Electronic Thesis and Dissertation Repository

3-23-2021 11:00 AM

Characterization and Discovery of Short Linear Motifs Mediating Protein Nuclear Import

Tanner M. Tessier, *The University of Western Ontario*

Supervisor: Mymryk, Joe S., *The University of Western Ontario*
A thesis submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree in Microbiology and Immunology
© Tanner M. Tessier 2021

Recommended Citation

Abstract

Protein-protein interactions (PPI) mediated through short linear motifs (SLiMs) are ubiquitous throughout the human proteome and are involved in many essential cellular processes. One such type of SLiM is the classical nuclear localization sequence (cNLS), which facilitates nuclear import by binding importin-α (Imp-α). This pathway is indispensable to many cellular processes and is extensively used by viral proteins that function within the nucleus of infected cells. Based on this, I demonstrated that the classical nuclear import pathway inhibitor, ivermectin, can inhibit replication of human adenovirus. Treatment with ivermectin blocks nuclear localization of the E1A protein, an essential viral nuclear protein that functions early during infection. I also demonstrate, for the first time, that ivermectin inhibits the Imp-α/cNLS interaction. Interestingly though, despite the classical nuclear import pathway being extensively studied, up to 50% of Imp-α cargo in yeast do not have a cNLS, as one would expect. However, whether this is true with humans remained unclear. To address this, I used currently available databases and datasets for human Imp-α PPIs and computationally searched for cNLSs. Using my approach, I found that 20–50% of Imp-α interactors do not have predictable cNLS. Furthermore, I found that the majority of proteins in the Mediator complex associate with Imp-α without having a predictable cNLS. Based on these findings I hypothesized that components of Mediator are likely to be using a "piggybacking" mechanism. These findings also demonstrated a need for identifying piggybacking mechanisms and/or novel NLSs. To explore these questions, I developed a yeast-based genetic selection to identify peptides conferring nuclear import. This system uses a large recombinant protein to express randomly generated peptides that can be subsequently selected for based on their ability to facilitate nuclear import in yeast. Peptides that I identified in this selection were also able to localize EGFP to the nucleus and interact with Imp-α in human cells. This approach also represents a novel strategy to identify SLiMs in a high throughput fashion, an area of SLiM discovery that currently lacks high throughput experimental methods.

Summary for Lay Audience

A human cell can be broken down into several different compartments. The two largest compartments are represented by the nucleus and the cytoplasm, which are physically separate from each other by the nuclear envelope (NE). This separation ultimately means that all human genes are expressed within the nucleus and translated into proteins within the cytoplasm. Proteins are generally regarded as the functional molecules of a cell and are responsible for carrying out cellular processes in every compartment, including the nucleus. For proteins to enter the nucleus, they must contain a nuclear localization sequence (NLS). Importantly, many viral proteins have NLSs as well, allowing these proteins to enter the nucleus and promote viral replication. Here, I show that nuclear import can be targeted by the drug ivermectin to block the replication of human adenovirus, a clinically important virus which currently lacks specific antiviral treatments. Additionally, I demonstrated that many nuclear proteins in humans do not have an NLS. Looking specifically at these proteins without an NLS, I provided evidence that proteins can "piggyback" into the nucleus. This means that a protein with an NLS can physically interact with a non-NLS protein and carry it into the nucleus. In particular, I provide evidence that an important cellular protein complex, Mediator, is likely to use piggybacking as a strategy. Since many nuclear proteins do not have an NLS, strategies for finding how they are transported into the nucleus are needed. To address this, I developed a genetic selection in yeast. This approach used an engineered protein that is too large to enter the nucleus unless it has an NLS. Expressing this protein in yeast with random protein sequences allowed me to select for those that could mediate nuclear import, since only these yeast were able to survive. Using this selection, I identified several protein sequences that look nothing like current NLSs. Interestingly, two of these NLSs were also able to function in human cells. Together, my findings demonstrate that additional strategies to gain access to the nucleus exist and that the nuclear import pathway can be targeted to inhibit viral replication.

Keywords

Nuclear import, nuclear localization signal, NLS, short linear motif, SLiM, genetic selection, *S. cerevisiae*, protein-protein interaction, human adenovirus, HAdV, ivermectin, piggybacking, proteomics

Co-Authorship Statement

Chapter 2: This chapter was published in *Journal of Virology*, 2020; 94(18), e00710-20. This paper was co-authored by myself and Dr. Cason King. I was responsible for immunofluorescence, qPCR, immunoprecipitation and genome replication experiments. I also participated in data analysis, writing, experimental planning, manuscript preparation and submission.

Chapter 3: This chapter was published in the journal *Biology*, 2020; 9(8): 188. I was the lead author of this manuscript and responsible for methodology, data analysis and writing.

Acknowledgments

First and foremost, there is no way I can thank my family members enough for how much they have supported me over the years. Despite never being able to fully understand what I'm doing; you've provided me with the resources and space to explore my interests and that I am truly grateful for.

I would like to thank my amazing partner, Courtney, who has been with me throughout my PhD studies. You have been incredibly supportive and beyond patient with me. Your constant source of encouragement has always helped to keep me motivated, even during my most doubtful and cynical phases.

I would like to thank current and past members of the Mymryk lab. I owe a special thanks to Drs. Michael Cohen, Greg Fonseca and Cason King, who were my mentors when I began working in the lab. I would also like thank Dr. Steven Gameiro, Martin Prusinkiewicz, Mackenzie Dodge, Katelyn MacNeil and Wyatt Anderson, who I shared much of my time with while in the lab. I appreciated your feedback and friendships over the years as well as the great times we've enjoyed together.

I also owe a special thank you to Dr. John Barrett for providing his expert advice and guidance. Working alongside you has undoubtedly benefited me in many ways as a scientist.

I would like to thank my committee members, Dr. Greg Dekaban and Dr. Chris Brandl, who have provided me with invaluable guidance throughout my PhD studies. I always appreciated your honest and critical feedback and I feel it has significantly contributed to my development as a scientist.

Finally, I would like to thank Joe for his mentorship. Your knowledge is as deep as it is wide, ranging from pure obscurity to in-depth scientific topics. Coming into the lab every day was nothing but pure enjoyment. The principles you instill in your students, including myself, is something I aim to replicate moving forward. I was hesitant when I began my graduate studies and unsure if it was what I wanted to do. My time spent in the lab working for Joe has made me confident in my abilities as a scientist and I am extremely grateful for being able to spend my formative years under your supervision.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

2A^{pro}	2A protease
3-AT	3-amino-1,2,4-triazole
AKAP	A-kinase anchoring protein
APC/C	Anaphase promoting complex
AR	Ankyrin repeat
ARM	Armadillo
BKPyV	BK polyomavirus
BME	β-mercaptoethanol
CBP	Creb-binding protein
CIRBP	Cold Inducible RNA Binding Protein
cNLS	Classical nuclear localization sequence
CoIP	Co-immunoprecipitation
DAPI	4',6-diamidino-2-phenylindole
DBD	DNA binding domain
DBP	DNA binding protein
ddH$_2$O	Double distilled water
DENV	Dengue virus
DMEM	Dulbecco's modified Eagle's medium
EBV	Epstein-Barr virus
FBS	Fetal bovine serum
EGFP	Enhanced green fluorescent protein
ELM	Eukaryotic linear motif
FDR	False discovery rate
FG	Phenylalanine glycine
G6PD	Glucose-6-phosphate dehydrogenase
Gal4$_{AD}$	Gal4 activation domain
Gal4$_{DBD}$	Gal4 DNA binding domain
HAdV	Human adenovirus
HBV	Hepatitis B virus
HCMV	Human cytomegalovirus

HCV	Hepatitis C virus
HeV	Hendra virus
HIV	Human immunodeficiency virus
HPA	Human Protein Atlas
Hpi	Hours post infection
HPV	Human papillomavirus
HRV	Human rhinovirus
HSV	Herpes simplex virus
HTLV	Human T-cell leukemia virus
HuRI	Human Reference Interactome
IAV	Influenza A virus
IBB	Importin beta binding
iCBS	Importin-α C-terminal binding segment
IDD	Intrinsically disordered domain
IDP	Intrinsically disordered protein
IDPR	Intrinsically disordered protein region
IFN	Interferon
Imp-α	Importin alpha
Imp-β1	Importin beta 1
IRF	Interferon regulatory factor
ISG	Interferon stimulated gene
IVM	Ivermectin
JAK	Janus kinase
Kapβ	Karyopherin beta
MBP	Maltose-binding protein
Mediator	Mediator of RNA polymerase II transcription
MERS-CoV	Middle East respiratory syndrome coronavirus
MOI	Multiplicity of infection
MoRF	Molecular recognition feature
NE	Nuclear envelope
NES	Nuclear export sequence
NLS	Nuclear localization sequence

NP	Nucleoplasmin
NPC	Nuclear pore complex
NTR	Nuclear transport receptor
Nup	Nucleoporin
NUPL2	Nucleoporin-like protein 2
ONPG	Ortho-nitrophenyl-β-galactoside
ORF	Open reading frame
PBS	Phosphate buffered saline
PEG	Polyethylene glycol
PIC	Preinitiation complex
PKA	Protein kinase A
PPI	Protein-protein interaction
PRIDE	Proteomics Identification Database
PRISMA	Protein interaction screen on peptide matrix
ProP-PD	Proteomic peptide phage display
PTM	Post translational modification
PV	Poliovirus
PY-NLS	Proline-tyrosine nuclear localization sequence
RanGAP1	Ran GTPase activating protein 1
RegEx	Regular expression
RFU	Relative fluorescence unit
RNAPII	RNA polymerase II
SD	Synthetic drop-out
SEM	Standard error of the mean
SLiM	Short linear motif
STAT	Signal transducer and activator of transcription
SV40	Simian virus 40
Syo1	Symportin-1
TAg	T antigen
TFIID	Transcription factor II D
TNPO1	Transportin 1
TNPO2	Transportin 2

TNPO3	Transportin 3
TREX	Transcription export adapter
U snRNP	Uridine-rich small ribonucleoprotein
VEEV	Venezuelan equine encephalitis virus
WCE	Whole cell extract
WT	Wild type
Xpo1	Exportin 1
Y2H	Yeast two-hybrid

Chapter 1

1 Introduction

1.1 General Introduction

The human proteome is composed of a collection of proteins that range in shape, size and form. These proteins function together to form complex biological networks that link underlying genetic information with the diverse phenotypes observed in human health and disease. Since the completion of the Human Genome Project it has become widely accepted that there are approximately 20,000 protein-coding genes; however, how this relates to the size of the human proteome is currently under question (Aebersold et al., 2018; Yates et al., 2020; Zahn-Zabal et al., 2020). Through cellular events like alternative splicing, RNA editing and post-translational modification (PTM), a combinatorial explosion of possible protein forms, or proteoforms, can be expressed from a single gene (Figure 1.1) (Aebersold et al., 2018; Hornbeck et al., 2019; Lau et al., 2019). Consideration of these factors, along with other sources of proteomic variation, has led to the estimation of up to 1,000,000 possible proteoforms within a cell at a given time (Aebersold et al., 2018). Understandably, how these proteins and proteoforms participate in protein-protein interactions (PPI) and how these PPIs are regulated is of particular importance to understanding many aspects of cell biology and systems biology.

Proteins represent the molecular workhorse of the cell and are involved in nearly every cellular task. Essential cellular processes such as signal transduction, transcription and translation require a myriad of PPIs to occur in specific and context-dependent settings. Importantly, physical interactions between proteins enable a cell to integrate information from both external and internal stimuli, which in turn allows a cell to make an appropriate decision based on that information. Therefore, to understand how the proteome functions as a whole, it is critical to understand how PPIs are mediated. Physical interactions between proteins can be thought of as a form of communication and how they communicate is embedded within their amino acid sequence. How the sequence of amino acids is arranged within a protein will determine the overall physical structure, which will ultimately impact
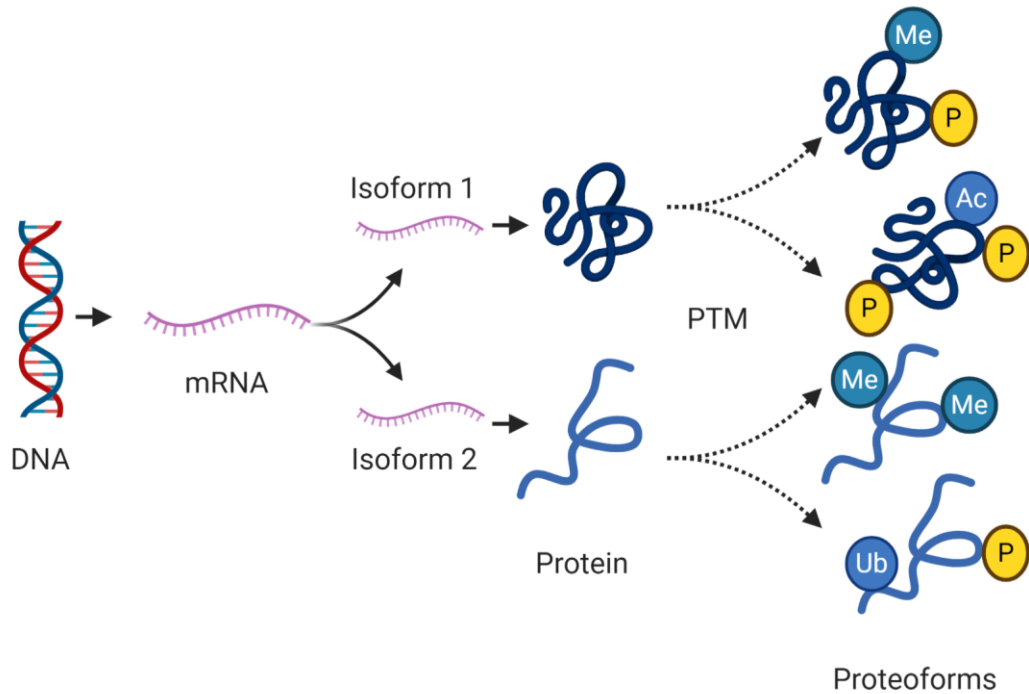
Figure 1.1 Sources of proteomic diversity.

A single gene can give rise to multiple different proteoforms through various cellular processes such as alternative splicing, RNA editing and post-translational modification. Alternative splicing and RNA editing of mRNA can give rise to multiple different spliceforms that differ in their primary amino acid sequence. Structural regions or regions mediating PPIs can be lost, giving rise to unique proteins with different functions or localizations. Additionally, the primary amino acid sequence can be chemically modified to produce a variety of different proteoforms which may bestow unique or highly specific functions. P = phosphorylation, Ub = ubiquitination, Me = methylation, Ac = acetylation. Created with BioRender.

the type of possible PPIs. In general, within the proteomic landscape, proteins can range from being highly structured to completely disordered. The latter are referred to as intrinsically disordered proteins (IDPs); however, many proteins are a mosaic of structured regions and intrinsically disordered protein regions (IDPRs) (A K Dunker et al., 2001; P E Wright & Dyson, 1999).

Well-folded globular domains represent highly structured interaction modules and represent the majority of studied interaction interfaces (Mosca et al., 2014; Pawson & Nash, 2003). Interestingly, prediction models dating back to 1998 demonstrated a significant portion of proteins from diverse species contain disordered regions, providing further evidence that a protein's function is not just dependent on possessing a stable structure (Romero et al., 1998). Today it is appreciated that approximately one-third of the human proteome is comprised of IDPRs; however, most of these regions are uncharacterized compared to globular domains (A K Dunker et al., 2001; Mosca et al., 2014). Despite their lack of characterization, it is now well understood that these regions represent an important regulatory framework by facilitating various cellular processes which include PPIs and more recently phase separation (Brocca et al., 2020; Peter E Wright & Dyson, 2015). The most common protein interaction module within IDPRs is the short linear motif (SLiM), sometimes referred to as a molecular recognition feature (MoRF), or minimotif, which are implicated in directing various cellular processes including PTM, protein localization, cleavage, docking and degradation (Mi et al., 2012; Mohan et al., 2006; Van Roey et al., 2014). There are an estimated 100,000 SLiMs (excluding PTMs) in the human proteome; however, only a few thousand of these have been identified to date (Kumar et al., 2020; Tompa et al., 2014). This disparity highlights a significant knowledge gap in our understanding of SLiMs and their importance to basic cell biology.

Fascinatingly, the nature of SLiMs and how they have emerged evolutionarily has more or less made them an Achilles heel. Pathogens, such as viruses have co-evolved to exploit host-SLiMs through molecular mimicry, allowing them to integrate into, and take advantage of host pathways (Davey et al., 2011). Since a diverse range of viruses can infect the same host, they are often faced with the same set of host factors. This level of selective

pressure is likely why diverse viruses have convergently evolved mechanisms to usurp key cellular processes like protein nuclear import and evade the host immune system (Via et al., 2014). These observations also highlight the importance of studying viral processes, as many SLiM-mediated interactions were first discovered through viral PPI studies (R. E. Jones et al., 1990; Daniel Kalderon et al., 1984).

1.2   Short Linear Motifs (SLiMs)

1.2.1   SLiMs are a class of protein-protein interaction

PPIs can be mediated through a range of binding modules that exist on a continuum between highly structured globular domains and intrinsically disordered SLiMs. In general, three classes of protein interaction modules have been proposed and these include: globular domains, intrinsically disordered domains (IDD) and SLiMs (Table 1.1) (Tompa et al., 2009; Van Roey et al., 2014). Globular domains are easily distinguished from IDDs and SLiMs based on their ability to form a stable tertiary structure in the absence of a binding partner (Han et al., 2007). In contrast, both IDDs and SLiMs are found within IDPRs and lack a well-defined tertiary structure under native conditions (Tompa et al., 2009). It is worth noting that even though SLiMs are almost exclusively found within IDPRs, they can also be found within disordered loops of globular domains. For example, phosphorylation sites have been frequently identified in this context (Via et al., 2009; Zanzoni et al., 2007). Additionally, both SLiMs and IDDs preferentially bind globular domains. However, IDDs have also been shown to bind other IDDs, such as with the p53 tetramerization domain, where two disordered regions adopt an ordered structure upon interaction (Fichó et al., 2017; Jeffrey et al., 1995).

Of the three classes, SLiM-mediated interactions are the weakest, with binding affinities in the low micromolar range, compared to IDD and globular domain interactions, which fall in the nanomolar and picomolar ranges, respectively (Hirschi et al., 2010; Kastritis et al., 2011; C. W. Lee et al., 2010). Although IDDs and SLiMs have several similarities, such as being found within IDPRs and forming transient interactions, a major discriminating factor is their sequence length (Davey et al., 2012). IDDs are substantially longer and range from 20-50 amino acids, while SLiMs are typically less than 10 amino

acids (Davey et al., 2012; Tompa et al., 2009). Finally, SLiM interaction interfaces are almost always monopartite and can function independently (Van Roey et al., 2014). This is an important distinction from IDDs, which interact in a multipartite fashion, meaning there are several distinct units that contribute together towards the function of an IDD.

1.2.2 Attributes of SLiMs

A distinguishing feature of SLiMs is their ability to make highly specific yet transient interactions using only a limited number of amino acids. Firstly, due to the inherent properties of SLiMs, their amino acid sequence information is most easily conveyed using the single-letter amino acid code in the form of a regular expression. To date, nearly 300 classes of SLiMs have been identified from a diverse range of biological processes and these are currently deposited within the Eukaryotic Linear Motif (ELM) resource (Kumar et al., 2020). In general, SLiMs almost exclusively bind globular domains; however, it's worth mentioning that examples of RNA binding and lipid binding motifs have also been documented (Kojima et al., 2004; Phan et al., 2011). Importantly, the ELM resource represents motifs that have been manually curated and experimentally validated over the past 20 years, greatly expanding our knowledge of SLiMs and allowing for comprehensive analyses of their attributes (Davey et al., 2012).

Analysis of motifs from the ELM resource have demonstrated that the majority of SLiMs are between 3–10 amino acids in length; however, instances of motifs over 20 amino acids, such as the bipartite nuclear localization signal (NLS), are well documented (Davey et al., 2012; Kumar et al., 2020). On average, only 2 to 4 amino acids within a motif are defined positions. These can be either fixed or degenerate, meaning certain positions, depending on the type of motif, can be substituted with highly similar amino acids. Additionally, the majority of motifs have several wild card positions, allowing for the accommodation of potentially any amino acid. In contrast, some positions within a motif cannot tolerate particular amino acids, particularly those that are physically and/or chemically incompatible with their corresponding binding site. This can be demonstrated with the monopartite NLS, which cannot accommodate aspartic acid or glutamic acid within or adjacent to the core motif (Kosugi et al., 2009). As a whole, most classes of SLiMs are found internally within a protein's linear amino acid sequence; however, certain classes

Table 1.1 Properties of protein-protein interaction modules.

Property	Globular domain	Intrinsically disordered domain (IDD)	Short linear motif (SLiM)
Length (amino acids)	50–200	20–50	3–10
Conformation	folded	disordered	disordered
Binding mode	multipartite	multipartite (linear)	monopartite
Affinity (with globular domain)	nanomolar-picomolar	nanomolar	micromolar
Binding partner	globular domains, IDDs, SLiMs	globular domains, IDDs	globular domains
Binding dynamics	stable	transient	transient

such as the PDZ domain binding motif or IAP-binding motif are specifically found at the C- and N-termini, respectively (Kumar et al., 2020).

1.2.3 Functions of SLiMs

It is well recognized that IDPRs play a significant role in cell signalling and regulation (Peter E Wright & Dyson, 2015). This can largely be attributed to the function of SLiMs, which are the most commonly found interaction module within IDPRs (Diella et al., 2008; Fuxreiter et al., 2007; Van Roey et al., 2014). According to the ELM resource, SLiMs can be categorized into one of six categories based on function (Figure 1.2) (Kumar et al., 2020). These classes include cleavage, degradation, docking, ligand binding, modification and targeting. SLiM-mediated interactions are involved in a multitude of diverse cellular processes as they can provide the necessary functional plasticity for multifunctional proteins (Kumar et al., 2020; Zanzoni et al., 2019). Additionally, SLiMs allow for the integration of different signals, acting as molecular switches to precisely control protein localization and/or function (Van Roey et al., 2013). One way this can be achieved is through PTMs like phosphorylation. For example, phosphorylation adjacent to the NFATC1 NLS by PKA acts a priming event for binding of GSK3 which results in further phosphorylation that completely blocks NFATC1 nuclear localization (Sheridan et al., 2002). Many examples of such switches exist and others have been implicated in regulating protein stability temporally and/or spatially and directing the assembly of signalling complexes, for example (Van Roey et al., 2013). Interestingly, many PTM classes found within IDPRs are specifically associated with regulatory or signalling regions, in contrast to PTMs targeted to ordered regions, which are mainly involved in changing catalytic function or conformational stability (Darling & Uversky, 2018; Xie et al., 2007). Importantly, PTMs like phosphorylation are reversible, allowing for SLiM functions to be turned on or off in the appropriate contexts, a key property of signalling pathways (Ardito et al., 2017).

Perhaps the most important aspect of IDPRs and their embedded SLiMs is their enrichment within hub proteins, or proteins which interact with hub proteins
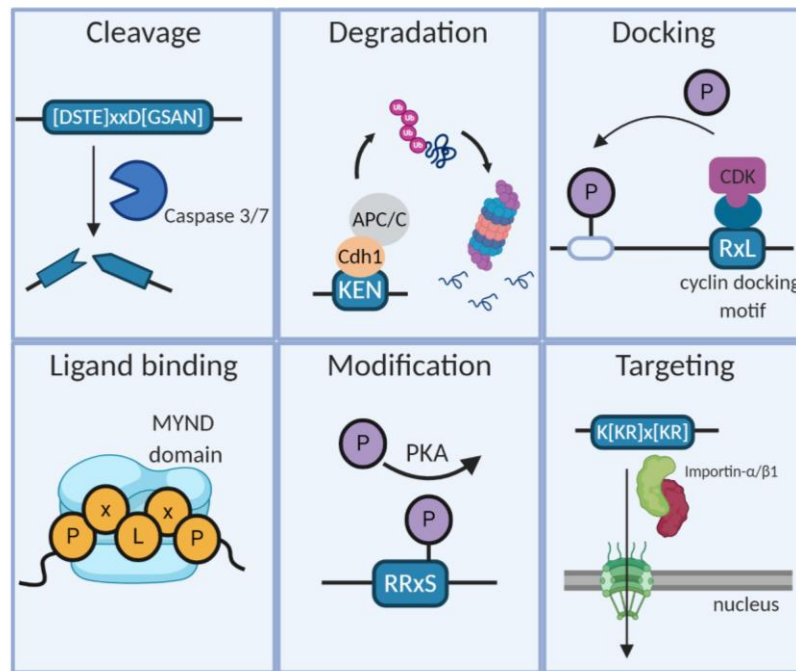
Figure 1.2 Categories of SLiMs.

SLiMs can be broadly categorized into six different categories based on function. Selected examples are depicted with corresponding SLiMs where "X" represents any amino acid and amino acids within square brackets can be substituted with one another. Cleavage motifs are recognized by enzymes, such as caspases, which irreversible cleave the protein. Degradation motifs are recognized by ubiquitin ligase complexes, such as the anaphase promoting complex (APC/C), and mediate proteosomal degradation. Docking motifs generally act as enzyme recruitment sites and increase substrate specificity. In the example above, cyclin binding to its docking motif increases substrate specificity for cyclin-dependent kinases (CDK). Ligand binding motifs are a broad category mediating interactions that do not result in modification upon binding, whereas modification sites themselves are directly chemically modified. Targeting motifs fall within the ligand binding category, however; their interactions result in localization to specific subcellular compartments such as the nucleus via NLSs. Created in BioRender.

(Dosztányi et al., 2006; A Keith Dunker et al., 2005; Haynes et al., 2006; Hu et al., 2017; Jespersen & Barbar, 2020). Within an interactome, proteins can interact with one another to varying degrees (Huttlin et al., 2015). Some proteins are highly connected, participating in possibly tens to hundreds of interactions, while others are less connected with only a few binding partners. These highly connected proteins have come to be recognized as hub proteins (A Keith Dunker et al., 2005; Ekman et al., 2006). Their importance in organizing PPI networks is based on evidence that genetic deletion of these proteins is most often lethal. This concept has been coined as "lethality and centrality" and is underscored by experimental evidence showing that highly connected proteins, which represent a small percentage of total PPIs, are most often essential for survival (Hu et al., 2017; Jeong et al., 2001; Zotenko et al., 2008). The presence of IDPRs within hub proteins provide a flexible backbone allowing them to dynamically sample the interaction space, a necessity for these multifunctional proteins. This can be exemplified with the highly disordered portions of the p53 protein (Uversky, 2016). Widely regarded as the "guardian of the genome" due to its tumor suppressor qualities, p53 is involved in a range of cellular processes (Lane, 1992; Sionov & Haupt, 1999). It's N- and C-terminal regions are highly disordered, enriched in PTMs and facilitate numerous SLiM-mediated interactions. The overall intrinsic disorder of p53 gives rise to multiple proteoforms which will favour certain interactions and PTMs (Uversky, 2016). Importantly, the lack of structure in IDPRs allows each proteoform to dynamically respond to cellular conditions and make the necessary PPIs without being locked in a particular conformation, demonstrating an important quality of hub proteins (Jespersen & Barbar, 2020).

1.2.4   SLiM evolution

Intrinsic disorder is a conserved feature of proteins among all living organisms, as well as viruses (Xue et al., 2012). Interestingly, computational studies have demonstrated that eukaryotes, both unicellular and multicellular, display a greater amount of intrinsic disorder than prokaryotes or archaea (A K Dunker et al., 2000; Ward et al., 2004; Xue et al., 2012). Why this is the case is likely due to the need for greater cell signalling and regulation in higher eukaryotes, as this is evidenced by signalling proteins having the greatest degree of IDPR enrichment (Iakoucheva et al., 2002). Since SLiMs are almost exclusively found

within IDPRs, their evolution is inextricably connected to the presence of IDPRs within proteomes.

In contrast to structured domain evolution, which predominantly involves gene duplication, IDPRs have been shown to evolve much more rapidly (Brown et al., 2002; Vogel et al., 2004). The presence of SLiMs within these rapidly evolving regions has now led to the interesting concept of *ex nihilo* evolution, meaning "from nothing" (Davey et al., 2015). Considering SLiMs only possess 2–4 critical amino acids that contribute the majority of affinity and specificity in binding, it seems reasonable that a single point mutation, or small insertion/deletion, could create a novel motif from nothing (Stein & Aloy, 2008). Through random mutation, identical motifs have convergently evolved *ex nihilo* in unrelated proteins, such as the PxIxIT calcineurin-binding motif and many others (Figure 1.3A) (Davey et al., 2015; Kumar et al., 2020). Post-duplication, if a protein does not acquire an advantageous function it can be eliminated from the gene pool due to genetic drift or negative selection. Therefore, through *ex nihilo* motif generation, homologous proteins have the ability to rapidly diversify their function post-duplication, as was the case with *S. cerevisiae* Ace2 and Swi5 (Figure 1.3B) (Nguyen Ba et al., 2014).

It's well documented that many disease-causing mutations are a result of point mutations which eliminate a motif (Kumar et al., 2020; Uyar et al., 2014). Interestingly though, several diseases have been attributed to *ex nihilo* motif generation. One such example is an a S→G mutation at position 2 of SHOC2, resulting in Noonan-like syndrome (Figure 1.3C) (Cordeddu et al., 2009). This mutation creates a novel N-myristoylation site in SHOC2, resulting in aberrant targeting to the plasma membrane and impaired nuclear import. Another example involves the generation of dileucine motifs within cytosolic tails of several transmembrane proteins; CACNA1H, GLUT1 and ITPR1 (K. Meyer et al., 2018). Investigation of GLUT1 demonstrated novel interactions with AP-1 and AP-2, resulting in an intracellular localization of GLUT1 (Figure 1.3D). Both of these examples highlight the ease at which a single point mutation can generate a novel SLiM.

In addition to novel motif evolution, the same process can allow for "tuning" of existing motifs (Davey et al., 2015). An important attribute of SLiMs is their ability to mediate
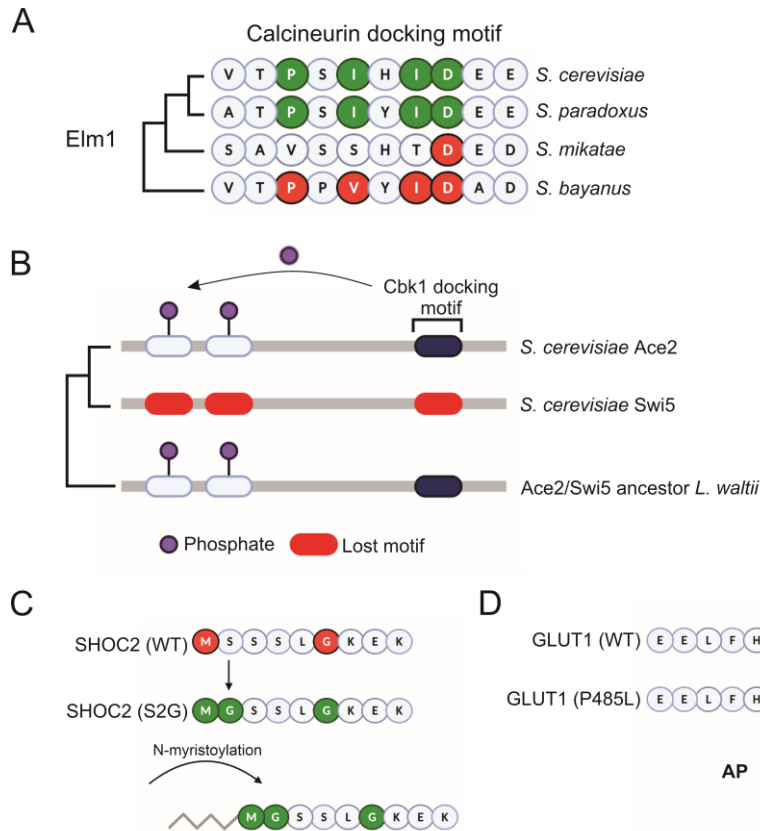
Figure 1.3 *Ex nihilo* evolution of SLiMs.

Due to the low sequence complexity of SLiMs, they can be gained or lost through single amino acid substitutions. *Ex nihilo*, meaning "from nothing", implies the generation of a motif from a peptide sequence otherwise devoid of SLiMs. **(A)** An example of *ex nihilo* evolution can be observed with the calcineurin docking motif of the yeast Elm1 protein (green). This motif likely evolved from a common ancestor of *Saccharomyces cerevisiae* and *Saccharomyces paradoxus*. **(B)** SLiMs can also be rapidly lost, altering a proteins functionality. Following whole genome duplication, *S. cerevisiae* Ace2 and Swi5 were retained; however, the Cbk1 docking motif and phosphorylation sites were lost in Swi5. **(C and D)** Single amino acid substitutions can also "knock in" motifs that cause disease. This has been documented with human SHOC2, where an S2G mutation creates an N-myristoylation site resulting in Noonan-like syndrome due to aberrant localization. Likewise, a P485L mutation in human GLUT1 creates a dileucine motif that is recognized by adapter proteins (AP), also causing aberrant intracellular localization. Created with BioRender.

transient interactions, an important attribute of signalling pathways in general. These transient binding properties have evidently been selected for, as artificially optimizing motifs for higher or lower affinity have a proportional effect on their respective signalling pathways (Marles et al., 2004; Roy et al., 2007). These observations highlight the idea that biological systems are optimized for efficiency. Through *ex nihilo* evolution, SLiMs can be acquired and then be further optimized, or tuned, in an evolutionarily rapid process.

1.2.5 Pathogenic molecular mimicry of SLiMs

The term molecular mimicry refers to structural or sequence similarities between pathogens and host proteins. This term was historically applied to shared antigens (antigenic mimicry) which results in cross-activation of the immune system (Kohm et al., 2003). However, this term is also broadly used to describe the sharing of any sequence/structure between pathogen and host (Damian, 1964). Pathogens like viruses, as well as some bacteria and parasites, replicate intracellularly within their host and therefore require an intimate relationship with host proteins to influence cellular processes such as cell cycle, metabolism, or detection by the immune system, for example. Since pathogens from diverse taxonomies infect the same organism, they are all faced with a similar range of host factors. Based on this, it is not surprising that they have acquired similar strategies, through convergent evolution, to manipulate host cellular processes. One widely used strategy that will be discussed further is the mimicry of host SLiMs by viral proteins (Davey et al., 2011).

As obligate intracellular parasites, viruses have evolved an intimate relationship with their host. As a result of their dependence on host cells for replication, viral genomes have evolved in a lockstep manner with their host (Simmonds et al., 2019). Importantly, each step of the viral replication cycle typically involves extensive host-pathogen PPIs, which is evident based on the number of viral PPI studies that dominate the literature compared to non-viral pathogens. These interaction studies have demonstrated the widespread presence of all six classes of SLiMs, as defined in the ELM resource, within viral proteomes (Davey et al., 2011; Kumar et al., 2020). In fact, early viral studies pioneered the discovery and characterization of SLiMs (R. E. Jones et al., 1990; D Kalderon et al.,

1984; Lyons et al., 1987). These early examples involved discovery of NLSs in human adenovirus (HAdV) E1A and simian virus 40 (SV40) large T antigen (TAg), as well as the retinoblastoma-binding SLiM in human papilloma virus (HPV) E7. Through extensive use of SLiMs, diverse viruses are able to target similar host pathways (Davey et al., 2011). These include intracellular transport via nuclear localization and export signals, immune evasion through signal transduction, altering host protein levels via proteosomal degradation, cell cycle regulation through transcriptional regulation and others such as host and viral PTM (Deng et al., 2004; Felsani et al., 2006; Horwitz et al., 2008; Stanley et al., 2008; Tessier et al., 2019; Thorley-Lawson, 2001; Welcker & Clurman, 2005).

Since cellular hub proteins are relatively few in number, the logical consequence of a random loss of a PPI is statistically less likely to have a detrimental effect on the overall system (Albert et al., 2000). The benefits of such a system are clear; however, this situation exposes a critical disadvantage where hub proteins can represent an Achilles heel within a signalling network, giving rapidly evolving pathogens, including viruses, the ability to hijack cellular pathways (Davey et al., 2011; Yu et al., 2007). It's now well established that viral proteins have evolved to specifically target cellular hub proteins (Calderwood et al., 2007; de Chassey et al., 2008; Dyer et al., 2008; Garamszegi et al., 2013). Computational analyses comparing endogenous (host-host interactions) and exogenous (virus-host) PPI binding surfaces revealed that viral proteins preferentially target SLiM-binding domains (Garamszegi et al., 2013). In fact, the probability of a cellular protein being a viral target positively correlates with the number of PPIs it makes endogenously, as well as with the number of SLiM-binding domains present within proteins that make a similar number of PPIs (Garamszegi et al., 2013). Together, this demonstrates that viral proteins have evolved to preferentially target cellular hub proteins with SLiM-binding domains. Furthermore, these viral proteins are enriched for disorder promoting amino acids and SLiMs on a per residue basis compared to human proteins.

Fascinatingly, some viral proteins participate in so many interactions that they themselves act as a hub protein. This is best demonstrated with the HAdV E1A protein which employs a multitude of SLiM-mediated interactions (Figure 1.4). Remarkably, using only
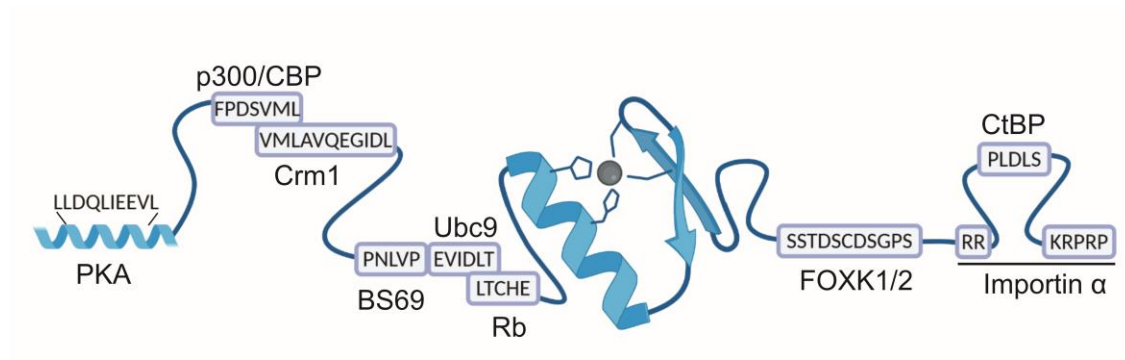
Figure 1.4 Molecular mimicry of SLiMs by HAdV-C5 E1A.

Depicted is the full length (289 amino acid) HAdV-C5 E1A protein. E1A is almost entirely disordered, allowing it to interact with a wide range of cellular proteins. Using a variety of SLiMs E1A can interact with host cellular proteins involved in different cellular processes such as intracellular transport, post-translational modification, transcriptional regulation and cell cycle regulation. Additionally, E1A is a classic example of how viral genomes, which have limited coding capacity, encode for proteins with extremely high functional density, as depicted by the many overlapping SLiMs. Created with BioRender.

289 amino acids, E1A can make up to 32 primary PPIs which are involved in altering cell cycle progression, transcriptional regulation, nuclear localization, relocalization of cellular targets to the nucleus, global PTM, and endogenous protein levels (King, Zhang, et al., 2018). However, in addition to E1As role as a viral hub protein, E1A can also target and manipulate cellular hub proteins such as CREB-binding protein (CBP) which participates in several hundred PPIs itself (Ferrari et al., 2014; Hu et al., 2017). These examples, along with others like human immune deficiency virus 1 (HIV-1) Nef, Epstein-Barr virus (EBV) LMP1 and SV40 TAg, highlight the ease at which viral proteins can usurp cellular processes through systematic targeting of cellular hub proteins and even function as hub proteins themselves (Davey et al., 2011).

In terms of viral SLiM evolution, the same process of *ex nihilo* motif generation used by cellular proteins can be applied. Given the high mutation rates for both RNA and DNA viruses, this process is probably accelerated and it's not unlikely that new motifs could rapidly emerge if a particular mutation is beneficial (Duffy et al., 2008; G. M. Jenkins et al., 2002). Since virus-host evolution is highly intertwined, it is not surprising that both have convergently evolved highly similar SLiMs (Davey et al., 2011). Analysis of Influenza A virus (IAV) H3N2 hemagglutinin protein from 1968 to 2003 is a fascinating example of this, where the number of Nx[TS] glycosylation motifs increased from 2 to 7 over this nearly 40 year period (Igarashi et al., 2008). Additionally, substitution of a single amino acid can dramatically alter SLiM-binding specificity. Rabies virus protein G from VIR and ATT viral strains both possess a PDZ-binding SLiM; however, VIR protein G promotes neuronal cell survival and ATT protein G results in neuronal cell death (Préhaud et al., 2010). These phenotypic differences are the result of a single amino acid change in their PDZ-binding SLiMs, allowing VIR (QRTL) and ATT (ERTL) protein G to outcompete cellular MAST2-PDZ and PTPN4-PDZ targets, respectively. In addition to altered specificity, viral SLiMs often differ in affinity compared to their cellular counterparts. For example, the HIV-1 Nef SH3 domain binding motif, PxxP, has a higher affinity for cellular Hck than host motifs due to additional contacts made outside of the core motif (Stangler et al., 2007). Likewise, the HAdV-C5 E1A protein is able to outcompete binding of cellular AKAP7 to PKA (King, Cohen, et al., 2016). In these

examples, a higher binding affinity was required for their respective function. However, a higher binding affinity is not always necessary and this has been shown with the HPV E6 interaction with cellular MAGI1, which binds with the same affinity as its cellular targets (Fournane et al., 2011).

One of the proposed properties of viral SLiMs that differentiates them from their eukaryotic counter-parts is the degree of intrinsic disorder within their core motif (Duro et al., 2015). Both human and viral SLiM containing regions show similar levels of intrinsic disorder when comparing the core SLiM region and flanking amino acids together. Interestingly, however, viral SLiMs tend to show greater disorder in their core motif compared to the flanking regions, presumably giving them greater local flexibility. Furthermore, human SLiMs tend to show a greater bias towards undergoing a disorder-order transition upon binding, while viral SLiMs appear more likely to remain unstructured upon binding (Duro et al., 2015). This has been demonstrated with the C-terminal domain of measles virus nucleoprotein, which remains disordered while bound to viral phosphoprotein (Bourhis et al., 2005). This ability to remain disordered while bound has been termed "fuzziness" (Tompa & Fuxreiter, 2008). Reconsidering viral proteins like HAdV E1A, which makes dozens of primary PPIs with only 289 amino acids, sufficient conformational adaptability is likely a requirement. By using "fuzzy" motifs viral proteins can exist in multiple proteoforms that are compatible with a wide range of PPIs.

1.2.6   SLiM detection methodologies

Both computational and experimental approaches have contributed significantly to the discovery of SLiMs. Computationally, significant challenges arise due to the limited information content embedded within SLiMs. With only 2 to 4 positions critical for binding, some of which have flexibility for similar amino acids, the number of false-positive matches hinders computational discovery (Davey et al., 2012; Gibson et al., 2015). One factor common to all computational approaches is the false discovery of motifs within structured regions. Since SLiMs are almost exclusively found within disordered regions, adding this filter is essential to reduce the number of false positives. Programs, such as the ELM resource, filter results through SMART, Pfam and GlobPlot to identify protein domains, as well as IUPred for identifying regions of predicted disorder. In general,

computational approaches can be categorized into i) identifying novel SLiM instances of a known motif or ii) identifying SLiMs *de novo*. Many software programs and packages such as MEME Suite and SLiMSuite have been developed for these purposes (Bailey et al., 2009; Edwards & Palopoli, 2015).

Identifying instances of SLiMs can be as straight forward as matching regular expressions of known or putative SLiMs against a protein sequence, or more complex using programs like SLiMSearch, which can filter instances based on their intrinsic disorder score and evolutionary conservation (Krystkowiak & Davey, 2017). Conversely, computational *de novo* SLiM discovery is not as straightforward. One of the most common approaches relies on convergent evolution, or the independent evolution of shared motifs from unrelated proteins (Edwards & Palopoli, 2015). Importantly, this approach most often leverages either user generated or publicly available PPI data. Commonly used programs such as SLiMFinder and DILIMOT will take a group of proteins that share a common binding partner and attempt to identify statistically significant SLiMs that are shared among those proteins (Davey et al., 2010; Edwards et al., 2007; Neduva & Russell, 2006). However, due to the limitations of computational methodologies in general, experimental validation is required. These approaches have been previously used, with experimental follow up, to identify novel SLiMs from several different model organisms, as well as humans (Neduva et al., 2005).

Experimental approaches for discovering novel SLiMs can be categorized into either genetic or biochemical screens (Blikstad & Ivarsson, 2015; K. Meyer & Selbach, 2020). Common genetic screening techniques have employed yeast two-hybrid (Y2H) and phage display (Blikstad & Ivarsson, 2015). In both cases, bait proteins are typically SLiM-binding domains, and these are screened against a library of peptides (prey). With Y2H, both bait and prey are expressed within yeast, unlike phage display where baits are immobilized and peptides are displayed externally on phage. The main advantage of both techniques is the ease at which large prey libraries can be constructed. More recently, advances in phage display has resulted in the specific expression of peptides representing particular regions of the proteome, a method referred to as proteomic peptide phage display (ProP-PD) (Davey et al., 2017). As an example, peptides specifically representing disordered regions

of the human proteome were displayed on phage and screened against several different immobilized baits (Davey et al., 2017). Importantly, both Y2H and phage display techniques have a similar bottleneck, as they are both limited by the number of bait proteins that can be tested.

Common biochemical strategies have mostly involved the use of peptide arrays or protein microarrays (K. Meyer & Selbach, 2020). The major advantage of peptide arrays over protein microarrays is the ability to synthesize hundreds different peptides at high density on cellulose membranes through SPOT-synthesis (Hilpert et al., 2007). From here, recombinantly expressed bait proteins can be incubated with the membrane to identify interacting peptides. Through the use of mass spectrometry, throughput of prey identification can be significantly increased by incubating peptide arrays with highly complex whole cell lysates. This was first done using immobilized peptides and has now evolved into testing over 200 peptides, via SPOT-synthesis, against cell lysates using a recently developed method termed protein interaction screen on peptide matrix (PRISMA) (Dittmar et al., 2019). However, even though significant advances have been made, each of these approaches remain relatively low throughput at either the bait or prey level. Given that there are likely thousands of additional motif classes to be discovered, the need for developing high throughput methodologies to identify these novel SLiMs is a fertile area of research.

1.3   Protein Nuclear Transport

1.3.1 Separation of nuclear and cytoplasmic compartments

Within eukaryotic cells the nuclear envelope (NE) provides a physical barrier that spatially separates the contents of the nucleus and cytoplasm. The NE confines key cellular processes like transcription and translation to the nuclear and cytoplasmic compartments, respectively. This separation allows for complex regulation of gene expression, as well as a variety of other distinguishing factors between prokaryotic and eukaryotic organisms (Devos et al., 2014; Martin & Koonin, 2006). Newly transcribed mRNA must be exported from the nucleus to the cytoplasm to be translated, while nuclear proteins such as transcription factors and histones must be imported into the nucleus. For complex

biological processes like signal transduction to occur, there must be dynamic spatial and temporal regulation of proteins and other macromolecules between the cytoplasmic and nuclear compartments, which ultimately requires passage across the NE in either direction.

The exchange of proteins and molecules between compartments occurs exclusively through nuclear pore complexes (NPC) which are embedded throughout the NE (Wente & Rout, 2010). NPCs are large macromolecular structures with an aqueous central channel that facilitates bidirectional movement of molecules. Importantly, the biochemical and physical properties of the central channel establish a permeability barrier that allows the NPC to act as molecular sieve (Hoelz et al., 2011). Small proteins can diffuse through the NPC; however, as protein size increases this becomes increasingly difficult (Timney et al., 2016). To maneuver the NPC, cells employ a variety of nuclear transport receptors (NTR) called karyopherins. These proteins facilitate bidirectional transport of proteins through the NPC in a rapid, energy-dependent, process (Cautain et al., 2015).

1.3.2 The nuclear pore complex

The NPC is an impressive biological structure built upon a framework comprising multiple copies of roughly 30 different nuclear pore proteins called nucleoporins (Nups) (Beck & Hurt, 2017; Schwartz, 2016). In general, Nups can be categorized as either scaffold Nups, which function in an architectural capacity, or FG-Nups that make up the inner aqueous channel (Beck & Hurt, 2017). The basic architecture of the NPC is composed of different substructures that are organized in a highly modular fashion. Using only a few common domain folds, Nups can oligomerize into stable higher-ordered structures, giving rise to an overall architecture that can be broken down into several elements which include: the inner pore ring, nuclear and cytoplasmic rings, nuclear basket and cytoplasmic filaments (Figure 1.5).

Depending on their role within the NPC, Nups can be classified as either scaffold- or FG-Nups (Beck & Hurt, 2017). Scaffold-Nups are primarily composed of folded protein domains and contribute to the overall architecture of the NPC. The inner ring and Y-complex substructure, which is the main component of the nuclear and cytoplasmic rings, are believed to provide the main scaffolding element of the NPC
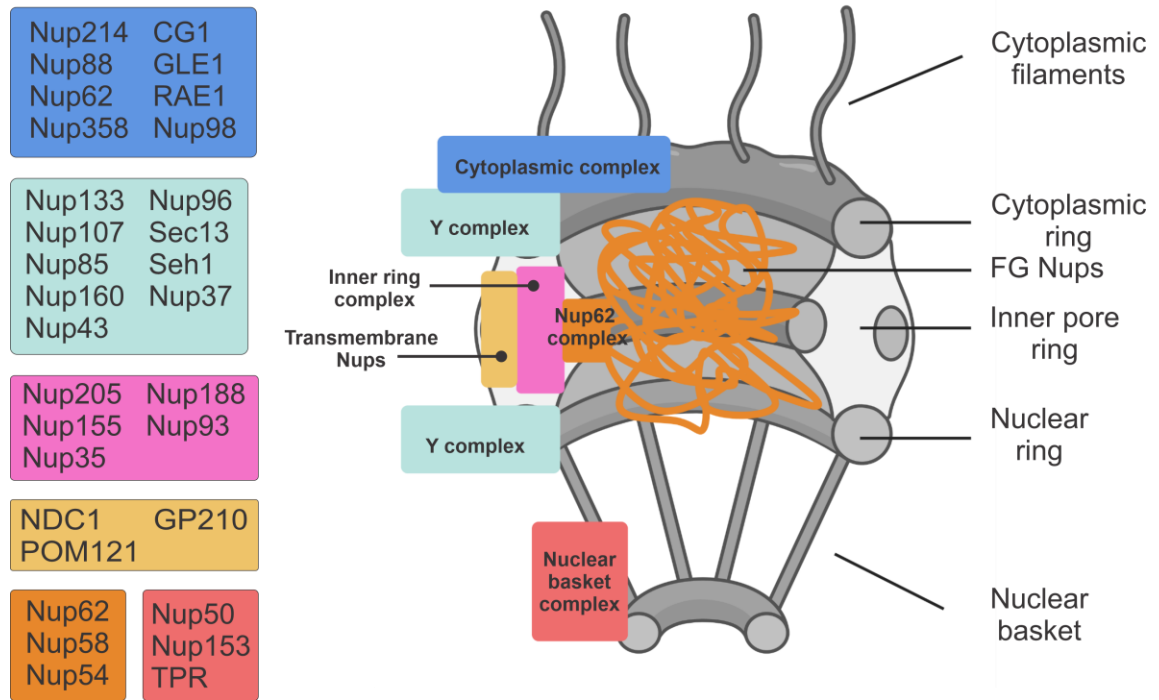
Figure 1.5 Structural organization of the nuclear pore complex.

The nuclear pore complex (NPC) is responsible for facilitating bidirectional transport of all molecules across the nuclear envelope (NE). Architecturally, the NPC can be categorized into several structures that include the cytoplasmic filaments, cytoplasmic ring, nuclear basket, nuclear ring and inner pore ring. The NPC is composed of nucleoporins (Nups) which are organized into various substructures, or modules, that fit within the NPCs overall architecture. Here, the various modules and corresponding Nups are colour coded and correspond to their location within the NPCs architecture. The Y complex and inner ring complex primarily make up the NPC scaffold, which is anchored to the NE via transmembrane Nups. Within the central channel is the Nup62 complex, which is primarily composed of phenylalanine glycine (FG) repeat containing Nups that bind NTRs and provide the permeability barrier of the NPC. Additionally, Nups within the nuclear basket and cytoplasmic complex contain FG Nups and aid in nuclear import and export. Figure adapted from Beck and Hurt, 2017 (Beck & Hurt, 2017). Created with BioRender.

(Beck & Hurt, 2017; Bui et al., 2013; Vollmer & Antonin, 2014). In contrast, FG-Nups contain large regions of intrinsic disorder that are rich in phenylalanine (F) and glycine (G) repeats (Terry & Wente, 2009). These repeats, which can range from 5–50, are separated by hydrophilic linker regions and come in several different flavours such as FxFG or GLFG. FG-Nups are found within both the cytoplasmic filaments and nuclear basket, which emanate from the cytoplasmic ring and nuclear ring, respectively, and are primarily responsible for recognizing NTRs (Bayliss et al., 2002). Additionally, FG-Nups line the inner pore ring within the central channel and create a sieve-like barrier. Here, FG-repeats interact with one another, forming a liquid phase separation that is proposed to contribute to the selectivity of the NPC (Celetti et al., 2020; Schmidt & Görlich, 2016).

Interestingly, it has been recently shown that connectivity within and between NPC subcomplexes is largely organized by SLiMs. Biochemical studies of the interactions from the inner ring complex of *C. thermophilum* demonstrated that SLiMs within the N- and C-termini of Nup53 mediated different interactions with other inner ring Nups (Amlacher et al., 2011). Furthermore, the inner ring Nup, Nup192, tethers to other Nups located within the outer ring scaffold and central channel FG-Nups (Fischer et al., 2015; D. H. Lin et al., 2016). An additional function of disordered FG repeats has also been implicated in stabilizing the structure of the NPC (Onischenko et al., 2017). As mentioned previously, a key feature that SLiMs bestow on their proteins is the possibility for regulation by PTM. This is indeed the case as many Nups are phosphorylated within disordered regions, resulting in the breakdown of the NPC during mitosis (Laurell et al., 2011). Non-mitotic phosphorylation of FG-Nups can also influence the permeability barrier of the NPC by decreasing the affinity between FG-repeats as well as decrease their affinity for NTRs (Kosako & Imamoto, 2010; Mishra et al., 2019). Overall, despite SLiMs interacting in a linear fashion they contribute significantly to the overall 3-dimensional structure of the NPC and are key regulators of NPC function.

1.3.3 Overview of nucleocytoplasmic transport cycle

Passage of proteins through the NPC is aided by NTRs called karyopherins, which can be functionally classified as either importins or exportins depending on the direction they

carry their respective cargo. It's also worth noting that not all karyopherins have been studied equally, therefore some of their roles are unclear. For example, exportin-7 has been historically assigned nuclear export functions, but has now recently been shown to be responsible for import of several cargos (Aksu et al., 2018). Most karyopherins belong to the highly conserved karyopherin-β (Kapβ) superfamily, which vary in number depending on the organism, ranging from 14 in the yeast *Saccharomyces cerevisiae* to at least 20 in humans (Harel & Forbes, 2004). Kapβ family members are composed 19-20 HEAT repeats, giving them an overall solenoid structure that allows them to interact directly with FG-Nups within NPC (Bayliss et al., 2002; Christie et al., 2016). Kapβ's can bind their cargo directly through the recognition of either a distinct NLS or nuclear export signal (Figure 1.6: left and right panel) (NES) (Xu et al., 2010). Alternatively, they interact with cargo via adapter karyopherins, such as importin-α (Imp-α), which also recognize distinct NLS sequences (Figure 1.6: middle panel) (Goldfarb et al., 2004b).

Interactions between karyopherins and their cargo is modulated through the binding of RanGTP, which is predominately localized within the nucleus, to the N-terminal HEAT repeats of Kapβ proteins (Vetter et al., 1999). The asymmetric localization of RCC1 (nuclear), a RanGTP exchange factor, and RanGAP1 (cytoplasmic), which stimulates the conversion of RanGTP to RanGDP, is responsible for establishing nuclear RanGTP and cytoplasmic RanGDP levels (Cavazza & Vernos, 2015; Stewart, 2007). Importantly, the distribution of RanGDP/GTP across the NE is essential for dictating the directionality of transport (D Görlich et al., 1996). Importin Kapβ's associate with their cargo in the cytoplasm and are rapidly transported through the NPC and subsequently release their cargo via conformational changes stimulated through binding of RanGTP (Christie et al., 2016). Conversely, exportin Kapβ's form a complex with RanGTP and their cargo within the nucleus and subsequent hydrolysis of RanGTP within the cytoplasm, stimulated by RanGAP1, causes the release of RanGDP and cargo dissociation (Bischoff et al., 1994). In order to drive further nuclear transport cycles, RanGDP is imported into the nucleus by NTF2, where RCC1 replenishes the RanGTP population (Ribbeck et al., 1998).
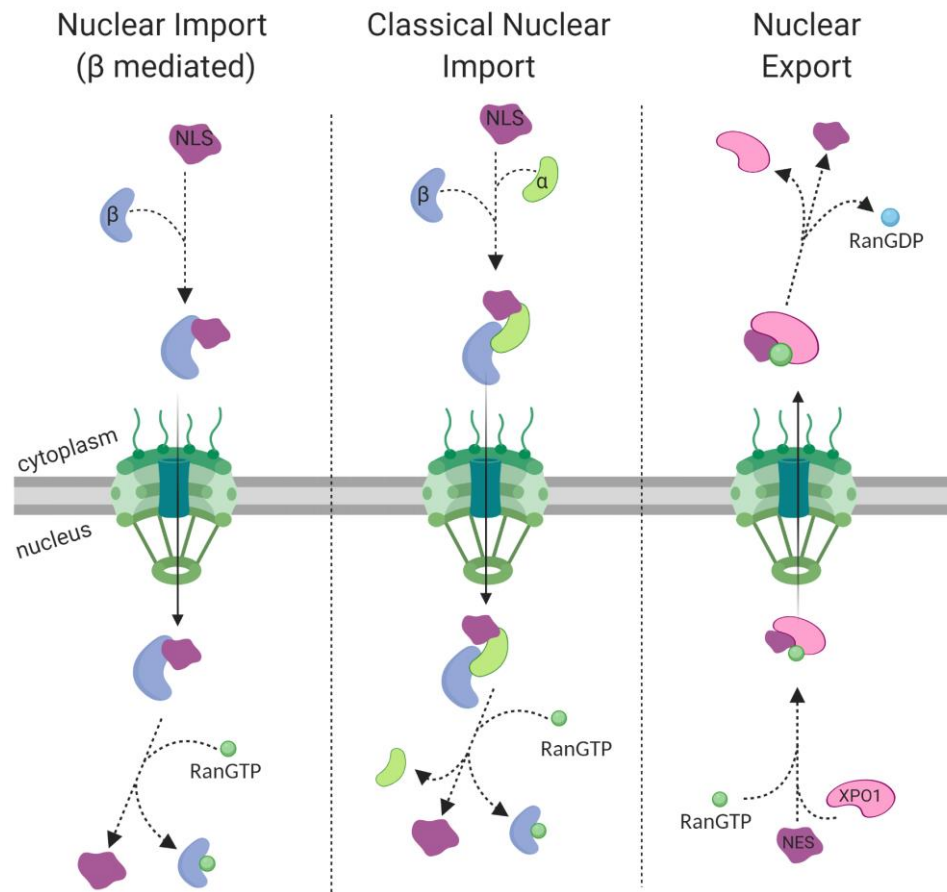
Figure 1.6 Protein nuclear import and export through the nuclear pore complex.

Proteins are transported through the nuclear pore complex (NPC) by binding importin or exportin karyopherins. Protein nuclear import is facilitated by importin karyopherin β's that recognize a proteins nuclear localization signal (NLS) within the cytoplasm (**left panel**). Of the importin karyopherin β's, Importin-β1(Imp-β1) recognizes it's cargo through the adapter Importin-α (**middle panel**). This pathway is formally known as the classical nuclear import pathway and is assumed to handle the majority of nuclear import. The nuclear import cycle is completed upon binding of RanGTP to Importin-β, causing cargo or Importin-α dissociation. Protein nuclear export primarily uses exportin-1 (XPO1), which recognizes a cargos nuclear export signal (NES), along with RanGTP, within the nucleus (**right panel**). In the cytoplasm, RanGTP hydrolysis is stimulated by the Ran GTPase activating protein 1 (RanGAP1), causing cargo release. Created with BioRender.

1.3.4   Imp-α dependent nuclear import: Classical nuclear import

Kapβ's are able to recognize their cargo directly through recognition of their NLS; however, Imp-β1 is unique in that it uses Imp-α as an adapter to bind cargo in addition to being able to bind cargo directly (Lam et al., 1999; Lange et al., 2007). This pathway has long been regarded as the classical nuclear import pathway, as both Imp-α and Imp-β1 were the first importins to be isolated and have thus contributed significantly to our mechanistic understanding of transport (Dirk Görlich et al., 1994; Lange et al., 2007; Moroianu et al., 1995). In this pathway, Imp-α recognizes a cargos NLS, which then allows Imp-β1 to bind Imp-α/NLS, forming a ternary complex that can move through the NPC (Figure 1.6: Middle panel). Components of the classical nuclear import pathway are also highly conserved across eukaryotes: *S. cerevisiae* express a single Imp-α, *D. melanogaster* express three isoforms, and the human genome encodes seven Imp-α isoforms (Pumroy & Cingolani, 2015). Importantly, the classical nuclear import pathway is generally regarded to handle the bulk of protein nuclear import.

Imp-α isoforms can be subdivided into three subfamilies known as α1, α2 and α3, all of which have a highly conserved architecture (Pumroy & Cingolani, 2015). The C-terminal portion of Imp-α isoforms are composed of 10 armadillo (ARM) repeats, which form the NLS-binding region while the N-terminal 70 amino acids function as an Imp-β binding (IBB) domain (Figure 1.7A) (Goldfarb et al., 2004a). The NLS-binding region is formed by two pockets between ARM repeats 2–4 and 6–8, which are referred to as the major and minor grooves, respectively (Fontes et al., 2000). Aside from linking Imp-α to Imp-β1, the IBB domain also plays an autoinhibitory role by binding these NLS-binding grooves, preventing futile import of empty Imp-α/β1 heterodimers (Kobe, 1999; Lott & Cingolani, 2011). In *S. cerevisiae*, the autoinhibitory function of the IBB domain is essential for survival; however, whether this is true for human Imp-α is unclear (Harreman, Hodel, et al., 2003). The autoinhibitory function of the IBB domain from different Imp-α isoforms can vary greatly. For example, Imp-α3 and -α7 have been shown to have reduced autoinhibition compared to Imp-α1 and this has been attributed to its affinity for the minor groove pocket (Pumroy et al., 2015; Pumroy & Cingolani, 2015). In general, efficient nuclear import requires complex interactions involving NLS recognition and Imp-β1
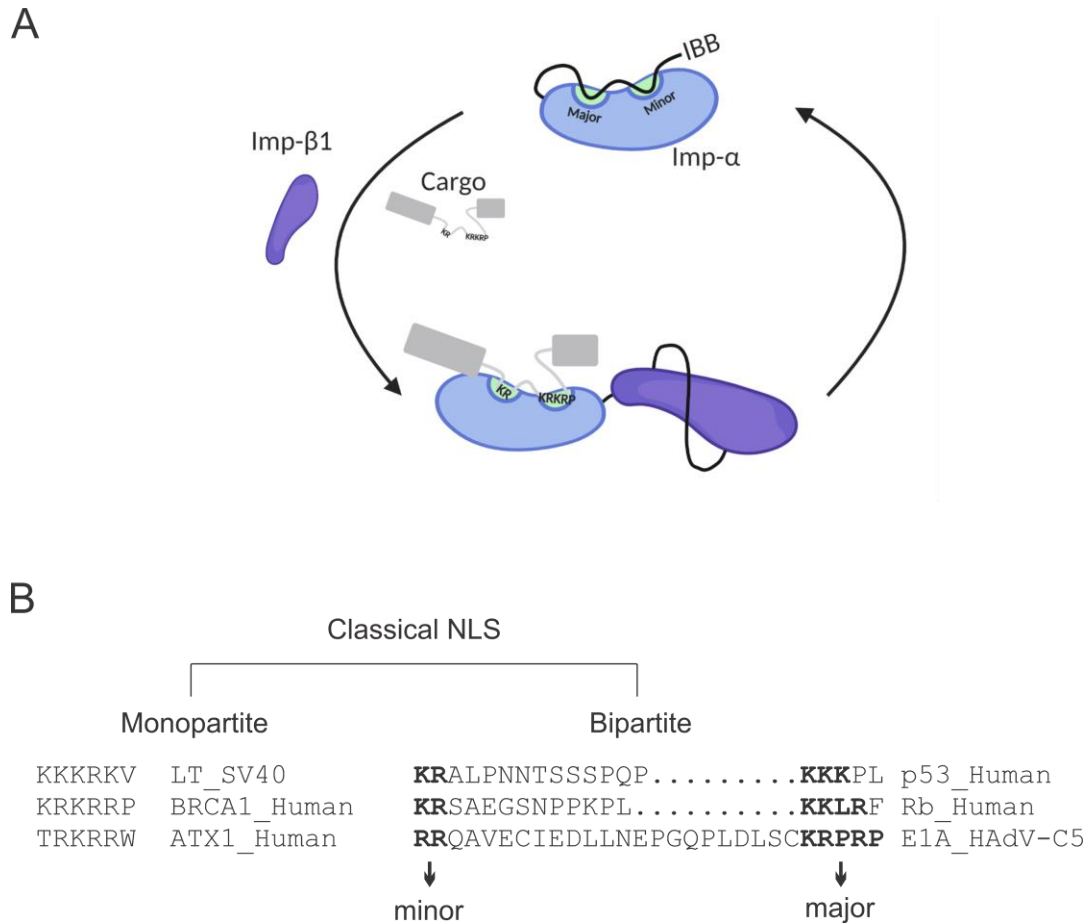
Figure 1.7 Classical nuclear localization sequence recognition by Importin-α.

(A) Importin-α (Imp-α) contains an N-terminal Importin-β binding domain (IBB) followed by a series of ARM repeats. Within the ARM repeat region two pockets are formed that recognize classical nuclear localization sequences (cNLS). ARM repeats 2–4 and 6–8 form the major and minor groove binding sites, respectively. When unbound, the ARM repeats of Imp-α are occupied by the IBB domain, preventing import of unloaded Imp-α. When recognizing a cargos cNLS, Imp-α is engaged by Importin-β1, forming a ternary complex that can shuttle through the nuclear pore complex. (B) cNLSs can be monopartite or bipartite with one or two clusters (in bold) of basic amino acids, respectively. Monopartite cNLSs typically bind the major binding groove of Imp-α, while bipartite cNLSs bind both the major and minor binding grooves. For bipartite cNLS recognition, the smaller basic cluster engages the minor site, while the larger C-terminal cluster engages the major site. Created with BioRender.

binding; however, the order in which these events occur may be dependent on additional factors including the IBB domain and/or NLS affinity. NLSs recognized by Imp-α are referred to as classical NLSs (cNLS) and represent a distinct class of SLiM. cNLSs are rich in basic amino acids and come in two forms, monopartite or bipartite (Figure 1.7B) (Lange et al., 2007). Monopartite cNLSs are comprised of a single cluster of basic amino acids while bipartite cNLS have two clusters of basic amino acids separated by a linker region of variable length and composition. These motifs can be exemplified by the SV40 TAg, PKKKRKV, and nucleoplasmin (NP), <u>KR</u>PAATKKAGQA<u>KKKK</u>, cNLS, respectively (D Kalderon et al., 1984; Robbins et al., 1991). Monopartite motifs primarily bind the major groove. However, several NLSs, both cellular and viral, can bind the minor groove exclusively (Kosugi et al., 2009; Lott et al., 2011; Nakada et al., 2015; Pang & Zhou, 2014). In contrast, bipartite cNLSs occupy both the major and minor grooves, with the smaller N-terminal basic cluster bound to the minor groove and the larger basic cluster bound to the major groove (Conti & Kuriyan, 2000).

1.3.5 Imp-α independent nuclear import: Non-classical import

Of the 20 or so human Kapβ proteins, roughly 10 mediate nuclear import, with an additional two (Importin-13 and exportin-4) capable of bidirectional transport (Kimura & Imamoto, 2014). Proteomic analysis of their cargos also suggests Kapβ members are linked to distinct cellular pathways (Kimura et al., 2017). However, unlike classical nuclear import mediated by Imp-α, only weak consensus NLSs have been defined for a few of these transport receptors (Chook & Süel, 2011; B. J. Lee et al., 2006; Maertens et al., 2014). Consequently, despite many cargos being identified for the Kapβ family members, including those from different species, little is known regarding the characteristics of the NLSs they recognize.

The most well characterized Kapβ import pathway is mediated by transportin 1 (TNPO1) and its yeast homolog Kap104p (Aitchison et al., 1996; Twyffels et al., 2014). This pathway is responsible for transporting numerous cargos into the nucleus, many of which are RNA binding proteins (Chook & Süel, 2011). TNPO1 recognizes proline-tyrosine NLSs (PY-NLS), which unlike compact and well-defined cNLSs, are defined by a set of

physical criteria and a loose consensus sequence (B. J. Lee et al., 2006; Soniat & Chook, 2015). The PY-NLS is composed of an N-terminal motif that can be either hydrophobic or basic and a C-terminal RX_{2-5}-PY motif. Due to the presence of either a hydrophobic or basic N-terminal region, the PY-NLS has been subdivided into two classes on this basis (Soniat & Chook, 2015). Additionally, these motifs are found within intrinsically disordered regions and have an overall positive charge. However, as more TNPO1 targets have been structurally resolved, additional variants have been identified that have a PL in place of PY, or even lack the PY motif altogether, highlighting the complexity of PY-NLSs (Soniat et al., 2013; Soniat & Chook, 2016). A highly similar Kapβ, named transportin 2 (TNPO2), was coincidentally discovered during the cloning of TNPO1; however, due to their similarity the majority of structural and biochemical work has been focused on TNPO1. An additional transportin, transportin 3 (TNPO3), has been shown to import cargos with arginine-serine (RS) repeat NLSs (Kataoka et al., 1999). Cargos of TNPO3 tend to be serine/arginine-rich (SR) splicing factors, which contain RS domains that consist of up to 50 RS dipeptide repeats that can be recognized by TNPO3 (Chook & Süel, 2011). However, it appears that many RS domains need to be phosphorylated (Yun et al., 2003). Furthermore, recent analysis of Cold Inducible RNA Binding Protein (CIRBP), which has no cNLS or PY-NLS, shows that it is able to bind both TNPO1 and TNPO3 (Bourgeois et al., 2020). Investigation of CIRBP showed that an RG/RGG (arginine-glycine) and RSY (arginine-serine-tyrosine) motif confers binding to TNPO1 and TNPO3, respectively, and binding is modulated by arginine methylation of the RG/RGG motif.

Additional transporters that function analogously to Imp-α as an adapter have also been shown to mediate nuclear import. One example is that of snurportin, which also functions as an Imp-β1 adapter, but is involved in the import of uridine-rich small ribonucleoprotein (U snRNP) (Mitrousis et al., 2008; Strasser et al., 2005). Snurportin has an N-terminal IBB domain like Imp-α; however, no other structural similarities exist (Huber et al., 1998). Fascinatingly, import by snurportin-1/Imp-β1 appears to be Ran-independent (Huber et al., 2002). An additional adapter responsible for the synchronous nuclear import of yeast ribosomal proteins Rpl5 and Rpl11, named symportin-1 (Syo1), has more recently been discovered (Bange et al., 2013; Kressler et al., 2012). Syo1 acts as an adapter to Kap104 (yeast homolog of TNPO1) by binding through an N-terminal PY-NLS. Interestingly,

structurally it appears to be a hybrid of Imp-α and Kapβ, as it contains a series of ARM repeats followed by a series of HEAT repeats (Kressler et al., 2012). Additional work on Syo1 is still limited to yeast and whether or not other cargo, or a functional human homolog exists remains unclear.

1.3.6 Alternative nuclear import pathways

Several unique nuclear import pathways that do not use Imp-α or any Kapβ have more recently been discovered. The first pathway utilizes an evolutionarily conserved protein called Hikeshi, which is responsible for nuclear import of Hsp70 during heat shock induced stress (Kose et al., 2012). During the heat shock response, Imp-α is retained within the nucleus, resulting in downregulation of the classical nuclear import pathway (Furuta et al., 2004). Additionally, RanGTP distribution is altered during cellular stressors, possibly inhibiting Kapβ mediated transport as well (Kelley & Paschal, 2007). However, during the heat shock response Hsp70 is imported into the nucleus by Hikeshi, which binds the NPC directly (Kose et al., 2012). The second novel nuclear import pathway was discovered through studies of the RanGDP nuclear import pathway mediated by NTF2 (Lu et al., 2014). Here, it was shown that several ankyrin repeat (AR) containing proteins are able to bind RanGDP via their ARs and complex with NTF2 indirectly.

Additional evidence of alternative nuclear import mechanisms has also come out of studying the classical nuclear import pathway. Bioinformatic analysis of nuclear proteins from several organisms has shown a large number of proteins do not have a predictable NLS (Bernhofer et al., 2017; Lange et al., 2007; Tessier et al., 2020). In addition, up to 50% of proteins that interact with Imp-α in yeast do not have a predictable cNLS (Lange et al., 2007). These examples highlight the possible existence of novel NLSs that can bind Imp-α directly, or piggyback on Imp-α cargo. In fact, both scenarios have already been shown to exist. For example, it's been shown that Senataxin and Smarca4 are able to bind Imp-α directly at ARM repeats 9-10, demonstrating that additional binding sites other than the major and minor grooves exist (Arjomand et al., 2014). Additionally, nuclear import of TAF10, a component of transcription factor II D, requires TAF3 or TAF8 to bind Imp-α/β1 (Soutoglou et al., 2005). Since TAF10 does not contain an NLS, piggybacking on TAF3 and TAF8 is required for efficient nuclear import.

1.3.7 Protein nuclear export

Protein nuclear export mainly utilizes the Kapβ CRM1, also known as exportin-1 (Xpo1), which recognizes leucine-rich NESs in the presence of RanGTP. These sequences are generally 8–15 amino acids long and contain 4–5 regularly spaced hydrophobic residues (Fornerod et al., 1997; Xu et al., 2012). In fact, the first NESs identified were of viral origin, and were originally discovered within the Rev proteins of HIV-1 and other lentiviruses (B. E. Meyer et al., 1996). Hydrophobic residues within an NES act as anchors that bind a hydrophobic pocket formed by HEAT repeats 11 and 12 of CRM1 (Fung et al., 2015, 2017). Interestingly, NES binding to CRM1 is generally conformationally unrestrained, as there is lack of contact with the NES backbone that allows hydrophobic anchor residues to bind in a variety of conformations. In addition, NESs are unusual in that they are able to bind CRM1 in either the N- to C-terminal or C- to N-terminal orientation, further enhancing the diversity of potential NESs (Fung et al., 2015, 2017).

Proteomic analysis of CRM1 cargo suggests it has a major role in helping maintain the separation of nuclear and cytoplasmic proteins (Kırlı et al., 2015). The NPC is not a perfect barrier, allowing cytoplasmic proteins to leak in, therefore, having active nuclear export pathways to sort these proteins, such as translation factors, back to the cytoplasm is integral to maintaining many cellular processes. Several other members of the Imp-β family are associated with nuclear export of protein or RNA, and these include exportin-2, exportin-5, exportin-6, and exportin-7, as well as bidirectional transporters importin-13 and exportin-4 (Kimura & Imamoto, 2014). However, in contrast to CRM1, the cargo proteins recognized by these karyopherins have not been as extensively studied (Kimura & Imamoto, 2014).

Nuclear export of mRNA on the other hand is typically handled by NXF1, which is not a karyopherin member, via the transcription-export (TREX) adaptor (Williams et al., 2018). During the early stages of mRNA biogenesis, TREX associates with the 5' end of mRNA. Following maturation, the TREX-mRNA complex can be exported via NXF1 (Cheng et al., 2006). Additionally, some mRNA also use CRM1-mediated export; however, CRM1 does not bind mRNA directly and instead uses adapters (Williams et al., 2018). Other subtypes of RNA, such as microRNA, can be directly bound by exportins, as is the case

with exportin-5 which represents the canonical microRNA export pathway (Bohnsack et al., 2004). Similarly, exportin-t is responsible for the export of 5'- and 3'-end processed tRNA and can bind those RNAs directly (A. Gupta et al., 2016).

1.4 Viral Manipulation of Protein Nuclear Transport

Each component of the nuclear transport system potentially represents a viable target that can be appropriated by a virus during infection to allow entry of viral genomic information, export of viral mRNA, and passage of viral proteins bidirectionally across the NE (Figure 1.8A–C). A common theme among many viruses is their limited coding capacity and, therefore, their absolute dependence on host proteins and pathways for a productive infection. Given this, it is unsurprising that viral proteins have evolved ways of interacting with the many components that make up the nuclear transport system.

1.4.1 Viral protein-karyopherin interactions

The most direct approach for viral proteins to traverse the NE is to target the NPC itself. Generally, this phenomenon is reserved for capsid interactions to bring viral genomic information into the nucleus. Herpes simplex virus 1 (HSV-1) UL36 is a preformed tegument protein, which aids in docking the viral capsid to the NPC by bridging the capsid with Nup358 of the NPC (Copeland et al., 2009). Similarly, the capsid protein of HIV-1 interacts with Nup153 to mediate import of the preinitiation complex (PIC) (Matreyek et al., 2013). For the most part, genomic studies have been primarily responsible for identifying components of the NPC important for the life cycle of many viruses; however, because of the nature of these experiments, it is often unclear which interactions directly involve the NPC (Le Sage & Mouland, 2013). Some viral proteins interact with the NPC directly, and several instances of viral proteins that are not components of the capsid directly binding the NPC have been documented. These include BGLF4 from EBV and HIV-1 Vpr (Chang et al., 2012; Fouchier et al., 1998; Y. Jenkins et al., 1998). *In vivo* and *in vitro* experiments demonstrated that the C-terminus of BGLF4 can directly associate with Nup62 and Nup153. Vpr, on the other hand, interacts with a poorly characterized nucleoporin CG1, also known as nucleoporin-like protein 2 (NUPL2) (Le Rouzic et al., 2002). During HIV-1 infection, Vpr is essential for nuclear import of the PIC, an essential
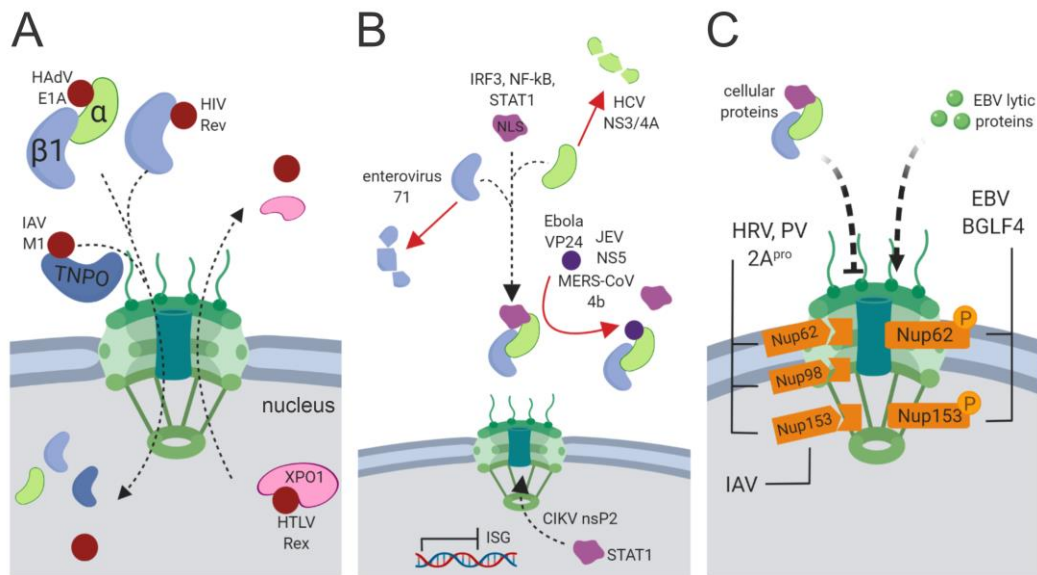
Figure 1.8 Viral manipulation of host nuclear transport pathways.

Viruses can use a number of strategies to manipulate host nuclear transport pathways. This allows them to achieve transport of viral proteins and/or perturb transport of host cellular proteins. **(A)** Viruses are known to utilize the classical Imp-α/β1 pathway as well as bind Imp-β1 directly, the nuclear pore complex (NPC), or transportin through a PY-nuclear localization signal (PY-NLS) for nuclear import as well as CRM1 for nuclear export. **(B)** Preventing nuclear import, or promoting export, of cellular proteins such as signal transducer and activator of transcription 1 (STAT1), interferon regulatory factor 3 (IRF3), or nuclear factor-kappa B (NF-κB) can block the antiviral innate immune response. **(C)** Viral proteins can perturb global nuclear transport by altering the dynamics of the NPC though the degradation or phosphorylation of nucleoporins (Nups). Created with BioRender.

step for replication in nondividing cells that also involves Vpr binding Imp-α (Bukrinsky et al., 1992; Kamata et al., 2005; Nitahara-Kasahara et al., 2007). Docking the PIC at the NPC is facilitated by Vpr's ability to bind the NPC through CG1 as well as Imp-α, essentially making it a functional mimic of Imp-β1 (Vodicka et al., 1998). A simple yet effective approach to targeting cellular importins is through molecular mimicry of a cNLS, which would allow viral proteins to interact with Imp-α (Figure 1.8A). Because of the sequence characteristics and predictability of these peptide motifs, many viral cNLSs have been discovered in a diverse range of viruses. For example, IAV nucleoprotein and PB2, HIV-1 integrase and Vpr, HAdV E1A, HPV E2, HSV-1 pUL30, and many more contain viral cNLSs (Alvisi et al., 2007; Ao et al., 2010; Bian & Wilson, 2010; Cohen et al., 2014; Hudjetz & Gabriel, 2012; Kohler et al., 2001; Nitahara-Kasahara et al., 2007; Pumroy et al., 2015).

An alternative approach for viral proteins to target the nuclear import machinery is to directly bind Imp-β (Figure 1.8A). Multiple viruses have been shown to bind Imp-β1 directly, surpassing the need for the adapter Imp-α. HIV-1 Rev was the first identified example of this, although similar examples can be seen with HIV-1 Tat, human T-cell leukemia virus (HTLV) Rex, HSV-1 capsid protein, hepatitis B virus (HBV) core protein, and HAdV protein VII (Arnold et al., 2006; C. Chen et al., 2016; Henderson & Percipalle, 1997; Ojala et al., 2000; Palmeri & Malim, 1999; Truant & Cullen, 1999; Wodrich et al., 2006). Additionally, there are other members of the Imp-β family, namely transportins, which can mediate import of proteins into the nucleus. The IAV M1 and human cytomegalovirus (HCMV) UL79 proteins both contain PY-NLSs that allow their interaction with the transportin nuclear import pathway (Miyake et al., 2019; Wang et al., 2012). Many viral proteins have been shown to utilize transportin; however, many have no identified PY-NLS. Specifically, HAdV core proteins pV and pVII, HIV-1 Rev, as well as HPV L2 and E6 have all been shown to interact with the transportin pathway, though none of these have identified PY-NLSs (Arnold et al., 2006; Darshan et al., 2004; Hindley et al., 2007; Le Roux & Moroianu, 2003; Wodrich et al., 2006).

1.4.2 Viral immune evasion through manipulation of nuclear import

A critical step in the viral replication cycle is avoiding detection by the immune system and one way this is frequently achieved is through disruption or manipulation of cellular protein trafficking (Figure 1.8B). Central to the innate antiviral immune response is the activation of the type I interferon (IFN) and NF-κB signaling pathways, which lead to the expression of IFN stimulated genes (ISGs) and the production of an array of proinflammatory cytokines that ultimately work to halt or delay viral replication. The production of type I IFN is triggered by the relocalization of interferon regulatory factor (IRF) 3, IRF7, and/or NF-κB transcription factors from the cytoplasm to the nucleus. The canonical type I IFN response involves activation of the Janus kinase (JAK) and the signal transducer and activator of transcription (STAT) pathways upon binding of IFN to its receptor. This triggers translocation of activated STAT proteins into the nucleus resulting in transcription of hundreds of ISGs (Ivashkiv & Donlin, 2014).

One set of components that can be targeted to block innate antiviral responses are the karyopherin proteins, which are directly responsible for cargo recognition. An example of this is the hepatitis C virus (HCV) NS3/4A protease, a well-studied protein involved in evading innate immunity (Morikawa et al., 2011). NS3/4A functions by interacting with and subsequently cleaving Imp-β1, resulting in its loss of function (Gagne et al., 2017). This study also demonstrated that Imp-β1 was the main nuclear import receptor for IRF3 and NF-κB, which are key transcription factors controlling antiviral innate immunity as described above. Thus, cleavage of Imp-β1 reduces nuclear transport of IRF3 and NF-κB p65 at early time points during infection, effectively delaying the IFN response.

A less direct way of blocking nuclear translocation of cellular proteins involved in innate immune signaling is through simple competition for Imp-α. Ebola virus VP24 can compete with tyrosine-phosphorylated STAT1, whose nuclear import is essential for transcriptional activation of ISGs, for binding to Imp-α5, α6, and α7 (Mateo et al., 2010; Reid et al., 2007). Japanese encephalitis virus NS5 protein is able to bind Imp-α3 and α4 to compete with IRF3 and NF-κB p65 binding, preventing their nuclear import and limiting their ability to activate IFN signaling (Ye et al., 2017). More recently it was shown that the 4b protein of Middle East respiratory syndrome coronavirus (MERS-CoV) is able to out compete NF-

κB p65 for binding to Imp-α3 (Canton et al., 2018). Beyond the examples listed here, numerous other viruses have evolved similar mechanisms to target importins, allowing them to evade host innate immunity. This can include sequestering importins in the cytoplasm and even expressing microRNAs that downregulate Imp-α expression (Frieman et al., 2007; Y. Liu et al., 2016; Yarbrough et al., 2014).

1.4.3 General disruption of host nucleocytoplasmic transport

Beyond evading innate immunity, the viral replicative cycle may require localization of cellular proteins to subcellular compartments different from their normal localization. In some cases, relocalization is specific to certain proteins, while in others this is implemented broadly, effecting a multitude of proteins (Figure 1.8C). Picornaviruses, such as human rhinovirus (HRV) and poliovirus (PV), target Nup153, a component of the NPC nuclear basket, as well as the FG-rich nucleoporins Nup62 and Nup98 and proteolytically cleave them via virally encoded 2A proteases (2A^{pro}) (Gustin & Sarnow, 2001, 2002; Watters et al., 2017). More specifically, 2A^{pro} from HRV and PV specifically cleaves the FG-rich region from Nup62, a region involved in recognizing karyopherins and forming the inner channel of the NPC (Park et al., 2010).

Although less well-studied, other viruses such as IAV, EBV, and Venezuelan equine encephalitis virus (VEEV) have similarly evolved strategies for targeting components of the NPC. EBV, a DNA virus, encodes the Ser/Thr protein kinase BGLF4, which induces phosphorylation of Nup62 and Nup153 (Figure 1D) (Chang et al., 2015). In the presence of BGLF4, nuclear targeting of Imp-β1 was attenuated, broadly inhibiting cNLS-mediated nuclear import. In contrast, the nuclear import of several non-NLS containing EBV lytic proteins was promoted in the presence of BGLF4. IAV, an RNA virus that replicates within the nucleus, induces enlargement of the NPC by exploiting cellular caspase activity (Muhlbauer et al., 2015). Caspase activation at later timepoints during infection results in degradation of Nup153, altering the structural integrity of the NPC and likely aiding in passive diffusion of ribonucleoprotein complexes across the nuclear envelope.

Establishing a state conducive to viral infection can also involve precise nucleocytoplasmic redistribution of select cellular proteins, a process best demonstrated by the HAdV E1A

protein. E1A itself has no intrinsic DNA binding or enzymatic capabilities and therefore relies on host proteins to carry out such functions (King, Zhang, et al., 2018). During HAdV infection, E1A interacts with the regulatory RIα and RIIα subunits of protein kinase A (PKA) and preferentially relocalizes them to the nucleus from the cytoplasm. Normally, the subcellular localization of PKA is regulated by cellular A-kinase anchoring proteins (AKAP). However, during infection, E1A appropriates this role through direct competition with cellular AKAPs by functioning as a viral AKAP itself (King, Cohen, et al., 2016; King, Gameiro, et al., 2018).

1.5   Thesis Overview

The central theme of the work presented within this thesis is focused on the analysis of SLiM-mediated PPIs. SLiMs are involved in a diverse range of cellular processes (Van Roey et al., 2014); however, the interactions under investigation here mainly involve those mediating protein nuclear import. I first show that targeting SLiM-mediated interactions in the classical nuclear import pathway is a viable antiviral strategy. Using HAdV-C5 as a model, I showed that ivermectin, an inhibitor of the classical nuclear pathway, inhibits viral replication. Additionally, I demonstrate ivermectin inhibits the interaction between E1A and Imp-α. Secondly, I tried to determine to what extent proteins piggyback into the nucleus, as the role of piggybacking in the classical nuclear import pathway remains understudied. To address this, I analyzed currently available PPI databases and datasets and showed that many nuclear proteins, which bind Imp-α, do not have an identifiable cNLS. Finally, building on the theme that many nuclear proteins do not have an identifiable NLS, I designed a novel yeast-based genetic system to identify SLiMs by exploiting the nuclear import process. Importantly, this system has several advantages over current experimental SLiM discovery approaches.

1.5.1   Chapter 2: Inhibition of human adenovirus replication by the importin α/β1 nuclear import inhibitor ivermectin

In this study, we explored the effect of the drug ivermectin on HAdV replication. Ivermectin is an inhibitor of the classical nuclear import pathway and is proposed to inhibit the interaction between Imp-α and Imp-β1 (Wagstaff et al., 2012). All HAdV utilize the

classical nuclear import pathway, particularly the E1A protein, which possesses a potent bipartite cNLS (Cohen et al., 2014; Lyons et al., 1987). Therefore, targeting an important host cellular process, like nuclear import, could represent an effective anti-viral approach. Using HAdV-C5 as a model, we show that ivermectin inhibits viral progeny production in a dose-dependent manner. Subcellular localization studies of HAdV-C5 infected cells treated with ivermectin showed an abrogation of nuclear import for key viral proteins, including E1A and DBP. Investigation of viral gene expression and protein synthesis during HAdV-C5 infection revealed a reduction of several viral early gene transcripts, as well as reduced early and late protein production. Additionally, viral genome replication efficiency was determined and ivermectin was shown to inhibit replication of both HAdV-C5 and -B3. Interestingly, co-immunoprecipitation experiments of E1A shows that ivermectin inhibits the Imp-α/cNLS interaction, and not the Imp-α/β1 interaction as previously proposed.

1.5.2 Chapter 3: Piggybacking on classical import and other non-classical mechanisms of nuclear import appear highly prevalent within the human proteome

How nuclear proteins without a cNLS use the classical nuclear import pathway remains an outstanding question within the field. Previous studies have established that many nuclear proteins in humans and yeast do not have a cNLS, and many yeast proteins that interact with Imp-α also do not possess a cNLS (Bernhofer et al., 2017; Lange et al., 2007). The focus of the work in this chapter aimed to estimate how widespread is the use of alternative mechanisms of nuclear import, such as piggybacking on classical nuclear import machinery. Importantly, I used currently available datasets and databases that were either not available at the time of previous studies or have not been used since. Using a list of nuclear proteins from The Human Protein Atlas I show that nearly 50% of human nuclear proteins do not have predicted mono or bipartite cNLS. For the first time, I also showed that 20–50% of cargo for each of the human Imp-α isoforms do not have a predictable cNLS. Using recently published data from the Human Reference Interactome project, I confirmed that many Imp-α cargo do not have an identifiable cNLS. In order to extend our analysis to Imp-α cargo that are not available in public interaction databases, I reanalyzed publicly available raw mass spectra files, which identified hundreds of Imp-α cargo. Many

of these cargos do not have an identifiable cNLS and interestingly, several of these proteins have already been shown to utilize piggybacking. Furthermore, I determined that the majority of proteins belonging to the Mediator complex do not have a cNLS yet bind Imp-α. Based on these findings I propose that Mediator utilizes piggybacking mechanisms.

1.5.3 Chapter 4: A novel protein nuclear import-based methodology for discovery of short linear motifs

In the final chapter, my goal was to exploit the nuclear import process in yeast to discover novel SLiMs. Using a genetic selection system, I expressed a library of recombinant proteins that were too large to diffuse into the nucleus and were fused with a randomly generated 10 amino acid peptide. If a peptide was able to mediate nuclear import, either through directly binding import machinery or indirectly piggybacking, it could be selected for genetically. This system is sensitive to many different cNLSs, as well as bipartite cNLSs. In a small-scale proof-of-principle experiment I identified several peptides that can mediate nuclear import in yeast. Expressing several of these peptides in a mammalian system as enhanced green fluorescent protein (EGFP) fusions showed nuclear localization, and this was confirmed using nuclear/cytoplasmic fractionation. Additionally, using co-immunoprecipitation I show that some of these peptides can interact with Imp-α.

Chapter 2

2   Inhibition of human adenovirus replication by the importin α/β1 nuclear import inhibitor ivermectin

2.1   Introduction

The continuous flow of molecules between the cytoplasm and nucleus of eukaryotic cells is essential for many cellular processes. Transport of macromolecules, especially proteins, across the nuclear envelope is a highly regulated process that requires passage through the NPC (Bauer et al., 2015; Cautain et al., 2015). As protein size increases, so does the difficulty in passing through the NPC. Larger proteins, typically over 40–50 kDa, require active transport mechanisms. This includes the classical protein nuclear import pathway which utilizes the Imp-α and Imp-β1 proteins. Imp-α recognizes a proteins NLS, while Imp-β1 serves to bridge the Imp-α/NLS complex with the NPC (Lange et al., 2007). Seven Imp-α isoforms exist in humans, all possessing a conserved IBB domain located on their N-terminus (Cingolani et al., 1999; Kelley et al., 2010; Pumroy & Cingolani, 2015). The IBB domain forms an intramolecular interaction with the NLS binding groove, preventing binding of Imp-β1 to an unloaded Imp-α, reducing futile nuclear translocation of empty import complexes (Lott & Cingolani, 2011). The Imp-α/β1 pathway defines classical nuclear import, where Imp-α recognizes NLSs containing a cluster of basic amino acids. These are collectively referred to as cNLSs (Kelley et al., 2010; Lange et al., 2007; Mason et al., 2009). By coevolving with eukaryotic hosts, viral pathogens have developed diverse mechanisms to usurp this important and highly coordinated pathway. These include mimicry of cellular NLSs by viral proteins and manipulation of these host factors at the molecular level (Tessier et al., 2019).

HAdVs are ubiquitous in the human population, particularly in children and young adults, and contribute to a significant portion of respiratory and gastrointestinal illnesses worldwide (Ghebremedhin, 2014; King, Zhang, et al., 2016; Lion, 2014). The most common HAdV species associated with disease are HAdV-C, -B and -E (Lion, 2014). In particular, HAdV-B3 and -B7 are commonly associated with acute respiratory illness and have been reported in several outbreaks worldwide (Jin et al., 2013; Lai et al., 2013; J. Lee

et al., 2010; W. J. Lee et al., 2015; Wo et al., 2015). HAdV-B55 has recently emerged as an epidemic strain in both Europe and Asia, causing acute respiratory illness in adults and outbreaks within civilian and military populations (Ko et al., 2019; Lafolie et al., 2016; Salama et al., 2016; Walsh et al., 2010). Additionally, immunocompromised individuals, such as transplant recipients, are at particular risk for severe illness or death from numerous HAdVs, including HAdV-C, -A, and -B (Khanal et al., 2018; Lion, 2014). Aside from cidofovir, a nucleotide analogue approved for cytomegalovirus-induced retinitis, and the cidofovir derivative brincidofovir, there are no antiviral drugs with clinically relevant activity against HAdV (Alvarez-Cardona et al., 2020; Khanal et al., 2018). Notably, HAdV gene expression, genome replication and virion assemble all take place within the nucleus and many of its gene products carry out crucial functions in this compartment to enhance its viral replication cycle (Berk, 2013; Pied & Wodrich, 2019). This raises the possibility that drugs interfering with nucleocytoplasmic transport might inhibit HAdV replication. In addition to providing new molecular insights into the replication stages of HAdV infections, understanding the relationship between HAdV and host nuclear import machinery may identify target points for anti-adenoviral therapies.

Ivermectin is a broad-spectrum anti-parasitic approved for use in humans to treat a variety of neglected diseases, such as onchocerciasis and lymphatic filariasis, and has been mass administered for the treatment of scabies (Babalola, 2011; Laing et al., 2017; Romani et al., 2015; Strycharz et al., 2008; Victoria & Trujillo, 2001). Through a high-throughput drug screening approach, ivermectin was also identified as a general inhibitor of Imp-α/β1-mediated nuclear import (Wagstaff et al., 2011). Several recent studies have explored ivermectin as a means of abrogating nuclear localization of viral proteins and as an inhibitor of viral replication (Tessier et al., 2019). Several RNA viruses, including HIV-1, dengue virus (DENV), and Hendra virus (HeV) were potently inhibited *in vitro* by ivermectin (Atkinson et al., 2018; Tay et al., 2013; Wagstaff et al., 2012). While findings on ivermectin's effects on DNA viruses are more limited, it has been shown to impair BK polyomavirus (BKPyV) infection *in vitro* (Bennett et al., 2015).

In this study we sought to extend the exploration of ivermectin's antiviral activity by examining its effects on HAdV-C5 infection. We found that ivermectin inhibits the overall

production of infectious HAdV-C5 progeny in a dose-dependent fashion. Multiple viral proteins produced early during infection exhibited impaired nuclear localization. Ivermectin treatment led to severely reduced levels of several HAdV-C5 mRNAs, protein products, and progeny genomes post-infection. In addition, we show that ivermectin targets the Imp-α/NLS interaction without disrupting Imp-β1 binding. Together, these findings offer insight into ivermectin's inhibitory effects on HAdV replication, as well as provide new mechanistic details regarding ivermectin's mode of action.

2.2 Materials and Methods

2.2.1 Cell lines, cell culture, and transfections

Human A549 (provided by Russ Wheeler, Molecular Pathology/Genetics, London Health Sciences Centre), HEK293, and HT-1080 (purchased from the American Type Culture Collection) cells were grown at 37°C with 5% CO_2 in Dulbecco's modified Eagle's medium (DMEM; Multicell Technologies) supplemented with 1% penicillin/streptomycin (Multicell) and 10% fetal bovine serum (FBS; Gibco). Transfections of DNA into HT-1080 cells were done using X-tremeGENE HP (Roche) using a 2:1 ratio of X-tremeGENE HP per μg of DNA. For treatment with ivermectin (MilliporeSigma, I8898) on transfected HT-1080 cells, media was replaced 16 hours post-transfection with fresh DMEM containing 10% FBS and the indicated concentration of drug (or the corresponding volume of DMSO as a control) and left for 1.5 hours prior to downstream harvesting for microscopy or immunoprecipitation.

2.2.2 Viruses and infection of cells

Wild-type (WT) HAdV-C5 (*dl309*) was previously described (N. Jones & Shenk, 1979).WT serotypes of HAdV-E3 (strain GB, lot 11W), -E4 (strain RI-67, lot 3W) were purchased from the ATCC via Cedarlane. A549 cells were infected at a multiplicity of infection (MOI) of 5 plaque-forming units (PFU)/mL. Cell cultures were infected at ~50% confluence for 1 hour at 37°C, and sub-confluent cells were collected at the indicated timepoints for downstream experiments. For plaque assays, confluent HEK293 cells infected with serially diluted samples for 1 hour at 37°C before being overlaid with DMEM containing 1% SeaPlaque agarose (Lonza). For treatment with ivermectin on infected A549

cells, infectious media was replaced after 1 hour with fresh DMEM containing 10% FBS and the indicated working concentrations of drug (or the corresponding volume of DMSO as a control) for the duration of the experiment.

2.2.3 Cell viability assay

Cells were seeded in 96-well plates at 2500 cells per well and left to adhere overnight. Media containing ivermectin was added at concentrations ranging from 10 nM to 150 µM and left for 72 hours before assessing viability. For each condition, three replicates were used. Viability was indirectly measured using PrestoBlue reagent (ThermoFisher). Normalized relative fluorescent units (RFUs) of the ivermectin-treated replicates were calculated as a percentage of the mean RFU of control (DMSO-only) treatment replicates. IC50 values, defined as the concentration at which the normalized RFU reached 50%, were calculated by non-linear regression (Prism® 8 Graphpad Software, Inc).

2.2.4 Plasmids

All constructs were expressed in vectors under the control of the HCMV promoter. Full-length E1A and TAg NLS constructs were built into pEGFP-C2. FLAG-tagged Qip1 was cloned into pcDNA3 and contains a full-length WT importin-α3 (KPNA4) and was previously described (Marshall et al., 2014). For a list of plasmids used in this chapter see Table 2.1.

2.2.5 Co-immunoprecipitation and western blotting

Cells were pretreated with 25 µM ivermectin for 1.5 hours prior to lysis. Transfected HT-1080 cells from 10cm plates were collected and lysed in 500 µL NP-40 lysis buffer (150 mM NaCl, 50 mM Tris-HCl pH 7.4, 1 mM EDTA, 1 mM EGTA, 0.1% NP-40, 10% glycerol) supplemented with protease inhibitor cocktail (MilliporeSigma, P8340) along with two freeze-thaw cycles on dry ice. Co-immunoprecipitation reactions were carried out at 4°C for 4 hours in lysis buffer supplemented with either 25 µM ivermectin or an equivalent volume of DMSO, using 20µl of washed magnetic FLAG beads (MilliporeSigma, M8823).

Table 2.1 Plasmids used in this chapter.

Clone	Vector	Lab ID
EGFP	pEGFP-C2	JMB4616
TAg NLS	pEGFP-C2	JMB4632
13S E1A	pEGFP-C2	JMB1449
FLAG/HA-Qip1	pcDNA3	JMB4390

Two percent of the sample was kept as input control. After washing with lysis buffer, samples were boiled in 25 µL of 2X LDS sample buffer (Thermofisher, NP0007) supplemented with DTT, for 5 minutes. Samples were separated on NuPage Bis-Tris gradient protein gels (Life Technologies, NP0321BOX) and transferred onto a polyvinylidene difluoride membrane (Amersham). Membranes were blocked in 5% skim milk constituted in Tris-buffered saline (TBS) with 0.1% Tween 20. Primary antibodies used include rabbit α-EGFP, mouse α-E1A, mouse α-DBP, mouse α-karyopherin β1, rabbit α-HAdV-C5 capsid (Table 2.2). Horseradish peroxidase-conjugated secondary antibody was detected using Luminata Crescendo or Forte substrate (Millipore).

2.2.6 Immunofluorescence microscopy

Cells were fixed in 3.7% paraformaldehyde at room temperature for 20 minutes, permeabilized on ice using 0.2% Triton X-100, and blocked using 3% BSA in phosphate-buffered saline (PBS). Samples were incubated in primary antibody that included rabbit α-EGFP, mouse α-E1A, mouse α-DBP (Table 2.2) for 1 hour at room temperature and another 30 minutes at room temperature with secondary antibody (Alexa Fluor 488; Life Technologies). Samples were mounted with Prolong Gold reagent containing 4',6-diamidino-2-phenylindole (DAPI; Life Technologies). Confocal images were acquired using a Fluoview 1000 laser scanning confocal microscope (Olympus Corp).

2.2.7 Reverse transcription and qPCR

Total RNA was prepared using the PureLink RNA Mini Kit (ThermoFisher) including on-column PureLink DNase treatment. RNA was reverse-transcribed into cDNA using the Superscript VILO Mastermix (ThermoFisher). Relative cDNA levels were measured by qPCR using Power SYBR green (ThermoFisher) with oligonucleotide sequences that specifically recognize HAdV-C5 E1A, E1B, E2, E3, E4. GAPDH was used as an endogenous cellular control for total cDNA along with no-RT and no-template negative controls. All qPCR primers used in this chapter are listed in Table 2.3. Results were normalized to cellular GAPDH, calculated using the ΔΔCt method, and set as relative to DMSO-treated samples.

Table 2.2 Primary antibodies used in this chapter.

Reactivity	Purpose	Description	Supplier
EGFP	IF, Western	Rabbit polyclonal	Takara (Living Colors)
M73 (E1A)	IF, Western	Mouse monoclonal	In-house
B68 (DBP)	IF, Western	Mouse monoclonal	In-house
HAdV capsid	Western	Rabbit polyclonal	Abcam
Actin	Western	Rabbit polyclonal	MilliporeSigma
FLAG (F1804)	Western	Mouse monoclonal	MilliporeSigma
FLAG (M8823)	IP	Mouse monoclonal	MilliporeSigma
Kpnb1 (H-7)	Western	Mouse monoclonal	Santa Cruz
Tubulin	Western	Mouse polyclonal	MilliporeSigma

2.2.8 Virus replication assay

A549 cells were infected with either HAdV-C5, -B3 or -E4 and then supplemented with media containing 10 µM ivermectin. Total cell DNA was purified at 6- and 48-hours post-infection using a DNeasy Blood and Tissue kit (Qiagen). Viral DNA levels were quantified by qPCR with Power SYBR green (ThermoFisher) using a forward (F) primer that recognizes a conserved sequence in E1A in combination with a HAdV-C5-specific reverse (R) primer Ad5E1A-R, HAdV-B3-specific Ad3E1A-R or HAdV-E4-specific Ad4E1A-R. Values were normalized to genomic GAPDH and the fold increase of viral copy number at 48 hours was calculated by normalizing to input viral DNA at 6 hours post-infection. All primers used for determining viral replication via qPCR are listed in Table 2.3. Viral replication efficiency in the presence of ivermectin was presented as the relative value compared to that for DMSO control-treated cells, which were normalized to 1.

2.2.9 Statistical analysis

All experiments were carried out with three biological replicates. Graphs represent means and standard errors of the means (SEM) for all biological replicates. For western blotting, a representative image was selected. Statistical significance of numerical differences was calculated using either t-tests or one-way analysis of variance and Holm-Sidak post hoc comparisons between experimental conditions.

Table 2.3 List of qPCR primers used in this chapter.

Method	Target	Orientation	Sequence	Lab ID
RT-qPCR	E1A HAdV-C5	forward	ACACCTCCTGAGATACACCC	JMO1238
	E1A HadV-C5	reverse	TTATTGCCCAGGCTCGTTAAGC	JMO1239
	E1B HadV-C5	forward	GACAATTACAGAGGATGGGC	JMO1240
	E1B HadV-C5	reverse	CACTCAGGACGGTGTCTGG	JMO1241
	E2 HadV-C5	forward	GGGGGTGGTTTCGCGCTGCTCC	JMO1262
	E2 HadV-C5	reverse	GCGGATGAGGCGGCGTATCGAG	JMO1263
	E3 HadV-C5	forward	GAGGCAGAGCAACTGCGCC	JMO1244
	E3 HadV-C5	reverse	GCTCTCCCTGGGCGGTAAGCCGG	JMO1245
	E4 HadV-C5	forward	GCCCCCATAGGAGGTATAAC	JMO1250
	E4 HadV-C5	reverse	GGCTGCCGCTGTGGAAGCGC	JMO1251
	GAPDH	forward	ACTGCTTAGCACCCCTGGCCAA	JMO1023
	GAPDH	reverse	ATGGCATGGACTGTGGTCATGAGTC	JMO1024
Genome replication	E1A (pan)	forward	AGAGGCCACTCTTGAGTGC	JMO1679
	HAdV-C5	reverse	CGTCACGTCTAAATCATAC	JMO1682
	HAdV-B3	reverse	TACAGATCGTGCAGCGTAGG	JMO1680
	HAdV-E4	reverse	AGCGAAGGTGTCTCAAATGG	JMO1681
	GAPDH	forward	ACTGCTTAGCACCCCTGGCCAA	JMO1023
	GAPDH	reverse	ATGGCATGGACTGTGGTCATGAGTC	JMO1024

2.3 Results

2.3.1 Ivermectin inhibits HAdV-C5 progeny production in a dose-dependent manner

Ivermectin has been previously demonstrated to function as an antiviral agent against several viruses that encode factors that rely on host Imp-α/β1-mediated nuclear transport (Wagstaff et al., 2011). Ivermectin has been described as a general inhibitor of this pathway and was shown to reduce nuclear import of viral proteins such as SV40 large TAg, HIV-1 integrase, DENV NS5, and BKPyV VP2 and VP3 (Atkinson et al., 2018; Bennett et al., 2015; Tay et al., 2013; Wagstaff et al., 2012). To build upon these initial studies, we sought to test this drug's ability to affect HAdV replication, as it too encodes many proteins with crucial nuclear functions (Charman et al., 2019).

A549 lung epithelial carcinoma cells were infected with HAdV-C5 at a multiplicity of infection (MOI) of 5 over a range of ivermectin concentrations and harvested at various timepoints post-infection to determine viral titers (Figure 2.1A). Compared to DMSO control-treated cells, production of infectious progeny from cells treated with ivermectin was significantly reduced. Observed differences emerged as early as 24 hours post-infection (hpi), continuing through to 60 hpi (Figure 2.1B). Notably, higher doses of ivermectin corresponded with more potent inhibition of progeny production, indicative of a dose-dependent response. The doses ranged from concentrations that did not affect cell health (1 μM), up to its IC_{50} as indicated by cell viability assays performed 72 hours post-administration (Figure 2.1C). For experiments where longer exposures (12-60 hours) of ivermectin were required, we chose 10 μM as our working concentration, whereas short-term experiments (<6 hours) used higher doses, consistent with existing literature (Fraser et al., 2014).
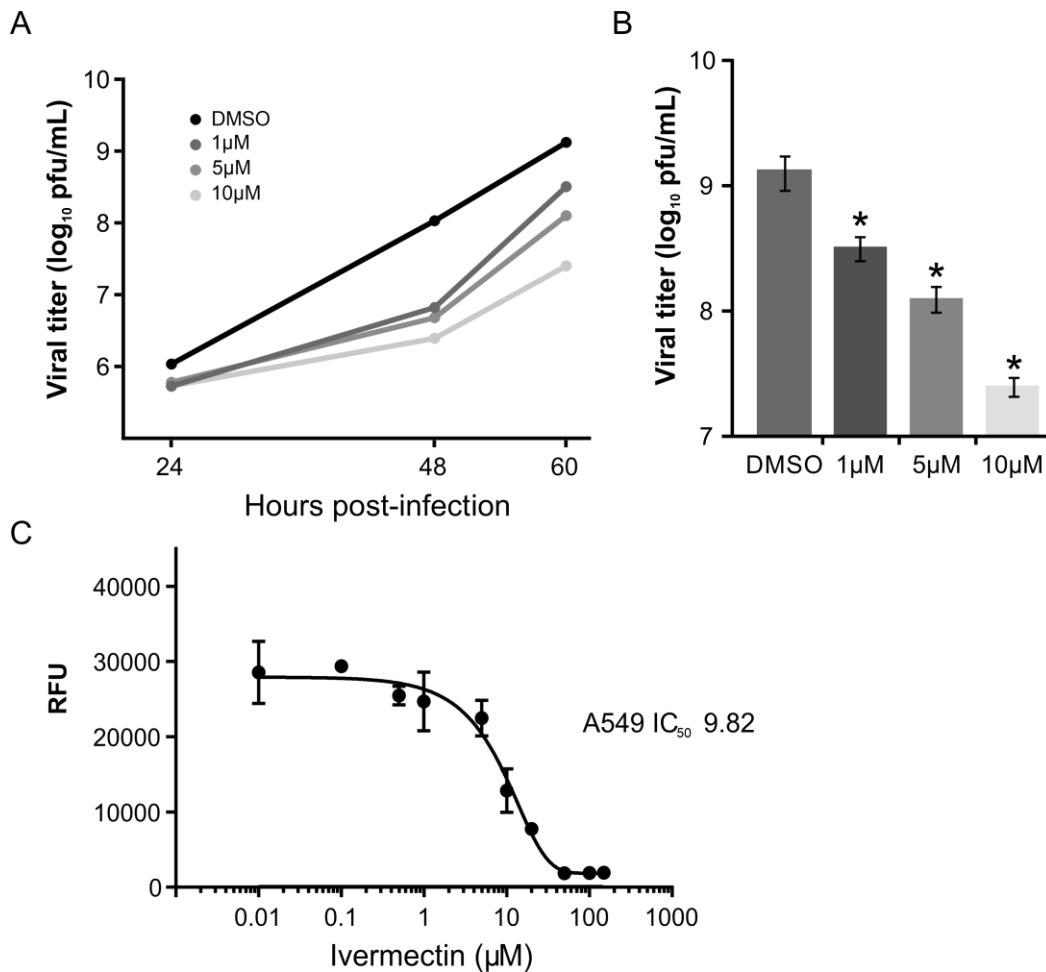
Figure 2.1 Ivermectin inhibits production of infectious HAdV-C5 progeny in a dose-dependent manner.

A549 cells were infected with HAdV-C5 (MOI 5) and treated with the indicated concentration of ivermectin. Cells were collected at various timepoints up to 60 hours post-infection, when cytopathic effect had cleared most cells. Production of infectious progeny virus quantitatively assayed by plaque formation on HEK293 cells. **A)** Data are shown over 24-60 hours for a range of ivermectin doses. **B)** Data at the 60-hour endpoint are highlighted and statistical significance is shown relative to DMSO control-treated cells. Values are represented as mean $\pm$ SEM. *$p<0.01$; n=3. **C)** A549 cells were subjected to a 72-hour dose of ivermectin at various concentrations to gauge long-term cell viability using a PrestoBlue assay.

2.3.2 Ivermectin abrogates nuclear localization of viral factors crucial for the HAdV-C5 replication cycle

To determine how ivermectin was inhibiting HAdV-C5 replication, we examined its impact on the subcellular localization of several HAdV-C5 proteins with known nuclear functions. As proof of principle, we first sought to reproduce previously published findings demonstrating ivermectin's ability to block nuclear import of an EGFP-tagged TAg NLS. The SV40 TAg NLS is one of the oldest and most well-characterized cNLSs and relies exclusively on Imp-α/β1 for import (D Kalderon et al., 1984). HT-1080 fibrosarcoma cells were transfected with either EGFP or EGFP-TAg NLS and subsequently treated for 1.5 hours with 25 µM ivermectin or DMSO control. As anticipated, EGFP localization was unaffected by treatment with ivermectin (Figure 2.2A). While EGFP-TAg NLS protein was predominantly nuclear in control treated cells, there was a noticeable shift to cytoplasmic accumulation in ivermectin treated cells (Figure 2.2B). Importantly, 25 µM ivermectin for 1.5 hours did not elicit any observable cytopathic effects with the HT-1080 cell line. This result confirms that ivermectin reduces nuclear import mediated by the classical SV40 TAg NLS in this experimental system.

Next, we examined ivermectin's effects on several HAdV-C5 early gene products during infection of A549 cells. We specifically targeted E1A and DNA-binding protein (DBP), two proteins which utilize the Imp-α/β1 machinery. Compared to control-treated cells, 10 µM ivermectin caused a dramatic increase in cytoplasmic localization of E1A at 20 hpi (Figure 2.2C). HAdV DBP is a single-stranded DNA-binding protein encoded by the viral E2A early gene and coats single-stranded viral DNA intermediates during HAdV genome replication (van Breukelen et al., 2003). In addition, visualization of DBP in HAdV-C5-infected cells serves as a surrogate marker for virus replication centres (Hidalgo & Gonzalez, 2019). Like E1A, DBP also contains a cNLS (Morin et al., 1989). While treatment of cells with 10 µM ivermectin did not elicit a drastic shift of DBP to the cytoplasm like it did with E1A (Figure 2.2D), noticeably smaller and fewer virus replication centres were present in the nuclei of infected cells at 20 hpi (Figure 2.2E).
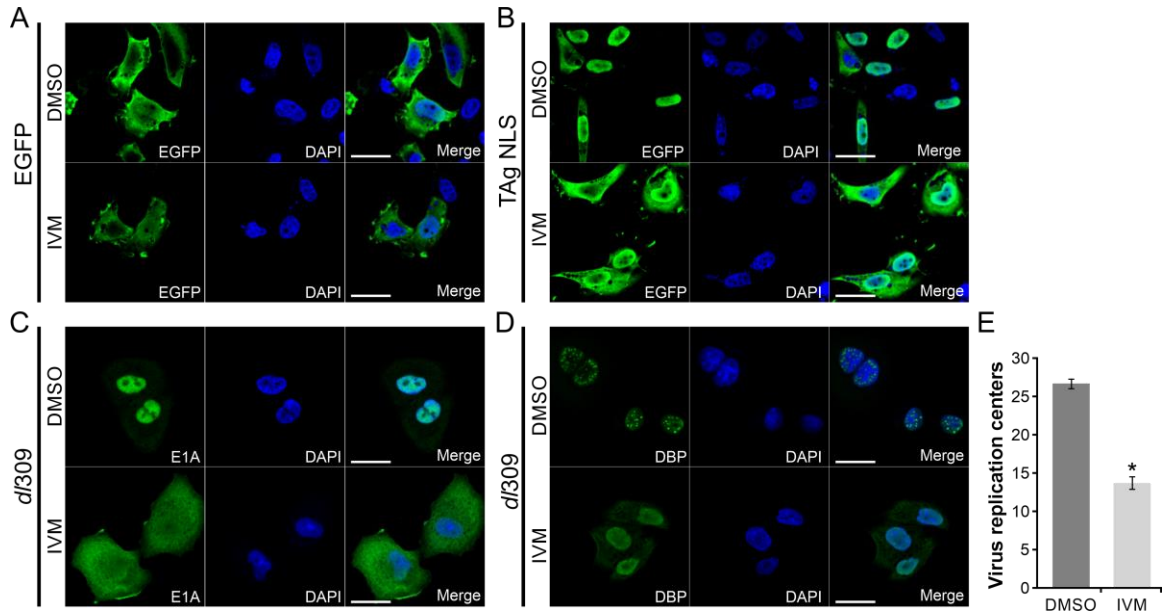
Figure 2.2 Ivermectin blocks Imp-α-mediated nuclear localization of proteins crucial for HAdV infection.

A and B) HT-1080 cells were transfected with either EGFP (vector) or EGFP-tagged T-antigen NLS and treated with 25 µM ivermectin (IVM) for 1.5 hours prior to fixation and immunofluorescence imaging. Cytoplasmic relocalization of T-antigen NLS in the presence of ivermectin serves as a positive control for IVM's function. **C and D)** A549 cells were infected with HAdV-C5 (MOI 5) and treated with 10 µM IVM for 20 hours until cells were fixed and processed for imaging. **C)** Subcellular localization of E1A during infection in the presence of ivermectin. **D)** DBP immunofluorescence, a surrogate for viral replication centres, reveals smaller and fewer virus replication centres. **E)** Quantification of viral replication centres in panel C as determined by DBP immunofluorescence (displayed as means ± SEM, *p<0.001; n=50). Scale bars represent 25 µm.

2.3.3 Ivermectin impairs viral transcription, genome replication, and protein synthesis

To further gauge ivermectin's effects on the molecular kinetics of HAdV-C5 infection, we analyzed multiple aspects of the HAdV-C5 replication cycle. The decreased levels of infectious progeny produced after ivermectin treatment (Figure 2.1) could be due to reduced expression of crucial viral proteins. We first examined overall levels of both early (E1A and DBP) and late (hexon, penton, protein V and VII) HAdV-C5 proteins across multiple infection timepoints (Figure 2.3A). We observed numerous defects in HAdV-C5 protein production in the presence of 10 µM ivermectin. While the initial burst (12–18 hpi) of E1A synthesis appeared unaffected, subsequent timepoints (24–36 hpi) displayed a large reduction in E1A protein levels as shown by western blot. Similarly, DBP and all late proteins were expressed at lower levels in cells treated with ivermectin across all time points.

We next tested if ivermectin was affecting replication of the viral DNA genome itself. As a consequence of the reduced levels of DBP and lower numbers of virus replication centres (Figure 2.2D), we suspected that viral genome copy numbers would be reduced. Indeed, when compared to DMSO treated cells, ivermectin treatment caused a significant decrease in viral genome replication efficiency at 48 hpi as measured by qPCR for HAdV-C5 (Figure 2.3B). To determine if ivermectin is effective against other clinically relevant HAdV species, we tested genome replication efficiency for both HAdV-B3 and HAdV-E4 (Figure 2.3C). Similar to HAdV-C5, a significant reduction in HAdV-B3 genome replication was observed. However, HAdV-E4 was unaffected by ivermectin in these conditions.

E1A is a necessary transcriptional activator of HAdV-C5 early gene expression (King, Zhang, et al., 2018). We hypothesized that ivermectin-mediated abrogation of its nuclear localization and overall expression (Figures 2.2C and 2.3A) would

Figure 2.3 Ivermectin impairs the molecular kinetics of HAdV infection on multiple levels.

A549 cells were infected (MOI 5) with HAdV-C5 (panels A, B, D), HAdV-B3 or -E4 (panel C) and treated with 10 µM ivermectin (IVM). **A)** Cells were harvested at 12, 18, 24, and 36 hpi and viral protein synthesis was assayed by western blot using antibodies against representative proteins from various HAdV-C5 transcription units. Actin was used a loading control. **B and C)** DNA was isolated at 6 hpi as a measure of viral input and at 48 hpi. Relative viral genomic DNA levels were quantified by qPCR using a forward primer recognizing a conserved sequence in the left end of the HAdV genome in combination with specific reverse primers for HAdV-C5, -B3 and -E4. To determine viral genome replication efficiency 48 hpi was normalize to the 6 hpi input. IVM treatment significantly impaired of HAdV-C5 and -B3 genome replication efficiency, represented as mean ± SEM, *p<0.005; n=3. **D)** RNA was isolated at 24 hpi and cDNA was generated using random primers. RT-qPCR was performed targeting HAdV-C5 early gene products and results were normalized to cellular GAPDH. Fold change as compared to control-treated cells is shown, displayed as mean ± SEM, *p<0.05, NS=not significant; n=3.
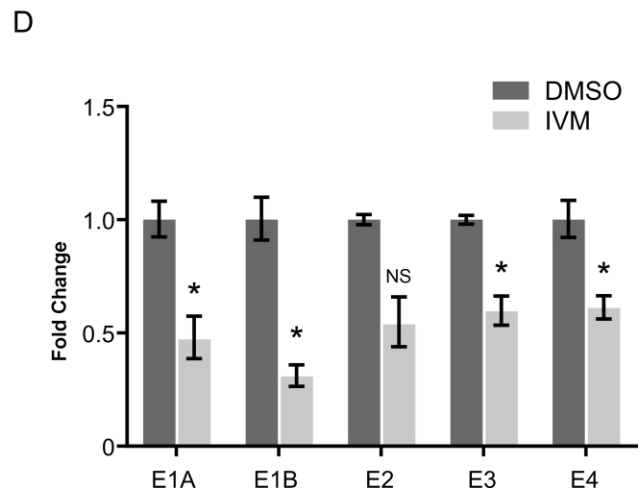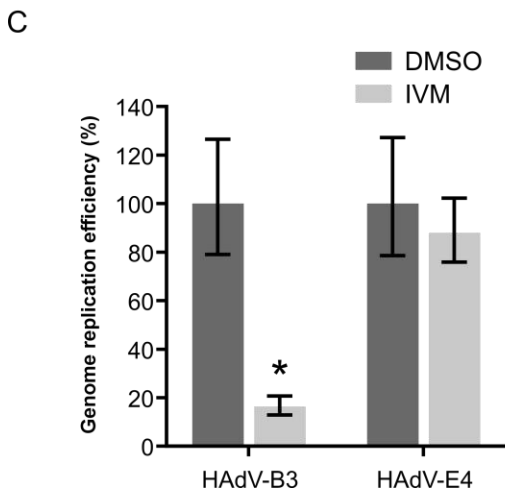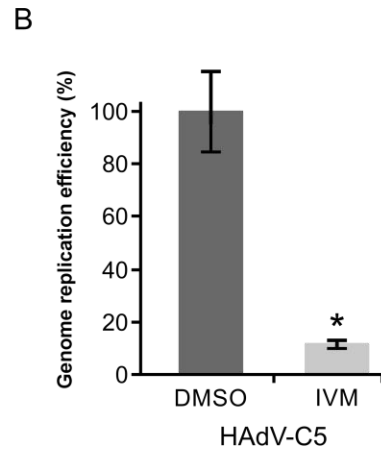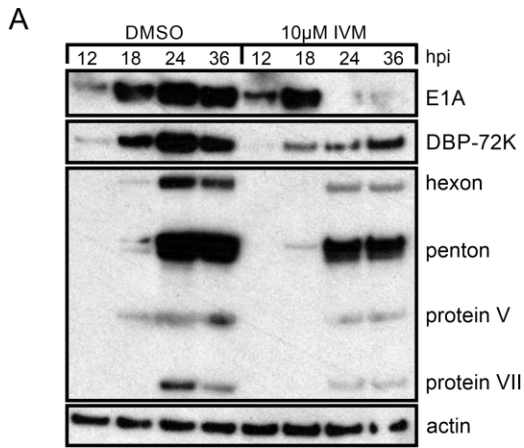
cause reduction of E1A-regulated viral transcripts. We examined infected cells at 24 hpi (when E1A protein levels showed a large difference between DMSO and ivermectin treatment [Figure 2.3A]) for expression of the E1A, E1B, E2, E3, and E4 HAdV-C5 transcription units using RT-qPCR (Figure 2.3C). All early genes displayed reduced levels of mRNA expression in the presence of ivermectin, consistent with our hypothesis.

2.3.4 Ivermectin inhibits the Imp-α/NLS interaction, but not the Imp-α/β1 interaction

Although ivermectin directly targets the classical nuclear import pathway, the exact molecular mechanism by which it inhibits nuclear import remains unclear. Ivermectin's mode of action could be via blocking NLS recognition by Imp-α, blocking the Imp-α/β1 interaction, or possibly both. To provide mechanistic insight into ivermectin's impact on nuclear import, we tested its effects on the protein-protein interaction between cNLSs and Imp-α. We co-transfected HT-1080 cells with FLAG-tagged Qip1 (Imp-α3) and either EGFP-tagged TAg NLS or HAdV-C5 E1A and subjected whole cell lysates to co-immunoprecipitation (CoIP) in the presence of ivermectin (Figure 2.4). Immunoprecipitation of Qip1 showed significantly reduced binding to either the TAg NLS or E1A in the presence of ivermectin compared to the DMSO control. Since nuclear import of cargo by Imp-α depends on Imp-β1, we also tested for Qip1's ability to pull down endogenous Imp-β1 (Kpnb1) using the same CoIP samples. In contrast to cargo interactions, the concentration of ivermectin sufficient to inhibit TAg NLS or E1A binding was not sufficient to block binding of endogenous Imp-β1 to Qip1. These results suggest that ivermectin specifically inhibits the ability of Imp-α to recognize a cNLS, without affecting the Imp-α/β1 interaction.
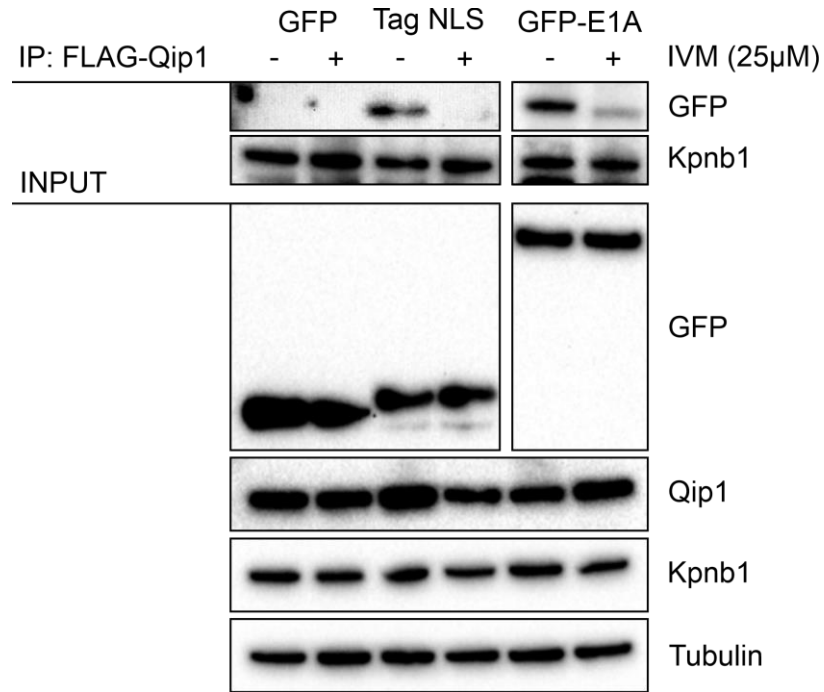
Figure 2.4 Ivermectin blocks NLS binding, but not the Imp-α and -β1 interaction.

HT-1080 cells were co-transfected with either EGFP-tagged TAg NLS or HAdV-C5 E1A as well as FLAG-tagged Qip1 (Imp-α3). Prior to co-immunoprecipitation cells were pre-treated with 25 μM ivermectin (IVM) or DMSO and then immunoprecipitated (FLAG) in the presence of 25μM IVM or DMSO. Western blots for both EGFP constructs and endogenous Kpnb1 (Imp-β1) were performed on the same FLAG-Qip1 immunoprecipitated samples. In the presence of ivermectin Qip1 is unable to recognize both TAg NLS and E1A protein. Under these conditions, the interaction between Qip1 and Imp-β1 was unaffected by IVM.

2.4   Discussion

Cellular pathways controlling nuclear-cytoplasmic transport of viral proteins have emerged as attractive targets for antiviral intervention. The search for new targets is partly due to existing limitations of directly-acting antivirals, which often utilize a "one drug, one bug" approach (Bekerman & Einav, 2015). Most of the approved antivirals target virally encoded enzymes (De Clercq & Li, 2016), which inherently have a narrow spectrum of activity when considering the vastness of viral diversity. Emerging viruses such as SARS-CoV, MERS-CoV, and SARS-CoV-2 have no directly-acting antiviral treatments, highlighting an area where broadly acting antivirals could have profound impact (De Wit et al., 2016; P. Zhou et al., 2020).

One approach to developing broadly acting antivirals is through targeting host factors upon which the virus depends. Evolution of resistance to a drug is expected to be much slower, or even non-existent, for a host factor than for a virus. Targeting these weak points of virus-host interfaces is a strategy being aggressively explored for numerous viruses (Baillie, 2014; S. M.-Y. Lee & Yen, 2012; Warfield et al., 2019). Success has even been achieved with antiretroviral compounds that target the host coreceptor CCR5, blocking HIV-1 replication (Brelot & Chakrabarti, 2018).

Intracellular pathways, such as protein nuclear import, are exploited by many RNA and DNA viruses. Many viruses replicate their genomes within the nucleus of infected cells and/or encode trans-acting viral proteins with important nuclear functions. Also, many protein components of nuclear transport pathways are directly hijacked by diverse viruses (King, Zhang, et al., 2018; Tessier et al., 2019). Among these is the classical nuclear import pathway mediated by Imp-α/β1, making it an actionable host target for broadly acting antivirals. Several drugs, including small molecules and peptides, have been shown to target this pathway (Kosyna & Depping, 2018). However, only two of these, ivermectin and mifepristone, are FDA approved, albeit for non-viral indications (Jans et al., 2019; Kosyna & Depping, 2018; S. Yang et al., 2019). A novel high-throughput screening approach identified ivermectin as a specific inhibitor of nuclear import of HIV-1 integrase via inhibition of this pathway (Wagstaff et al., 2011). To date, ivermectin has been shown

to inhibit nuclear import of viral proteins from a range of viruses *in vitro* (Jans & Martin, 2018; Lv et al., 2018; Tay et al., 2013; Thomas et al., 2018; Varghese et al., 2016; Wagstaff et al., 2012; S. Yang et al., 2019; S. N. Y. Yang et al., 2020) and is currently being used in a clinical trial for the treatment of pediatric dengue virus patients (NCT03432442). Further highlighting the potential application of ivermectin is recent evidence demonstrating that ivermectin can inhibit replication of SARS-CoV-2 (Caly et al., 2020). Together, these findings map a new direction for studying these classes of drugs, particularly ivermectin, as potentially useful broad-spectrum antiviral agents.

Here we sought to examine ivermectin's effects on HAdV infection, using HAdV-C5 as a model. Treatment of cells with ivermectin after infection with HAdV-C5 resulted in a dose-dependent decrease in the production of infectious progeny virus as determined by plaque assay. This reduction was most severe in the presence of 10 µM ivermectin, which caused a nearly 2-fold log reduction in viral titers at 60 hpi compared to DMSO treated control cells. While this dose of ivermectin did begin to negatively affect cell health at 72 hours post-administration, this is unlikely to account for the near 100-fold reduction in viral output, especially as experimental samples were collected earlier than 72 hours. These results parallel recent findings demonstrating ivermectin can inhibit production of infectious progeny from other viruses, including flaviviruses such as West Nile virus, Zika virus, and DENV (S. N. Y. Yang et al., 2020).

To begin to understand how ivermectin inhibits HAdV-C5 replication, we explored the impact of ivermectin on intracellular localization of HAdV proteins during infection. We first confirmed ivermectin's ability to block Imp-α-mediated nuclear import by using the well-characterized SV40 TAg cNLS as a control. Like TAg, the HAdV-C5 E1A protein also contains a cNLS that utilizes Imp-α/β1 to drive nuclear import (Cohen et al., 2014). After infection with HAdV-C5, treatment with 10 µM ivermectin caused a shift in E1A subcellular localization from its predominantly nuclear localization to a nuclear/cytoplasmic one, similar to that observed for a mutant E1A containing a deletion of its C-terminal NLS sequence (King, Gameiro, et al., 2018). While most E1A was relocalized to the cytoplasm, some remained in the nucleus. This may be due to insufficient drug, or E1A's small size, which may allow some passive diffusion into the nucleus.

Additionally, E1A has several non-cNLSs, which may allow entry into the nucleus during inhibition of the classical nuclear import pathway (Marshall et al., 2014). Nevertheless, use of such alternative nuclear import pathways is clearly not sufficient for the virus to overcome the negative effects caused by ivermectin mediated inhibition of classical nuclear import.

We also examined the localization of HAdV-C5 DBP, a single stranded DNA-binding protein. DBP is a larger (~72kDa) protein crucial for replication of the viral genome and is frequently used as a marker for visualizing virus replication centres (van Breukelen et al., 2003). Interestingly, ivermectin did not affect DBP nuclear localization as dramatically as E1A, despite the known presence of two cNLSs within its N-terminus. This was not entirely unexpected, as a previous study showed that mutant DBP lacking its NLSs was still strongly nuclear (Morin et al., 1989). Since ivermectin only targets Imp-α cargo, it remains possible that DBP utilizes an Imp-α-independent nuclear import pathway. Nevertheless, ivermectin treatment affected the formation of virus replication centres, as they were smaller and fewer in number, in concordance with the observed lower levels of viral progeny production.

In addition to infectious progeny, we surveyed the impact of ivermectin treatment on HAdV-C5 early and late proteins as well as replication of its DNA genome. On the protein level, synthesis of E1A at early timepoints during infection remained unchanged. This may be due to E1A's initial expression being driven by a strong, constitutive enhancer sequence in the left-end of the HAdV genome (Hearing & Shenk, 1983). It also suggests that the earliest parts of HAdV-C5 infection may be unaffected by disruption of the Imp-α/β1 pathway. This is different from the effect of ivermectin on infection by BKPyV, another small DNA tumor virus. BKPyV utilizes cNLSs to facilitate nuclear entry of incoming viral particles (Bennett et al., 2015), while HAdV capsids transit to the nucleus via microtubules in an NLS-independent manner (Bremner et al., 2009; Dodding & Way, 2011). Despite the initial burst of E1A expression in the presence of ivermectin, subsequent timepoints showed a severe loss of E1A protein synthesis after 24 hours. This suggests that the positive feedback loop whereby newly synthesized E1A protein would enter the nucleus to trans-activate additional expression of itself may be disrupted by ivermectin. Levels of DBP were

decreased across infection, consistent with the lower amounts of virus replication centres previously observed. Expression of various late proteins, including hexon, penton, pV and pVII, were also greatly reduced in the presence of ivermectin. As expected, overall replication of the HAdV-C5 genome was severely decreased upon ivermectin treatment. This aligns with observations of smaller and fewer virus replication centres and overall lower expression levels of DBP protein. The combination of reduced viral genome templates and early proteins also likely accounts for the subsequent reduction in HAdV-C5 late gene expression.

In addition to HAdV-C5, we tested genome replication efficiency for HAdV-B3 and -E4, both of which represent clinically relevant HAdV species. Like HAdV-C5, ivermectin dramatically reduced HAdV-B3 genome replication. HAdV-B3 is commonly associated with acute respiratory illness, therefore ivermectin could have significant clinical relevance. For HAdV-E4, overall levels of replication were lower than for -B3 and -C5 even in DMSO-treated cells, but its relative genome replication of HAdV-E4 was unaffected by ivermectin treatment. Although initially surprising, the lower replicative ability of this virus could obscure ivermectin's effects. Alternatively, HAdV-E4 is unique among HAdVs as it more closely resembles simian adenoviruses (Dehghan et al., 2013). Past studies examining HAdV-E4 genome replication suggest that HAdV-E4 relies on a different repertoire of host factors than other HAdV species (King, Gameiro, et al., 2018). For example, knockdown of host protein kinase A severely inhibits genome replication of HAdV-B3, -C5, -D9 and -A12, but not -E4. Furthermore, *in vitro* evidence has demonstrated that the E1A protein of HAdV-E4 uses a non-cNLS located in conserved region 3. Indeed, E1A from HAdV-B3 and -C5 also possess this non-cNLS however, nuclear import studies have shown this region of HAdV-E4 E1A to be a much more potent stimulator of nuclear import than -B3 or C5 (Marshall et al., 2014).

Lastly, we probed for relative expression levels of mRNAs from all HAdV-C5 early transcription units (E1A, E1B, E2, E3, and E4). At 24 hpi, we detected lower levels of mRNA for each of these early genes. This is expected given the lower levels of E1A expression, which would otherwise drive higher expression of these genes (King, Zhang, et al., 2018). Together, these findings suggest ivermectin's inhibition of HAdV-C5

replication stems mostly from lower expression of crucial viral gene products, particularly E1A, via disruption of its nuclear localization.

Despite the growing interest in studying ivermectin as an antiviral agent, relatively little is known about its molecular mode of action. The original high-throughput drug screen unequivocally demonstrated ivermectin's ability to inhibit the binding of HIV-1 integrase and TAg NLS to the Imp-α/β1 complex. Recent *in vitro* data published by Yang *et al*., suggests that ivermectin induces structural changes in Imp-α that prevents NLS recognition, as well as binding to Imp-β1 (S. N. Y. Yang et al., 2020). Our data derived from CoIP experiments confirm that ivermectin blocks NLS recognition. However, our results suggest that concentrations of ivermectin sufficient to block NLS binding to Imp-α are not sufficient to disrupt formation of the Imp-α/β1 complex. The IBB domain of Imp-α acts as an auto-inhibitory NLS that binds its own NLS binding region in a fashion that mimics cNLS binding (Harreman, Cohen, et al., 2003). Ivermectin could theoretically block or weaken this intramolecular interaction the same way it blocks NLS binding, while leaving the αIBB/Imp-β1 intermolecular interaction unaffected. It would be interesting to determine if Imp-β1 is still carrying unloaded Imp-α into the nucleus in the presence of ivermectin, causing futile cycles of import that may have consequence beyond simple inhibition of NLS/Imp-α interactions. Although we examined Qip1 (Imp-α3) in detail, as it is the preferred importin for E1A (Kohler et al., 2001), it will be important to consider how ivermectin may influence NLS binding to other Imp-α isoforms. Differences among isoforms can influence both the binding of NLSs and the auto-inhibitory potential of the IBB (Pumroy & Cingolani, 2015). This could be achieved by studying the broader effects of ivermectin with a panel of different Imp-α isoforms and a variety of cNLSs.

Based on the data presented here, we conclude that the effect of ivermectin on nuclear import of E1A can explain the significant reduction in viral replication (Figure 2.5). Upon HAdV infection, E1A is the first viral gene transcribed and is responsible for setting the stage for viral replication to proceed (Montell et al., 1982; Winberg & Shenk, 1984). Inhibition of E1A nuclear import at early time points during infection would impair its ability to trans-activate viral early genes, including itself, ultimately resulting in the
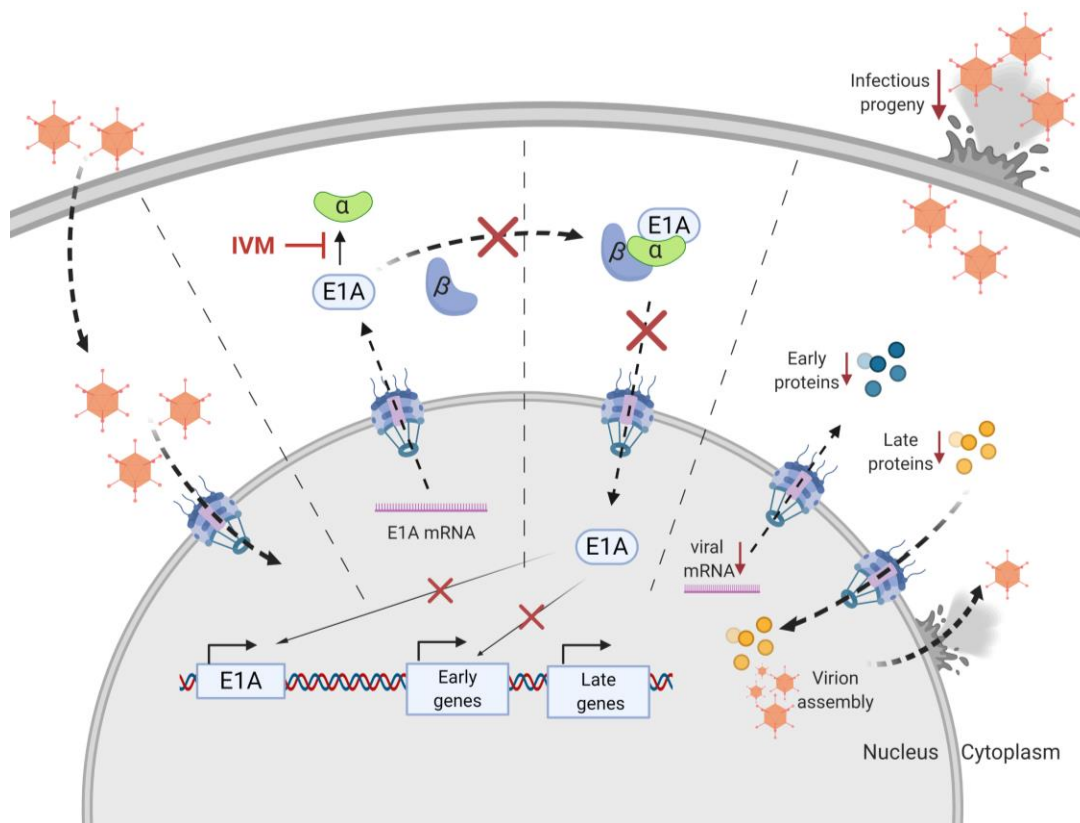
Figure 2.5 Ivermectin-mediated inhibition of the HAdV replication cycle.

Proposed model of how ivermectin (IVM) inhibits adenovirus replication. Black arrows indicate normal viral processes. Those processes inhibited directly or indirectly by IVM are labelled with red. **A)** HAdV virions are trafficked to the nuclear pore complex (NPC) via microtubules and viral genomes are released into the nucleus. **B)** E1A is transcribed immediately and its corresponding mRNA is translated within the cytoplasm. In the presence of IVM, E1A protein cannot bind importin alpha (α, green), therefore preventing the formation of a competent E1A/importin-α/β1 import complex. **C)** In the presence of IVM E1A protein is not efficiently imported into the nucleus, indirectly preventing further transcription of E1A and induction of other viral early genes. **D)** Reduced expression of viral gene products leads to reduced genome replication, late proteins, and ultimately the assembly of viral progeny within the nucleus.

downstream consequences observed here: impaired production of other early proteins, genome replication, late proteins, and infectious progeny. Supporting this, orthogonal experiments using HAdV encoding an NLS-deleted E1A showed a similar reduction in overall viral replication as we see here with ivermectin (Crisostomo et al., 2017).

In summary, this study provides evidence that disruption of classical Imp-α/β1-mediated nuclear import has promise in combating at least a subset of HAdV infections. Importantly, these data support the growing body of literature suggesting ivermectin has utility as a broadly acting antiviral agent. In addition, these findings demonstrate the strategic advantage of targeting host factors for antimicrobial action. Genetically distinct viruses often share important, potentially druggable features as a result of their co-dependence on host processes, including the classical nuclear import pathway (Bennett et al., 2015; Howley & Livingston, 2009; Tao et al., 2003; Tessier et al., 2019). In addition to repurposing ivermectin for treatment of viral infections that have no current therapeutic options (including HAdV), these results support the idea that targeting host nuclear import by other drugs or means could be a valid strategy as well.

Chapter 3

3   Piggybacking on classical import and other non-classical mechanisms of nuclear import appear highly prevalent within the human proteome

3.1   Introduction

Nucleocytoplasmic transport of proteins across the nuclear envelope is an essential cellular process unique to eukaryotic organisms. The nuclear envelope spatially separates the contents of the nucleus from the cytoplasm and provides a physical mechanism for regulating numerous cellular events such as transcription, translation and the cell cycle. Transport of proteins and RNA across the nuclear envelope is a tightly orchestrated process that requires all molecules to pass through the NPC, a large multimeric complex built from multiple copies of approximately 30 different proteins called nucleoporins (Hoelz et al., 2011; Knockenhauer & Schwartz, 2016). In essence, the NPC functions as a semi-permeable barrier, selectively allowing passage of certain molecules while simultaneously preventing passage of others (Timney et al., 2016). Proteins of varying sizes are able to diffuse through the central channel of the NPC; however, this is influenced by several factors. Notably, the rate at which a protein can diffuse into the nucleus is inversely related to its size. As protein size increases from less than 30–40 kDa, its ability to diffuse diminishes rapidly (Mohr et al., 2009; Schmidt & Görlich, 2016). Other factors such as a proteins shape and surface composition also have a significant effect on passive diffusion (Frey et al., 2018; Schmidt & Görlich, 2016). Nevertheless, despite proteins having the capacity to diffuse through the NPC, proteins of all sizes employ the assistance of NTRs for rapid and efficient nuclear import or export (C. F. Chen et al., 1996; Ribbeck et al., 1998).

Bidirectional transport of proteins through the NPC is carried out by a group of soluble transport receptor proteins belonging to the Kapβ superfamily, which can be further subdivided into importins or exportins (O'Reilly et al., 2011). The human genome encodes 20 Kapβ proteins; 10 are importins that shuttle proteins into the nucleus, 7 are exportins and shuttle proteins out of the nucleus, 2 are bidirectional transporters and 1 currently has

no known function (Kimura & Imamoto, 2014). Kapβ proteins recognize a cargo's NLS for import, or NES for proteins which undergo export (Soniat & Chook, 2015; Wen et al., 1995). Following cargo binding, Kapβ can facilitate transport through the NPC where cargo is released into either the cytoplasm or nucleus. Importantly, the loading and un-loading of cargo onto Kapβ is aided by the RanGTPase system, where cargo is released within the nucleus upon RanGTP binding to Kapβ (Cautain et al., 2015). Conversely, binding of export cargo is aided by binding of RanGTP, where subsequent release of cargo into the cytoplasm is triggered by RanGTP hydrolysis.

Not all Kapβ's recognize their cargo directly. For example, Imp-β1 mainly recognizes its cargo through the adapter Imp-α. This pathway is commonly referred to as the classical nuclear import pathway and involves recognition of a cargo's NLS by one of the 7 human Imp-α isoforms, which are then shuttled through the NPC by Imp-β1 as a heterotrimeric complex (Goldfarb et al., 2004b; Dirk Görlich et al., 1995). NLSs recognized by Imp-α are referred to as cNLSs, are rich in basic amino acids and come in two forms, monopartite or bipartite (Lange et al., 2007). Monopartite cNLSs are comprised of a single cluster of basic amino acids while bipartite cNLS have two clusters of basic amino acids separated by a linker region of variable length and composition. These motifs can be exemplified by the SV40 large TAg (PKKKRKV) and NP (KRPAATKKAGQAKKKK) cNLS, respectively (D Kalderon et al., 1984; Robbins et al., 1991). Structurally, Imp-α is composed of 10 ARM domains which form two pockets, referred to as the major and minor groove (Conti et al., 1998; Conti & Kuriyan, 2000). These grooves accommodate the basic clusters of amino acids characteristic of a cNLS. Monopartite motifs primarily bind the major groove. However, several NLSs, both cellular and viral, can bind the minor groove exclusively (Kosugi et al., 2009; Lott et al., 2011; Nakada et al., 2015; Pang & Zhou, 2014). Bipartite cNLSs occupy both the major and minor grooves, with the smaller N-terminal basic cluster bound to the minor groove and the larger basic cluster bound to the major groove (Conti & Kuriyan, 2000).

Another class of NLS that has been characterized is the PY-NLS, which is recognized by TNPO1 and TNPO2, both importin Kapβ members (B. J. Lee et al., 2006; Soniat & Chook, 2015). The PY-NLS is not as well characterized as the cNLS; however, some general rules

have emerged. These include an N-terminal motif, either hydrophobic or basic, a C-terminal $R/K/H-X_{2-5}-PY$ motif and are found within a structurally disordered region with overall basic charge (B. J. Lee et al., 2006). An additional Kapβ, transportin-SR2 (also known as TRN-SR or TNPO3) has been shown to import SR splicing factors by binding to their arginine-serine (RS) domains (Maertens et al., 2014). In regard to exportin Kapβs, only one class of NES has been characterized and this is mediated by exportin-1 (XPO1, also known as CRM1), which recognizes 8-15 amino acid long leucine rich motifs (Dong et al., 2009; Fornerod et al., 1997; Fukuda et al., 1997). Similar to the classical nuclear import pathway, XPO1/CRM1-mediated export has been extensively studied with hundreds of characterized cargo from human and model organisms (Kırlı et al., 2015).

The classical nuclear import pathway is assumed to handle the majority of nuclear import as it is the best characterized and has many documented cargos and cNLSs (Bernhofer et al., 2017; Lange et al., 2007). Sequence attributes of cNLSs have made them highly predictable and this has led to the development of numerous NLS prediction programs (Bernhofer et al., 2017; Brameier et al., 2007; Kosugi et al., 2009; J. Lin & Hu, 2013; Nguyen Ba et al., 2009). Interestingly, early estimates using PSORT II demonstrated that only ~55% of nuclear proteins in *S. cerevisiae* have a predictable cNLS (Lange et al., 2007). As more NLS prediction programs emerged, the fraction of yeast and human nuclear proteins with predictable cNLSs unexpectedly remained between 30–40% (Bernhofer et al., 2017; Marfori et al., 2011). These observations likely reflect a combination of non-classical import pathways, alternative cNLSs and piggybacking into the nucleus indirectly via physical interaction with other proteins that directly bind the nuclear transport apparatus. In fact, data from yeast shows that up to 50% of proteins that bind Srp1 (the only yeast Imp-α) do not have a predictable cNLS, providing strong circumstantial evidence that their association with Srp1 and subsequent nuclear import occurs via piggybacking or alternative NLSs (Lange et al., 2007).

Whether or not these observations hold true in humans has not been explored. With the development of databases such as the Human Protein Atlas (HPA) and Human Reference Interactome (HuRI), as well as the abundance of publicly available high-throughput mass spectrometry data, it may be possible to establish a more accurate picture of nuclear

transport mechanisms (Deutsch et al., 2016; Luck et al., 2020; Thul et al., 2017; Uhlén et al., 2015). While examples of piggybacking into the nucleus using Imp-α have been documented for several distinct nuclear proteins, widespread identification of potential piggybacking proteins or estimates of the extent to which this nuclear import strategy is used remain poorly characterized (Asally & Yoneda, 2005; Bange et al., 2013; Czeko et al., 2011; Di Croce, 2011; Kressler et al., 2012; Trowitzsch et al., 2015). Using a collection of resources and NLS prediction programs, I aimed to acquire information regarding the prevalence of piggybacking and use of alternative nuclear import pathways in eukaryotic cells. My analyses show that nearly 50% of nuclear proteins in the human proteome do not have a predictable cNLS. I identify a large cohort of proteins found in both the nucleus and cytoplasm which have a predicted NES, but not a predicted cNLS. Examination of binary interactions for 6 of the 7 known Imp-α isoforms demonstrates that 20–50% of interactors also do not have a predictable cNLS. Furthermore, a reanalysis of publicly available mass spectra files for protein interactions mediated by several Imp-α isoforms showed that up to 50% of cargos do not have a predictable cNLS. Finally, using this data I specifically focus on several nuclear protein complexes involved in transcription, and show that the majority of proteins belonging to the mediator of RNA polymerase II transcription (Mediator) complex interact with at least one Imp-α, yet do not have a predictable cNLS.

3.2   Materials and Methods

3.2.1   Datasets for nuclear, cytoplasmic and nucleocytoplasmic proteins

Proteins with experimental evidence of being localized to the nucleus or cytoplasm were downloaded from the HPA (Thul et al., 2017). According to the HPA nuclear localization dataset, this includes the nucleoplasm, nuclear speckles and nuclear bodies, while cytoplasmic localization includes the aggresome, cytosol, cytoplasmic bodies, rods and rings. Proteins present in both the nucleus and cytoplasm were additionally grouped together as nucleocytoplasmic proteins. From here all canonical protein sequences were retrieved from UniprotKB/Swiss-Prot (UniProt Consortium, 2021).

Proteins known to associate with yeast Srp1 were downloaded from BioGrid while interactors for human Imp-α1, α3, α4, α5, α6 and α7 were retrieved from IntAct (Orchard et al., 2014; Oughtred et al., 2019). Only physical interactions were kept, removing any interactions identified through genetic studies or post-translational modifications. Proteins with binary, or direct, interactions with Imp-α were retrieved from the HuRI database for Imp-α isoforms α1, α3, α4, α5, α6 and α7 (Luck et al., 2020). All interactors were combined together, and redundant interactors were removed before screening against the HPA nuclear dataset to obtain those with evidence of nuclear localization only.

3.2.2 NLS and NES prediction

For identifying cNLSs using regular expression matching, I used the regular expressions provided through the ELM database corresponding to monopartite core (ELME000270), monopartite core with C-terminal preferences (ELME000278), monopartite core with N-terminal preferences (ELME000271) and bipartite (ELME000276) (Kumar et al., 2020). The criteria for having a cNLS only required a protein to have at least one of these regular expressions satisfied.

For predicting NLSs using NLStradamus, each protein was searched for both monopartite and bipartite NLSs using a threshold score of 0.6 (Nguyen Ba et al., 2009). All hits were combined and duplicate protein ID matches were removed to end up with a list of unique proteins containing either a monopartite or bipartite NLS. For searches using cNLS Mapper, the default cut-off score of 0.5 was used and included searching the entire region of the protein for bipartite NLS with a long linker region (Kosugi et al., 2009). Since NLSdb searches for matches within its own library of potential or experimentally confirmed NLSs no threshold or cut-off scores could be used a priori and any matches to NLSdb were taken as a hit (Bernhofer et al., 2017). Predictions made with NESmapper used the default threshold score of 2 to identify potential NESs (Kosugi et al., 2014). Similar to NLS prediction, any duplicate protein IDs were removed to obtain a list of unique proteins with at least one predicted NES. For more detailed cNLS analysis PONDR and DisEMBL were used to screen for intrinsic disorder (Linding et al., 2003; Peng et al., 2006). PONDR (VSL2) was used to find short (<30 residues) or long (>30 residues) regions

of predicted disorder where a given cNLS was located. Similarly, DisEMBL was used to identify regions with a high degree of mobility that overlap the cNLS in question.

3.2.3   Identification of novel motifs with MEME and SLiMSearch

Imp-α interactors from the HuRI dataset without a cNLS were combined into a single group and all duplicate protein identifications were removed as well as any nucleoporins or Imp-α proteins. This produced a group of 10 unique proteins which were then analyzed using the motif elicitation program MEME (Bailey et al., 2009). MEME settings were set to identify three clusters with zero or one occurrence per sequence with a minimum length of 6 amino acids. The top scoring cluster was subsequently used for further analysis. Using the motif defined as KxRxHxK I searched the human proteome with SLiMSearch (Krystkowiak & Davey, 2017) for any motif matches that are located within a predicted IDR (disorder cut-off set to 0.4). Identifications retrieved with SLiMSearch then underwent gene ontology analysis using Metascape (Y. Zhou et al., 2019) to identify enriched cellular processes.

3.2.4   Proteomics analysis

Raw mass spectra were downloaded from the Proteomics Identification database (PRIDE) corresponding to project PXD007976 titled "Landscape of nuclear transport receptor cargo specificity" (Mackmull et al., 2017; Perez-Riverol et al., 2019). Specifically, mass spectra corresponding to wild-type control, BirA* control, Imp-α1 (N- and C-terminal BirA* tag), Imp-α5 (N- and C-terminal BirA* tag), and Imp-α6 (C-terminal BirA* tag) were retrieved for samples that were digested on-bead. Tandem mass spectra were searched using MS-GF+ against the human Swiss-Prot entries from UniProtKB (release 03/2020, 20,305 entries) and included common contaminants in addition to BirA and streptavidin. Additionally, a reverse decoy database was used for false discovery rate estimation (Elias & Gygi, 2007). MS-GF+ search parameters were as follows: full tryptic specificity, precursor mass tolerance of 20 ppm, and dynamic modifications for methionine oxidation, N-terminal acetylation and biotinylation of lysine.

Proteins were identified using a target-decoy strategy with IDPicker and filtered at a false discovery rate (FDR) of 1% and a minimum of two unique peptides per protein (Ma et al.,

2009). Experiments where Imp-α was expressed with BirA* either on the N- or C-terminus were combined to establish a unique set of interactors encompassing both experiments. Proteins identified from each experimental sample were analysed separately with SAINTexpress (Teo et al., 2014) using wild-type and BirA* samples as controls. To further increase the statistical strength of identifying co-purifying bait proteins I used additional controls provided through the CRAPome (CC532) (Mellacheruvu et al., 2013). Peptides identified through SAINTexpress with an FDR less than 5% were considered statistically relevant. Finally, all interactors for Imp-α1, α5 and α6 were combined and reduced into a list of non-redundant proteins (available upon request) that could be screened against the HPA for evidence of nuclear localization, unless otherwise stated. Only those with evidence of nuclear localization were used for analysis. For TAF-Imp-α interactions, spectral counts and FDR values produced by SAINTexpress were submitted to ProHits-viz for visualization (Knight et al., 2017).

3.3   Results

3.3.1   Many nuclear localized proteins do not have a predictable cNLS

The classical nuclear import pathway is assumed to handle the majority of protein nuclear import. Extensive research into this pathway has established a defined set of rules for the cNLS-Imp-α interaction, making them highly amenable to computational prediction (Kosugi et al., 2009; Marfori et al., 2011). To estimate the fraction of nuclear proteins with a predictable cNLS, I first retrieved a list of proteins from the HPA that are localized to the nucleus. Additionally, I collected proteins that localize to the cytoplasm in order to capture proteins present in both compartments that could potentially shuttle bidirectionally across the nuclear envelope. From the HPA, 6542 nuclear and 4493 cytoplasmic proteins were identified. These nuclear and cytoplasmic proteins demonstrated substantial overlap, with over 2100 proteins found in both the nucleus and cytoplasm (Figure 1A). This number represents almost a third of nuclear proteins and one-half of cytoplasmic proteins that localize to both cellular compartments.

Analysis of proteins which localize to the nucleus, nucleus and cytoplasm, and cytoplasm using NLStradamus for NLS prediction and NESmapper for NES prediction indicated that NLSs are more frequently predicted in nuclear proteins and NESs are more frequently predicted in cytoplasmic proteins, as anticipated (Figure 1B). The difference in frequency of identifying a predicted cNLSs for nuclear proteins or a predicted NESs for cytoplasmic proteins is particularly intriguing. Over 80% of cytoplasmic proteins have a predictable NES, while only ~40% of nuclear proteins have a predictable cNLS.

As less than half of the nuclear proteins have a predicted cNLS, I questioned if this was due to NLS prediction being too specific. To search for cNLSs with greater sensitivity, I used simple regular expression (RegEx) matching to search all nuclear proteins, including those that are nucleocytoplasmic. For RegEx matching I used experimentally validated motifs corresponding to monopartite core, monopartite N-extended, monopartite C-extended, and bipartite from the ELM resource (Kumar et al., 2020). As expected, RegEx matching increased cNLS prediction sensitivity; however, putative cNLSs were still identified in only 53% of nuclear proteins, compared to 37% using NLStradamus (Figure 1C). Comparison of proteins with a predicted cNLS from RegEx matching or NLStradamus shows significant overlap; with the majority of NLStradamus hits also being identified by RegEx matching (Figure 1D). My prediction with NLStradamus agrees with previous findings (Bernhofer et al., 2017), and less stringent searches for cNLSs using RegEx matching still fail to predict a cNLS in almost 50% of nuclear proteins.

While other NLSs such as the PY-NLS exist, only a limited number of PY-NLSs have been characterized in detail, and no reliable prediction models exist. Most PY-NLSs characterized to date possess the sequence motif $R\backslash K\backslash H\text{-}X_{2\text{-}5}\text{-}PY$, where a positively charged amino acid (Arg, Lys, His) can be found up to 5 amino acids N-terminal to a PY motif (Xu et al., 2010). While this motif is one of several PY-NLS attributes, it is not sufficient to predict a PY-NLS and on its own would be highly over-predictive. Nevertheless, I used this motif to search proteins that do not contain a cNLS using RegEx matching and found only 30% of these proteins contained this minimal PY-NLS motif (Figure 1E), leaving a substantial portion of the nuclear proteome without a predictable cNLS or PY-NLS.

Despite the HPA characterizing protein subcellular localization in several cell lines, this does not rule out additional cytoplasmic proteins that could potentially localize to the nucleus under different cellular conditions, or other cell types not captured by the HPA. As previously demonstrated, putative cNLSs can be found in over 20% of cytoplasmic proteins (Figure 1B). With this subset of cytoplasmic proteins, I used NucPred to predict each proteins probability of localizing to the nucleus (Brameier et al., 2007). NucPred scores range from 0 to 1, with higher scores having a greater probability of a protein being nuclear. As expected, most cytoplasmic proteins had a lower NucPred score; however, many proteins still scored greater that 0.8 (Figure 1F). As these scores are only probabilities, NucPred performance is further enhanced if a protein also has a predicted NLS. Proteins with a NucPred score greater than 0.8 and a predicted NLS have been shown to be correctly identified as nuclear with over 90% accuracy (Brameier et al., 2007). Taking the 380 cytoplasmic proteins that scored equal or greater than 0.8 and filtering with NLStradamus resulted in nearly 25% of these proteins having a potential cNLS. Indeed, it's possible many of these proteins have nuclear functions despite being classified as cytoplasmic based on Protein Atlas data. Nevertheless, these observations suggest that substantially more cytosolic proteins may have under characterized, context specific occupancies within the nucleus than anticipated, which cannot be captured by immunofluorescence alone.

Based on data from the HPA, these findings point to a conservative estimate where almost 50% of nuclear proteins lack a predictable cNLS, and this estimate increased to over 60% using more stringent NLS prediction programs. Furthermore, analysis of cytoplasmic proteins using nuclear localization and NLS prediction demonstrates a substantial portion of cytoplasmic proteins may have currently uncharacterized, potentially context dependent roles within the nucleus. Taken together, these findings emphasize the discrepancy in cNLS prediction for established human nuclear proteins and highlight an intriguing inconsistency between the frequencies of NES and NLS prediction.

Figure 3.1 The majority of human nuclear proteins do not contain a predictable cNLS.

A) Nuclear and cytoplasmic proteins were retrieved from the Human Protein Atlas (HPA) for the identification of distinct nuclear, cytoplasmic and nucleocytoplasmic proteins. **B)** cNLS and NES prediction of nuclear, nucleocytoplasmic (Nuc/Cyto) and cytoplasmic proteins from the HPA using NLStradamus and NESmapper, respectively. The majority of Nuc/Cyto and cytoplasmic proteins have a predictable NES in contrast to nuclear and Nuc/Cyto proteins where the majority do not have a predictable cNLS. **C)** Comparison of cNLS prediction approaches using NLStradamus and regular expression matching (RegEx) on nuclear proteins from the HPA. All motifs corresponding to cNLSs in the ELM resource were used for RegEx matching where any protein with at least one match is counted as a hit. Comparing approaches shows that somewhere between 47–63% of nuclear proteins do not have a predictable cNLS. **D)** Prediction of proteins with a cNLS using either RegEx matching or NLStradamus demonstrates significant overlap. The majority of NLStradamus predictions are also predicted by RegEx matching. **E)** Proteins without a cNLS, as determined by RegEx matching, were searched for a minimal PY-NLS (R/H/K-X$_{2-5}$-PY), demonstrating that a substantial portion of nuclear proteins also do not contain a PY-NLS. **F)** Cytoplasmic proteins were analyzed with NucPred to predict nuclear localization and those with a score greater than 0.8 were searched for cNLSs. Those with a cNLS are considered to have a high probability of nuclear localization, highlighting the potential for additional nucleocytoplasmic localizations not supported by the HPA.
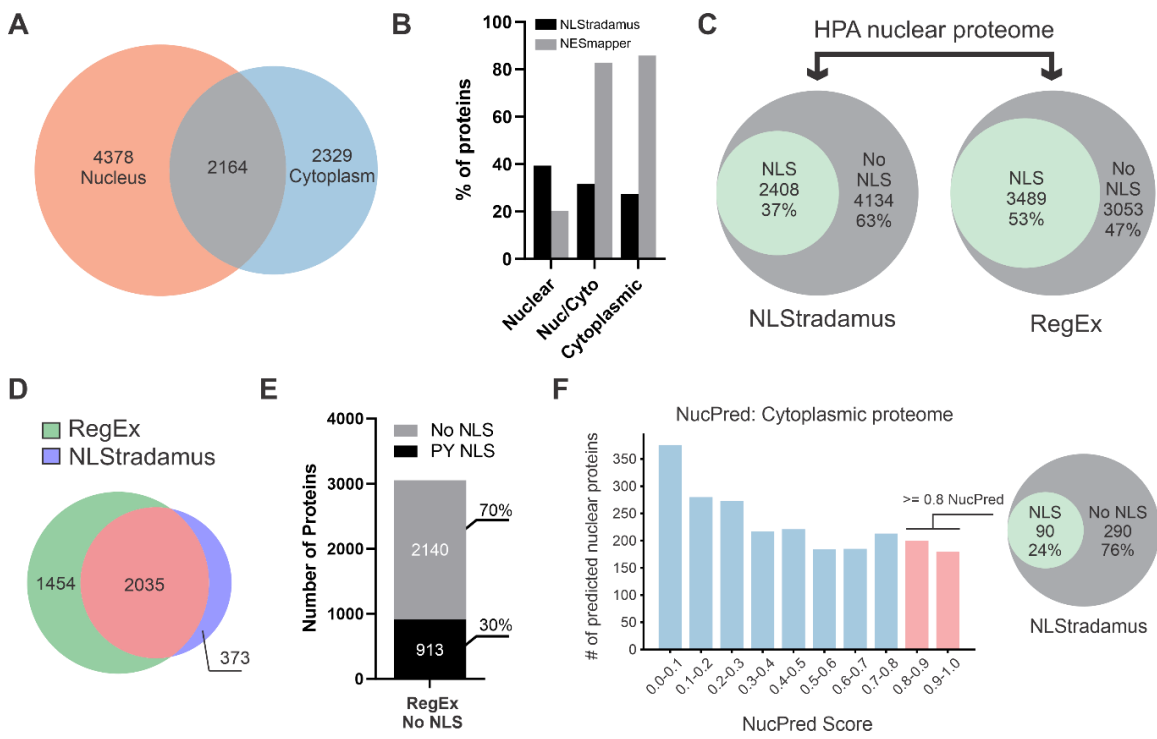
3.3.2 Many Imp-α binding partners do not have a predictable cNLS

Protein nuclear import is mediated by a variety of different importins, ranging from Imp-α and the classical nuclear import pathway to alternative import pathways using importin Kapβs (Lange et al., 2007; Xu et al., 2010). The lack of predictable cNLSs in nuclear proteins may partly be reflected by the diversity of nuclear import pathways; however, previous observations in yeast have shown that up to 50% of proteins which bind Imp-α do not have a predictable cNLS (Lange et al., 2007). To evaluate if this holds true for human Imp-α isoforms, I specifically looked at proteins that have documented interactions with an Imp-α family member. To obtain a list of these physical interactors, proteins were retrieved from BioGrid and IntAct databases for the only yeast Imp-α, Srp1, and all seven human Imp-α isoforms (Orchard et al., 2014; Oughtred et al., 2019). To determine which cargos have a predictable cNLS, I used the less stringent RegEx matching to come up with a list of proteins that interact with Imp-α, but do not have a predicted cNLS (Figure 2A). In yeast, approximately 50% of the proteins which associate with Srp1 have a predictable NLS and this is in agreement with previous reports (Lange et al., 2007; Marfori et al., 2011). Human Imp-α1 shows a similar trend to Srp1, where just over 50% of interactors have a predictable NLS. This is in contrast to Imp-α3, α4, α5, α6 and α7, where roughly 25% of their identified interactions do not have a predictable cNLS. Based on these findings, I conservatively estimate that roughly 25–50% of Imp-a cargo in humans do not have a predictable cNLS.

Not all protein interactions reported in databases such as BioGrid or IntAct are binary, making it difficult to determine if a protein is directly binding Imp-α or by indirectly piggybacking on a protein that interacts directly with Imp-α. To evaluate direct binding partners of Imp-α, I explored the recently published HuRI database (Luck et al., 2020). This project involved a Y2H pipeline that tested roughly 17,000 human open reading frames (ORFs) in an 'all-by-all' format. From this dataset I was able to retrieve 102 non-redundant binary interactions from all Imp-α isoforms, except Imp-α8 which had no data available. Further refinement ultimately reduced this down to 59 proteins, as only 67 have evidence of nuclear localization from the HPA and another 8 of which are either

Figure 3.2 Many Imp-α cargo do not have a predictable cNLS.

A) Physical protein interactions, either direct or indirect, for yeast Srp1 and the indicated human Imp-α isoforms were retrieved from BioGrid and IntAct and analyzed for cNLSs using RegEx matching. Prediction shows between 50 and 80% do not have a predictable cNLS. **B)** Direct protein interactions for all Imp-α isoforms except Imp-α8 were retrieved from the Human Reference Interactome (HuRI). Interactors were pooled to remove redundant proteins and checked against the HPA for evidence of nuclear localization before cNLS prediction. Several prediction programs were used to determine a range of predicted cNLSs, which demonstrated between 50 and 80% do not have a cNLS. **C)** Proteins without a predicted cNLS from any of the prediction programs were processed with MEME to identify novel motifs common amongst each protein that might interact with Imp-α. Several of the motifs identified were rich in basic amino acids but did not resemble a cNLS. Disorder prediction using PONDR (VSL2) and DisEMBL shows these motifs are also found within predicted disordered protein regions. **D)** The motif KxRxHxK was searched against the human proteome using SLiMSearch, identifying 37 proteins, which were then analyzed with Metascape. Proteins bearing this motif are most enriched in core nuclear processes like RNA pol II transcription and DNA repair. **E)** Proteins with the KxRxHxK motif were also checked against the HPA for evidence of subcellular localization. Of the 30 proteins with localization information, two-thirds have evidence of nuclear localization. **F)** Reanalysis of tandem mass spectra for protein interactions corresponding to Imp-α1, α5 and α6 from the Nuclear Landscape dataset. All significant interactions were checked against the HPA for nuclear localization before cNLS prediction. Between 50 and 75% of Imp-α cargo do not have predictable cNLS when analyzed with NLStradamus and RegEx matching, respectively. **G)** Identified proteins from the Nuclear Landscape dataset were compared to those from HuRI and IntAct. Comparison shows that the majority of protein identifications from the Nuclear Landscape dataset are not represented within these databases and that these interactions show similar results in the number of proteins without predictable cNLSs.

A

	NLS	No NLS
Srp1	127	112
Imp-α1	66	54
Imp-α3	37	12
Imp-α4	61	21
Imp-α5	61	26
Imp-α6	12	6
Imp-α7	39	14

Percent (NLS vs No NLS)

B HuRI

NLS / No NLS

	RegEx	NLSdb	NLStradamus	cNLS Mapper
No NLS	10	29	23	12
NLS	49	30	36	47

Percent (NLS vs No NLS)

C

bits

ID	Sequence	PONDR	DisEMBL
GEMI2	KHRSHWK	0.6620	Loop/Coil
CYHR1	MDQSHRK	0.5664	
MACD1	KKIPTWK	<0.5	Hot-loop
ARL4A	KLRPLWK	<0.5	Loop/Coil
SC5AB	KNLSHAK	0.5523	
TAD2A	KERQRRK	0.5761	Hot-loop
PLAP	KTVKLWK	<0.5	
SGK1	SSNPHAK	0.8232	Loop/Coil
MDFI	GPKAHRK	0.8928	Loop/Coil
RBPMS	ATSQGWK	0.6434	Loop/Coil

D

transcription elongation
double-strand break repair
tRNA metabolic process
stem cell maintenance
nucleosome organization
DNA replication
skeletal system development
sensory organ development
regulation of cell development

-log10(P)

E

Nuclear / Non-nuclear

Percent

F Nuclear Landscape

NLS / No NLS

	RegEx	NLStradamus
No NLS	94	104
NLS	309	299

Percent (NLS vs No NLS)

G

Nuclear Landscape 366
HuRI 9
6
10
42
21
105
IntAct

76

nucleoporins or importins. Searching these proteins for potential cNLSs using several approaches revealed that between 20–50% do not have a predictable cNLS (Figure 2B). Both RegEx matching and cNLS Mapper predicted cNLSs in roughly 80% of proteins, while NLSdb and NLStradamus predicted 50–60% with a cNLS, likely putting the range of true cNLSs somewhere between the two extremes.

Despite the HuRI dataset being relatively small compared to the number of potential nuclear proteins that may bind Imp-α, these findings demonstrate that a potentially large fraction of Imp-α binary interactions may be mediated by a non-typical cNLS. To explore this idea further, I used the motif elicitation program MEME to determine if this group of proteins from the HuRI dataset have any common motifs (Bailey et al., 2009). First, proteins without a predictable cNLS from each prediction program were combined and reduced into a group of 10 non-redundant proteins. These proteins were then evaluated using MEME to look for minimal motifs that occur once in each protein. Interestingly, the top scoring motif was still enriched with positively charged amino acids despite no resemblance to a true cNLS (Figure 2C). This 7 amino acid motif has the strongest preference for Lys at positions 1 and 7 and His at position 5. Position 6 was consistently either Trp, Arg or Ala and position 3 has a minor preference for Arg. Since short motifs, like the cNLS, are most frequently found within intrinsically disordered regions of a protein, I next searched each protein using the disorder prediction programs DisEMBL and PONDR (Linding et al., 2003; Peng et al., 2006; Van Roey et al., 2014). Results from PONDR (VSL2) show that many of the motifs are within a predicted region of disorder, based on their score being greater than 0.5. Analysis with DisEMBL was similar, with most motifs residing in predicted disordered loops/coils or hot-loops. Overall, this data from a small subset of proteins shows that an alternative motif, divergent from a cNLS, yet possessing several basic residues, may be present.

From the motif generated with MEME, I searched the human proteome for KxRxHxK, since these were the prominent basic amino acids, using SLiMSearch (Krystkowiak & Davey, 2017). This resulted in 37 proteins where this motif could be found within a predicted IDR. Gene ontology analysis of these proteins using Metascape (Y. Zhou et al.,

2019) shows that they are most enriched for core nuclear processes involving RNA polymerase II transcription and DNA repair (Figure 2D). In total, 30 of the 37 proteins had subcellular localization data from the HPA, with 20 having evidence of nuclear localization (Figure 2E). Taken together, these findings suggest that most proteins bearing the KxRxHxK motif are likely nuclear.

Proteins known to associate with Imp-α that can be collected from databases such as IntAct or BioGrid likely only represent a fraction of Imp-α cargo. To extend these findings further I explored datasets which were not available, or not utilized, during previous attempts at characterizing the classical nuclear import pathway in this manner (Bernhofer et al., 2017; Lange et al., 2007). To do this, I reanalyzed publicly available raw mass spectra files published by Mackmull *et al.*, that were obtained through PRIDE and are referred to here as "Nuclear Landscape" (Mackmull et al., 2017). This dataset includes interaction data for Imp-α1, -α5 and -α6 that was acquired through *in situ* proximity ligation (BioID). In their experiments, Imp-α1 and -α5 were expressed as N- and C-terminal BirA* fusions, while Imp-α6 was only expressed with C-terminal BirA*. This approach is highly sensitive and allows protein-protein interactions to be mapped under normal cellular conditions. Briefly, raw tandem mass spectra were searched using MS-GF+ with a reverse target-decoy strategy and the resulting peptides were assembled into proteins using IDPicker, with a global protein FDR < 1% (Elias & Gygi, 2007; Kim & Pevzner, 2014; Ma et al., 2009). Statistically significant interactions were identified using SAINTexpress, with additional background controls provided through the CRAPome, ultimately resulting in a combined 502 high-confidence interactions (Mellacheruvu et al., 2013; Teo et al., 2014). This list of interactors was then compared to proteins localized to the nucleus according to the HPA, resulting in a final list of 403 interactors. Many of the proteins omitted have evidence of nuclear localization; however, for consistency only proteins with evidence in the HPA were used. To establish an estimate of cargos without a predictable cNLS, I used RegEx matching and NLStradamus to determine that roughly 20–25% and 50% of proteins did not have a predicted cNLS, respectively (Figure 2F). These findings echo the results obtained from the HuRI and IntAct datasets (Figure 2A and B), which show a similar number of proteins without a cNLS when using both prediction approaches. It remained possible that these similarities arise due to the analysis of overlapping/redundant proteins

within their respective datasets. However, a comparison of Imp-α interactors from each source demonstrated minimal overlap between proteins identified through my reanalysis and IntAct or HuRI (Figure 2G). Thus, reanalysis of the Nuclear Landscape dataset using both a different mass spectrometry pipeline and statistical protein-protein interaction analysis identified significantly more Imp-α cargos, many of which are novel, yet also do not have a predictable cNLS.

Overall, using several different cNLS prediction programs, 20–50% of proteins which directly bind Imp-α do not have a predicted cNLS. Importantly, these observations are independently observed in the reanalysis of the Nuclear Landscape dataset, which represents hundreds of new Imp-α cargos. When taken together, these data highlight potentially new Imp-α binding motifs and are also highly suggestive that piggybacking strategies are used extensively for Imp-α interactions.

3.3.3 Identification of putative piggybacking proteins

As shown above, roughly 50–60% of proteins known to localize to the nucleus do not have a predictable cNLS. Some of these proteins without a cNLS may instead target one of the importin Kapβs directly. However, there remain many nuclear proteins that associate with Imp-α as determined by proteomic studies, which do not have a predictable cNLS. One situation that would satisfy nuclear import via Imp-α, without the use of a cNLS is through piggybacking, which is simply the indirect association with Imp-α via an intermediary protein (Czeko et al., 2011). Despite a few specific examples of piggybacking as a mechanism of nuclear import, the prevalence of this process remains poorly characterized.

To identify putative piggybacking proteins I used Metascape to first establish a general overview of the cellular processes associated with non-cNLS nuclear proteins from the HPA (Y. Zhou et al., 2019). The rational being that proteins involved in similar cellular processes are most likely to function together. Of the top 10 non-redundant enriched clusters I identified, the top three processes were RNA polymerase II transcription initiation, DNA repair and RNA splicing, which are all nuclear processes (Figure 3A). Further inspection of members within the RNA polymerase II transcription initiation cluster revealed multiple groups of proteins with related functions (Figure 3B). The first

major group consists of proteins belonging to the type-II nuclear receptor family, a class of ligand-regulated transcription factors (Sever & Glass, 2013). Interestingly, most nuclear receptor proteins identified have a predicted cNLS by cNLS Mapper. However, these predicted cNLSs do not align with those identified by experimentation and are likely incorrect. Interestingly, many nuclear receptors have been shown to contain an NLS within their DNA binding domain, specifically within the linker region between zinc-finger domains (Chopin-Delannoy et al., 2003; Hsieh et al., 1998; Prüfer & Barsony, 2002). These motifs appear in many nuclear receptors; however, they do not resemble any previously identified cNLS or non-cNLS (Figure 3C).

Additionally, several proteins were identified that function together in large multi-protein complexes, including subunits of RNA polymerase II (RNAPII), transcription factor II D (TFIID) and Mediator. Identification of RNAPII subunits is encouraging, as RNAPII is already suspected to assemble within the cytoplasm prior to nuclear import (Boulon et al., 2010; Czeko et al., 2011; Di Croce, 2011). Similar to RNAPII, the assembly of TFIID has been proposed to occur within the cytoplasm and subsequently enter the nucleus through a piggybacking mechanism. Specifically, the cTAF subcomplex, consisting of TAF2-TAF8-TAF10 has been shown to shuttle into the nucleus via Imp-α1 (Trowitzsch et al., 2015). My reanalysis of the Nuclear Landscape dataset using SAINTexpress supports these observations, showing statistically significant interactions (FDR < 0.05) between Imp-α1 and several TAF proteins, including TAF2 and TAF8 (Figure 3D). Visualization of these interactions using ProHits-viz shows the highest number of spectral counts between Imp-α1 (C-terminal BirA* fusion) and various TAF proteins (Knight et al., 2017). The N-terminal BirA* fusion of Imp-α1 produced many similar interactions, but with fewer spectral counts. Likewise, Imp-α5 N- and C-terminal BirA* constructs identified similar hits with varying spectral counts, while Imp-α6 produced the fewest hits overall. Despite positive identification of peptides corresponding to TAF10, the interaction between Imp-α1 and TAF10 was not statistically significant according to SAINTexpress. However, with prior knowledge of a TAF2-TAF8-TAF10 complex and a number of other interactions between Imp-α1 and several

Figure 3.3 Identification of putative piggybacking proteins in non-cNLS nuclear proteins.

A) Nuclear proteins from the HPA without a predicted cNLS were analyzed using Metascape to identify enriched cellular processes. **B)** Many of the proteins within the RNA polymerase II (RNAPII) transcription initiation cluster belong to related protein groups, such as the nuclear receptors, and distinct multi-protein complexes like transcription factor II D (TFIID), RNAPII and Mediator. Underlined proteins have a predicted cNLS as determined by cNLS Mapper. RNAPII is known to use piggybacking as well as several subunits of TFIID. **C)** Although not suspected to piggyback, multiple sequence alignment with Clustal Omega of the identified nuclear receptors shows conservation of a motif (red) that has been previously shown to mediate nuclear import (bolded black) (Sievers et al., 2011). **D)** Visualization of TAF interactions with Imp-α1 and 5 (N- and C-terminal BirA* fusions) and Imp-α6 (C-terminal BirA* fusion) shows many TAFs are strongly associated with Imp-α1. Several have a cNLS Mapper score ≥ 7 (green) while others have weaker cNLS Mapper scores that are < 7 but still greater than 5 (yellow). Those in red have scores below 5. Importantly, many of these predicted cNLSs are found within disordered regions (green) as determined by PONDR (VSL2).

A

GO:0006367: RNA pol IIpromoter transcription initiation
GO:0006281: DNA repair
GO:0008380: RNA splicing
R-HSA-381340: Transcriptional regulation of white adipocyte
R-HSA-8951664: Neddylation
GO:0001655: urogenital system development
GO:0031647: regulation of protein stability
GO:0061061: muscle structure development
GO:0048511: rhythmic process
GO:0048732: gland development

-log10(P)

B

NR0B1	POLR2G	TAF9	TEAD2	MED4
RUNX2	POLR2I	TAF10	MED23	NRBF2
CCNC	POLR2J	TBP	MED17	TAF9B
CDK4	RARG	TBX5	MED20	CAND1
CREB1	RORB	TEAD1	NR1D1	MED25
ERCC2	RORC	TEAD4	NR1I3	MED10
GTF2B	RXRA	NR2E1	NR1D2	MED30
NR4A1	RXRB	NR2C1	MED6	SGF29
HNF4G	RXRG	NR2C2	TRIM28	MED8
NR4A2	SOX9	ZNF45	YAP1	
POLR2D	TAF6	TAF15	WWTR1	

Nuclear receptor   Mediator
TFIID   RNAPII

C

RXRA_HUMAN	YGVYSCEGCKGFF**KR**TV**RK**DLTY-TCRDN
RXRB_HUMAN	YGVYSCEGCKGFF**KR**TI**RK**DLTY-SCRDN
RXRG_HUMAN	YGVYSCEGCKGFF**KR**TI**RK**DLIY-TCRDN
RARG_HUMAN	YGVSSCEGCKGFF**RR**SIQKNMVY-TCHRD
RORB_HUMAN	YGVITCEGCKGFF**RR**SQQNNASY-SCPRQ
RORG_HUMAN	YGVITCEGCKGFF**RR**SQRCNAAY-SCTRQ
NR1D1_HUMAN	YGVHACEGCKGFF**RR**SIQQNIQY**KR**CLKN
NR1D2_HUMAN	YGVHACEGCKGFF**RR**SIQQNIQY**KK**CLKN
NR1I3_HUMAN	FNALTCEGCKGFF**RR**TVSKSIGP-TCPFA
NR2C1_HUMAN	YGAVTCEGCKGFF**KR**SI**RK**NLVY-SCRGS
NR2C2_HUMAN	YGAVSCEGCKGFF**KR**SV**RK**NLTY-SCRSN
NR2E1_HUMAN	YGVYACDGCSGFF**KR**SI**RR**NRTY-VCKSG
NR4A1_HUMAN	YGVRTCEGCKGFF**KR**TVQKNAKY-ICLAN
NR4A2_HUMAN	YGVRTCEGCKGFF**KR**TVQKNAKY-VCLAN

D

TAF9B TAF1 TAF12 TAF4 TAF3 TAF8 TAF6 TAF6L TAF5L TAF9 TAF2 BTAF1 TAF7 TAF5 TAF15 TAF10

Imp-α6c
Imp-α5c
Imp-α5N
Imp-α1c
Imp-α1N

cNLS
IDR

Spectral Count
0   50

Relative abundance

FDR
≤ 0.01   ≤ 0.05   ≥ 0.05

TAFs, the Imp-α1-TAF10 interaction is likely accurate. Additionally, many individual subunits of the 5TAF (TAF4, 5, 6, 9, 12) and sTAF (TAF1, 7, 11, 13 and TBP) subcomplexes appear to preferentially associate with Imp-α1. Interestingly, despite ample evidence in support of piggybacking, many TFIID subunits have predictable cNLSs within a predicted intrinsically disordered region. With the exception of TAF15 which has a PY-NLS, only TAF6 has no predictable cNLS (Marko et al., 2012).

Based on these findings, my analysis of nuclear proteins without a predictable cNLS identified protein subunits of RNAPII and TFIID already shown to piggyback into the nucleus. In contrast, many of the Mediator proteins identified do not have a predictable cNLS, a particular area that has remained largely unexplored and could possibly represent a novel example of piggybacking.

3.3.4 Mediator proteins associate with Imp-α and do not have a predictable cNLS

Mediator, like RNAPII and TFIID, is a multiprotein complex consisting of up to 30 subunits. Despite being relatively well characterized with respect to its role in transcriptional coactivation, nuclear import of Mediator proteins has not been studied extensively. Furthermore, evidence of cytoplasmic assembly prior to nuclear import via a piggybacking mechanism has not been previously proposed.

To investigate the Mediator complex further, I first inspected each Mediator subunit for a predictable cNLS using RegEx matching, NLStradamus and cNLS Mapper (Figure 4A). Of the 30 Mediator subunits evaluated, RegEx matching was the most sensitive, identifying 12 proteins with a cNLS, most of which were confirmed with NLStradamus and/or cNLS Mapper. For the remaining 18 Mediator subunits without a RegEx predicted cNLS, only 3 were predicted to have a cNLS using one of the other prediction programs. Overall, using each cNLS prediction method only 11 of the 30 proteins have a cNLS predicted by at least two approaches, suggesting many subunits may use alternative nuclear import pathways, alternative cNLSs, or possibly piggyback into the nucleus.

In addition to cNLS prediction, I inspected the Nuclear Landscape dataset along with a literature search for interactions between Mediator proteins and NTRs. In addition to Imp-α1, α5 and α6, the Nuclear Landscape dataset also contains information for other NTRs and includes several importin Kapβ proteins (Kpnb1, IPO4, IPO5, IPO11 and IPO13) as well exportin Kapβs (NXT1, NXT2, XPO1, XPO2 and XPO7). From this dataset, 22 components of the Mediator complex were identified as having an association with at least one NTR (Figure 4A). Although most Mediator subunits interact with at least one Imp-α protein, many do not have anything resembling a cNLS. For MED7 and MED27 specifically, putative cNLSs were identified using RegEx, but not by NLStradamus and cNLS Mapper, suggesting these cNLSs may not be valid. Interestingly, while many Mediator proteins associate with multiple Imp-α isoforms, or importin Kapβ transporters like TNPO1 and 2, and IPO4, 5 and 11, none exclusively associate with only Kapβ proteins. In other words, these Mediator associations always co-occur with an Imp-α.

Due to the physical limitations imposed by the NPC, nuclear import of larger proteins requires facilitated nuclear transport pathways. Individual Mediator components range from 13kDa to over 200kDa. Not surprisingly, as molecular weight increases, so does the likelihood of a protein having a predictable cNLS (Figures 4A & B). Most Mediator subunits without a predictable cNLS are less than 50kDa, and in theory may enter the nucleus via passive diffusion. In contrast, both MED23 and MED25 exceed the NPC diffusion limit and lack a predicted cNLS. Given the extensive number of interactions made within the Mediator complex (Figure 4C), it's plausible that MED23, MED25 and many of the smaller components lacking cNLSs piggyback into the nucleus with the larger cNLS-bearing subunits.

Based on these analyses, it appears that the classical nuclear import pathway is responsible for nuclear import of the majority of Mediator subunits, while alternative pathways using Kapβs may be used to a lesser extent. It's particularly interesting that most Mediator subunits associate with the classical NTR Imp-α, yet do not have anything resembling a cNLS, suggesting that Mediator components may piggyback into the nucleus as complexes as described for RNAPII and TFIID.

Figure 3.4 Mediator complex subunits may utilize a piggybacking mechanism for nuclear import.

A) Mediator subunits were analyzed for cNLSs using different cNLS prediction programs, and data was tabularized using Microsoft Excel. Many subunits have a predicted cNLS (green) from more than one program while the majority do not have a predicted cNLS (red). Data from the Nuclear Landscape dataset and other published NTR interactions show that many subunits associate with Imp-α, as well as transportin (TNPO). **B)** Mediator subunits vary in molecular weight, with larger subunits more frequently having a predicted cNLS. Subunits with a cNLS predicted from two programs or more are shaded in dark green (2 NLS) and those with a prediction from only one program are shaded in light blue (1 NLS). Although imprecise, a passive diffusion limit of 50kDa (dotted line) shows many subunits without a cNLS are below this cut-off. **C)** A model figure of Mediator was adapted from Soutourina, 2018, to show corresponding subunits with predicted cNLSs as well as Imp-α associations.

A

NLS Prediction

Mediator Subunit	RegEx	NLStradamus	cNLS Mapper	NTR	NTR Source
MED13	green	green	green	No evidence	
MED14	green	red	green	IMA1/5/6, IPO11, TNPO1/2	Nuclear landscape
MED27	green	red	red	IMA1/5, IPO11, TNPO1/2	Nuclear landscape
MED31	green	red	green	No evidence	
MED7	green	red	red	No evidence	
MED1	green	red	red	IMA1/5, TNPO1/2	Nuclear landscape
MED12	green	green	green	IMA1/5/6	Nuclear landscape
MED15	green	red	green	IMA1, TNPO1	Nuclear landscape
MED19	green	green	green	IMA1, TNPO2, IPO5	Nuclear landscape
MED26	green	green	green	IMA3	Sato et al., 2004
CDK8	green	red	green	IMA5	Nuclear landscape
MED24	green	red	green	Kpnb1	Hein et al., 2015
MED11	red	red	red	IMA1	Nuclear landscape
MED17	red	red	green	IMA1/5/6, TNPO2	Nuclear landscape
MED18	red	red	red	No evidence	
MED10	red	red	red	IMA1/6	Nuclear landscape
MED16	red	red	green	No evidence	
MED20	red	red	red	IMA1/5, TNPO2	Nuclear landscape
MED21	red	red	red	IMA1	Nuclear landscape
MED22	red	red	red	IMA1/6, IPO11	Nuclear landscape
MED23	red	red	red	IMA5	Nuclear landscape
MED25	red	red	red	No evidence	
MED28	red	red	red	IMA1/5	Nuclear landscape
MED29	red	red	red	No evidence	
MED30	red	red	red	IMA1/5	Nuclear landscape
MED4	red	red	red	IMA1/6, IPO4	Nuclear landscape
MED6	red	green	green	IMA1	Nuclear landscape
MED8	red	red	red	IMA1/5/6 TNPO1	Nuclear landscape
MED9	red	red	red	IMA1/5	Nuclear landscape
CCNC	red	red	red	No evidence	

B

Molecular Weight (kDa)

- 2 predicted NLS
- 1 predicted NLS
- No predicted NLS

MED11, MED21, MED10, MED31, MED9, MED28, MED30, MED29, MED22, MED20, MED18, MED19, MED7, MED6, MED8, MED4, CCNC, MED27, CDK8, MED26, MED17, MED25, MED15, MED16, MED24, MED23, MED14, MED1, MED13, MED12

C

13/13L

12/12L

CCNC

CDK8

- ○ Imp-α association
- Predicted NLS in 1 program
- Predicted NLS in 2 programs

19, 26, 10, 7, 31, 21, 6, 4, 9, 8, 22, 17, 11, 14, 1, 18, 20, 28, 30, 29, 27, 15, 16, 24, 25, 23

3.4 Discussion

Here, I performed a general analysis of protein nuclear import and highlight several novel and interesting observations. In general, my results extend previous findings found in model organisms and provides evidence that nuclear import signals are absent in a major fraction of the human nuclear proteome.

Overall, my approach used RegEx matching to identify predicted cNLSs within the human nuclear proteome, which demonstrated that approximately 50% of nuclear proteins from the HPA have a predictable cNLS. Importantly, these findings are based on the assumption that each predicted cNLS is accessible to Imp-α and resides within a predicted IDR. Indeed, many predicted cNLSs likely reside within an IDR and are non-functional. However, the primary objective was to identify proteins without a cNLS. Applying IDR prediction to proteins without a predictable cNLS would not provide any additional information therefore, these assumptions were necessary for creating a high confidence, conservative list of non-cNLS bearing nuclear proteins. Based on these findings, NLS predictions using RegEx matching and NLStradamus suggests that somewhere between 47–63% of nuclear proteins from the HPA do not have a predictable cNLS.

Analysis of nuclear proteins obtained through the HPA showed a large discrepancy between the presence of predicted cNLSs (<40%) in nuclear proteins and predicted NESs (>80%) in cytoplasmic proteins. It's unlikely cNLS prediction is simply worse than NES prediction, given the fact that both types of motifs have been extensively studied. Rather, this could reflect the diversity in pathways that control protein import or export. All proteins are translated within the cytoplasm, therefore nuclear proteins require a process to reliably pass through the NPC, in contrast to cytoplasmic proteins that function in the same subcellular compartment they are translated in. Interestingly, over 80% of cytoplasmic proteins have a predictable NES, when in theory this is unnecessary. Possibly, many NESs serve to simply export cytoplasmic proteins that may drift into the nucleus or become localized to the nucleus upon nuclear envelope reformation after mitosis. In these instances, it's possible that the XPO1 pathway is responsible for dealing with these scenarios. In fact, this line of reasoning is supported by experimental evidence suggesting XPO1-mediated

export functions as a countermeasure to help define the nuclear and cytoplasmic compartments (Kırlı et al., 2015).

In contrast to NESs, less than 40% of nuclear proteins have a predicted cNLS. NLStradamus uses a relatively stringent statistical model to predict NLSs and this includes both cNLSs and non-cNLSs that could bind Kapβs. For this reason, I also used a non-statistical approach that uses simple RegEx matching, which likely over-predicts many cNLSs. Paradoxically, the over-predictive nature of this approach is well suited for finding proteins without anything resembling a cNLS. Surprisingly, RegEx matching only identified cNLSs in 53% of nuclear proteins, whereas NLStradamus predicted cNLSs in only 38% of proteins, which is similar to results obtained from analysis of 2163 human nuclear proteins with NLSdb (Bernhofer et al., 2017). Based on these findings, I conservatively estimate that at least 50% of nuclear proteins in humans do not have a cNLS. Roughly one-third of these proteins are predicted to meet one of the requirements of a PY-NLS by having a R/H/K-X_{2-5}-PY motif. However, this is only one of the criteria of a PY-NLS and the large majority of these are probably not true PY-NLSs (Soniat & Chook, 2015). Nevertheless, even if these were true PY-NLSs, this leaves a substantial portion of the nuclear proteome without any predictable NLS. Other variants of the PY-NLS exist that do not have the PY motif, or instead have PL in place of PY; however, only a few examples of these exist and there is no way to determine how abundant these motifs are within the nuclear proteome (Kressler et al., 2012; Soniat et al., 2013; Soniat & Chook, 2016).

The discrepancy between cNLS and NES prediction is also apparent in nucleocytoplasmic proteins. Hypothetically, these proteins should possess both targeting motifs; however, roughly only 30% contain a cNLS while ~80% have an NES (Gama-Carvalho & Carmo-Fonseca, 2001). Interestingly, this leaves more than 10% of nucleocytoplasmic proteins without either a predictable cNLS or NES. In addition, up to 25% of cytoplasmic proteins have predicted nuclear localization, as well as a putative cNLS. Although immunofluorescent imaging is highly informative for protein localization, a single image — or even several — only provide information at a particular point in time and context. It is possible that at least some proteins documented as cytoplasmic have short tenures within

the nucleus in response to a particular stress or stimulus that is not captured through tissue culture-based experiments.

To date, cNLS prediction on human Imp-α binding partners has not been performed. In yeast, it has been shown that 50% of Srp1 binding partners do not have a cNLS (Lange et al., 2007). However, because these data were collected using yeast proteins in a yeast system, indirect binding to Srp1 cannot be ruled out. Analysis of Imp-α interactors from BioGrid and IntAct show that between 50–70% do not have a predictable cNLS. However, whether or not these interactions are direct remains unclear. Imp-α data taken from the HuRI database is less likely to be impacted by indirect binding, since binary interactions of human proteins were tested in yeast, and it is less likely for yeast proteins to facilitate human protein interactions. Using either RegEx matching or cNLS Mapper, I determined that at least ~20% of human Imp-α interactors do not have a predictable cNLS. This raises the possibility that a novel, as yet unidentified binding motif is responsible for a subset of Imp-α interactions.

A widely used computational approach for identifying novel motifs is based upon the assumption that multiple unrelated proteins that interact with the same protein are likely to use the same, or highly similar, interaction motif (Davey et al., 2010; Gibson et al., 2015; Neduva et al., 2005). Using proteins from the HuRI dataset without a putative cNLS, I attempted to find a consensus motif using MEME (Bailey et al., 2009). This identified a motif that was rich in positively charged amino acids. This motif doesn't conform to a typical cNLS; however, it shares important properties such as basic amino acids and localization within predicted disordered protein regions. Using this consensus, I searched the human proteome for the motif KxRxHxK. Interestingly, proteins with this motif are enriched in nuclear processes and have evidence of nuclear localization. The consensus motif identified should be taken with careful consideration, since many of the positions do not have a clear amino acid preference. Additionally, the sequences identified in Figure 2C may be reminiscent of importin-α C-terminal binding segment (iCBS)-NLSs, which bind a C-terminal region of Imp-α instead of the major or minor grooves, are rich in basic amino acids, but do not appear to conform to any regular pattern (Arjomand et al., 2014).

Data available through resources such as IntAct and HuRI only provide a limited number of Imp-α interactions. For example, the HuRI dataset tested each Imp-α isoform against 17,000 human ORFs, yet reported only ~250 interactions for all Imp-α isoforms combined. These numbers are surprisingly low, considering thousands of proteins are localized to the nucleus. This prompted us to search for additional Imp-α interactions by reanalyzing proteomic data from mass spectrometry repositories, such as PRIDE. Here, I identified a dataset (referred to as Nuclear Landscape) that used BioID, a proximity ligation technique designed to capture protein interactions *in situ*, with Imp-α1, -α5 and -α6 (Gingras et al., 2019; Mackmull et al., 2017). In these experiments, Imp-α was expressed as a fusion protein with BirA* on either the N- or C-terminus. In the presence of exogenously supplied biotin, Imp-α-BirA* will biotinylate proximal proteins *in vivo,* which can then be subsequently identified through streptavidin-based affinity purification and mass spectrometry. Since proximal and directly interacting proteins are biotinylated directly, this approach is more sensitive to detecting piggybacking interactions than standard affinity purification approaches, as stable interactions are not required during sample preparation. The majority of Imp-α-associated proteins identified through this reanalysis are not represented in the HuRI or IntAct datasets. Importantly, RegEx matching shows at least 20–25% do not have a predictable cNLS, in agreement with cNLS prediction performed on proteins retrieved from HuRI and IntAct. Thus, this independent, experimentally based method of detecting Imp-α-associated proteins confirms that many nuclear proteins do not have a predicted cNLS. This may reflect piggybacking into the nucleus, since these interactions would not necessarily be detected through binary interaction studies performed in yeast.

Having established that many nuclear proteins and Imp-α cargo do not have a predictable cNLS, I next wanted to identify putative piggybacking proteins. A Metascape analysis of the cellular processes enriched with nuclear proteins without a predictable cNLS identified the RNA polymerase II transcription initiation pathway. Within this group were many proteins belonging to a subfamily of the nuclear receptors. Although not suspected of piggybacking, alignment of the region located between zinc-fingers shows conservation of an experimentally validated NLS. This NLS has been shown to be active in other nuclear receptors like the vitamin D receptor, RXR and NR1D1/2 (Rev-Erbα/β) (Chopin-Delannoy

et al., 2003; Hsieh et al., 1998; Prüfer & Barsony, 2002). The non-classical appearance of this motif, and divergence from other NLSs in general, makes it interesting from a nuclear import perspective and warrants further investigation.

Intriguingly, several proteins represented within the group of cargo without predicted cNLSs are already known to use piggybacking and these mainly belong to the RNAPII complex, where assembly has been shown to take place within the cytoplasm prior to nuclear import (Boulon et al., 2010; Di Croce, 2011). Additionally, several TAF proteins belonging to the TFIID complex were identified in this group. TAF8 and TAF10 assemble co-translationally within the cytoplasm and shuttle into the nucleus along with TAF2 (Kamenova et al., 2019; Soutoglou et al., 2005; Trowitzsch et al., 2015). Similarly, TAF6 and TAF9, as well as TAF1 and TBP assemble co-translationally and may also piggyback into the nucleus (Antonova et al., 2018; Kamenova et al., 2019; Patel et al., 2020). Analysis of TAF proteins identified in my reanalysis shows that most subunits have a cNLS with a cNLS Mapper score greater than 7, which is considered sufficient to localize EGFP to the nucleus. This is an interesting observation, considering many of these proteins are suspected to piggyback into the nucleus as subunits of larger multi-protein complexes. Whether or not these cNLSs are functional or even accessible to Imp-α is unknown. However, based on my reanalysis of the Nuclear Landscape dataset, the majority of these subunits associate with Imp-α, suggesting some of these cNLSs may be accessible for binding. It's possible that pre-assembled TFIID is imported into the nucleus in a manner where multiple pre-assembled subunits are able to independently contact Imp-α.

In contrast to RNAPII and TFIID, components of the Mediator complex have not been reported to piggyback into the nucleus. Mediator is an evolutionarily conserved multi-protein complex composed of up to 30 subunits and is a key component of transcription regulation (Bourbon, 2008; Soutourina, 2018). Mediator's main function is to bridge interactions with transcription factors at enhancer regions with transcriptional machinery assembled at promoters (Poss et al., 2013). The composition of Mediator can be subdivided into the head (MED6/8/11/17/18/20/22/28/30), middle (MED1/4/7/9/10/19/21/26/31), tail (MED15/16/23/24/25/27/29) and kinase module (MED12/13, CCNC and CDK8 or CDK19). Intriguingly, the large majority of Mediator proteins do not have a predictable

cNLS but were still observed to associated with Imp-α according to my analysis of the Nuclear Landscape dataset. Based on these findings, it is highly likely that Mediator subunits utilize a piggybacking mechanism similar to that employed by RNAPII and TFIID. Furthermore, there appears to be a trend across all Mediator modules where smaller subunits may piggyback on their larger cNLS-bearing binding partners. Although these smaller subunits could diffuse into the nucleus, active transport via piggybacking may preserve stoichiometric ratios and import rates necessary for this essential function. Furthermore, association between the smaller Mediator subunits and Imp-α clearly support active transport and not passive diffusion. Of the individual modules, the head module may represent a good starting point to explore piggybacking, as it had the fewest subunits with a predicted cNLS. MED14, which links the head, middle and tail modules contains a cNLS and could possibly nucleate piggybacking of several Mediator proteins as well (Tsai et al., 2014).

The fact that RNAPII, TFIID and potentially Mediator use piggybacking for nuclear localization is interesting, given that they all function in formation of the PIC. The assembly of such multi-subunit complexes in the cytoplasm and subsequent co-transport via piggybacking into the nucleus suggests that this may be important for their respective functions. Transport through the NPC is rapid; however, proteins of different sizes transport at different rates (Lolodi et al., 2016). Pre-assembled complexes can traffic at a uniform rate and arrive at the nucleus as a complete functional unit, rather than import individually at different rates with subsequent piece-by-piece assembly at an enhancer or promoter.

Overall, my data highlights several interesting observations regarding nuclear transport. Using just RegEx's, more nuclear proteins with a predicted cNLS are identified than previously reported. However, at least ~50% of human nuclear proteins do not have a predictable cNLS. I also showed for the first time that at least 20% of proteins that bind a variety of human Imp-α isoforms do not have a predictable cNLS. Taken together, many nuclear proteins likely localize by extensive use of non-classical nuclear import pathways, as well as by piggyback mechanisms. Analysis of nuclear proteins without cNLSs provides additional support for piggybacking of the TFIID complex into the nucleus and suggests

that the Mediator complex similarly piggybacks into the nucleus. Overall, these results demonstrate the need for deeper investigation into alternative NLSs and nuclear piggybacking mechanisms.

Chapter 4

4 A novel protein nuclear import-based approach for discovery of short linear motifs

4.1 Introduction

The molecular composition and functional diversity of a cell consists of several layers that extend far beyond that of the genome. Notably, 'omics'-based approaches have revealed a non-linear relationship in complexity between the genome, transcriptome, proteome and interactome (Bludau & Aebersold, 2020). The human genome for instance, which encodes roughly 20,000 protein-coding genes, is currently estimated to express >84,000 annotated protein-coding transcripts (Frankish et al., 2019). At the proteomic level, the number of proteoforms easily increases into the hundreds of thousands considering the number of possible PTMs and additional sources of protein variation (Aebersold et al., 2018). Most importantly, how these proteins interact with one another to make a network, or interactome, is a fundamental attribute underlying the complexity of biological organisms. Efforts in characterizing the human protein interactome, such as through the HuRI, has produced a map of over 60,000 binary PPIs (Luck et al., 2020). Similarly, the most recent update from the Bioplex Interactome project now includes nearly 120,000 PPIs from roughly 15,000 human proteins (Huttlin et al., 2015). Importantly, these efforts also reveal that our understanding of the protein interactome is still within its infancy given that statistical estimates suggest there are roughly 650,000 PPIs (Stumpf et al., 2008).

Historically, PPI studies have been dominated by globular protein domains, which form stable, well-folded tertiary structures. More recently however, it has become clear that protein functionality can exist independent of structure. Roughly one-third of the human proteome is intrinsically disordered and yet despite being the smaller fraction, this is where much of the proteomes signalling information is routed via events such as PTMs and PPIs (Darling & Uversky, 2018; Iakoucheva et al., 2004; Van Roey et al., 2014). The most common functional units found within IDPRs are stretches of short amino acid sequences, often less than 10 amino acids in length, known as SLiMs (Davey et al., 2012). Within a typical SLiM, only a few amino acids are necessary for function. This makes SLiMs

evolutionarily plastic, and likely explains why they are essential for both cellular and viral protein function (Davey et al., 2011; Hu et al., 2017). According to the ELM resource, whose aim is to catalogue experimentally verified SLiMs, there are over 3500 motif instances, spanning a wide range of cellular processes (Kumar et al., 2020).

In contrast to studying stable PPIs between two globular protein domains, interactions mediated through SLiMs are often transient in nature, making them much less amenable to experimental discovery (Davey et al., 2012; Gibson et al., 2015). In light of these difficulties, it's anticipated that roughly 100,000 SLiMs (not including PTMs) may reside within the human proteome, presenting a large discrepancy between the number of currently identified and predicted SLiMs (Tompa et al., 2014). This incongruity highlights an important need for high-throughput SLiM discovery approaches. Indeed, several SLiM discovery methods employing bioinformatic and experimental approaches have been utilized to great success (Neduva et al., 2005). However, computational power cannot be fully realized when so few of the predicted motifs have been identified, further emphasizing the need for high-throughput experimental approaches to identify novel functional SLiMs.

Experimental approaches such as protein/peptide arrays, phage display of peptides derived from disordered regions of the human proteome (ProP-PD) or spectrally encoded beads presenting peptides (MRBLE-pep) allow for high-throughput screening of peptides that can bind a specific protein domain (Davey et al., 2017; K. Meyer & Selbach, 2020; Nguyen et al., 2019). These approaches employ a "one-vs-many" framework, where one protein domain is screened against a library of peptides. Given that thousands of protein domain families exist within the human proteome a "one-vs-many" model is not well suited to address these questions on a larger scale (El-Gebali et al., 2019). More recently, an approach termed protein interaction screen on peptide matrix (PRISMA) allows for "many-vs-many" peptide-protein interaction screens (Dittmar et al., 2019). This approach significantly increases throughput compared to previously established methods. However, only a few hundred different peptides have been tested so far, and like peptide/protein-arrays these experiments are carried out *in vitro*. Based on these limitations, I have developed a novel "many-vs-many" approach using the model organism *S. cerevisiae*. This

system exploits the cellular protein nuclear import pathway and is built on the assumption that if a peptide can mediate nuclear import, or interact with a nuclear protein, it can be genetically selected for and identified.

Transport across the nuclear envelope occurs exclusively through the NPC (Wente & Rout, 2010). This process is highly regulated and uses a group of proteins called karyopherins, which have the unique ability to shuttle proteins through the NPC (Christie et al., 2016; Otsuka et al., 2008). The most well-defined transport pathway is the highly conserved classical nuclear import pathway, which is facilitated by the karyopherin proteins Imp-α and Imp-β1 (Goldfarb et al., 2004b; Lange et al., 2007). Yeast express a single Imp-α (Srp1) while humans express 7 different isoforms, all of which bind Imp-β1 through their IBB domain (Lott & Cingolani, 2011; Pumroy & Cingolani, 2015). Imp-α recognizes a protein's NLS, a class of SLiM characterized by one (monopartite) or two (bipartite) clusters of basic amino acids. These NLSs are formally known as cNLSs. Following cNLS recognition, the Imp-α-cNLS complex is transported through the NPC via Imp-β1 (Moroianu et al., 1995). Although other karyopherin-mediated import pathways exist, the classical pathway is widely accepted to be responsible for the majority of protein nuclear import (Christie et al., 2016; Lange et al., 2007). Interestingly, however, many nuclear proteins do not have predictable cNLSs, yet still bind Imp-α in yeast and human cells. These observations suggest that alternative pathways such as piggybacking, or novel SLiMs which could facilitate nuclear import, are present (Lange et al., 2007; Tessier et al., 2020).

Based on these observations, I designed a selection-based assay in yeast that exploits protein nuclear import to identify novel SLiMs. Here, randomly generated peptides are selected for their ability to either mediate nuclear import or possibly bind a nuclear protein and enter the nucleus via a "piggybacking" mechanism. To enrich selection for non-cNLS motifs, peptides are generated based on a semi-degenerate codon library that eliminates arginine and most lysine codons. This novel "many-vs-many" approach is unique since it allows for high throughput testing of peptides against the yeast nuclear proteome in a functional, *in vivo* setting.

4.2   Material and Methods

4.2.1   Yeast culture, transformation and β-galactosidase assay

Prior to transformation, yeast (Table 4.1) were streaked onto plates containing standard YEP media (yeast extract, peptone and 2% glucose) and grown at 30°C until individual colonies became fully visible. A single colony was chosen and grown overnight in YEP liquid culture at 30°C. The following morning, cultures were diluted if overgrown, allowing them to grow until reaching an OD_{600} between 0.5 and 1. For a single transformation, 1 mL of overnight culture was washed in 1 mL of sterile double distilled water (ddH$_2$O) and resuspended in 100 mM lithium acetate; which was then incubated at 30°C for 5 minutes. Yeast were pelleted and resuspended in 50 μL ddH$_2$O before adding a solution containing 240 μL of 50% polyethylene glycol (PEG) 3350, 36 μL of 1 M lithium acetate, 25 μL of 2 mg/mL salmon sperm DNA, and 100–500 ng of plasmid (see Table 4.2 for list of plasmids used in this chapter). This mixture was vortexed and incubated at 42°C for 30 minutes. Following incubation, yeast were pelleted and resuspended in 100 μL ddH$_2$O and plated on appropriate synthetic drop-out (SD) media containing 2% glucose.

For the Gal4 repression assays, W303-1A yeast were transformed with 100–500 ng of both a Gal4 responsive *lacZ* reporter plasmid (JMB1404) and Gal4$_{DBD}$ peptide fusions and plated on SD media containing glucose and grown at 30°C until colonies were large enough to pick (1-2 mm). The selected colonies were grown overnight at 30°C in SD media containing 2% raffinose. The following morning yeast cultures were diluted 1:5 into fresh SD media with raffinose and then supplemented with galactose (0.25% final concentration) for 4 hours to induce endogenous Gal4 expression.

For β-galactosidase assays, transformed yeast colonies were picked and grown overnight in 5 mL of appropriate liquid culture at 30°C on a rotating drum. The following morning, OD_{600} values were recorded for yeast cultures and if necessary, cultures were diluted to an OD_{600} between 0.2–0.4 and allowed to grow until values were between 0.5–1.0. For each

Table 4.1 Yeast strains used in this chapter.

Strain	Genotype	Purpose
L40	MATa leu2 his3 trp1 ade2 GAL4 gal80 LYS2::(lexAop)4-HIS3 URA3::(lexAop)-lacZ	pNIA2 nuclear import
W303-1A	MATa leu2-3,112 trp1-1 can1-100 ura3-1 ade2-1 his3-11,15	Gal4 repression assay

Table 4.2 Plasmids used in this chapter.

Clone	Vector	Description
LexA-Gal4 (2µ)	pNIA (Rhee et al., 2000)	Original vector obtained for Rhee et al., 2000
LexA-Gal4-MBP (2µ)	pNIA	
LexA-Gal4 TAg NLS (2µ)	pNIA	
LexA-Gal4 (CEN)	pNIA (Marshall et al., 2007)	Backbone created by Marshall et al., 2007 for expressing pNIA in a low copy CEN plasmid
LexA-Gal4-MBP (CEN)	pNIA (Marshall)	
LexA-Gal4 TAg NLS (CEN)	pNIA (Marshall)	
pNIA2		Reassembled pNIA system used for expressing peptides internally
pNIA2-TAg NLS (Internal)	pNIA2	
pNIA2-TAg NLS (C-terminal)	pNIA2	
pNIA2-E1A NLS (Internal)	pNIA2	
pNIA2-E1A NLS (C-terminal)	pNIA2	
pNIA2-NP	pNIA2	
pNIA2-cMYC	pNIA2	Human cMYC NLS
pNIA2-Tat	pNIA2	HIV-1 Tat protein NLS
pNIA2-SGSG	pNIA2	Serine-glycine (SG) linker
pNIA2-AIR2	pNIA2	
pNIA2-PAP2	pNIA2	
pNIA2 NM1–34	pNIA2	Isolated plasmid from pNIA2 screen, with peptides.
Gal4$_{DBD}$	pAS1	
Gal4$_{DBD}$ NM1–34	pAS1	Peptides from pNIA2 selection cloned onto Gal4$_{DBD}$
Gal4$_{DBD}$-E1A (1-82)	pAS1	Transactivation positive control
Gal4$_{DBD}$-E1A (1-29)	pAS1	Transactivation negative control
Gal4$_{DBD}$-TAg NLS	pAS1	
lacZ reporter		Used in Gal4 repression assay, contains 5 Gal4 binding sites. Acquired from Dr. Chris Brandl.
EGFP	pEGFP-C2	Multiple cloning site of EGFP (JMB925) has been adjusted so BamHI is in-frame.
EGFP TAg NLS	pEGFP-C2	NLS from simian virus 40 large T antigen
EGFP Rev NES	pEGFP-C2	NES from HIV-1 Rev
EGFP NM-9	pEGFP-C2	
EGFP NM-34	pEGFP-C2	
FLAG-Imp-α1	pcDNA3	
FLAG-Imp-α3	pcDNA3	
FLAG-Imp-α5	pcDNA3	

NLS (nuclear localization signal); NES (nuclear export signal); 2µ (high copy); CEN (low copy)

β-galactosidase experiment, 500 μL of liquid culture was pelleted and resuspended in 1 mL Z-buffer (60 mM $Na_2HPO_4\bullet7H_2O$, 40 mM $NaH_2PO_4\bullet H_2O$, 10 mM g KCl, 0.1 mM, $MgSO_4\bullet7H_2O$, in dd H_2O). β-mercaptoethanol (BME) (27 μL BME per mL Z-buffer) was added fresh to the appropriate amount of Z-buffer before use. Next, 20 μL each of 0.1% SDS and chloroform were added to each sample, vortexed for 30 seconds and incubated at 30°C for 15 minutes. Following incubation, 200 μL of ortho-nitrophenyl-β-galactoside (ONPG; 4 mg/ml diluted in Z-buffer with BME) was added, samples were vortexed for 30 seconds and incubated at 30°C until samples developed a visible yellow colour. At this point, time (minutes) was recorded and 500 μL of 1 M sodium carbonate was added to stop the reaction, samples were centrifuged at max speed for 5 minutes and OD_{420} was recorded using 200 μL of each sample's supernatant. β-galactosidase activity was expressed in Miller units using the formula:

$$Miller\ unit = 1000 * \frac{OD420}{OD600 * \text{volume (ml)} * \text{time (min)}}$$

4.2.2 Plasmid construction

Construction of pNIA2 followed the same methods used to build pNIA (Marshall et al., 2007). However, MBP was cloned onto the C-terminus of a $LexA_{DBD}$-$Gal4_{AD}$ recombinant protein. Positive control NLSs (TAg, E1A, cMYC, NP and Tat) and negative control peptides (AIR2, PAP2, SGSG) were cloned into pNIA2 using self-annealed oligonucleotides with EcoRI compatible overhangs (Table 4.3).

For Gal4 repression assays and EGFP immunofluorescence, the region encoding peptides selected for in pNIA2 were amplified using forward (Fw) and reverse (Rv) NHS BamHI primers (Table 4.4) PCR products were gel purified and digested with BamHI and cloned into the same sites of $Gal4_{DBD}$ and EGFP. This produces $Gal4_{DBD}$ and EGFP C-terminal peptide fusions with flanking SGSG linkers.

4.2.3 Yeast plasmid isolation and western blotting

Plasmid isolation from yeast was performed using the "smash and grab" method (Robzyk & Kassir, 1992). From the yeast cultures grown for β-galactosidase assays, 2 mL was

Table 4.3 Self-annealing oligonucleotides used in this chapter.

Oligonucleotide	Sequence	Amino acid (motif)
TAg NLS	AATTGTCTGGATCAGGTCCTCCAAAAAAGAAGAGAAAGGTATCAGGATCTGGTC	PPKKKRKV
E1A NLS	AATTGTCTGGATCAGGTCTGTCTTGTAAACGCCCCAGGCCATCTGGATCAGGAC	LSCKRPRP
cMYC NLS	AATTGTCTGGATCAGGTCCGGCGGCCAAAAGGGTGAAATTAGATTCTGGATCAGGTC	PAAKRVKLD
SGSG linker	AATTGTCTGGATCAGGTTCAGGTC	SGSGSG
Tat NLS	AATTCTCTGGATCAGGTAGGAAGAAGCGGAGACAGCGACGAAGATCTGGATCAGGTG	RKKRRQRRR
Nucleoplasmin NLS	AATTCTCTGGATCAGGTGCTGTTAAAAGACCAGCTGCAACTAAAAAGGCAGGTCAAGCTAAAAAGAAGAAATTGGATTCTGGATCAGGTG	AVKRPAATKKAGQAKKKKLD
AIR2 ELM accession ELMI002850	AATTCTCTGGATCAGGTTTAAGAGCTCTTAGAGGGCAGGGTAGATATTTTGGCGTAAGCGATGATGACAAGGATGCCTCTGGATCAGGTG	LRALRGQGRYFGVSDDDKDA
PAP2 ELM accession ELMI002849	AATTCTCTGGATCAGGTACATATATCACTGTCTCTAGCGAAGATGATGATGAAGATGGATATAATCCTTATACCCTTTCTGGATCAGGTG	TYITVSSEDDDEDGYNPYTL

Table 4.4 Primers used in this chapter.

Primer	Sequence	Description
NHS Ultramer	TGAAGGGCTGGCGGTTGGGGTTATTCGCAACGGCGACTGGCTGGAATTCT CTGGATCAGGTNHS(x10)TCTGGATCAGGTGAATTCAATTTTAATCAAAGTGG GAATATTGCTGATAGCTCATTGTC	
Fw Ultramer	TGAAGGGCTGGCGGTT	Used to amplify NHS Ultramer
Rv Ultramer	GACAATGAGCTATCAGCAATATTCCC	
Fw NHS BamHI	AACGGCGACTGGCTGGGATCCTCT	PCR NHS codons from pNIA2 with BamHI site
Rv NHS BamHI	TATTCCCACTTTGATTAAAATTGGATCCACC	

pelleted and resuspended in 100 µL STET (8% sucrose, 50 mM Tris-HCl pH 8, 50 mM EDTA, 5% Triton X-100) and an equal volume of acid washed glass beads (MilliporeSigma; G8772). Samples were vortex for 30 seconds, followed by 30 seconds on ice, ten times. An additional 100 µL of STET was added and samples were vortexed for an additional 30 seconds before incubation in a 95°C heat block for 5 minutes. Samples were cooled briefly on ice and centrifuged at max speed in a microcentrifuge for 10 minutes at 4°C. One hundred µL of supernatant was added to 200 µL of cold 95% ethanol, vortexed and incubated on ice for 2 minutes. DNA was pelleted by centrifuging samples at 4°C for 15 minutes at max speed. Supernatant was removed and the pellet was rinsed with 200 µL of cold 70% ethanol. Finally, samples were dried at room temperature, resuspended in 20 µL ddH$_2$O and transformed into competent DH5α *E. coli*.

For western blot analysis of protein extracts from yeast, the appropriate amount of liquid culture was used that corresponded to an OD$_{600}$ of 0.5 per mL. For example, if liquid cultures were measured to have an OD$_{600}$ of 1.0, then 500 µL would be used for lysis since diluting 500 µL in 1 mL would produce an OD$_{600}$ of 0.5. This process ensured a relatively equal number of cells were used for each protein preparation. Next, protein was isolated according to the method outlined by von der Haar, 2007 (von der Haar, 2007). In brief, cells were resuspended in 200 µL lysis buffer (0.1 M NaOH, 0.05 M EDTA, 2% SDS, 2% BME) and heated at 95°C for 10 minutes. Next, 5 µL of 4 M acetic acid was added to each sample and vortexed for 30 seconds before heating to 95°C for an additional 10 minutes. Cells were then pelleted, and supernatant was transferred to a fresh tube containing 50 µL loading buffer (0.25 M Tris HCl pH 6.8, 50% glycerol, 0.05% bromophenol blue).

4.2.4 Ultramer™ amplification and peptide selection with pNIA2

Random peptides were generated from oligonucleotide Ultramers™ (Integrated DNA Technologies) that contained 10 NHS codons, where N = any base, H = A, C, or T and S = G or C (Table 4.5). When synthesized, each individual oligonucleotide will have a region with 10 random codons. Ultramers were amplified by standard PCR using 2 µL NHS Ultramer (2 µM) and 2 µL forward and reverse Ultramer primers (5 µM each) in 50 µL total

Table 4.5 NHS codon table.

Amino acid	Codons (unavailable codons are crossed out)
Alanine	~~GCT~~ GCC ~~GCA~~ GCG
Arginine	~~CGT~~ ~~CGC~~ ~~CGA~~ ~~CGG~~ ~~AGA~~ ~~AGG~~
Asparagine	~~AAT~~ AAC
Aspartic acid	~~GAT~~ GAC
Cysteine	~~TGT~~ ~~TGC~~
Glutamic acid	~~GAA~~ GAG
Glutamine	~~CAA~~ CAG
Glycine	~~GGT~~ ~~GGC~~ ~~GGA~~ ~~GGG~~
Histidine	~~CAT~~ ~~CAC~~
Isoleucine	~~ATT~~ ATC ~~ATA~~
Leucine	~~CTT~~ CTC ~~CTA~~ CTG ~~TTA~~ TTG
Lysine	~~AAA~~ AAG
Methionine	ATG
Phenylalanine	~~TTT~~ TTC
Proline	~~CCT~~ CCC ~~CCA~~ CCG
Serine	~~TCT~~ TCC ~~TCA~~ TCG ~~AGC~~ ~~AGT~~
Threonine	~~ACT~~ ACC ~~ACA~~ ACG
Tryptophan	~~TGG~~
Tyrosine	~~TAT~~ TAC
Valine	~~GTT~~ GTC ~~GTA~~ GTG
Stop codon	~~TAA~~ TAG ~~TGA~~

reaction volume using Phusion DNA polymerase (New England Biolabs). Samples were denatured for 1 min at 98°C for 1 min, followed by 20 cycles of 98°C/60°C/72°C for 8 seconds each and a final extension stage of 2 minutes. Ten PCR reactions were done in parallel and amplification products were pooled. For expression of random peptides in pNIA2, 1 mL of L40 yeast (OD_{600} = 0.8) were transformed with 400 ng of EcoRI digested pNIA2 and 64 ng amplified NHS Ultramer. This corresponds to an 8:1 molar ratio of Ultramer to pNIA2. Yeast were plated directly onto SD media lacking leucine and histidine.

4.2.5   Human cell lines, cell culture, and transfections

Human HT-1080 cells (purchased from the American Type Culture Collection) were grown at 37°C with 5% CO_2 in Dulbecco's modified Eagle's medium (DMEM; Wisent) supplemented with 1% penicillin/streptomycin (Wisent) and 10% FBS (Wisent). Transfections of DNA into HT-1080 cells was done using X-tremeGENE HP (Roche) using a 2:1 ratio of X-tremeGENE HP per μg of DNA.

4.2.6   Immunofluorescent microscopy

HT-1080 cells were seeded onto glass cover slips prior to transfection. Following transfection, cells were fixed in 3.7% paraformaldehyde at room temperature for 20 minutes, permeabilized on ice using 0.2% Triton X-100, and blocked using 3% BSA in PBS. Samples were incubated with rabbit α-EGFP for 1 hour at room temperature, washed three times for 5 minutes in PBS, followed with 30 minutes at room temperature with secondary antibody (Alexa Fluor 488; Life Technologies), and finally another three 5 minute washes in PBS (For a list of primary antibodies used in this chapter see Table 4.6). Cover slips were then mounted on glass slides using ProLong™ Gold Antifade Mountant with DAPI (Thermofisher). Confocal images were taken with a Nikon Eclipse Ti2 under 60x magnification using the NIS Elements acquisition software.

4.2.7   Nuclear fractionation and quantification

HT-1080 cells were seeded onto 10 cm plates in triplicate and transfected the following day when cells reached approximately 80% confluency. Eighteen to 24 hours after

Table 4.6 Primary antibodies used in this chapter.

Reactivity (clone)	Purpose	Description	Supplier	Product number
LexA	Western	Rabbit polyclonal	MilliporeSigma	06-719
HA (3F10)	Western	Rat monoclonal	MilliporeSigma	11867423001
G6PD	Western	Rabbit polyclonal	MilliporeSigma	A9521
EGFP	IF, Western	Rabbit polyclonal	Takara (Living Colors)	632592
FLAG (M2)	Western	Mouse monoclonal	MilliporeSigma	F1804
FLAG (M2)	IP	Mouse monoclonal	MilliporeSigma	M8823
Lamin A/C	Western	Rabbit polyclonal	Proteintech	10298-1-AP
Tubulin (DM1A)	Western	Mouse monoclonal	MilliporeSigma	T9026

transfection cells were harvested for isolation of nuclear and cytoplasmic compartments using the REAP method (Suzuki et al., 2010). First, cells were pelleted at 10,000 rpm in a microcentrifuge for 30 seconds and resuspended in 900 µL lysis buffer (0.1% NP40 in PBS). From here, 300 µL was collected and labelled as whole cell extract (WCE). Samples were immediately centrifuged and 300 µL of supernatant was collected and labelled as cytoplasm. The remaining supernatant was aspirated, and pellets were resuspended in 1 mL lysis buffer. Samples were centrifuged once again, the supernatant was aspirated and the pellet was resuspended in 150 µL 1X LDS sample buffer (Thermofisher) and labelled as nuclear. Importantly, each resuspension step involved slowly pipetting up and down exactly ten times. This ensured that pellets were fully resuspended in a time efficient manner. Fifty µL of WCE and cytoplasm were mixed with 50 µL 2X LDS sample buffer. WCE and nuclear fractions were sonicated using the Bioruptor Pico (Diagenode) set to 4 cycles of 30 seconds on/30 seconds off at 4°C. Finally, all samples were boiled for 2 minutes and 12 µL nuclear, 18 µL cytoplasmic, and 18 µL WCE (corresponding to a roughly 2:1:1 ratio of extract) was used for loading samples on protein gels.

4.2.8 Co-immunoprecipitation and western blotting

Transfected HT-1080 cells from 10 cm plates were collected and lysed in 500 µL NP-40 lysis buffer (150 mM NaCl, 50 mM Tris-HCl pH 7.4, 1 mM EDTA, 1 mM EGTA, 0.1% NP-40, 10% glycerol) supplemented with protease inhibitor cocktail (MilliporeSigma, P8340) and sonicated as described above. Two percent of sample was kept for use as input control. Co-immunoprecipitation reactions were carried out at 4°C for 4 hours in lysis buffer using 20 µl of washed magnetic FLAG beads (MilliporeSigma, M8823). Samples were washed twice in 1 mL lysis buffer (no protease inhibitor) and boiled in 25 µL of 2X LDS sample buffer (Thermofisher, NP0007) supplemented with DTT (0.2 M final), for 5 minutes.

4.3 Results

4.3.1 Construction of a yeast-based protein nuclear import assay

Previously, the Mymryk laboratory modified and enhanced a yeast-based system for measuring nuclear import activity of proteins and/or specific regions within a protein of interest (Marshall et al., 2007; Rhee et al., 2000). This system, referred to as pNIA, utilizes plasmid-based expression of a recombinant protein consisting of the LexA DNA-binding domain (DBD) and maltose binding protein (MBP) from *E. coli*, and the *S. cerevisiae* Gal4 transcriptional activation domain (Gal4$_{AD}$), from N- to C-terminus, respectively (Figure 4.1A). Here, pNIA refers to both the plasmid vector and recombinant protein expressed from it. Importantly, this system is designed to work in the *S. cerevisiae* L40 yeast strain, which contains a genomic *HIS3* selectable marker and *lacZ* reporter that are both under the control of a LexA responsive promoter. Together, these features allow for the selection and quantification of nuclear import, respectively. With the addition of MBP, pNIA is greater (~80 kDa) than the passive diffusion limit of the NPC and unable to enter the nucleus; therefore, nuclear import and subsequent activation of expression of lacZ only results by fusing a nuclear targeting signal onto the C-terminus of pNIA.

The pNIA system was initially designed to test larger protein segments of known sequence for their ability to direct nuclear import; therefore, to make this system suitable for screening short random peptides several adjustments were necessary. The first adjustment involved identifying a putative location for expressing peptides. Using IUPred2A to predict IDPRs, it was determined that placing LexA in frame with the Gal4$_{AD}$ created a region of predicted disorder within the linker region connecting these domains (Figure 4.1B). The second adjustment was to relocate MBP to the C-terminus to preserve the overall size. To reflect these modifications, I have renamed this system pNIA2. Importantly, since the peptides to be expressed and tested are randomly generated, occurrences of stop codons and cleavage sites would result in expression of truncated motifs in the context of pNIA. Therefore, when expressed internally the presence of a stop codon or cleavage site should prevent pNIA2-mediated reporter gene expression due to the absence of the Gal4$_{AD}$.

Figure 4.1 Construction and testing of pNIA2.

A) Depiction of the recombinant protein, pNIA, that was previously developed by Marshall et al, 2007. Only in the nucleus can pNIA bind and activate reporter gene expression via the LexA DBD and a Gal4$_{AD}$. The addition of MBP increases the size of pNIA (~80 kDa) to prevent diffusion into the nucleus. **B)** pNIA was reorganized to express candidate test peptides internally and renamed pNIA2. The IUPred2A disorder prediction program shows a region of predicted disorder between the LexA DBD and Gal4$_{AD}$ where peptides could likely be expressed with minimal influence by the structures of the adjacent LexA DBD and Gal4$_{AD}$. **C)** Reconstruction of pNIA2 shows that expression from a low copy CEN plasmid and the addition of MBP significantly decreases background β-galactosidase activity compared to expression from a high copy 2μ plasmid. When expressed with the SV40 TAg cNLS, pNIA2 nuclear import activity increase over 100x compared to pNIA2 without an NLS. **D)** To assess potential positional effects of NLSs expressed in pNIA2, I tested the HAdV-C5 E1A cNLS and SV40 TAg cNLS, which are natively located C-terminally and internally, respectively. Both NLSs function in either position in pNIA2; however, they work best when expressed in a location similar to their natural position. **E and F)** Several NLSs (E) and non-NLSs (F) were expressed in pNIA2 to determine specificity for nuclear localization signals. NLSs tested included a bipartite cNLS (NP), two monopartite cNLS (SV40 TAg and HAdV-C5 E1A) and a non-cNLS (HIV-1 Tat), all of which produced significant nuclear import activity. Non-NLS motifs, including an SGSG linker, did not show significant nuclear import activity compared to pNIA2. Experiments were performed in triplicate and results were analyzed using a one-way ANOVA and Dunnett's multiple comparisons test where each NLS/peptide was compared to pNIA2.
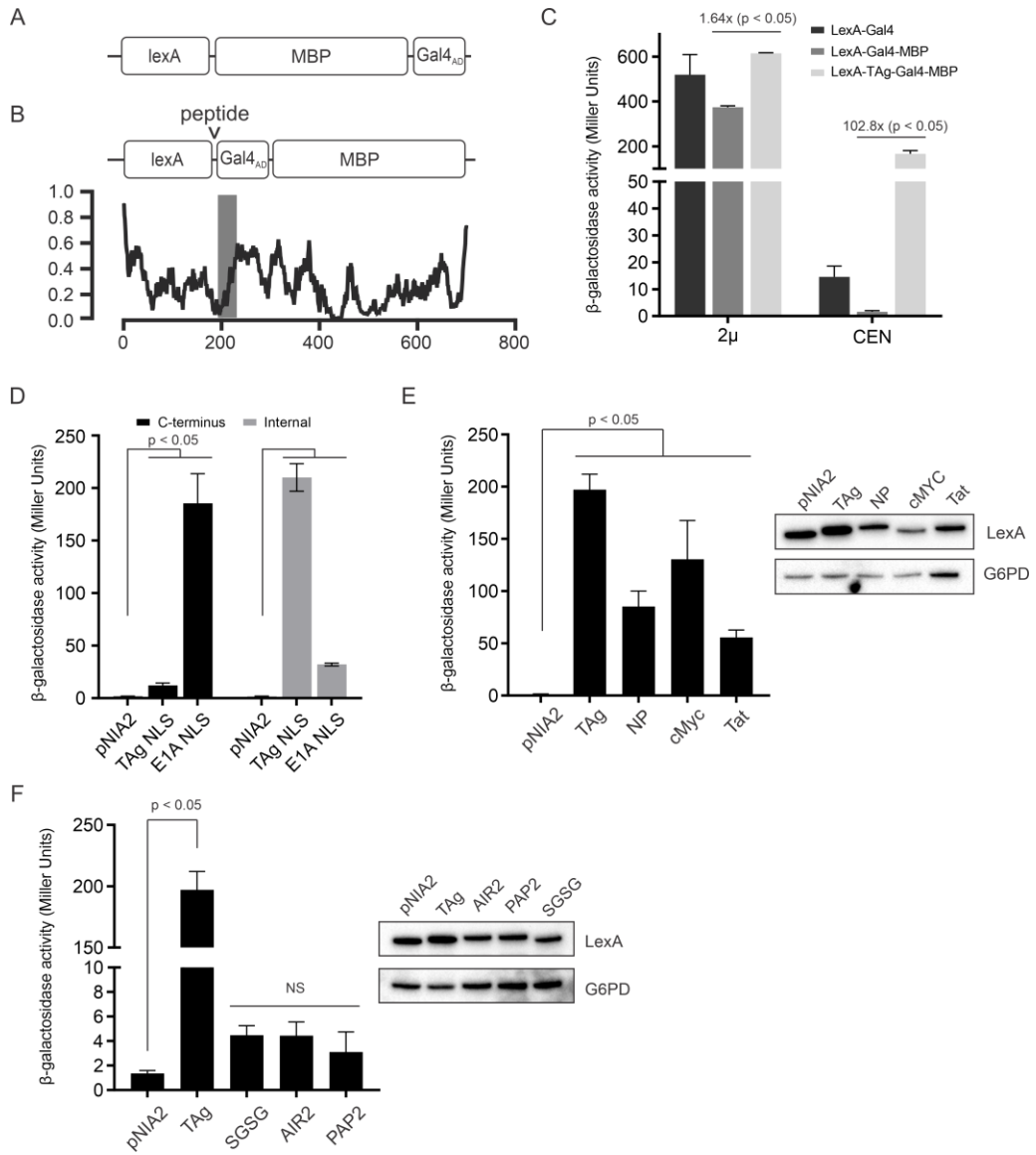
Initially, I built this system into a high-copy-2μ based shuttle plasmid. However, high-level expression of just the LexA-Gal4$_{AD}$ fusion from this backbone resulted in significant background activity. The addition of MBP minimally decreased this background activity and the addition of the strong and well characterized SV40 large TAg monopartite cNLS, PKKKRKV, resulted in only a modest, 1.6 fold increase in activity compared to the MBP construct (Figure 4.1C).

Expressing the pNIA2 cassette from a low-copy CEN shuttle plasmid showed much lower activity overall compared to the 2μ-based expression system. Importantly, the addition of MBP to LexA-Gal4$_{AD}$ reduced background activity to nearly undetectable levels and when expressed with the SV40 TAg cNLS, activity increased over 100-fold (Figure 4.1C). These results demonstrate that combining expression from a low copy plasmid and the addition of MBP to limit diffusion through the NPC creates a system with minimal background and high sensitivity necessary for identifying SLiMs that confer nuclear localization.

The region within a given protein sequence where an NLS may reside is not restricted to any particular area. Based on this, I next evaluated if positional effects could influence NLS activity when expressed in pNIA2. To evaluate this, I compared the SV40 TAg cNLS, representing an internally derived motif, and the well-studied human adenovirus 5 (HAdV-C5) E1A cNLS (LCSKRPRP), a motif located at the extreme C-terminus of E1A. Both cNLSs were expressed internally within pNIA2, as well as on the C-terminus, and respective nuclear import activity measured (Figure 4.1D). As expected, both cNLSs performed well when expressed in a position that reflected their native locations. When expressed in the alternative positions, nuclear import activity was significantly lower; however, cNLS activity for both motifs was still easily detectable compared to pNIA2 control vector lacking any NLS. These findings indicate that the linker between the LexA DBD and the Gal4$_{AD}$ of pNIA2 is a suitable position to test motifs from both internal and C-terminal derived motifs. Importantly, pNIA2 is sensitive enough to detect established motifs like the E1A cNLS even when localized within a suboptimal position.

In order to test the specificity of pNIA2 to detect nuclear import activity of peptides conferring nuclear import, I expressed a range of experimentally validated cNLSs within

pNIA2 and tested their respective nuclear import activity by β-galactosidase assay (Figure 4.1D). In addition to the TAg cNLS, I tested the bipartite NP (KRPAATKKAGQAKKKK) and monopartite cMYC (PAAKRVKLD) cNLS (Dang & Lee, 1988; Robbins et al., 1991). Each cNLS tested demonstrated significant nuclear import activity compared to pNIA2 control (Figure 4.1E). Interestingly, the HIV-1 Tat protein NLS (RKKRRQRRR), is a non-canonical NLS that binds Imp-β1 directly, bypassing Imp-α (Truant & Cullen, 1999). *S. cerevisiae* express several different Imp-β proteins that are conserved in humans. Since the Tat NLS can mediate nuclear import, this further reinforces the high degree of conservation of nuclear import across eukaryotes. These results demonstrate that different cNLSs function when expressed in pNIA2 and that this system is also sensitive enough to detect non-classical nuclear import pathways.

So far, only well characterized NLSs I expected to work had been tested and it remained unclear whether or not pNIA2 is specific to nuclear import. To test this, I expressed several other irrelevant, non-NLS SLiMs, as well as a comparably sized serine-glycine linker (SGSG) (Figure 4.1F). Yeast derived SLiMs, from proteins AIR2 and PAP2, that have no relation to nuclear import were also chosen from the ELM resource and expressed in pNIA2. Neither PAP2, AIR2, nor the SGSG linker showed significant nuclear import activity (Figure 4.1E and F). Lack of nuclear import activity from these motifs is not due to lack of protein expression, as each motif was comparably expressed. Taken together, these results demonstrate that the pNIA2 system appears selective for peptides with the ability to mediate nuclear import.

4.3.2 Selection of peptides conferring nuclear import using pNIA2

Having established pNIA2 can reproducibly detect nuclear import mediated by multiple different NLSs, I next extended this system to develop a selection for random peptides conferring nuclear import in the L40 yeast strain. I took advantage of yeast homologous recombination to recombine a duplex DNA fragment containing randomly synthesized codons into the linker region between the LexA DBD and Gal4$_{AD}$ (Figure 4.2A). Using standard PCR, DNA fragments were amplified from an oligonucleotide Ultramer containing 10 codons designated NHS. Additionally, I flanked both ends of the NHS codon

Figure 4.2 Outline of peptide selection with pNIA2.

A) NHS Ultramers have homology to the LexA DBD-Gal4$_{AD}$ linker region within pNIA2, allowing for recombination when co-transformed into L40 yeast. Since pNIA2 carries the *LEU2* marker, successful recombinants can be selected for on media lacking leucine. When expressed, the pNIA2 fusion protein will contain an in-frame fusion of 10 random amino acids flanked with an SGSG linker on both sides. **B)** The first step in selecting for peptides involves co-transforming L40 yeast with linearized pNIA2 and amplified NHS Ultramer. pNIA2 is digested with EcoRI, which cuts specifically between the LexA DBD and Gal4$_{AD}$. **C)** Transformed yeast are plated onto media without leucine and histidine and supplemented with 10 mM 3-AT to increase the stringency of *HIS3* selection. If a random peptide confers nuclear localization, pNIA2 will activate expression of *HIS3* allowing growth on selective media. **D)** Yeast which have been directly selected for on media lacking leucine and histidine undergo two sequential β-galactosidase assay screens for nuclear import activity. The first screen directly tests yeast which have undergone selection for nuclear import on media lacking leucine and histidine. After this screen plasmids are isolated, sequenced, and retransformed into L40 and grown under non-selective conditions (media contains histidine). These yeasts are used for a secondary nuclear import activity screen to identify statistically significant peptides. **E)** Background activity was sampled by randomly picking L40 yeast with recombined pNIA2 and NHS Ultramer which and have not undergone selection for nuclear import on media lacking histidine. pNIA2-NHS was compared to pNIA2 using the Welch's t test ($p < 0.05$). **F)** To determine if this approach selects for nuclear import, L40 yeast were transformed with either pNIA2, a 10:1 mixture of pNIA2 and pNIA2-TAg, or pNIA2-TAg. Yeast were grown directly in liquid culture and passaged twice, measuring nuclear import activity by β-galactosidase activity after each passage. After the second passage the pNIA2/pNIA2-TAg samples showed similar activity to pNIA2-TAg, indicating nuclear import is being selected for. Experiments were done in triplicate and analyzed using a one-way ANOVA and Tukey multiple comparison test ($p < 0.05$).
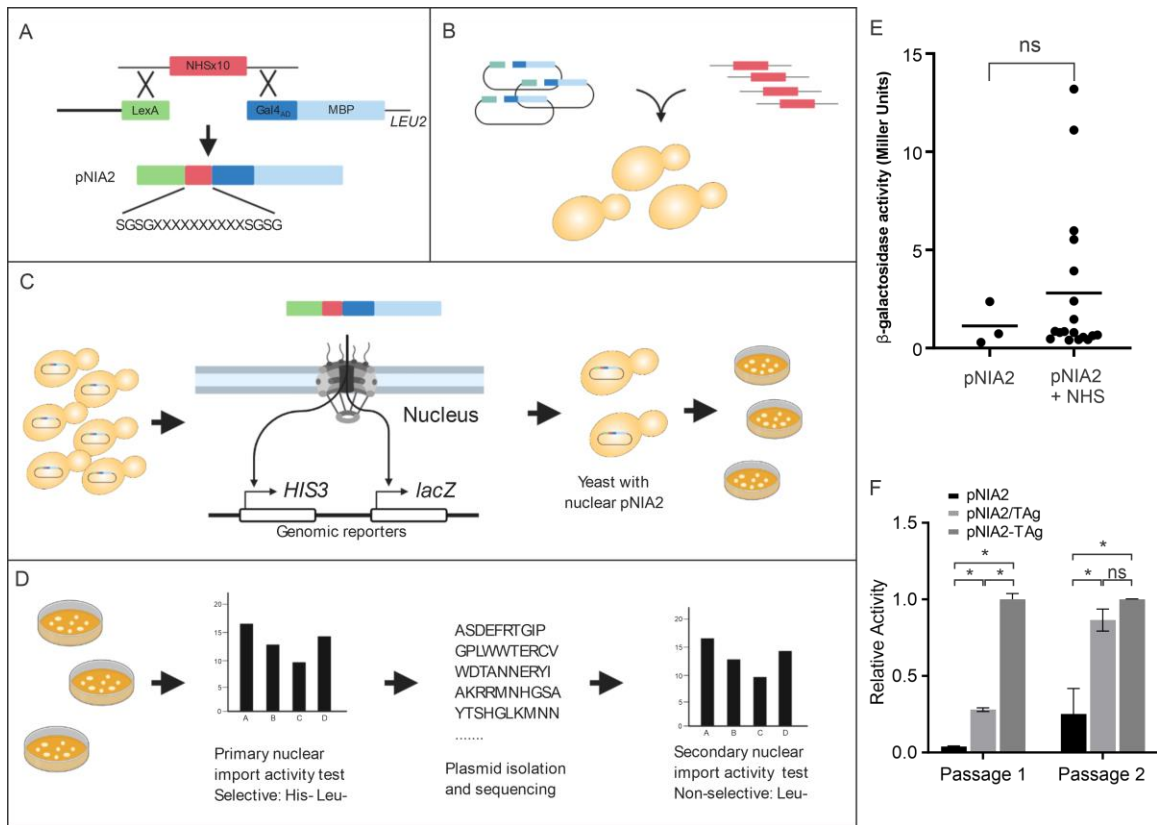
region with SGSG-linkers to avoid potential accessibility issues and 60 nucleotides of homologous flanking sequence corresponding to the LexA DBD and Gal4$_{AD}$ region of pNIA2. Since this system utilizes nuclear import, I also developed a strategy to minimize the occurrence of motifs representing cNLSs, as this pathway is the predominant protein nuclear import pathway and would likely be overrepresented during selection. To overcome this, I specifically used oligonucleotide templates with NHS codons (NHS Ultramer) where: N = any base, H = A, C, or T and S = G or C. This eliminates codons representing the amino acids arginine, cysteine, glycine and tryptophan while also limiting the occurrence of lysine codons (Table 4.5). This approach effectively eliminates the possibility of motifs to arise that represent a cNLS, while maintaining many of the codons for nearly all other amino acids. Another benefit of this strategy is that 2 of the 3 possible stop codons are also eliminated.

Transformation of yeast using a pool of amplified NHS Ultramer and linearized pNIA2 will trigger gap repair and the resulting recombination will generate yeast colonies expressing a pNIA2 protein with a unique 10 amino acid sequence insert between the LexA DBD and Gal4$_{AD}$ (Figure 4.2B and C). Successful recombinants were selected using media lacking leucine, as the pNIA2 plasmid contains a *LEU2* selectable marker. To select for successful recombinants with peptides mediating nuclear import, yeast were grown on media lacking both leucine and histidine (Figure 4.2C). Only if pNIA2 can enter the nucleus will it drive expression of the genomically integrated *HIS3* and *lacZ* reporter cassettes.

Yeast selected for based on pNIA2 nuclear import underwent a primary and secondary activity screen that measures nuclear import by β-galactosidase assay (Figure 4.2D). At this stage, pNIA2 plasmid could be easily isolated from corresponding yeast cultures and sequenced to determine the amino acid sequence encoded by the random NHS region. Since primary nuclear import activity is derived from a single colony, biological replicates cannot be tested appropriately. Additionally, the pNIA2 negative control cannot grow on media lacking histidine. Therefore, to resolve this issue the sequenced pNIA2 plasmids were transformed back into L40 yeast and grown on media lacking only leucine. This

additional step avoids any selective pressure that nuclear import may impose when grown on media lacking both leucine and histidine, reflecting nuclear import activity more accurately.

To detect if homologous recombination introduced any obvious background activity, L40 yeast were first transformed with amplified NHS Ultramer and linearized pNIA2 and grown under non-selective conditions (media containing histidine, but lacking leucine). Twenty colonies from the NHS-transformed group of yeast were chosen at random and compared to pNIA2 control for nuclear import activity using β-galactosidase assays (Figure 4.2E). Overall, no statistical differences were observed; however, several individual colonies chosen from the NHS transformed group displayed slightly higher activity that would still be considered background based on previous experiments using the SGSG linker alone.

To demonstrate if the conditions selective for nuclear import actually select for peptides capable of mediating nuclear import, L40 yeast were transformed with plasmid corresponding to either pNIA2, pNIA2-TAg cNLS and a 10:1 mixture of pNIA2 and pNIA2-TAg cNLS. Transformed yeast were grown directly in liquid culture and passaged twice, consecutively, from non-selective to selective media (without histidine) with nuclear import activity being measured after each passage (Figure 4.2F). Following the first passage, the mixture of pNIA2/TAg showed roughly 25% of the activity of pNIA2-TAg alone. However, after a second passage the pNIA2/TAg mixture had nearly equal activity to that of pNIA2-TAg, indicating that yeast expressing pNIA2-TAg cNLS are being selected for under these growth conditions. Taken together, these results demonstrate that the pNIA2 system produces minimal background when tested under non-selective conditions for nuclear import and that when placed under selective pressure pNIA2 containing inserts that confer nuclear import can be effectively selected for.

4.3.3   pNIA2 selects for novel peptide motifs conferring nuclear import

Using my outlined pNIA2 selection protocol, I next performed a small-scale selection to determine if novel peptides with the ability to mediate nuclear import can be selected for.

Transformed L40 yeast were plated onto leu⁻his⁻ media containing 10 mM 3-amino-1,2,4-triazole (3-AT) and primary nuclear import activity was determined for several dozen colonies, named NM-1 through 34 (Figure 4.3A). Interestingly, almost all colonies sampled showed a similar degree of nuclear import activity compared to pNIA2-TAg cNLS. Plasmids were isolated from their respective yeast cultures and sequenced so that each individual peptide sequence could be inferred (Table 4.7). As expected, based on my NHS Ultramer design, none of these peptides showed any resemblance to a cNLS. Finally, all plasmids were transformed back into L40 yeast for a secondary nuclear import test with appropriate controls and biological replicates under non-selective nuclear import conditions (Figure 4.3B). For greater stringency all motifs were compared to pNIA2 with an SGSG linker (pNIA2-SG), as this construct shows slightly higher background activity than pNIA2. Under non-selective conditions (leu⁻his⁺) most motifs did not exhibit statistically significant nuclear import activity; however, several motifs, NM-9, -15, -21, -27, -28, -30, -33 and -34, continued to demonstrate significantly greater nuclear import activity. Furthermore, nuclear import activity did not appear to correlate with protein expression, as many motifs showed similar expression levels, but did not mediate nuclear import (Figure 4.3C). For example, NM-12, -13, -9 and -34 all show similar protein expression levels; however, statistically significant nuclear import activity was only observed with NM-9 and -34.

4.3.4 Several peptides identified appear to function modularly

The motifs generated thus far have been selected for in the context of pNIA2. To extend my findings, I next wanted to determine if any of these putative nuclear import motifs are modular. Here, motifs deemed as modular will be able to function in the context of a different protein that shares no sequence similarity to pNIA2. To test this, I expressed each of these motifs fused to the C-terminus of the Gal4 DNA-binding domain (Gal4$_{DBD}$) in the galactose inducible W303-1A yeast strain. Growth in media supplemented with galactose will induce expression of endogenous Gal4, which in turn will drive expression of a plasmid based Gal4 responsive *lacZ* reporter (Figure 4.4A, left panel). Expressing the Gal4$_{DBD}$ with an NLS, in place of the activation domain, will lead to Gal4$_{DBD}$ nuclear accumulation and competition with endogenous Gal4 for DNA

Figure 4.3 Primary and secondary nuclear import activity screen.

A) Primary nuclear import activity screen: L40 yeast transformed with linearized pNIA2 and amplified NHS Ultramer were selected for on media lacking leucine and histidine. Colonies were picked and grown overnight in media lacking leucine and histidine to test nuclear import activity by β-galactosidase activity (n = 1). At this stage statistically significant changes cannot be determined, and peptide sequences are unknown. **B)** Plasmids were isolated from yeast cultures, sequenced, and retransformed into L40 yeast and grown under non-selective conditions in media lacking only leucine. β-galactosidase activity was determined (n = 3) for each sample and compared to pNIA2-SG (SGSG linker) using t tests and not correcting for multiple comparisons (p < 0.05). Peptides with statistically significant increases (p < 0.05) are indicated and labelled with their corresponding amino acid sequence. **C)** Protein lysates from each sample were analyzed by western blot using LexA DBD primary antibody and G6PD (glucose-6-phosphate dehydrogenase) as loading control. pNIA2 protein expression was variable but doesn't appear to qualitatively correlate with nuclear import activity.

A

B

NM-9 * MQADKVMEPT
NM-15 * TPLMTTDLTP
NM-21 * DPFSEYIPDA
NM-27 * EEVMAPADQP
NM-28 * LALVPNEADM
NM-30 * ASVAHAYESF
NM-33 * SDNTHKFPTH
NM-34 * NKVLMLAYSD

β-galactosidase activity (Miller Units)

C
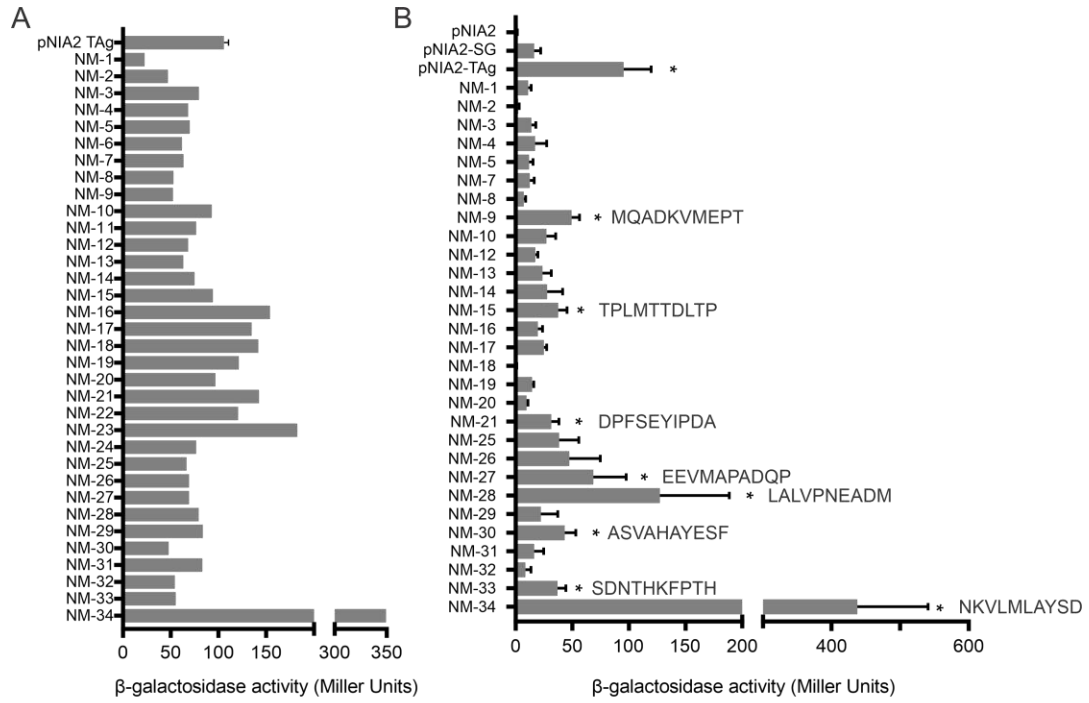
Table 4.7 Peptides identified in primary

nuclear import screen.

Peptide	Amino acid sequence
NM-1	HVDHFTTEDQ
NM-3	SVMSPDPLAE
NM-4	IPLPTMHHNV
NM-5	SKLATVEQDS
NM-7	ALPVSATKSK
NM-8	SMVHVLLLMV
NM-9	MQADKVMEPT
NM-10	SHPDLATTDS
NM-12	EIQTPTDSTS
NM-13	PAYTNQEMAK
NM-14	HLDDDTSQVL
NM-15	TPLMTTDLTP
NM-16	QLDLAQEYPS
NM-17	PLSELPSLEP
NM-19	DVDQVVVSEA
NM-20	AELHPLLHMD
NM-21	DPFSEYIPDA
NM-22	LSIPPDAKHA
NM-25	EQVMDKAQFS
NM-26	AHETATKDTA
NM-27	EEVMAPADQP
NM-28	LALVPNEADM
NM-29	LKVTEMTDLA
NM-30	ASVAHAYESF
NM-31	MVDIEAHPAS
NM-32	FDFYNAAEMK
NM-33	SDNTHKFPTH
NM-34	NKVLMLAYSD

binding (Figure 4.4A, right panel). Therefore, based on this system the reduction in β-galactosidase activity will be a functional measure of nuclear import activity.

As a positive control for this second confirmatory assay, I tested the SV40 large TAg NLS. As expected, expression of the Gal4$_{DBD}$-TAg cNLS fusion resulted in a significant reduction in β-galactosidase activity compared to just the Gal4$_{DBD}$ (Figure 4.4B). Interestingly, several of the motifs identified in the pNIA2 selection also repress β-galactosidase activity when fused to the Gal4$_{DBD}$. Thus, they appear to mediate nuclear import in this system, and function modularly in both systems (Figure 4.4B, indicated in green). In addition, this assay also identified motifs which mediate nuclear import that were not significant when tested in the secondary pNIA2 screen (NM-3, -16, -19 and -31). These observations possibly reflect the positional effects I observed with pNIA2 when expressing C-terminally derived NLSs internally. However, unlike expressing these motifs in pNIA2, expression on Gal4$_{DBD}$ produced inconsistent protein expression (Figure 4.4C). Nevertheless, peptides like NM-9 appear to function despite lower protein expression levels.

Based on my results from the pNIA2 selection and Gal4 competition assay, I chose NM-9 and NM-34 for further analysis. Motif NM-9 functioned in both experiments and although NM-34 nuclear import could not be reproduced when expressed on Gal4$_{DBD}$, I chose to pursue this motif due the significant activity observed with pNIA2. However, it also remains possible that some of these motifs themselves function as transactivation domains. Therefore, before pursuing these peptides further, I first wanted to rule out whether transactivation was a factor. To test this, I repeated the Gal4 competition assay using Gal4$_{DBD}$ NM-9 and -34 except I did not induce endogenous Gal4 expression (Figure 4.4D). In doing so, I could test these motifs for transactivation directly. In comparison to the HAdV-C5 E1A N-terminal region (1-82), a potent transactivator, as well as region 1-29 which lacks this ability, NM-9 and -34 do not appear to have any capacity to transactivate (Yousef et al., 2009). This is particularly important since performing screens with pNIA2 could alternatively select for motifs which function as unusually potent transactivators, rather than facilitate PPIs.

Figure 4.4 Testing NHS derived peptides in a Gal4 repression assay shows some function in a modular fashion.

A) Using the galactose inducible W303-1A yeast strain, endogenous Gal4 will bind and activate expression from a plasmid-based *lacZ* reporter when grown in media containing galactose (left panel). Expressing the Gal4$_{DBD}$, which lacks the transactivation domain, with peptides that facilitate nuclear localization will result in nuclear accumulation and competition with endogenous Gal4, resulting in a decrease of *lacZ* expression (right panel). **B and C)** Peptides identified in the pNIA2 selection were cloned onto the C-terminus of Gal4$_{DBD}$ and tested for their ability to compete with endogenous Gal4 by β-galactosidase assay (n = 3) (B). Peptides which functioned in both pNIA2 and on Gal4$_{DBD}$ are shaded in green. Protein expression was determined by western blot using αGal4 antibody (C). Statistically significant decreases (p < 0.05), compared to Gal4$_{DBD}$, were identified using t tests and not correcting for multiple comparisons. **D)** Peptides NM-9 and -34 were tested for their ability to transactivate (n = 3). Experimental conditions were the same as in (A), except yeast were grown in media without galactose; therefore, expression of *lacZ* depends on transactivation by the exogenously expressed Gal4$_{DBD}$ fusion. Amino acids 1-82 and 1-29 of HAdV-C5 E1A were used as a positive and negative control for transactivation, respectively. Neither NM-9 nor -34 were able to transactivate. Statistically significant increases (p < 0.05) were determined using a one-way ANOVA and Dunnett's multiple comparison test.
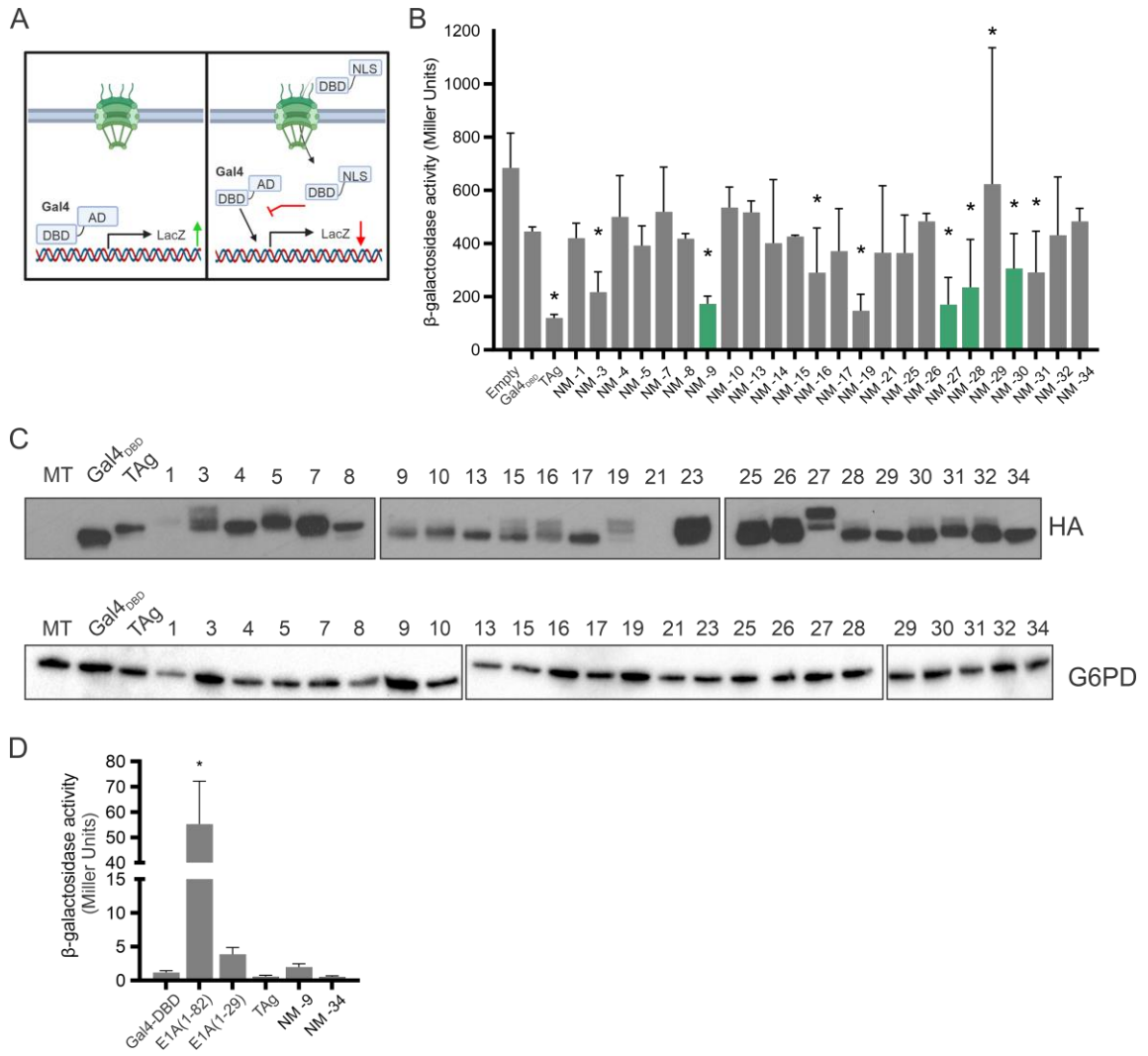
Overall, my data show that this screening approach is able to select for short, 10 amino acid peptide motifs which can putatively mediate nuclear import in yeast. Importantly, none of the identified peptides resemble a cNLS and several appear to be modular based on their ability to function as fusions to both pNIA2 and Gal4$_{DBD}$.

4.3.5 Peptides NM-9 and NM-34 can localize EGFP to the nucleus in mammalian cells

Nuclear import pathways among eukaryotes are highly conserved (Mason et al., 2009; O'Reilly et al., 2011; Pumroy & Cingolani, 2015). Based on this, I next determined if peptides NM-9 and -34 can mediate nuclear import in mammalian cells. I transfected HT-1080 cells with C-terminal EGFP fusions expressing NM-9 and NM-34 to determine subcellular localization via confocal microscopy (Figure 4.5A). Intriguingly, both NM-9 and NM-34 have a nucleocytoplasmic localization, compared to EGFP, which is almost exclusively cytoplasmic in HT-1080 cells. In comparison to EGFP-TAg cNLS, which is highly nuclear as expected, NM-9 and -34 appear to act as weak NLSs. Whether or not these are true, direct acting NLSs is unclear, as it's possible these peptides confer interactions with other nuclear proteins that confer piggyback transport into the nucleus.

To alternatively confirm the subcellular localization of NM-9 and -34 EGFP fusions, I biochemically isolated nuclear and cytoplasmic compartments of transfected HT-1080 cells using the well-established REAP method (Nabbi & Riabowol, 2015; Suzuki et al., 2010). Cells expressing either EGFP, EGFP TAg cNLS, HIV-1 Rev nuclear export signal (NES), NM-9 and NM-34 were analyzed by western blot at a 2:1:1 ratio of nuclear, cytoplasmic and WCE (Figure 4.5B). Standard cytoplasmic and nuclear compartment controls corresponding to tubulin and lamin were used, respectively. However, to better reflect my experimental approach, I also used EGFP TAg cNLS and EGFP Rev NES as additional nuclear and cytoplasmic compartment controls. As expected, EGFP TAg showed a strong nuclear localization, while EGFP and EGFP-Rev did not. Importantly, the absence of EGFP Rev NES, as well as tubulin, from the nucleus was an important indicator demonstrating leakage from cytoplasm to nucleus was not an issue during sample preparation. Comparison of nuclear and cytoplasmic compartments from

Figure 4.5 Peptides NM-9 and NM-34 can localize EGFP to the nucleus in human cells.

A) Human HT-1080 cells were transfected with either EGFP or EGFP C-terminal fusions with TAg NLS, NM-9 or NM-34. Cells were stained with DAPI to determine nuclear localization by confocal immunofluorescence microscopy. **B and C)** Nuclear (N), cytoplasmic (C) and whole cell extracts (W) from transfected HT-1080 cells were analyzed by western blot at a 2:1:1 ratio of N:C:W. SV40 TAg NLS and HIV-1 Rev NES were used as positive and negative controls for nuclear localization, respectively. Tubulin and lamin A/C western blots were used to test sample preparation efficiency. Additionally, 2% of whole cell extract (WCE) from each sample was tested simultaneously by western blot to compare protein expression. **D)** Western blots in (B) were quantified using ImageJ and nuclear localization was expressed as percentage of N and W. Statistically significant increases were identified using a one-way ANOVA and Dunnett's multiple comparison test. **E)** Nuclear extract from EGFP and EGFP NM-34 samples were analyzed by western blot at a 2, 4 and 6:1:1 ratio of N, C and W.
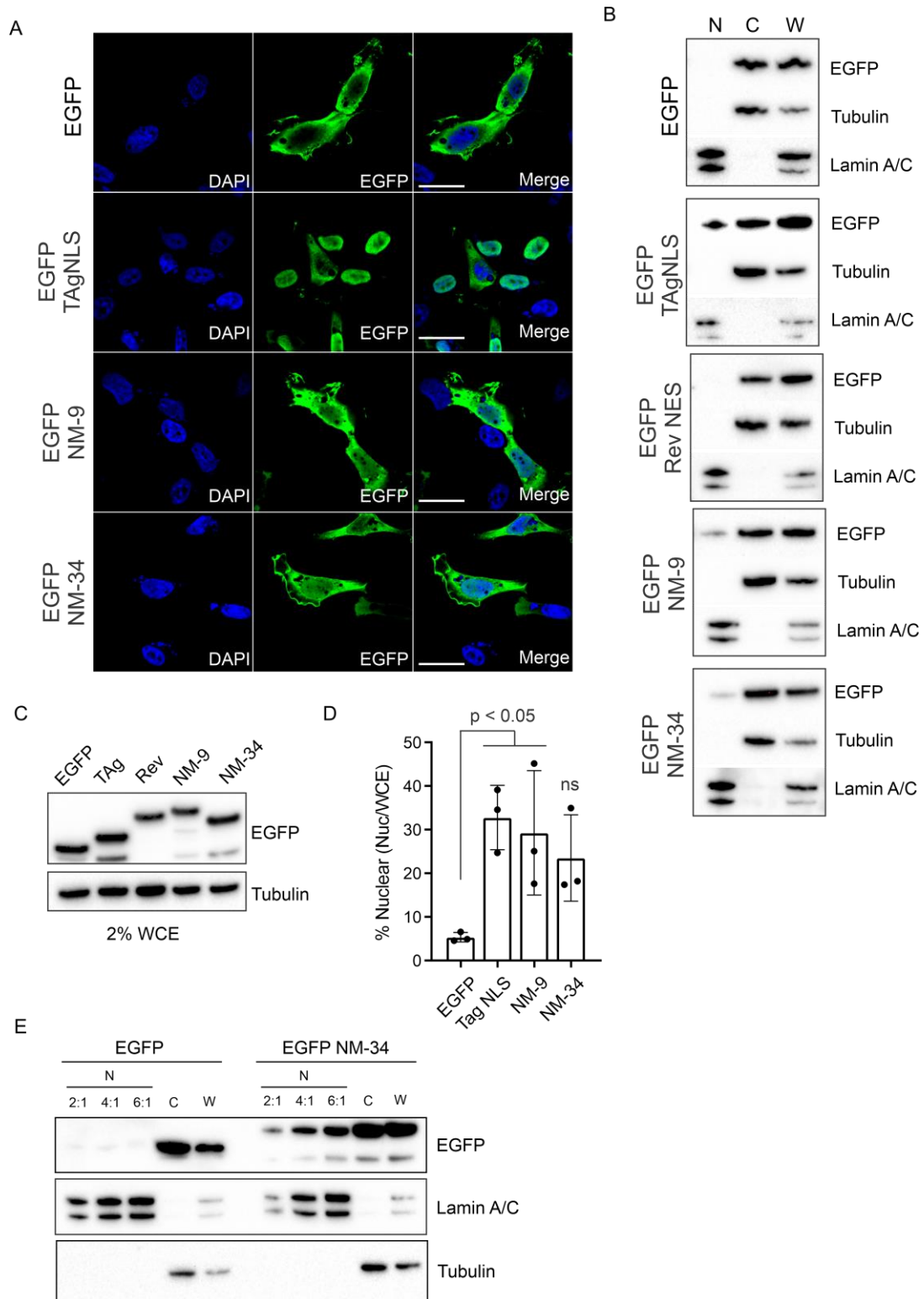
EGFP NM-9 and -34 samples confirmed their nuclear localization. Although nuclear localization was weak, these findings agree with my observations based on confocal microscopy and yeast-based assays. These results also do not appear to be an artifact of expression, as each construct was expressed at roughly equal levels (Figure 4.5C). Additionally, western blots were quantified from three biological replicates using ImageJ software. Comparing nuclear and WCE at a 2:1:1 ratio shows significantly greater nuclear localization for EGFP NM-9, but not NM-34, compared to EGFP (Figure 4.5D).

At a 2:1:1 ratio, nuclear EGFP NM-9 and -34 were only weakly detectable by western blot. To test this further, I loaded increasing amounts of nuclear extract from EGFP and EGFP NM-34 samples. At a 6:1:1 ratio EGFP NM-34, and likely NM-9 as well, could be easily detected within nuclear extracts (Figure 4.5E). Interestingly, EGFP was still not detectable within the nucleus at a 6:1:1 ratio despite roughly equal levels of nuclear lamin compared to EGFP NM-34. Interestingly, despite western blot quantification showing EGFP NM-34 was not significant, western blots of increasing amounts of nuclear extract clearly indicate EGFP NM-34 is more nuclear than EGFP.

4.3.6 Peptides NM-9 and NM-34 can associate with Imp-α

Evidence so far demonstrates that NM-9 and NM-34 are able to localize EGFP to the nucleus; however, it's unclear whether or not these peptides are promoting active nuclear import or simply binding a nuclear protein after diffusing into the nucleus. It's also possible these peptides are piggybacking into the nucleus on a protein which uses one of the many possible nuclear import pathways.

Since the classical nuclear import pathway mediated by Imp-α is well established, I first decided to investigate its possible role, as this would be a logical first step before exploring additional interactors. Humans express seven Imp-α isoforms that can be categorized into three subfamilies based on sequence similarity: α1 (Imp-α1, and -α8), α2 (Imp-α3 and -α4), and α3 (Imp-α5, -α6 and -α7) (Pumroy & Cingolani, 2015). To determine if NM-9 and NM-34 interact with any of these Imp-α isoforms, I tested a representative member from each subfamily for their ability to interact with EGFP NM-9 and -34 by co-

immunoprecipitation. HT-1080 cells were co-transfected with either FLAG-tagged Imp-α1, -α3 or -α5 and EGFP, EGFP TAg NLS, NM-9 or -34, and immunoprecipitated using anti-FLAG magnetic beads (Figure 4.6). These data show a strong interaction between the TAg cNLS and each of the Imp-α isoforms tested. To my surprise, NM-9 and -34 appear to interact with specific Imp-α isoforms. Neither peptide appear to interact with Imp-α3, although evidence of a weak interaction might be inferred; however, both EGFP NM-9 and -34 are able to bind Imp-α5. Furthermore, it appears that only EGFP NM-34 has specificity towards Imp-α1.

These results are especially interesting due to the fact that neither NM-9 nor NM-34 resemble a cNLS. Taken together, these results demonstrate that peptides identified in my yeast-based selection can localize EGFP to the nucleus in mammalian cells. Perhaps more interesting is their ability to associate with Imp-α. How exactly they interact is unknown; however, whether it is via a direct or indirect interaction, it's evident these peptides are mediating PPIs.
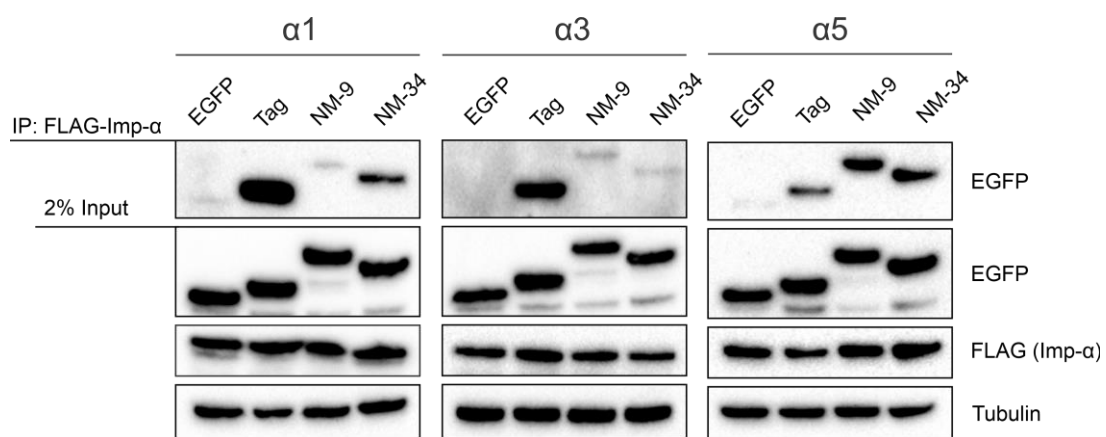
Figure 4.6 Peptides NM-9 and NM-34 interact with human importin alpha.

Human HT-1080 cells were co-transfected with FLAG-tagged Imp-α isoforms 1, 3 or 5, and either EGFP, EGFP NM-9 or EGFP NM-34. Samples were immunoprecipitated with anti-FLAG antibody and western blots were performed using anti-EGFP antibody.

4.4 Discussion

The protein interactome of living cells reflects a multitude of different PPIs. Characterization of *what* proteins interact with one another greatly exceeds our understanding of *how* these proteins actually interact. This makes our understanding of how PPIs are facilitated an important step in realizing the complexity of network regulation within living systems. One particular type of PPI module belongs to the class represented by SLiMs (Van Roey et al., 2014). These are short motifs, often less than 10 amino acids, that are integral to a variety of cellular processes. Roughly 100,000 SLiMs are predicted to be found within the human proteome; however, only a few thousand have been documented (Kumar et al., 2020; Tompa et al., 2014).

Here, I have developed a genetic, yeast-based selection system, that exploits protein nuclear import to identify novel SLiMs. Additionally, I performed a small-scale selection and demonstrated this system can successfully identify SLiMs in yeast that can localize EGFP to the nucleus in human cells. This system was based off of a previous iteration, pNIA, which was not suited to test random peptides (Marshall et al., 2007). To adapt this system for expressing random peptides, I reorganized the protein domains to express peptides internally, rather than on the C-terminus, and renamed it pNIA2. The major advantage with this approach was that if an insert happened to encode a cleavage motif, or contain a stop codon, the Gal4$_{AD}$ would no longer be tethered to the promoters of the reporter construct and would not be recovered by the selection. This modification, as one would theoretically predict, should result in reduced false positives identifications.

A variety of different NLSs were tested and shown to work in pNIA2. This also included many cNLSs as well as the HIV-1 Tat NLS, which is known to bind Imp-β1 directly, indicating this system is sensitive enough to detect non-cNLSs. Since *S. cerevisiae* express a number of different Imp-β proteins that are functionally conserved with their human homologs, it's important that peptides targeting non-classical pathways can still be isolated. Another advantage of this system was the extremely low background activity and wide range of activity. When expressed in L40 yeast without an NLS, background activity was almost undetectable, whereas addition of an NLS could increase activity over 100-fold.

Intriguingly, two peptides, NM-9 and -34, identified in yeast by a small-scale selection were demonstrated to function in human HT-1080 cells, by localizing EGFP to the nucleus using both immunofluorescence microscopy and biochemical fractionation of subcellular compartments. Additionally, co-immunoprecipitation experiments showed that both of these peptides were able to interact with human Imp-α. In particular, both peptides were able to interact with Imp-α5, which has the greatest sequence similarity to Srp1, the only Imp-α in *S. cerevisiae* (Pumroy & Cingolani, 2015). My focus was to develop an assay that could provide candidate SLiMs for further investigation. For this reason, I did not investigate these PPIs further and it is unknown if these peptides directly bind Imp-α. Nevertheless, these peptides appear to be facilitating PPIs involved in nuclear localization.

Several other advantages of this approach are worth discussing. First, because this system works in yeast, it is inexpensive and highly accessible. Only oligonucleotides are required, with no need for specialized equipment. This is a far simpler approach compared to protein or peptide arrays. Secondly, this is an *in vivo* system, albeit in yeast, that is also functional since it exploits protein nuclear import. This is a major advantage over current SLiM discovery methods, like phage display or peptide/protein arrays, which are exclusively *in vitro*. Lastly, because this is a genetic system, no prior knowledge of the peptides being tested is required. Plasmids can be easily isolated, like with phage display, from yeast and sequenced to determine peptides which have been selected for. Pools of selected recombinants could be sequenced in bulk using next-generation approaches, and in combination with enrichment type analyses could identify thousands of high confidence PPI motifs.

While this system is designed to select for peptides mediating protein nuclear import it remains possible that alternative scenarios exist where selection may not reflect nuclear import. For example, since LexA functions as a homodimer it's possible that peptides which enhance dimerization will be selected (Mohana-Borges et al., 2000). Alternatively, peptides which promote nuclear import but negatively influence dimerization could be missed due to poor reporter gene expression. Another scenario which could lead to the false discovery of peptides is through selection of peptides with the ability to transactivate. Transactivation domains are often enriched in acidic and aromatic/hydrophobic amino

acids, leaving open the possibility these amino acids could be selected since they are present within the NHS library design (Regier et al., 1993; Sigler, 1988; Staller et al., 2018).

An important limitation to this approach is that it selects for nuclear localization. This means that PPIs will be limited to nuclear proteins, in theory. However, despite this limitation it's important to consider that nearly 1500 proteins in *S. cerevisiae* localize to the nucleus and it's estimated that approximately 87% of protein domains in yeast are also found within humans (Huh et al., 2003; Peterson et al., 2013). This number of potential interactors is significant and greatly exceeds similar approaches. Phage display and protein/peptide arrays typically utilize a one-vs-many approach, where a few peptides are tested individually against many baits, or vice versa. A recently developed technique, called PRISMA, has significantly expanded the number of peptides and baits which can be tested (many-vs-many); however, experimental design still requires prior knowledge of the peptides being tested (Dittmar et al., 2019).

Given the large discrepancy between the number of predicted and documented SLiMs within the human proteome, more high-throughput techniques are required to fill this gap. The system outlined here combines many of the advantages of other approaches. For example, it has the genetic tractability of phage display, allowing thousands of peptides to be screened and sequenced. Like PRISMA, it is high throughput, except significantly more peptides can be tested with my approach. Although I tested randomly generated peptides, my approach is highly complementary to other recently used methodologies. For example, experiments expressing the human disordered proteome as overlapping short peptides using phage display could be easily replicated with this system (Davey et al., 2017). Furthermore, this approach can lend itself to fields such as synthetic biology. Using randomly generated peptides, as I did here, completely unique motifs can be easily identified.

Overall, the system I present here is an additional tool that can be added to the molecular biologist's toolbox. Depending on one's goals, hundreds to thousands of random peptides can be screened and selected for based on nuclear localization in yeast. Importantly, this

can represent not only active nuclear import but also simply binding a nuclear protein and piggybacking into the nucleus. Since this system utilizes a genetic selection, potential candidate peptides can be rapidly identified for further investigation, as opposed to screening only approaches. Although this approach is purposely biased towards nuclear localization, of the tens of thousands of SLiMs yet to be documented, many will undoubtedly have roles within the nucleus.

Chapter 5

5 General Discussion

5.1 Thesis Summary

The work described within my thesis explores protein nuclear import from several different perspectives. In Chapter 2 of this thesis, I demonstrated that targeting the classical nuclear import pathway, mediated by Imp-α and -β1, using ivermectin, inhibits replication of HAdV-C5 and -B3. To explore the effects of ivermectin in greater detail I used HAdV-C5 as a model. Treating infected A549 cells *in vitro* with ivermectin reduced viral early gene transcription (transcripts for E1A, E1B, E3 and E4) as well as protein expression of several early proteins (E1A and DBP) and late proteins (capsid proteins). Additionally, infected cells treated with ivermectin no longer show the characteristic nuclear localization of E1A. Interestingly immunoprecipitation experiments of Qip1 (Imp-α3) show that ivermectin specifically targets the Imp-α-NLS interaction. Based on these findings I hypothesized that the effect of ivermectin is likely a result of E1A's inability to function within the nucleus.

In Chapter 3, I focused specifically on the classical nuclear import pathway in a cellular context; however, due the intimate relationship between virus and host, these concepts could also be applied to viruses. Chapter 3 is an extension of a previous analysis done in yeast, which showed many nuclear proteins do not have a predicted cNLS, and many proteins that bind Imp-α also do not have a cNLS (Lange et al., 2007). In my work, I extended these findings to the human nuclear proteome for the first time. Using the latest publicly available databases and datasets, I showed that at least 50% of nuclear proteins in human do not have a predicted cNLS. Looking specifically at cargo for each Imp-α isoform I also showed that many cargos also do not have a predicted cNLS. Thus, my results using human proteins are remarkably similar to those found in the simple eukaryote *S. cerevisiae*. To complement these findings, I reanalyzed raw mass spectrometry data from PPI studies with several Imp-α isoforms, which generated very similar conclusions. Within this group of nuclear proteins without a cNLS, I specifically identified the Mediator protein complex and proposed that it is highly likely to use a piggyback mechanism to access the nucleus.

In Chapter 4, my final data chapter, I used the connections between nuclear import and SLiM-mediated PPI interactions via a yeast-based genetic selection assay to identify novel SLiMs that also facilitate nuclear localization. Using this selection assay, I identified several peptides that do not bear any resemblance to a cNLS. Further analysis with two of these peptides showed they are able to localize EGFP to the nucleus in the human HT-1080 cell line. Furthermore, I demonstrated through co-immunoprecipitation experiments that both of these peptides are able to interact with Imp-α.

Overall, my research showed that targeting the nuclear import pathway, specifically the Imp-α-NLS interaction, is an effective target to inhibit viral replication of HAdV. Additionally, I highlighted that the classical nuclear import pathway needs deeper characterization, and this highlighted potential widespread use of piggybacking mechanisms, as well as suggested that many potentially novel NLSs exist.

5.2   Nuclear import is an effective target for virus-host SLiM-mediated interactions

Viruses are obligate intracellular parasites and therefore have an intimate relationship with their host. Classically, antiviral drugs have been developed against virus-specific factors like virally-encoded enzymes, such as polymerases, reverse transcriptases and proteases, for example (De Clercq, 2009; De Clercq & Li, 2016). Although this approach has been successful in many instances, drugs developed against virally-encoded factors can be rendered obsolete due to the rapid emergence of resistance (Van Poelvoorde et al., 2020). Alternatively, anti-viral drugs can be developed to target host factors, such as those that can modulate the immune system and activate innate immune responses (Kaufmann et al., 2018).

The inherent host dependency which viruses exhibit makes virus-host PPIs attractive targets for treating viral infections. During HAdV infection, cellular PPI networks become entangled with virus-host protein interactions (King, Zhang, et al., 2018). This results in a completely unique intracellular environment, potentially exposing new vulnerabilities that wouldn't otherwise exist (Mast et al., 2020). One cellular pathway which many viruses, both DNA and RNA, exploit is the protein nuclear import pathway (Tessier et al., 2019;

Yarbrough et al., 2014). In Chapter 2 of this thesis, I demonstrated that the classical protein nuclear import inhibitor, ivermectin, can inhibit replication of HAdV-C5 and -B3. These findings add to increasing body of evidence suggesting ivermectin as a potential antiviral (Fraser et al., 2014; Varghese et al., 2016; Wagstaff et al., 2012). Though I did not test the effectiveness of ivermectin on additional HAdV species, HAdV-C5 and -B3 are both clinically relevant viruses (Lion, 2014).

Importantly, these results only point towards a potential antiviral function and do not conclusively prove that ivermectin will be clinically useful. Importantly, ivermectin is widely used as an antiparasitic and has a well-established safety profile, representing a drug that could be theoretically repurposed for the treatment of additional diseases with minimal effort and cost (Laing et al., 2017). However, whether or not an effective concentration can be achieved to treat viral infections is unknown. If ivermectin concentrations sufficient to function as an anti-viral cannot be achieved clinically, it remains possible that ivermectin can work synergistically with other antivirals. For example, the relocalization of cellular proteins to the nucleus by E1A enhances viral replication (King, Cohen, et al., 2016). Through ivermectin-mediated inhibition of E1A, nuclear import of interacting cellular proteins will also be inhibited. In doing so, based on the concept of synthetic lethality, new vulnerabilities are likely to be exposed (Mast et al., 2020). Following this line of reasoning, future experiments should involve testing ivermectin in combination with other antivirals, including those that target other host processes.

Additionally, I was able to show for the first time that ivermectin targets the Imp-α-NLS interaction as a mechanism to inhibit import. Previously, based on *in vitro* assays performed by another group, the function of ivermectin was assumed to target the Imp-α/β1 interaction (Wagstaff et al., 2012). However, my results using co-immunoprecipitation experiments show that the Imp-α-NLS interaction was lost, while the Imp-β1 interaction was maintained. Exactly why these results differ from previous studies is unclear; however, recently these findings have been supported based on structural modelling of Imp-α and ivermectin, which showed ivermectin binding maps to the NLS binding groove (P. S. Sen Gupta et al., 2020). Further experiments are still required to fully understand ivermectin's mechanism of action. In my work, I only tested the interactions

between SV40 TAg cNLS and HAdV-C5 E1A with Qip1. As a class of SLiM, cNLSs bind Imp-α in a similar fashion; however, residues outside the core motif can modulate SLiM-mediated interactions. As a result, ivermectin may not influence binding of all Imp-α-NLS interactions equally. Since I tested full-length E1A and not just it's cNLS, these results also suggest ivermectin can inhibit bipartite cNLS binding. E1A's cNLS has classically been referred to as a monopartite cNLS; however, more recently it was shown to in fact be a bipartite cNLS. Additionally, I did not test cNLSs that specifically bind the minor groove of Imp-α. Therefore, additional mutational studies, as well as interactions with additional Imp-α family members using a variety of cNLSs will help to better understand ivermectin's mode of action.

Overall, the data presented in Chapter 2 demonstrates that virus-host SLiM-mediated interactions can be targeted to interfere with virus infection. Many viral proteins from a diverse range of viruses rely on SLiM-mediated PPIs. Targeting host proteins, like targeting Imp-α versus targeting a viral NLS directly, could be an effective strategy for finding new antivirals.

5.3   Identifying piggyback mechanism and novel NLSs

The classical nuclear import pathway is assumed to be responsible for handling the majority of protein nuclear import. Interestingly though, it was previously demonstrated in *S. cerevisiae* that up to 50% of Imp-α cargos do not have a predictable cNLS (Lange et al., 2007). The cNLS SLiM has been extensively studied and this revealed well defined rules for what may constitute a cNLS, making those observations particularly interesting (Kosugi et al., 2009). Since these findings had not been extended to the human Imp-α isoforms, this prompted me to address these questions with currently available datasets and databases with human Imp-α PPI data.

Using a computational approach, I looked at nuclear proteins from the HPA, as well Imp-α interactors from various databases and datasets. Rather than using NLS prediction programs, I specifically chose to look for regular expressions corresponding to cNLSs. This approach was inherently over predictive and undoubtedly identified many false positives, as many hits would likely be within structured and/or inaccessible regions. This means that

proteins without a hit, truly represent those that do not have a motif that fulfills our current definition of a cNLS. Using this approach, I estimated that roughly 50% of nuclear proteins in human do not have an amino acid sequence resembling a cNLS.

Interestingly, when specifically looking at Imp-α interactors from the IntAct database, I found that approximately 50–80% also do not have a predicted cNLS. I was also able to confirm these results using an independent dataset of Imp-α interactors that involved reanalysis of raw mass spectrometry files. When specifically looking at Imp-α interactors without a cNLS, I identified several protein complexes which may use piggybacking as a strategy to gain access to the nucleus. Intriguingly, RNAPII and TFIID already had existing evidence in support of piggybacking-based mechanisms to access the nucleus; however, a third group of proteins belonging to the Mediator complex did not (Di Croce, 2011; Trowitzsch et al., 2015). Based on these findings, I specifically looked at each subunit of Mediator to try and determine which protein subunits have a predicted cNLS and found that over half of Mediator subunits do not have a predictable cNLS, yet interact with Imp-α. Since other protein complexes, also involved in RNAPII transcription initiation, appear to use piggybacking, I proposed that nuclear import of Mediator may be mechanistically similarly.

Conceptually, the concept of piggybacking into the nucleus is highly advantageous. In theory, piggybacking can effectively increase the "mileage" of how many proteins a single Imp-α, or nuclear transporter, can import. Since nuclear import is an energy dependent process, one nuclear import cycle can effectively transport multiple proteins. Another important factor to consider is the effect of size on the rate of nuclear import. As protein size increases, the rate of transport through the NPC decreases (Lolodi et al., 2016). This introduces a potential dilemma for multi-subunit protein complexes like Mediator, which in this case is composed of proteins that range in size from less than 20 kDa to over 200 kDa. Having subunits import into the nucleus at different rates could have highly detrimental effects on transcription initiation. A scenario where subunits are preassembled in the cytoplasm, like RNAPII and TFIID, and piggyback into the nucleus could provide a uniform rate of import and ensure correct stoichiometry of the subunits in this large complex.

Overall, the results presented in Chapter 3 demonstrate a significant number of proteins associate with Imp-α, but do not have a cNLS. This highlights a need for identifying piggybacking proteins, as well as novel NLSs in future studies of nuclear trafficking.

5.4 Exploiting protein nuclear import to discover novel PPIs

As previously mentioned, many nuclear proteins in yeast, including those that interact with Imp-α, do not have a predictable cNLS (Lange et al., 2007). These findings suggest additional mechanisms by which proteins could be interacting with the nuclear import machinery. This could be through direct interactions with Imp-α, or possibly piggybacking on Imp-α cargos. Regardless of the scenario, both situations will inevitably involve PPIs. However, how these PPIs might be mediated and how they can be discovered falls under a much broader area of investigation, which is developing high throughput approaches to discover SLiM-mediated interactions.

In order to address these questions experimentally, I identified a convenient scenario that used a single approach to answer two not so disparate questions: 1) are there SLiMs other than cNLSs that can mediate nuclear localization and 2) can I develop a high throughput approach to identify these novel SLiMs. The approach that I developed involved a yeast-based (*S. cerevisiae*) genetic selection assay that uses a recombinant protein, pNIA2, that is too large to diffuse into the nucleus. Only upon addition of a functional NLS to the chimeric protein can pNIA2 enter the nucleus. Using this approach with randomly generated peptides, I identified several functional peptides that bear no resemblance to cNLSs. Interestingly, two of these peptides were able to localize EGFP to the nucleus and interact with Imp-α in a human cell line, showing the high degree of conservation between nuclear import across eukaryotes.

Experimental approaches that can be used to identify novel SLiMs follow a "one-vs-many" or a "few-vs-many" approach (Blikstad & Ivarsson, 2015; K. Meyer & Selbach, 2020). Current technologies, whether it be phage display or peptide/protein arrays, typically classify peptide-domain interactions as bait and prey (or vice versa) and are thus limited to examining only a few baits or preys at a time. Of course, these approaches can be scaled

up to increase throughput; however, this comes with a proportional increase in labour. The approach which I have outlined within Chapter 4 can effectively screen thousands of peptides, via selection, in a single experiment that can be scaled with a nearly negligible increase in labour. These advantages make this approach highly economical, as well as equitable, since it doesn't require specialized technology. Additionally, since this approach is functional, it uses the yeast nuclear proteome as "prey", making this a "many-vs-many" approach.

However, like Y2H screens, this approach generated a significant number of false positives (Blikstad & Ivarsson, 2015). During primary β-galactosidase assays, nuclear import activity is noticeably higher than in the secondary screen under non-selective conditions. Since nuclear import activity is most reflective under non-selective conditions (not selecting for nuclear import), this presents a work-flow bottleneck. To address this, further optimizations will be required. These could include replica plating yeast from media selective for nuclear import onto non-selective media before screening for nuclear import activity, or serially passaging selected yeast colonies in non-selective media. However, despite the limitations with Y2H screens, this approach has been invaluable for performing high-throughput proteome-wide screens, especially when combined with next-generation sequencing technology (Suter et al., 2015). These advantages are also applicable to the approach I have outlined in Chapter 4. Large-scale yeast transformations and selection, followed by bulk PCR amplification of the NHS region within pNIA2, would allow for high throughput identification of peptides.

My primary goal was to perform a small scale, proof-of-principle selection to determine if this approach was viable. For this reason, I did not perform detailed follow up experiments with the peptides identified in this study. Since they were generated at random, they bear little resemblance to human cellular proteins. Before pursuing more detailed experiments, mutational analysis should be performed to identify the amino acids necessary for function. With this information, it will be easier to identify cellular proteins that share the same core amino acids. Furthermore, *in vitro* binding assays using purified protein could be carried out to determine if these peptides interact directly with Imp-α. If this was the case, these peptides could be used as tools to explore Imp-α-NLS binding constraints. For example,

neither peptides bind Imp-α1, but will both bind Imp-α5; furthermore, only NM-34 can bind Imp-α3. Exploring why this is the case could potentially lead to a deeper understanding of Imp-α-NLS recognition. Conversely, these peptides may not bind Imp-α directly. To identify a list of potential interactors, affinity purification-mass spectrometry would likely prove useful in this scenario.

Since the approach I took used random peptides, it's also possible these motifs and others may have no similarity to those found in cellular proteins. If this is the case and these peptides are functional, they may be useful in unrelated areas of research, such as synthetic biology. For instance, peptide NM-34 shows potent nuclear import activity in yeast, an organism that has been at the forefront of synthetic biology (Z. Liu et al., 2019). Such a peptide could be used to direct recombinant proteins to the nucleus to achieve any number of desired functions.

Overall, whether peptides isolated using this approach directly bind importins is irrelevant. All that is needed is a PPI that localizes the recombinant test protein to the nucleus. From here, a list of peptides to pursue further can be rapidly generated. Importantly, this approach is another strategy that can be used to identify novel SLiMs.

5.5 Concluding remarks

The work presented here highlights various aspects of protein nuclear import. Together, my work aimed to address several outstanding questions within the field. I demonstrated that inhibition of the classical nuclear import pathway, via ivermectin, can inhibit HAdV replication. Interest in ivermectin as an antiviral has been steadily increasing; however, the mechanism behind its ability to inhibit nuclear import has been unclear. With respect to HAdV-C5 E1A and TAg cNLS, I demonstrated for the first time that ivermectin can disrupt the Imp-α-NLS interaction. Ivermectin was first proposed to inhibit the Imp-α/β1 interaction; therefore, evidence that it can inhibit Imp-α/cNLS binding could be an important factor when considering its use. As an antiviral, this drug is inhibiting the interface between a virus and host PPI. If ivermectin blocked the Imp-α/β1 interaction, one could assume that any viral NLS that binds Imp-α could be effectively inhibited. However, if ivermectin is in fact inhibiting the Imp-α/cNLS interaction this could exhibit some level

of selectivity that might have important implications. Specifically, determining if different viral cNLS/Imp-α interactions are affected differently by ivermectin will be an important step in determining its utility as a general antiviral.

Based on observations in yeast, I showed for the first time in humans that many Imp-α cargos do not have a cNLS. These observations focus attention on the idea of piggybacking and potentially new NLSs. Piggybacking into the nucleus would be a highly efficient process; however, studying this process on a large scale would be difficult. One approach to identify particular protein complexes was outlined in Chapter 3, which identified Mediator as a candidate for piggybacking. Identifying and studying these potential examples of piggybacking can provide deeper insight into basic, overlooked, cellular processes. These observations also highlighted the idea of potentially novel NLSs that can bind Imp-α. To address this, I developed a yeast-based selection to identify novel NLSs/SLiMs. This novel selection is significant in that it allows high throughput identification of short peptides (SLiMs) in a "many-vs-many" fashion, an area of SLiM discovery devoid of such approaches. Using this genetic selection and increasing its scale, one could identify a large number of motifs that facilitate nuclear localization through novel, uncharacterized interactions. Given the immense disparity of annotated and predicted SLiMs in the human proteome, such an approach, albeit with its own limitations, can be used to help close this knowledge gap.

References

Aebersold, R., Agar, J. N., Amster, I. J., Baker, M. S., Bertozzi, C. R., Boja, E. S., Costello, C. E., Cravatt, B. F., Fenselau, C., Garcia, B. A., Ge, Y., Gunawardena, J., Hendrickson, R. C., Hergenrother, P. J., Huber, C. G., Ivanov, A. R., Jensen, O. N., Jewett, M. C., Kelleher, N. L., … Zhang, B. (2018). How many human proteoforms are there? *Nature Chemical Biology*, *14*(3), 206–214. https://doi.org/10.1038/nchembio.2576

Aitchison, J. D., Blobel, G., & Rout, M. P. (1996). Kap104p: a karyopherin involved in the nuclear transport of messenger RNA binding proteins. *Science (New York, N.Y.)*, *274*(5287), 624–627. https://doi.org/10.1126/science.274.5287.624

Aksu, M., Pleiner, T., Karaca, S., Kappert, C., Dehne, H.-J., Seibel, K., Urlaub, H., Bohnsack, M. T., & Görlich, D. (2018). Xpo7 is a broad-spectrum exportin and a nuclear import receptor. *The Journal of Cell Biology*, *217*(7), 2329–2340. https://doi.org/10.1083/jcb.201712013

Albert, R., Jeong, H., & Barabasi, A. L. (2000). Error and attack tolerance of complex networks. *Nature*, *406*(6794), 378–382. https://doi.org/10.1038/35019019

Alvarez-Cardona, J. J., Whited, L. K., & Chemaly, R. F. (2020). Brincidofovir: understanding its unique profile and potential role against adenovirus and other viral infections. *Future Microbiology*, doi: 10.2217/fmb-2019-0288. https://doi.org/10.2217/fmb-2019-0288

Alvisi, G., Musiani, D., Jans, D. A., & Ripalti, A. (2007). An importin alpha/beta-recognized bipartite nuclear localization signal mediates targeting of the human herpes simplex virus type 1 DNA polymerase catalytic subunit pUL30 to the nucleus. *Biochemistry*, *46*(32), 9155–9163. https://doi.org/10.1021/bi7002394

Amlacher, S., Sarges, P., Flemming, D., van Noort, V., Kunze, R., Devos, D. P., Arumugam, M., Bork, P., & Hurt, E. (2011). Insight into structure and assembly of the nuclear pore complex by utilizing the genome of a eukaryotic thermophile. *Cell*, *146*(2), 277–289. https://doi.org/10.1016/j.cell.2011.06.039

Antonova, S. V, Haffke, M., Corradini, E., Mikuciunas, M., Low, T. Y., Signor, L., van Es, R. M., Gupta, K., Scheer, E., Vos, H. R., Tora, L., Heck, A. J. R., Timmers, H. T. M., & Berger, I. (2018). Chaperonin CCT checkpoint function in basal transcription factor TFIID assembly. *Nature Structural & Molecular Biology*, *25*(12), 1119–1127. https://doi.org/10.1038/s41594-018-0156-z

Ao, Z., Danappa Jayappa, K., Wang, B., Zheng, Y., Kung, S., Rassart, E., Depping, R., Kohler, M., Cohen, E. A., & Yao, X. (2010). Importin alpha3 interacts with HIV-1 integrase and contributes to HIV-1 nuclear import and replication. *Journal of Virology*, *84*(17), 8650–8663. https://doi.org/10.1128/JVI.00508-10

Ardito, F., Giuliani, M., Perrone, D., Troiano, G., & Lo Muzio, L. (2017). The crucial role of protein phosphorylation in cell signaling and its use as targeted therapy (Review). *International Journal of Molecular Medicine*, *40*(2), 271–280. https://doi.org/10.3892/ijmm.2017.3036

Arjomand, A., Baker, M. A., Li, C., Buckle, A. M., Jans, D. A., Loveland, K. L., & Miyamoto, Y. (2014). The α-importome of mammalian germ cell maturation provides novel insights for importin biology. *FASEB Journal : Official Publication of the Federation of American Societies for Experimental Biology*, *28*(8), 3480–3493. https://doi.org/10.1096/fj.13-244913

Arnold, M., Nath, A., Hauber, J., & Kehlenbach, R. H. (2006). Multiple importins function as nuclear transport receptors for the Rev protein of human immunodeficiency virus type 1. *The Journal of Biological Chemistry*, *281*(30), 20883–20890. https://doi.org/10.1074/jbc.M602189200

Asally, M., & Yoneda, Y. (2005). Beta-catenin can act as a nuclear import receptor for its partner transcription factor, lymphocyte enhancer factor-1 (lef-1). *Experimental Cell Research*, *308*(2), 357–363. https://doi.org/10.1016/j.yexcr.2005.05.011

Atkinson, S. C., Audsley, M. D., Lieu, K. G., Marsh, G. A., Thomas, D. R., Heaton, S. M., Paxman, J. J., Wagstaff, K. M., Buckle, A. M., Moseley, G. W., Jans, D. A., & Borg, N. A. (2018). Recognition by host nuclear transport proteins drives disorder-to-order transition in Hendra virus V. *Scientific Reports*, *8*(1), 358. https://doi.org/10.1038/s41598-017-18742-8

Babalola, O. E. (2011). Ocular onchocerciasis: Current management and future prospects. *Clinical Ophthalmology*, *5*(1), 1479–1491. https://doi.org/10.2147/opth.s8372

Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., Ren, J., Li, W. W., & Noble, W. S. (2009). MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Research*, *37*(Web Server issue), W202-8. https://doi.org/10.1093/nar/gkp335

Baillie, J. K. (2014). Translational genomics. Targeting the host immune response to fight infection. *Science (New York, N.Y.)*, *344*(6186), 807–808. https://doi.org/10.1126/science.1255074

Bange, G., Murat, G., Sinning, I., Hurt, E., & Kressler, D. (2013). New twist to nuclear import: When two travel together. *Communicative & Integrative Biology*, *6*(4), e24792. https://doi.org/10.4161/cib.24792

Bauer, N. C., Doetsch, P. W., & Corbett, A. H. (2015). Mechanisms Regulating Protein Localization. *Traffic (Copenhagen, Denmark)*, *16*(10), 1039–1061. https://doi.org/10.1111/tra.12310

Bayliss, R., Littlewood, T., Strawn, L. A., Wente, S. R., & Stewart, M. (2002). GLFG and FxFG nucleoporins bind to overlapping sites on importin-beta. *The Journal of*

Biological Chemistry, *277*(52), 50597–50606. https://doi.org/10.1074/jbc.M209037200

Beck, M., & Hurt, E. (2017). The nuclear pore complex: understanding its function through structural insight. *Nature Reviews. Molecular Cell Biology*, *18*(2), 73–89. https://doi.org/10.1038/nrm.2016.147

Bekerman, E., & Einav, S. (2015). Combating emerging viral threats. *Science*, *348*(6232), 282–283. https://doi.org/10.1126/science.aaa3778

Bennett, S. M., Zhao, L., Bosard, C., & Imperiale, M. J. (2015). Role of a nuclear localization signal on the minor capsid proteins VP2 and VP3 in BKPyV nuclear entry. *Virology*, *474*, 110–116. https://doi.org/10.1016/j.virol.2014.10.013

Berk, A. (2013). Adenoviridae. In *Fields Virology* (pp. 1704–1731).

Bernhofer, M., Goldberg, T., Wolf, S., Ahmed, M., Zaugg, J., Boden, M., & Rost, B. (2017). NLSdb—major update for database of nuclear localization signals and nuclear export signals. *Nucleic Acids Research*, *46*(D1), D503–D508. https://doi.org/10.1093/nar/gkx1021

Bian, X.-L., & Wilson, V. G. (2010). Common importin alpha specificity for papillomavirus E2 proteins. *Virus Research*, *150*(1–2), 135–137. https://doi.org/10.1016/j.virusres.2010.02.011

Bischoff, F. R., Klebe, C., Kretschmer, J., Wittinghofer, A., & Ponstingl, H. (1994). RanGAP1 induces GTPase activity of nuclear Ras-related Ran. *Proceedings of the National Academy of Sciences of the United States of America*, *91*(7), 2587–2591. https://doi.org/10.1073/pnas.91.7.2587

Blikstad, C., & Ivarsson, Y. (2015). High-throughput methods for identification of protein-protein interactions involving short linear motifs. In *Cell Communication and Signaling*. https://doi.org/10.1186/s12964-015-0116-8

Bludau, I., & Aebersold, R. (2020). Proteomic and interactomic insights into the molecular basis of cell functional diversity. *Nature Reviews. Molecular Cell Biology*, *21*(6), 327–340. https://doi.org/10.1038/s41580-020-0231-2

Bohnsack, M. T., Czaplinski, K., & Gorlich, D. (2004). Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs. *RNA (New York, N.Y.)*, *10*(2), 185–191. https://doi.org/10.1261/rna.5167604

Boulon, S., Pradet-Balade, B., Verheggen, C., Molle, D., Boireau, S., Georgieva, M., Azzag, K., Robert, M.-C., Ahmad, Y., Neel, H., Lamond, A. I., & Bertrand, E. (2010). HSP90 and its R2TP/Prefoldin-like cochaperone are involved in the cytoplasmic assembly of RNA polymerase II. *Molecular Cell*, *39*(6), 912–924. https://doi.org/10.1016/j.molcel.2010.08.023

Bourbon, H.-M. (2008). Comparative genomics supports a deep evolutionary origin for the large, four-module transcriptional mediator complex. *Nucleic Acids Research*, *36*(12), 3993–4008. https://doi.org/10.1093/nar/gkn349

Bourgeois, B., Hutten, S., Gottschalk, B., Hofweber, M., Richter, G., Sternat, J., Abou-Ajram, C., Gobl, C., Leitinger, G., Graier, W. F., Dormann, D., & Madl, T. (2020). Nonclassical nuclear localization signals mediate nuclear import of CIRBP. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(15), 8503–8514. https://doi.org/10.1073/pnas.1918944117

Bourhis, J.-M., Receveur-Bréchot, V., Oglesbee, M., Zhang, X., Buccellato, M., Darbon, H., Canard, B., Finet, S., & Longhi, S. (2005). The intrinsically disordered C-terminal domain of the measles virus nucleoprotein interacts with the C-terminal domain of the phosphoprotein via two distinct sites and remains predominantly unfolded. *Protein Science : A Publication of the Protein Society*, *14*(8), 1975–1992. https://doi.org/10.1110/ps.051411805

Brameier, M., Krings, A., & MacCallum, R. M. (2007). NucPred--predicting nuclear localization of proteins. *Bioinformatics (Oxford, England)*, *23*(9), 1159–1160. https://doi.org/10.1093/bioinformatics/btm066

Brelot, A., & Chakrabarti, L. A. (2018). CCR5 Revisited: How Mechanisms of HIV Entry Govern AIDS Pathogenesis. *Journal of Molecular Biology*, *430*(17), 2557–2589. https://doi.org/10.1016/j.jmb.2018.06.027

Bremner, K. H., Scherer, J., Yi, J., Vershinin, M., Gross, S. P., & Vallee, R. B. (2009). Adenovirus Transport via Direct Interaction of Cytoplasmic Dynein with the Viral Capsid Hexon Subunit. *Cell Host and Microbe*, *6*(6), 523–535. https://doi.org/10.1016/j.chom.2009.11.006

Brocca, S., Grandori, R., Longhi, S., & Uversky, V. (2020). Liquid-Liquid Phase Separation by Intrinsically Disordered Protein Regions of Viruses: Roles in Viral Life Cycle and Control of Virus-Host Interactions. *International Journal of Molecular Sciences*, *21*(23). https://doi.org/10.3390/ijms21239045

Brown, C. J., Takayama, S., Campen, A. M., Vise, P., Marshall, T. W., Oldfield, C. J., Williams, C. J., & Dunker, A. K. (2002). Evolutionary rate heterogeneity in proteins with long disordered regions. *Journal of Molecular Evolution*, *55*(1), 104–110. https://doi.org/10.1007/s00239-001-2309-6

Bui, K. H., von Appen, A., DiGuilio, A. L., Ori, A., Sparks, L., Mackmull, M.-T., Bock, T., Hagen, W., Andrés-Pons, A., Glavy, J. S., & Beck, M. (2013). Integrated structural analysis of the human nuclear pore complex scaffold. *Cell*, *155*(6), 1233–1243. https://doi.org/10.1016/j.cell.2013.10.055

Bukrinsky, M. I., Sharova, N., Dempsey, M. P., Stanwick, T. L., Bukrinskaya, A. G., Haggerty, S., & Stevenson, M. (1992). Active nuclear import of human immunodeficiency virus type 1 preintegration complexes. *Proceedings of the*

National Academy of Sciences of the United States of America, *89*(14), 6580–6584.

Calderwood, M. A., Venkatesan, K., Xing, L., Chase, M. R., Vazquez, A., Holthaus, A. M., Ewence, A. E., Li, N., Hirozane-Kishikawa, T., Hill, D. E., Vidal, M., Kieff, E., & Johannsen, E. (2007). Epstein-Barr virus and virus human protein interaction maps. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(18), 7606–7611. https://doi.org/10.1073/pnas.0702332104

Caly, L., Druce, J. D., Catton, M. G., Jans, D. A., & Wagstaff, K. M. (2020). The FDA-approved Drug Ivermectin inhibits the replication of SARS-CoV-2 in vitro. *Antiviral Research*, 104787. https://doi.org/10.1016/J.ANTIVIRAL.2020.104787

Canton, J., Fehr, A. R., Fernandez-Delgado, R., Gutierrez-Alvarez, F. J., Sanchez-Aparicio, M. T., Garcia-Sastre, A., Perlman, S., Enjuanes, L., & Sola, I. (2018). MERS-CoV 4b protein interferes with the NF-kappaB-dependent innate immune response during infection. *PLoS Pathogens*, *14*(1), e1006838. https://doi.org/10.1371/journal.ppat.1006838

Cautain, B., Hill, R., de Pedro, N., & Link, W. (2015). Components and regulation of nuclear transport processes. *FEBS Journal*, *282*(3), 445–462. https://doi.org/10.1111/febs.13163

Cavazza, T., & Vernos, I. (2015). The RanGTP Pathway: From Nucleo-Cytoplasmic Transport to Spindle Assembly and Beyond. *Frontiers in Cell and Developmental Biology*, *3*, 82. https://doi.org/10.3389/fcell.2015.00082

Celetti, G., Paci, G., Caria, J., VanDelinder, V., Bachand, G., & Lemke, E. A. (2020). The liquid state of FG-nucleoporins mimics permeability barrier properties of nuclear pore complexes. *The Journal of Cell Biology*, *219*(1). https://doi.org/10.1083/jcb.201907157

Chang, C.-W., Lee, C.-P., Huang, Y.-H., Yang, P.-W., Wang, J.-T., & Chen, M.-R. (2012). Epstein-Barr virus protein kinase BGLF4 targets the nucleus through interaction with nucleoporins. *Journal of Virology*, *86*(15), 8072–8085. https://doi.org/10.1128/JVI.01058-12

Chang, C.-W., Lee, C.-P., Su, M.-T., Tsai, C.-H., & Chen, M.-R. (2015). BGLF4 kinase modulates the structure and transport preference of the nuclear pore complex to facilitate nuclear import of Epstein-Barr virus lytic proteins. *Journal of Virology*, *89*(3), 1703–1718. https://doi.org/10.1128/JVI.02880-14

Charman, M., Herrmann, C., & Weitzman, M. D. (2019). Viral and cellular interactions during adenovirus DNA replication. *FEBS Letters*, *593*(24), 3531–3550. https://doi.org/10.1002/1873-3468.13695

Chen, C. F., Li, S., Chen, Y., Chen, P. L., Sharp, Z. D., & Lee, W. H. (1996). The nuclear localization sequences of the BRCA1 protein interact with the importin-alpha subunit of the nuclear transport signal receptor. *The Journal of Biological*

Chemistry, *271*(51), 32863–32868. https://doi.org/10.1074/jbc.271.51.32863

Chen, C., Wang, J. C.-Y., Pierson, E. E., Keifer, D. Z., Delaleau, M., Gallucci, L., Cazenave, C., Kann, M., Jarrold, M. F., & Zlotnick, A. (2016). Importin beta Can Bind Hepatitis B Virus Core Protein and Empty Core-Like Particles and Induce Structural Changes. *PLoS Pathogens*, *12*(8), e1005802. https://doi.org/10.1371/journal.ppat.1005802

Cheng, H., Dufu, K., Lee, C.-S., Hsu, J. L., Dias, A., & Reed, R. (2006). Human mRNA export machinery recruited to the 5' end of mRNA. *Cell*, *127*(7), 1389–1400. https://doi.org/10.1016/j.cell.2006.10.044

Chook, Y. M., & Süel, K. E. (2011). Nuclear import by karyopherin-βs: recognition and inhibition. *Biochimica et Biophysica Acta*, *1813*(9), 1593–1606. https://doi.org/10.1016/j.bbamcr.2010.10.014

Chopin-Delannoy, S., Thénot, S., Delaunay, F., Buisine, E., Begue, A., Duterque-Coquillaud, M., & Laudet, V. (2003). A specific and unusual nuclear localization signal in the DNA binding domain of the Rev-erb orphan receptors. *Journal of Molecular Endocrinology*, *30*(2), 197–211. https://doi.org/10.1677/jme.0.0300197

Christie, M., Chang, C.-W., Róna, G., Smith, K. M., Stewart, A. G., Takeda, A. A. S., Fontes, M. R. M., Stewart, M., Vértessy, B. G., Forwood, J. K., & Kobe, B. (2016). Structural Biology and Regulation of Protein Import into the Nucleus. *Journal of Molecular Biology*, *428*(10 Pt A), 2060–2090. https://doi.org/10.1016/j.jmb.2015.10.023

Cingolani, G., Petosa, C., Weis, K., & Müller, C. W. (1999). Structure of importin-β bound to the IBB domain of importin-α. *Nature*, *399*(6733), 221–229. https://doi.org/10.1038/20367

Cohen, M. J., King, C. R., Dikeakos, J. D., & Mymryk, J. S. (2014). Functional analysis of the C-terminal region of human adenovirus E1A reveals a misidentified nuclear localization signal. *Virology*, *468–470*, 238–243. https://doi.org/10.1016/j.virol.2014.08.014

Conti, E., & Kuriyan, J. (2000). Crystallographic analysis of the specific yet versatile recognition of distinct nuclear localization signals by karyopherin alpha. *Structure (London, England : 1993)*, *8*(3), 329–338. https://doi.org/10.1016/s0969-2126(00)00107-6

Conti, E., Uy, M., Leighton, L., Blobel, G., & Kuriyan, J. (1998). Crystallographic analysis of the recognition of a nuclear localization signal by the nuclear import factor karyopherin alpha. *Cell*, *94*(2), 193–204.

Copeland, A. M., Newcomb, W. W., & Brown, J. C. (2009). Herpes simplex virus replication: roles of viral proteins and nucleoporins in capsid-nucleus attachment. *Journal of Virology*, *83*(4), 1660–1668. https://doi.org/10.1128/JVI.01139-08

Cordeddu, V., Di Schiavi, E., Pennacchio, L. A., Ma'ayan, A., Sarkozy, A., Fodale, V., Cecchetti, S., Cardinale, A., Martin, J., Schackwitz, W., Lipzen, A., Zampino, G., Mazzanti, L., Digilio, M. C., Martinelli, S., Flex, E., Lepri, F., Bartholdi, D., Kutsche, K., … Tartaglia, M. (2009). Mutation of SHOC2 promotes aberrant protein N-myristoylation and causes Noonan-like syndrome with loose anagen hair. *Nature Genetics*, *41*(9), 1022–1026. https://doi.org/10.1038/ng.425

Crisostomo, L., Soriano, A. M., Frost, J. R., Olanubi, O., Mendez, M., & Pelka, P. (2017). The influence of E1A C-terminus on adenovirus replicative cycle. *Viruses*, *9*(12), E387. https://doi.org/10.3390/v9120387

Czeko, E., Seizl, M., Augsberger, C., Mielke, T., & Cramer, P. (2011). Iwr1 Directs RNA Polymerase II Nuclear Import. *Molecular Cell*, *42*(2), 261–266. https://doi.org/10.1016/j.molcel.2011.02.033

Damian, R. T. (1964). Molecular mimicry: Antigen sharing by parasite and host and its consequences. *American Naturalist*, *98*, 129–149.

Dang, C. V, & Lee, W. M. (1988). Identification of the human c-myc protein nuclear translocation signal. *Molecular and Cellular Biology*, *8*(10), 4048–4054. https://doi.org/10.1128/mcb.8.10.4048

Darling, A. L., & Uversky, V. N. (2018). Intrinsic Disorder and Posttranslational Modifications: The Darker Side of the Biological Dark Matter. *Frontiers in Genetics*, *9*, 158. https://doi.org/10.3389/fgene.2018.00158

Darshan, M. S., Lucchi, J., Harding, E., & Moroianu, J. (2004). The l2 minor capsid protein of human papillomavirus type 16 interacts with a network of nuclear import receptors. *Journal of Virology*, *78*(22), 12179–12188. https://doi.org/10.1128/JVI.78.22.12179-12188.2004

Davey, N. E., Cyert, M. S., & Moses, A. M. (2015). Short linear motifs - ex nihilo evolution of protein regulation. *Cell Communication and Signaling : CCS*, *13*(1), 43. https://doi.org/10.1186/s12964-015-0120-z

Davey, N. E., Haslam, N. J., Shields, D. C., & Edwards, R. J. (2010). SLiMFinder: a web server to find novel, significantly over-represented, short protein motifs. *Nucleic Acids Research*, *38*(Web Server issue), W534-9. https://doi.org/10.1093/nar/gkq440

Davey, N. E., Seo, M.-H., Yadav, V. K., Jeon, J., Nim, S., Krystkowiak, I., Blikstad, C., Dong, D., Markova, N., Kim, P. M., & Ivarsson, Y. (2017). Discovery of short linear motif-mediated interactions through phage display of intrinsically disordered regions of the human proteome. *The FEBS Journal*, *284*(3), 485–498. https://doi.org/10.1111/febs.13995

Davey, N. E., Travé, G., & Gibson, T. J. (2011). How viruses hijack cell regulation. *Trends in Biochemical Sciences*, *36*(3), 159–169. https://doi.org/10.1016/j.tibs.2010.10.002

Davey, N. E., Van Roey, K., Weatheritt, R. J., Toedt, G., Uyar, B., Altenberg, B., Budd, A., Diella, F., Dinkel, H., & Gibson, T. J. (2012). Attributes of short linear motifs. *Molecular BioSystems*, *8*(1), 268–281. https://doi.org/10.1039/c1mb05231d

de Chassey, B., Navratil, V., Tafforeau, L., Hiet, M. S., Aublin-Gex, A., Agaugué, S., Meiffren, G., Pradezynski, F., Faria, B. F., Chantier, T., Le Breton, M., Pellet, J., Davoust, N., Mangeot, P. E., Chaboud, A., Penin, F., Jacob, Y., Vidalain, P. O., Vidal, M., … Lotteau, V. (2008). Hepatitis C virus infection protein network. *Molecular Systems Biology*, *4*, 230. https://doi.org/10.1038/msb.2008.66

De Clercq, E. (2009). Anti-HIV drugs: 25 compounds approved within 25 years after the discovery of HIV. *International Journal of Antimicrobial Agents*, *33*(4), 307–320. https://doi.org/10.1016/j.ijantimicag.2008.10.010

De Clercq, E., & Li, G. (2016). Approved antiviral drugs over the past 50 years. *Clinical Microbiology Reviews*, *29*(3), 695–747. https://doi.org/10.1128/CMR.00102-15

De Wit, E., Van Doremalen, N., Falzarano, D., & Munster, V. J. (2016). SARS and MERS: Recent insights into emerging coronaviruses. *Nature Reviews Microbiology*, *14*(8), 523–534. https://doi.org/10.1038/nrmicro.2016.81

Dehghan, S., Seto, J., Liu, E. B., Walsh, M. P., Dyer, D. W., Chodosh, J., & Seto, D. (2013). Computational analysis of four human adenovirus type 4 genomes reveals molecular evolution through two interspecies recombination events. *Virology*, *443*(2), 197–207. https://doi.org/10.1016/j.virol.2013.05.014

Deng, W., Lin, B. Y., Jin, G., Wheeler, C. G., Ma, T., Harper, J. W., Broker, T. R., & Chow, L. T. (2004). Cyclin/CDK regulates the nucleocytoplasmic localization of the human papillomavirus E1 DNA helicase. *Journal of Virology*, *78*(24), 13954–13965. https://doi.org/10.1128/JVI.78.24.13954-13965.2004

Deutsch, E. W., Csordas, A., Sun, Z., Jarnuczak, A., Perez-Riverol, Y., Ternent, T., Campbell, D. S., Bernal-Llinares, M., Okuda, S., Kawano, S., Moritz, R. L., Carver, J. J., Wang, M., Ishihama, Y., Bandeira, N., Hermjakob, H., & Vizcaíno, J. A. (2016). The ProteomeXchange consortium in 2017: supporting the cultural change in proteomics public data deposition. *Nucleic Acids Research*, *45*(D1), D1100–D1106. https://doi.org/10.1093/nar/gkw936

Devos, D. P., Graf, R., & Field, M. C. (2014). Evolution of the nucleus. *Current Opinion in Cell Biology*, *28*, 8–15. https://doi.org/10.1016/j.ceb.2014.01.004

Di Croce, L. (2011). Regulating the Shuttling of Eukaryotic RNA Polymerase II. *Molecular and Cellular Biology*, *31*(19), 3918–3920. https://doi.org/10.1128/mcb.06093-11

Diella, F., Haslam, N., Chica, C., Budd, A., Michael, S., Brown, N. P., Trave, G., & Gibson, T. J. (2008). Understanding eukaryotic linear motifs and their role in cell signaling and regulation. *Frontiers in Bioscience : A Journal and Virtual Library*,

13, 6580–6603. https://doi.org/10.2741/3175

Dittmar, G., Hernandez, D. P., Kowenz-Leutz, E., Kirchner, M., Kahlert, G., Wesolowski, R., Baum, K., Knoblich, M., Hofstätter, M., Muller, A., Wolf, J., Reimer, U., & Leutz, A. (2019). PRISMA: Protein Interaction Screen on Peptide Matrix Reveals Interaction Footprints and Modifications- Dependent Interactome of Intrinsically Disordered C/EBPβ. *IScience*, *13*, 351–370. https://doi.org/10.1016/j.isci.2019.02.026

Dodding, M. P., & Way, M. (2011). Coupling viruses to dynein and kinesin-1. *EMBO Journal*, *30*(17), 3527–3539. https://doi.org/10.1038/emboj.2011.283

Dong, X., Biswas, A., Süel, K. E., Jackson, L. K., Martinez, R., Gu, H., & Chook, Y. M. (2009). Structural basis for leucine-rich nuclear export signal recognition by CRM1. *Nature*, *458*(7242), 1136–1141. https://doi.org/10.1038/nature07975

Dosztányi, Z., Chen, J., Dunker, A. K., Simon, I., & Tompa, P. (2006). Disorder and sequence repeats in hub proteins and their implications for network evolution. *Journal of Proteome Research*, *5*(11), 2985–2995. https://doi.org/10.1021/pr060171o

Duffy, S., Shackelton, L. A., & Holmes, E. C. (2008). Rates of evolutionary change in viruses: patterns and determinants. *Nature Reviews. Genetics*, *9*(4), 267–276. https://doi.org/10.1038/nrg2323

Dunker, A K, Lawson, J. D., Brown, C. J., Williams, R. M., Romero, P., Oh, J. S., Oldfield, C. J., Campen, A. M., Ratliff, C. M., Hipps, K. W., Ausio, J., Nissen, M. S., Reeves, R., Kang, C., Kissinger, C. R., Bailey, R. W., Griswold, M. D., Chiu, W., Garner, E. C., & Obradovic, Z. (2001). Intrinsically disordered protein. *Journal of Molecular Graphics & Modelling*, *19*(1), 26–59. https://doi.org/10.1016/s1093-3263(00)00138-8

Dunker, A K, Obradovic, Z., Romero, P., Garner, E. C., & Brown, C. J. (2000). Intrinsic protein disorder in complete genomes. *Genome Informatics. Workshop on Genome Informatics*, *11*, 161–171.

Dunker, A Keith, Cortese, M. S., Romero, P., Iakoucheva, L. M., & Uversky, V. N. (2005). Flexible nets. The roles of intrinsic disorder in protein interaction networks. *The FEBS Journal*, *272*(20), 5129–5148. https://doi.org/10.1111/j.1742-4658.2005.04948.x

Duro, N., Miskei, M., & Fuxreiter, M. (2015). Fuzziness endows viral motif-mimicry. *Molecular BioSystems*, *11*, 2821–2829. https://doi.org/10.1039/C5MB00301F

Dyer, M. D., Murali, T. M., & Sobral, B. W. (2008). The landscape of human proteins interacting with viruses and other pathogens. *PLoS Pathogens*, *4*(2), e32. https://doi.org/10.1371/journal.ppat.0040032

Edwards, R. J., Davey, N. E., & Shields, D. C. (2007). SLiMFinder: a probabilistic method for identifying over-represented, convergently evolved, short linear motifs in proteins. *PloS One*, *2*(10), e967. https://doi.org/10.1371/journal.pone.0000967

Edwards, R. J., & Palopoli, N. (2015). Computational prediction of short linear motifs from protein sequences. *Methods in Molecular Biology (Clifton, N.J.)*, *1268*, 89–141. https://doi.org/10.1007/978-1-4939-2285-7_6

Ekman, D., Light, S., Björklund, A. K., & Elofsson, A. (2006). What properties characterize the hub proteins of the protein-protein interaction network of Saccharomyces cerevisiae? *Genome Biology*, *7*(6), R45. https://doi.org/10.1186/gb-2006-7-6-r45

El-Gebali, S., Mistry, J., Bateman, A., Eddy, S. R., Luciani, A., Potter, S. C., Qureshi, M., Richardson, L. J., Salazar, G. A., Smart, A., Sonnhammer, E. L. L., Hirsh, L., Paladin, L., Piovesan, D., Tosatto, S. C. E., & Finn, R. D. (2019). The Pfam protein families database in 2019. *Nucleic Acids Research*, *47*(D1), D427–D432. https://doi.org/10.1093/nar/gky995

Elias, J. E., & Gygi, S. P. (2007). Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nature Methods*, *4*(3), 207–214. https://doi.org/10.1038/nmeth1019

Felsani, A., Mileo, A. M., & Paggi, M. G. (2006). Retinoblastoma family proteins as key targets of the small DNA virus oncoproteins. *Oncogene*, *25*(38), 5277–5285. https://doi.org/10.1038/sj.onc.1209621

Ferrari, R., Gou, D., Jawdekar, G., Johnson, S. A., Nava, M., Su, T., Yousef, A. F., Zemke, N. R., Pellegrini, M., Kurdistani, S. K., & Berk, A. J. (2014). Adenovirus small E1A employs the lysine acetylases p300/CBP and tumor suppressor Rb to repress select host genes and promote productive virus infection. *Cell Host & Microbe*, *16*(5), 663–676. https://doi.org/10.1016/j.chom.2014.10.004

Fichó, E., Reményi, I., Simon, I., & Mészáros, B. (2017). MFIB: a repository of protein complexes with mutual folding induced by binding. *Bioinformatics (Oxford, England)*, *33*(22), 3682–3684. https://doi.org/10.1093/bioinformatics/btx486

Fischer, J., Teimer, R., Amlacher, S., Kunze, R., & Hurt, E. (2015). Linker Nups connect the nuclear pore complex inner ring with the outer ring and transport channel. *Nature Structural & Molecular Biology*, *22*(10), 774–781. https://doi.org/10.1038/nsmb.3084

Fontes, M. R., Teh, T., & Kobe, B. (2000). Structural basis of recognition of monopartite and bipartite nuclear localization sequences by mammalian importin-alpha. *Journal of Molecular Biology*, *297*(5), 1183–1194. https://doi.org/10.1006/jmbi.2000.3642

Fornerod, M., Ohno, M., Yoshida, M., & Mattaj, I. W. (1997). CRM1 is an export receptor for leucine-rich nuclear export signals. *Cell*, *90*(6), 1051–1060.

Fouchier, R. A., Meyer, B. E., Simon, J. H., Fischer, U., Albright, A. V, Gonzalez-Scarano, F., & Malim, M. H. (1998). Interaction of the human immunodeficiency virus type 1 Vpr protein with the nuclear pore complex. *Journal of Virology*, *72*(7), 6004–6013.

Fournane, S., Charbonnier, S., Chapelle, A., Kieffer, B., Orfanoudakis, G., Travé, G., Masson, M., & Nominé, Y. (2011). Surface plasmon resonance analysis of the binding of high-risk mucosal HPV E6 oncoproteins to the PDZ1 domain of the tight junction protein MAGI-1. *Journal of Molecular Recognition : JMR*, *24*(4), 511–523. https://doi.org/10.1002/jmr.1056

Frankish, A., Diekhans, M., Ferreira, A.-M., Johnson, R., Jungreis, I., Loveland, J., Mudge, J. M., Sisu, C., Wright, J., Armstrong, J., Barnes, I., Berry, A., Bignell, A., Carbonell Sala, S., Chrast, J., Cunningham, F., Di Domenico, T., Donaldson, S., Fiddes, I. T., … Flicek, P. (2019). GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Research*, *47*(D1), D766–D773. https://doi.org/10.1093/nar/gky955

Fraser, J. E., Rawlinson, S. M., Wang, C., Jans, D. A., & Wagstaff, K. M. (2014). Investigating dengue virus nonstructural protein 5 (NS5) nuclear import. *Methods in Molecular Biology*, *1138*, 301–328. https://doi.org/10.1007/978-1-4939-0348-1_19

Frey, S., Rees, R., Schunemann, J., Ng, S. C., Funfgeld, K., Huyton, T., & Gorlich, D. (2018). Surface Properties Determining Passage Rates of Proteins through Nuclear Pores. *Cell*, *174*(1), 202–217.

Frieman, M., Yount, B., Heise, M., Kopecky-Bromberg, S. A., Palese, P., & Baric, R. S. (2007). Severe acute respiratory syndrome coronavirus ORF6 antagonizes STAT1 function by sequestering nuclear import factors on the rough endoplasmic reticulum/Golgi membrane. *Journal of Virology*, *81*(18), 9812–9824. https://doi.org/10.1128/JVI.01012-07

Fukuda, M., Asano, S., Nakamura, T., Adachi, M., Yoshida, M., Yanagida, M., & Nishida, E. (1997). CRM1 is responsible for intracellular transport mediated by the nuclear export signal. *Nature*, *390*(6657), 308–311. https://doi.org/10.1038/36894

Fung, H. Y. J., Fu, S.-C., Brautigam, C. A., & Chook, Y. M. (2015). Structural determinants of nuclear export signal orientation in binding to exportin CRM1. *ELife*, *4*. https://doi.org/10.7554/eLife.10034

Fung, H. Y. J., Fu, S.-C., & Chook, Y. M. (2017). Nuclear export receptor CRM1 recognizes diverse conformations in nuclear export signals. *ELife*, *6*. https://doi.org/10.7554/eLife.23961

Furuta, M., Kose, S., Koike, M., Shimi, T., Hiraoka, Y., Yoneda, Y., Haraguchi, T., & Imamoto, N. (2004). Heat-shock induced nuclear retention and recycling inhibition of importin alpha. *Genes to Cells : Devoted to Molecular & Cellular Mechanisms*, *9*(5), 429–441. https://doi.org/10.1111/j.1356-9597.2004.00734.x

Fuxreiter, M., Tompa, P., & Simon, I. (2007). Local structural disorder imparts plasticity on linear motifs. *Bioinformatics (Oxford, England)*, *23*(8), 950–956. https://doi.org/10.1093/bioinformatics/btm035

Gagne, B., Tremblay, N., Park, A. Y., Baril, M., & Lamarre, D. (2017). Importin beta1 targeting by hepatitis C virus NS3/4A protein restricts IRF3 and NF-kappaB signaling of IFNB1 antiviral response. *Traffic (Copenhagen, Denmark)*, *18*(6), 362–377. https://doi.org/10.1111/tra.12480

Gama-Carvalho, M., & Carmo-Fonseca, M. (2001). The rules and roles of nucleocytoplasmic shuttling proteins. *FEBS Letters*, *498*(2–3), 157–163. https://doi.org/10.1016/s0014-5793(01)02487-5

Garamszegi, S., Franzosa, E. a, & Xia, Y. (2013). Signatures of pleiotropy, economy and convergent evolution in a domain-resolved map of human-virus protein-protein interaction networks. *PLoS Pathogens*, *9*(12), e1003778. https://doi.org/10.1371/journal.ppat.1003778

Ghebremedhin, B. (2014). Human adenovirus: Viral pathogen with increasing importance. *European Journal of Microbiology and Immunology*, *4*(1), 26–33. https://doi.org/10.1556/EuJMI.4.2014.1.2

Gibson, T. J., Dinkel, H., Van Roey, K., & Diella, F. (2015). Experimental detection of short regulatory motifs in eukaryotic proteins: tips for good practice as well as for bad. *Cell Communication and Signaling : CCS*, *13*, 42. https://doi.org/10.1186/s12964-015-0121-y

Gingras, A.-C., Abe, K. T., & Raught, B. (2019). Getting to know the neighborhood: using proximity-dependent biotinylation to characterize protein complexes and map organelles. *Current Opinion in Chemical Biology*, *48*, 44–54. https://doi.org/10.1016/j.cbpa.2018.10.017

Goldfarb, D. S., Corbett, A. H., Mason, D. A., Harreman, M. T., & Adam, S. A. (2004a). Importin α: A multipurpose nuclear-transport receptor. In *Trends in Cell Biology*. https://doi.org/10.1016/j.tcb.2004.07.016

Goldfarb, D. S., Corbett, A. H., Mason, D. A., Harreman, M. T., & Adam, S. A. (2004b). Importin alpha: a multipurpose nuclear-transport receptor. *Trends in Cell Biology*, *14*(9), 505–514. https://doi.org/10.1016/j.tcb.2004.07.016

Görlich, D, Panté, N., Kutay, U., Aebi, U., & Bischoff, F. R. (1996). Identification of different roles for RanGDP and RanGTP in nuclear protein import. *The EMBO Journal*, *15*(20), 5584–5594.

Görlich, Dirk, Kostka, S., Kraft, R., Dingwall, C., Laskey, R. A., Hartmann, E., & Prehn, S. (1995). Two different subunits of importin cooperate to recognize nuclear localization signals and bind them to the nuclear envelope. *Current Biology*, *5*(4), 383–392. https://doi.org/10.1016/S0960-9822(95)00079-0

Görlich, Dirk, Prehn, S., Laskey, R. A., & Hartmann, E. (1994). Isolation of a protein that is essential for the first step of nuclear protein import. *Cell*. https://doi.org/10.1016/0092-8674(94)90067-1

Gupta, A., Kailasam, S., & Bansal, M. (2016). Insights into the Structural Dynamics of Nucleocytoplasmic Transport of tRNA by Exportin-t. *Biophysical Journal*, *110*(6), 1264–1279. https://doi.org/10.1016/j.bpj.2016.02.015

Gustin, K., & Sarnow, P. (2001). Effects of poliovirus infection on nucleo-cytoplasmic trafficking and nuclear pore complex composition. *The EMBO Journal*, *20*(1–2), 240–249. https://doi.org/10.1093/emboj/20.1.240

Gustin, K., & Sarnow, P. (2002). Inhibition of nuclear import and alteration of nuclear pore complex composition by rhinovirus. *Journal of Virology*, *76*(17), 8787–8796.

Han, J.-H., Batey, S., Nickson, A. A., Teichmann, S. A., & Clarke, J. (2007). The folding and evolution of multidomain proteins. In *Nature reviews. Molecular cell biology* (Vol. 8, Issue 4, pp. 319–330). https://doi.org/10.1038/nrm2144

Harel, A., & Forbes, D. J. (2004). Importin beta: conducting a much larger cellular symphony. *Molecular Cell*, *16*(3), 319–330. https://doi.org/10.1016/j.molcel.2004.10.026

Harreman, M. T., Cohen, P. E., Hodel, M. R., Truscott, G. J., Corbett, A. H., & Hodel, A. E. (2003). Characterization of the auto-inhibitory sequence within the N-terminal domain of importin α. *Journal of Biological Chemistry*, *278*(24), 21361–21369. https://doi.org/10.1074/jbc.M301114200

Harreman, M. T., Hodel, M. R., Fanara, P., Hodel, A. E., & Corbett, A. H. (2003). The auto-inhibitory function of importin α is essential in vivo. *Journal of Biological Chemistry*. https://doi.org/10.1074/jbc.M210951200

Haynes, C., Oldfield, C. J., Ji, F., Klitgord, N., Cusick, M. E., Radivojac, P., Uversky, V. N., Vidal, M., & Iakoucheva, L. M. (2006). Intrinsic disorder is a common feature of hub proteins from four eukaryotic interactomes. *PLoS Computational Biology*, *2*(8), e100. https://doi.org/10.1371/journal.pcbi.0020100

Hearing, P., & Shenk, T. (1983). The adenovirus type 5 E1A transcriptional control region contains a duplicated enhancer element. *Cell*, *33*(3), 695–703.

Henderson, B. R., & Percipalle, P. (1997). 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*, *274*(5), 693–707. https://doi.org/10.1006/jmbi.1997.1420

Hidalgo, P., & Gonzalez, R. A. (2019). Formation of adenovirus DNA replication compartments. *FEBS Letters*, *593*(24), 3518–3530. https://doi.org/10.1002/1873-3468.13672

Hilpert, K., Winkler, D. F. H., & Hancock, R. E. W. (2007). Peptide arrays on cellulose support: SPOT synthesis, a time and cost efficient method for synthesis of large numbers of peptides in a parallel and addressable fashion. *Nature Protocols*, *2*(6), 1333–1349. https://doi.org/10.1038/nprot.2007.160

Hindley, C. E., Lawrence, F. J., & Matthews, D. A. (2007). A role for transportin in the nuclear import of adenovirus core proteins and DNA. *Traffic (Copenhagen, Denmark)*, *8*(10), 1313–1322. https://doi.org/10.1111/j.1600-0854.2007.00618.x

Hirschi, A., Cecchini, M., Steinhardt, R. C., Schamber, M. R., Dick, F. A., & Rubin, S. M. (2010). An overlapping kinase and phosphatase docking site regulates activity of the retinoblastoma protein. *Nature Structural & Molecular Biology*, *17*(9), 1051–1057. https://doi.org/10.1038/nsmb.1868

Hoelz, A., Debler, E. W., & Blobel, G. (2011). The Structure of the Nuclear Pore Complex. *Annual Review of Biochemistry*, *80*, 613–643. https://doi.org/10.1146/annurev-biochem-060109-151030

Hornbeck, P. V, Kornhauser, J. M., Latham, V., Murray, B., Nandhikonda, V., Nord, A., Skrzypek, E., Wheeler, T., Zhang, B., & Gnad, F. (2019). 15 years of PhosphoSitePlus®: integrating post-translationally modified sites, disease variants and isoforms. *Nucleic Acids Research*, *47*(D1), D433–D441. https://doi.org/10.1093/nar/gky1159

Horwitz, G. A., Zhang, K., McBrian, M. A., Grunstein, M., Kurdistani, S. K., & Berk, A. J. (2008). Adenovirus Small e1a Alters Global Patterns of Histone Modification. *Science*, *321*(5892), 1084–1085. https://doi.org/10.1126/science.1155544

Howley, P. M., & Livingston, D. M. (2009). Small DNA tumor viruses: Large contributors to biomedical sciences. *Virology*, *384*(2), 256–259. https://doi.org/10.1016/j.virol.2008.12.006

Hsieh, J. C., Shimizu, Y., Minoshima, S., Shimizu, N., Haussler, C. A., Jurutka, P. W., & Haussler, M. R. (1998). Novel nuclear localization signal between the two DNA-binding zinc fingers in the human vitamin D receptor. *Journal of Cellular Biochemistry*, *70*(1), 94–109.

Hu, G., Wu, Z., Uversky, V. N., & Kurgan, L. (2017). Functional Analysis of Human Hub Proteins and Their Interactors Involved in the Intrinsic Disorder-Enriched Interactions. *International Journal of Molecular Sciences*, *18*(12). https://doi.org/10.3390/ijms18122761

Huber, J., Cronshagen, U., Kadokura, M., Marshallsay, C., Wada, T., Sekine, M., & Lührmann, R. (1998). Snurportin1, an m3G-cap-specific nuclear import receptor with a novel domain structure. *The EMBO Journal*, *17*(14), 4114–4126. https://doi.org/10.1093/emboj/17.14.4114

Huber, J., Dickmanns, A., & Lührmann, R. (2002). The importin-beta binding domain of

snurportin1 is responsible for the Ran- and energy-independent nuclear import of spliceosomal U snRNPs in vitro. *The Journal of Cell Biology*, *156*(3), 467–479. https://doi.org/10.1083/jcb.200108114

Hudjetz, B., & Gabriel, G. (2012). Human-like PB2 627K influenza virus polymerase activity is regulated by importin-alpha1 and -alpha7. *PLoS Pathogens*, *8*(1), e1002488. https://doi.org/10.1371/journal.ppat.1002488

Huh, W.-K., Falvo, J. V, Gerke, L. C., Carroll, A. S., Howson, R. W., Weissman, J. S., & O'Shea, E. K. (2003). Global analysis of protein localization in budding yeast. *Nature*, *425*(6959), 686–691. https://doi.org/10.1038/nature02026

Huttlin, E. L., Ting, L., Bruckner, R. J., Gebreab, F., Gygi, M. P., Szpyt, J., Tam, S., Zarraga, G., Colby, G., Baltier, K., Dong, R., Guarani, V., Vaites, L. P., Ordureau, A., Rad, R., Erickson, B. K., Wühr, M., Chick, J., Zhai, B., … Gygi, S. P. (2015). The BioPlex Network: A Systematic Exploration of the Human Interactome. *Cell*, *162*(2), 425–440. https://doi.org/10.1016/j.cell.2015.06.043

Iakoucheva, L. M., Brown, C. J., Lawson, J. D., Obradović, Z., & Dunker, A. K. (2002). Intrinsic disorder in cell-signaling and cancer-associated proteins. *Journal of Molecular Biology*, *323*(3), 573–584. https://doi.org/10.1016/s0022-2836(02)00969-5

Iakoucheva, L. M., Radivojac, P., Brown, C. J., O'Connor, T. R., Sikes, J. G., Obradovic, Z., & Dunker, A. K. (2004). The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Research*, *32*(3), 1037–1049. https://doi.org/10.1093/nar/gkh253

Igarashi, M., Ito, K., Kida, H., & Takada, A. (2008). Genetically destined potentials for N-linked glycosylation of influenza virus hemagglutinin. *Virology*, *376*(2), 323–329. https://doi.org/10.1016/j.virol.2008.03.036

Ivashkiv, L. B., & Donlin, L. T. (2014). Regulation of type I interferon responses. *Nature Reviews. Immunology*, *14*(1), 36–49. https://doi.org/10.1038/nri3581

Jans, D. A., & Martin, A. J. (2018). Nucleocytoplasmic Trafficking of Dengue Non-structural Protein 5 as a Target for Antivirals. *Advances in Experimental Medicine and Biology*, *1062*, 199–213. https://doi.org/10.1007/978-981-10-8727-1_15

Jans, D. A., Martin, A. J., & Wagstaff, K. M. (2019). Inhibitors of nuclear transport. *Current Opinion in Cell Biology*, *58*, 50–60. https://doi.org/10.1016/j.ceb.2019.01.001

Jeffrey, P. D., Gorina, S., & Pavletich, N. P. (1995). Crystal structure of the tetramerization domain of the p53 tumor suppressor at 1.7 angstroms. *Science (New York, N.Y.)*, *267*(5203), 1498–1502. https://doi.org/10.1126/science.7878469

Jenkins, G. M., Rambaut, A., Pybus, O. G., & Holmes, E. C. (2002). Rates of molecular

evolution in RNA viruses: a quantitative phylogenetic analysis. *Journal of Molecular Evolution*, *54*(2), 156–165. https://doi.org/10.1007/s00239-001-0064-3

Jenkins, Y., McEntee, M., Weis, K., & Greene, W. C. (1998). Characterization of HIV-1 vpr nuclear import: analysis of signals and pathways. *The Journal of Cell Biology*, *143*(4), 875–885.

Jeong, H., Mason, S. P., Barabási, A. L., & Oltvai, Z. N. (2001). Lethality and centrality in protein networks. *Nature*, *411*(6833), 41–42. https://doi.org/10.1038/35075138

Jespersen, N., & Barbar, E. (2020). Emerging Features of Linear Motif-Binding Hub Proteins. *Trends in Biochemical Sciences*, *45*(5), 375–384. https://doi.org/10.1016/j.tibs.2020.01.004

Jin, Y., Zhang, R. F., Xie, Z. P., Yan, K. L., Gao, H. C., Song, J. R., Yuan, X. H., Hou, Y. De, & Duan, Z. J. (2013). Prevalence of adenovirus in children with acute respiratory tract infection in Lanzhou, China. *Virology Journal*, *10*. https://doi.org/10.1186/1743-422X-10-271

Jones, N., & Shenk, T. (1979). Isolation of adenovirus type 5 host range deletion mutants defective for transformation of rat embryo cells. *Cell*, *17*(3), 683–689.

Jones, R. E., Wegrzyn, R. J., Patrick, D. R., Balishin, N. L., Vuocolo, G. A., Riemen, M. W., Defeo-Jones, D., Garsky, V. M., Heimbrook, D. C., & Oliff, A. (1990). Identification of HPV-16 E7 peptides that are potent antagonists of E7 binding to the retinoblastoma suppressor protein. *The Journal of Biological Chemistry*, *265*(22), 12782–12785.

Kalderon, D, Roberts, B. L., Richardson, W. D., & Smith, A. E. (1984). A short amino acid sequence able to specify nuclear location. *Cell*, *39*(3 Pt 2), 499–509.

Kalderon, Daniel, Roberts, B. L., Richardson, W. D., & Smith, A. E. (1984). A short amino acid sequence able to specify nuclear location. *Cell*, *39*(3 PART 2), 499–509. https://doi.org/10.1016/0092-8674(84)90457-4

Kamata, M., Nitahara-Kasahara, Y., Miyamoto, Y., Yoneda, Y., & Aida, Y. (2005). Importin-alpha promotes passage through the nuclear pore complex of human immunodeficiency virus type 1 Vpr. *Journal of Virology*, *79*(6), 3557–3564. https://doi.org/10.1128/JVI.79.6.3557-3564.2005

Kamenova, I., Mukherjee, P., Conic, S., Mueller, F., El-Saafin, F., Bardot, P., Garnier, J.-M., Dembele, D., Capponi, S., Timmers, H. T. M., Vincent, S. D., & Tora, L. (2019). Co-translational assembly of mammalian nuclear multisubunit complexes. *Nature Communications*, *10*(1), 1740. https://doi.org/10.1038/s41467-019-09749-y

Kastritis, P. L., Moal, I. H., Hwang, H., Weng, Z., Bates, P. A., Bonvin, A. M. J. J., & Janin, J. (2011). A structure-based benchmark for protein-protein binding affinity. *Protein Science : A Publication of the Protein Society*, *20*(3), 482–491.

https://doi.org/10.1002/pro.580

Kataoka, N., Bachorik, J. L., & Dreyfuss, G. (1999). Transportin-SR, a nuclear import receptor for SR proteins. *The Journal of Cell Biology*, *145*(6), 1145–1152. https://doi.org/10.1083/jcb.145.6.1145

Kaufmann, S. H. E., Dorhoi, A., Hotchkiss, R. S., & Bartenschlager, R. (2018). Host-directed therapies for bacterial and viral infections. *Nature Reviews. Drug Discovery*, *17*(1), 35–56. https://doi.org/10.1038/nrd.2017.162

Kelley, J. B., & Paschal, B. M. (2007). Hyperosmotic stress signaling to the nucleus disrupts the Ran gradient and the production of RanGTP. *Molecular Biology of the Cell*, *18*(11), 4365–4376. https://doi.org/10.1091/mbc.e07-01-0089

Kelley, J. B., Talley, A. M., Spencer, A., Gioeli, D., & Paschal, B. M. (2010). Karyopherin α7 (KPNA7), a divergent member of the importin α family of nuclear import receptors. *BMC Cell Biology*, *11*(63). https://doi.org/10.1186/1471-2121-11-63

Khanal, S., Ghimire, P., & Dhamoon, A. S. (2018). The repertoire of adenovirus in human disease: The innocuous to the deadly. *Biomedicines*, *6*(1), E30. https://doi.org/10.3390/biomedicines6010030

Kim, S., & Pevzner, P. A. (2014). MS-GF+ makes progress towards a universal database search tool for proteomics. *Nature Communications*, *5*, 5277. https://doi.org/10.1038/ncomms6277

Kimura, M., & Imamoto, N. (2014). Biological significance of the importin-β family-dependent nucleocytoplasmic transport pathways. *Traffic (Copenhagen, Denmark)*, *15*(7), 727–748. https://doi.org/10.1111/tra.12174

Kimura, M., Morinaka, Y., Imai, K., Kose, S., Horton, P., & Imamoto, N. (2017). Extensive cargo identification reveals distinct biological roles of the 12 importin pathways. *ELife*, *6*, e21184. https://doi.org/10.7554/eLife.21184

King, C. R., Cohen, M. J., Fonseca, G. J., Dirk, B. S., Dikeakos, J. D., & Mymryk, J. S. (2016). Functional and Structural Mimicry of Cellular Protein Kinase A Anchoring Proteins by a Viral Oncoprotein. *PLoS Pathogens*, *12*(5), e1005621. https://doi.org/10.1371/journal.ppat.1005621

King, C. R., Gameiro, S. F., Tessier, T. M., Zhang, A., & Mymryk, J. S. (2018). Mimicry of Cellular A Kinase-Anchoring Proteins Is a Conserved and Critical Function of E1A across Various Human Adenovirus Species. *Journal of Virology*, *92*(8), e01902-17. https://doi.org/10.1128/JVI.01902-17

King, C. R., Zhang, A., & Mymryk, J. S. (2016). The Persistent Mystery of Adenovirus Persistence. *Trends in Microbiology*, *24*(5), 323–324. https://doi.org/10.1016/j.tim.2016.02.007

King, C. R., Zhang, A., Tessier, T. M., Gameiro, S. F., & Mymryk, J. S. (2018). Hacking the Cell: Network Intrusion and Exploitation by Adenovirus E1A. *MBio*, *9*(3), e00390-18. https://doi.org/10.1128/mBio.00390-18

Kırlı, K., Karaca, S., Dehne, H. J., Samwer, M., Pan, K. T., Lenz, C., Urlaub, H., & Görlich, D. (2015). A deep proteomics perspective on CRM1-mediated nuclear export and nucleocytoplasmic partitioning. *ELife*, *4*, e11466. https://doi.org/10.7554/eLife.11466

Knight, J. D. R., Choi, H., Gupta, G. D., Pelletier, L., Raught, B., Nesvizhskii, A. I., & Gingras, A.-C. (2017). ProHits-viz: a suite of web tools for visualizing interaction proteomics data. *Nature Methods*, *14*(7), 645–646. https://doi.org/10.1038/nmeth.4330

Knockenhauer, K. E., & Schwartz, T. U. (2016). The Nuclear Pore Complex as a Flexible and Dynamic Gate. *Cell*, *164*(6), 1162–1171. https://doi.org/10.1016/j.cell.2016.01.034

Ko, J.-H., Woo, H., Oh, H. S., Moon, S. M., Choi, J. Y., Lim, J. U., Kim, D., Byun, J., Kwon, S.-H., Kang, D., Heo, J. Y., & Peck, K. R. (2019). Ongoing outbreak of human adenovirus-associated acute respiratory illness in the Republic of Korea military, 2013 to 2018. *The Korean Journal of Internal Medicine*, DOI: 10.3904/kjim.2019.092. https://doi.org/10.3904/kjim.2019.092

Kobe, B. (1999). Autoinhibition by an internal nuclear localization signal revealed by the crystal structure of mammalian importin α. *Nature Structural Biology*. https://doi.org/10.1038/7625

Kohler, M., Gorlich, D., Hartmann, E., & Franke, J. (2001). Adenoviral E1A protein nuclear import is preferentially mediated by importin alpha3 in vitro. *Virology*, *289*(2), 186–191. https://doi.org/10.1006/viro.2001.1151

Kohm, A. P., Fuller, K. G., & Miller, S. D. (2003). Mimicking the way to autoimmunity: an evolving theory of sequence and structural homology. *Trends in Microbiology*, *11*(3), 101–105. https://doi.org/10.1016/s0966-842x(03)00006-4

Kojima, C., Hashimoto, A., Yabuta, I., Hirose, M., Hashimoto, S., Kanaho, Y., Sumimoto, H., Ikegami, T., & Sabe, H. (2004). Regulation of Bin1 SH3 domain binding by phosphoinositides. *The EMBO Journal*, *23*(22), 4413–4422. https://doi.org/10.1038/sj.emboj.7600442

Kosako, H., & Imamoto, N. (2010). Phosphorylation of nucleoporins: signal transduction-mediated regulation of their interaction with nuclear transport receptors. *Nucleus (Austin, Tex.)*, *1*(4), 309–313. https://doi.org/10.4161/nucl.1.4.11744

Kose, S., Furuta, M., & Imamoto, N. (2012). Hikeshi, a nuclear import carrier for Hsp70s, protects cells from heat shock-induced nuclear damage. *Cell*, *149*(3), 578–

589. https://doi.org/10.1016/j.cell.2012.02.058

Kosugi, S., Hasebe, M., Matsumura, N., Takashima, H., Miyamoto-Sato, E., Tomita, M., & Yanagawa, H. (2009). Six classes of nuclear localization signals specific to different binding grooves of importinα. *Journal of Biological Chemistry*, *284*(1), 478–485. https://doi.org/10.1074/jbc.M807017200

Kosugi, S., Yanagawa, H., Terauchi, R., & Tabata, S. (2014). NESmapper: Accurate Prediction of Leucine-Rich Nuclear Export Signals Using Activity-Based Profiles. *PLOS Computational Biology*, *10*(9), e1003841. https://doi.org/10.1371/journal.pcbi.1003841

Kosyna, F. K., & Depping, R. (2018). Controlling the Gatekeeper: Therapeutic Targeting of Nuclear Transport. *Cells*, *7*(11), E221. https://doi.org/10.3390/cells7110221

Kressler, D., Bange, G., Ogawa, Y., Stjepanovic, G., Bradatsch, B., Pratte, D., Amlacher, S., Strauß, D., Yoneda, Y., Katahira, J., Sinning, I., & Hurt, E. (2012). Synchronizing nuclear import of ribosomal proteins with ribosome assembly. *Science (New York, N.Y.)*, *338*(6107), 666–671. https://doi.org/10.1126/science.1226960

Krystkowiak, I., & Davey, N. E. (2017). SLiMSearch: A framework for proteome-wide discovery and annotation of functional modules in intrinsically disordered regions. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkx238

Kumar, M., Gouw, M., Michael, S., Sámano-Sánchez, H., Pancsa, R., Glavina, J., Diakogianni, A., Valverde, J. A., Bukirova, D., Čalyševa, J., Palopoli, N., Davey, N. E., Chemes, L. B., & Gibson, T. J. (2020). ELM-the eukaryotic linear motif resource in 2020. *Nucleic Acids Research*, *48*(D1), D296–D306. https://doi.org/10.1093/nar/gkz1030

Lafolie, J., Mirand, A., Salmona, M., Lautrette, A., Archimbaud, C., Brebion, A., Regagnon, C., Chambon, M., Mercier-Delarue, S., Le Goff, J., & Henquell, C. (2016). Severe pneumonia associated with adenovirus type 55 infection, France, 2014. *Emerging Infectious Diseases*, *22*(11), 2012–2014. https://doi.org/10.3201/eid2211.160728

Lai, C. Y., Lee, C. J., Lu, C. Y., Lee, P. I., Shao, P. L., Wu, E. T., Wang, C. C., Tan, B. F., Chang, H. Y., Hsia, S. H., Lin, J. J., Chang, L. Y., Huang, Y. C., & Huang, L. M. (2013). Adenovirus Serotype 3 and 7 Infection with Acute Respiratory Failure in Children in Taiwan, 2010-2011. *PLoS ONE*, *8*(1), e53614. https://doi.org/10.1371/journal.pone.0053614

Laing, R., Gillan, V., & Devaney, E. (2017). Ivermectin - Old Drug, New Tricks? *Trends in Parasitology*, *33*(6), 463–472. https://doi.org/10.1016/j.pt.2017.02.004

Lam, M. H., Briggs, L. J., Hu, W., Martin, T. J., Gillespie, M. T., & Jans, D. A. (1999). Importin beta recognizes parathyroid hormone-related protein with high affinity and

mediates its nuclear import in the absence of importin alpha. *The Journal of Biological Chemistry*, *274*(11), 7391–7398. https://doi.org/10.1074/jbc.274.11.7391

Lane, D. P. (1992). Cancer. p53, guardian of the genome. In *Nature* (Vol. 358, Issue 6381, pp. 15–16). https://doi.org/10.1038/358015a0

Lange, A., Mills, R. E., Lange, C. J., Stewart, M., Devine, S. E., & Corbett, A. H. (2007). Classical Nuclear Localization Signals: Definition, Function, and Interaction with Importin. *Journal of Biological Chemistry*, *282*(8), 5101–5105. https://doi.org/10.1074/jbc.R600026200

Lau, E., Han, Y., Williams, D. R., Thomas, C. T., Shrestha, R., Wu, J. C., & Lam, M. P. Y. (2019). Splice-Junction-Based Mapping of Alternative Isoforms in the Human Proteome. *Cell Reports*, *29*(11), 3751-3765.e5. https://doi.org/10.1016/j.celrep.2019.11.026

Laurell, E., Beck, K., Krupina, K., Theerthagiri, G., Bodenmiller, B., Horvath, P., Aebersold, R., Antonin, W., & Kutay, U. (2011). Phosphorylation of Nup98 by multiple kinases is crucial for NPC disassembly during mitotic entry. *Cell*, *144*(4), 539–550. https://doi.org/10.1016/j.cell.2011.01.012

Le Roux, L. G., & Moroianu, J. (2003). Nuclear entry of high-risk human papillomavirus type 16 E6 oncoprotein occurs via several pathways. *Journal of Virology*, *77*(4), 2330–2337. https://doi.org/10.1128/jvi.77.4.2330-2337.2003

Le Rouzic, E., Mousnier, A., Rustum, C., Stutz, F., Hallberg, E., Dargemont, C., & Benichou, S. (2002). Docking of HIV-1 Vpr to the nuclear envelope is mediated by the interaction with the nucleoporin hCG1. *The Journal of Biological Chemistry*, *277*(47), 45091–45098. https://doi.org/10.1074/jbc.M207439200

Le Sage, V., & Mouland, A. J. (2013). Viral subversion of the nuclear pore complex. *Viruses*, *5*(8), 2019–2042. https://doi.org/10.3390/v5082019

Lee, B. J., Cansizoglu, A. E., Suel, K. E., Louis, T. H., Zhang, Z., & Chook, Y. M. (2006). Rules for nuclear localization sequence recognition by karyopherin beta 2. *Cell*, *126*(3), 543–558. https://doi.org/10.1016/j.cell.2006.05.049

Lee, C. W., Ferreon, J. C., Ferreon, A. C. M., Arai, M., & Wright, P. E. (2010). Graded enhancement of p53 binding to CREB-binding protein (CBP) by multisite phosphorylation. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(45), 19290–19295. https://doi.org/10.1073/pnas.1013078107

Lee, J., Choi, E. H., & Lee, H. J. (2010). Comprehensive serotyping and epidemiology of human adenovirus isolated from the respiratory tract of Korean children over 17 consecutive years (1991-2007). *Journal of Medical Virology*, *82*(4), 624–631. https://doi.org/10.1002/jmv.21701

Lee, S. M.-Y., & Yen, H.-L. (2012). Targeting the host or the virus: current and novel

concepts for antiviral approaches against influenza virus infection. *Antiviral Research*, *96*(3), 391–404. https://doi.org/10.1016/j.antiviral.2012.09.013

Lee, W. J., Jung, H. D., Cheong, H. M., & Kim, K. (2015). Molecular epidemiology of a post-influenza pandemic outbreak of acute respiratory infections in Korea caused by human adenovirus type 3. *Journal of Medical Virology*, *87*(1), 10–17. https://doi.org/10.1002/jmv.23984

Lin, D. H., Stuwe, T., Schilbach, S., Rundlet, E. J., Perriches, T., Mobbs, G., Fan, Y., Thierbach, K., Huber, F. M., Collins, L. N., Davenport, A. M., Jeon, Y. E., & Hoelz, A. (2016). Architecture of the symmetric core of the nuclear pore. *Science (New York, N.Y.)*, *352*(6283), aaf1015. https://doi.org/10.1126/science.aaf1015

Lin, J., & Hu, J. (2013). SeqNLS: Nuclear Localization Signal Prediction Based on Frequent Pattern Mining and Linear Motif Scoring. *PLOS ONE*, *8*(10), e76864. https://doi.org/10.1371/journal.pone.0076864

Linding, R., Jensen, L. J., Diella, F., Bork, P., Gibson, T. J., & Russell, R. B. (2003). Protein disorder prediction: implications for structural proteomics. *Structure (London, England : 1993)*, *11*(11), 1453–1459. https://doi.org/10.1016/j.str.2003.10.002

Lion, T. (2014). Adenovirus infections in immunocompetent and immunocompromised patients. *Clinical Microbiology Reviews*, *27*(3), 441–462. https://doi.org/10.1128/CMR.00116-13

Liu, Y., Sun, J., Zhang, H., Wang, M., Gao, G. F., & Li, X. (2016). Ebola virus encodes a miR-155 analog to regulate importin-alpha5 expression. *Cellular and Molecular Life Sciences : CMLS*, *73*(19), 3733–3744. https://doi.org/10.1007/s00018-016-2215-0

Liu, Z., Zhang, Y., & Nielsen, J. (2019). Synthetic Biology of Yeast. *Biochemistry*, *58*(11), 1511–1520. https://doi.org/10.1021/acs.biochem.8b01236

Lolodi, O., Yamazaki, H., Otsuka, S., Kumeta, M., & Yoshimura, S. H. (2016). Dissecting in vivo steady-state dynamics of karyopherin-dependent nuclear transport. *Molecular Biology of the Cell*, *27*(1), 167–176. https://doi.org/10.1091/mbc.E15-08-0601

Lott, K., Bhardwaj, A., Sims, P. J., & Cingolani, G. (2011). A minimal nuclear localization signal (NLS) in human phospholipid scramblase 4 that binds only the minor NLS-binding site of importin alpha1. *The Journal of Biological Chemistry*, *286*(32), 28160–28169. https://doi.org/10.1074/jbc.M111.228007

Lott, K., & Cingolani, G. (2011). The importin beta binding domain as a master regulator of nucleocytoplasmic transport. *Biochimica et Biophysica Acta*, *1813*(9), 1578–1592. https://doi.org/10.1016/j.bbamcr.2010.10.012

Lu, M., Zak, J., Chen, S., Sanchez-Pulido, L., Severson, D. T., Endicott, J., Ponting, C.

P., Schofield, C. J., & Lu, X. (2014). A code for RanGDP binding in ankyrin repeats defines a nuclear import pathway. *Cell*, *157*(5), 1130–1145. https://doi.org/10.1016/j.cell.2014.05.006

Luck, K., Kim, D.-K., Lambourne, L., Spirohn, K., Begg, B. E., Bian, W., Brignall, R., Cafarelli, T., Campos-Laborie, F. J., Charloteaux, B., Choi, D., Cote, A. G., Daley, M., Deimling, S., Desbuleux, A., Dricot, A., Gebbia, M., Hardy, M. F., Kishore, N., … Calderwood, M. A. (2020). A reference map of the human binary protein interactome. *Nature*, *580*(7803), 402–408. https://doi.org/10.1038/s41586-020-2188-x

Lv, C., Liu, W., Wang, B., Dang, R., Qiu, L., Ren, J., Yan, C., Yang, Z., & Wang, X. (2018). Ivermectin inhibits DNA polymerase UL42 of pseudorabies virus entrance into the nucleus and proliferation of the virus in vitro and vivo. *Antiviral Research*, *159*, 55–62. https://doi.org/10.1016/j.antiviral.2018.09.010

Lyons, R. H., Ferguson, B. Q., & Rosenberg, M. (1987). Pentapeptide nuclear localization signal in adenovirus E1a. *Molecular and Cellular Biology*, *7*(7), 2451–2456.

Ma, Z.-Q., Dasari, S., Chambers, M. C., Litton, M. D., Sobecki, S. M., Zimmerman, L. J., Halvey, P. J., Schilling, B., Drake, P. M., Gibson, B. W., & Tabb, D. L. (2009). IDPicker 2.0: Improved protein assembly with high discrimination peptide identification filtering. *Journal of Proteome Research*, *8*(8), 3872–3881. https://doi.org/10.1021/pr900360j

Mackmull, M., Klaus, B., Heinze, I., Chokkalingam, M., Beyer, A., Russell, R. B., Ori, A., & Beck, M. (2017). Landscape of nuclear transport receptor cargo specificity. *Molecular Systems Biology*, *13*(12), 962. https://doi.org/10.15252/msb.20177608

Maertens, G. N., Cook, N. J., Wang, W., Hare, S., Gupta, S. S., Öztop, I., Lee, K., Pye, V. E., Cosnefroy, O., Snijders, A. P., KewalRamani, V. N., Fassati, A., Engelman, A., & Cherepanov, P. (2014). Structural basis for nuclear import of splicing factors by human Transportin 3. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(7), 2728–2733. https://doi.org/10.1073/pnas.1320755111

Marfori, M., Mynott, A., Ellis, J. J., Mehdi, A. M., Saunders, N. F. W., Curmi, P. M., Forwood, J. K., Bodén, M., & Kobe, B. (2011). Molecular basis for specificity of nuclear import and prediction of nuclear localization. *Biochimica et Biophysica Acta*, *1813*(9), 1562–1577. https://doi.org/10.1016/j.bbamcr.2010.10.013

Marko, M., Vlassis, A., Guialis, A., & Leichter, M. (2012). Domains involved in TAF15 subcellular localisation: dependence on cell type and ongoing transcription. *Gene*, *506*(2), 331–338. https://doi.org/10.1016/j.gene.2012.06.088

Marles, J. A., Dahesh, S., Haynes, J., Andrews, B. J., & Davidson, A. R. (2004). Protein-protein interaction affinity plays a crucial role in controlling the Sho1p-mediated

signal transduction pathway in yeast. *Molecular Cell*, *14*(6), 813–823. https://doi.org/10.1016/j.molcel.2004.05.024

Marshall, K. S., Cohen, M. J., Fonseca, G. J., Todorovic, B., King, C. R., Yousef, A. F., Zhang, Z., & Mymryk, J. S. (2014). Identification and characterization of multiple conserved nuclear localization signals within adenovirus E1A. *Virology*, *454–455*, 206–214. https://doi.org/10.1016/j.virol.2014.02.020

Marshall, K. S., Zhang, Z., Curran, J., Derbyshire, S., & Mymryk, J. S. (2007). An improved genetic system for detection and analysis of protein nuclear import signals. *BMC Molecular Biology*, *8*, 6. https://doi.org/10.1186/1471-2199-8-6

Martin, W., & Koonin, E. V. (2006). Introns and the origin of nucleus-cytosol compartmentalization. *Nature*, *440*(7080), 41–45. https://doi.org/10.1038/nature04531

Mason, D. A., Stage, D. E., & Goldfarb, D. S. (2009). Evolution of the metazoan-specific importin α gene family. *Journal of Molecular Evolution*, *68*(4), 351–365. https://doi.org/10.1007/s00239-009-9215-8

Mast, F. D., Navare, A. T., van der Sloot, A. M., Coulombe-Huntington, J., Rout, M. P., Baliga, N. S., Kaushansky, A., Chait, B. T., Aderem, A., Rice, C. M., Sali, A., Tyers, M., & Aitchison, J. D. (2020). Crippling life support for SARS-CoV-2 and other viruses through synthetic lethality. *The Journal of Cell Biology*, *219*(10). https://doi.org/10.1083/jcb.202006159

Mateo, M., Reid, S. P., Leung, L. W., Basler, C. F., & Volchkov, V. E. (2010). Ebolavirus VP24 binding to karyopherins is required for inhibition of interferon signaling. *Journal of Virology*, *84*(2), 1169–1175. https://doi.org/10.1128/JVI.01372-09

Matreyek, K. A., Yucel, S. S., Li, X., & Engelman, A. (2013). Nucleoporin NUP153 phenylalanine-glycine motifs engage a common binding pocket within the HIV-1 capsid protein to mediate lentiviral infectivity. *PLoS Pathogens*, *9*(10), e1003693. https://doi.org/10.1371/journal.ppat.1003693

Mellacheruvu, D., Wright, Z., Couzens, A. L., Lambert, J.-P., St-Denis, N. A., Li, T., Miteva, Y. V, Hauri, S., Sardiu, M. E., Low, T. Y., Halim, V. A., Bagshaw, R. D., Hubner, N. C., Al-Hakim, A., Bouchard, A., Faubert, D., Fermin, D., Dunham, W. H., Goudreault, M., … Nesvizhskii, A. I. (2013). The CRAPome: a contaminant repository for affinity purification-mass spectrometry data. *Nature Methods*, *10*(8), 730–736. https://doi.org/10.1038/nmeth.2557

Meyer, B. E., Meinkoth, J. L., & Malim, M. H. (1996). Nuclear transport of human immunodeficiency virus type 1, visna virus, and equine infectious anemia virus Rev proteins: identification of a family of transferable nuclear export signals. *Journal of Virology*, *70*(4), 2350–2359.

Meyer, K., Kirchner, M., Uyar, B., Cheng, J.-Y., Russo, G., Hernandez-Miranda, L. R., Szymborska, A., Zauber, H., Rudolph, I.-M., Willnow, T. E., Akalin, A., Haucke, V., Gerhardt, H., Birchmeier, C., Kühn, R., Krauss, M., Diecke, S., Pascual, J. M., & Selbach, M. (2018). Mutations in Disordered Regions Can Cause Disease by Creating Dileucine Motifs. *Cell*, *175*(1), 239-253.e17. https://doi.org/10.1016/j.cell.2018.08.019

Meyer, K., & Selbach, M. (2020). Peptide-based Interaction Proteomics. *Molecular & Cellular Proteomics : MCP*, *19*(7), 1070–1075. https://doi.org/10.1074/mcp.R120.002034

Mi, T., Merlin, J. C., Deverasetty, S., Gryk, M. R., Bill, T. J., Brooks, A. W., Lee, L. Y., Rathnayake, V., Ross, C. A., Sargeant, D. P., Strong, C. L., Watts, P., Rajasekaran, S., & Schiller, M. R. (2012). Minimotif Miner 3.0: database expansion and significantly improved reduction of false-positive predictions from consensus sequences. *Nucleic Acids Research*, *40*(Database issue), D252-60. https://doi.org/10.1093/nar/gkr1189

Mishra, A., Sipma, W., Veenhoff, L. M., Van der Giessen, E., & Onck, P. R. (2019). The Effect of FG-Nup Phosphorylation on NPC Selectivity: A One-Bead-Per-Amino-Acid Molecular Dynamics Study. *International Journal of Molecular Sciences*, *20*(3). https://doi.org/10.3390/ijms20030596

Mitrousis, G., Olia, A. S., Walker-Kopp, N., & Cingolani, G. (2008). Molecular basis for the recognition of snurportin 1 by importin beta. *The Journal of Biological Chemistry*, *283*(12), 7877–7884. https://doi.org/10.1074/jbc.M709093200

Miyake, Y., Keusch, J. J., Decamps, L., Ho-Xuan, H., Iketani, S., Gut, H., Kutay, U., Helenius, A., & Yamauchi, Y. (2019). Influenza virus uses transportin 1 for vRNP debundling during cell entry. *Nature Microbiology*, *4*(4), 578–586. https://doi.org/10.1038/s41564-018-0332-2

Mohan, A., Oldfield, C. J., Radivojac, P., Vacic, V., Cortese, M. S., Dunker, A. K., & Uversky, V. N. (2006). Analysis of molecular recognition features (MoRFs). *Journal of Molecular Biology*, *362*(5), 1043–1059. https://doi.org/10.1016/j.jmb.2006.07.087

Mohana-Borges, R., Pacheco, A. B., Sousa, F. J., Foguel, D., Almeida, D. F., & Silva, J. L. (2000). LexA repressor forms stable dimers in solution. The role of specific dna in tightening protein-protein interactions. *The Journal of Biological Chemistry*, *275*(7), 4708–4712. https://doi.org/10.1074/jbc.275.7.4708

Mohr, D., Frey, S., Fischer, T., Güttler, T., & Görlich, D. (2009). Characterisation of the passive permeability barrier of nuclear pore complexes. *The EMBO Journal*, *28*(17), 2541–2553. https://doi.org/10.1038/emboj.2009.200

Montell, C., Fisher, E. F., Caruthers, M. H., & Berk, A. J. (1982). Resolving the functions of overlapping viral genes by site-specific mutagenesis at a mRNA splice site.

Nature, *295*(5848), 380–384. https://doi.org/10.1038/295380a0

Morikawa, K., Lange, C. M., Gouttenoire, J., Meylan, E., Brass, V., Penin, F., & Moradpour, D. (2011). Nonstructural protein 3-4A: the Swiss army knife of hepatitis C virus. *Journal of Viral Hepatitis*, *18*(5), 305–315. https://doi.org/10.1111/j.1365-2893.2011.01451.x

Morin, N., Delsert, C., & Klessig, D. F. (1989). Nuclear localization of the adenovirus DNA-binding protein: requirement for two signals and complementation during viral infection. *Molecular and Cellular Biology*, *9*(10), 4372–4380.

Moroianu, J., Blobel, G., & Radu, A. (1995). Previously identified protein of uncertain function is karyopherin alpha and together with karyopherin beta docks import substrate at nuclear pore complexes. *Proceedings of the National Academy of Sciences of the United States of America*, *92*(6), 2008–2011. https://doi.org/10.1073/pnas.92.6.2008

Mosca, R., Céol, A., Stein, A., Olivella, R., & Aloy, P. (2014). 3did: a catalog of domain-based interactions of known three-dimensional structure. *Nucleic Acids Research*, *42*(Database issue), D374-9. https://doi.org/10.1093/nar/gkt887

Muhlbauer, D., Dzieciolowski, J., Hardt, M., Hocke, A., Schierhorn, K. L., Mostafa, A., Muller, C., Wisskirchen, C., Herold, S., Wolff, T., Ziebuhr, J., & Pleschka, S. (2015). Influenza virus-induced caspase-dependent enlargement of nuclear pores promotes nuclear export of viral ribonucleoprotein complexes. *Journal of Virology*, *89*(11), 6009–6021. https://doi.org/10.1128/JVI.03531-14

Nabbi, A., & Riabowol, K. (2015). Rapid Isolation of Nuclei from Cells In Vitro. *Cold Spring Harbor Protocols*, *2015*(8), 769–772. https://doi.org/10.1101/pdb.prot083733

Nakada, R., Hirano, H., & Matsuura, Y. (2015). Structure of importin-alpha bound to a non-classical nuclear localization signal of the influenza A virus nucleoprotein. *Scientific Reports*, *5*, 15055. https://doi.org/10.1038/srep15055

Neduva, V., Linding, R., Su-Angrand, I., Stark, A., De Masi, F., Gibson, T. J., Lewis, J., Serrano, L., & Russell, R. B. (2005). Systematic discovery of new recognition peptides mediating protein interaction networks. *PLoS Biology*. https://doi.org/10.1371/journal.pbio.0030405

Neduva, V., & Russell, R. B. (2006). DILIMOT: discovery of linear motifs in proteins. *Nucleic Acids Research*, *34*(Web Server issue), W350-5. https://doi.org/10.1093/nar/gkl159

Nguyen Ba, A. N., Pogoutse, A., Provart, N., & Moses, A. M. (2009). NLStradamus: a simple Hidden Markov Model for nuclear localization signal prediction. *BMC Bioinformatics*, *10*, 202. https://doi.org/10.1186/1471-2105-10-202

Nguyen Ba, A. N., Strome, B., Hua, J. J., Desmond, J., Gagnon-Arsenault, I., Weiss, E. L., Landry, C. R., & Moses, A. M. (2014). Detecting functional divergence after gene duplication through evolutionary changes in posttranslational regulatory sequences. *PLoS Computational Biology*, *10*(12), e1003977. https://doi.org/10.1371/journal.pcbi.1003977

Nguyen, H. Q., Roy, J., Harink, B., Damle, N. P., Latorraca, N. R., Baxter, B. C., Brower, K., Longwell, S. A., Kortemme, T., Thorn, K. S., Cyert, M. S., & Fordyce, P. M. (2019). Quantitative mapping of protein-peptide affinity landscapes using spectrally encoded beads. *ELife*. https://doi.org/10.7554/eLife.40499

Nitahara-Kasahara, Y., Kamata, M., Yamamoto, T., Zhang, X., Miyamoto, Y., Muneta, K., Iijima, S., Yoneda, Y., Tsunetsugu-Yokota, Y., & Aida, Y. (2007). Novel nuclear import of Vpr promoted by importin alpha is crucial for human immunodeficiency virus type 1 replication in macrophages. *Journal of Virology*, *81*(10), 5284–5293. https://doi.org/10.1128/JVI.01928-06

O'Reilly, A. J., Dacks, J. B., & Field, M. C. (2011). Evolution of the karyopherin-β family of nucleocytoplasmic transport factors; ancient origins and continued specialization. *PloS One*, *6*(4), e19308. https://doi.org/10.1371/journal.pone.0019308

Ojala, P. M., Sodeik, B., Ebersold, M. W., Kutay, U., & Helenius, A. (2000). Herpes simplex virus type 1 entry into host cells: reconstitution of capsid binding and uncoating at the nuclear pore complex in vitro. *Molecular and Cellular Biology*, *20*(13), 4922–4931. https://doi.org/10.1128/mcb.20.13.4922-4931.2000

Onischenko, E., Tang, J. H., Andersen, K. R., Knockenhauer, K. E., Vallotton, P., Derrer, C. P., Kralt, A., Mugler, C. F., Chan, L. Y., Schwartz, T. U., & Weis, K. (2017). Natively Unfolded FG Repeats Stabilize the Structure of the Nuclear Pore Complex. *Cell*, *171*(4), 904-917.e19. https://doi.org/10.1016/j.cell.2017.09.033

Orchard, S., Ammari, M., Aranda, B., Breuza, L., Briganti, L., Broackes-Carter, F., Campbell, N. H., Chavali, G., Chen, C., del-Toro, N., Duesbury, M., Dumousseau, M., Galeota, E., Hinz, U., Iannuccelli, M., Jagannathan, S., Jimenez, R., Khadake, J., Lagreid, A., … Hermjakob, H. (2014). The MIntAct project--IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Research*, *42*(Database issue), D358-63. https://doi.org/10.1093/nar/gkt1115

Otsuka, S., Iwasaka, S., Yoneda, Y., Takeyasu, K., & Yoshimura, S. H. (2008). Individual binding pockets of importin-beta for FG-nucleoporins have different binding properties and different sensitivities to RanGTP. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(42), 16101–16106. https://doi.org/10.1073/pnas.0802647105

Oughtred, R., Stark, C., Breitkreutz, B.-J., Rust, J., Boucher, L., Chang, C., Kolas, N., O'Donnell, L., Leung, G., McAdam, R., Zhang, F., Dolma, S., Willems, A., Coulombe-Huntington, J., Chatr-Aryamontri, A., Dolinski, K., & Tyers, M. (2019).

The BioGRID interaction database: 2019 update. *Nucleic Acids Research*, *47*(D1), D529–D541. https://doi.org/10.1093/nar/gky1079

Palmeri, D., & Malim, M. H. (1999). Importin beta can mediate the nuclear import of an arginine-rich nuclear localization signal in the absence of importin alpha. *Molecular and Cellular Biology*, *19*(2), 1218–1225. https://doi.org/10.1128/mcb.19.2.1218

Pang, X., & Zhou, H.-X. (2014). Design rules for selective binding of nuclear localization signals to minor site of importin alpha. *PloS One*, *9*(3), e91025. https://doi.org/10.1371/journal.pone.0091025

Park, N., Skern, T., & Gustin, K. E. (2010). Specific cleavage of the nuclear pore complex protein Nup62 by a viral protease. *The Journal of Biological Chemistry*, *285*(37), 28796–28805. https://doi.org/10.1074/jbc.M110.143404

Patel, A. B., Greber, B. J., & Nogales, E. (2020). Recent insights into the structure of TFIID, its assembly, and its binding to core promoter. *Current Opinion in Structural Biology*, *61*, 17–24. https://doi.org/10.1016/j.sbi.2019.10.001

Pawson, T., & Nash, P. (2003). Assembly of cell regulatory systems through protein interaction domains. *Science (New York, N.Y.)*, *300*(5618), 445–452. https://doi.org/10.1126/science.1083653

Peng, K., Radivojac, P., Vucetic, S., Dunker, A. K., & Obradovic, Z. (2006). Length-dependent prediction of protein intrinsic disorder. *BMC Bioinformatics*, *7*, 208. https://doi.org/10.1186/1471-2105-7-208

Perez-Riverol, Y., Csordas, A., Bai, J., Bernal-Llinares, M., Hewapathirana, S., Kundu, D. J., Inuganti, A., Griss, J., Mayer, G., Eisenacher, M., Pérez, E., Uszkoreit, J., Pfeuffer, J., Sachsenberg, T., Yilmaz, S., Tiwary, S., Cox, J., Audain, E., Walzer, M., … Vizcaíno, J. A. (2019). The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Research*, *47*(D1), D442–D450. https://doi.org/10.1093/nar/gky1106

Peterson, T. A., Park, D., & Kann, M. G. (2013). A protein domain-centric approach for the comparative analysis of human and yeast phenotypically relevant mutations. *BMC Genomics*, *14 Suppl 3*(Suppl 3), S5. https://doi.org/10.1186/1471-2164-14-S3-S5

Phan, A. T., Kuryavyi, V., Darnell, J. C., Serganov, A., Majumdar, A., Ilin, S., Raslin, T., Polonskaia, A., Chen, C., Clain, D., Darnell, R. B., & Patel, D. J. (2011). Structure-function studies of FMRP RGG peptide recognition of an RNA duplex-quadruplex junction. *Nature Structural & Molecular Biology*, *18*(7), 796–804. https://doi.org/10.1038/nsmb.2064

Pied, N., & Wodrich, H. (2019). Imaging the adenovirus infection cycle. *FEBS Letters*, *593*(24), 3419–3448. https://doi.org/10.1002/1873-3468.13690

Poss, Z. C., Ebmeier, C. C., & Taatjes, D. J. (2013). The Mediator complex and transcription regulation. *Critical Reviews in Biochemistry and Molecular Biology*, *48*(6), 575–608. https://doi.org/10.3109/10409238.2013.840259

Préhaud, C., Wolff, N., Terrien, E., Lafage, M., Mégret, F., Babault, N., Cordier, F., Tan, G. S., Maitrepierre, E., Ménager, P., Chopy, D., Hoos, S., England, P., Delepierre, M., Schnell, M. J., Buc, H., & Lafon, M. (2010). Attenuation of rabies virulence: takeover by the cytoplasmic domain of its envelope protein. *Science Signaling*, *3*(105), ra5. https://doi.org/10.1126/scisignal.2000510

Prüfer, K., & Barsony, J. (2002). Retinoid X Receptor Dominates the Nuclear Import and Export of the Unliganded Vitamin D Receptor. *Molecular Endocrinology*, *16*(8), 1738–1751. https://doi.org/10.1210/me.2001-0345

Pumroy, R. A., & Cingolani, G. (2015). Diversification of importin-alpha isoforms in cellular trafficking and disease states. *The Biochemical Journal*, *466*(1), 13–28. https://doi.org/10.1042/BJ20141186

Pumroy, R. A., Ke, S., Hart, D. J., Zachariae, U., & Cingolani, G. (2015). Molecular determinants for nuclear import of influenza A PB2 by importin alpha isoforms 3 and 7. *Structure (London, England : 1993)*, *23*(2), 374–384. https://doi.org/10.1016/j.str.2014.11.015

Regier, J. L., Shen, F., & Triezenberg, S. J. (1993). Pattern of aromatic and hydrophobic amino acids critical for one of two subdomains of the VP16 transcriptional activator. *Proceedings of the National Academy of Sciences of the United States of America*, *90*(3), 883–887. https://doi.org/10.1073/pnas.90.3.883

Reid, S. P., Valmas, C., Martinez, O., Sanchez, F. M., & Basler, C. F. (2007). Ebola virus VP24 proteins inhibit the interaction of NPI-1 subfamily karyopherin alpha proteins with activated STAT1. *Journal of Virology*, *81*(24), 13469–13477. https://doi.org/10.1128/JVI.01097-07

Rhee, Y., Gurel, F., Gafni, Y., Dingwall, C., & Citovsky, V. (2000). A genetic system for detection of protein nuclear import and export. *Nature Biotechnology*, 433–437.

Ribbeck, K., Lipowsky, G., Kent, H. M., Stewart, M., & Görlich, D. (1998). NTF2 mediates nuclear import of Ran. *The EMBO Journal*, *17*(22), 6587–6598. https://doi.org/10.1093/emboj/17.22.6587

Robbins, J., Dilworth, S. M., Laskey, R. A., & Dingwall, C. (1991). Two interdependent basic domains in nucleoplasmin nuclear targeting sequence: identification of a class of bipartite nuclear targeting sequence. *Cell*, *64*(3), 615–623.

Robzyk, K., & Kassir, Y. (1992). A simple and highly efficient procedure for rescuing autonomous plasmids from yeast. *Nucleic Acids Research*, *20*(14), 3790. https://doi.org/10.1093/nar/20.14.3790

Romani, L., Whitfeld, M. J., Koroivueta, J., Kama, M., Wand, H., Tikoduadua, L., Tuicakau, M., Koroi, A., Andrews, R., Kaldor, J. M., & Steer, A. C. (2015). Mass drug administration for scabies control in a population with endemic disease. *New England Journal of Medicine*, *373*(24), 2305–2313. https://doi.org/10.1056/NEJMoa1500987

Romero, P., Obradovic, Z., Kissinger, C. R., Villafranca, J. E., Garner, E., Guilliot, S., & Dunker, A. K. (1998). Thousands of proteins likely to have long disordered regions. *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*, 437–448.

Roy, J., Li, H., Hogan, P. G., & Cyert, M. S. (2007). A conserved docking site modulates substrate affinity for calcineurin, signaling output, and in vivo function. *Molecular Cell*, *25*(6), 889–901. https://doi.org/10.1016/j.molcel.2007.02.014

Salama, M., Amitai, Z., Amir, N., Gottesman-Yekutieli, T., Sherbany, H., Drori, Y., Mendelson, E., Carmeli, Y., & Mandelboim, M. (2016). Outbreak of adenovirus type 55 infection in Israel. *Journal of Clinical Virology*, *78*, 31–35. https://doi.org/10.1016/j.jcv.2016.03.002

Schmidt, H. B., & Görlich, D. (2016). Transport Selectivity of Nuclear Pores, Phase Separation, and Membraneless Organelles. *Trends in Biochemical Sciences*, *41*(1), 46–61. https://doi.org/10.1016/j.tibs.2015.11.001

Schwartz, T. U. (2016). The Structure Inventory of the Nuclear Pore Complex. *Journal of Molecular Biology*, *428*(10 Pt A), 1986–2000. https://doi.org/10.1016/j.jmb.2016.03.015

Sen Gupta, P. S., Biswal, S., Panda, S. K., Ray, A. K., & Rana, M. K. (2020). Binding mechanism and structural insights into the identified protein target of COVID-19 and importin-α with in-vitro effective drug ivermectin. *Journal of Biomolecular Structure & Dynamics*, 1–10. https://doi.org/10.1080/07391102.2020.1839564

Sever, R., & Glass, C. K. (2013). Signaling by nuclear receptors. *Cold Spring Harbor Perspectives in Biology*, *5*(3), a016709–a016709. https://doi.org/10.1101/cshperspect.a016709

Sheridan, C. M., Heist, E. K., Beals, C. R., Crabtree, G. R., & Gardner, P. (2002). Protein kinase A negatively modulates the nuclear accumulation of NF-ATc1 by priming for subsequent phosphorylation by glycogen synthase kinase-3. *The Journal of Biological Chemistry*, *277*(50), 48664–48676. https://doi.org/10.1074/jbc.M207029200

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

Sigler, P. B. (1988). Transcriptional activation. Acid blobs and negative noodles. In *Nature* (Vol. 333, Issue 6170, pp. 210–212). https://doi.org/10.1038/333210a0

Simmonds, P., Aiewsakun, P., & Katzourakis, A. (2019). Prisoners of war - host adaptation and its constraints on virus evolution. *Nature Reviews. Microbiology*, *17*(5), 321–328. https://doi.org/10.1038/s41579-018-0120-2

Sionov, R. V., & Haupt, Y. (1999). The cellular response to p53: the decision between life and death. *Oncogene*, *18*(45), 6145–6157. https://doi.org/10.1038/sj.onc.1203130

Soniat, M., & Chook, Y. M. (2015). Nuclear localization signals for four distinct karyopherin-β nuclear import systems. *The Biochemical Journal*, *468*(3), 353–362. https://doi.org/10.1042/BJ20150368

Soniat, M., & Chook, Y. M. (2016). Karyopherin-beta2 Recognition of a PY-NLS Variant that Lacks the Proline-Tyrosine Motif. *Structure (London, England : 1993)*, *24*(10), 1802–1809. https://doi.org/10.1016/j.str.2016.07.018

Soniat, M., Sampathkumar, P., Collett, G., Gizzi, A. S., Banu, R. N., Bhosle, R. C., Chamala, S., Chowdhury, S., Fiser, A., Glenn, A. S., Hammonds, J., Hillerich, B., Khafizov, K., Love, J. D., Matikainen, B., Seidel, R. D., Toro, R., Rajesh Kumar, P., Bonanno, J. B., … Almo, S. C. (2013). Crystal structure of human Karyopherin β2 bound to the PY-NLS of Saccharomyces cerevisiae Nab2. *Journal of Structural and Functional Genomics*, *14*(2), 31–35. https://doi.org/10.1007/s10969-013-9150-1

Soutoglou, E., Demény, M. A., Scheer, E., Fienga, G., Sassone-Corsi, P., & Tora, L. (2005). The nuclear import of TAF10 is regulated by one of its three histone fold domain-containing interaction partners. *Molecular and Cellular Biology*, *25*(10), 4092–4104. https://doi.org/10.1128/MCB.25.10.4092-4104.2005

Soutourina, J. (2018). Transcription regulation by the Mediator complex. *Nature Reviews. Molecular Cell Biology*, *19*(4), 262–274. https://doi.org/10.1038/nrm.2017.115

Staller, M. V, Holehouse, A. S., Swain-Lenz, D., Das, R. K., Pappu, R. V, & Cohen, B. A. (2018). A High-Throughput Mutational Scan of an Intrinsically Disordered Acidic Transcriptional Activation Domain. *Cell Systems*, *6*(4), 444-455.e6. https://doi.org/10.1016/j.cels.2018.01.015

Stangler, T., Tran, T., Hoffmann, S., Schmidt, H., Jonas, E., & Willbold, D. (2007). Competitive displacement of full-length HIV-1 Nef from the Hck SH3 domain by a high-affinity artificial peptide. *Biological Chemistry*, *388*(6), 611–615. https://doi.org/10.1515/BC.2007.075

Stanley, B. J., Ehrlich, E. S., Short, L., Yu, Y., Xiao, Z., Yu, X.-F., & Xiong, Y. (2008). Structural insight into the human immunodeficiency virus Vif SOCS box and its role in human E3 ubiquitin ligase assembly. *Journal of Virology*, *82*(17), 8656–8663.

https://doi.org/10.1128/JVI.00767-08

Stein, A., & Aloy, P. (2008). Contextual specificity in peptide-mediated protein interactions. *PloS One*, *3*(7), e2524. https://doi.org/10.1371/journal.pone.0002524

Stewart, M. (2007). Molecular mechanism of the nuclear protein import cycle. *Nature Reviews. Molecular Cell Biology*, *8*(3), 195–208. https://doi.org/10.1038/nrm2114

Strasser, A., Dickmanns, A., Lührmann, R., & Ficner, R. (2005). Structural basis for m3G-cap-mediated nuclear import of spliceosomal UsnRNPs by snurportin1. *The EMBO Journal*, *24*(13), 2235–2243. https://doi.org/10.1038/sj.emboj.7600701

Strycharz, J. P., Yoon, K. S., & Clark, J. M. (2008). A new ivermectin formulation topically kills permethrin-resistant human head lice (Anoplura: Pediculidae). *Journal of Medical Entomology*, *45*(1), 75–81.

Stumpf, M. P. H., Thorne, T., de Silva, E., Stewart, R., An, H. J., Lappe, M., & Wiuf, C. (2008). Estimating the size of the human interactome. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(19), 6959–6964. https://doi.org/10.1073/pnas.0708078105

Suter, B., Zhang, X., Pesce, C. G., Mendelsohn, A. R., Dinesh-Kumar, S. P., & Mao, J.-H. (2015). Next-Generation Sequencing for Binary Protein-Protein Interactions. *Frontiers in Genetics*, *6*, 346. https://doi.org/10.3389/fgene.2015.00346

Suzuki, K., Bose, P., Leong-Quong, R. Y., Fujita, D. J., & Riabowol, K. (2010). REAP: A two minute cell fractionation method. *BMC Research Notes*, *3*, 294. https://doi.org/10.1186/1756-0500-3-294

Tao, M., Kruhlak, M., Xia, S., Androphy, E., & Zheng, Z.-M. (2003). Signals that dictate nuclear localization of human papillomavirus type 16 oncoprotein E6 in living cells. *Journal of Virology*, *77*(24), 13232–13247.

Tay, M. Y. F., Fraser, J. E., Chan, W. K. K., Moreland, N. J., Rathore, A. P., Wang, C., Vasudevan, S. G., & Jans, D. A. (2013). Nuclear localization of dengue virus (DENV) 1-4 non-structural protein 5; protection against all 4 DENV serotypes by the inhibitor Ivermectin. *Antiviral Research*, *99*(3), 301–306. https://doi.org/10.1016/j.antiviral.2013.06.002

Teo, G., Liu, G., Zhang, J., Nesvizhskii, A. I., Gingras, A.-C., & Choi, H. (2014). SAINTexpress: improvements and additional features in Significance Analysis of INTeractome software. *Journal of Proteomics*, *100*, 37–43. https://doi.org/10.1016/j.jprot.2013.10.023

Terry, L. J., & Wente, S. R. (2009). Flexible gates: dynamic topologies and functions for FG nucleoporins in nucleocytoplasmic transport. *Eukaryotic Cell*, *8*(12), 1814–1827. https://doi.org/10.1128/EC.00225-09

Tessier, T. M., Dodge, M. J., Prusinkiewicz, M. A., & Mymryk, J. S. (2019). Viral Appropriation: Laying Claim to Host Nuclear Transport Machinery. *Cells*, *8*(6), E559. https://doi.org/10.3390/cells8060559

Tessier, T. M., MacNeil, K. M., & Mymryk, J. S. (2020). Piggybacking on Classical Import and Other Non-Classical Mechanisms of Nuclear Import Appear Highly Prevalent within the Human Proteome. *Biology*, *9*(8). https://doi.org/10.3390/biology9080188

Thomas, D. R., Lundberg, L., Pinkham, C., Shechter, S., DeBono, A., Baell, J., Wagstaff, K. M., Hick, C. A., Kehn-Hall, K., & Jans, D. A. (2018). Identification of novel antivirals inhibiting recognition of Venezuelan equine encephalitis virus capsid protein by the Importin alpha/beta1 heterodimer through high-throughput screening. *Antiviral Research*, *151*, 8–19. https://doi.org/10.1016/j.antiviral.2018.01.007

Thorley-Lawson, D. A. (2001). Epstein-Barr virus: exploiting the immune system. *Nature Reviews. Immunology*, *1*(1), 75–82. https://doi.org/10.1038/35095584

Thul, P. J., Akesson, L., Wiking, M., Mahdessian, D., Geladaki, A., Ait Blal, H., Alm, T., Asplund, A., Björk, L., Breckels, L. M., Bäckström, A., Danielsson, F., Fagerberg, L., Fall, J., Gatto, L., Gnann, C., Hober, S., Hjelmare, M., Johansson, F., … Lundberg, E. (2017). A subcellular map of the human proteome. *Science*, *356*(6340), eaal3321. https://doi.org/10.1126/science.aal3321

Timney, B. L., Raveh, B., Mironska, R., Trivedi, J. M., Kim, S. J., Russel, D., Wente, S. R., Sali, A., & Rout, M. P. (2016). Simple rules for passive diffusion through the nuclear pore complex. *The Journal of Cell Biology*, *215*(1), 57–76. https://doi.org/10.1083/jcb.201601004

Tompa, P., Davey, N. E., Gibson, T. J., & Babu, M. M. (2014). A million peptide motifs for the molecular biologist. *Molecular Cell*, *55*(2), 161–169. https://doi.org/10.1016/j.molcel.2014.05.032

Tompa, P., & Fuxreiter, M. (2008). Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions. *Trends in Biochemical Sciences*, *33*(1), 2–8. https://doi.org/10.1016/j.tibs.2007.10.003

Tompa, P., Fuxreiter, M., Oldfield, C. J., Simon, I., Dunker, A. K., & Uversky, V. N. (2009). Close encounters of the third kind: disordered domains and the interactions of proteins. *BioEssays : News and Reviews in Molecular, Cellular and Developmental Biology*, *31*(3), 328–335. https://doi.org/10.1002/bies.200800151

Trowitzsch, S., Viola, C., Scheer, E., Conic, S., Chavant, V., Fournier, M., Papai, G., Ebong, I.-O., Schaffitzel, C., Zou, J., Haffke, M., Rappsilber, J., Robinson, C. V, Schultz, P., Tora, L., & Berger, I. (2015). Cytoplasmic TAF2-TAF8-TAF10 complex provides evidence for nuclear holo-TFIID assembly from preformed submodules. *Nature Communications*, *6*, 6011. https://doi.org/10.1038/ncomms7011

Truant, R., & Cullen, B. (1999). The arginine-rich domains present in human immunodeficiency virus type 1 Tat and Rev function as direct importin beta-dependent nuclear localization signals.pdf. *Molecular and Cellular Biology*, *19*(2), 1210–1217.

Tsai, K.-L., Tomomori-Sato, C., Sato, S., Conaway, R. C., Conaway, J. W., & Asturias, F. J. (2014). Subunit architecture and functional modular rearrangements of the transcriptional mediator complex. *Cell*, *157*(6), 1430–1444. https://doi.org/10.1016/j.cell.2014.05.015

Twyffels, L., Gueydan, C., & Kruys, V. (2014). Transportin-1 and Transportin-2: protein nuclear import and beyond. *FEBS Letters*, *588*(10), 1857–1868. https://doi.org/10.1016/j.febslet.2014.04.023

Uhlén, M., Fagerberg, L., Hallström, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, Å., Kampf, C., Sjöstedt, E., Asplund, A., Olsson, I., Edlund, K., Lundberg, E., Navani, S., Szigyarto, C. A.-K., Odeberg, J., Djureinovic, D., Takanen, J. O., Hober, S., … Pontén, F. (2015). Proteomics. Tissue-based map of the human proteome. *Science*, *347*(6220), 1260419. https://doi.org/10.1126/science.1260419

UniProt Consortium. (2021). UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Research*, *49*(D1), D480–D489. https://doi.org/10.1093/nar/gkaa1100

Uversky, V. N. (2016). p53 Proteoforms and Intrinsic Disorder: An Illustration of the Protein Structure-Function Continuum Concept. *International Journal of Molecular Sciences*, *17*(11). https://doi.org/10.3390/ijms17111874

Uyar, B., Weatheritt, R. J., Dinkel, H., Davey, N. E., & Gibson, T. J. (2014). Proteome-wide analysis of human disease mutations in short linear motifs: Neglected players in cancer? *Molecular BioSystems*. https://doi.org/10.1039/c4mb00290c

van Breukelen, B., Brenkman, A. B., Holthuizen, P. E., & van der Vliet, P. C. (2003). Adenovirus type 5 DNA binding protein stimulates binding of DNA polymerase to the replication origin. *Journal of Virology*, *77*(2), 915–922.

Van Poelvoorde, L. A. E., Saelens, X., Thomas, I., & Roosens, N. H. (2020). Next-Generation Sequencing: An Eye-Opener for the Surveillance of Antiviral Resistance in Influenza. *Trends in Biotechnology*, *38*(4), 360–367. https://doi.org/10.1016/j.tibtech.2019.09.009

Van Roey, K., Dinkel, H., Weatheritt, R. J., Gibson, T. J., & Davey, N. E. (2013). The switches.ELM resource: a compendium of conditional regulatory interaction interfaces. *Science Signaling*, *6*(269), rs7. https://doi.org/10.1126/scisignal.2003345

Van Roey, K., Uyar, B., Weatheritt, R. J., Dinkel, H., Seiler, M., Budd, A., Gibson, T. J., & Davey, N. E. (2014). Short linear motifs: ubiquitous and functionally diverse protein interaction modules directing cell regulation. *Chemical Reviews*, *114*(13),

6733–6778. https://doi.org/10.1021/cr400585q

Varghese, F. S., Kaukinen, P., Glasker, S., Bespalov, M., Hanski, L., Wennerberg, K., Kummerer, B. M., & Ahola, T. (2016). Discovery of berberine, abamectin and ivermectin as antivirals against chikungunya and other alphaviruses. *Antiviral Research*, *126*, 117–124. https://doi.org/10.1016/j.antiviral.2015.12.012

Vetter, I. R., Arndt, A., Kutay, U., Görlich, D., & Wittinghofer, A. (1999). Structural view of the Ran-Importin beta interaction at 2.3 A resolution. *Cell*, *97*(5), 635–646. https://doi.org/10.1016/s0092-8674(00)80774-6

Via, A., Gould, C. M., Gemünd, C., Gibson, T. J., & Helmer-Citterich, M. (2009). A structure filter for the Eukaryotic Linear Motif Resource. *BMC Bioinformatics*, *10*, 351. https://doi.org/10.1186/1471-2105-10-351

Via, A., Uyar, B., Brun, C., & Zanzoni, A. (2014). How pathogens use linear motifs to perturb host cell networks. *Trends in Biochemical Sciences*, *40*(1), 36–48. https://doi.org/10.1016/j.tibs.2014.11.001

Victoria, J., & Trujillo, R. (2001). Topical ivermectin: A new successful treatment for scabies. *Pediatric Dermatology*, *18*(1), 63–65. https://doi.org/10.1046/j.1525-1470.2001.018001063.x

Vodicka, M. A., Koepp, D. M., Silver, P. A., & Emerman, M. (1998). HIV-1 Vpr interacts with the nuclear transport pathway to promote macrophage infection. *Genes & Development*, *12*(2), 175–185.

Vogel, C., Bashton, M., Kerrison, N. D., Chothia, C., & Teichmann, S. A. (2004). Structure, function and evolution of multidomain proteins. *Current Opinion in Structural Biology*, *14*(2), 208–216. https://doi.org/10.1016/j.sbi.2004.03.011

Vollmer, B., & Antonin, W. (2014). The diverse roles of the Nup93/Nic96 complex proteins - structural scaffolds of the nuclear pore complex with additional cellular functions. *Biological Chemistry*, *395*(5), 515–528. https://doi.org/10.1515/hsz-2013-0285

von der Haar, T. (2007). Optimized protein extraction for quantitative proteomics of yeasts. *PloS One*, *2*(10), e1078. https://doi.org/10.1371/journal.pone.0001078

Wagstaff, K. M., Rawlinson, S. M., Hearps, A. C., & Jans, D. A. (2011). An AlphaScreen(R)-based assay for high-throughput screening for specific inhibitors of nuclear import. *Journal of Biomolecular Screening*, *16*(2), 192–200. https://doi.org/10.1177/1087057110390360

Wagstaff, K. M., Sivakumaran, H., Heaton, S. M., Harrich, D., & Jans, D. A. (2012). Ivermectin is a specific inhibitor of importin alpha/beta-mediated nuclear import able to inhibit replication of HIV-1 and dengue virus. *The Biochemical Journal*, *443*(3), 851–856. https://doi.org/10.1042/BJ20120150

Walsh, M. P., Seto, J., Jones, M. S., Chodosh, J., Xu, W., & Seto, D. (2010). Computational analysis identifies human adenovirus type 55 as a re-emergent acute respiratory disease pathogen. *Journal of Clinical Microbiology*, *48*(3), 991–993. https://doi.org/10.1128/JCM.01694-09

Wang, L., Li, M., Cai, M., Xing, J., Wang, S., & Zheng, C. (2012). A PY-nuclear localization signal is required for nuclear accumulation of HCMV UL79 protein. *Medical Microbiology and Immunology*, *201*(3), 381–387. https://doi.org/10.1007/s00430-012-0243-4

Ward, J. J., Sodhi, J. S., McGuffin, L. J., Buxton, B. F., & Jones, D. T. (2004). Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *Journal of Molecular Biology*, *337*(3), 635–645. https://doi.org/10.1016/j.jmb.2004.02.002

Warfield, K. L., Schaaf, K. R., DeWald, L. E., Spurgers, K. B., Wang, W., Stavale, E., Mendenhall, M., Shilts, M. H., Stockwell, T. B., Barnard, D. L., Ramstedt, U., & Das, S. R. (2019). Lack of selective resistance of influenza A virus in presence of host-targeted antiviral, UV-4B. *Scientific Reports*, *9*(1), 7484. https://doi.org/10.1038/s41598-019-43030-y

Watters, K., Inankur, B., Gardiner, J. C., Warrick, J., Sherer, N. M., Yin, J., & Palmenberg, A. C. (2017). Differential Disruption of Nucleocytoplasmic Trafficking Pathways by Rhinovirus 2A Proteases. *Journal of Virology*, *91*(8). https://doi.org/10.1128/JVI.02472-16

Welcker, M., & Clurman, B. E. (2005). The SV40 large T antigen contains a decoy phosphodegron that mediates its interactions with Fbw7/hCdc4. *The Journal of Biological Chemistry*, *280*(9), 7654–7658. https://doi.org/10.1074/jbc.M413377200

Wen, W., Meinkoth, J. L., Tsien, R. Y., & Taylor, S. S. (1995). Identification of a signal for rapid export of proteins from the nucleus. *Cell*, *82*(3), 463–473. https://doi.org/10.1016/0092-8674(95)90435-2

Wente, S. R., & Rout, M. P. (2010). The nuclear pore complex and nuclear transport. In *Cold Spring Harbor perspectives in biology*. https://doi.org/10.1101/cshperspect.a000562

Williams, T., Ngo, L. H., & Wickramasinghe, V. O. (2018). Nuclear export of RNA: Different sizes, shapes and functions. *Seminars in Cell & Developmental Biology*, *75*, 70–77. https://doi.org/10.1016/j.semcdb.2017.08.054

Winberg, G., & Shenk, T. (1984). Dissection of overlapping functions within the adenovirus type 5 E1A gene. *The EMBO Journal*, *3*(8), 1907–1912.

Wo, Y., Lu, Q. Bin, Huang, D. D., Li, X. K., Guo, C. T., Wang, H. Y., Zhang, X. A., Liu, W., & Cao, W. C. (2015). Epidemical features of HAdV-3 and HAdV-7 in pediatric pneumonia in Chongqing, China. *Archives of Virology*, *160*(3), 633–638.

https://doi.org/10.1007/s00705-014-2308-8

Wodrich, H., Cassany, A., D'Angelo, M. A., Guan, T., Nemerow, G., & Gerace, L. (2006). Adenovirus core protein pVII is translocated into the nucleus by multiple import receptor pathways. *Journal of Virology*, *80*(19), 9608–9618. https://doi.org/10.1128/JVI.00850-06

Wright, P E, & Dyson, H. J. (1999). Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *Journal of Molecular Biology*, *293*(2), 321–331. https://doi.org/10.1006/jmbi.1999.3110

Wright, Peter E, & Dyson, H. J. (2015). Intrinsically disordered proteins in cellular signalling and regulation. *Nature Reviews. Molecular Cell Biology*, *16*(1), 18–29. https://doi.org/10.1038/nrm3920

Xie, H., Vucetic, S., Iakoucheva, L. M., Oldfield, C. J., Dunker, A. K., Obradovic, Z., & Uversky, V. N. (2007). Functional anthology of intrinsic disorder. 3. Ligands, post-translational modifications, and diseases associated with intrinsically disordered proteins. *Journal of Proteome Research*, *6*(5), 1917–1932. https://doi.org/10.1021/pr060394e

Xu, D., Farmer, A., & Chook, Y. M. (2010). Recognition of nuclear targeting signals by Karyopherin-beta proteins. *Current Opinion in Structural Biology*, *20*(6), 782–790. https://doi.org/10.1016/j.sbi.2010.09.008

Xu, D., Grishin, N. V, & Chook, Y. M. (2012). NESdb: a database of NES-containing CRM1 cargoes. *Molecular Biology of the Cell*, *23*(18), 3673–3676. https://doi.org/10.1091/mbc.E12-01-0045

Xue, B., Dunker, A. K., & Uversky, V. N. (2012). Orderly order in protein intrinsic disorder distribution: disorder in 3500 proteomes from viruses and the three domains of life. *Journal of Biomolecular Structure & Dynamics*, *30*(2), 137–149. https://doi.org/10.1080/07391102.2012.675145

Yang, S., Atkinson, S., Fraser, J., Wang, C., Maher, B., Roman, N., Forwood, J., Wagstaff, K., Borg, N., & Jans, D. (2019). Novel Flavivirus Antiviral That Targets the Host Nuclear Transport Importin α/β1 Heterodimer. *Cells*, *8*(3), E281. https://doi.org/10.3390/cells8030281

Yang, S. N. Y., Atkinson, S. C., Wang, C., Lee, A., Bogoyevitch, M. A., Borg, N. A., & Jans, D. A. (2020). The broad spectrum antiviral ivermectin targets the host nuclear transport importin alpha/beta1 heterodimer. *Antiviral Research*, *177*, 104760. https://doi.org/10.1016/j.antiviral.2020.104760

Yarbrough, M. L., Mata, M. a, Sakthivel, R., & Fontoura, B. M. a. (2014). Viral subversion of nucleocytoplasmic trafficking. *Traffic (Copenhagen, Denmark)*, *15*(2), 127–140. https://doi.org/10.1111/tra.12137

Yates, A. D., Achuthan, P., Akanni, W., Allen, J., Allen, J., Alvarez-Jarreta, J., Amode, M. R., Armean, I. M., Azov, A. G., Bennett, R., Bhai, J., Billis, K., Boddu, S., Marugán, J. C., Cummins, C., Davidson, C., Dodiya, K., Fatima, R., Gall, A., … Flicek, P. (2020). Ensembl 2020. *Nucleic Acids Research*, *48*(D1), D682–D688. https://doi.org/10.1093/nar/gkz966

Ye, J., Chen, Z., Li, Y., Zhao, Z., He, W., Zohaib, A., Song, Y., Deng, C., Zhang, B., Chen, H., & Cao, S. (2017). Japanese Encephalitis Virus NS5 Inhibits Type I Interferon (IFN) Production by Blocking the Nuclear Translocation of IFN Regulatory Factor 3 and NF-kappaB. *Journal of Virology*, *91*(8). https://doi.org/10.1128/JVI.00039-17

Yousef, A. F., Brandl, C. J., & Mymryk, J. S. (2009). Requirements for E1A dependent transcription in the yeast Saccharomyces cerevisiae. *BMC Molecular Biology*, *10*, 32. https://doi.org/10.1186/1471-2199-10-32

Yu, H., Kim, P. M., Sprecher, E., Trifonov, V., & Gerstein, M. (2007). The importance of bottlenecks in protein networks: correlation with gene essentiality and expression dynamics. *PLoS Computational Biology*, *3*(4), e59. https://doi.org/10.1371/journal.pcbi.0030059

Yun, C. Y., Velazquez-Dones, A. L., Lyman, S. K., & Fu, X.-D. (2003). Phosphorylation-dependent and -independent nuclear import of RS domain-containing splicing factors and regulators. *The Journal of Biological Chemistry*, *278*(20), 18050–18055. https://doi.org/10.1074/jbc.M211714200

Zahn-Zabal, M., Michel, P.-A., Gateau, A., Nikitin, F., Schaeffer, M., Audot, E., Gaudet, P., Duek, P. D., Teixeira, D., Rech de Laval, V., Samarasinghe, K., Bairoch, A., & Lane, L. (2020). The neXtProt knowledgebase in 2020: data, tools and usability improvements. *Nucleic Acids Research*, *48*(D1), D328–D334. https://doi.org/10.1093/nar/gkz995

Zanzoni, A., Ausiello, G., Via, A., Gherardini, P. F., & Helmer-Citterich, M. (2007). Phospho3D: a database of three-dimensional structures of protein phosphorylation sites. *Nucleic Acids Research*, *35*(Database issue), D229-31. https://doi.org/10.1093/nar/gkl922

Zanzoni, A., Ribeiro, D. M., & Brun, C. (2019). Understanding protein multifunctionality: from short linear motifs to cellular functions. *Cellular and Molecular Life Sciences : CMLS*, *76*(22), 4407–4412. https://doi.org/10.1007/s00018-019-03273-4

Zhou, P., Yang, X. Lou, Wang, X. G., Hu, B., Zhang, L., Zhang, W., Si, H. R., Zhu, Y., Li, B., Huang, C. L., Chen, H. D., Chen, J., Luo, Y., Guo, H., Jiang, R. Di, Liu, M. Q., Chen, Y., Shen, X. R., Wang, X., … Shi, Z. L. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*, *579*(7798), 270–273. https://doi.org/10.1038/s41586-020-2012-7

Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A. H., Tanaseichuk, O., Benner, C., & Chanda, S. K. (2019). Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nature Communications*, *10*(1), 1523. https://doi.org/10.1038/s41467-019-09234-6

Zotenko, E., Mestre, J., O'Leary, D. P., & Przytycka, T. M. (2008). Why do hubs in the yeast protein interaction network tend to be essential: reexamining the connection between the network topology and essentiality. *PLoS Computational Biology*, *4*(8), e1000140. https://doi.org/10.1371/journal.pcbi.1000140

Curriculum Vitae
Tanner Tessier

Education

Microbiology and Immunology, PhD The University of Western Ontario Supervisor: Joseph Mymryk	Sept 2013 – Mar 2021
Honours Double Major Biochemistry and Medical Science The University of Western Ontario	Sept 2009 – Apr 2013

Academic Employment

Teaching Assistantship (x3) MICROIMM 3100A, University of Western Ontario Course coordinator: Drs. Carole Creuzenet and Stephen Barr	Sept – Dec, 2014-2017
Laboratory Manager/Ordering Laboratory: Joseph Mymryk	Sept 2016 – Dec 2020
Summer Research Assistant	May – Aug 2013

Publications

Dodge MD, MacNeil KM, **Tessier TM**, Weinberg JB, Mymryk JS. Emerging antiviral therapeutics for human adenovirus infection: Recent developments and novel strategies. Antiviral Res. 2021 Apr;188:105034

Ghasemi F*, **Tessier TM***, Gameiro SF*, Maciver AH, Cecchini MJ, Mymryk JS. High MHC-II expression in Epstein-Barr virus-associated gastric cancers suggests that tumor cells serve an important role in antigen presentation. Sci Rep. 2020 Sep 8;10(1):14786

King CR*, **Tessier TM***, Dodge MJ, Weinberg JB, Mymryk JS. Inhibition of Human Adenovirus Replication by the Importin α/β1 Nuclear Import Inhibitor Ivermectin. J Virol. 2020 Aug 31;94(18):e00710-20

Tessier TM, MacNeil KM, Mymryk JS. Piggybacking on Classical Import and Other Non-Classical Mechanisms of Nuclear Import Appear Highly Prevalent within the Human Proteome. Biology (Basel). 2020 Jul 23;9(8):188

Ghasemi F*, Gameiro SF*, **Tessier TM***, Maciver AH, Mymryk JS. High Levels of Class I Major Histocompatibility Complex mRNA Are Present in Epstein-Barr Virus-Associated Gastric Adenocarcinomas. Cells. 2020 Feb 21;9(2):499

Tessier TM, Dodge MJ, Prusinkiewicz MA, Mymryk JS. Viral Appropriation: Laying Claim to Host Nuclear Transport Machinery. Cells. 2019 Jun 8;8(6):559

Zhang A, **Tessier TM***, Galpin KJC, King CR, Gameiro SF, Anderson WW, Yousef AF, Qin WT, Li SSC, Mymryk JS. The Transcriptional Repressor BS69 is a Conserved Target of the E1A Proteins from Several Human Adenovirus Species. Viruses. 2018 Nov 22;10(12):662

King CR, Zhang A, **Tessier TM**, Gameiro SF, Mymryk JS. Hacking the Cell: Network Intrusion and Exploitation by Adenovirus E1A. mBio. 2018 May 1;9(3):e00390-18

King CR, Gameiro SF, **Tessier TM**, Zhang A, Mymryk JS. Mimicry of Cellular A Kinase-Anchoring Proteins Is a Conserved and Critical Function of E1A across Various Human Adenovirus Species. J Virol. 2018 Mar 28;92(8):e01902-17

King CR, **Tessier TM**, Mymryk JS. Color Me Infected: Painting Cellular Chromatin with a Viral Histone Mimic. Trends Microbiol. 2016 Oct;24(10):774-776

Presentations

2019, ICGEB DNA Tumor Virus Meeting – 50th Anniversary. Trieste, Italy. Poster title: "Investigating Conserved Region 2 of human Adenovirus E1A using proteomics and transcriptomic approaches"

2016, Nuclear Organization & Function. Cold Spring Harbor, NY. Poster title: "A Discovery Based Approach for Identifying Novel Nuclear Import Mechanisms".

2015, ICGEB DNA Tumor Virus Meeting. Trieste, Italy. Poster title: "A Discovery Based Approach for Identifying Novel Nuclear Import Mechanisms".

2015, American Society for Virology. University of Western Ontario. Poster title: "A Discovery Based Approach for Identifying Novel Nuclear Import Mechanisms".

2014, Infection and Immunity Research Forum. London Ontario. Poster title: "A Discovery Based Approach for Identifying Novel Nuclear Import Mechanisms".

2014, DNA Tumor Virus Meeting. University of Madison Wisconsin. Poster title: "A Discovery Based Approach for Identifying Novel Nuclear Import Mechanisms".

2014, Oncology Research Day. London Ontario. Poster title: "A Discovery Based Approach for Identifying Novel Nuclear Import Mechanisms".

2014, London Health Research Day. London Ontario. Poster title: "Coordinated Regulation of Cellular pRb, BS69, and UBC9 by the Adenovirus E1A protein".

2013, Infection and Immunity Research Forum. London Ontario. Poster title: " Coordinated Regulation of Cellular pRb, BS69, and UBC9 by the Adenovirus E1A protein"

Scholarships and Awards

2017 Ontario Graduate Scholarship
Awarded to graduate students in Ontario based on academic achievement and research merit
Value: $15,000 over one year

2016 RGE Murray Graduate Scholarship in Microbiology and Immunology
Awarded to full-time student based on academic achievement and research merit
Value: $10,000

2016, Dr. Frederick W. Luney Graduate Travel Award in Microbiology & Immunology
Awarded to full-time student based on academic achievement and research merit
Value: $2,000

2015, Dr. FW Luney Graduate Entrance Scholarship in Microbiology & Immunology.
Value: $3,000

Volunteer Activities

London Hospice   2011 – 2018
Volunteer Visitor
Provide in-home respite support for palliative clients

Infection and Immunity Research Forum organizing committee   2014 – 2017
University of Western Ontario

American Society for Virology   July 2015
University of Western Ontario
Event volunteer

Microbiology and Immunology social committee   2013 – 2016
University of Western Ontario