on an intact C-terminal X-ORF [92] . With this knowledge it was hypothesised that PA-X required interactions with host cellular proteins to aid nuclear localization and targeting of RNA pol II transcripts. Transient and stable interactions between cellular proteins and the X-ORF were identified using the BioID system. To create a high confidence list of candidate interacting proteins, divergent 41 and 61 amino acid X-ORFs from different IAV strains were

tested. Candidate interacting proteins were eliminated if bound to a mutated X-ORF that fails to accumulate in the cell nucleus. RNA silencing was used to determine whether candidate proteins were required for PA-X host shutoff activity. This combination of approaches rendered a short list of high-value target proteins that will be the subject of further investigation.

Chapter 2: Materials and Methods

2.1 Cell culture

Cell lines were passaged at 1:10 every 2-3 days. Cells were not permitted to exceed 90% confluency. All cells were washed with phosphate-buffered saline (PBS) (Wisent Inc.), trypsinized with 0.05% Trypsin -EDTA (Invitrogen) and cultured in Dulbecco's Modified Eagle's Medium (DMEM; Invitrogen) supplemented with 10% heat inactivated fetal bovine serum (FBS) (Invitrogen), 1% penicillin – streptomycin (Invitrogen), and 2mM glutamine (Invitrogen); this media will hereafter be referred to as normal growth media. Cells were grown in an incubator at 37°C and 5% carbon dioxide. All cells were cultured in this manner unless specified otherwise. Cell lines used within this study include human embryonic kidney (HEK) 293A, 293T, and HeLa Tet-Off cells.

2.2 BioID

2.2.1 BioID: A proximity-labelling proteomics method

The BioID system, described by Roux *et al.,* utilizes the biotin ligase from *E. coli* with the R118G mutation (BirA*), which releases 5'-AMP-biotin within a 10 nm radius [104,113] . Fused to a bait protein of interest, BirA* freely biotinylates nearby proteins on available Lys residues [104] . Promiscuous biotin ligase (BirA*) (Addgene) from *E. coli* was N-terminally fused to three X-ORFs of Influenza A virus designed and constructed by Dr. Denys Khaperskyy A/Puerto Rico/8/1934 (H1N1) X-ORF (X61), A/Puerto Rico/8/1934 (H1N1) X-ORF substitutions K198A, R199A,

K202A and K203A (X61 (4A)) and A/California/7/2009 (H1N1) X-ORF (X41) (Figure 5); all fusion proteins were cloned into pCR3.1-myc [123] . In all pCR3.1 plasmids BirA* was inserted between *Kpn*I and *Eco*RI to generate the pCR3.1: BirA*-myc plasmid. Into pCR3.1BirA* -myc X61 and X61 (4A) were inserted between *Eco*RI and *Mlu*I to generate pCR3.1: BirA*-X61-myc or pCR3.1: BirA*-X61 (4A). Finally, into pCR3.1: BirA*-myc X41 was inserted between *Eco*RI and *Xho*I restriction sites to generate pCR3.1: BirA*-X41.

A detailed protocol for BioID labelling was provided by Dr. Andrew Leidal (University of California, San Francisco). HEK 293T cells were seeded into 10 cm dishes and after one day of rest were transfected with 3 μg pUC19, 3 μg BirA* construct, and 18 μl polyethylenimine (PEI) in 1 mL OPTI-MEM (Invitrogen) in serum-free and antibiotic-free growth media. Media was changed 6 hours post transfection to antibiotic-free regular growth media supplemented with 50 μM biotin (Sigma) for 18 hours. 24 hours post-transfection cells were washed with PBS, scraped off dishes in ice cold PBS, centrifuged at 250 x g for 5 minutes at 4°C, resuspended in RIPA buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 1% igepal-CA630 (Sigma), 1 mM ethylenediaminetetraacetic acid (EDTA), 1 mM ethylene-bis(oxyethylenenitriolo)tetraacetic acid (EGTA), 0.1% SDS, 1:500 Protease Inhibitor P8340 (Sigma) and 0.5% sodium deoxycholate), and agitated at 4°C for 1 hour. Lysates were mechanically disrupted using a 21-gauge needle and cleared by centrifugation at 4°C for 20 minutes at 20,000 x g, prior to western blotting or pull-downs.

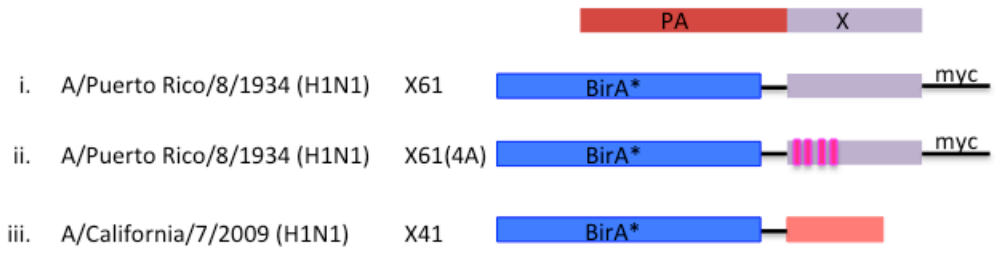


Figure 5: BioID fusion proteins used in this study
BirA* was fused to the X-ORF from A/Puerto Rico/1934 (H1N1) or A/California/7/2009 (H1N1). Three separate fusion proteins were designed: BirA*-X61, BirA*-X61(4A), and BirA*-X41. Only BirA*-X61 fusion proteins bore C-terminal myc epitope tags.

2.2.2 Western blotting

Proteins from whole cell lysates were separated on a 10% acrylamide gel (29:1 acrylamide/bis-acrylamide; BioRad). Prior to loading samples, 300 μM dithiothreitol (DTT) and 2.5 μl saturated bromophenol blue solution were added to each sample, which were then boiled at 95°C for 5 minutes, vortexed, and centrifuged at 12000 x g for 1 minute at room temperature. A voltage of 140 V was applied to acrylamide gels for approximately 90 minutes or until the bromophenol blue dye front traversed the entirety of the gel. Separated proteins were transferred from gels to PVDF (polyvinylidene difluroide) membranes using the BioRad Trans-Blot® Turbo ™ RTA Transfer kit (LF (low-fluorescence) PVDF membranes) using the standard mixed molecular weight setting on the BioRad Turbo-blot.

Membranes were blocked in 5% (w/v) non-fat skim milk in TBS (Tris-buffered saline, 50 mM Tris-HCl, and 150 mM NaCl)-T (0.05% Tween 20) for 1 hour at room temperature. For detection of biotinylated proteins, membranes were rinsed with 1X TBS-T for 5 minutes and incubated with streptavidin-HRP (1:2000; 3999S Cell Signalling) in 5% non-fat skim milk in TBS-T for 1 hour at room temperature.

The anti-myc blot was washed three times with TBS-T prior to addition of rabbit anti-myc (1:1000; 2278S Cell Signalling) in 5% BSA TBS-T on a rocker overnight at 4°C, washed three times, five minutes per wash, in TBS-T before the addition of anti-rabbit HRP (1:5000; 7074S Cell Signalling) in 5% non-fat skim milk TBST-T for 1 hour at room temperature. Prior to visualization, both streptavidin and

myc blots were washed three times in TBS-T for 5 minutes and one final wash in TBS (no Tween20). Membranes were developed with Clarity ™ Western ECL substrate Peroxide and Luminol/enhancer solutions (1:1) and imaged on the ChemiDoc (BioRad).

2.2.3 Immunofluorescence

HeLa Tet-Off cells were seeded onto cover slips at a density of 1.0 x 10^5 cells per well in regular growth media. Cells were transfected with 500ng BirA* construct (BirA*-X61, BirA*-X61 (4A) or BirA* alone) and 500 ng pUC19 per well via PEI transfection described above. New media supplemented with 50 μM biotin was added 6 hours post-transfection. Cover slips were washed with PBS before being fixed with 4% paraformaldehyde for 15 minutes on a rocker, permeabilized with 500 μl of ice-cold methanol per well for 10 minutes on a rocker, and blocked with 5% (w/v) BSA (PBS) per well for one hour at room temperature. Primary antibody, rabbit α- myc (1:400), in BSA was added to the coverslips for 24 hours. Cover slips were washed with PBS before secondary staining with Alexa-Fluor 555 Donkey α-rabbitt (1:1000; Invitrogen), Streptavidin-Alexa-Fluor 488 conjugate (1:1000; Invitrogen) and Hoescht (1:7000) for 1 hour in the dark prior to being mounted onto glass microscope slides using ProLong Gold Antifade reagent (Invitrogen). Images were captured using AXIO Imager Z.2 (40x magnification), in the Cellular and Molecular Digital Imaging Facility (https://medicine.dal.ca/research-dal-med/facilities/cellular-molecular-digital-imaging.html).

2.2.4 Neutravidin pull-down

60 μl high capacity Neutravidin® Agarose Beads (Thermo Scientific) 50% slurry and 50% beads were used per sample, washed 3 times in RIPA buffer (no protease inhibitor cocktail or deoxycholate) on an end-over-end rotation for 10 minutes at 4°C before centrifuged at 400 x g for 1 minute.

RNaseA (QIAGEN) treatment was done as follows: 1 μl 500x RNaseA was added to each sample, 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 an end-over-end 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 (no PIC, deoxycholate or igepal-CA630) three times, followed by three washes with TAP buffer (50 mM HEPES-KOH pH 8.0, 100 mM KCl and 10% glycerol).

2.2.6 Mass spectrometry

Beads were resuspended in 50 mM triethylammonium bicarbonate (TEAB) buffer (Sigma). 24 mM DTT and 32 mM $IAcNH_2$ were added sequentially, per tube then incubated for 30 minutes at 37°C. Beads were washed with 50 mM TEAB and centrifuged for 1 min at 400 x g before resuspending in 50mM 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 the addition of 3 μl

Formic acid until a pH lower than 3 was achieved. A hole was punctured in the bottom of the 1.5 mL Eppendorf tube using a 30-gauge needle and placed in a 2 mL Eppendorf tube spun at 2000 rpm for 1 minute at room temperature to collect trypsinized peptides. Beads were washed with 50mM TEAB prior to de-salting.

Samples were desalted with Oasis/SepPak Desalting columns and conditioned with successive washes of methanol, 50% acetonitrile (ACN) with 0.1% TFA, and finally 0.1% TFA. Once the samples were loaded four washes with 0.1%TFA occurred prior to elution in 1 mL 50% ACN/0.1% TFA and 500 μl 70% ACN/0.1% TFA. Samples were dried in the Thermo SPDIIIV speed vacuum and frozen at -20°C.

Quantitative mass spectrometry analysis via reductive dimethylation enabled measurements of relative abundance of biotinylated proteins in each experimental condition. In reductive dimethylation, formaldehyde molecules with different combinations of stable hydrogen and carbon isotopes are conjugated to peptide samples with sodium cyanoborohydride ($NaCNBH_3$), which labels terminal amines [124]. The isotopes used in this thesis were: formaldehyde (light), D2-formaldehyde (medium), and D2-C13-formaldehyde (heavy).

Dried samples were resuspended by sonication for 15 minutes in 50 mM TEAB. In separate tubes, BirA*, BirA*-XORF (X61 or X41) and BirA*-X61(4A) were labelled light, medium and heavy respectively. To the resuspended samples 8 μl formaldehyde (Sigma) was added to the light, 15 μl D2-formaldehyde (Cambridge Isotope Laboratories Inc.) was added to the medium and 15 μl D2-C13 formaldehyde (Aldrich) was added to the heavy samples. Reactions were mixed and

incubated for 5 minutes at room temperature. Once incubation was completed, 0.51 M $NaCNBH_3$ (Fluka) was added to the light and medium samples, while 0.51 M $NaCNBD_3$ (sodium cyanoborodeuteride; Aldrich) was added to the heavy reaction. 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 as described above.

The combined sample was desalted and dried as previously described. Samples were prepared for LC-MS/MS by being resuspended in 3% ACN/0.1% formic acid and sonication for 15 min. Dr. Alejandro Cohen of the Proteomics CORE Facility (https://medicine.dal.ca/research-dal-med/facilities/proteomics.html) facilitated the mass spectrometry and analysis of peptide identification.

2.2.7 Analysing mass spectrometry data

Dr. Alejandro Cohen of the Proteomics CORE Facility completed peptide identification from the mass spectrometry experiment using Proteome Discoverer™ (Thermo Fisher Scientific). Critical analysis of identified proteins was as follows: any protein with less than 2 unique peptides identified in the mass spectrometry data set was deemed insignificant and not included in analysis. Those with a greater than 2-fold increase in 2 out of 3 samples were chosen as targets of interest or with 1.5 increase X61 or X41/BirA in all three rounds of mass spectrometry were selected for further analysis.

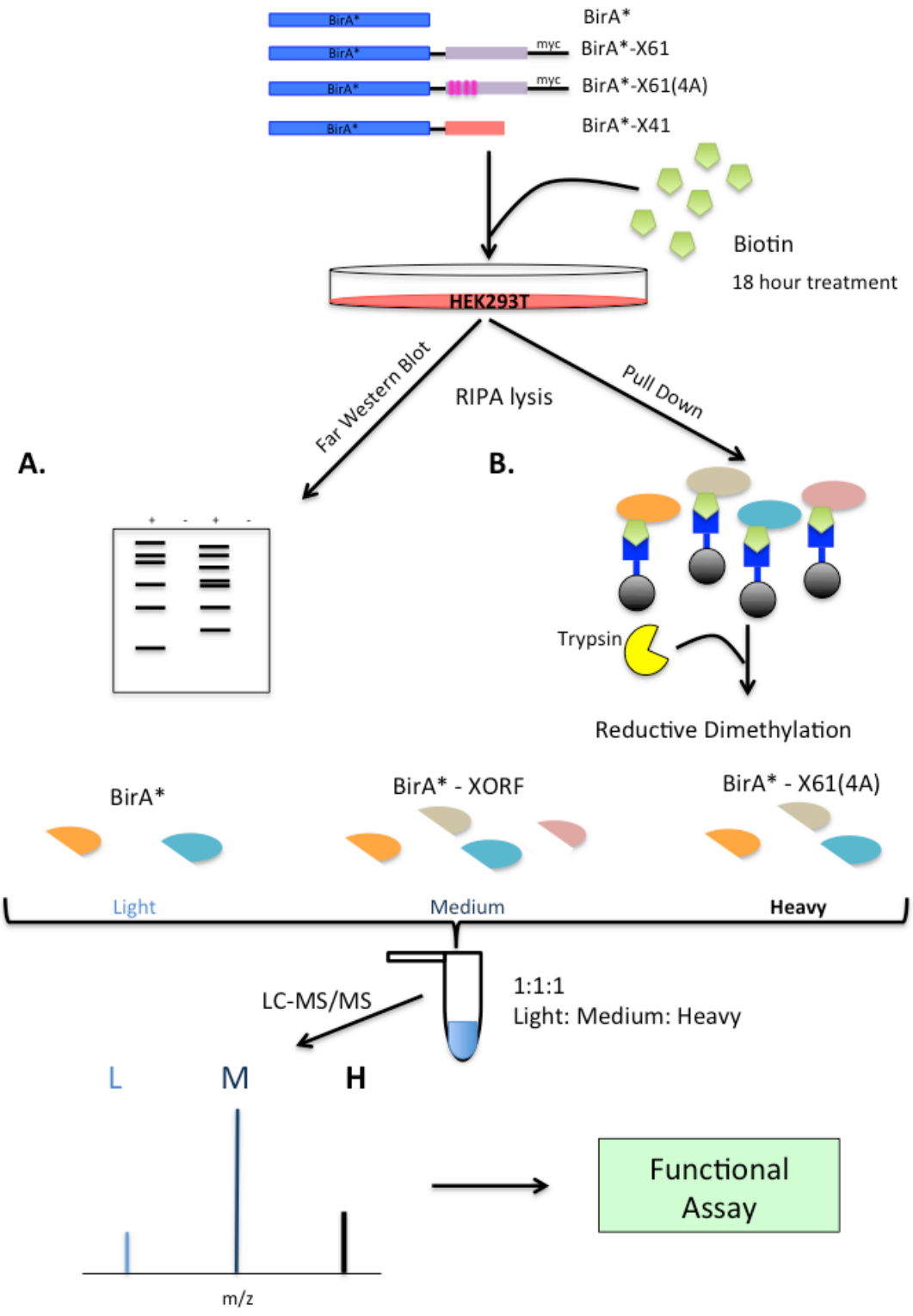


Figure 6: BioID workflow

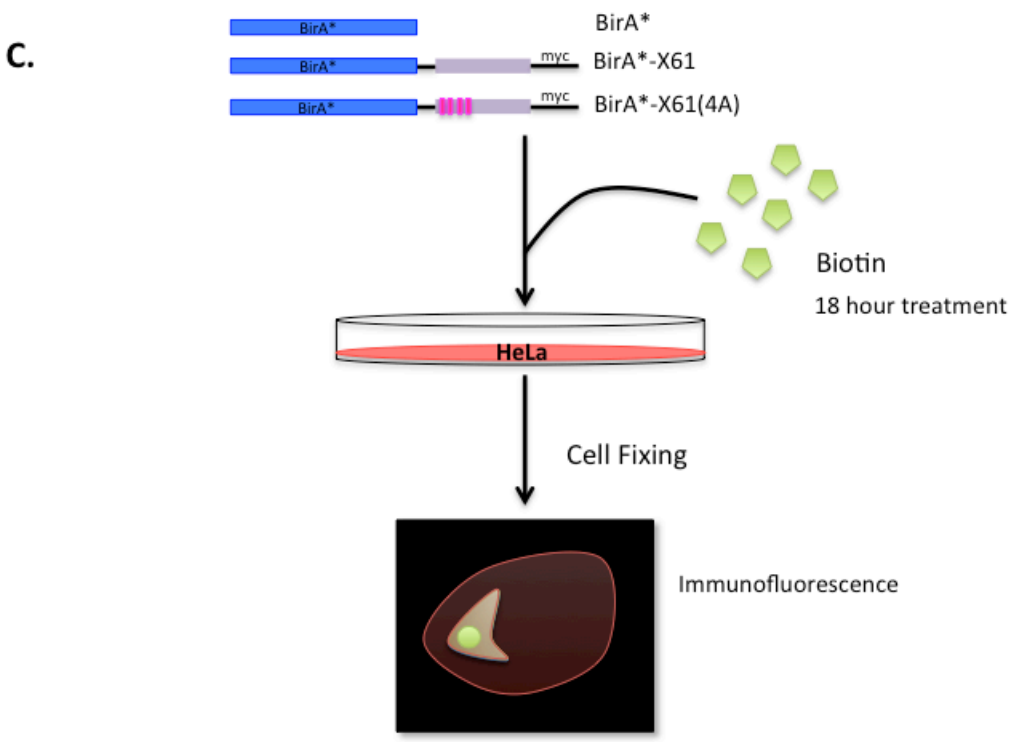


Figure 6: BioID workflow (Continued)
Plasmids containing BirA*, BirA*-X61, BirA*-X61(4A), and BirA*-X41 were transfected into HEK293T cells via PEI transfection for 6 hours prior to addition of 50 μM biotin. 18 hours later, cell lysates were harvested in RIPA buffer for (A) far-western blot analysis and (B) neutravidin bead pull-down for proteomic analysis. Three separate mass spectrometry trials were conducted with the following samples: (1 & 2) BirA*-X61, BirA*-X61 (4A), and BirA* or (3) BirA*-X41, BirA*-X61 (4A), and BirA*. Relative abundances of biotinylated proteins were identified using reductive dimethylation labelling. For each round of mass spectrometry, labelling occurred as flows: all BirA* samples were labeled light (formaldehyde), BirA*-X-ORF were labeled medium (D2-formaldyhe), and BirA*-X61 (4A) were labeled heavy (D2-C^{13}-formaldyde). Samples were combined into one tube at a 1:1:1 ratio and analyzed through LC-MS/MS. Candidate X-ORF interacting proteins were identified by stringent criteria and subsequently were evaluated in functional assays. (C.) Sub-cellular localization of biotinylated proteins was determined via immunofluorescence. 24 hours post transfection cells were fixed with PFA and stained with rabbit α-myc/donkey-α-rabbit-555, Strep-Alex Fluor 488, and Hoechst, before imaging on AXIO Imager Z.2 (40x magnification).

2.3 Dual- Glo luciferase assay

2.3.1 shRNA cloning

Using the RNAi consortium web portal (http://portals.broadinstitute.org/gpp/public/gene/search) 3 sequences were selected for each gene of interest that were compatible for cloning into pLKO.1; with the exception of the non-targeting scrambled control shRNA obtained from Addgene; hereafter, the non-targeting scrambled control will be referred to only as the scramble control. Forward and reverse oligonucleotides were ordered from Invitrogen (Table 1). 6 μg of pLKO.1 was digested with 0.5 μl *Age*I (New England Biolabs) and *Eco*RI-HF (New England Biolabs) in NEB CutSmart buffer at 37 °C for 2 hours before being purified via QIAQuick Gel Extraction Kit reagents and protocol. Oligonucleotides were annealed under the following cycle: 95°C for 1 min, 94°C for 30 sec, repeated 64 times -1°C per cycle, in a reaction consisting of 2.5X T4 Ligase Buffer (NEB), 25 μM forward oligomer, 25 μM reverse oligomer and nuclease-free water.

Ligation reactions were made with 5 μl of annealed oligonucleotide reactions diluted 1:50 or 1:100 to approximately 300 ng of digested pLKO.1 and 1 μl T4 DNA ligase (NEB) prior to incubation at 16 °C overnight. 5 μl of ligation reactions were transformed in 50 μl of Stbl3 chemically competent *E. coli* cells. Ligation reaction sat on ice for 30 minutes prior to being heat shocked at 42°C for 30 seconds, rested with 250 μl LB Broth, and shaken for 1 hour at 37°C. Transformed *E. coli* were plated onto LB Agar plates with 0.1 mg/mL (Invitrogen) and grown at 37°C

overnight. A single colony was selected from each plate and grown in 4 mL of LB Broth with 0.1 mg/mL carbenicillin at 37°C overnight. DNA was isolated via the QIAprep Spin Miniprep protocol and kit and sent for sequencing at Genewiz using the forward pLKO.1 sequencing primer. (5' – GACTATCATATGCTTACCGT-3' (Weinberg Lab; Addgene)).

Table 1: shRNA oligonucleotide sequences from RNAi consortium

Name	Forward (5' – 3')	Reverse (5' - 3')
NCL-1	CCGGGCACTTGGAGTGGTGAATCAACTCGAGTTGATTCACCACTCCAAGTGCTTTTTG	AATTCAAAAAGCACTTGGAGTGGTGAATCAACTCGAGTTGATTCACCACTCCAAGTGC
NCL-2	CCGGGCGATCTATTTCCCTGTACTACTCGAGTAGTACAGGGAAATAGATCGCTTTTTG	AATTCAAAAAGCGATCTATTTCCCTGTACTACTCGAGTAGTACAGGGAAATAGATCGC
NCL-3	CCGGCGGTGAAATTGATGGAAATAACTCGAGTTATTTCCATCAATTTCACCGTTTTTG	AATTCAAAAACGGTGAAATTGATGGAAATAACTCGAGTTATTTCCATCAATTTCACCG
NPM1-1	CCGGGCAAAGGATGAGTTGCACATTCTCGAGAATGTGCAACTCATCCTTTGCTTTTTG	AATTCAAAAAGCAAAGGATGAGTTGCACATTCTCGAGAATGTGCAACTCATCCTTTGC
NPM1-2	CCGGGCCGACAAAGATTATCACTTTCTCGAGAAAGTGATAATCTTTGTCGGCTTTTTG	AATTCAAAAAGCCGACAAAGATTATCACTTTCTCGAGAAAGTGATAATCTTTGTCGGC
NPM1-3	CCGGGCGCCAGTGAAGAAATCTATACTCGAGTATAGATTTCTTCACTGGCGCTTTTTG	AATTCAAAAAGCGCCAGTGAAGAAATCTATACTCGAGTATAGATTTCTTCACTGGCGC
TRIM28 -1	CCGGGAGAATTATTTCATGCGTGATCTCGAGATCACGCATGAAATAATTCTCTTTTTG	AATTCAAAAAGAGAATTATTTCATGCGTGATCTCGAGATCACGCATGAAATAATTCTC
TRIM28 -2	CCGGGACCACCAGTACCAGTTCTTACTCGAGTAAGAACTGGTACTGGTGGTCTTTTTG	AATTCAAAAAGACCACCAGTACCAGTTCTTACTCGAGTAAGAACTGGTACTGGTGGTC
TRIM28 -3	CCGGGAGGACTACAACCTTATTGTTCTCGAGAACAATAAGGTTGTAGTCCTCTTTTTG	AATTCAAAAAGAGGACTACAACCTTATTGTTCTCGAGAACAATAAGGTTGTAGTCCTC
XPO1-1	CCGGCCATTGTAAAGCGACTTCAAACTCGAGTTTGAAGTCGCTTTACAATGGTTTTTG	AATTCAAAAACCATTGTAAAGCGACTTCAAACTCGAGTTTGAAGTCGCTTTACAATGG
XPO1-2	CCGGCCTGCTTTCAAGGAACATTTACTCGAGTAAATGTTCCTTGAAAGCAGGTTTTTG	AATTCAAAAACCTGCTTTCAAGGAACATTTACTCGAGTAAATGTTCCTTGAAAGCAGG
XPO1-3	CCGGCAGCTATATTTGCCCATGTTACTCGAGTAACATGGGCAAATATAGCTGTTTTTG	AATTCAAAAACAGCTATATTTGCCCATGTTACTCGAGTAACATGGGCAAATATAGCTG
HNRNPK-1	CCGGTGCCAGTGTTTCAGTCCCAGACTCGAGTCTGGGACTGAAACACTGGCATTTTTG	AATTCAAAAATGCCAGTGTTTCAGTCCCAGACTCGAGTCTGGGACTGAAACACTGGCA
HNRNPK-2	CCGGCTCTGTTGTATGTTGGATTATCTCGAGATAATCCAACATACAACAGAGTTTTTG	AATTCAAAAACTCTGTTGTATGTTGGATTATCTCGAGATAATCCAACATACAACAGAG
HNRNPK -3	CCGGCCAAAGATTTGGCTGGATCTACTCGAGTAGATCCAGCCAAATCTTTGGTTTTTG	AATTCAAAAACCAAAGATTTGGCTGGATCTACTCGAGTAGATCCAGCCAAATCTTTGG
RPL26-1	CCGGCCCACATTCGAAGGAAGATTACTCGAGTAATCTTCCTTCGAATGTGGGTTTTTG	AATTCAAAAACCCACATTCGAAGGAAGATTACTCGAGTAATCTTCCTTCGAATGTGGG
RPL26-2	CCGGCCAGGTTTACAGGAAGAAATACTCGAGTATTTCTTCCTGTAAACCTGGTTTTTG	AATTCAAAAACCAGGTTTACAGGAAGAAATACTCGAGTATTTCTTCCTGTAAACCTGG
RPL26-3	CCGGAGTCCAGGTTTACAGGAAGAACTCGAGTTCTTCCTGTAAACCTGGACTTTTTTG	AATTCAAAAAAGTCCAGGTTTACAGGAAGAACTCGAGTTCTTCCTGTAAACCTGGACT
CSE1L-1	CCGGCCGTCATGAATTTAAGTCAAACTCGAGTTTGACTTAAATTCATGACGGTTTTTG	AATTCAAAAACCGTCATGAATTTAAGTCAAACTCGAGTTTGACTTAAATTCATGACGG
CSE1L-2	CCGGCCTGGGTTACTAGGTGTCTTTCTCGAGAAAGACACCTAGTAACCCAGGTTTTTG	AATTCAAAAACCTGGGTTACTAGGTGTCTTTCTCGAGAAAGACACCTAGTAACCCAGG
CSE1L-3	CCGGCGCTGACAAGTATCTGTGAAACTCGAGTTTCACAGATACTTGTCAGCGTTTTTG	AATTCAAAAACGCTGACAAGTATCTGTGAAACTCGAGTTTCACAGATACTTGTCAGCG
CPSF6-1	CCGGGGTGATTATGGGAGTGCTATTCTCGAGAATAGCACTCCCATAATCACCTTTTTG	AATTCAAAAAGGTGATTATGGGAGTGCTATTCTCGAGAATAGCACTCCCATAATCACC
CPSF6-2	CCGGGCAACTTTATTACTGGTTATACTCGAGTATAACCAGTAATAAAGTTGCTTTTTG	AATTCAAAAAGCAACTTTATTACTGGTTATACTCGAGTATAACCAGTAATAAAGTTGC
CPSF6-3	CCGGGCAATCTCAAGCAGTGCTATTCTCGAGAATAGCACTGCTTGAGATTGCTTTTTG	AATTCAAAAAGCAATCTCAAGCAGTGCTATTCTCGAGAATAGCACTGCTTGAGATTGC
IPO7-1	CCGGGCACTGACTCACGGTCTTAATCTCGAGATTAAGACCGTGAGTCAGTGCTTTTTG	AATTCAAAAAGCACTGACTCACGGTCTTAATCTCGAGATTAAGACCGTGAGTCAGTGC
IPO7-2	CCGGGCTAACAAGAAGATGTCTGATCTCGAGATCAGACATCTTCTTGTTAGCTTTTTG	AATTCAAAAAGCTAACAAGAAGATGTCTGATCTCGAGATCAGACATCTTCTTGTTAGC
IPO7-3	CCGGCCCTGTTGATGAGTATCAGATCTCGAGATCTGATACTCATCAACAGGGTTTTTG	AATTCAAAAACCCTGTTGATGAGTATCAGATCTCGAGATCTGATACTCATCAACAGGG
BTF3-1	CCGGCCCAGCATCTTAAACCAGCTTCTCGAGAAGCTGGTTTAAGATGCTGGGTTTTTG	AATTCAAAAACCCAGCATCTTAAACCAGCTTCTCGAGAAGCTGGTTTAAGATGCTGGG
BTF3-2	CCGGTGAGTCAACTTCTGAAGATAACTCGAGTTATCTTCAGAAGTTGACTCATTTTTG	AATTCAAAAATGAGTCAACTTCTGAAGATAACTCGAGTTATCTTCAGAAGTTGACTCA
BTF3-3	CCGGGCAGCGAACACTTTCACCATTCTCGAGAATGGTGAAAGTGTTCGCTGCTTTTTG	AATTCAAAAAGCAGCGAACACTTTCACCATTCTCGAGAATGGTGAAAGTGTTCGCTGC
RPL24-1	CCGGCCTCTACAGAAGGAAGCACAACTCGAGTTGTGCTTCCTTCTGTAGAGGTTTTTG	AATTCAAAAACCTCTACAGAAGGAAGCACAACTCGAGTTGTGCTTCCTTCTGTAGAGG
RPL24-2	CCGGCCTGAAGTTAGAAAGGCTCAACTCGAGTTGAGCCTTTCTAACTTCAGGTTTTTG	AATTCAAAAACCTGAAGTTAGAAAGGCTCAACTCGAGTTGAGCCTTTCTAACTTCAGG

Name	Forward (5' – 3')	Reverse (5' - 3')
RPL24-3	CCGGGCTTTCCTTTCCAAGAGGAATCTCGAGATTCCTCTTGGAAAGGAAAGCTTTTTG	AATTCAAAAAGCTTTCCTTTCCAAGAGGAATCTCGAGATTCCTCTTGGAAAGGAAAGC
PSME3-1	CCGGGAGAAATGTAACACGGTCAAACTCGAGTTTGACCGTGTTACATTTCTCTTTTTG	AATTCAAAAAGAGAAATGTAACACGGTCAAACTCGAGTTTGACCGTGTTACATTTCTC
PSME3-2	CCGGAGAGCCAAATTGGTTTCTAAACTCGAGTTTAGAAACCAATTTGGCTCTTTTTTG	AATTCAAAAAAGAGCCAAATTGGTTTCTAAACTCGAGTTTAGAAACCAATTTGGCTCT
PSME3-3	CCGGGCAGAAGACTTGGTGGCAAATCTCGAGATTTGCCACCAAGTCTTCTGCTTTTTG	AATTCAAAAAGCAGAAGACTTGGTGGCAAATCTCGAGATTTGCCACCAAGTCTTCTGC
NACA-1	CCGGCCTGCTTCAGATACTTACATACTCGAGTATGTAAGTATCTGAAGCAGGTTTTTG	CCGGCCTGCTTCAGATACTTACATACTCGAGTATGTAAGTATCTGAAGCAGGTTTTTG
NACA-2	CCGGGAGCCCTGCTTCAGATACTTACTCGAGTAAGTATCTGAAGCAGGGCTCTTTTTG	AATTCAAAAAGAGCCCTGCTTCAGATACTTACTCGAGTAAGTATCTGAAGCAGGGCTC
NACA-3	CCGGGCCCTGAAGAACAACAGTAATCTCGAGATTACTGTTGTTCTTCAGGGCTTTTTG	AATTCAAAAAGCCCTGAAGAACAACAGTAATCTCGAGATTACTGTTGTTCTTCAGGGC
PRPF4-1	CCGGCGTTGTATCATGTTCTTAGAACTCGAGTTCTAAGAACATGATACAACGTTTTTG	AATTCAAAAACGTTGTATCATGTTCTTAGAACTCGAGTTCTAAGAACATGATACAACG
PRPF4-2	CCGGGCAAAGTGATGGGCCTAGATACTCGAGTATCTAGGCCCATCACTTTGCTTTTTG	AATTCAAAAAGCAAAGTGATGGGCCTAGATACTCGAGTATCTAGGCCCATCACTTTGC
PRPF4-3	CCGGAGGACGTTGTATCATGTTCTTCTCGAGAAGAACATGATACAACGTCCTTTTTTG	AATTCAAAAAAGGACGTTGTATCATGTTCTTCTCGAGAAGAACATGATACAACGTCCT
ATP5C1-1	CCGGCGTCAGTCATTGCCCTTGAATCTCGAGATTCAAGGGCAATGACTGACGTTTTTG	AATTCAAAAACGTCAGTCATTGCCCTTGAATCTCGAGATTCAAGGGCAATGACTGACG
ATP5C1-2	CCGGGATCTTTAGCTCTGTATGAAACTCGAGTTTCATACAGAGCTAAAGATCTTTTTG	AATTCAAAAAGATCTTTAGCTCTGTATGAAACTCGAGTTTCATACAGAGCTAAAGATC
ATP5C1-3	CCGGGCTGACAGCATGAGTATCTATCTCGAGATAGATACTCATGCTGTCAGCTTTTTG	AATTCAAAAAGCTGACAGCATGAGTATCTATCTCGAGATAGATACTCATGCTGTCAGC
CAND1-1	CCGGTCCATAATCCAGAGGTTGTAACTCGAGTTACAACCTCTGGATTATGGATTTTTG	AATTCAAAAATCCATAATCCAGAGGTTGTAACTCGAGTTACAACCTCTGGATTATGGA
CAND1-2	CCGGCGGTTGATGATGGTCTGGATACTCGAGTATCCAGACCATCATCAACCGTTTTTG	AATTCAAAAACGGTTGATGATGGTCTGGATACTCGAGTATCCAGACCATCATCAACCG
CAND1-3	CCGGGTACTCTTTCTGCCCTTGATACTCGAGTATCAAGGGCAGAAAGAGTACTTTTTG	AATTCAAAAAGTACTCTTTCTGCCCTTGATACTCGAGTATCAAGGGCAGAAAGAGTAC
CCAR2-1	CCGGCGGGTCTTCACTGGTATTGTTCTCGAGAACAATACCAGTGAAGACCCGTTTTTG	AATTCAAAAACGGGTCTTCACTGGTATTGTTCTCGAGAACAATACCAGTGAAGACCCG
CCAR2-2	CCGGCCTCTGAAGCAGATTAAGTTTCTCGAGAAACTTAATCTGCTTCAGAGGTTTTTG	AATTCAAAAACCTCTGAAGCAGATTAAGTTTCTCGAGAAACTTAATCTGCTTCAGAGG
CCAR2-3	CCGGGCATTGATTTGAGCGGCTGTACTCGAGTACAGCCGCTCAAATCAATGCTTTTTG	AATTCAAAAAGCATTGATTTGAGCGGCTGTACTCGAGTACAGCCGCTCAAATCAATGC
EIF3I-1	CCGGGCGGTCCATTACGCAGATTAACTCGAGTTAATCTGCGTAATGGACCGCTTTTTG	AATTCAAAAAGCGGTCCATTACGCAGATTAACTCGAGTTAATCTGCGTAATGGACCGC
EIF3I-2	CCGGCGAAGATGGTTACGTCCGTATCTCGAGATACGGACGTAACCATCTTCGTTTTTG	AATTCAAAAACGAAGATGGTTACGTCCGTATCTCGAGATACGGACGTAACCATCTTCG
EIF3I-3	CCGGCCATTACTTCGACCCACAGTACTCGAGTACTGTGGGTCGAAGTAATGGTTTTTG	AATTCAAAAACCATTACTTCGACCCACAGTACTCGAGTACTGTGGGTCGAAGTAATGG
GTF2I-1	CCGGCCAGCAGAAGATGATGATTATCTCGAGATAATCATCATCTTCTGCTGGTTTTTG	AATTCAAAAACCAGCAGAAGATGATGATTATCTCGAGATAATCATCATCTTCTGCTGG
GTF2I- 2	CCGGCCACAGAAGATTCTGGCATTTCTCGAGAAATGCCAGAATCTTCTGTGGTTTTTG	AATTCAAAAACCACAGAAGATTCTGGCATTTCTCGAGAAATGCCAGAATCTTCTGTGG
GTF2I-3	CCGGCCAGCAGAAGATTCTACTCAACTCGAGTTGAGTAGAATCTTCTGCTGGTTTTTG	AATTCAAAAACCAGCAGAAGATTCTACTCAACTCGAGTTGAGTAGAATCTTCTGCTGG
HUWE1-1	CCGGCCGCACTGTGTTAAACCAGATCTCGAGATCTGGTTTAACACAGTGCGGTTTTTG	AATTCAAAAACCGCACTGTGTTAAACCAGATCTCGAGATCTGGTTTAACACAGTGCGG
HUWE1-2	CCGGCCACACTTTCACAGATACTATCTCGAGATAGTATCTGTGAAAGTGTGGTTTTTG	AATTCAAAAACCACACTTTCACAGATACTATCTCGAGATAGTATCTGTGAAAGTGTGG
HUWE1-3	CCGGCCACAAATATGCCATGATGTTCTCGAGAACATCATGGCATATTTGTGGTTTTTG	AATTCAAAAACCACAAATATGCCATGATGTTCTCGAGAACATCATGGCATATTTGTGG
LRPPRC-1	CCGGGCCTCATCATAACGCAAGTTACTCGAGTAACTTGCGTTATGATGAGGCTTTTTG	AATTCAAAAAGCCTCATCATAACGCAAGTTACTCGAGTAACTTGCGTTATGATGAGGC
LRPPRC-2	CCGGCCGTGAACTTTCTAACGCATACTCGAGTATGCGTTAGAAAGTTCACGGTTTTTG	AATTCAAAAACCGTGAACTTTCTAACGCATACTCGAGTATGCGTTAGAAAGTTCACGG
LRPPRC-3	CCGGCGCAGCTTTAAGAGGTGAAATCTCGAGATTTCACCTCTTAAAGCTGCGTTTTTG	AATTCAAAAACGCAGCTTTAAGAGGTGAAATCTCGAGATTTCACCTCTTAAAGCTGCG
MCM5-1	CCGGCACGGGCTTCACCTTCAAATACTCGAGTATTTGAAGGTGAAGCCCGTGTTTTTG	AATTCAAAAACACGGGCTTCACCTTCAAATACTCGAGTATTTGAAGGTGAAGCCCGTG

Name	Forward (5' – 3')	Reverse (5' - 3')
MCM5-2	CCGGGAAACTGAAGAACCGCTACATCTCGAGATGTAG CGGTTCTTCAGTTTCTTTTTG	AATTCAAAAAGAAACTGAAGAACCGCTACATCTCGAGATG TAGCGGTTCTTCAGTTTC
MCM5-3	CCGGGTACTGGATTGAGGTGGAGATCTCGAGATCTCC ACCTCAATCCAGTACTTTTTG	AATTCAAAAAGTACTGGATTGAGGTGGAGATCTCGAGATC TCCACCTCAATCCAGTAC
NASP-1	CCGGGCTAAGAAACTATTGGGTTTACTCGAGTAAACC CAATAGTTTCTTAGCTTTTTG	AATTCAAAAAGCTAAGAAACTATTGGGTTTACTCGAGTAA ACCCAATAGTTTCTTAGC
NASP-2	CCGGCCCGAAATTAGAGAGAAGATACTCGAGTATCTT CTCTCTAATTTCGGGTTTTTG	AATTCAAAAACCCGAAATTAGAGAGAAGATACTCGAGTAT CTTCTCTCTAATTTCGGG
NASP-3	CCGGCGGAAAGATGATGCAAAGAAACTCGAGTTTCTT TGCATCATCTTTCCGTTTTTG	AATTCAAAAACGGAAAGATGATGCAAAGAAACTCGAGTT TCTTTGCATCATCTTTCCG
NOP2-1	CCGGGACGATGCTGATACGGTAGATCTCGAGATCTAC CGTATCAGCATCGTCTTTTTG	AATTCAAAAAGACGATGCTGATACGGTAGATCTCGAGATC TACCGTATCAGCATCGTC
NOP2-2	CCGGGCTAAAGCAACGATCACCTAACTCGAGTTAGGT GATCGTTGCTTTAGCTTTTTG	AATTCAAAAAGCTAAAGCAACGATCACCTAACTCGAGTTA GGTGATCGTTGCTTTAGC
NOP2-3	CCGGGCCGTGAAGACTAACAAGGATCTCGAGATCCTT GTTAGTCTTCACGGCTTTTTG	AATTCAAAAAGCCGTGAAGACTAACAAGGATCTCGAGATC CTTGTTAGTCTTCACGGC
NUDT21-1	CCGGCAGCGCATGAGGGAAGAATTTCTCGAGAAATTC TTCCCTCATGCGCTGTTTTTG	AATTCAAAAACAGCGCATGAGGGAAGAATTTCTCGAGAAA TTCTTCCCTCATGCGCTG
NUDT21-2	CCGGGAACCTCCTCAGTATCCATATCTCGAGATATGG ATACTGAGGAGGTTCTTTTTG	AATTCAAAAAGAACCTCCTCAGTATCCATATCTCGAGATA TGGATACTGAGGAGGTTC
NUDT21-3	CCGGCATTGACGATTGCATTGGTAACTCGAGTTACCA ATGCAATCGTCAATGTTTTTG	AATTCAAAAACATTGACGATTGCATTGGTAACTCGAGTTA CCAATGCAATCGTCAATG
PSMC5-1	CCGGGCACAGAGGAACGAACTAAATCTCGAGATTTAG TTCGTTCCTCTGTGCTTTTTG	AATTCAAAAAGCACAGAGGAACGAACTAAATCTCGAGATT TAGTTCGTTCCTCTGTGC
PSMC5-2	CCGGCAAGGTTATCATGGCTACTAACTCGAGTTAGTA GCCATGATAACCTTGTTTTTG	AATTCAAAAACAAGGTTATCATGGCTACTAACTCGAGTTA GTAGCCATGATAACCTTG
PSMC5-3	CCGGGCTCTGAACTGGTACAGAAATCTCGAGATTTCT GTACCAGTTCAGAGCTTTTTG	AATTCAAAAAGCTCTGAACTGGTACAGAAATCTCGAGATT TCTGTACCAGTTCAGAGC
PSMD3-1	CCGGGCCGCAAAGTGTTACTATTATCTCGAGATAATA GTAACACTTTGCGGCTTTTTG	AATTCAAAAAGCCGCAAAGTGTTACTATTATCTCGAGATA ATAGTAACACTTTGCGGC
PSMD3-2	CCGGGCTGCGGAATTACCTACACTACTCGAGTAGTGT AGGTAATTCCGCAGCTTTTTG	AATTCAAAAAGCTGCGGAATTACCTACACTACTCGAGTAG TGTAGGTAATTCCGCAGC
PSMD3-3	CCGGCCATGAGGTTTCCTCCCAAATCTCGAGATTTGG GAGGAAACCTCATGGTTTTTG	AATTCAAAAACCATGAGGTTTCCTCCCAAATCTCGAGATT TGGGAGGAAACCTCATGG
RBM39-1	CCGGCCAAGGGATTTGGAAGAGTTTCTCGAGAAACTC TTCCAAATCCCTTGGTTTTTG	AATTCAAAAACCAAGGGATTTGGAAGAGTTTCTCGAGAAA CTCTTCCAAATCCCTTGG
RMB39 -2	CCGGCCAAGGGATATGGATTTATTACTCGAGTAATAA ATCCATATCCCTTGGTTTTTG	AATTCAAAAACCAAGGGATATGGATTTATTACTCGAGTAA TAAATCCATATCCCTTGG
RBM39-3	CCGGGCTGGACCTATGAGGCTTTATCTCGAGATAAAG CCTCATAGGTCCAGCTTTTTG	AATTCAAAAAGCTGGACCTATGAGGCTTTATCTCGAGATA AAGCCTCATAGGTCCAGC
TCOF1-1	CCGGGACGCTTCATATAGATGTGTACTCGAGTACACA TCTATATGAAGCGTCTTTTTG	AATTCAAAAAGACGCTTCATATAGATGTGTACTCGAGTAC ACATCTATATGAAGCGTC
TCOF1-2	CCGGCCCAAGACTAGCATCTACCAACTCGAGTTGGTA GATGCTAGTCTTGGGTTTTTG	AATTCAAAAACCCAAGACTAGCATCTACCAACTCGAGTTG GTAGATGCTAGTCTTGGG
TCOF1-3	CCGGCGTAACCCTTCTGGACATCTACTCGAGTAGATG TCCAGAAGGGTTACGTTTTTG	AATTCAAAAACGTAACCCTTCTGGACATCTACTCGAGTAG ATGTCCAGAAGGGTTACG
PUF60-1	CCGGCGTCCCAAGATGCTGTGTCTTCTCGAGAAGACA CAGCATCTTGGGACGTTTTTG	AATTCAAAAACGTCCCAAGATGCTGTGTCTTCTCGAGAAG ACACAGCATCTTGGGACG
PUF60-2	CCGGGTCATCATCTACCAAGAGAAACTCGAGTTTCTC TTGGTAGATGATGACTTTTTG	AATTCAAAAAGTCATCATCTACCAAGAGAAACTCGAGTTT CTCTTGGTAGATGATGAC
PUF60-3	CCGGTGCAGAAATCATTGTCAAGATCTCGAGATCTTG ACAATGATTTCTGCATTTTTG	AATTCAAAAATGCAGAAATCATTGTCAAGATCTCGAGATC TTGACAATGATTTCTGCA

2.3.2 Lentivirus and stable cell line production

HEK 293T cells were seeded into 6-well plates at a density of 5.6 x10^5 cells per well and cultured in serum- and antibiotic-free DMEM. These cells were co-transfected with 550 ng shRNA sequence cloned in pLKO.1 transfer plasmid, 333 ng of pSPAX2 (Addgene) packaging plasmid and 166 ng of pMD2.G (Addgene) envelope plasmid per well. Transfection was carried out using 3 μl PEI in 400 μl OPTI-MEM per well. At 6 hours post-transfection, media was changed to DMEM/10%FBS/0.1%. Cell supernatants containing lentiviruses were collected at 48 hours post-transfection, centrifuged at 2500x g for 5 minutes at room temperature, and aliquoted into 500 μl aliquots for storage at -80°C.

For the luciferase-based host shutoff assay, inconsistent measurements from the cells transduced with the scramble control shRNA lentivirus led to optimizing various aspects of the assay, including the lentiviral transduction protocol itself. All protocols are described in detail below, for a pictorial depiction of the timeline of experimental design refer to Figure 16A (Round 1), C (Round 2), and E (Round 3).

Round one, HEK 293A cells were seeded into 12-well dish at a confluency of 1.0 x10^5 cells per well, cultured as described in Section 2.1. Prior to transduction, polybrene was diluted 1:3000 in normal growth media. Cells were transduced with a 1:2 dilution of virus in two wells. At one-day post transduction media was changed for each virus into one of two scenarios: one well was treated with 1 μg/mL puromycin ((+) puro; Invitrogen) and the other was treated with normal growth media ((-) puro). (-) Puro treated cells were seeded for the luciferase assay upon

observation that the kill control in the (+) puro cells had reached complete cell death, approximately two days post-treatment.

Due to inconsistent host shutoff measurements in cells that received the scramble control shRNA lentivirus, the following modifications were made to the protocol for the second round. HEK 293A cells seeded into 6 well dishes at a confluency of 2.6 $x10^5$ cells per well and cultured as described in Section 2.1. Virus was diluted to 1:10 virus per well. Cells were transduced with one virus per well. Media was changed to 1 μg/mL Puromycin in normal growth media the day following transduction. Puromycin selection was carried out for 2 days before cells were rested with normal growth media for 24 hours prior to seeding for the luciferase assay.

In the third round, HEK293A cells were seeded and transduced as described in round two. After the 24-hour rest period, cells were seeded into a 12 well dish as mentioned in Section 2.1 with the exception of the cell seeding being 1:5. Cells were permitted 48 hours of growth before being seeded for the luciferase assay.

2.3.3 Luciferase assay

All luciferase assays were carried out following the same protocol. Stable cell lines were diluted to 2.0 $x10^5$ cells per mL with 50 μL seeded per well of a 96 well opaque flat-bottom dish. Each set of transduced cells were seeded in 6 well cluster dishes in regular growth media. One day post seeding, cells were transfected with 50 μl OPTI-MEM containing 0.3 μl FuGENE (Promega) per well and 40 ng pCR3.1: PA-X or pCR3.1, 40 ng pCMV: FF-luciferase-intron (Addgene), and 20 ng pCR3.1:

mRFP per well. Three wells were transfected with each condition (PA-X or empty vector) per cell line. Cells were removed from the incubator at 23 hours post transfection to cool to room temperature for 1 hour before 50 μl of Dual-Glo® Luciferase Assay System (Promega) substrate was added. Cells were rested for 10 minutes before being read on a Tecan M200Pro Plate Reader.

2.3.4 Host shutoff luciferase assay

HEK293A cells were seeded and transduced as described above for round three of the original luciferase assay screen. However, they were given an additional two days of growth. Cells were seeded at 5.2 x 10^4 cells per well in a 12 well dish, 4 wells were seeded per cell line. Transfection was performed via PEI, with 200 ng pCR3.1: PA-X or pCR3.1, 200 ng pCMV: FF-luciferase-intron and 100 ng pCR3.1: mRFP was transfected with 1.5 μl PEI in 200 μl OPTI-MEM per well in serum and antibiotic free media growth media. Media was changed 6 hours post transfection to regular growth media lacking antibiotics. 24 hours post transfection cells were washed with sterile PBS and lysed with 200 μl 1 x Passive lysis buffer (Promega) and rocked for 30 minutes at room temperature. 5 μl of cellular lysate was added to 25 μl of LAR II reagent from Promega and read on the GLO-MAX luminometer.

Protein lysates harvested using 2X Laemmli buffer (10% SDS, glycerol and 1M Tris-HCl pH6.8) were subjected to SDS-PAGE and immunoblotting using the above protocol for myc blotting. Changes to the protocol include blocking in 5% BSA TBS-T and 1° antibody was α-Influenza A PA (1:1000, GTX125932 Gene Tex).

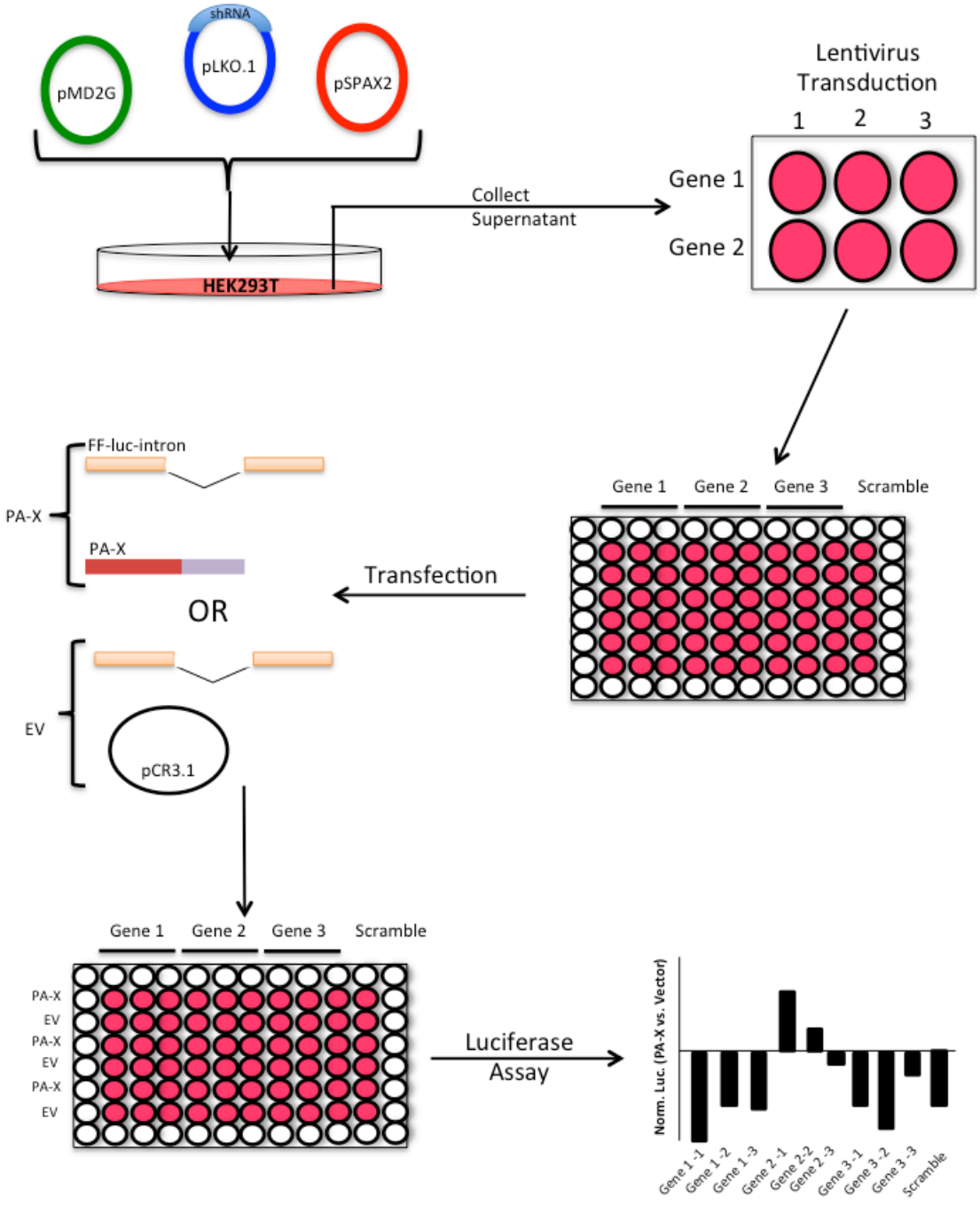


Figure 7: Schematic for RNA silencing and luciferase-based host shutoff assay

Figure 7: Schematic for RNA silencing and luciferase-based host shutoff assay (Continued)

Three shRNA sequences targeting each of the 29 genes that encode candidate X-ORF interacting proteins were selected from the RNAi consortium and individually cloned into the pLKO.1 plasmid. HEK293T cells were co-transfected with pLKO.1-shRNA plasmid and pMD2G and pSPAX2 lentiviral packaging plasmids, and lentiviruses were harvested from cell supernatants 48 hours later. HEK293A cells were transduced with shRNA-bearing lentiviruses; stable tranductants were selected with puromycin. These cells were re-seeded into 96-well cluster dishes and transfected with firefly-luciferase containing β-globin intron (FF-luc-intron) and PA-X or empty vector (EV; pCR3.1) for 24 hours. Cells were harvested and screened for luciferase activity via Dual-Glo® Luciferase Assay System (Promega). Normalized luciferase data was expressed as a ratio of relative light units (RLUs)from PA-X-containing samples to RLUs from control samples, and graphed on a $\log_2$ scale.

2.3.5 PA-X(D108A)-GFP localization

Confocal microscopy was performed in a similar manner to the immunofluorescence described in the context of BioID. HEK293A cells transduced with the shRNA CPSF6-1 and scramble control lentiviruses were seeded onto coverslips after puromycin selection and rest at a density described above. Cells were transfected via PEI transfection with pCR3.1: GFP- D108A PA-X from Khaperskyy *et al.* [92] . Due to only needing the Hoechst stain for the nucleus, at 24 hours post transfection cells were blocked with 5% BSA in sterile PBS and stained in the previously described manner for Hoechst. Once mounted to glass slides, Dr. Denys Khaperskyy imaged clover slips at 40x magnification on the Zeiss Axioplan II.

2.4 Measurement of RNA silencing

2.4.1 RNA isolation and cDNA production

Cells transduced with shRNA lentiviruses as previously described in Section 2.3.2 Round 3 and Section 2.3.4, were seeded in 12-well dishes, one well per cell line, at a confluency of 5.2 x10^4 cells/well. Transfection of cells also occurred as mentioned in Section 2.3.4. Qiagen RLT plus buffer was used to lyse cells 24 hours post transfection. RNA isolation occurred using Qiagen RNEasy® Plus Mini Kit from Qiagen following the protocol from the manufacturer. Isolated RNA was diluted in 20 μl at a final concentration of 10 ng/μl. Reverse transcriptase reactions were performed using the Quanta BioSciences qScript™ cDNA Super Mix Kit and protocol.

2.4.2 qPCR analysis of knockdowns

Primers were generated using the NCBI Primer Blast tool (https://www.ncbi.nlm.nih.gov/tools/primer-blast/). Each primer pair was designed to target all isoforms of the silenced target and were as follows: CPSF6 Forward: 5'- AGA CGT GAA CGA TCA AGA GAG A -3', CPSF6 Reverse: 5'- ACA AGC GTC ATT TTT CCC CC – 3', CPSF5 Forward: 5'- GTA CAT GAG CAC CGG CTA CC – 3', and CPSF5 Reverse: 5'- AAA CTC CAT CCT GAC GAC CC - 3'. Primers were ordered from Invitrogen. Protocol was carried out by Dr. Denys Khaperskyy using GoTaq® qPCR Master Kit adding 10 ng equivalents of RNA from the RT reaction to each mix; CPSF5, CPSF6, PRP4 and scrambled shRNA constructs were tested against CPSF5 and CPSF6 primers.

Chapter 3: Results

3.1 BioID identifies X-ORF interacting proteins

The BioID proximity proteomic method was used to identify X-ORF interacting proteins of the IAV host shutoff protein PA-X. Constructs were designed to replace the PA endonuclease domain of PA-X with the promiscuous biotin ligase BirA* (Figure 8A). Three separate N-terminal BirA* fusion proteins were designed by Dr. Denys Khaperskyy and cloned into the pCR3.1-myc vector. BirA*-X61 and BirA*-X61(4A) contained the 61 amino acid long X-ORF from A/Puerto Rico/8/1934 (H1N1); BirA*-X61(4A) had the following amino acid substitutions: K198A, R199A, K202A and K203A (Figure 8A). Because these four mutations prevent nuclear accumulation of PA-X, BirA*-X61(4A) serves as a negative control for these experiments [92] . The final construct, BirA*-X41, was the naturally occurring truncated X-ORF from A/California/7/2009 (H1N1) (Figure 8A). Both X61 and X61(4A) contained a C-terminal myc tag, while the truncated X41 did not. PA-X constructs from the different IAV strains were selected to identify a conserved mechanism of action for all isoforms of PA-X.

Plasmid constructs were individually transfected into 100 mm dishes containing HEK293T cells using the PEI method, treated with 50 μM biotin at 6 hours post transfection and harvested with RIPA buffer 18 hours later. Biotinylated proteins in the lysates were identified through far-western blotting analysis with streptavidin-conjugated HRP (Figure 8B and C). A differential banding pattern was

observed in BirA*-X61, BirA*-X61(4A) and BirA*-X41 samples relative to BirA* alone. Distinct bands were observed for BirA* -X61 at approximately 100 kDa, 45 kDa and 40 kDa, highlighted in Figure 8B with red arrows. BirA* has been shown to self-biotinylate and therefore the biotinylated 45 kDa species in both the BirA*-X61 and BirA*-X61(4A) lanes likely correlates with the bands observed for both BirA*-X61 and BirA*-X61(4A) on the anti-myc western blot, indicating the biotinylated fusion proteins (Figure 8B, bottom panel) [111] . There was also a distinct loss of bands observed at the top of the blot and bands approximately 50 kDa and 32 kDa for BirA*-X61, highlighted in Figure 8B with blue arrows. Loss of biotinylated proteins likely indicates differential subcellular localization of the BirA* construct; the fusion to the X-ORF influences the location and protein labelling by BirA* biotinylation. The 32 kDa band present in the BirA* alone lane is likely the self-biotinylated BirA* protein, as it correlates to the 32 kDa band in the anti-myc blot (Figure 8B, bottom panel). A similar differential banding pattern was seen for the BirA*-X61(4A) sample as seen with BirA*-X61 (Figure 8B); however, the band at 100 kDa is dramatically reduced and the band at 120 kDa is more predominant. This indicates that there are specific cellular proteins selectively interacting with the X-ORF compared to BirA* alone. More specifically, there are distinct proteins that interact with the wild type X61 compared to the X61(4A) mutant either through the amino acids that had been substituted in the 4A construct or are interacting with the X61 X-ORF due to nuclear localization.

As observed in the X61 constructs, the naturally occurring truncated version of the X-ORF of the A/California/7/2009 (H1N1) had a differential banding pattern

compared to the BirA* alone sample (Figure 8C). Distinct bands were observed at approximately 120 kDa, 100 kDa, 55 kDa and 40 kDa compared to BirA* alone. However, the band at 40 kDa is likely the autobiotinylated BirA*-X41 fusion protein (Figure 8C).

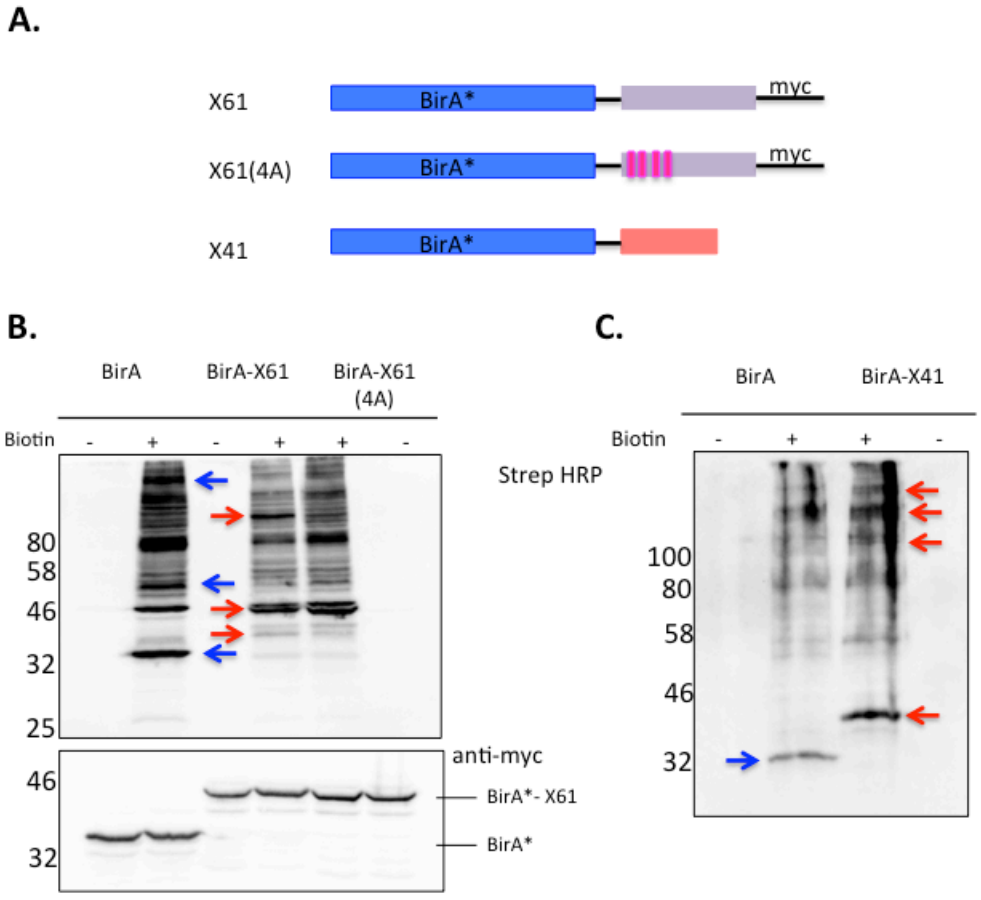


Figure 8: Host protein biotinylation by BirA*-X-ORF fusion proteins
Three separate BirA*-X-ORF constructs were used in the BioID system (A) BirA*-X61, BirA*-X61(4A), and BirA*-X41. Plasmids containing BirA* fusion proteins were transfected into HEK293A cells (B) and HEK293T cells (C) via PEI transfection. At 6 hours post-transfection, cells were treated with 50 μM biotin for 18 hours. 24 hours post-transfection, cells were lysed with RIPA buffer and analyzed via far-western (streptavidin-HRP; StrepHRP) and western blot (anti-myc). Biotinylated proteins from RIPA buffer whole cell lysates of (B) 293A (X61 and X61(4A)) and (C) 293T cells (X41) were separated by SDS-PAGE. Far-western blotting identified a differential banding pattern in X-ORF samples vs. BirA* alone. Blue arrows highlight loss of interacting proteins compared to BirA* alone, while red arrows signify distinct interacting proteins of the X-ORF.

The BirA* fusion proteins were predicted to localize in separate subcellular compartments; BirA*-X61 and BirA*-X41 were predicted to localize to the nucleus, whereas BirA*-X61(4A) alone was predicted to remain in the cytoplasm. Plasmids containing BirA*-X61-myc and BirA*-X61(4A)- myc were transfected into HeLa Tet-Off cells, treated with 50 μM biotin for 18 hours, and fixed and processed for immunofluorescence at 24 hours post-transfection (Figure 9). Cells were probed for biotin (green), myc tag (red) and nuclei (blue). BirA*-X61 was found predominantly in the nucleus of the cells, whereas BirA*-X61(4A) displayed a higher cytoplasmic retention. This corresponded to the majority of the biotinylated products as well; in the BirA*-X61 sample, a majority of the biotinylated products were nuclear, whereas in the BirA*-X61(4A) they were cytoplasmic. Interestingly, BirA*-X61 expressing cells saw the accumulation of biotinylated proteins in the nucleolus, but not the BirA*-X61 fusion protein itself (Figure 9B bottom panel). This observation suggested that the X-ORF transiently interacted with a protein that subsequently accumulated in nucleoli.

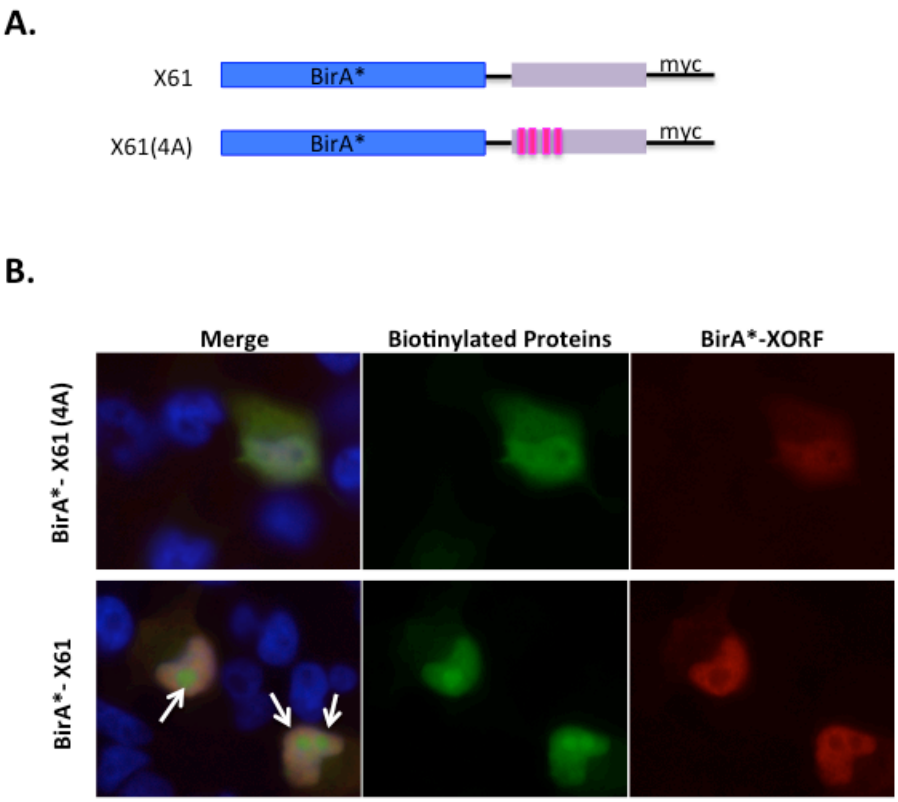


Figure 9: X-ORF transiently interacts with proteins that accumulate in nucleoli
BirA*-X61 and BirA*-X61(4A) plasmids (A) were transfected into HeLa Tet-Off cells seeded onto glass coverslips. At 6 hours post-transfection, cells were treated with 50 μM biotin. At 24 hours post-transfection cells were fixed and stained for (B) immunofluorescence microscopy: biotinylated proteins (green), myc tag fusion protein (red) and nuclear Hoechst stain (blue). Biotinylated proteins co-localized with fusion proteins, with the exception of BirA*-X61. White arrows in BirA*-X61 merge channel indicate nucleolar biotinylated proteins. Images were captured at 40x magnification on an AXIO Imager Z.2 microscope.

3.2 Discovery of candidate X-ORF interacting proteins by quantitative proteomics

Due to the observed differential banding pattern in the far-western blot analysis and the nucleolar staining in the immunofluorescence assay (Figure 8B and C, 9B), mass spectrometry analysis was selected to identify the biotinylated proteins. A quantitative mass spectrometry approach was selected to determine the relative abundances of identified biotinylated proteins in each of the experimental conditions. Each of the constructs were delivered to HEK293T cells via PEI transfection, treated with 50 μM biotin for 18 hours and harvested using RIPA buffer 24 hours post-transfection. Biotinylated proteins were separated from whole cell lysates through affinity purification with Neutravidin beads and washed with RIPA and TAP buffers. Prior to on bead digestion with trypsin, two of the three experimental conditions were treated with RNaseA to identify RNA-independent interactions. Mass spectrometry analysis was conducted with three separate conditions: (1) BirA*, BirA*-X61 and BirA*-X61 (4A) without RNaseA (X61 (-)), (2) BirA*, BirA*-X61 and BirA*-X61 (4A) with RNaseA (X61), and (3) BirA*, BirA*-X41 and BirA*-X61 (4A) with RNaseA (X41).

To derive a high confidence list of interacting proteins, the quantitative approach of reductive dimethylation was selected. This approach allows for the relative abundances of each identified protein to be compared in each of the separate conditions; this is because differential labelling permits a single run of mass spectrometry for comparison. This is advantageous because the conditions of

all samples remained the same, with the same amount of protein added to the beads for affinity purification, amount of differential labelling, and finally the ratio of samples pooled for analysis via mass spectrometry was 1:1:1. Throughout three different rounds of mass spectrometry labelling remained consistent; the light sample was BirA*, medium was the X-ORF (X61 or X41) and heavy was X61 (4A). Samples were prepared under the advice and guidance of Dr. Alejandro Cohen of the proteomics' CORE facility at Dalhousie University. Peptide identification data was acquired and prepared by Dr. Cohen. Upon obtaining the data set of identified proteins, any with less than 2 unique peptides were automatically excluded from further analysis.

Further analysis was conducted on the data sets by examining the relative abundance of each protein relative to the control BirA* alone sample. Section B of Figures 10-12 highlight the relative abundances of each of the target proteins identified in each data set through comparison of their fold increase of the X-ORF to BirA* alone (x-axis) and the X-ORF to the mutated X61(4A) (y-axis). The minimal inclusion criteria were met if a candidate interacting protein had a fold increase greater than or equal to 1.5 X-ORF vs. BirA*. Each data set had distinct outlying interacting proteins that are found only within a single trial. For example, PSME3 has a high fold increase X-ORF to BirA* alone and the X-ORF to the mutated X61(4A) within the X61 (-) sample, but is not found with the same abundances in either the X61 or X41 subsets. Largely, the target proteins fell within a tight cluster in the subsets, ranging from 1.5- to 4-fold increase in the X-ORF fusion protein samples compared to BirA* alone. Both X61 samples saw the cluster ranging from 0.5- to 3-

fold increase in the X-ORF compared to the mutated X61(4A) (Figure 10 and 11). However, X41 showed proteins that had higher levels of specificity of binding the four charged residues of the X-ORF, as the cluster peaked from 0.5- to 10-fold increase X-ORF compared to the mutated X61(4A) (Figure 12).

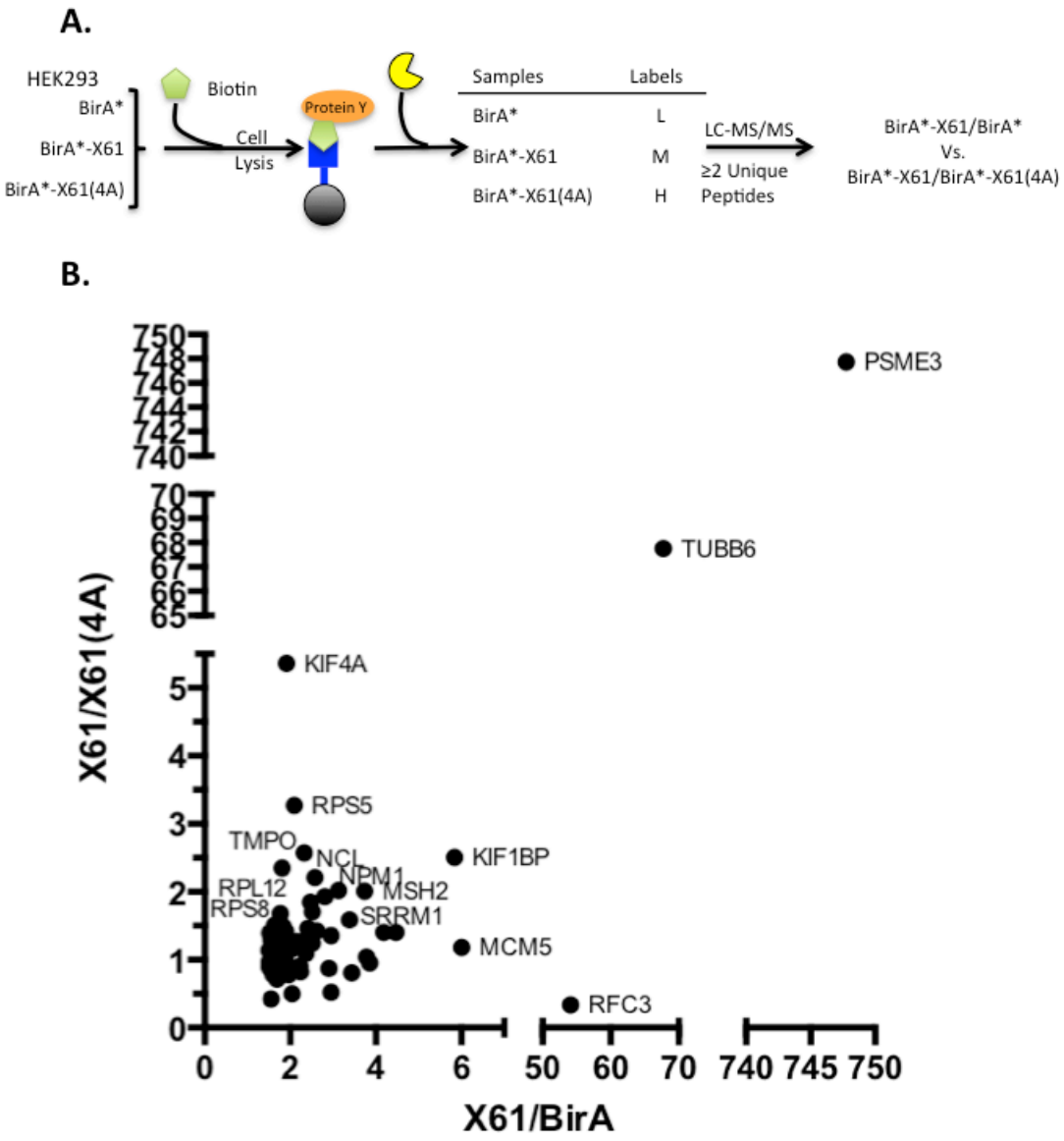


Figure 10: Identification of biotinylated X-ORF interacting proteins by mass spectrometry

Figure 10: Identification of biotinylated X-ORF interacting proteins by mass spectrometry (Continued)
Visualization of the first round of mass spectrometry data with X61 X-ORF interacting proteins relative abundances of X61/BirA* vs. X61/X61(4A). (A) Schematic depiction of methods used to obtain data set in (B). HEK293T cells were transfected with each BioID construct and treated with 50 μM biotin (green) for 18 hours. Cells were RIPA buffer lysed and harvested 24 hours post-transfection. Proteins were captured on Neutravidin beads (black and blue) before digestion with trypsin (yellow). Peptide abundance was determined through reductive dimethylation labeling of the samples of light (BirA*), medium (X61) and heavy (X61(4A)) before mass spectrometry analysis. Inclusion criteria included any protein with 2 or more unique peptides and a 1.5-fold increase X61/BirA*. X61/BirA* indicates proteins with specificity to the X-ORF over BirA* alone, while proteins with a greater abundance in X61/X61(4A) show specificity to four basic amino acids or nuclear localization. A majority of identified proteins are clustered between 1.5- to 4-fold increase in both X61/BirA* and X61/X61(4A).

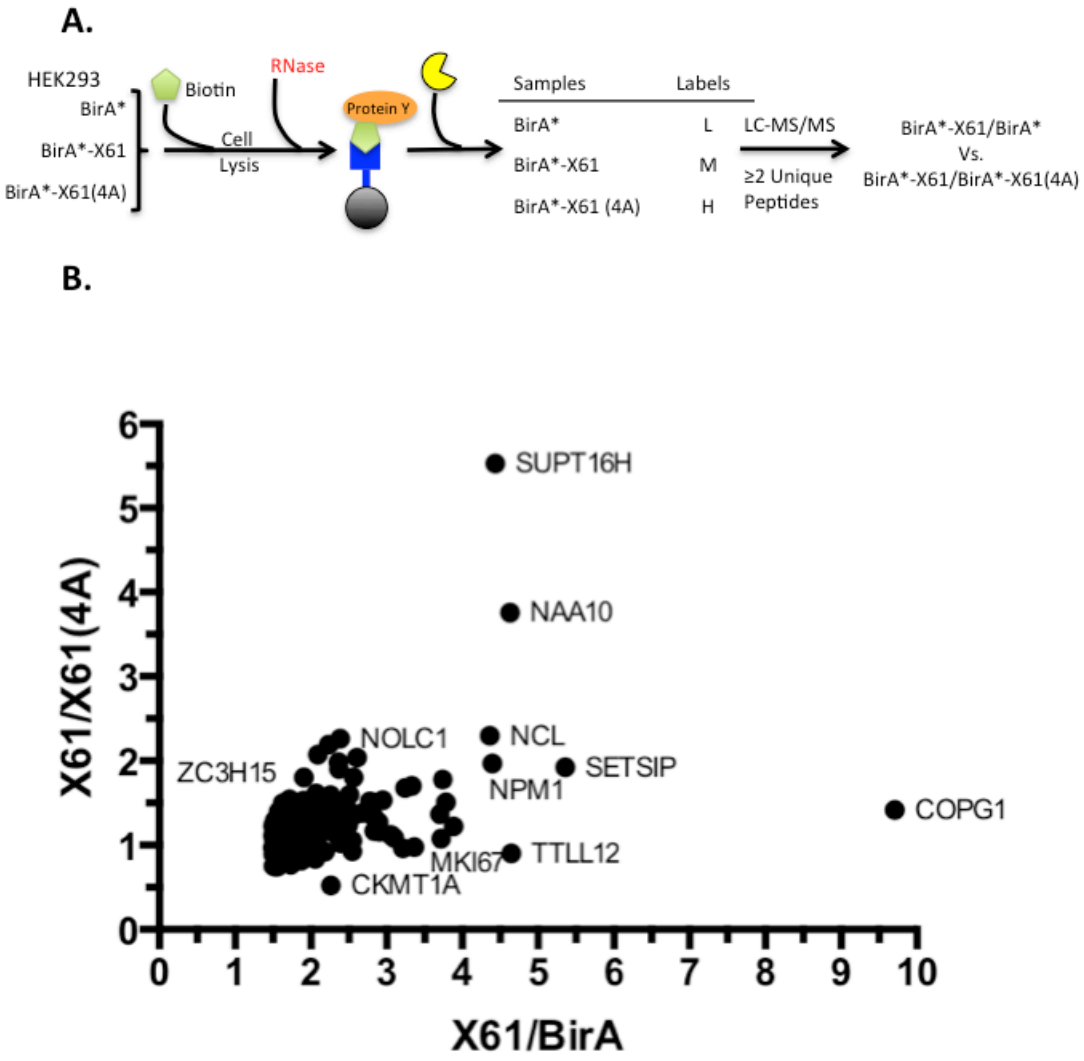


Figure 11: RNase treatment reveals RNA-dependent interactions between the X-ORF and host proteins

Figure 11: RNase treatment reveals RNA-dependent interactions between the X-ORF and host proteins (Continued)

Visualization of the second round of mass spectrometry data of biotinylated interacting proteins for RNA-independent interactions between BirA*-X61 and host proteins. (A) HEK293T cells were transfected with each BioID construct and treated with 50 μM biotin (green) for 18 hours. 24 hours post-transfection cells were lysed with RIPA buffer and harvested. Prior to Neutravidin bead pull-down, RNaseA treatment was conducted on biotinylated samples to eliminate proteins associated with RNA from the data set, as indicated in red (A). Proteins captured on the Neutravidin beads (black and blue) proteins were digested with trypsin (yellow). Reductive dimethylation labeling with light (BirA*), medium (X61), or heavy (X61(4A)) formaldehyde enabled determination of relative abundance of X-ORF interacting proteins. Data was processed with the following inclusion criteria: greater than or equal to 2 unique peptides identified and a relative abundance of greater than 1.5 fold increase X61/BirA*. Relative abundances were compared to identify X-ORF specific interactions (X61/BirA*) vs. basic amino acid interactions or nuclear localized interaction (X61/X61(4A)). Most interacting proteins identified in this round of mass spectrometry fell between 1.5- to 4-fold increase X61/BirA* and X61/X61(4A).

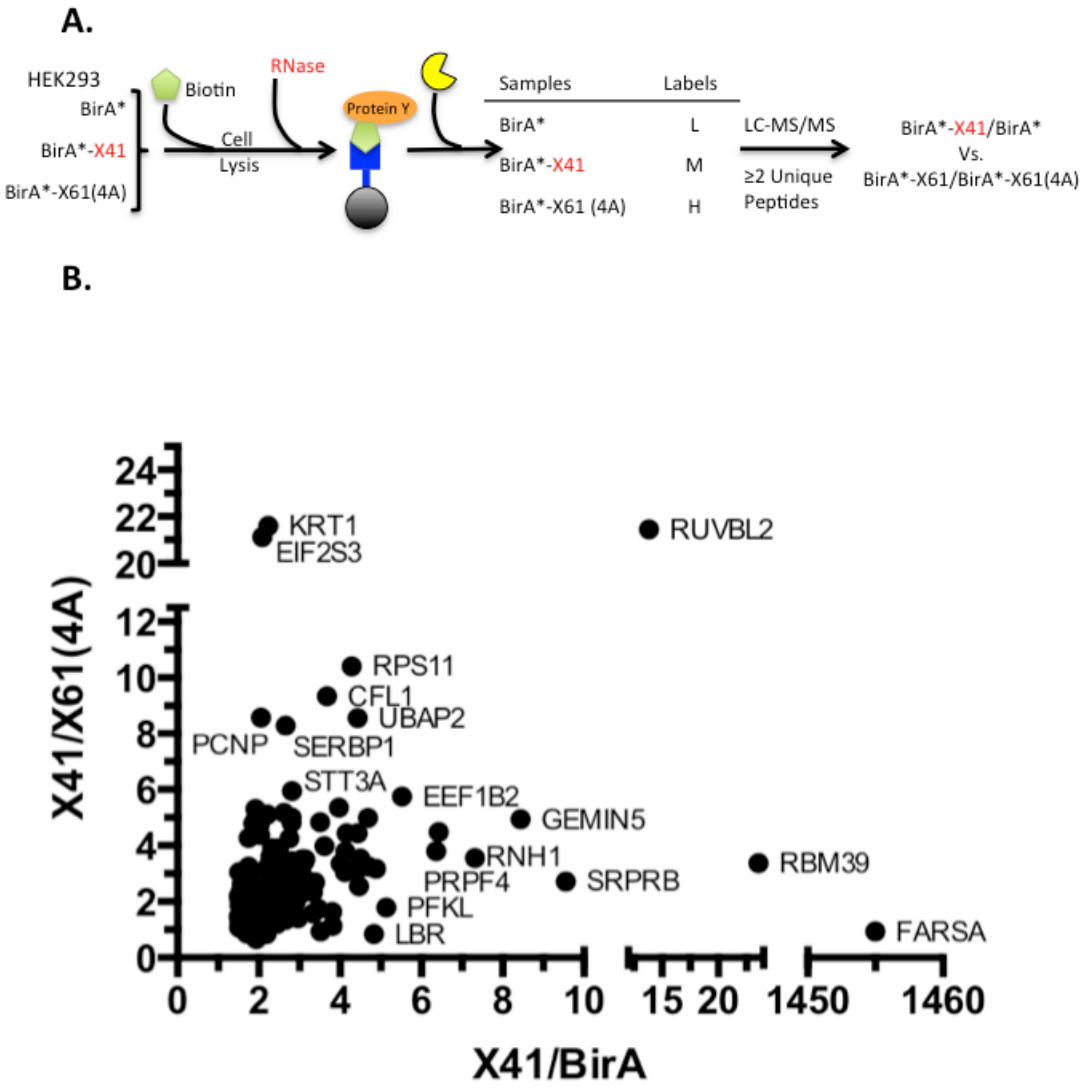


Figure 12: X41 has different outlying interacting proteins

Figure 12: X41 has different outlying interacting proteins (Continued)
Graphed data of the third round of mass spectrometry data identified interacting proteins of the naturally truncated X41 X-ORF. (A) X41 (red) without a myc tag was used to identify common interacting proteins among all X-ORF isoforms. HEK293T cells were transfected with BirA* fusion constructs and treated with 50 μM biotin (green) for 18 hours. Cells were lysed with RIPA buffer and harvested at 24 hours post-transfection. Prior to Neutravidin bead pull-down, lysates were treated with RNaseA (red) to eliminate RNA-bridged interactions. Biotinylated proteins were captured on Neutravidin beads (black and blue) and digested with trypsin (yellow). Biotinylated protein relative abundances were determined through reductive dimethylathion: light (BirA*), medium (X41) and heavy (X61(4A)). (B) Identified interacting proteins were plotted only if 2 or more unique peptides were identified and contained a greater than 1.5-fold increase X41/BirA*. X-axis, X41/BirA*, indicates X-ORF specific interactions, whereas, y-axis, X41/X61(4A) indicates interactions specific to nuclear localization or the four basic amino acid residues. A majority of the proteins identified fall within the range of 1.5-8-fold increase X41/BirA* and X41/X61(4A).

Of the identified proteins only 159 were common among all three samples, as shown in the Venn diagram in Figure 13A. Upon closer investigation of the target protein lists 80, 212, and 162 proteins were identified to have greater than 1.5-fold increase X-ORF to BirA* in the X61 (no RNase), X61 (RNase), and X41 samples respectively (Figure 13B). Of the proteins within each data set, only an overlap of 18 proteins was found amongst all three (Figure 13B). To aid in narrowing the list of interacting proteins, without eliminating possible proteins of interest due to experimental error, the criteria was set as follows: a 1.5 fold increase in 2 out of the 3 samples, with an average of greater than 1.7 fold increase in X-ORF to BirA* and the protein must have appeared in all three mass spectrometry data sets, or the target proteins must have had a greater than 2 fold increase in X-ORF to BirA* in 2 out of the 3 mass spectrometry data sets. The list was further narrowed for proteins that may have implications on PA-X due to the known host shutoff activity of the protein. This resulted in a list of 29 target proteins that can be found in both heat and STRING map formats (Figure 14 and 15).

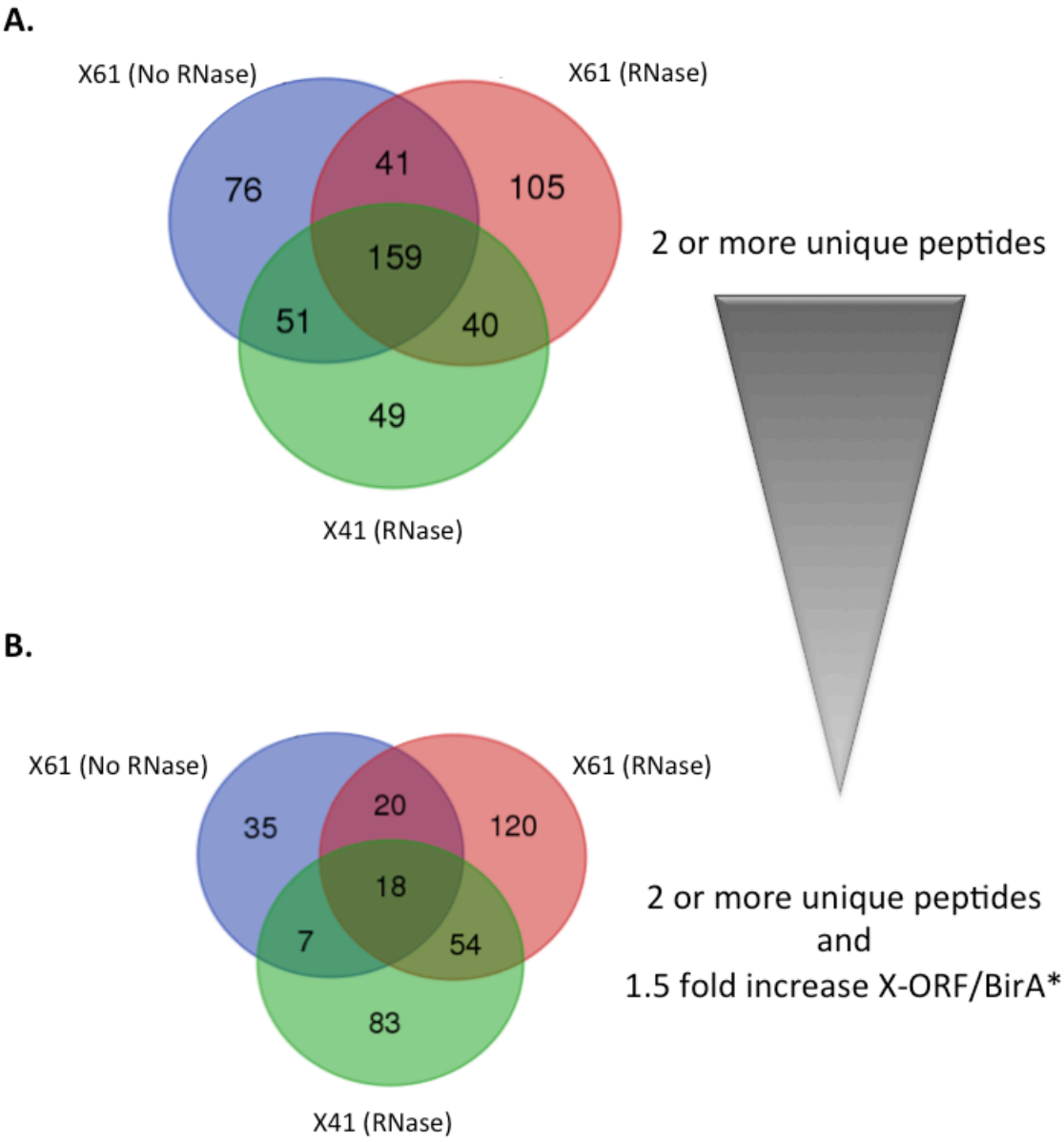


Figure 13: Stringent filtering yields a short list of high-confidence candidate interacting proteins

Comparing all three rounds of mass spectrometry to each other via Venn diagrams revealed common interacting proteins. (A) An overlap of 159 interacting proteins was initially identified with proteins having 2 or more unique peptides. When the criteria included specific parameters for X-ORF-interacting proteins, 2 or more unique peptides per identified protein and greater than or equal to 1.5 fold increase X-ORF/BirA*, (B) the number of conserved interacting proteins is reduced to 18. Black arrow indicates the narrowing of criteria from Venn diagram (A.) to (B.). Venn diagrams were produced using [125] .

As seen in the heat and STRING maps (Figure 14 and 15), the list of 29 proteins can be subdivided into 7 different clusters: apoptosis (blue), transcription/translation (green), DNA interactor (purple), nuclear proteins (teal), nuclear import/export (gold), mRNA processing (red), and miscellaneous (not attached). This same clustering system was used for heat map analysis of the target proteins relative abundances relative to each other (Figure 14); the highest abundances are coloured red and the lowest blue. The scale was set from 1 to 4 due to a majority of the target proteins falling within that range for observed fold increase. However, some outliers do exist such as seen in Figure 10, PSME3 had a fold-increase of greater than 700 in both the X-ORF/BirA* alone and the X-ORF/X61(4A). This high relative abundance was only observed for PSME3 in the non-RNase treated sample. However, PSME3 was present in all experimental conditions, indicating RNA-bridged interactions between the X-ORF and PSME3 (Figure 14). There were also a few proteins that had a relative abundance ratio between 0.5- and 1-fold difference, however to ensure the relative abundances were clearly and uniformly mapped in the colour range, 1 to 4 was selected. High abundances of biotinylated interacting proteins were identified in each functional node in all experimental conditions, these include: the mRNA processing node (PUF60, CPSF6, NUDT21 (CPSF5) and PRPF4), transcription/translation (RPL24, NACA and BTF3), DNA interactors (NASP and MCM5), and nuclear proteins (NCL and NPM1) (Figure 14). Grey squares on the heat map indicate no protein abundance data available; these interacting proteins met the criteria of 2 fold increase X-ORF/BirA* and only found in 2 of the 3 mass spectrometry experiments.

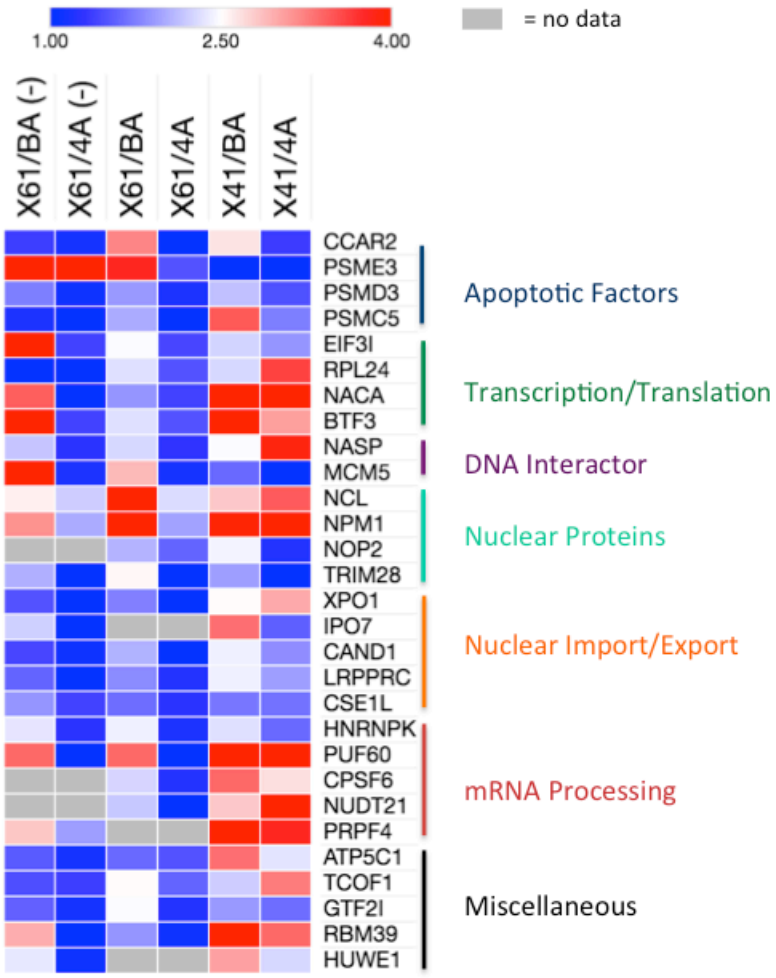


Figure 14: Proteins with mRNA processing and transcription/translation functions are highly-enriched amongst candidate X-ORF interacting proteins
Heat map of relative abundances for the 29 candidate X-ORF interacting proteins across each BioID trial. The list of 29 proteins was determined from the following criteria: 2 or more unique peptides identified for all proteins, > 1.5 fold increase X-ORF/BirA* in 2 of 3 samples but the interacting proteins appears in all 3 trials, or > 2 fold increase X-ORF/BirA* with the interacting protein appearing at a minimum of 2 out of 3 samples. Proteins are grouped by STRING functional groupings identified in Figure 15. A range was set from 1-to 4-fold to illustrate the differences between all interacting proteins relative abundances in the mass spectrometry data. Grey on the heat map indicates no peptides were detected for that candidate protein. Heat map designed using the Morpheus heat map tool [126] .

Figure 15: STRING analysis identified seven functional groupings of candidate X-ORF interacting proteins

Figure 15: STRING analysis identified seven functional groupings of candidate X-ORF interacting proteins (Continued)

STRING map of the list of 29 proteins identified as interacting protein of the X-ORF under the criteria of: 2 or more unique peptides identified for all proteins, > 1.5 fold increase X-ORF/BirA* in 2 of 3 samples but the interacting proteins appears in all 3 trials, or > 2 fold increase X-ORF/BirA* with the interacting proteins appearing at a minimum of 2 out of 3 samples. Each node was individually coloured as follows: apoptotic factors (blue), transcription/translation (green), DNA interactors (purple), nuclear proteins (teal), nuclear import/export (gold), mRNA processing (red) and miscellaneous (no strings, multi-coloured). Proteins with known structures are indicated with large spheres containing protein structure, whereas small spheres indicate no available protein structural information. STRING map designed using STRING software [127] .

3.3 PA-X host shutoff activity does not rely on all interacting proteins

Investigation of functional interactions between candidate X-ORF interacting host proteins and PA-X was conducted by RNA silencing and host shutoff measurements. HEK293A cells were transduced with lentiviruses bearing shRNAs targeting each candidate host gene. Three shRNAs were selected for each gene of interest. A luciferase assay screen was designed to test the PA-X host shutoff activity by using a RNA pol II driven luciferase construct containing a β-globin intron. An intron was included in this construct because we have unpublished data indicating PA-X sensitivity correlates with mRNA splicing potential (M. Gaglia, personal communication). The pCMV-FF-luc-intron construct was therefore the best candidate to determine PA-X host shutoff function in the knockdown cell lines.

Three different protocols for identifying interacting proteins crucial for PA-X host shutoff activity were set up. This was a result of optimization of the luciferase-assay screen by processing the list of 29 identified interacting proteins. Acceptable values for PA-X host shutoff activity in a luciferase assay screen lie between 0.13- and 0.5-fold change of normalized luciferase relative light units (RLUs) of PA-X containing samples to empty vector containing samples, as previously reported by Khaperskyy *et al.* with firefly luciferase and Renilla luciferase respectively [92] . Due to scramble control values outside of this range for several trials of the screen, an optimization process was conducted. The optimization process progressed as follows: round 1 (Figure 16B) had a 1:2 virus dilution with no Puromycin selection and two day incubation period prior to seeding into a 96 well dish, round 2 (Figure 16D) had a 1:10 virus dilution with puromycin selection for two days and one day

rest prior to seeding into a 96 well dish, and round 3 (Figure 16F-I) had a 1:10 virus dilution with puromycin selection for two days and three days of rest prior to seeding into 96 well dish. All shRNA expressing cell lines were transiently transfected with PA-X and firefly luciferase the day after cells were seeded into 96 well dishes. 24 hours post-transfection, cells were lysed and the intensity of the firefly luciferase was read.

Figure 16 shows the result of the screen for PA-X host shutoff activity. All data is normalized to the empty vector controls for each of the cell lines, and represented as a ratio of PA-X to empty vector on a $\log_2$ scale. The scramble control in all rounds of the screen behaved similarly to what was expected with values of each scramble control hovering around 0.5, indicating PA-X was inhibiting the luciferase production in the PA-X conditions over the empty vector control.

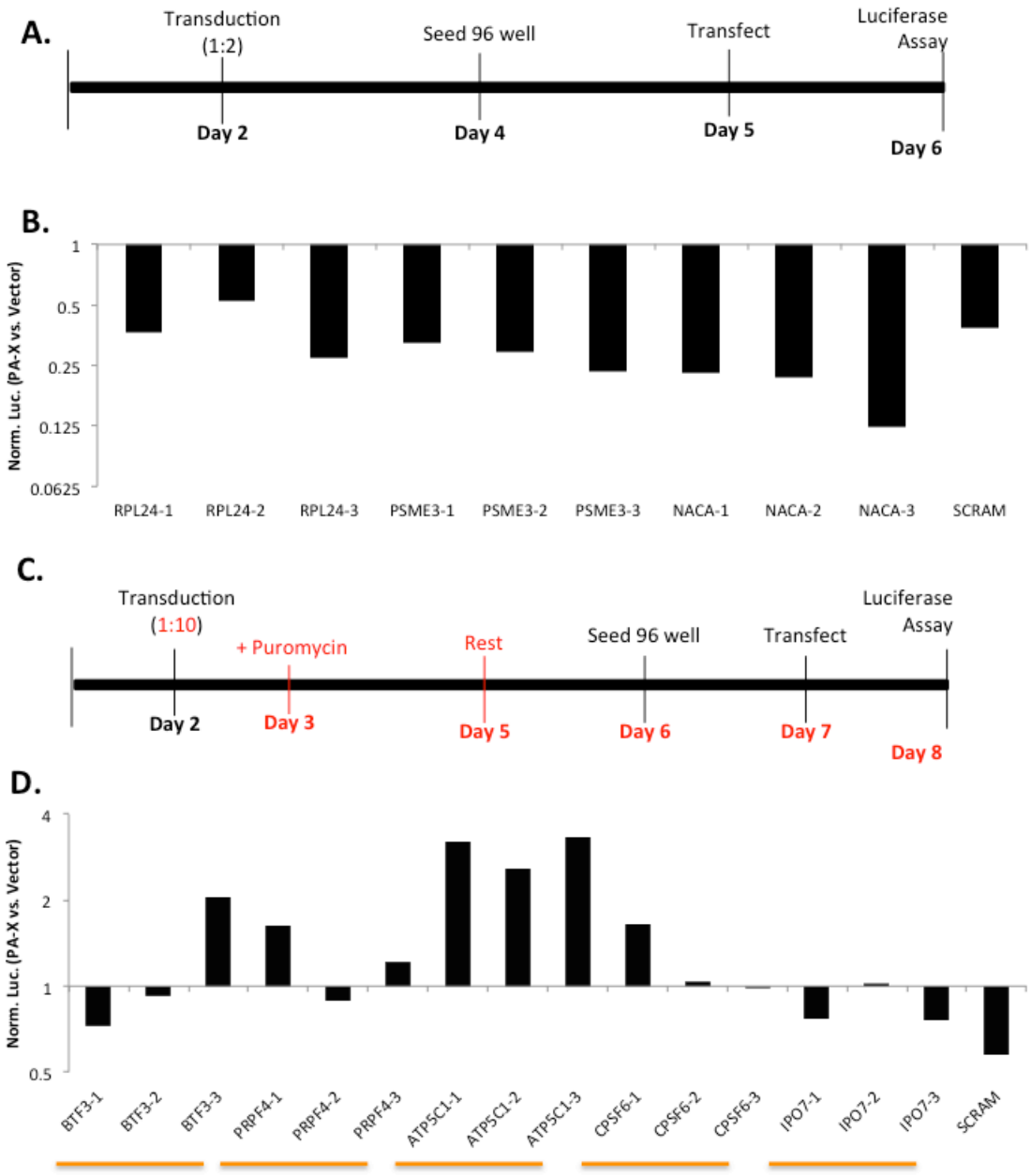


Figure 16: Identification of protein-protein interactions crucial for PA-X host shutoff activity

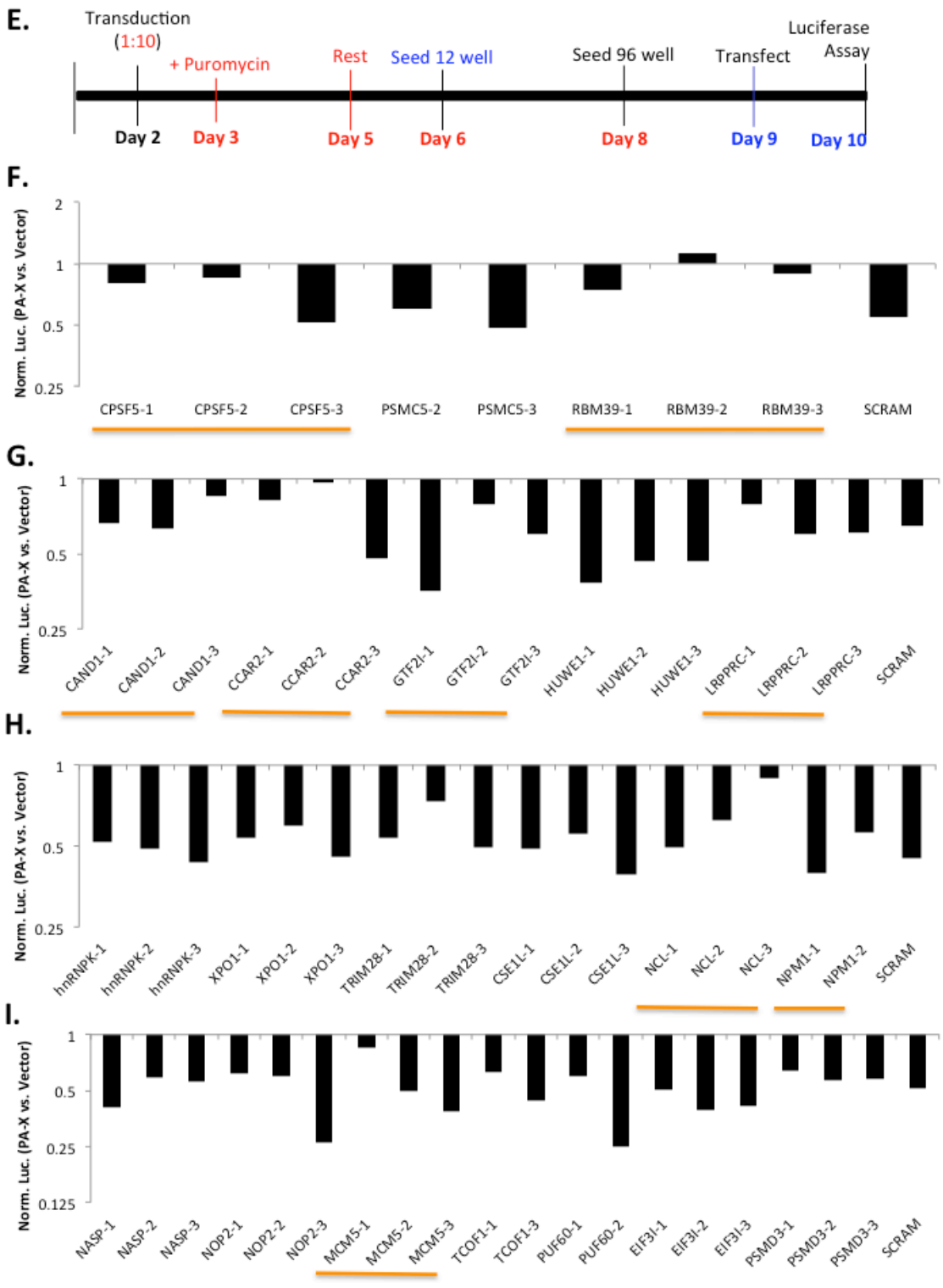


Figure 16: Identification of protein-protein interactions crucial for PA-X host shutoff activity (Continued)

Figure 16: Identification of protein-protein interactions crucial for PA-X host shutoff activity (Continued)
HEK293A cells were transduced with lentiviruses containing shRNAs targeting host genes that encode 29 candidate X-ORF-interacting proteins. Three shRNA sequences were selected for each gene for optimal silencing and denoted "gene name" 1-3. Luciferase assays were conducted as described in (A) for round 1, (C) for round 2, and (E) for round 3. Changes to each round are indicated with different colours, changes in round 2 are highlighted in red and changes occurring in round 3 are highlighted in blue. The three rounds of luciferase assays were completed on cells transduced with lentivirus (B) 1:2 virus dilution with no puromycin selection on cells or rest days, (D) 1:10 virus dilution, puromycin selection and one day rest, and (F)-(I) 1:10 virus dilution, puromycin selection followed by one day rest prior to splitting into 12-well dish, allowing for two days of growth. In all experimental conditions, shRNA expressing cell lines were seeded in 96 well dishes and transfected with PA-X or pCR3.1 and luciferase-containing β-globin intron for 24 hours prior to luciferase assay. Values in the luciferase assay are the average of three technical replicates and normalized PA-X to empty vector ratio on a $\log_2$ scale. Orange underlines throughout the luciferase assay indicate candidates examined further in the Discussion. PA-X host shutoff function was assessed by the ability of PA-X to reduce luciferase expression via a reduction in RLUs in PA-X samples vs. empty vector. Scramble control was determined to be acceptable at 0.5-fold. Proteins that had a fold-change between 0.5 and 1 were categorized as weak candidates, while those around 1 or above the line were strong candidates. Proteins that were below the scramble control were considered host shutoff enhancing candidates.

Round 1 RNA silencing targets had little effect on PA-X host shutoff activity, as in their absence PA-X behaved similar to the scramble control (Figure 16B). The NACA-3 shRNA expressing cell line was the exception, as it did appear to slightly increase PA-X host shutoff activity.

RNA silencing targets in round 2 had the greatest impact on PA-X host shutoff activity. Only BTF3-1 behaved similarly to the scramble control (Figure 16D), all other targets in this screen indicate an interaction that promotes host shutoff behaviour of PA-X. When silenced through shRNAs, these proteins greatly hinder the ability of PA-X to degrade luciferase mRNA. This is represented with the higher output of RLUs for the luciferase construct, represented in a fold increase of normalized luciferase PA-X to empty vector ratio (Figure 16). Though most genes show one shRNA target to have the strongest effects (BTF3-3, PRP4-1 and CPSF6-1), all three of *ATP5C1* targets appear to be similar in inhibiting the effectiveness of PA-X host shutoff activity (Figure 16D). This difference in effectiveness could possibly indicate that the three shRNAs of ATP5C1 are functional at silencing the target, while the other gene targets only have one strong shRNA.

Targets that saw a slight increase of luciferase production due to lack of PA-X degradation, for example the values of PA-X to empty vector ratio hovered around 1, were deemed to be "weak" hits. Weak interactions were the most common within Round 3(Figure 16F-I). shRNAs for CPSF5-1/2, as well as, RBM39 1-3 were considered weak hits. PSMC5-1 had no data collected due to the lack of viable cells in the assay. However, both PSMC5-2 and PSMC5-3 indicated similar behaviour to the scramble control of PA-X and therefore deemed not a factor in PA-X host shutoff.

The targets screened in Figure 16G, for the most part, behaved similarly to the scramble control of PA-X. However, there were a few silenced genes that indicated limited PA-X host shutoff function; those include CAND1-3, CCAR2-1 CCAR2-2, GTF2I-2 and LRPPRC-1. Interestingly, there were three shRNA expressing cell lines that appear to increase PA-X functionality; those include CCAR2-3, GTF2I-1, HUWE1-1, HUWE1-2 and HUWE1-3, as seen by the greater fold decrease of normalized RLUs compared to the scramble control.

Similarly, the targets screened in Figure 16H saw PA-X behaving similarly to the scramble control in all shRNA expressing cell lines, with the exception of TRIM28-2 and NCL-3. As seen in the previous screens, NPM1-3 was not included in the screen due to the lack of viable cells for the screening process.

Following a similar pattern to candidate proteins in Figure 16H, the candidates in Figure 16I saw little change to PA-X host shutoff capability. MCM5-1 shRNA expressing cells were the exception and MCM5 deemed a weak functional interacting protein of the X-ORF with respect to PA-X host shutoff activity. TOCF-2 and PUF60-3 were not included in the screen due to a lack of viability in cells for the screen to occur.

Due to the observed results, it was hypothesised that the X-ORF interacting proteins that saw a negative impact on PA-X host shutoff activity, seen by an increase in luciferase production over the empty vector control, were crucial PA-X functional interacting proteins. Therefore, these targets were selected for validation and further evaluation of the results found. However, based upon reported data of

PA-X interaction with mRNA 3'-end processing the focus was decided to be primarily on the mRNA processing node (Figure 15) [92] .

3.4 Validation luciferase assay identifies CPSF6 as integral partner of PA-X

The mRNA processing node was chosen in the validation luciferase assay as it contained interacting proteins that had aided (PRP4, CPSF6 and CPSF5), as well as those that did not aid (PUF60 and hnRNPK), PA-X host shutoff activity, respectively (Figure 16F, D, H and I). ATP5C1 shRNAs were also included in the screen due to the robust luciferase expression in the presence of PA-X for all three of the *ATP5C1* silenced cell lines (Figure 16D). HEK293A cells were transduced similar to the method outlined for Figure 16F-I; however, cells were afforded an additional 2 days rest to allow for more robust cell growth. Similar to the previous luciferase screen the RNA silenced cell lines were transfected with PA-X or empty vector and pCMV-FF-luc-intron for 24 hours before harvesting and detection of RLUs based off luciferase production.

The scramble control was performed in the same manner as seen in the previous assay and in accordance to the acceptable range of normalized luciferase RLUs PA-X to empty vector ratio (Figure 17). A majority of the validation results mirrored the previous screen, however inconsistencies were present for some functional candidates. Most notably, this includes all three shRNAs for ATP5C1. Figure 16D indicates RNA silenced *ATP5C1* cells prevented PA-X from performing normal host shutoff activities, as more luciferase was produced relative to the empty vector control. However, in the validation luciferase assay none of these *ATP5C1* silenced cells provided the same phenotype. The RNA silenced cells for

ATP5C1 showed no effect on PA-X host shutoff activity and shared a similar ratio of luciferase production to that of the scramble control (Figure 17). PRPF4-1 and CPSF6-2 also saw phenotype reversal of PA-X host shutoff activity in the validation luciferase assay compared to their previous screen (Figure 16D).

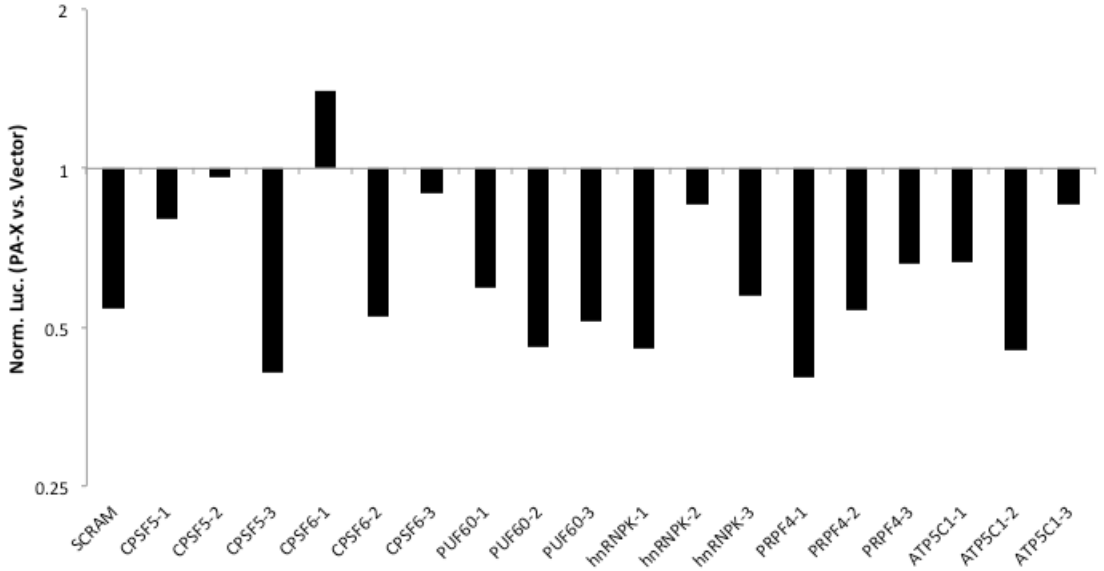


Figure 17: CFIm complex proteins CPSF5 and CPSF6 are required for PA-X host shutoff activity

A host shutoff luciferase assay to validate findings from Figure 16 was conducted. HEK 293A cells were transduced with lentiviruses bearing the indicated shRNAs and stable transformants were selected with puromycin. Following one day of rest, cells were seeded into 12-well cluster dishes. An additional two days growth was permitted before seeding shRNA expressing cell lines into 12-well cluster dishes in duplicate. Cells were transfected with PA-X or empty vector and luciferase containing β-globin plasmid for 24 hours prior to harvesting lysates for luciferase assays. Data plotted is the average of two technical replicates and normalized RLUs for the PA-X containing samples to the empty vector control-containing sample on a $\log_2$ scale. The majority of the shRNA-expressing cell lines repeated the results found in Figure 16, however a distinct difference present in this assay is the reversal of the phenotype for ATP5C1.

Because multiple host shutoff luciferase reporter assays indicated that CPSF6 was required for PA-X host shutoff activity (Figure 16D and Figure 17), it was selected for further investigation, along with CPSF5 and PRPF4. Reverse transcriptase- quantitative polymerase chain reaction (RT-qPCR) was used to confirm silencing of target genes (Figure 18A). Cells containing the *PRPF4*, *CPSF5*, and *CPSF6* shRNAs were lysed and harvested for RNA using the Qiagen RNeasy kit. cDNA from each of the cell lines was produced using a RT reaction. The cDNA produced in the RT reaction was used in the qPCR reactions. Primers for CPSF5 and CPSF6 were tested against all samples, including the scramble control. CPSF5 was effectively silenced by shRNA (Figure 18A). However, CPSF6 was up-regulated in CPSF5 shRNA-expressing cells. Moreover, a similar phenotype was observed for CPSF5 in the CPSF6-expressing cells; CPSF5 was up-regulated when CPSF6 was silenced (Figure 18A). An exception was CPSF6-1 expressing cells; these cells saw both a reduction in CPSF5 and CPSF6. By contrast, CPSF5 or CPSF6 mRNA levels were unaffected by PRPF4 silencing (Figure 18A). These findings suggest that there may be a mechanism that responds to changes in CPSF5/6 levels and regulates their biogenesis. Control immunoblotting experiments confirmed that PA-X protein accumulation was generally unperturbed by silencing of CPSF5 or CPSF6 (Figure 18B and C).

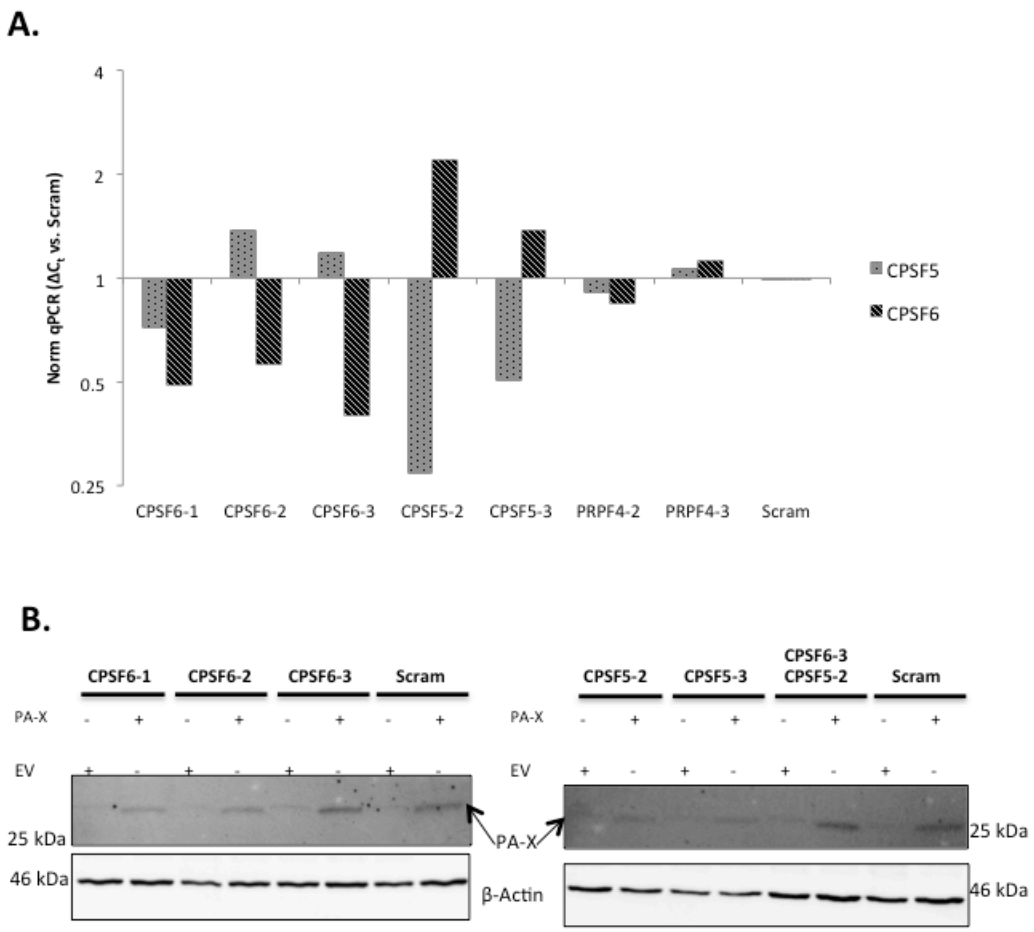


Figure 18: Confirmation of CPSF5 and CPSF6 silencing via RT-qPCR
(A.) RNA was harvested from cells stably transduced with lentiviruses bearing the indicated shRNAs (as well as PA-X expression vector or empty vector control and luciferase reporter plasmid), and processed for RT-PCR with gene-specific primers (CPSF5, grey bars with black polka dots; CPSF6, black and white striped bars). RT-qPCR was performed by Dr. Denys Khaperskyy (B). Whole cell lysates were processed for immunoblotting with antibodies for PA-X and β-actin. Cell lysates from separate experiments were processed in (A) and (B).

3.5 CPSF6 is not required for nuclear accumulation of PA-X

It was hypothesized that the BioID screen could reveal X-ORF-interacting proteins required for the nuclear accumulation of PA-X. Among the 29 high-confidence hits from the BioID screen, CPSF6 was particularly interesting because it was required for efficient PA-X-mediated host shutoff (Figure 16D and 17). Interestingly, CPSF6 binds HIV capsid proteins, thereby aiding nuclear translocation of the HIV genome [128] . Therefore, I hypothesized that CPSF6 might similarly be required for nuclear accumulation of PA-X. To test this directly, a catalytically inactive PA-X point mutant (D108A) protein fused to GFP was used to monitor PA-X trafficking in cells deficient in CPSF6 or scrambled control cells. The D108A mutant was chosen because it is known to traffic to the nucleus normally, but fails to cleave RNA pol II-driven transcripts, including transcripts generated from plasmids [92] . Thus, the advantage of the D108A mutant PA-X protein is that it will not degrade the transcripts that encode it, thereby ensuring synthesis of the fusion protein. CPSF6 silenced cells and scrambled control cells were seeded onto coverslips, and transfected with the PA-X(D108A)-GFP construct for 24 hours before fixation and processing for epifluorescence microscopy. The PA-X(D108A)-GFP construct accumulated in nuclei, which were identified by staining with Hoechst dye (Figure 19). *CPSF6* silencing had no effect of the nuclear accumulation of the fusion protein. These findings suggest that CPSF6 is required for some aspect of PA-X host shutoff, but it is dispensable for PA-X translocation into the nucleus. The punctate nuclear staining pattern suggests that the fusion protein accumulates in a sub-nuclear compartment in a CPSF6 independent manner.

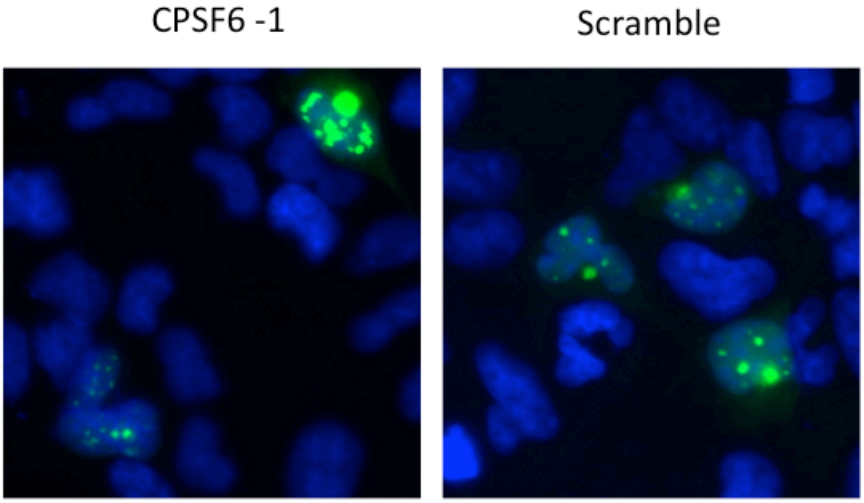


Figure 19: CPSF6 is not required for PA-X nuclear import
HEK293A cells expressing CPSF6-1 shRNA or scrambled control shRNA were transfected with a plasmid encoding PA-X(D108A)-GFP; this catalytically-inactive PA-X construct is suitable for protein trafficking studies because it translocates to the nucleus and fails to degrade mRNA. After 24 hours, cells were fixed with para-formaldehyde, permeabilized and nuclei were stained with Hoechst. Epifluorescence microscopy revealed that PA-X(D108A)-GFP (green) accumulates in puncta within the nucleus (blue; Hoechst) in CPSF6-1 shRNA and scramble control shRNA-expressing cell lines. Merged images of the green and blue channels are represented to identify location of PA-X(D108A)-GFP within the cell. Images captured on a Zeiss Axioplan II at 40X magnification by Dr. Denys Khaperskyy.

Chapter 4: Discussion

4.1 Identification of a network of PA-X interacting proteins

4.1.1 BioID reveals that PA-X interacting partners accumulate in nucleus

The precise mechanism of action of the IAV host shutoff protein PA-X remains elusive. It is known that PA-X selectively targets host RNA pol II transcripts, while sparing transcripts of RNA pol I and III, as well as viral RdRp-generated transcripts [92] . The absence of 5'-m^7G cap and poly (A) tail on RNA pol I and III transcripts may protect them against PA-X degradation. Previous studies by Khaperskyy *et al.* demonstrated that PA-X translocates to the nucleus and targets canonical 3'-end poly (A) tail containing transcripts, suggesting that PA-X substrate recognition may be linked to interactions with the mRNA 3'-end processing machinery [92] . The goal of this study was to elucidate the mechanism of action of PA-X through identification of host cell protein interactions with the X-ORF. Proteomic studies were conducted using the BioID method. In BioID, a promiscuous biotin ligase from *E. coli* is fused to a bait protein of choice, thereby guiding the enzyme to biotinylate a subset of cellular proteins. This method was chosen to identify proteins influencing PA-X host shutoff activity through either transient or direct protein-protein interactions. PA-X from two distinct strains of IAV were selected because each naturally contained either a 61 and 41 amino acid long X-ORF; A/Puerto Rico/8/1934 (H1N1) and A/California/7/2009 (H1N1), respectively. This allowed for identification of conserved protein-protein interactions between

host proteins and the X-ORF. Four basic amino acid residues in the X-ORF serve as a putative NLS for PA-X nuclear import and were therefore mutated in a separate X61 construct, to serve as a negative control. A quantitative mass spectrometry approach allowed for confident identification of X-ORF interacting proteins. A luciferase-based host shutoff functional assay indicated the effect interacting proteins had on PA-X host shutoff activity. This study was the first to confidently identify PA-X interacting host proteins and link these physical interactions to effects on PA-X host shutoff activity.

In agreement with experiments performed by Khaperskyy *et al.*, BirA*-X61 was found within the nucleus of the cell, whereas BirA*-X61(4A) was not [92] . A majority of the biotinylated interacting partners of BirA*-X61 accumulated in the same nuclear compartment as the fusion protein (Figure 9B). However, a striking accumulation of biotinylated proteins occurred in nucleoli, where the BirA*-X61 fusion protein did not accumulate (Figure 9B). This suggests that a group of cellular proteins associated with the X61 X-ORF in the 10 nm labelling radius of BirA* prior to translocation to nucleoli. The biotinylated interacting partners of BirA*-X61(4A) and the fusion protein itself had a dispersed cytoplasmic localization. BirA*-X61(4A) did not have any protein accumulation in the nucleolus (Figure 9B). This suggests the interaction with nucleolar partners observed in the X61 sample are either dependent on X-ORF nuclear localization or, require interaction with the four basic amino acids of the X-ORF.

4.1.2 Mass spectrometry analysis identified unique interacting partners

Reductive dimethylation was used to identify biotinylated X-ORF interacting proteins in a quantitative fashion. Each mass spectrometry experiment included differentially labelled biotinylated proteins from BirA*, BirA*-X-ORF (X61 or X41), and BirA*-X61(4A). Proteins identified with greater than 1 unique peptide were kept for further analysis. Proteins that met a threshold minimum relative abundance X-ORF/BirA* of greater than or equal to 1.5-fold increase were selected for further analysis. Interacting partners were deemed a protein of interest if they were biotinylated by all fusion proteins, were biotinylated by the nuclear localized X61 or X41 fusion proteins, or had known implications in mRNA 3'-end processing.

Plotting the relative abundances of the X-ORF vs. BirA* and X-ORF vs. X61(4A) visually represented the interacting partners identified through the mass spectrometry analysis. The X-ORF vs. BirA* values indicated proteins with greater specificity to the X-ORF while, the X-ORF vs. X61(4A) values indicated protein specificity of the 4 basic amino acids within the N-terminal of the X-ORF or are a result of the X-ORF nuclear translocation. However, hits identified do not imply direct binding to the X-ORF but rather indicate the protein was within the 10 nm radius of BirA* for labeling and had a freely available lysine residue.

Figures 10-12 demonstrate the specificity of interacting partners to the X-ORF and four basic residues based on their relative abundances. In each experimental condition outliers were observed. For example, the proteasome activator complex subunit 3, PSME3 (PA28γ), was identified as a distinct outlier in the X61 non-RNase treated sample (Figure 10). This hit shows great relative

abundance for the X-ORF either through nuclear localization-based interactions or interactions with the basic amino acids in the X-ORF. However, this specificity was only present under the non-RNase treatment and did not have the large relative abundances in the repeated X61 sample with the RNase treatment or the X41 sample. Together these findings suggest that the X-ORF and PSME3 interact in an RNA-dependent manner. Subsequent RNaseA treatment was applied to remaining conditions for mass spectrometry to minimize the amount of RNA-bridged interactions (Figures 11 and 12).

4.1.3 X-ORF length affects interactions with host proteins

Some candidate interacting partners were more strongly biotinylated by the BirA*-X61 fusion protein compared to the BirA*-X41, and vice versa. For example, RBM39 was highly represented in the X41 sample, but less in the X61 samples (Figure 14). The relative abundance graph in Figure 20 clearly suggests that the composition of the X41 and X61 X-ORFs affect interactions with host proteins. Such differences may ultimately inform our understanding of the increased pathogenesis of viruses bearing the shorter X41 PA-X variant [94] .

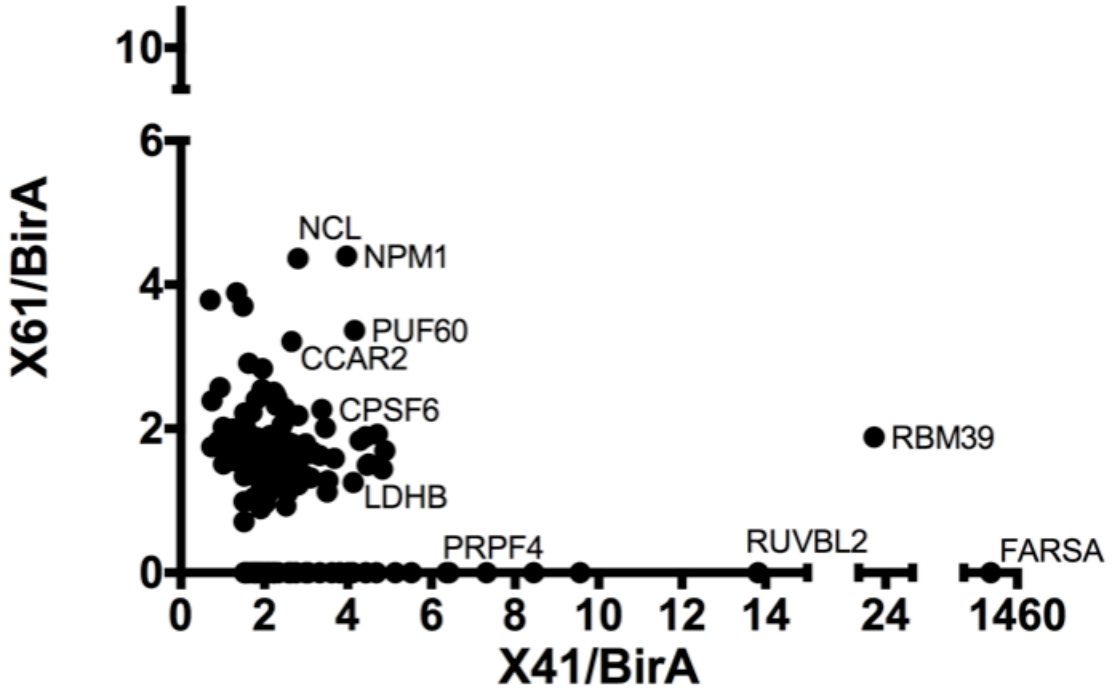


Figure 20: Selective interaction of host proteins with X-ORF variants
Comparative analysis of relative abundances of RNA-independent interactions between X-ORF variants and host proteins indicates specific interacting partners of each X-ORF. Slight variation does occur between X61 and X41 X-ORF amino acid residues.

A majority of interacting partners were clustered with relative abundances between 1.5- and 4-fold increase X-ORF vs. BirA* and X-ORF vs. X61(4A) (Figures 10-12). Within these clusters, NCL and NPM1 were two of the most prominent proteins identified. Increased biotinylation of NCL and NPM1 in RNase-treated X61 and X41 samples suggests that these RNA-binding proteins can interact with the X-ORF in an RNA-independent manner (Figure 14).

4.2 Identification of 29 high-confidence X-ORF interacting proteins

Under the criteria described in Section 3.2, the list of candidate X-ORF interacting partners was narrowed to 29 proteins. Inclusion criteria for the 29 candidate functional partners were as follows:

1) 2 or more unique peptides identified for all proteins, or
2) > 1.5 fold increase X-ORF/BirA* in 2 of 3 samples but the interacting partner appears in all 3 trials, or
3) > 2 fold increase X-ORF/BirA* with the interacting partner appearing at a minimum of 2 out of 3 samples.

The relative abundance of each of the candidate proteins under each experimental condition are depicted in the heat map in Figure 14.

The host shutoff assay was designed using a luciferase construct containing a β-globin intron. In this assay, normal PA-X activity is characterized by a ratio fold change between 0.13- and 0.5-fold normalized luciferase RLUs of PA-X to empty vector, as seen in the scramble control (Figure 16) [92] . Therefore, strong hits in the screen were defined by a ≥1-fold increase in normalized luciferase RLUs of PA-X compared to empty vector; PA-X host shutoff function was greatly inhibited with the

candidate gene silenced. Weak hits in the screen were defined by a normalized luciferase PA-X to empty vector ratio between 1 and the scramble control; indicating a slight hindrance to PA-X activity with the candidate gene silenced. The majority of identified interacting partners did not affect the host shutoff function of PA-X. This, in part, could be due to the shRNAs not fully silencing the gene of interest. However, it is also possible these X-ORF-interacting proteins were not required for PA-X host shutoff activity. Functional studies were conducted to narrow the list of candidate host proteins, and to help set priorities for future work. Eventually, it will be useful to characterize the physical interactions of PA-X with these 29 candidate X-ORF-interacting proteins, to determine whether the interactions are direct or bridged by other proteins or nucleic acids.

4.3 PA-X function greatly affected by interaction with the CFIm complex

4.3.1 CFIm complex regulates alternative sites of polyadenylation and length of 3'-UTR

Using BioID, CPSF5 and CPSF6 were identified as PA-X interacting proteins associated with mRNA processing. The two proteins make up the CFIm complex, a component of the mRNA 3'-end processing machinery [74]. The CFIm complex is a tetramer that consists of a CPSF5 (CFIm25/NUDT21) dimer and either a CPSF7 (CFIm59) dimer or CPSF6 (CFIm68) dimer [129] . The preference of the CFIm complex for CPSF6 or CPSF7 arise because of distinct interactions of the CFIm complex within the cell due to slight variations in amino acid sequence between CPSF6 and CPSF7 [129] .

The CFIm tetramer consists of two distinct domains: the RNA recognition motif and Nudix domain [130] . CFIm binds to the 3'-end of pre-mRNAs in a sequence specific manner to regulate cleavage of the poly (A) signal. Specifically, the CPSF5 domain recognizes two sequences of 5'-UGUAA-3', which permits CPSF6 looping of the RNA [131] . The length between UGUAA sequences binding by CFIm is variable, which enables the RNA looped by CPSF6 to also be of varied length [130] . Alternative polyadenylation sites are selected due the RNA looping by CPSF6, which can encompass the preferred polyadenylation signal sequence [130] . The binding of the CFIm complex to the 3'-end of pre-mRNA is crucial for recruitment of other 3'-end processing machinery [132] .

4.3.2 CFIm complex proteins are required for PA-X host shutoff activity

CPSF5 and CPSF6 were identified as individually interacting with PA-X through the BioID screen (Figure 14 and 15). PA-X has previously been shown to target RNA pol II transcripts that are processed by canonical 3'-end cleavage and polyadenylation machinery [92] . The identification of interactions between the X-ORF and CFIm complex components supports this model of PA-X function. It is possible that PA-X binding to the CFIm complex brings it in close proximity to target mRNAs for subsequent cleavage. The data collected to date suggests that CPSF6 is more strongly associated with PA-X host shutoff activity than CPSF5 (Figure 16D and 17). CPSF6 has been shown to cause looping of the mRNA for further RNA 3'-end processing [130] . It is possible PA-X requires the looping action to target and cleave the 3'-end of mRNA. Future work should directly examine sites of PA-X cleavage in relation to CFIm complex binding sites.

Interestingly, when either *CPSF5* or *CPSF6* were silenced, an up-regulation of the other gene product was observed; this was most evident for cells that expressed the CPSF5-2 shRNA (Figure 18D). This intriguing finding suggests that the cell could have a compensation mechanism to supply the sufficient levels of CPSF proteins to support mRNA biogenesis. In this experiment, only the CPSF6-1 shRNA did not cause increased CPSF5 mRNA levels (Figure 18D); instead *CPSF6* silencing caused down-regulation of CPSF5 mRNA levels, similar to the phenotype reported by Gruber *et al*. in an siRNA experiment [129] .

In addition to controlling cleavage and polyadenylation site selection, the CFIm complex also regulates splicing. The arginine and serine rich (RS) domains of CPSF6 have been shown to interact with spliceosomal SR proteins [133] . This is consistent with our emerging model for PA-X function, as RNA-sequencing experiments have demonstrated that PA-X selectively targets highly-spliced mRNAs irrespective of mRNA length (M. Gaglia, personal communication). It is possible that CFIm complex proteins help direct PA-X to target mRNAs as they are being processed, although much work remains to be done to investigate this model.

4.3.1 CPSF6 is not required for nuclear import of PA-X

CPSF6 has been shown to bind the HIV capsid protein and facilitate nuclear import of the HIV genome through interaction with the nuclear import factor TNPO3 (transportin-3) [128] . This property of CPSF6 is mapped to the RS domain [133] . A stable interaction between the HIV capsid protein and CPSF6 is mediated through hydrogen bonding [134] . Two key residues responsible for the hydrogen bonding

include, the moderately conserved Q179 and highly conserved K182 of the capsid protein [134]

The X-ORF contains conserved Q200 and K203 residues with the same relative spacing as the HIV capsid protein (Figure 3); K203 is one of the 6 highly conserved basic residues required for PA-X function, and one of the four lysine residues required for nuclear import [92,102] . I hypothesized that CPSF6 was responsible for targeting PA-X to the nucleus due to the similarity in spacing and conservation of Lys in the X-ORF and the identified interaction with CPSF6. Moreover, a failure to traffic to the nucleus would be predicted to inhibit PA-X host shutoff activity, consistent with the observations in CSPF6 silenced cell lines (Figure 16B and 17C). To test this hypothesis directly, a plasmid encoding a catalytically-inactive GFP-PA-X fusion protein was transfected into CPSF6 silenced cell lines or scramble control cells. The GFP-PA-X fusion protein accumulated in the nucleus of CPSF6-deffiecnt cells, indicating that CPSF6 is unlikely to play a role in nucleocytoplasmic trafficking of PA-X (Figure 19). Therefore, CPSF6 must be required for another aspect of PA-X hot shutoff. .

4.4 Weak candidate interacting partners identified within each node

Interacting partners from all STRING map nodes were identified to have implications in PA-X host shutoff activity (Figure 16). While much work remains to be done to validate these candidate genes, including a thorough examination of efficiency of shRNA-mediated gene silencing, some compelling candidate interacting proteins are worth further discussion in the sections that follow.

4.4.1 Miscellaneous interacting partners

Candidate proteins with no known functional association with the other identified interacting partners in the STRING map were grouped together as "miscellaneous interacting partners" (Figure 15). The majority of these proteins were found to have the greatest interactions with X41 (Figure 14). However, the BioID analysis indicated RBM39 to be the only protein in this grouping to show a higher level of interaction with X61 across all experiments (Figure 14). The luciferase-based screen indicated that RBM39 silencing inhibited PA-X host shutoff, (specifically shRNAs 2 and 3) (Figure 16F). It is possible that RBM39 could mediate localization of PA-X to the 3'-end of RNA pol II transcripts. RBM39 has been shown to be a regulator of alternative splicing and thought to be an essential component of the spliceosome via interactions with U2AF65 [135] . Furthermore, RBM39 has been show to preferentially bind the 5'- or 3'- end of pre-mRNAs [135] .

The general transcription factor II-I (GTF2I) is another candidate interacting partner of interest within this group. BioID revealed relatively low biotinylation of GTF2I by all fusion proteins. However, GTF2I had a higher relative rate of biotinylation by BirA*-X41 and GTF2I silencing prevented efficient PA-X host shutoff (Figure 16G). Only a single shRNA implicated GTF2I in PA-X host shutoff activity (GTF2I-2) (Figure 16G). GTF2I, also known as TFII-I, is a signal-dependent transcription mediator [136] . It also interacts with viral proteins including, the early region 4 (E4)– ORF 3 (E4-ORF3) of adenovirus type 5 which marks GTF2I for degradation in the proteasome through SUMOylation, enhancing its turnover rate

[137] . Like E4-ORF3, PA-X may mark GTF2I for proteasomal degradation. Consistent with this notion, the BioID screen identified some X-ORF interacting proteasome proteins, including ,PSME3, PSMD3, and PSMC5 (Figure 14).

ATP5C1, the γ-subunit of the mitochondrial ATP synthase, strongly inhibited PA-X host shutoff activity. Implications of ATP5C1 and PA-X interaction will be discussed in greater detail in Section 4.5.

4.4.2 The DNA helicase MCM5 is a candidate X-ORF interacting protein

BioID revealed that MCM5, a component of the DNA helicase complex, was selectively biotinylated by BirA*-X61 fusion proteins but not the BirA*-X41 fusion protein (Figure 14). Although its role in PA-X host shutoff activity is not defined, the luciferase assay indicated one of the shRNAs resulted in the inhibition of PA-X host shutoff activity. (Figure 16I). Interestingly, the MCM complex interacts with PA during synthesis of cRNA [138] . Because PA lacks the X-ORF, it is likely that the interaction between PA-X and MCM5 is novel. The MCM proteins have also been co-purified with protein complexes containing RNA pol II and other members of the TFII general transcription factors [139] . By interacting with MCM5, PA-X may access newly-synthesized RNA pol II transcripts.

4.4.3 X-ORF interacting protein BTF3 is required for host shutoff activity

Basic transcription factor 3 (BTF3) is a part of the nascent polypeptide associated complex (NAC) known as NAC-β [140] . It is involved in a wide variety of cellular functions including cell cycle regulations, apoptosis and mitochondrial protein targeting [140] . Most notably, BTF3 directly binds to RNA pol II and

controls transcription initiation [141] . BioID revealed that BirA*-X61 interacted with BTF3 in an RNA-dependent manner (Figure 14), and BTF3 was required for PA-X host shutoff (Figure 16D). Therefore it is likely the interaction between BTF3 and PA-X guides PA-X to nascent RNA pol II transcripts.

4.4.4 Apoptotic factors differentially modulate PA-X function

Cell cycle and apoptosis regulator protein, CCAR2 (also known as DBC1) was identified in the BioID assay to interact with BirA*-X61 and BirA*-X41 in an RNA-independent manner, irrespective of nuclear accumulation (Figure 14). PA-X host shutoff activity was only mildly inhibited in CCAR2 shRNAs expressing cells (Figure 16G). CCAR2 forms a complex with ZNF326, a ZNF-protein known to interact with nuclear mRNPs, known as DBIRD [142] . The DBIRD complex is involved in alternative splicing of mRNAs due to a bridging interaction between mRNPs and RNA pol II during transcription elongation [142] . In light of these known activities, it is possible that PA-X interaction with CCAR2 may aid targeting of elongating RNA pol II transcripts.

CCAR2 is grouped in the apoptosis node of the STRING map (Figure 15) due to its role in p53 regulation via interaction with the NAD-dependent protein deacetylase sirtuin-1 (SIRT1); in the presence of CCAR2, SIRT1 cannot deacetylate p53 [143] . Further investigations are required to determine whether CCAR2-mediated apoptosis regulation is related to PA-X cytotoxicity.

4.4.5 NCL weakly promotes PA-X host shutoff activity

Interactions between the X-ORF and nucleolar proteins were of special interest because of the observed accumulation of biotinylated proteins in the nucleolus of cells expressing the BirA*-X61 fusion protein and not in the BirA*-X61(4A) mutant (Figure 9B). Because the BirA*-X61 fusion protein itself did not accumulate in nucleoli, it was hypothesized that these nucleolar biotinylated proteins transiently interacted with the BirA*-X61 fusion protein. BioID revealed that NCL was biotinylated by all BirA*-X-ORF fusion proteins, with especially strong labeling in the BirA*-X41 and RNA-independent BirA*-X61 samples (Figure 14). However, the luciferase assay screen indicated NCL only had a moderate effect on PA-X host shutoff activity; NCL shRNAs 2 and 3 were weak hits in the screen (Figure 16H).

NCL has been previously implicated in interactions with viral endonucleases. Muller *et al.* showed that NCL binds to the 3'-UTR of IL-6 (interleukin-6) mRNA in KSHV-infected cells, resulting in the protection of IL-6 from SOX-mediated degradation [144] . NCL bound to IL-6 mRNA translocates to the cytoplasm and is able to interact with eIF4H to ensure the IL-6 translation, effectively allowing IL-6 mRNA to escape host shutoff [144] . The binding occurs at a 200-nt region in the 3'-UTR known as the SOX resistance element; a similar element was also identified in GADD45B [144,145] . However, this element did not prevent PA-X-mediated degradation of IL-6 mRNA. Thus, NCL aids selective escape of certain mRNAs from herpesviral shutoff endonucleases, but conversely, our studies to date suggest that NCL may aid PA-X host shutoff activity.

4.4.6 PA-X nuclear localization may be controlled by IPO7 and CAND1

PA-X is targeted to the nucleus because of the X-ORF, so it is not surprising that top X-ORF interacting partners included nuclear import and export machinery (Figure 14) [92] . However, none of the interacting proteins in the BioID dataset with roles in nuclear import or export had specificity for the four basic amino acid residues required for nuclear import of PA-X [92] . IPO7, importin 7, and CAND1, a regulator of the E3 ubiquitin ligase complex SCF (Skp1-Cul1-Fbox), had similar abundance patterns in the BioID assay, compared to the described pattern observed for the entire node. IPO7 bound BirA*-X61 in an RNA-dependent manner, whereas CAND1 bound in an RNA-independent manner (Figure 14). In the host shutoff assays, IPO7 shRNA-expressing cells weakly inhibited PA-X host shutoff activity (Figure 16D). CAND1 shRNA-expressing cells were not as effective as IPO7 shRNA-expressing cells at inhibiting PA-X host shutoff activity (Figure 16G).

Albeit speculative, IPO7 is a good candidate for the PA-X import factor. IPO7, originally named RanBP7, is an importin β-like protein that can mediate the direct binding and nuclear import of cellular and viral proteins [146,147] . Binding to IPO7 is not done through the classic NLS [146] . This is observed with the binding of the HIV Rev protein directly, and with greater specificity, to importin-β than to importin-α [147] . The NLS of Rev is more similar to the importin-β binding domain (IBB) of importin-α than it is to the canonical NLS [147] . Interestingly, the IBB of importin-α is made up of highly conserved basic residues [148] . Due to the basic nature of the IBB, it is probable that PA-X binds IPO7 for direct import into the

nucleus, in a manner similar to Rev. Further experimentation must be conducted to test this hypothesis.

CAND1 regulates the E3 ubiquitin ligase complex SCF (Skp1-Cul1-Fbox) via regulatory binding to Cul1 [149] . When Cul1 is bound to CAND1 it is unable to form the complex and ubiquitination of the substrate cannot occur [149] . Ubiquitination of a lysine residue regulates the function of the paramyxovirus matrix protein bipartite NLS [150] . When Lys was mutated to Arg a decrease in matrix protein ubiquitination, which resulted in nuclear retention, was observed [150] . It is possible that PA-X is modified by post-translational modifications, such as ubiquitination, and the interaction between CAND1 and PA-X regulates the level of ubiquitination of PA-X, thereby altering nucleocytoplasmic trafficking and degradation of PA-X. Ubiquitination of PA-X on X-ORF lysines could be a convenient method to promote nuclear exclusion and PA-X protein degradation. If PA-X promotes the association of CAND1 with Cul1 it would render the SCF complex unavailable to ubiquitinate substrates, and if it were under a similar control as the paramyxovirus matrix proteins, this would then promote the nuclear retention of PA-X.

4.4.7 mRNA 3'-end processing factors affect PA-X host shutoff activity

The mRNA processing node was one of two nodes containing candidate proteins with high relative abundance across most experimental conditions, with the BioID assay indicating the highest degree of specificity with BirA*-X41 (Figure 14). This is consistent with previous work from Khaperskyy *et al.* that demonstrated PA-X targeting to RNA pol II transcripts competent for 3'-end processing by the

canonical cellular machinery [92] . The U2AF65-like protein involved in alternative splicing, PUF60, was the candidate interacting protein with the highest relative abundance across all experimental conditions for this node (Figure 14) [135] . However, PUF60 was not required for PA-X host shutoff activity (Figure 16I).

PRPF4, pre-mRNA processing factor 4 (also known as snRNP60 or hPrp4), a spliceosomal component [151] was the only other protein from the mRNA processing node required for PA-X host shutoff activity. BioID revealed that PRPF4 interacts with BirA*-X61 in an RNA-dependent manner, but interactions with the BirA*-X41 fusion protein were RNA-independent, and required nuclear localization of the fusion protein (Figure 14). Additional experiments will be required to clearly determine whether X-ORF interactions with PRPF4 are bridged by RNA. Remarkably, all three PRPF4 shRNAs disrupted PA-X host shutoff activity (Figure 16D).

4.5 Mitochondrial PA-X-interacting proteins

4.5.1 X-ORF interacts with ATP synthase F_1 subunits

The identification of the γ-subunit of the ATP synthase complex, ATP5C1, as an X-ORF interacting protein was an interesting and puzzling discovery, because PA-X has never been observed in mitochondria. In the initial PA-X host shutoff activity screen, ATP5C1 silencing via any of the three shRNAs tested strongly inhibited PA-X host shutoff activity (Figure 16D). Careful inspection of BioID datasets revealed many putative interactions between the X-ORF and mitochondrial proteins, including other components of the ATP synthase complex (Table 2).

Table 2: Relative abundances of key mitochondrial interactors of PA-X

		Relative Abundance (X-ORF/BirA*)			
Protein Family	**Protein**	**X61 (-)**	**X61**	**X41**	**Average**
ATP Synthase	ATP5A1	1.25	1.47	1.42	1.38
	ATP5B	1.59	1.26	0.98	1.28
	ATP5C1	1.54	1.63	3.33	2.17
	ATP5F1	--	2.06	--	NA
	ATP5L	--	0.81	1.42	1.12
Mitochondria Translocase Complex	TOM40	--	0.93	2.4	1.67
Voltage Dependent Anion Channel	VDAC2	0.95	1.16	0.50	0.87
	VDAC3	--	1.36	2.25	1.80
Leucine Rich PPR motif containing protein, mitochondrial	LRPPRC	1.61	1.82	2.40	1.94

ATPase subunits α, β and γ (ATP5A1, ATP5B and ATP5C1 respectively) were biotinylated across all experimental conditions (Table 2). In particular, the γ-subunit was strongly enriched in all BioID experiments, followed by the α- and β-subunits (Table 2). I hypothesised that PA-X plays a role in ATP synthesis or hydrolysis during infection due the identified interactions between the α-, β- and γ-subunits of the ATPase complex. An alternative hypothesis could be that mitochondrial ATP generation may be required for efficient PA-X-mediated host shutoff. Indeed, as mentioned previously, *ATP5C1* silencing strongly inhibited host shutoff (Figure 16D). The γ-subunit of the ATPase complex is required for rotation of the synthase complex; silencing of ATP5C1 could therefore prevent the rotation the α- and β- subunits [152] . Proper positioning of ATP5C1 in the complex is critical for ATP synthesis and hydrolysis [153] . This enzyme complex is sensitive to external forces. For example, cross-linking of β- and γ-subunits can stop or slow hydrolysis, whereas external mechanical forces can implement rotation for ATP synthesis [153,154] . Therefore, it is possible PA-X influences the rotation of the ATPase complex. PA-X may also simply interfere with mitochondrial localization of this ATPase complex.

Unfortunately, the robust inhibition of PA-X host shutoff activity in *ATP5C1* silenced cells was not reproduced in the subsequent validation luciferase assays (Figure 17). This could be due to incomplete shRNA silencing in the validation assay. Additional trials will be required to determine whether ATP5C1 is a good candidate PA-X-interacting protein.

4.5.2 LRPPRC influences mRNA stability in mitochondria

LRPPRC (leucine-rich PPR motif containing protein, mitochondrial) was associated with the nuclear import and export node of the STRING map (Figure 15) and showed low-level functional interactions with the X-ORF throughout these experiments (Figure 14). PA-X host shutoff activity was weakly suppressed in cells expressing LRPPRC-1 shRNA (Figure 16G). LRPPRC binds and stabilizes mitochondrial mRNA, and has been implicated in mRNA poly (A) tail formation [155] . Interestingly, LRPPRC regulates the organization and activity of mitochondrial ATP synthase complexes [156] . LRPPRC loss prevents the ATPase complex from coupling hydrolysis and H^+ translocation [156] . Considering the relatively weak interactions between the X-ORF and LRPPRC compared to stronger interactions with ATPase components, it is possible that LRPPRC only scored in the BioID assay because it was bridged to the X-ORF fusion protein by mitochondrial ATPase subunits. Interactions between PA-X and mitochondrial mRNAs should be further studied to determine whether PA-X affects mitochondrial mRNA stability.

4.5.3 PA-X translocation into the mitochondria through TOMM40

Mitochondrial protein transport machinery was identified in preliminary BioID data sets of proteins containing 2 or more unique peptides. BioID revealed TOMM40 (translocase of the outer mitochondrial membrane) interacted with both X61 and X41 in an RNA-independent manner (Table 2); TOMM40 was specifically enriched in the X41 samples. TOMM40 is the β-barrel pore member of the translocase of the outer mitochondrial membrane responsible for bringing proteins targeted to the mitochondria into the inner membrane space [157] . No other

mitochondrial translocase proteins were identified in this study, so it remains unclear how PA-X accesses TOMM40.

Given the interaction of the X-ORF with TOMM40 and the α-, β- and γ-ATPase subunits, it is possible that PA-X is targeted to the mitochondria and interferes with cellular energy production. However, it is also possible PA-X interacts with the mitochondrial proteins in the cytoplasm, prior to their integration in the mitochondria. Further investigations are required to validate these interactions and map their intracellular location.

X-ORF interactions with mitochondrial proteins are worthy of further investigation. The pro-apoptotic IAV PB1-F2 protein is targeted to mitochondria. [158] . PB1-F2 is a small protein ranging between 87 and 90 amino acids in different IAV strains [158] . An α-helix encompassing residues 54-62, as well as Lys63 and Arg75, act as a mitochondrial import sequence for PB1-F2 [159] . Mitochondrial import is mediated via direct interaction between PB1-F2 and TOMM40, eliminating the need for interactions with the precursor proteins TOMM20 and TOMM22 [160] Inside the mitochondria, PB1-F2 disrupts mitochondria-mediated innate immune response and promotes an apoptotic response by associating with the inner mitochondrial membrane and disrupting membrane potential [160,161] . Membrane potential is dissipated due to the interaction with ANT3, adenine nucleotide translocator 3, ultimately leading to apoptosis [161] . Zamarin *et al.* speculated that PB1-F2-mediated loss of mitochondrial membrane integrity and apoptosis are promoted through complex formation with ANT3 and VDAC (adenine

nucleotide translocator 3 dependent anion channel); PB1-F2 has previously been reported to interact with VDAC1 [161] .

PB1-F2 and PA-X share interacting partners like VDAC1 and TOMM40 (Table 2). If PB1-F2 is imported directly into mitochondria via TOMM40 without the assistance of other translocase proteins, perhaps PA-X can be directly imported as well [160] . It is also possible that IAV exerts a concerted effort by PA-X and PB1-F2 to undermine mitochondrial functions and promote apoptosis during later stages of IAV infection. Further work will be required to directly investigate the interplay between PA-X and PB1-F2.

4.6 Known PA-X interacting proteins

Li *et al.* previously identified PA-X-interacting proteins via co-immunoprecipitation and tandem mass-tag mass spectrometry [162] . They identified 56 high-confidence PA-X interacting proteins, including NCL and NPM1, which were also identified in our BioID experiments. Interestingly, Li *et al.* also identified ATP5B, the β-subunit of the ATP synthase complex. Congruent findings between these two discovery methods provide strong motivation to further investigate common hits.

The McCormick lab previously conducted a yeast 2-hybrid screen to identify PA-X interacting proteins. The bait for this screen was the full-length catalytically-inactive PA-X (D108APAX), which was used to screen a human cDNA library. The short list of 22 hits from the yeast 2-hybrid screen did not overlap with the BioID hit list presented in this work, but there were some common themes. For example, different RNA-binding hnRNP proteins were identified in these screens.

4.7 Comparative Proteomic Approaches

The BioID system, first described by Roux *et al.* in 2012, uses a mutated biotin ligase from *E. coli*, (BirA) [104] . The mutation, R118G relaxes substrate constraints and causing the release of biotin-5'-AMP that can react with any available lysine residue in a 10 nm radius [110,112,113] . This approach identifies proteins that interact stably or transiently with the bait protein of choice. Compared to other approaches, BioID is quite sensitive, and does permit the identification of low-abundance interacting proteins [117] . A major limitation of BioID is the inefficient labeling of proteins with low numbers of lysines. Another limitation to this approach is the relatively long labelling times required to accumulate biotinylated proteins.

APEX, ascorbate peroxidase, is an alternative method to the BioID system. APEX was first described and used by Rhee *et al*. in 2013 where they targeted APEX to the matrix of the mitochondria and labelled proteins within it [163] . In the APEX system, hydrogen peroxide treatment activates the APEX enzyme to convert phenol derivatives into phenoxyl radicals, which are rapidly covalently attached to Tyr, Trp, His and Cys residues on nearby proteins [163] . Rhee *et al.* used a biotin-modified phenol in their experiment and pulse-labelled for only one minute [163] . This short time frame was sufficient for robust labelling in a small, confined area of the cell, enabling characterization of proteins in the mitochondrial matrix [163] . The APEX method enables dynamic mapping of the proteomes in discrete cellular compartments. In the case of IAV infection, APEX could identify differential interactions between host shutoff proteins and host proteins at different stages of

infection. APEX is also advantageous because of the ability to label more residues than just lysines. Kim and Roux highlight some drawbacks to the APEX system; for example, brief-labeling periods may prevent labeling of low-abundance or transiently interacting proteins [164] . Moreover, hydrogen peroxide and oxygen radicals elicited during APEX labeling may disrupt signalling pathways and protein structure, which may significantly alter protein labeling [164] .

Affinity purification using conventional epitope tags is commonly used to identify stable protein-protein interactions [165] . However, for an interaction to be identified, interacting proteins of the bait it must stably and tightly bind such that the interaction is not disrupted during cellular lysis [165] The Gingras group typically has used both affinity purification and the BioID system in tandem. In a study by Lambert *et al.,* BioID screens were preformed in parallel with conventional affinity purification mass spectrometry screens, which determined little overlap between these two datasets; BioID produced a richer dataset, perhaps by capturing more transient protein-protein interactions [117] .

The yeast 2-hybrid system employs reporter genes that are not activated unless the bait, which is fused to the DNA binding domain, and the prey, fused to the activating domain, come into close contact to promote reporter gene transcription [166] . Large prey cDNA libraries are often comprised of partial or full-length ORFs [166] . Yeast 2-hybrid is thought to primarily reveal direct protein-protein interactions. These screens also require both bait and prey to be translocated to the yeast cell nucleus, which may confound attempts to detect natural interactions between proteins in the cytosol and other compartments [166] .

4.8 Future directions

The molecular mechanism of action of PA-X remains to be fully elucidated, but the list candidate PA-X-interacting host proteins described in this thesis represent a starting point for future investigations. These candidate proteins must be subjected to more confirmatory experiments. For example, additional biological replicates of the luciferase-based host shutoff reporter assay must be performed, along with RT-qPCR confirmation of candidate gene silencing. Furthermore, studies of protein-protein interactions between PA-X and top-priority candidates, like the CFIm complex proteins, should be performed to determine whether these interactions are direct or indirect. Moreover, binding surfaces on both the X-ORF and host proteins must be mapped.

Many of the candidate PA-X-interacting proteins identified in this study have known roles in RNA biogenesis and processing, by employing technologies like photoactivatable ribonucleoside enhanced crosslinking and immunoprecipitation (PAR-CLIP), it should be possible to map interactions between nascent RNA pol II transcripts and the RNA-binding proteins bound by PA-X. Such an approach could aid mapping of PA-X cleavage sites, and provide insight into potential links between PA-X host shutoff and RNA splicing, cleavage and polyadenylation.

Putative interactions between PA-X and mitochondrial proteins can be further investigated via confirmatory protein-protein interaction studies as described above. It will also be useful to determine the subcellular trafficking of PA-X and bound host proteins in mitochondrial sub-compartments. As these studies progress, it may be possible to investigate effects of mitochondrial ATPase complex

activity on PA-X host shutoff, and vice versa. A concerted effort will be required to validate these protein-protein interactions in IAV infected cells. Finally, it will be interesting to determine whether PB1-F2 and PA-X collaborate to promote mitochondrial apoptosis in later stages of IAV infection.

4.9 Conclusion

In these studies, the objective was to advance mechanistic understanding of PA-X host shutoff by identifying host X-ORF-binding proteins required for efficient host shutoff. Candidate interacting proteins were identified via BioID, and functionally-validated by a luciferase-based host shutoff assay; a summary of top candidate interacting proteins identified can be found in Table 3. Consistent with recently published findings, PA-X accumulated in the nucleus and host shutoff activity appeared to depend on interactions with mRNA 3'-end processing machinery [92] , most notably the CFIm complex (Figure 21). It is not yet fully understood how PA-X translocates into the nucleus, but BioID did reveal a strong candidate import factor in IPO7 (Figure 21).

Unexpectedly, BioID identified interactions between the X-ORF and mitochondrial ATP synthase and translocase proteins, and some of these interactions appear to be important for PA-X host shutoff (Table 2). These surprising findings suggest that mitochondrial biology may be relevant to PA-X host shutoff, but it also may mean that PA-X performs distinct tasks in mitochondria during infection. Much more must be done to fully characterize interactions between PA-X and mitochondrial proteins, and to determine the functional consequences of these interactions.

Taken together, the data presented in this thesis clearly demonstrate the value of BioID as a discovery tool, and that the interactions between the X-ORF and certain host proteins are required for efficient trafficking of PA-X to the nucleus and shutoff of host gene expression. This work represents an important step forward in full elucidation of the mechanism of action of PA-X.

Table 3: Summary of candidate X-ORF interacting partners

STRING node	Candidate protein	Function in the literature	Hypothesized function with PA-X	Literature reference
mRNA processing	CPSF5	- CFIm complex - Regulates 3'-UTR length - Binds UGUA sequences	- Guiding interaction of PA-X to 3'-end cleavage site	[74,129]
	CPSF6	- CFIm complex - Regulates 3'-UTR length - Loops RNA	- Guiding interaction of PA-X to 3'-end cleavage site	[74,129,130]
	PRPF4 (hPrp4)	- Splicing	- Spliceosome interactor with PA-X	[151]
Miscellaneous	ATP5C1	- γ-subunit of the ATPase complex - ATP synthesis or hydrolysis	- Interference with cellular ATP production	[153]
	RBM39	- Alternative splicing regulator - 5'- and 3'- end binding	- Guiding interaction to 3'-end processing machinery	[135]
	GTF2I (TFII-I)	- Transcription mediator	- Guiding interaction for PA-X to targets	[136]
DNA Interactor	MCM5	- DNA helicase - Isolated with RNA pol II	- Proximity interaction to RNA pol II	[139]
Transcription/ Translation	BTF3 (NAC-β)	- Nascent polypeptide associate complex - RNA pol II direct binding	-Proximity interaction to RNA pol II	[140,141]
Apoptotic Factors	CCAR2 (DBC1)	- mRNA and RNA poll II bridging in DBIRD complex - Inhibition of p53 transcription via SIRT1 interaction	- Proximity interaction to RNA pol II	[142,143]
Nuclear Proteins	NCL	- Protecting factor of IL-6 from SOX degradation	- Possible transient nucleolar interacting partner	[144,145]
Nuclear Import/ Export	IPO7	- Importin-β like protein - Directly imports HIV Rev to the nucleus	- Nuclear import factor for PA-X	[146,147]

STRING node	Candidate protein	Function in the literature	Hypothesized function with PA-X	Literature reference
Nuclear Import/ Export	CAND1	- SCF E3 ubiquitin ligase regulator	- Regulate PA-X localization via ubiquitination	[149]
	LRPPRC	- Mitochondrial mRNA 3'-end processing - Organization of ATPase complex	- Guiding interaction to the mitochondria - Regulatory interaction with ATPase complex	[155,156]

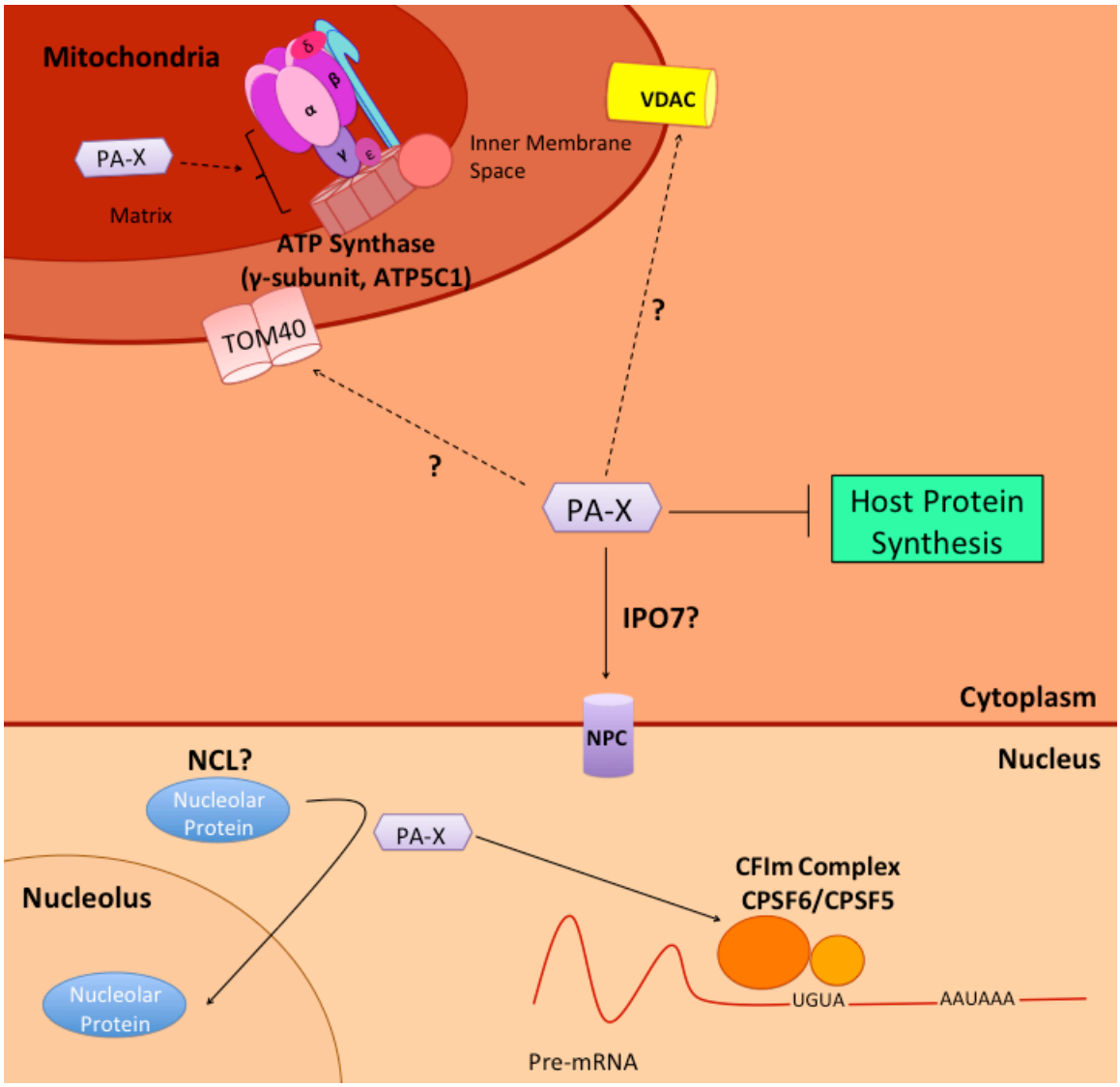


Figure 21: Proposed model of PA-X function
PA-X selectively targets host RNA pol II transcripts through interactions mediated by host proteins and its C-terminal X-ORF. PA-X is translocated to the nucleus, possibly through interactions with IPO7, where proteomic analysis identified interaction with CFIm complex proteins CPSF5 and CPSF6. This data is concordant with previous findings of PA-X targeting RNA pol II transcripts processed via the canonical 3'-end processing machinery. Transient nucleolar interactions with the X-ORF were identified, with the probable candidate being NCL. This study indicated X-ORF interaction with mitochondrial proteins, most notably the γ-subunit of the ATP synthase complex. It is not clear if these interactions occur within the mitochondria or if these interactions are prior to mitochondrial protein translocation. Solid arrows in this model indicate known pathways of PA-X either through previous studies or this thesis. Dashed arrows indicate prospective movement of PA-X based upon identified X-ORF interactions described in this thesis.

References

[1] Eisfeld AJ, Neumann G, Kawaoka Y: **At the centre: influenza A virus ribonucleoproteins.** *Nat Rev Microbiol* 2015, 13:28-41.

[2] Stencel-Baerenwald JE, Reiss K, Reiter DM, Stehle T, Dermody TS: **The sweet spot: defining virus-sialic acid interactions.** *Nat Rev Microbiol* 2014, 12:739-749.

[3] Lakadamyali M, Rust MJ, Zhuang X: **Endocytosis of influenza viruses.** *Microbes Infect* 2004, 6:929-936.

[4] Wohlbold TJ, Krammer F: **In the shadow of hemagglutinin: a growing interest in influenza viral neuraminidase and its role as a vaccine antigen.** *Viruses* 2014, 6:2465-2494.

[5] **A revision of the system of nomenclature for influenza viruses: a WHO memorandum.** *Bull World Health Organ* 1980, 58:585-591.

[6] White J, Kartenbeck J, Helenius A: **Membrane fusion activity of influenza virus.** *EMBO J* 1982, 1:217-222.

[7] Schrauwen EJ, Fouchier RA: **Host adaptation and transmission of influenza A viruses in mammals.** *Emerg Microbes Infect* 2014, 3:e9.

[8] Gottschalk A, Lind PE: **Product of interaction between influenza virus enzyme and ovomucin.** *Nature* 1949, 164:232.

[9] Haff RF, Stewart RC: **Role of Sialic Acid Receptors in Adsorption of Influenza Virus to Chick Embryo Cells.** *Journal of Immunology* 1965, 94:842-851.

[10] White J, Kartenbeck J, Helenius A: **Membrane fusion activity of influenza virus.** *EMBO J* 1982, 1:217-222.

[11] Skehel JJ, Wiley DC: **Receptor binding and membrane fusion in virus entry: the influenza hemagglutinin.** *Annu Rev Biochem* 2000, 69:531-569.

[12] Bottcher E, Matrosovich T, Beyerle M, Klenk HD, Garten W, Matrosovich M: **Proteolytic activation of influenza viruses by serine proteases TMPRSS2 and HAT from human airway epithelium.** *J Virol* 2006, 80:9896-9898.

[13] Connor RJ, Kawaoka Y, Webster RG, Paulson JC: **Receptor specificity in human, avian, and equine H2 and H3 influenza virus isolates.** *Virology* 1994, 205:17-23.

[14] Ito T, Couceiro JN, Kelm S, Baum LG, Krauss S, Castrucci MR, Donatelli I, Kida H, Paulson JC, Webster RG, Kawaoka Y: **Molecular basis for the generation in pigs of influenza A viruses with pandemic potential.** *J Virol* 1998, 72:7367-7373.

[15] Dourmashkin RR, Tyrrell DA: **Electron microscopic observations on the entry of influenza virus into susceptible cells.** *J Gen Virol* 1974, 24:129-141.

[16] Patterson S, Oxford JS, Dourmashkin RR: **Studies on the mechanism of influenza virus entry into cells.** *J Gen Virol* 1979, 43:223-229.

[17] Maeda T, Kawasaki K, Ohnishi S: **Interaction of influenza virus hemagglutinin with target membrane lipids is a key step in virus-induced hemolysis and fusion at pH 5.2.** *Proc Natl Acad Sci U S A* 1981, 78:4133-4137.

[18] Bizebard T, Gigant B, Rigolet P, Rasmussen B, Diat O, Bosecke P, Wharton SA, Skehel JJ, Knossow M: **Structure of influenza virus haemagglutinin complexed with a neutralizing antibody.** *Nature* 1995, 376:92-94.

[19] Worch R: **Structural biology of the influenza virus fusion peptide.** *Acta Biochim Pol* 2014, 61:421-426.

[20] Lee KK: **Architecture of a nascent viral fusion pore.** *EMBO J* 2010, 29:1299-1311.

[21] Wang J, Qiu JX, Soto C, DeGrado WF: **Structural and dynamic mechanisms for the function and inhibition of the M2 proton channel from influenza A virus.** *Curr Opin Struct Biol* 2011, 21:68-80.

[22] Bukrinskaya AG, Vorkunova NK, Kornilayeva GV, Narmanbetova RA, Vorkunova GK: **Influenza virus uncoating in infected cells and effect of rimantadine.** *J Gen Virol* 1982, 60:49-59.

[23] Hay AJ, Wolstenholme AJ, Skehel JJ, Smith MH: **The molecular basis of the specific anti-influenza action of amantadine.** *EMBO J* 1985, 4:3021-3024.

[24] Wang C, Takeuchi K, Pinto LH, Lamb RA: **Ion channel activity of influenza A virus M2 protein: characterization of the amantadine block.** *J Virol* 1993, 67:5585-5594.

[25] Deyde VM, Xu X, Bright RA, Shaw M, Smith CB, Zhang Y, Shu Y, Gubareva LV, Cox NJ, Klimov AI: **Surveillance of resistance to adamantanes among influenza A(H3N2) and A(H1N1) viruses isolated worldwide.** *J Infect Dis* 2007, 196:249-257.

[26] Hsu M, Parvin JD, Gupta S, Krystal M, Palese P: **Genomic RNAs of influenza viruses are held in a circular conformation in virions and in infected cells by a**

terminal panhandle. *Proceedings of the National Academy of Sciences of the United States of America* 1987, 84:8140-8144.

[27] Compans RW, Content J, Duesberg PH: **Structure of the ribonucleoprotein of influenza virus.** *J Virol* 1972, 10:795-800.

[28] Martin K, Helenius A: **Transport of incoming influenza virus nucleocapsids into the nucleus.** *J Virol* 1991, 65:232-244.

[29] Wang P, Palese P, O'Neill RE: **The NPI-1/NPI-3 (karyopherin alpha) binding site on the influenza a virus nucleoprotein NP is a nonconventional nuclear localization signal.** *J Virol* 1997, 71:1850-1856.

[30] Wu WW, Sun YH, Pante N: **Nuclear import of influenza A viral ribonucleoprotein complexes is mediated by two nuclear localization sequences on viral nucleoprotein.** *Virol J* 2007, 4:49-422X-4-49.

[31] Mark GE, Taylor JM, Broni B, Krug RM: **Nuclear Accumulation of Influenza Viral RNA Transcripts and the Effects of Cycloheximide, Actinomycin D, and a-Amanitin.** *Journal of Virology* 1979, 29:744-752.

[32] Krug RM, Broni BA, Lafiandra AJ, M, M.A., Shantkin AJ: **Priming and inhibitory activities of RNAs for the influenza viral transcriptase do not require base pairing with the virion template RNA.** *Proceedings of the National Academy of Sciences of the United States of America* 1980, 77:5874-5878.

[33] Blaas D, Patzelt E, Kuechler E: **Identification of the cap binding protein of influenza virus.** *Nucleic Acids Res* 1982, 10:4803-4812.

[34] Li ML, Ramirez BC, Krug RM: **RNA-dependent activation of primer RNA production by influenza virus polymerase: different regions of the same protein subunit constitute the two required RNA-binding sites.** *EMBO J* 1998, 17:5844-5852.

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

[36] Poon LLM, Pritlove DC, Fodor E, Brownlee GG: **Direct Evidence that the Poly(A) Tail of Influenza A Virus mRNA Is Synthesized by Reiterative Copying of a U Track in the Virion RNA Template.** *Journal of Virology* 1999, 73:3473-3476.

[37] Lamb RA, Lai CJ, Choppin PW: **Sequences of mRNAs derived from genome RNA segment 7 of influenza virus: colinear and interrupted mRNAs code for overlapping proteins.** *Proc Natl Acad Sci U S A* 1981, 78:4170-4174.

[38] Smith DB, Inglis SC: **Regulated production of an influenza virus spliced mRNA mediated by virus-specific products.** *EMBO J* 1985, 4:2313-2319.

[39] Bier K, York A, Fodor E: **Cellular cap-binding proteins associate with influenza virus mRNAs.** *J Gen Virol* 2011, 92:1627-1634.

[40] Larsen S, Bui S, Perez V, Mohammad A, Medina-Ramirez H, Newcomb LL: **Influenza polymerase encoding mRNAs utilize atypical mRNA nuclear export.** *Virol J* 2014, 11:154-422X-11-154.

[41] Yanguez E, Nieto A: **So similar, yet so different: selective translation of capped and polyadenylated viral mRNAs in the influenza virus infected cell.** *Virus Res* 2011, 156:1-12.

[42] de la Luna S, Fortes P, Beloso A, Ortin J: **Influenza virus NS1 protein enhances the rate of translation initiation of viral mRNAs.** *J Virol* 1995, 69:2427-2433.

[43] Enami K, Sato TA, Nakada S, Enami M: **Influenza virus NS1 protein stimulates translation of the M1 protein.** *J Virol* 1994, 68:1432-1437.

[44] Burgui I, Aragon T, Ortin J, Nieto A: **PABP1 and eIF4GI associate with influenza virus NS1 protein in viral mRNA translation initiation complexes.** *J Gen Virol* 2003, 84:3263-3274.

[45] York A, Hengrung N, Vreede FT, Huiskonen JT, Fodor E: **Isolation and characterization of the positive-sense replicative intermediate of a negative-strand RNA virus.** *Proc Natl Acad Sci U S A* 2013, 110:E4238-45.

[46] Vreede FT, Brownlee GG: **Influenza virion-derived viral ribonucleoproteins synthesize both mRNA and cRNA in vitro.** *J Virol* 2007, 81:2196-2204.

[47] Olson AC, Rosenblum E, Kuchta RD: **Regulation of influenza RNA polymerase activity and the switch between replication and transcription by the concentrations of the vRNA 5' end, the cap source, and the polymerase.** *Biochemistry* 2010, 49:10208-10215.

[48] Newcomb LL, Kuo RL, Ye Q, Jiang Y, Tao YJ, Krug RM: **Interaction of the influenza a virus nucleocapsid protein with the viral RNA polymerase potentiates unprimed viral RNA replication.** *J Virol* 2009, 83:29-36.

[49] Perez JT, Zlatev I, Aggarwal S, Subramanian S, Sachidanandam R, Kim B, Manoharan M, tenOever BR: **A small-RNA enhancer of viral polymerase activity.** *J Virol* 2012, 86:13475-13485.

[50] O'Neill RE, Talon J, Palese P: **The influenza virus NEP (NS2 protein) mediates the nuclear export of viral ribonucleoproteins.** *EMBO J* 1998, 17:288-296.

[51] Amorim MJ, Bruce EA, Read EK, Foeglein A, Mahen R, Stuart AD, Digard P: **A Rab11- and microtubule-dependent mechanism for cytoplasmic transport of influenza A virus viral RNA.** *J Virol* 2011, 85:4143-4156.

[52] Sansom MS, Kerr ID, Smith GR, Son HS: **The influenza A virus M2 channel: a molecular modeling and simulation study.** *Virology* 1997, 233:163-173.

[53] McCown MF, Pekosz A: **The influenza A virus M2 cytoplasmic tail is required for infectious virus production and efficient genome packaging.** *J Virol* 2005, 79:3595-3605.

[54] Bos TJ, Davis AR, Nayak DP: **NH2-terminal hydrophobic region of influenza virus neuraminidase provides the signal function in translocation.** *Proc Natl Acad Sci U S A* 1984, 81:2327-2331.

[55] Tate MD, Job ER, Deng YM, Gunalan V, Maurer-Stroh S, Reading PC: **Playing hide and seek: how glycosylation of the influenza virus hemagglutinin can modulate the immune response to infection.** *Viruses* 2014, 6:1294-1316.

[56] Ali A, Avalos RT, Ponimaskin E, Nayak DP: **Influenza virus assembly: effect of influenza virus glycoproteins on the membrane association of M1 protein.** *J Virol* 2000, 74:8709-8719.

[57] Scheiffele P, Roth MG, Simons K: **Interaction of influenza virus haemagglutinin with sphingolipid-cholesterol membrane domains via its transmembrane domain.** *EMBO J* 1997, 16:5501-5508.

[58] Barman S, Ali A, Hui EK, Adhikary L, Nayak DP: **Transport of viral proteins to the apical membranes and interaction of matrix protein with glycoproteins in the assembly of influenza viruses.** *Virus Res* 2001, 77:61-69.

[59] Rossman JS, Jing X, Leser GP, Lamb RA: **Influenza virus M2 protein mediates ESCRT-independent membrane scission.** *Cell* 2010, 142:902-913.

[60] Chlanda P, Mekhedov E, Waters H, Sodt A, Schwartz C, Nair V, Blank PS, Zimmerberg J: **Palmitoylation Contributes to Membrane Curvature in Influenza A Virus Assembly and Hemagglutinin-Mediated Membrane Fusion.** *J Virol* 2017, 91:10.1128/JVI.00947-17. Print 2017 Nov 1.

[61] Rossman JS, Lamb RA: **Influenza virus assembly and budding.** *Virology* 2011, 411:229-236.

[62] Barman S, Adhikary L, Chakrabarti AK, Bernas C, Kawaoka Y, Nayak DP: **Role of transmembrane domain and cytoplasmic tail amino acid sequences of influenza a virus neuraminidase in raft association and virus budding.** *J Virol* 2004, 78:5258-5269.

[63] Neumann G, Noda T, Kawaoka Y: **Emergence and pandemic potential of swine-origin H1N1 influenza virus.** *Nature* 2009, 459:931-939.

[64] Dias A, Bouvier D, Crepin T, McCarthy AA, Hart DJ, Baudin F, Cusack S, Ruigrok RW: **The cap-snatching endonuclease of influenza virus polymerase resides in the PA subunit.** *Nature* 2009, 458:914-918.

[65] Yuan P, Bartlam M, Lou Z, Chen S, Zhou J, He X, Lv Z, Ge R, Li X, Deng T *et al.*: **Crystal structure of an avian influenza polymerase PA(N) reveals an endonuclease active site.** *Nature* 2009, 458:909-913.

[66] Hara K, Schmidt FI, Crow M, Brownlee GG: **Amino acid residues in the N-terminal region of the PA subunit of influenza A virus RNA polymerase play a critical role in protein stability, endonuclease activity, cap binding, and virion RNA promoter binding.** *J Virol* 2006, 80:7789-7798.

[67] Engelhardt OG, Smith M, Fodor E: **Association of the influenza A virus RNA-dependent RNA polymerase with cellular RNA polymerase II.** *J Virol* 2005, 79:5812-5818.

[68] Khaperskyy DA, McCormick C: **Timing Is Everything: Coordinated Control of Host Shutoff by Influenza A Virus NS1 and PA-X Proteins.** *J Virol* 2015, 89:6528-6531.

[69] Lin D, Lan J, Zhang Z: **Structure and function of the NS1 protein of influenza A virus.** *Acta Biochim Biophys Sin (Shanghai)* 2007, 39:155-162.

[70] Clancy S: **RNA splicing: introns, exons and spliceosome.** *Nature Education* 2008, 1:31.

[71] Matera AG, Wang Z: **A day in the life of the spliceosome.** *Nat Rev Mol Cell Biol* 2014, 15:108-121.

[72] Fortes P, Lamond AI, Ortin J: **Influenza virus NS1 protein alters the subnuclear localization of cellular splicing components.** *J Gen Virol* 1995, 76 (Pt 4):1001-1007.

[73] Qiu Y, Nemeroff M, Krug RM: **The influenza virus NS1 protein binds to a specific region in human U6 snRNA and inhibits U6-U2 and U6-U4 snRNA interactions during splicing.** *RNA* 1995, 1:304-316.

[74] Tian B, Graber JH: **Signals for pre-mRNA cleavage and polyadenylation.** *Wiley Interdiscip Rev RNA* 2012, 3:385-396.

[75] Nemerof ME, Barabino SML, Li Y, Keller W, Krug RM: **Influenza Virus NS1 Protein Interacts with the Cellular 30 kDa Subunit of CPSF and Inhibits 3′ End Formation of Cellular Pre-mRNAs.** *Molecular Cell* 1998, 1:991-1000.

[76] Chen Z, Li Y, Krug RM: **Influenza A virus NS1 protein targetspoly(A)-binding protein II of the cellular 3′-end processing machinery.** *The EMBO journal* 1999, 18:2273-2283.

[77] Wickramasinghe VO, Laskey RA: **Control of mammalian gene expression by selective mRNA export.** *Nat Rev Mol Cell Biol* 2015, 16:431-442.

[78] Satterly N, Tsai PL, van Deursen J, Nussenzveig DR, Wang Y, Faria PA, Levay A, Levy DE, Fontoura BM: **Influenza virus targets the mRNA export machinery and the nuclear pore complex.** *Proc Natl Acad Sci U S A* 2007, 104:1853-1858.

[79] Hatada E, Saito S, Fukuda R: **Mutant Influenza Viruses with a Defective NS1 Protein Cannot Block the Activation of PKR in Infected Cells.** *Journal of virology* 1999, 73:2425-2433.

[80] Hale BG, Randall RE, Ortin J, Jackson D: **The multifunctional NS1 protein of influenza A viruses.** *J Gen Virol* 2008, 89:2359-2376.

[81] Dauber B, Wolff T: **Activation of the Antiviral Kinase PKR and Viral Countermeasures.** *Viruses* 2009, 1:523-544.

[82] Nanduri S, Rahman F, Williams BRG, Qin J: **Characterization of the Solution Complex between the Interferoninduced, Double-stranded RNA-activated Protein Kinase and HIV-I Trans-activating Region RNA.** *The EMBO journal* 2000, 19:5567-5574.

[83] Carpick BW, Graziano V, Schneider D, Maitra RK, Lee X, Williams BRG: **Characterization of the Solution Complex between the Interferoninduced, Double-stranded RNA-activated Protein Kinase and HIV-I Trans-activating Region RNA.** *Journal of Biological Chemistry* 1997, 272:9510-9516.

[84] Kapp LD, Lorsch JR: **The molecular mechanics of eukaryotic translation.** *Annu Rev Biochem* 2004, 73:657-704.

[85] Holcik M, Sonenberg N: **Translational control in stress and apoptosis.** *Nat Rev Mol Cell Biol* 2005, 6:318-327.

[86] Li S, Min J, Krug RM, Sen GC: **Binding of the influenza A virus NS1 protein to PKR mediates the inhibition of its activation by either PACT or double-stranded RNA.** *Virology* 2006, 349:13-21.

[87] Jagger BW, Wise HM, Kash JC, Walters KA, Wills NM, Xiao YL, Dunfee RL, Schwartzman LM, Ozinsky A, Bell GL *et al.*: **An overlapping protein-coding region in influenza A virus segment 3 modulates the host response.** *Science* 2012, 337:199-204.

[88] Gog JR, Afonso Edos S, Dalton RM, Leclercq I, Tiley L, Elton D, von Kirchbach JC, Naffakh N, Escriou N, Digard P: **Codon conservation in the influenza A virus genome defines RNA packaging signals.** *Nucleic Acids Res* 2007, 35:1897-1907.

[89] Eisinger J, Feuer B, Yamane T: **Codon-anticodon binding in tRNAphe.** *Nat New Biol* 1971, 231:126-128.

[90] Shi M, Jagger BW, Wise HM, Digard P, Holmes EC, Taubenberger JK: **Evolutionary conservation of the PA-X open reading frame in segment 3 of influenza A virus.** *J Virol* 2012, 86:12411-12413.

[91] Oishi K, Yamayoshi S, Kawaoka Y: **Mapping of a Region of the PA-X Protein of Influenza A Virus That Is Important for Its Shutoff Activity.** *J Virol* 2015, 89:8661-8665.

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

[93] Khaperskyy DA, Emara MM, Johnston BP, Anderson P, Hatchette TF, McCormick C: **Influenza a virus host shutoff disables antiviral stress-induced translation arrest.** *PLoS Pathog* 2014, 10:e1004217.

[94] Xu G, Zhang X, Sun Y, Liu Q, Sun H, Xiong X, Jiang M, He Q, Wang Y, Pu J *et al.*: **Truncation of C-terminal 20 amino acids in PA-X contributes to adaptation of swine influenza virus in pigs.** *Sci Rep* 2016, 6:21845.

[95] Gao H, Sun Y, Hu J, Qi L, Wang J, Xiong X, Wang Y, He Q, Lin Y, Kong W *et al.*: **The contribution of PA-X to the virulence of pandemic 2009 H1N1 and highly pathogenic H5N1 avian influenza viruses.** *Sci Rep* 2015, 5:8262.

[96] Gao H, Xu G, Sun Y, Qi L, Wang J, Kong W, Sun H, Pu J, Chang KC, Liu J: **PA-X is a virulence factor in avian H9N2 influenza virus.** *J Gen Virol* 2015, 96:2587-2594.

[97] Hayashi T, MacDonald LA, Takimoto T: **Influenza A Virus Protein PA-X Contributes to Viral Growth and Suppression of the Host Antiviral and Immune Responses.** *J Virol* 2015, 89:6442-6452.

[98] Lee J, Yu H, Li Y, Ma J, Lang Y, Duff M, Henningson J, Liu Q, Li Y, Nagy A *et al.*: **Impacts of different expressions of PA-X protein on 2009 pandemic H1N1 virus replication, pathogenicity and host immune responses.** *Virology* 2017, 504:25-35.

[99] Rivas HG, Schmaling SK, Gaglia MM: **Shutoff of Host Gene Expression in Influenza A Virus and Herpesviruses: Similar Mechanisms and Common Themes.** *Viruses* 2016, 8:102.

[100] Gaglia MM, Rycroft CH, Glaunsinger BA: **Transcriptome-Wide Cleavage Site Mapping on Cellular mRNAs Reveals Features Underlying Sequence-Specific Cleavage by the Viral Ribonuclease SOX.** *PLoS Pathog* 2015, 11:e1005305.

[101] Desmet EA, Bussey KA, Stone R, Takimoto T: **Identification of the N-terminal domain of the influenza virus PA responsible for the suppression of host protein synthesis.** *J Virol* 2013, 87:3108-3118.

[102] Oishi K, Yamayoshi S, Kawaoka Y: **Mapping of a Region of the PA-X Protein of Influenza A Virus That Is Important for Its Shutoff Activity.** *J Virol* 2015, 89:8661-8665.

[103] Bavagnoli L, Cucuzza S, Campanini G, Rovida F, Paolucci S, Baldanti F, Maga G: **The novel influenza A virus protein PA-X and its naturally deleted variant show different enzymatic properties in comparison to the viral endonuclease PA.** *Nucleic Acids Res* 2015, 43:9405-9417.

[104] Roux KJ, Kim DI, Raida M, Burke B: **A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells.** *J Cell Biol* 2012, 196:801-810.

[105] Eisenberg MA, Prakash O, Hsiung SC: **Purification and properties of the biotin repressor. A bifunctional protein.** *J Biol Chem* 1982, 257:15167-15173.

[106] Chapman-Smith A, Cronan JE: **Molecular Biology of Biotin Attachment to Proteins.** *The Journal of Nutrition* 1999, 129:477S-484S.

[107] Chapman-Smith A, Cronan JE: **The enzymatic biotinylation of proteins: a post-translational modification of exceptional specificity.** *Trends in Biochemical Sciences* 1999, 24:359-363.

[108] Kosow DP, Huang SC, Lane MD: **Propionyl holocarboxylase synthesis. I. Preparation and properties of the enzyme system.** *J Biol Chem* 1962, 237:3633-3639.

[109] Lane MD, Rominger KL, Young DL, Lynen F: **The Enzymatic Synthesis of Holotranscarboxylase from Apotranscarboxylase and (+)-Biotin. Ii. Investigation of the Reaction Mechanism.** *J Biol Chem* 1964, 239:2865-2871.

[110] Kwon K, Beckett D: **Function of a conserved sequence motif in biotin holoenzyme synthetases.** *Protein Science* 2000, 9:1530-1539.

[111] Choi-Rhee E, Schulman H, Cronan JE: **Promiscuous protein biotinylation by Escherichia coli biotin protein ligase.** *Protein Sci* 2004, 13:3043-3050.

[112] Streaker ED, Beckett D: **Nonenzymatic biotinylation of a biotin carboxyl carrier protein: unusual reactivity of the physiological target lysine.** *Protein Sci* 2006, 15:1928-1935.

[113] Kim DI, Birendra KC, Zhu W, Motamedchaboki K, Doye V, Roux KJ: **Probing nuclear pore complex architecture with proximity-dependent biotinylation.** *Proc Natl Acad Sci U S A* 2014, 111:E2453-61.

[114] Morriswood B, Havlicek K, Demmel L, Yavuz S, Sealey-Cardona M, Vidilaseris K, Anrather D, Kostan J, Djinovic-Carugo K, Roux KJ, Warren G: **Novel bilobe components in Trypanosoma brucei identified using proximity-dependent biotinylation.** *Eukaryot Cell* 2013, 12:356-367.

[115] Firat-Karalar EN, Rauniyar N, Yates JR,3rd, Stearns T: **Proximity interactions among centrosome components identify regulators of centriole duplication.** *Curr Biol* 2014, 24:664-670.

[116] Couzens AL, Knight JD, Kean MJ, Teo G, Weiss A, Dunham WH, Lin ZY, Bagshaw RD, Sicheri F, Pawson T *et al.*: **Protein interaction network of the mammalian Hippo pathway reveals mechanisms of kinase-phosphatase interactions.** *Sci Signal* 2013, 6:rs15.

[117] Lambert JP, Tucholska M, Go C, Knight JD, Gingras AC: **Proximity biotinylation and affinity purification are complementary approaches for the interactome mapping of chromatin-associated protein complexes.** *J Proteomics* 2015, 118:81-94.

[118] Gupta GD, Coyaud E, Goncalves J, Mojarad BA, Liu Y, Wu Q, Gheiratmand L, Comartin D, Tkach JM, Cheung SW *et al.*: **A Dynamic Protein Interaction Landscape of the Human Centrosome-Cilium Interface.** *Cell* 2015, 163:1484-1499.

[119] Coyaud E, Mis M, Laurent EM, Dunham WH, Couzens AL, Robitaille M, Gingras AC, Angers S, Raught B: **BioID-based Identification of Skp Cullin F-box (SCF)beta-TrCP1/2 E3 Ligase Substrates.** *Mol Cell Proteomics* 2015, 14:1781-1795.

[120] Uuskula-Reimand L, Hou H, Samavarchi-Tehrani P, Rudan MV, Liang M, Medina-Rivera A, Mohammed H, Schmidt D, Schwalie P, Young EJ *et al.*: **Topoisomerase II beta interacts with cohesin and CTCF at topological domain borders.** *Genome Biol* 2016, 17:182-016-1043-8.

[121] Johnston WL, Krizus A, Ramani AK, Dunham W, Youn JY, Fraser AG, Gingras AC, Dennis JW: **C. elegans SUP-46, an HNRNPM family RNA-binding protein that prevents paternally-mediated epigenetic sterility.** *BMC Biol* 2017, 15:61-017-0398-y.

[122] Gu B, Lambert JP, Cockburn K, Gingras AC, Rossant J: **AIRE is a critical spindle-associated protein in embryonic stem cells.** *Elife* 2017, 6:10.7554/eLife.28131.

[123] Khaperskyy DA, Hatchette TF, McCormick C: **Influenza A virus inhibits cytoplasmic stress granule formation.** *FASEB J* 2012, 26:1629-1639.

[124] Hsu JL, Huang SY, Chow NH, Chen SH: **Stable-isotope dimethyl labeling for quantitative proteomics.** *Anal Chem* 2003, 75:6843-6852.

[125] Bioinformatics and Evolutionary Genomics: **Calculate and draw ustom Venn diagrams.** 2017.

[126] **Morpheus.** 2017.

[127] Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP *et al.*: **STRING v10: protein-protein interaction networks, integrated over the tree of life.** *Nucleic Acids Res* 2015, 43:D447-52.

[128] Price AJ, Fletcher AJ, Schaller T, Elliott T, Lee K, KewalRamani VN, Chin JW, Towers GJ, James LC: **CPSF6 Defines a Conserved Capsid Interface that Modulates HIV-1 Replication.** *PLoS Pathogens* 2012, 8:e1002896.

[129] Gruber AR, Martin G, Keller W, Zavolan M: **Cleavage factor Im is a key regulator of 3′ UTR length.** *RNA Biology* 2012, 9:1405-1412.

[130] Yang Q, Coseno M, Gilmartin GM, Doublie´ S: **Crystal Structure of a Human Cleavage Factor CFIm25/CFIm68/RNA Complex Provides an Insight into Poly(A) Site Recognition and RNA Looping.** *Structure* 2011, 19:368-377.

[131] Brown KM, Gilmartin GM: **A mechanism for the regulation of pre-mRNA 3' processing by human cleavage factor Im.** *Mol Cell* 2003, 12:1467-1476.

[132] Venkataraman K, Brown KM, Gilmartin GM: **Analysis of a noncanonical poly(A) site reveals a tripartite mechanism for vertebrate poly(A) site recognition.** *Genes Dev* 2005, 19:1315-1327.

[133] Dettwiler S, Aringhieri C, Cardinale S, Keller W, Barabino SM: **Distinct sequence motifs within the 68-kDa subunit of cleavage factor Im mediate RNA binding, protein-protein interactions, and subcellular localization.** *J Biol Chem* 2004, 279:35788-35797.

[134] Bhattacharya A, Alam SL, Fricke T, Zadrozny K, Sedzicki J, Taylor AB, Demeler B, Pornillos O, Ganser-Pornillos BK, Diaz-Griffero F *et al.*: **Structural basis of HIV-1 capsid recognition by PF74 and CPSF6.** *Proc Natl Acad Sci U S A* 2014, 111:18625-18630.

[135] Mai S, Qu X, Li P, Ma Q, Cao C, Liu X: **Global regulation of alternative RNA splicing by the SR-rich protein RBM39.** *Biochimica et Biophysica Acta* 2016, 1859:1014-1024.

[136] Roy AL: **Biochemistry and biology of the inducible multifunctional transcription factor TFII-I: 10 years later.** *Gene* 2012, 492:32-41.

[137] Bridges RG, Sohn SY, Wright J, Leppard KN, Hearing P: **The Adenovirus E4-ORF3 Protein Stimulates SUMOylation of General Transcription Factor TFII-I to Direct Proteasomal Degradation.** *MBio* 2016, 7:e02184-15.

[138] Kawaguchi A, Nagata K: **De novo replication of the influenza virus RNA genome is regulated by DNA replicative helicase, MCM.** *EMBO J* 2007, 26:4566-4575.

[139] Yankulov K, Todorov I, Romanowski P, Licatalosi D, Cilli K, McCracken S, Laskey R, Bentley DL: **MCM proteins are associated with RNA polymerase II holoenzyme.** *Mol Cell Biol* 1999, 19:6154-6163.

[140] Jamil M, Wang W, Xu M, Tu J: **Exploring the roles of basal transcription factor 3 in eukaryotic growth and development.** *Biotechnol Genet Eng Rev* 2015, 31:21-45.

[141] Zheng XM, Black D, Chambon P, Egly JM: **Sequencing and expression of complementary DNA for the general transcription factor BTF3.** *Nature* 1990, 344:556-559.

[142] Close P, East P, Dirac-Svejstrup AB, Hartmann H, Heron M, Maslen S, Chariot A, Soding J, Skehel M, Svejstrup JQ: **DBIRD complex integrates alternative mRNA splicing with RNA polymerase II transcript elongation.** *Nature* 2012, 484:386-389.

[143] Zhao W, Kruse J, Tang Y, Jung SY, Qin J, Gu W: **Negative regulation of the deacetylase SIRT1 by DBC1.** *Nature* 2008, 451:587-590.

[144] Muller M, Hutin S, Marigold O, Li KH, Burlingame A, Glaunsinger BA: **A ribonucleoprotein complex protects the interleukin-6 mRNA from degradation by distinct herpesviral endonucleases.** *PLoS Pathog* 2015, 11:e1004899.

[145] Muller M, Glaunsinger BA: **Nuclease escape elements protect messenger RNA against cleavage by multiple viral endonucleases.** *PLoS Pathog* 2017, 13:e1006593.

[146] Jakel S, Gorlich D: **Importin beta, transportin, RanBP5 and RanBP7 mediate nuclear import of ribosomal proteins in mammalian cells.** *EMBO J* 1998, 17:4491-4502.

[147] Henderson BR, Percipalle P: **Interactions between HIV Rev and Nuclear Import and Export Factors: The Rev Nuclear Localisation Signal Mediates Specific Binding to Human Importin-beta.** *Journal of Molecular Biology* 1997, 274:693-707.

[148] Cingolani G, Petosa C, Weis K, Mu¨ller CW: **Structure of importin-beta bound to the IBB domain of importin-a.** *Nature* 1999, 399:221-229.

[149] Lee EK, Diehl JA: **SCFs in the new millennium.** *Oncogene* 2014, 33:2011-2018.

[150] Pentecost M, Vashisht AA, Lester T, Voros T, Beaty SM, Park A, Wang YE, Yun TE, Freiberg AN, Wohlschlegel JA, Lee B: **Evidence for ubiquitin-regulated nuclear and subnuclear trafficking among Paramyxovirinae matrix proteins.** *PLoS Pathog* 2015, 11:e1004739.

[151] Will CL, Luhrmann R: **Spliceosome structure and function.** *Cold Spring Harb Perspect Biol* 2011, 3:10.1101/cshperspect.a003707.

[152] Nakanishi-Matsui M, Sekiya M, Nakamoto RK, Futai M: **The mechanism of rotating proton pumping ATPases.** *Biochim Biophys Acta* 2010, 1797:1343-1352.

[153] Itoh H, Takahashi A, Adachi K, Noji H, Yasuda R, Yoshida M, Kinosita K: **Mechanically driven ATP synthesis by F1-ATPase.** *Nature* 2004, 427:465-468.

[154] Baylis Scanlon JA, Al-Shawi MK, Nakamoto RK: **A Rotor-Stator Cross-link in the F1-ATPase Blocks the Rate-limiting Step of Rotational Catalysis.** *Journal of Biological Chemistry* 2008, 283:26228-26240.

[155] Ruzzenente B, Metodiev MD, Wredenberg A, Bratic A, Park CB, Camara Y, Milenkovic D, Zickermann V, Wibom R, Hultenby K *et al.*: **LRPPRC is necessary for polyadenylation and coordination of translation of mitochondrial mRNAs.** *EMBO J* 2012, 31:443-456.

[156] Mourier A, Ruzzenente B, Brandt T, Kuhlbrandt W, Larsson NG: **Loss of LRPPRC causes ATP synthase deficiency.** *Hum Mol Genet* 2014, 23:2580-2592.

[157] Wiedemann N, Frazier AE, Pfanner N: **The protein import machinery of mitochondria.** *J Biol Chem* 2004, 279:14473-14476.

[158] Chakrabarti AK, Pasricha G: **An insight into the PB1F2 protein and its multifunctional role in enhancing the pathogenicity of the influenza A viruses.** *Virology* 2013, 440:97-104.

[159] Yamada H, Chounan R, Higashi Y, Kurihara N, Kido H: **Mitochondrial targeting sequence of the influenza A virus PB1-F2 protein and its function in mitochondria.** *FEBS Lett* 2004, 578:331-336.

[160] Yoshizumi T, Ichinohe T, Sasaki O, Otera H, Kawabata S, Mihara K, Koshiba T: **Influenza A virus protein PB1-F2 translocates into mitochondria via Tom40 channels and impairs innate immunity.** *Nat Commun* 2014, 5:4713.

[161] Zamarin D, Garcia-Sastre A, Xiao X, Wang R, Palese P: **Influenza virus PB1-F2 protein induces cell death through mitochondrial ANT3 and VDAC1.** *PLoS Pathog* 2005, 1:e4.

[162] Li Q, Yuan X, Wang Q, Chang G, Wang F, Liu R, Zheng M, Chen G, Wen J, Zhao G: **Interactomic landscape of PA-X-chicken protein complexes of H5N1 influenza A virus.** *Journal of Proteomics* 2016, 148:20-25.

[163] Rhee HW, Zou P, Udeshi ND, Martell JD, Mootha VK, Carr SA, Ting AY: **Proteomic mapping of mitochondria in living cells via spatially restricted enzymatic tagging.** *Science* 2013, 339:1328-1331.

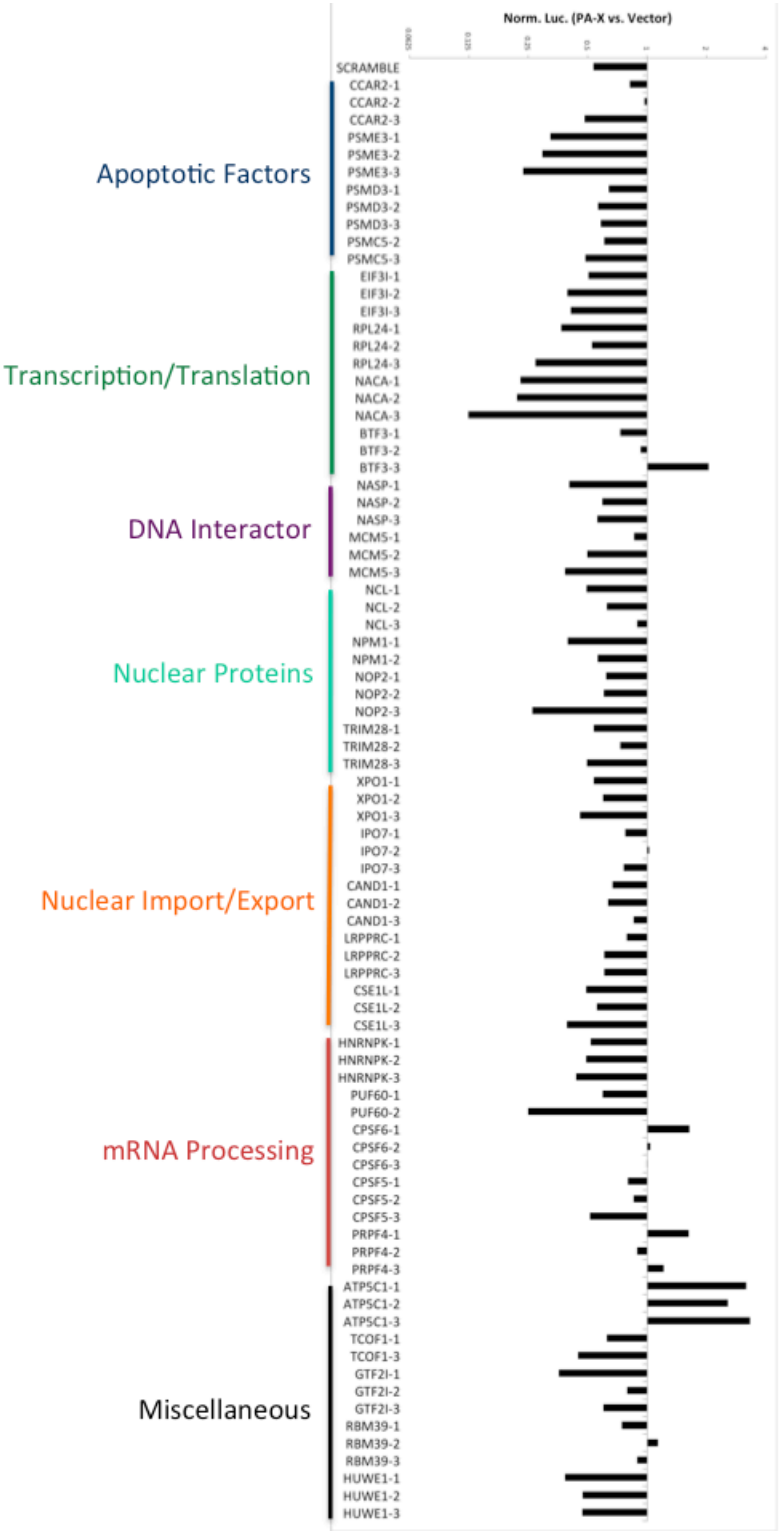
[164] Kim DI, Roux KJ: **Filling the Void: Proximity-Based Labeling of Proteins in Living Cells.** *Trends Cell Biol* 2016, 26:804-817.

[165] Morris JH, Knudsen GM, Verschueren E, Johnson JR, Cimermancic P, Greninger AL, Pico AR: **Affinity purification-mass spectrometry and network analysis to understand protein-protein interactions.** *Nat Protoc* 2014, 9:2539-2554.

[166] Brückner A, Polge C, Lentze N, Auerbach D, Schlattner U: **Yeast Two-Hybrid, a Powerful Tool for Systems Biology.** *International Journal of Molecular Sciences* 2009, 10:2763-2788.

Appendix

Appendix A1



Appendix A1: Grouped data from luciferase screen

Appendix A1: Grouped data from luciferase screen

All collected luciferase assay data from Figure 16 grouped into STRING map assigned groupings. shRNA expressing cell lines of the list of 29 interacting proteins were transfected with PA-X and FF-luciferase containing a β-globin intron or empty and FF-luciferase containing β-globin intron for 24 hours. Data is normalized to the luciferase reading of the empty vector control for each of the cell lines and graphed on a $\log_2$ scale. All luciferase screens were done in technical triplicates. The scramble control is the average of all scramble controls from Figure 16, and was within an acceptable range. Proteins that had a fold change between 0.5 and 1 were deemed weak candidates, while those around 1 or above the line were deemed to be strong candidates. Proteins that were below the scramble control were considered inhibiting candidates.