

UNIVERSITY OF PADOVA

DEPARTMENT OF MOLECULAR MEDICINE



Master's thesis

# **A comprehensive analysis of viral classical nuclear localization signals**

Supervisor: prof. Gualtiero Alvisi

Student: Nasim Akbari

Matricola:1220376

Academic year: 2021/2022

## TABLE OF CONTENTS

Acknowledgements.....	v
Abstract.....	1
1. Introduction.....	3
1.1. NUCLEAR TRANSPORT.....	3
1.1.1. Host nuclear transport system.....	3
1.1.2. Different classes of nuclear localization signals.....	4
1.1.3. IMPa structure.....	4
1.1.4. Extra regulating factors in nuclear import activity.....	5
1.1.5. Viral mechanisms targeting nucleocytoplasmic trafficking.....	5
1.1.6. Inhibitors of nuclear transport.....	6
1.2. <i>Polyomaviruses</i> .....	7
1.2.1. Infectious cycle of <i>polyomaviruses</i> based on study on SV40 virions.....	7
1.2.2. Large T Antigens of Polyomaviruses.....	7
1.2.3. Structure of Large T antigen.....	8
1.2.4. Cell transformation and tumorigenesis mediated by LTA.....	8
1.3. <i>Poxviruses</i> .....	9
1.3.1. DNA-sensing signal pathways in vaccinia-infected cells.....	10
1.3.2. Immune evasion strategies in vaccinia virus.....	11
1.4. Vaccinia N2 protein.....	12
1.5. Vaccinia A19 protein.....	14
1.6. previous results.....	15
1.7. Aim of the project.....	16
2. MATERIALS AND METHODS.....	20
2.1. MOLECULAR CLONING TECHNIQUES.....	20
2.1.1 Primer resuspension.....	20
2.1.2 Polymerase chain reaction for Gateway® Technology cloning.....	20
2.1.3 Agarose gel electrophoresis.....	21
2.1.4 Purification of PCR products.....	21
2.1.5 BP recombination reactions.....	21
2.1.6 LR recombination reactions.....	22
2.1.7 TOPO TA cloning.....	23
2.1.8 Bacterial transformation.....	23
2.1.9 Bacteria inoculation.....	24
2.1.10 Plasmid isolation.....	24
2.1.11 DNA quantification.....	24

2.1.12 Restriction enzyme (RE) digestion analysis.....	24
2.1.13 Site directed mutagenesis.....	25
2.1.14 Sequencing.....	25
2.1.15 Site-specific mutagenesis primers design.....	26
2.1.16 Glycerol stocks preparation.....	26
2.2 Cell culture.....	26
2.2.1 Cell lines, media and maintenance.....	26
2.2.2 Cell freezing.....	27
2.2.3 Cell Thawing.....	27
2.3 Confocal laser scanning microscopy.....	27
2.3.1 Coverslips treatment with poly-L.....	27
2.3.2 Cellular Transfection with Lipofectamine 2000.....	27
2.3.3 Cellular ATP depletion.....	28
2.3.4. Confocal Laser Scanning Microscopy.....	28
2.3.5 Image analysis.....	28
2.4. Buffers and media.....	29
2.4.1 No phenol red DMEM medium.....	29
2.4.2 Opti-MEM medium.....	29
2.4.3 Fluoromount-G® mounting medium.....	29
2.4.4 PHEM 4x.....	29
2.4.5 Paraformaldehyde 4%.....	29
3-Results.....	30
3.1. Viral proteins classification based on bioinformatic analysis.....	30
3.2. Large T antigen <i>polyomaviruses</i> .....	30
3.2.1. Comparative bioinformatic analysis identified putative cNLSs on <i>Polyomaviridae</i> LTA.	30
3.2.2. cNLSs on LTAs from different HPyVs are heterogeneously distributed.....	31
3.2.3. Nuclear subcellular localization of newly identified cNLSs on all human <i>Polyomaviruses</i> LTA.....	31
3.2.4. Large T antigens in STL, KI and WUPyV have a potential bipartite cNLS.....	32
3.2.5. HPyV7 LTA accumulates into the nucleus in a concentration dependent manner and contains a functional cNLS, which its activity is modulated by upstream residues.....	33
3.2.6. Nuclear accumulation of HPyV LTAs is dependent on the IMP / heterodimer.....	33
3.2.7. cNLS in MWPy and MCPyV LTA works as a bipartite.....	34
3.3. Twenty-six previously undescribed viral cNLSs mediate active transport of GFP to the cell nucleus.....	35
3.4. A functional bipartite cNLS is conserved across Poxviridae N2 proteins.....	36
3.5. The bipartite cNLS of Vaccinia virus N2 is responsible for IMP / mediated nuclear import.....	37

3.6 A19 is actively transported into the nucleus relying on IMP $\alpha$ / $\beta$ 1 complex.....	38
4.Discussion.....	39
5- Figures and Legends.....	47
Figure 5.1.....	48
Figure 5.2.....	49
Figure 5.3.....	50
Figure 5.4.....	51
Figure 5.5.....	52
Figure 5.6.....	53
Figure 5.7.....	54
Figure 5.8.....	55
Figure 5.9.....	56
Figure 5.10.....	57
Figure 5.11.....	58
Figure 5.12.....	59
Figure 5.13.....	60
Figure 5.14.....	61
Figure 5.15.....	62
Legends.....	63
6-Appendix.....	68
Reference.....	84

## Acknowledgements

I would like to express my sincere gratitude to my supervisor, professor Gualtiero Alvisi for his consistent support and guidance during the running of this project. Thank you for your patience and your encouraging words and thoughtful, detailed feedback.

Thanks to the University of Padova for providing me with the financial means to complete this project.

I cannot forget to thank my parents for the unconditional support and love in this very intense academic year

## Abstract

Several viral proteins are translocated in the host cell nucleus during the infection to accomplish several different tasks. The latter include genome nuclear transport, expression, and replication, as well as manipulating host functions, such as immune response, cell growth, and survival.

Active transport of proteins is dependent on the presence of nuclear localization signals (NLSs) which are recognized by specific cellular transporters belonging to the karyopherin superfamily, called importins (IMPs). The first NLS was identified in Simian Virus 40 (SV40) large T antigen (LTA), as a highly basic amino acid stretch essential to confer nuclear localization. Similar sequences were subsequently identified in several other nuclear proteins. Such highly basic signals were dubbed classical (c)NLSs and are recognized by IMP $\beta$  1 through the IMP $\alpha$  adapter. cNLS can either be monopartite or bipartite. Monopartite cNLSs have a single cluster of basic amino acids, whereas bipartite cNLSs have two cluster of basic amino acids separated by 10-12 amino acids linker. Several other type of NLSs – known as a non-classical (nc)NLSs have been described. Such ncNLSs, are not necessarily basic and are directly recognized by IMP $\beta$  1 or one of its several homologues.

Recently, specific inhibitors of nuclear import pathways have been developed, with potential implications for anti-cancer and antiviral drug development. Among them, the best-characterized is Ivermectin, an FDA-approved broad-spectrum antiparasitic drug, which has also been shown to function as inhibitor of IMP $\alpha$   $\beta$  mediated nuclear import. Several recent studies have been shown the ability of Ivermectin to abrogate nuclear localization of viral proteins and therefore viral replication in several RNA and DNA families.

Therefore, identification of new viral nuclear proteins and characterization of their nuclear import pathway would not only enable a better understanding of virus-host interaction but also, pave the way to the development of new antivirals preventing the accumulation of viral protein in the nucleus.

Keeping this in mind, our lab previously performed a bioinformatic analysis to identify about 200 novel viral proteins potentially translocated into the nucleus of infected cells via IMP $\alpha$   $\beta$  pathway due to their putative cNLSs.

In this study, we began the functional validation of such hypothetical nuclear proteins with two different approaches. First, we focused on viral proteins which were already known to be nuclear in some species. but were not characterized as nuclear protein in their orthologous belonging to the same viral family.

This approach was used to study the nuclear import process of LTAs from all Human Polyomaviruses (HPyVs). By combining phylogenetic and bioinformatic analyses, we identified at least one putative

cNLS that could be responsible for the nuclear import of each LTA. The functionality of such cNLSs was investigated by quantitative confocal laser scanning microscopy (CLSM) of cells expressing GFP-NLS fusions under several conditions, including co-expression with the IMP $\alpha$   $\beta$  inhibitor Bimax2 and energy depletion. For selected LTAs, nuclear transport was further characterized by site-specific mutagenesis targeting key residues within or close to identified cNLSs in the context of full-length proteins. Our analyses revealed that most LTAs bear a strong monopartite cNLS, while Saint Louis (STL), KI, WU, and Malawi (MW) polyomaviruses are more likely to have a bipartite cNLS. In addition, Merkel cell polyomavirus (MCPyV) possesses two functional cNLSs, which can potentially synergize as a bipartite cNLS. The weakest cNLS was found in HPyV7, which indeed showed a variable localization pattern, with nuclear accumulation inversely proportional to protein expression levels.

Secondly, we functionally validated the 26 top ranking newly identified viral nuclear proteins based on their cNLS strength. Validation was performed by quantifying their nuclear targeting properties as GFP fusions in the presence or absence of IMP $\alpha$   $\beta$  inhibitors and ATP.

Interestingly, 11 of such cNLSs belong to members of the Poxviridae family. Intriguingly, three of them are orthologs of the A19 and four of the N2 Vaccinia virus proteins. Both proteins have been previously described as nuclear proteins, but their NLS and nuclear transport mechanism have not been characterized so far. Previous studies identified the Vaccinia N2 protein as an inhibitor of interferon regulatory factor (IRF3). By combining site specific mutagenesis and CLSM analysis, we were able to show that N2 protein and its orthologous have a bipartite NLS, which is sufficient and necessary for protein nuclear accumulation. We also characterized N2 nuclear import, showing its dependence on the IMP $\alpha$   $\beta$  heterodimer. A similar approach was followed for the identification of A19 NLS and characterization of its nuclear import pathway as a cNLS, energy, and IMP $\alpha$   $\beta$  dependent process. Our results thus implications for the understanding of the evolution of bipartite cNLSs, the development of new antivirals to interfere with the life cycle of *Polyomaviridae* and *Poxviridae* members and for the characterization of the nuclear role of A19 and N2 proteins.

# 1. Introduction

## 1.1. NUCLEAR TRANSPORT

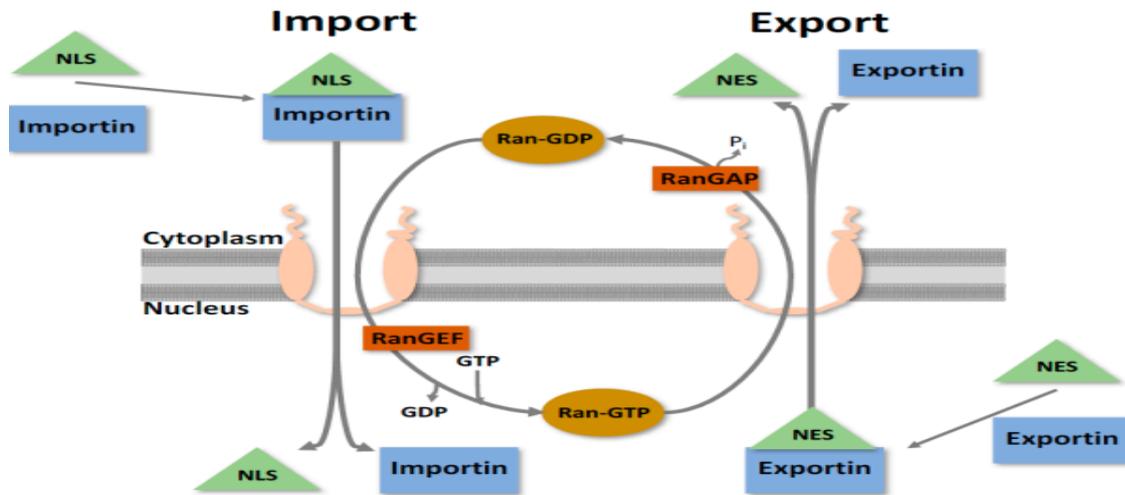
### 1.1.1. Host nuclear transport system

The nucleus is separated from the cytoplasm by the nuclear envelope (1). The nuclear envelope consists of a double layer and functions as a barrier to prevent the free passage of molecules between the nucleus and cytoplasm (2). However, some proteins and RNAs need to pass across the nuclear membrane and this transport is tightly regulated by nuclear pore complex (NPC,3). NPC is an essential function that lubricates the transport of wide range of proteins and various macromolecules into the nucleus. the NPC itself is a large structure, made up of about 30 different nucleoporins in multiple copies leading to its huge structure of 400 polypeptides (3).

Small molecules such as ions, metabolites, and proteins below 60 KDa passively diffused through the open channels intrinsic to the NPC, However, larger proteins and RNAs unable to pass through the open channels and they require the active transport. This active transport usually guided by specific amino acids sequence called nuclear localization signals (NLS). Which most of these sequenced are characterized by short stretches of basic amino acids (4).

NLS mediate the nuclear import of proteins by binding to their receptors known as importins, so nuclear import of proteins is initiated by the formation of a ternary complex with  $IMP\alpha$ ,  $IMP\beta$  and a cargo (5). Once NLS bound to its respective  $IMP\alpha$  subunit, the receptor-subunit complex binds to  $IMP\beta$  subunit and This complex is targeted to the nuclear pore by binding to NPC (2). When  $IMP\beta$  docks the complex to the NPC, release the cargo into the nucleus through the binding of Ran-GTP to the  $IMP\beta$  (5). This step is energy dependent; a high concentration of Ran-GDP is required on cytosolic side for formation of complex between  $IMP\alpha$ ,  $IMP\beta$  and cargo then this complex transport to the nucleus where there is a high concentration of Ran-GTP. Ran-GTP then binds to  $IMP\beta$ , releasing  $IMP\alpha$  and its substrate (1). The Ran-GTP complexed with IMP is immediately exported back into the cytoplasm (6). In the cytoplasm, GTP is hydrolyzed to GDP and releases  $IMP\beta$  which is then available (Figure 1.1;1).

**Figure1.1**



**Figure1.1. Nucleocytoplasmic protein transport.**

The NLS within the protein is recognized by importins. Ran-GDP binds to the complex then complex transported into the nucleus. Inside the nucleus, Ran-GTP complexes with Importin releasing the IMP subunit and the cargo protein. Ran-GTP hydrolyzed to Ran-GDP and IMPs are recycled to the cytoplasm

### 1.1.2. Different classes of nuclear localization signals

A recent study identified six classes of NLS that specifically bind to distinct binding pocket of IMP  $\alpha$  (7). The classical monopartite NLS (class 1 and 2) bind to the major pocket of IMP $\alpha$ , while class 3 and 4 which are known as a noncanonical NLS bind to minor binding pocket of IMP $\alpha$ . Two other classes including class 5 NLS, and bipartite NLS require the region outside the core basic residues for their activity (7). Monopartite NLSs have a single cluster of basic amino acids and bipartite NLSs have two cluster of basic amino acids which are separated by 10-12 amino acids linker (6). A putative consensus sequence of bipartite NLS has been defined as (K/R) (K/R) X<sub>10-12</sub>(K/R)<sub>3/5</sub>. Linker region has been found to be tolerant to amino acids conversion (8).

There are different kinds of NLSs according to their residue composition and IMPs recognizing them. The best characterized NLSs are the classic (c)NLSs that are highly basic sequences recognized by IMP $\alpha$   $\beta$ . another type of NLS is represented by Arginine-rich NLSs (R-rich NLSs) which are directly recognized by IMP $\beta$ . In contrast, PY-NLSs have diverse sequence and larger structure compared to cNLSs and are directly recognized by IMP $\beta$ 2 by multiple interactions between the weak NLS and IMP $\beta$ 2.

### 1.1.3. IMP $\alpha$ structure

IMP $\alpha$  is responsible for the initial recognition of the cargo by binding their NLSs. It is a 60 kDa protein consisting of two functional domains. In N-terminal region there is an importin beta binding domain (IBB) for binding to IMP $\beta$ . This part has at least 41 essential amino acids. The remaining of the protein is made up of

eight 43-residue repeated motifs. These repeats recognize the basic amino acids of NLS (10). Such repeated motifs form two binding sites, the major binding site binds to the classical monopartite NLS, whereas both (minor and major pockets) are required for binding to bipartite NLS (9). The IBB domain is rich in basic amino acids. Thanks to these basic residues in the IBB domain, when  $\text{IMP}\beta$  is absent, these basic amino acids fold inward and occupy the NLS binding sites of  $\text{IMP}\alpha$ . Because of this,  $\text{IMP}\alpha$  bound  $\text{IMP}\beta$  has a higher NLS affinity than  $\text{IMP}\alpha$  alone (11).

#### **1.1.4. Extra regulating factors in nuclear import activity**

Five years after the discovery of LTA SV40NLS, clusters of proximal phosphorylation sites were shown to modulate the rate of nuclear import. Protein kinase CK2-mediated phosphorylation Eleven residues upstream of LTA NLS enhanced the transport, whereas cdk1 phosphorylation adjacent to the NLS inhibited it (12). The role of phosphorylation around or within an NLS for regulating nuclear transport has been shown in number of regulatory proteins in yeast (12). It has been shown that acidic amino acids adjacent basic core of cNLSs inhibits NLS activity, and phosphorylation around NLS mimics the role of acidic amino acids to impair or decrease NLS activity (5). Proteins with cdk1-mediated phosphorylation sites adjacent to the NLS are controlled and imported to the nucleus dependent on the cdk1 phosphorylation, and these proteins with cdk1 sites close to the NLS are cell cycle-dependent nucleocytoplasmic shuttling (5). In bipartite NLS, the terminal of linker is more sensitive to cdk1 phosphorylation rather than the central part of the linker. If we substitute the serin or threonine of the cdk1 site with Alanine, we can see that the protein, which was nuclear just in the G0 phase of the cell cycle, became nuclear during the cell cycle (5).

#### **1.1.5. Viral mechanisms targeting nucleocytoplasmic trafficking**

The nucleus has crucial roles in different cellular processes, including cell survival. Several Viruses for replication and virulence target the nucleopore complex as well as import proteins (13). Many different viruses target import protein to prevent activation of the innate immune system, whereas some viruses benefit nuclear transport to transport viral mRNAs or proteins into the nucleus to complete their lifecycle (13).

For example, Poliovirus is a positive-stranded RNA virus that replicates exclusively in the cytoplasm, but still, it has some nuclear proteins which have roles in inhibiting the cellular process (14). For example, some host nuclear proteins contain NLS trapped in the cytoplasm during poliovirus infection, suggesting general inhibition of host nuclear trafficking (14). Therefore, it was hypothesized that these single-strand RNA viruses disrupt the nuclear import pathway to prevent the start of an antiviral response (13). As a result of the degradation of the nuclear pore complex during

poliovirus infection, the hnRNP C1/C2 protein, which has a role in host mRNA transcription processing, remains in the cytoplasm (15). Another example is *encephalomyocarditis* virus which encodes EMCV L protein. This protein is responsible for the hyperphosphorylation of nuclear pore proteins, eventually degrading nuclear import. Moreover, EMCV L also binds to Ran-GTPase, which is required to release and bind cargo to importins (16).

The severe acute respiratory syndrome coronavirus (SARS-COV) is another example of positive single-strand RNAs, utilizing different mechanisms to impair the nuclear transport system. The ORF6 protein binds to IMP $\alpha$  and IMP $\beta$  and conceals them. In consequence, preventing the translocation of STAT1 transcription factor into the nucleus, thus preventing an antiviral response (13).

Inhibition of nuclear import pathway is not limited to RNAs viruses. For example, the *herpes simplex virus* (HSV) is a double-stranded DNA virus that replicates in the infected cell's nucleus. HSV encodes the ICP27 protein, which directly interacts with the nuclear pore protein and blocks the nuclear import of IMP $\alpha/\beta$  (17).

### 1.1.6. Inhibitors of nuclear transport

Recently specific small molecules have been developed which target IMPs and therefore inhibit nuclear import. The simple fact that the inhibitors target transporters essential for cell function means that toxicity is an inevitable corollary of their use, limiting clinical application (18). The best characterized inhibitor of IMP $\alpha$  is Ivermectin, a compound that was FDA approved for parasitic infections such as river blindness in humans (19). Ivermectin was identified in 2011 in a high through-put screen, as able to inhibit nuclear import of HIV-1 Integrase (IN), but also of simian virus SV40 large tumor antigen and other IMP $\alpha$   $\beta$ -dependent nuclear import of specific viral proteins including Dengue virus (DENV) and related flaviviruses, influenza and Venezuelan equine encephalitis virus (18).

Importazole is another small molecule with anti-tumor potential and it was identified for the first time by in silico screening to target the overlapping binding sites on IMP $\beta$  1 for IMP $\alpha$  and RanGTP and it has been shown to inhibit IMP $\beta$  1- dependent nuclear import (20). Since the nuclear transport inhibitors target essential IMPs/EXPs and thereby general transport pathways of the cell, their use is likely to be completely linked with toxicity. By contrast, specific inhibitors to particular cargoes are great interest, since global effects on multiple cellular proteins-functions are spared in this scenario

ADDIN CSL\_CITATION {"citationItems":[{"id":"ITEM-1","itemData":{"id":"ITEM-1","issued":{"date-parts":[["0"]]},"title":"cargo inhibitors.pdf","type":"article"},"uris":["http://www.mendeley.com/documents/?uuid=c800ff68-

```
9a4f-40bc-b24c-d860bcff331a"]}], "mendeley": {"formattedCitation": "(21)", "manualFormatting": "{Formatting Citation}", "plainTextFormattedCitation": "(21)", "previouslyFormattedCitation": "(21)", "properties": {"noteIndex": 0}, "schema": "https://github.com/citation-style-language/schema/raw/master/csl-citation.json"} (22). Overall, targeting the host–pathogen interface to avoid issues of resistance and toxicity seems an attractive possibility.
```

## **1.2. Polyomaviruses**

Polyomaviruses are small DNA viruses with a circular double-stranded genome coding for 5–9 proteins. It has long been known that polyomaviruses, as the name indicates, can induce tumor formation (23). At present, 14 human polyomaviruses have been identified, which four of them have been linked to human diseases. BKV (BK polyomavirus), JCV (JC polyomavirus), TSV (Trichodysplasia spinulosa–associated polyomavirus), and MCV (Merkel cell polyomavirus). The Merkel cell polyomavirus (MCPyV) is the first virus of the polyomavirus family to be established as a causal factor of human cancer. Merkel cell carcinoma is a rare but very aggressive skin cancer with a high mortality rate (24). Except for avian viruses, polyomaviruses are asymptomatic, persisting for long time in their host. Nevertheless, unfortunately, we do not have efficient cell culture for in-vitro studying of different polyomaviruses. Therefore, SV40 has been used as a model for understanding polyomavirus infection and the host cell biology (24). Polyomaviruses genome can be divided into three parts, an early region that encodes the T antigens; a late region that encodes the capsid proteins and, in some viruses, auxiliary proteins; and a regulatory region that contains the origin of viral DNA replication (23).

### **1.2.1. Infectious cycle of polyomaviruses based on study on SV40 virions**

After virion-receptor interaction, virion enters the cell via the endoplasmic reticulum. Following translocation of viral chromatin in the nucleus, the early promoter recruits the cellular transcription factors and initiates the expression of the early gene. All polyomaviruses encode two common early proteins, large T antigen, and small T antigen protein, which are involved in viral genome expression and replication as well as in modulating host cell functions (25).

### **1.2.2. Large T Antigens of Polyomaviruses**

Large T antigens (LTA) are essential for viral replication, but they can initiate and maintain cell transformation and tumorigenesis in several different cells in human (24). LTA is one of the most important proteins to make an infective virion in polyomaviruses. It drives infected cells into the S phase and induces a DNA damage response to facilitate viral DNA replication. Since the DNA repair mechanism in infected cells, would activate the p53 pathway LTA binds and blocks p53, avoiding growth arrest and premature cell death. Accumulating data suggests that SV40 hijacks DDR enzymes rather than the normal cellular DNA replication machinery to replicate viral DNA (26). LTA then functions in viral DNA replication, capturing the host's replication machinery, and it binds to the viral origin of replication (ori) and assembles into double hexamer DNA helicase to unwind template dsDNA for replication. The complete assembly of this dodecamer requires binding to origin DNA and the presence of ATP or ADP (23). Finally, LTA activates the late viral promoter.

### **1.2.3. Structure of Large T antigen**

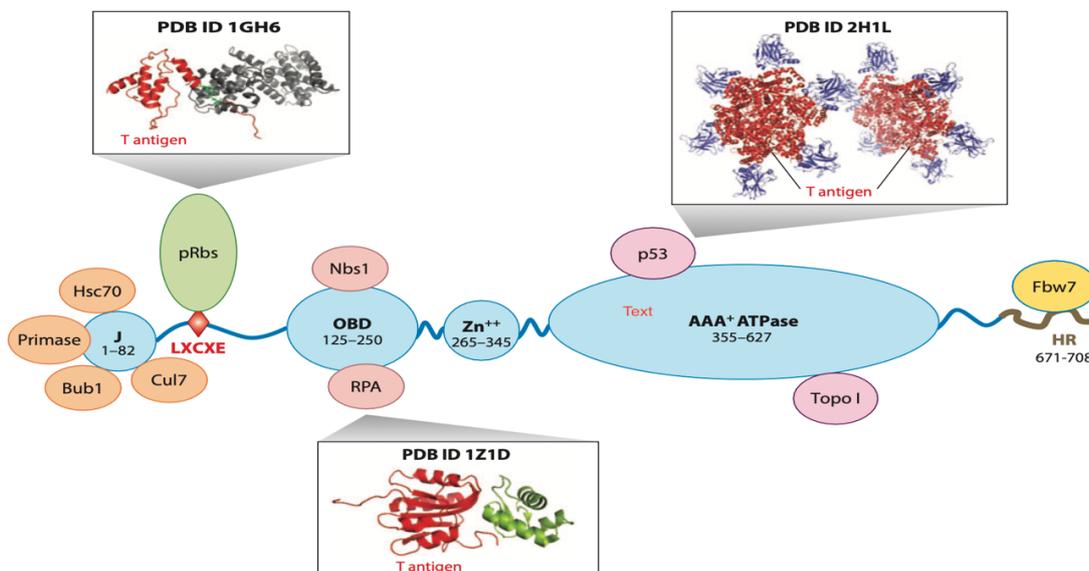
All polyomavirus LTAs contain four conserved domains: J domain, origin-binding domain (OBD), zinc (Zn)-binding domain, and ATPase domain (Figure 1.2; 18). The J domain is at the amino terminus of LTA, which has a chaperon activity. LTA binds to p130 RB via its LXCXE motif, and then energy derived from J domain-stimulated ATP hydrolysis by Hsc70 disrupts the complex of RB and EF2, then EF2 can be free. OBD is a domain that detects a specific sequence in ori that could be essential for initiating viral replication. OBD also can bind to several transcription factors. The Zn-binding domain is responsible primarily for forming LTA hexamers to simulate helicase activity. The ATPase domain provides the energy for helicase activity. LTAs have a flexible, partially non-structured region between the J domain and the OBD (43 amino acids in SV40 T antigen). This region harbors the NLS as well as several binding motifs for cellular proteins. The Rb proteins, the checkpoint kinase Bub1, and the cellular complex anchored by Cul7 associate with SV40 LTA through motifs in this region. This flexible region in different polyomavirus LTA is different, leading to different or extra motifs (19).

### **1.2.4. Cell transformation and tumorigenesis mediated by LTA**

LTA is necessary and sometimes sufficient to induce cell transformation and tumorigenesis in multiple cell types from several hosts and sources (19). This transforming activity blocks tumor suppressor proteins such as p53 and retinoblastoma (pRB). Additionally, LTA can bind to several other cellular factors, such as transcriptional co-activators p300 and CBP, which might contribute to its transformation function (21).

The RB protein family negatively controls cell proliferation via binding and inhibiting the E2F transcription factors. When we have growth positive stimulation, they lead to phosphorylate RB, and RB proteins cannot bind to the E2Fs Transcription factors anymore. However, E2F can bind to the promoters of E2F-dependent genes, whose expression is necessary for S-phase entry and progression (24). LTA preferentially binds to hypo phosphorylated pRB, resulting in sequestration of the form of pRB capable of binding to E2F. The interaction between RB protein and LTA usually occurs between the pocket domain in RB (residues 373–772 in pRB) and LxCxE motif (residues 103–107 in LTA). In addition, the N-terminal J domain of SV40 LTA has a role in pRB inactivation (20). To inactivate these tumor suppressor proteins, especially RB protein, LTA must enter the nucleus where RB proteins bind to E2F transcription factors. The nuclear import of most of the Polyomavirus's LTA is due to the small peptide sequences in the target proteins termed nuclear localization sequences (NLS) which are recognized and bound by IMP $\alpha$ , and in a complex with IMP $\beta$  the "cargo" protein will transfer to the nucleus (8).

**Figure 1.2**



**Figure 2.2** Domain structure and biological activities of SV40 large T antigen. SV40 large T antigen consists of four domains. J domain binds to Hsc70 and functions as its co-chaperone. J domain also interacts with DNA polymerase (Pol)  $\alpha$ . Downstream of J domain there is LXCXE motif that is important for the interaction between the T antigen and the pRb proteins. The OBD binds to the SV40 replication origin. The Zn-binding domain mediates T antigen hexamer. ATPase domain is essential during viral DNA replication which binds and hydrolyzes ATP to unwind template DNA.

### 1.3. Poxviruses

Poxviruses comprise a large group of double-stranded DNA (dsDNA) viruses that replicate exclusively in the cytoplasm of infected cells. Most poxvirus virions are typically a slightly flattened barrel with overall dimensions of 360 x 270 x 250 nm containing a single dsDNA genome of 130-380

kb with AT-rich covalently closed hairpins Poxviruses infect a wide variety of hosts, with the most famous member of the Variola virus, the causative agent of smallpox (27). Variola has two subtypes: variola major and Variola minor, with humans being the only natural host of Variola virus (28). Historically, these viruses were estimated to be responsible for 300 -500 million deaths in the twentieth century alone (29). The earliest method of smallpox prevention utilized a live Variola virus (VARV) to immunize the human population, but using the live virus posed a high risk for patients (30). Subsequent immunization campaigns used the closely related Vaccinia virus (VACV) as the smallpox vaccine (30). This immunization technique proved very effective and by 1979 the World Health Organization declared the global eradication of naturally occurring smallpox (29). Despite its eradication, the United States military continues to vaccinate soldiers in fear of VARV being released as a bioterrorism agent. Although smallpox is unlikely to be released back into the general population, other members of the poxvirus family still remain circulating in nature in many different hosts. These viruses raise the potential for zoonotic spread and the possibility of poxviruses jumping from their zoonotic hosts to the human population. Such cases have been reported with both monkeypox (MPXV) and cowpox (CPXV) (31). This is of importance as VARV, MPXV, CPXV, and VACV all belong to the same group of *Orthopoxviruses*.

### **1.3.1. DNA-sensing signal pathways in vaccinia-infected cells**

#### 1-cGAS

It is a universal cytoplasmic DNA sensor upstream of STING which recognizes many cytoplasmic DNA viruses (HSV-1, KSHV, and VACV) and retroviruses (HIV-1, HIV-2). Once CGAS binds to the DNA will be active and start catalyzing the production of 2'3'-cGAMP from ATP and GTP, resulting in the binding of cGMP to STING. Subsequently, STING recruits TBK1, which phosphorylates IRF. IRF3 then translocates to the nucleus and induces the IFN response (32).

#### 2-TLR9

TLR9 specifically recognizes the unmethylated CpG motif in dsDNA (CpG DNA), which is common in bacterial and viral genomes (33). TLR9 recruits the adaptor protein MyD88 subsequently tumor necrosis factor receptor-associated factor6 (TRAF6) and I $\kappa$ B kinase (IKK) complexes which leads to the activation of IRF7, then production of type I IFN, also activation of NF- $\kappa$ B (34).

#### 3-AIM2

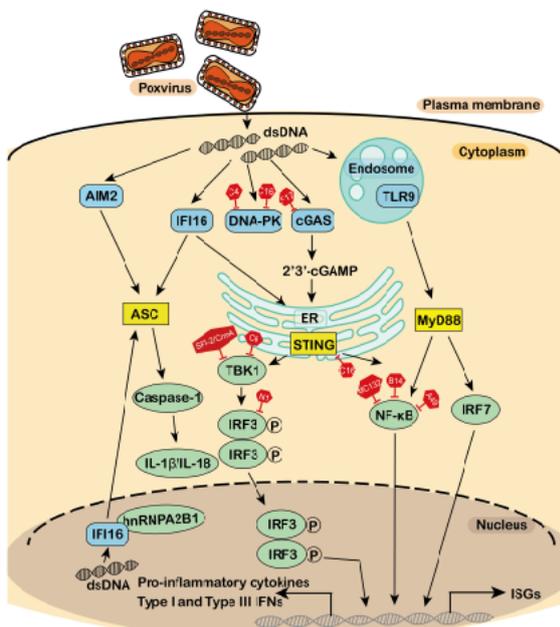
AIM2 senses viral DNA and can activate the inflammasome pathway which has a crucial role to clear the infected cells through pyroptosis. AIM2 binds to DNA through its HIN200 domain and recruits caspase-1 leading to the production of inflammatory cytokines, including IL-1 $\beta$  and IL-18 (32).

#### 4-IFI16

IFI16 is mainly located in the nucleus but can also shuttle between the cytoplasm and nucleus in different types of cells. IFI16 could bind to the dsDNA also can interact with STING to induce the TBK1-dependent IFN- $\beta$  response. Nuclear IFI16 in the cells expose to the viral DNA activates the inflammasome pathway through ASC and caspase-1 which leads to the production of IL-1 $\beta$  and IL-18 (35).

Both IFI16 and cGAS are crucial for activation of STING to promote its phosphorylation and translocation.

**Figure 3.3**



**Figure 4.3** summary of the mechanism of action of different DNA sensors during viral infection.

the cytosolic DNA sensor activates the adaptor, which in turn activates a series of downstream effectors to produce interferons, cytokines, and interleukins for an antiviral immune response.

### 1.3.2. Immune evasion strategies in vaccinia virus

Vaccinia virus (VACV) is a prototype member of the *Orthopoxvirus* genus of the *Poxviridae* family and has been used as a live vaccine for smallpox eradication (32). Interest in VACV persists because it is an excellent model for studying host-pathogen interactions and cell biology

In VACV, 33-50% of the genome is dedicated to virulence factors that modulate the host's immune response (36). The majority of these proteins are expressed early in infection to prevent the antiviral

response of host innate immune system (36).

NF $\kappa$ B is a crucial transcription factor that controls the cell's fate, and the NF $\kappa$ B pathway is one of the critical targets for viruses. For example, the varicella-zoster virus (VZV) blocks the NF $\kappa$ B pathway to inhibit antiviral response (37). VACV also encodes multiple proteins that inhibit NF $\kappa$ B activation. These proteins can halt interferon binding, complement, and chemokines on the cell surface, or some intracellular proteins can block the activation of pro-inflammatory signaling or apoptosis (43). A46 is one of these proteins which binds to adapter molecules associated with TLRs and IL-1R that prevent the activation of MAP kinases in signaling cascade (38). B14 is another essential virulence factor encoded with the VACV, and has a role in inhibiting the NF $\kappa$ B pathway by binding to IKK $\beta$  and inhibiting the phosphorylation of I $\kappa$ B $\alpha$  (40). A49 with other strategy inhibit the activation of NF $\kappa$ B. A49 binds to BTrCP, an E3 ubiquitin ligase, so in the presence of A49 even if I $\kappa$ B $\alpha$  is phosphorylated, it cannot be degraded (41). Clearly, VACV employs several diverse strategies to nullify the NF $\kappa$ B pathway. Another VACV protein that has an essential role in the virulence of viruses is N1 protein, which forms homodimers to inhibit both apoptosis and NF $\kappa$ B activation. Different surfaces of the protein mediate these functions (39). For example, a hydrophobic groove on the surface of protein is responsible for anti-apoptotic activity, while anti-NF $\kappa$ B activity requires an intact dimer interface (39). This anti-NF $\kappa$ B activity of N1 protein has a crucial role in the virulence of virus more than its anti-apoptotic function (39).

#### 1.4. Vaccinia N2 protein

Each mammalian viruses have at least one mechanism against the innate immune system. This is evident in both RNA and DNA viruses, but particularly evident in large DNA viruses. Thanks to their greater coding capacity, they express many proteins that target immune response during the infection (42).

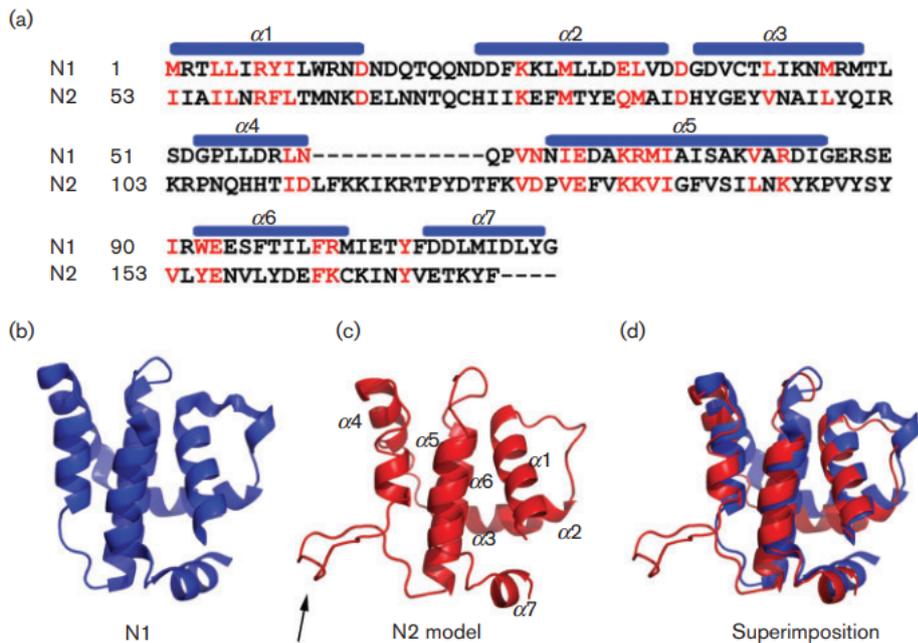
Interestingly, VACCV N1 protein belongs to BCL2 family, along with 9 more VACCV proteins (45). B14, A52, K7 have at least one binding partner in host cells and contribute to virulence (43). Another two VACV Bcl-2 proteins, C6 and A46 has been shown to contribute to virulence (45). N2, B22 and C1 are other predicted BCL2 proteins that are still uncharacterized.

N2 protein has 175aa with a predicted mass of 20.8 kDa. Yeast hybrid screen reported that N2 binds to IMP $\alpha$  so that N2 might be a member of the BCL2 family (46). Moreover, bioinformatic analysis of N2 protein showed that there are very highly conserved orthologues (95 % aa identity) of the VACV N2 protein in the majority of other sequenced Orthopoxviruses such as cowpox virus, variola virus, horsepox virus, and monkeypox virus (42). The

conservation of N2 in many *orthopoxviruses* suggest a pivotal role for N2 during the infection. Comparison of N2 sequences with other vaccinia proteins showed that there are non-poxvirus homologs to the N2 protein but is most closely related to the VACV N1 protein, which it shares 14% aa identity (42). Alignment of N1 and N2 aa sequenced showed that N2 has 52 aa extensions in N terminal, and internal insertion of 12 aa relative to N1. The majority of conserved residues are hydrophobic that have role to stabilize the alpha helix (37, Figure 1.4).

It was previously published that N2 protein is a nuclear protein that expresses early during the infection. The immunoblotting experiment showed that N2 was expressed 4 hours after infection even in the presence of cytosine arabinoside (42), and immunofluorescence experiment of infected cells by V N2 -tap in different post infection time points, showed a strong nuclear fluorescence for N2 -TAP (42). In addition, through luciferase reporter assays, N2 was shown to inhibit the activation of the IFN $\beta$  promoter (42). Luciferase reporter assays provided evidence that N2 may inhibit the activation of IRF3. It was supposed that might N2 prevents IRF3 phosphorylation or its translocation into the nucleus, however immunoblotting showed IRF3 phosphorylation was unaltered in the presence of N2 protein, and still IRF3 translocated into the nucleus during an infection in presence of N2 (42). By examining the virulence properties of the mutant viruses, which N2 open reading frame is deleted it was conclude that N2 is a virulence factor. Two animal models were tested an intranasal (i.n.) mouse model where they infected the mice at  $5 \times 10^3$  pfu and an intradermal (i.d.) model where the mice were injected with  $10^4$  pfu (42). in this study the N2L deletion mutant showed less signs of infection and less weight loss rather than control group. Along with virus was cleared mire rapidly from the infected tissue (42). Virulence in the i.d. model was determined by the size of the lesion and the time of lesion healing. The conclusions were consistent with the results obtained from the i.n. infection (42). The N2L deletion virus formed smaller lesions and less time took for healing the lesion. These data confirmed that N2 protein is a virulence factor that effects on outcome of infection. In summary N2 protein is an intracellular virulence factor that inhibits IRF3 activity in the nucleus (42). The goal of our experiment was to confirm N2 protein has a potential bipartite cNLS that helps N2 protein to actively transport to the nucleus, although it has a small molecular weight, and it can easily go to the nucleus by diffusion. Furthermore, based on bioinformatics analysis, we performed confocal scanning microscopy to confirm active functional bipartite cNLS in N2 and its orthologues as well in monkeypox, cowpox, and horsebox viruses. Accordingly, we hypothesized that N2 uses active transport to prevent the nuclear translocation of transcription factors which are essential for amounting effective innate immune responses. As we can see in many other viruses which use the ability of interference with host nuclear transport to promote virulence (47).

**Figure 1.4**



**Figure 1.4** A comparison of the N2 sequence with VACV N1 protein.

A model of the N2 tertiary structure based on N1 crystal structure showed the extra loop in N2 indicated with an arrow. That represents internal insertion of 12 aa relative to N1.

## 1.5. Vaccinia A19 protein

VACV A19 protein is a small protein (9-kDa) that is synthesized after viral DNA replication and is phosphorylated independently of the vaccinia virus F10 kinase (48). This protein is required for the late steps in morphogenesis, for the transition of spherical immature virions to barrel-shaped infectious mature virions (49). VACV morphogenesis is a complex process that remains to be fully elucidated. The first step in morphogenesis is the formation of a crescent-shaped membrane structure stabilized by trimers of the D13 protein, which forms a honeycomb lattice on the cytoplasmic side of the membrane. The crescents engulf core proteins and enlarge to become spherical immature virions

(IVs) containing the DNA genome. The subsequent transition to barrel-shaped infectious mature virions (MVs) involves the disruption of the D13 scaffold, proteolytic processing of certain membrane and core proteins, and the formation of intramolecular disulfide bonds (50). In the absence of A19, noninfectious particles will accumulate. But the specific role of A19 on the stage which is required for morphogenesis still is unclear.

A19 is conserved among *chordopoxviruses*, Amino acid sequences of A19 orthologs from each representative genus of *chordopoxviruses* were compared using the clustalW program and indicated conservation of a series of lysine and arginine, which could serve as a nuclear localization signal (Figure 1.5, 44).

VACV genome replication and post replicative transcription and translation occur within specialized compartments of the cytoplasm. Consistent with the cytoplasmic replication of VACV, there are few nuclear proteins that have vital role on virulence of virus, such as N2 protein (Figure 1.3). Immunofluorescence experiment on replication competent recombinant VACV with a HA tag, determined A19 was distributed in the nucleus and in cytoplasmic factories as well (50).

## 1.6. previous results

Nucleo-cytoplasmic transport of macromolecules is critical for both cell physiology and pathology. Consequently, investigating its regulation and disease-associated alterations can reveal novel therapeutic approaches to fight human diseases. To this end, the proteomes of all human viruses present on viral zone were downloaded from Uniprot along with individual proteins annotations. Sequences were scanned with three algorithms (Psort II, cNLS mapper and Deep Loc) to predict the localization of all these proteins. Based on their predicted subcellular localization and pathway used for nuclear import, proteins were classified in three main categories: Confirmed Nuclear, Putative Nuclear and Cytosolic proteins. Based on this study several other comparisons were done regarding the different requirements for nuclear import of different viral proteins. For example, the percentage of confirmed nuclear in cytoplasmic and nuclear replicating viruses were compared and 15% (532 proteins) of all viral proteins were confirmed nuclear, which 80% belonged to nuclear replicating viruses (corresponding to 40% of their proteome) and the remaining 20% belonged to cytoplasmic replicating ones (corresponding to 3% of their proteome). By considering all confirmed and putative viral nuclear proteins, nuclear proteins in nuclear and cytoplasmic replicating viruses increased up to 50 and 20%, respectively, suggesting that several nuclear proteins were misannotated on Uniprot (Figure 1.6). Another division to evaluate the results is according to Baltimore classes and the family. In Baltimore classes the highest percentages of nuclear proteins were found in II, VII classes where

the virus replicate in the nucleus. In family division we achieved the same results, the quantity of nuclear proteins is higher when the site of replication is the nucleus, and the percentage increased by considering all confirmed and putative viral nuclear proteins (Figure 1.7). Another interesting analysis was done comparing the proteome size with the quantity of nuclear proteins, and it was an inverse correlation between the proteome size and percentage of nuclear proteins. Major concentration of nuclear proteins in small proteome size compared to the larger, where bigger the virus, less the percentage of nuclear proteins. It can be noticed that the percentage of nuclear proteins increases if all confirmed and putative nuclear proteins are considered. In particular, large cytoplasmic replicating Poxviridae family exhibit an increase from nearly no nuclear protein to a 40% of nuclear proteins if all confirmed and hypothetical nuclear proteins are considered (Figure 1.8). The comparative analyses revealed that several proteins were wrongly annotated on Uniprot. Overall, such databases allowed us to identify nearly 200 new viral nuclear proteins from 67 different viruses.

### 1.7. Aim of the project

As mentioned above our research group previously performed bioinformatic analysis from uniprot in combined with other three algorithms (Psort II, cNLS mapper and Deep Loc) to predict the localization of all human viral proteins. This allowed identification of nuclear proteins which are classified as confirmed nuclear proteins ( $\alpha$   $\beta$  dependent proteins in some species) and hypothetical nuclear ( $\alpha$   $\beta$  dependent).

During my thesis I went further with hypothetical nuclear proteins which were ranked based on cNLS score of strongest cNLS present, selected putative cNLS and full length proteins were functionally analyzed by CLSM subcellular localization analysis, following functional signals were further characterized by subcellular localization assay in the absence of ATP to discriminate active transport from passive diffusion or intracellular binding, in the presence of Bimax2 inhibitor to confirm contribution of IMP  $\alpha$   $\beta$  dependent proteins for nuclear transportation. For selected cNLS and viral proteins, nuclear transport was further characterized by site-specific mutagenesis targeting key residues within or close to identified CNLS. This study can lead to develop Therapeutic agents that attempt to target nuclear viral protein localization such as ivermectin that target IMPs or other types of inhibitors which targeting host- pathogen interface such as *N*-(4-hydroxyphenyl) retinamide.

**Figure 1.5**

VACV	-----MDSTNVRSGMKSRRK-----KPKTTVIDDD-----DD	33
VARV	-----MDSTNACSGMKSRRK-----KPKTTVIDED-----D-	33
DPV	-----MEDTGGAKRIRKR-----KPKTTIQDND-----D-	31
YMTV	-----MDDMGGAKR-KKR-----KPKIAIQNNN-----D-	30
CRV	-----MDAGAKQKRRRKR-----ARTTVEDEATTASGTGR	30
SWPV	-----MDTAGSKR-KKR-----KPKTTIKDDD-----	29
ORFV	MSAVKAKAGAKGGSKGGVVDVAAGNKRRRR-----RVTTVVE-----DG	46
MOCV	-----MDVSGAKQRRRKR-----KPRTTVEEPPA-----DS	31
MYXV	-----MEDDAGVKRRKKR-----KPKAEVEDD-----	30
SPPV	-----MEDSAGNKRRKKR-----KPKTTVQDEE-----	31
FWPV	-----MADSTAGAKRKKRSTSATSTRKEPPTVIPEDE-----	41
	. . . . .	
VACV	CMTCSACQSKLVKISDITKVSLDYINT-MR----GNTLACAACGSSLKLLNDFAS---	77
VARV	CMTCSVCQSKLVKISDITKVSLDYINT-MR----GNTLACTACGSSLKLLNDFAS---	76
DPV	CVTCSSCYSKLIKVSDITKVSLDLYKVSQK----GNTLSACAACGSELRLNDFVS---	75
YMTV	CVTCSSCHSKLVKVS DITKISLDELKVAGK----GNVLTCSACGSELRLLSGFVS---	74
CRV	KTTCSLCDSKLLTFSSLSNSSFRKIQPCGGVAGAGTALRCSACNSVLTTLRPTQTPAV	88
SWPV	CMTCSSCYKLVKMSDITTVSLSQYKVVGK----GNTLSCSACGSELRTLNDVFH---	72
ORFV	DPVCSSCNSRLVSIKDVDRSLSTLSLGLACS-----STLSACAACGSALTPLRDLAR---	90
MOCV	CTTCSVCQSRLAFFSDVSKLQASPLMATPG---PDTLHCAACGSALCPLSEFAR---	78
MYXV	CVTCSSCYSKLVKVS DITKVSLHEYKVAGK----GNTLSACAACGSELRLNLFVN---	73
SPPV	CMICSSCYSKLVKVS DITKVSLNEYKVGK----GNILTCSACGSELRMLNLFVN---	74
FWPV	CTTCSICQSKLVMFSGVSKYKLSDYLNLTGKVF-TNSNIRCKACGSSLCHLRDLKSKS--	88
	* * * * *	

**Figure 1.5 ClustalW alignment of A19 chordopoxvirus orthologs.**

Amino acid sequences of A19 orthologs from each representative genus of Chordopoxvirinae: VACV, vaccinia virus Western Reserve; VARV, variola virus Bangladesh; DPV, deerpox virus strain W-1170-84; YMTV, Yaba monkey tumor virus Amano; CRV, crocodilepox virus Zimbabwe; SWPV, swinepox virus Nebraska; ORFV, Orf virus OV-SA00; MOCV, molluscum contagiosum virus subtype1; MYXV, myxoma virus Lausanne; SPPV, sheeppox virus strain NISKHI; FWPV, fowlpox virus Iowa. Asterisks represent completely conserved amino acids and highly conserved amino acids are denoted by dots.

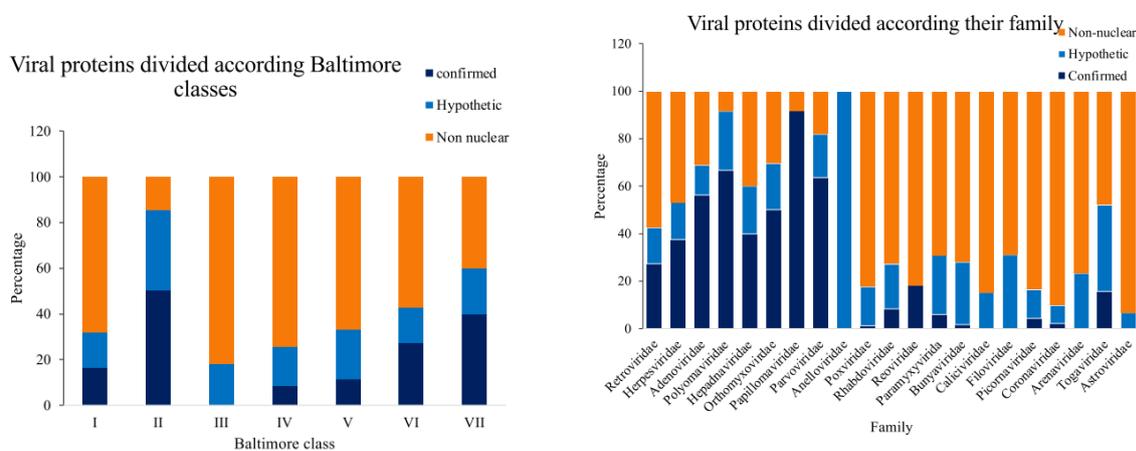
**Figure 1.6**

Division	Prot size	Confirmed	Confirmed+Hypothetic
<b>Viral Proteome</b>	3 690	14.01%	30.92%
<b>Human Proteome</b>	76 481	7.10%	10.03%
<b>Rep. Nucleus</b>	1104	40.49%	55.07%
<b>Rep. Cytoplasm</b>	2586	3.28%	19.33%

**Figure 1.6 Percentage of Nuclear Viral Proteins increases if all hypothetical and confirmed nuclear proteins are considered.**

Bioinformatic analyses highlighted that 14.01% of all viral proteins localizes in the host cell nucleus, of which 40.5% belong to nuclear replicating virus and 3.3% to cytoplasmic replicating ones. In all of these cases the addition of hypothetical nuclear proteins raises the percentage.

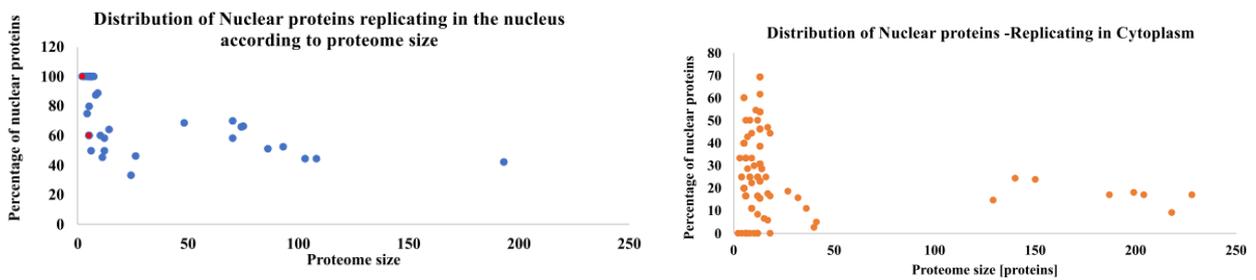
**Figure 1.7**



**Figure 1.7 Viral proteins divided according to Baltimore classes and their families.**

Another division to evaluate the results is according to Baltimore classes and the family. In Baltimore classes the highest percentages of nuclear proteins were found in II, VII classes where the virus replicate in the nucleus. In family division we achieved the same results, the quantity of nuclear proteins is higher when the site of replication is the nucleus and grows taking into account the hypothetical.

**Figure 1.8**



**Figure 1.8 Correlation between proteome size and percentage of nuclear proteins in first Baltimore class of viruses.** The percentage of nuclear proteins is inversely proportional to viral proteome size. Poxviruses confirmed nuclear proteins are almost zero, but the percentage increases if all hypothetical and confirmed nuclear proteins are considered.

## 2. MATERIALS AND METHODS

### 2.1. MOLECULAR CLONING TECHNIQUES

**2.1.1 Primer resuspension.** Lyophilized primers were centrifuged for 10 min at 4°C and 13'000 rpm, and subsequently resuspended in an appropriate volume of Tris-HCl pH7.8 10mM under a PCR cabinet, in order to obtain a final concentration of 100 µM. Stocks were subsequently vortexed and briefly centrifuged at max speed. For TOPO® cloning reactions (**section 2.1.7**), primer stocks were diluted to 1 µM final concentration in Tris-HCl pH7.8 10mM. For PCR amplification (**section 2.1.2**) prior to BP Gateway® reactions (**section 2.1.5**), primer stocks were diluted to 5 µM final concentration in Tris-HCl pH7.8 10 mM. All primer tubes were labeled and subsequently stored at -20°C until needed.

**2.1.2 Polymerase chain reaction for Gateway® Technology cloning.** Polymerase chain reaction (PCR) was used to amplify DNA fragments encoding for open reading frames (ORF) of interest, flanked by attB sites, to allow recombination in plasmid pDNR207 by Gateway BP reactions (**section 2.1.5**). Forward and reverse primers contained *attB* sequences (*attB1*: 5'-GGGGACAAGTTTGTACAAAAAAGCAGGCT-3', and *attB2*: 5'-GGGGACCACTTTGTACAAGAAAGCTGGG-3', respectively). The PCR reaction mix included 1x Gold Buffer (Applied Biosystem #4311806), 200 mM dNTPs mix, 2 mM MgCl<sub>2</sub>, 0.2 µM FWD primer, 0.2 µM REV primer, and 1.25U/50 ml rxn AmpliTaq Gold™ (Applied Biosystem #4311806), and 100 ng of template DNA.

Amplification was carried out following such steps in a thermal cycler:

- i) initial denaturation at 95°C for 5 min
- ii) 30 cycles of:
  - a. denaturation at 95°C for 30 sec
  - b. primer annealing at 56°C for 30 sec
  - c. primer extension at 72°C for 1min/Kb
- iii) 7 min extension at 72°C.

A list of primers and templates used for PCR *Gateway*<sup>®</sup> *Technology* cloning is indicated in the *Appendix (Table 6.6)*.

PCR products were separated from nonspecific amplification products by electrophoresis on a 2% agarose gel (**Section 2.1.3**). Specific PCR products were visualized using a UV transilluminator and excised using a razor blade before being gel purified using a GenElute<sup>™</sup> Extraction Kit (SIGMA #NA1111-1KT), according to the manufacturer recommendations (**Section 2.1.4**). Purified PCR products were quantified using Nanodrop spectrophotometer (ND-2000, ThermoFisher) (**Section 2.1.11**).

**2.1.3 Agarose gel electrophoresis.** DNAs molecules (Plasmids, PCR products and QuikChange reactions) were visualized after electrophoretic separation on agarose gels using an image station (UVITEC Cambridge Alliance). An appropriate volume of sample was mixed with loading buffer to a 1x final concentration [6% glycerol (v/v); 5mM Tris HCl pH 7.5], briefly centrifuged and loaded onto a TAE 0.5 x agarose gel, before being electrophoretically separated in TAE buffer 0.5x at 80 V for 30 minutes. EuroSafe Nucleic acid stain (EuroClone, #EMR440001) was added directly to the gel at a final concentration of 1x. Final agarose concentration ranged between 0.7% and 2 % depending on the size of DNA fragments to be separated.

**2.1.4 Purification of PCR products.** PCR fragments containing attB-flanked coding sequences of interest were purified by GenElute<sup>™</sup> Extraction Kit (SIGMA #NA1111-1KT). To this end, gel slices containing the DNA fragments of interest were solubilized by adding 3 volumes of *Gel solubilization* and incubation at 50-60°C until the gel was completely dissolved. A spin column was inserted into a provided *collection tube* and 0.5 ml of *preparation solution* was added to each spin column and centrifuged at 12'000 rpm for 1 min. Subsequently, 1 gel volume of 100 % isopropanol was added to the sample and mixed until homogenous. The mixture was transferred to the *binding column* and centrifuged at max speed for 1 min. The flow through was discarded, the column was washed with *wash solution* ethanol. Columns were further centrifuged for 2 min at max speed to completely remove undesired material from the silica membrane. Finally, DNA was eluted with 55  $\mu$ l of *elution solution*.

**2.1.5 BP recombination reactions.** Gateway<sup>®</sup> BP reactions were performed between an attB-flanked DNA fragment and an attP- containing DONOR vector (pDNR207), to generate an ENTRY clone.

BP recombination reactions were performed by adding the following reagents to a 1.5 ml microcentrifuge tube.

pDONR 207 (GW5; 75ng/ $\mu$ l)	1 $\mu$ l
AttB-containing fragment (75ng/ $\mu$ l)	1 $\mu$ l
TE	2 $\mu$ l

Subsequently BP clonase<sup>™</sup> enzyme was placed on ice, briefly vortexed, and 1  $\mu$ l was added to the reaction. Reactions were incubated at 25 °C overnight. The day after, reactions were inactivated by adding 0.5  $\mu$ l of proteinase K, followed by an incubation at 37 °C for 10 min.

ENTRY clones were transformed into chemically competent *E. coli* and transformants selected by plating on LB-agar plates containing Gentamycin (Table 6.4, 6.5; Section 2.1.4). For each BP reaction, three individual colonies were inoculated with a sterile loop in LB broth supplemented with Gentamycin (SIGMA ALDRICH, #G1272-10ML) (Section 2.1.9) and cultured overnight at 37 °C on an orbital shaker. Subsequently plasmid DNA was purified using a GenElute<sup>™</sup> Plasmid Miniprep (SIGMA, #PLN70) (Section 2.1.6). Integrity of plasmid DNAs was analyzed by restriction analysis to (Section 2.1.12) and their sequence verified by Sanger sequencing (Section 2.1.14).

**2.1.6 LR recombination reactions.** Gateway<sup>®</sup> LR reactions were performed between attL-containing ENTRY clones and the attR-containing DESTINATION vector pDESTnt-YFP (GW22) to generate EXPRESSION clones allowing expression of the gene of interest fused to the c-terminus of YFP. LR recombination reactions were performed by adding the following reagents to a 1.5 ml microcentrifuge tube.

pDESTnt-YFP (75ng/ $\mu$ l)	1 $\mu$ l
ENTRY clone (75ng/ $\mu$ l)	1 $\mu$ l
TE	2 $\mu$ l

Subsequently, LR clonase<sup>™</sup> enzyme was placed on ice, briefly vortexed, and 1  $\mu$ l was added to the reaction. Reactions were incubated at 25 °C overnight. The day after, reactions were inactivated by adding 0.5  $\mu$ l of proteinase K, followed by an incubation at 37 °C for 10 min.

LR reactions were transformed into chemically competent *E. coli DH5a* strain and transformants selected on LB agar plates containing Ampicillin (**Table 6.4, 6.5; Section 2.1.8**). For each LR reaction, three individual colonies were inoculated with a sterile loop in LB broth supplemented with Ampicillin antibiotic (SIGMA ALDRICH, #A9518-25G) (**Section 2.1.9**), and cultured overnight at 37°C on an orbital shaker. Plasmid DNAs were purified using a *GenElute™ Plasmid Miniprep* (SIGMA, #PLN70) (**Section 2.1.6**) and analyzed by restriction analysis (**Section 2.1.12**).

Plasmids generated with the Gateway® Technology are indicated in Appendix, in (**Table 6.8**).

### 2.1.7 TOPO TA cloning.

**2.1.7.1 Duplex formation.** DNA duplexes were formed by adding 1 µl of forward (FWD) primer 1 µM, 1 µl of reverse (REV) primer 1 µM (**Section 2.1.1**) to 122 µl of TE buffer in a to obtain a final primer concentration of 8.2 nM. The mixture subsequently incubated at 95°C for 5 min and then gradually cooled at RT. Duplexes were used for TOPO® reactions, and subsequently stored at -20°C.

**2.1.7.2 TOPO® reactions.** Cloning of cDNAs smaller than 80 kbps was performed by annealing appropriate oligo pairs to generate 3'-A overhangs into plasmid pcDNA3.1-NT-GFP-TOPO (ThermoFisher Scientific, #K4810-01). Reactions were performed by mixing 1 µl of duplex (**Section 2.1.7.1**), 1 µl of salt solution [diluted 1:2 in TE buffer (v/v)], and 1 µl of pcDNA3.1-NT-GFP-TOPO vector [diluted 1:3 in TE buffer (v/v)]. The TOPO reaction was incubated at room temperature for 30 min. Reactions were used to transform chemically competent *E. coli* cells as described in (**Section 2.1.8**) (**Table 6.5**). Colonies were picked and grown up in LB broth supplemented with the appropriate antibiotic (**Section 2.1.9**), then Plasmid DNA was isolated using a *GenElute™ Plasmid Miniprep* (SIGMA, #PLN70) (**Section 2.1.6**). Sequences were further verified by Sanger Sequencing (**Section 2.1.14**).

All vectors generated by TOPO® reaction are indicated in Appendix, in **Table 6.9**.

**2.1.8 Bacterial transformation.** Plasmid DNAs (either products of BP, LR, QuikChange or TOPO reactions) were transformed into chemically competent *E. coli* strains. Different *E. coli* strains were used according to the nature of transforming DNA (**Table 6.5**).

In all cases, an appropriate amount of plasmid DNA was added to chemically competent *E. coli* strains and incubated on ice for 30 min, subsequently bacteria were heat shocked for 45 sec at 42°C, and placed on ice for additional 2 min. Subsequently, 200-500 µl of S.O.C recovery media (ThermoFisher #15544034) were added to the cells. Bacteria were incubated at 37°C for 1 hour in a

horizontal shaker to allow expression of antibiotic resistance gene products. An aliquot of transformed bacteria was subsequently spread onto LB plates containing the appropriate antibiotic.

**2.1.9 Bacteria inoculation.** Single bacteria colonies from *E. Coli* transformation were picked with a sterile loop and incubated in 10 ml of LB containing and appropriate antibiotic at 37°C overnight using an orbital shaker (New Brunswick™ Excella® E25R Incubator Shaker, Eppendorf) at 200 rpm.

**2.1.10 Plasmid isolation.** Plasmid DNA was isolated from overnight bacterial cultures using a GenElute™ Plasmid Miniprep (SIGMA, #PLN70) as per the manufacturer's instructions. Briefly, 1 ml of bacterial cultures (**Section 2.1.9**) was stored at 4°C for glycerol stock preparation, and (**Section 2.1.16**). The remaining 9 ml was centrifuged for 10 min at 4'000 rpm at 4°C. The supernatant was discarded, the pellet resuspended in 200 µl of *Resuspension Solution* containing RNase A (100 µg/ml) and transferred to a 1.5 Eppendorf tube. Subsequently, 200 µl of *Lysis Solution* were added and tubes mixed until the solution did become transparent, indicating complete bacterial lysis, when 350 µl of *Neutralizing Solution* were added to each sample. Tubes were further inverted four times to promote precipitation of chromosomal DNA and debris were pelleted by centrifugation at 13'000 rpm for 10 min at 4°C. and the debris-clean lysate supernatant was transferred to a *Binding Column* containing the silica membrane which had been previously activated with 500 µl of *Colum Preparation Solution*. After a centrifugation at 13'000 prm for 1 min at 4°C, the flow-through was discarded and the column washed 2 times with 500 µl of *Washing Solutions* with different saline concentration. A void centrifuge of 1 min at max speed at 4°C was performed to ensure that undesired material was removed from the silica membrane. Plasmid DNAs were finally eluted by adding 55 µl of the *Elution Buffer* and centrifuging the tube at 13'000 rpm for 1 min at 4°C. The highly purified DNA was quantified using a spectrophotometer (**Section 2.1.11**).

**2.1.11 DNA quantification.** DNA concentration was determined by spectrophotometry using NanoDrop® ND-2000 Spectrophotometer (Nanodrop Technologies). The 260/280 nm and 260/230 nm ratios were also annotated. To this end 1.5 µl of each sample was placed on the spectrophotometer electrode, and absorbance at specific wavelengths recorded. before blanking with an appropriate buffer.

**2.1.12 Restriction enzyme (RE) digestion analysis.** Prior to sequencing, recombinant clones underwent a first screening through enzymatic digestion and electrophoretic separation on 1-2% Agarose gels, depending on the size of the DNA fragments. Based on the theoretical plasmid map with ApE software, specific restriction enzymes were selected for the enzymatic digestion. Reactions were performed in 20 µl reactions, which consisted of the appropriate buffer, 500 ng of plasmid

DNA, and 0.5-1  $\mu$ l of the appropriate restriction enzyme. Digestions was incubated at 37°C for 2 hrs. Subsequently, 4  $\mu$ l of Gel Loading Dye 6x were added to the reaction mix and DNA fragments separated by agarose gel electrophoresis (**Section 2.1.2**).

**2.1.13 Site directed mutagenesis.** Point mutations were inserted in plasmid DNA with QuikChange site-directed mutagenesis Kit (Stratagene, #210519<sup>b</sup>).

Mutagenesis reactions were prepared as indicated below:

Component	Volume per reaction ( $\mu$ l)	Concentration
Reaction buffer	5	10x
dsDNA template	1	25ng/ $\mu$ l
FORWARD primer	1	125ng/ $\mu$ l
REVERSE primer	1	125ng/ $\mu$ l
dNTPs mix	1	10mM
Quik solution reagent	1.5	
H <sub>2</sub> O MQ	to final volume 50 $\mu$ l	

Finally, 1  $\mu$ l of Quikchange lightning enzyme was added to each sample.

The PCR cycling parameters of each reaction are outlined in the table below:

Segment	Cycles	Temperature	Time
1	1	95°C	2 min
2	18	95°C	20 sec
		60°C	10 sec
		68°C	45 sec/kb
3	1	68°C	5 min

When the PCR reaction was completed, 2  $\mu$ l of Dpn1 (10U/ $\mu$ l) restriction enzyme was directly added to each amplification reaction, then immediately incubated at 37°C for 20 min to digest the parental supercoiled dsDNA.

Subsequently, 5  $\mu$ l of the Dpn1-treated DNA from each sample reaction was used to transform to 45  $\mu$ l of XL-10 Gold ultracompetent cells previously treated with  $\beta$ -mercaptoethanol (**Section 2.1.8**) (Table 6.7), while 20  $\mu$ l of amplification reaction was used for electrophoresis analysis (**Section 2.1.3**).

**2.1.14 Sequencing.** The nucleotide sequence of generated clones was determined by sanger sequencing method using appropriate primers and a private sequencing service (BMR Genomics S.r.l). To this end, 700 ng of each sample were loaded in a 0.5 microcentrifuge tube, briefly centrifuged, and the pellet dried at 70 °C for 30 min. Subsequently tubes were labeled according to the sequencing service company. Sequencing results were used to verify individual colonies using software a plasmid editor (APE; University of Utah). Bacterial cultures from which verified sequences were obtained, were finally used to prepare glycerol stocks (**Section 2.1.16**).

**2.1.15 Site-specific mutagenesis primers design.** Mutagenic oligonucleotide primers were designed considering the following criteria:

- a) Both of the mutagenic primers contained the desired mutation and they annealed to the same sequence on opposite strands of the plasmid
- b) Primer lengths did not exceed 45 bases in length
- c) Primers optimally should have a minimum GC content of 40%
- d) Following formulas was used to calculate the melting temperature

$$T_m = 81.5 + 0.41(\text{GC}\%) - (675/N) - \% \text{mismatches}$$

\*N is primer length

\* Primers need not be 5' phosphorylated

(The list of mutagenic primers is indicated in appendix, **Table 6.2**).

**2.1.16 Glycerol stocks preparation.** Bacterial cultures containing plasmids which had been confirmed by DNA sequencing or restriction analysis were used to prepare glycerol stocks. To this end, 1 ml of bacterial cultures were centrifuged for 10 min at 2000 rpm, the supernatant removed, and the pellet gently resuspended in 500 µl of Glycerol/LB (30% v/v). Samples were labelled and stored at -80°C.

## 2.2 Cell culture

**2.2.1 Cell lines, media and maintenance.** HEK 293A (ATCC® CRL-1573TM) were grown in Dulbecco's Modified Eagle's Medium (DMEM) which was supplemented (cpt DMEM) with 10% Fetal Bovine Serum (FBS) 1% L-glutamine, 1% non-essential amino acids, and Penicillin and Streptomycin (100U/ml; all reagents from Gibco), at 37°C and 5% CO<sub>2</sub> in a humidified incubator. Cells were passaged when reached >80% confluence. To this end, cells were briefly washed with PBS 1x to remove non-adherent cell and FBS that can inactivate trypsin. Subsequently, cells were

rinsed with warm 0.05% trypsin-EDTA for 3-5 min. Cells were subsequently resuspended in cpt DMEM and an appropriate volume of cells was transferred to a new flask containing required amount of cpt DMEM. Different volumes of PBS, trypsin and DMEM were used depending on the format of the flasks.

Cells were cultured in 6 and 13 ml of DMEM was added to T25 and T75 respectively. On the other hand, 1 ml of trypsin-EDTA solution was used to detach cells from T75 flasks, and 0.2 ml from T25 flasks.

**2.2.2 Cell freezing.** For long term storage, cells were stored in liquid nitrogen in cryovials (Corning® Cryogenic Vials with Orange Cap, #100-0091). Before being frozen, cells were washed in 1xPBS, trypsinized and resuspended in DMEM. Cells were subsequently centrifuged for 5 min at 700 rpm, the supernatant discarded, and the cell pellet was finally resuspended on ice in freezing medium [FBS/10% DMSO (v/v)]. Each 25cm<sup>2</sup> of cultured cells were resuspended in 1.8 ml of freezing medium and aliquoted in pre-chilled cryovials, and incubated 10 min on ice, 2 hrs at -20°C, and ON at -80°C before being transferred to liquid nitrogen.

**2.2.3 Cell Thawing.** For each thawed cryovial one 15 ml Falcon® tube was prepared with 6ml of DMEM cpt. Cryovials were rapidly transferred from liquid nitrogen under the hood and thawed by resuspending the cells in warm DMEM cpt medium already present in the 15ml Falcon tube. Cells were centrifuged at 700 rpm for 5 min and the supernatant discarded. Cells were finally resuspended in 6ml DMEM cpt medium, seeded in the appropriate T25 flask and incubated at 37°C and 5% CO<sub>2</sub> in a humidified incubator.

## **2.3 Confocal laser scanning microscopy**

**2.3.1 Coverslips treatment with poly-L.** Sterile glass coverslips (12 mm round, 1001/12 BIOSCIENTIFICA) were treated with poly-Lysine in a 24 well plate prior to cell seeding, to allow better attachment of HEK293A to glass coverslips. Poly-Lysine (poly- L) solution was prepared by resuspending 5mg of poly-lysine (SIGMA 6282) in 50 ml of sterile MilliQ water in a sterile environment, and stored at +4°C. After placing the sterile coverslips in a 24 well plate, 500 µl of poly-L were added to each well and the coverslips were pushed to the bottom of the well with a pipette tip to remove air bubbles, and the 24 well plate was incubated at 37°C and 5% CO<sub>2</sub> in a humidified incubator for at least 30 min. Subsequently, the poly-L was carefully removed, and the coverslips were rinsed with sterile MilliQ water. Coverslips were tilted 45 degrees and dried overnight in order

to avoid sticking to the plate.

**2.3.2 Cellular Transfection with Lipofectamine 2000.**  $5 \times 10^4$  cell/well HEK 293A cells were cultured on poly-L pre-treated glass coverslips (**Section 2.3.2**) in 24 well plates and incubated at 37°C for 24 hours. The day after cells transfection were performed with lipofectamine 2000 (Thermofisher Scientific, #11668030). The appropriate amount of DNA needed for transfection (5-250 ng) was added to 50  $\mu$ l of Opti-MEM Reduced Serum Media (Gibco, Thermofisher Scientific #11058021), in a 1.5 ml microcentrifuge tube. Subsequently 50  $\mu$ l of lipofectamine 2000 diluted 1:50 in Opti-MEM Reduced Serum Media were added to each tube, and incubated for 20 min. In the meanwhile, cell culture media was replaced with media without antibiotic. Finally, 100  $\mu$ l of each mixture was transferred to the corresponding well. Cells were incubated at 37°C with 5% CO<sub>2</sub>. 24h post transfection cell culture media was replaced with 500  $\mu$ l cpt DMEM and protein expression was monitored with an inverted fluorescence microscope (Leica, #DFC420C). At 48 hrs post transfection cells were incubated for 30 min with DRAQ5™ (Thermofisher Scientific, #62251; 1:5000 in DMEM no phenol red) and washed with 500  $\mu$ l PHEM 1x (**Section 2.4.4**). Cells were fixed with 400  $\mu$ l of 3% paraformaldehyde in PHEM 1x (**Section 2.4.5**) for 10 min at RT, before being washed with 1ml of PHEM 1x. Cells were mounted on object holder slide with mounting media Fluoromount G (Bioscience, #00-4958-02) (**Section 2.4.3**). Samples were stored at 4°C protected from light.

**2.3.3 Cellular ATP depletion.** For inhibition of active nuclear transport, 48 hrs post transfection, cells were incubated for 1 h at 37°C in a humidified incubator, with a cellular ATP depletion media consisting of No phenol-red, glucose-free DMEM (Gibco™ ThermoFischer Scientific, #A1443001) containing 10% Fetal Bovine Serum (FBS), L-glutamine (2mM), Penicillin and Streptomycin (100U/ml), HEPES 1M and freshly added sodium azide (10mM) and (2-deoxy D-glucose 6mM), along DRAQ5 (Thermofisher Scientific, #62251; 1:5000 in DMEM no phenol red).

**2.3.4. Confocal Laser Scanning Microscopy.** Subcellular localization of spontaneously fluorescent fusion proteins was evaluated by a Nikon A1 confocal laser scanning microscope (Nikon), equipped with a 60x oil immersion objective. Samples were scanned at 200 HZ, with an image resolution of 1024 x 1024 pixels (or, alternatively, of 524 x 524 pixels). For each sample four fields were acquired with a zoom of 1, and for each field several images at zoom of 2.5 were acquired, with variable settings in order to obtain in focus and unsaturated images of each cell. GFP and YFP fusions were excited at 488 or 514 nm using an argon laser, respectively. Mcherry fusions were excited at 568 nm with a HeNe laser, while DNA-bound DRAQ 5 was excited with a 633 nm laser. Digital images were

saved in tiff format for further analysis.

**2.3.5 Image analysis.** Fn/c values were determined by FiJI public domain software (NIH). In order to quantify nuclear accumulation of individual proteins from digital images, a specific ROI was set to measure the mean fluorescent signal in nucleus (Fn) and in the cytoplasm (Fc) from individual cells. Such values were corrected by subtracting the background fluorescence (Fb). Subsequently the levels of nuclear accumulation of the protein of interest were calculated, according to the following formula:  $Fn/c=(Fn-Fb)/(Fc-Fb)$ . Data were statistically analyzed using Graphpad Prism 9 (GraphPad Software) software to perform Student's T-test or ANOVA.

## **2.4. Buffers and media**

### **2.4.1 No phenol red DMEM medium**

To avoid any signal interference during cell imaging with CLSM, DRAQ 5 staining was diluted in cell culture media phenol-red free. Media (Gibco, ThermoFischer Scientific, #21063029) was supplemented with 10% Fetal Bovine Serum (FBS), 1% non-essential aminoacids, 1% L-glutamine (2mM), Penicillin and Streptomycin (100U/ml). Stored at +4°C.

**2.4.2 Opti-MEM medium.** Opti-MEM reduced serum media (Gibco, ThermoFischer Scientific #11058021) was used for lipofectamine based transfection. This is a modification of Eagle minimum essential medium, buffered with HEPES and sodium bicarbonate, and supplemented with hypoxanthine, thymidine, sodium pyruvate, L-glutamine, and growth factors.

**2.4.3 Fluoromount-G® mounting medium.** Coverslips were mounted on objector holder slide with 7 µl of Fluoromount-G® (FG) mounting medium (Bioscience #00-4958-02).

**2.4.4 PHEM 4x.** The stock was prepared with 240mM PIPES (SIGMA, #10011028244), 100 mM HEPES (stock 1M), 40mM EGTA (SIGMA, #111K5411) and 16mM MgSO<sub>4</sub> (SIGMA, #102215511). The pH was adjusted to 6.7 with NaOH 2M and MilliQ water was added to reach a final stock volume of 100ml. Subsequently the buffer was filtered in a sterile environment with a 0.22 µm filter. For long period storage it was kept at -20°C and for the short term at +4°C.

**2.4.5 Paraformaldehyde 4%.** Paraformaldehyde (SIGMA, #158127; 8 g) was dissolved in 140 ml of MilliQ water. Subsequently, 20 ml of PBS 10x was added to the solution and the pH adjusted to 7.4 with in the presence of NaOH 1M. MilliQ was added to reach a final volume of 200 ml and the solution aliquoted in 50ml falcons for storage at -20°C.

## 3-Results

### 3.1. Viral proteins classification based on bioinformatic analysis.

As mentioned before (section 1), our previous studies developed an automated pipeline that combines UniProt annotations of all human viral proteins with the two most popular cNLS mapping softwares (PSORTII and cNLS mapper) to identify novel putative viral classical nuclear localization signals (cNLSs) (**Figure 5.1**). This approach allowed classifying every protein from human viruses into a specific category. Specifically, a viral protein could be classified as:

- i) **Confirmed nuclear viral proteins relying on the IMP $\alpha$   $\beta$  nuclear import pathway**, annotated as nuclear on Uniprot and with a cNLS predicted by at least one algorithm.
- ii) **Confirmed nuclear viral proteins not relying on the IMP $\alpha$   $\beta$  nuclear import pathway**: annotated as nuclear on Uniprot but without a cNLS predicted by any algorithm.
- iii) **Hypothetic nuclear viral proteins relying on the IMP $\alpha$   $\beta$  nuclear import pathway**: not annotated as nuclear on Uniprot but with at least one cNLS predicted by each algorithm.
- iv) **Non-nuclear proteins**: not annotated as nuclear on Uniprot, cNLS predicted by one or fewer algorithms, and no nuclear localization predicted by any algorithms.

This allowed us to identify more than 200 viral proteins for follow up studies aimed at elucidating the virus-host cell interface. Moreover, such database can also be used to study how different viral families interact with the host cell nuclear transport apparatus.

In this study, we started the functional validation of potential candidates with two different approaches. As a first approach, we focused on viral proteins which are known to be nuclear in certain viral species, but not in their orthologous belonging to the same viral family. In this respect we investigated the nuclear transport process of all human *Polyomaviridae* large tumor antigens. As a second approach, we functionally validated the 26 viral cNLS with the highest score according to the software cNLS mapper. Interestingly 11 out of such proteins belong to the *Poxviridae* family.

### 3.2. Large T antigen *polyomaviruses*

#### 3.2.1. Comparative bioinformatic analysis identified putative cNLSs on *Polyomaviridae* LTA.

As an example of the first approach, we focused on members of the Polyomaviridea family, which are small non-enveloped viruses replicating in the host cell nucleus. Accordingly, most of their proteins localize into the host cell nucleus during virus replication. Our Bioinformatic analysis on five members of the *Polyomaviridae* family showed that  $68 \pm 21.68$  % of the overall viral proteins

localized in the nucleus. However, this percentage would be raised if we also consider hypothetical nuclear proteins (**Figure 5.2**). Intriguingly, the Large Tumor Antigens (LTA) from BK and KI *Polyomavirus* were classified as “confirmed nuclear with cNLSs”, while those for WU, JC, and MCP were classified as “hypothetical nuclear proteins with cNLS”.

### **3.2.2. cNLSs on LTAs from different HPyVs are heterogeneously distributed.**

Given the role of HPyVs LTAs in viral genome expression and replication, all HPyVs LATs should localize to the cell nucleus during viral infection. Therefore, our results imply miss annotations on uniprot data. To shed light on this issue, we set out to search for putative cNLS in the primary sequence from LTAs of all the 14 known HPyVs. By combining the phylogenetic and bioinformatic analysis, we identified at least one putative cNLS for each LTAs. Interestingly, 50% of all these proteins have more than one putative monopartite cNLS (**Figure 5.3**). Indeed, several HPyVs (WUPyV, MCPyV, and HPyV12), possess two putative cNLSs, while LipyV has three putative cNLSs. In JCPyV and BKPyV, the cNLS from SV40 (PKKKRKV-132) is completely conserved. Along with the different comparative bioinformatic analysis performed in the previous study, we noticed that the cdc2 T phosphorylation sites, upstream of the P2-P5 basic residues, was conserved in all of the identified cNLS, which modulate the rate of nuclear import on SV40 LTA. Additionally, CK2 phosphorylation site, roughly eleven residues upstream of LTA NLS, which enhanced the transport in most identified cNLS, was also conserved (**Figure 5.3**).

### **3.2.3. Nuclear subcellular localization of newly identified cNLSs on all human *Polyomaviruses* LTA.**

In order to discriminate functional from non-functional newly identified cNLSs, we generated recombinant plasmids expressing such sequences fused to GFP to visualize fusion protein subcellular localization in transfected cells. Depending on the length of the cNLS of interest, we used two different cloning techniques to generate expression plasmids. Indeed, sequences shorter than 30 amino acids were cloned via TOPO® reaction (**section 2.1.7**), while larger sequences were cloned with the Gateway® Technology (**section 2.1.2**). Plasmids were transfected into HEK 293 A cells, and 48 hours post-transfection, cells were stained with DRACQ5 to visualize nuclei, before being fixed. Subsequently, subcellular localization of the GFP-cNLSs fusions was quantitatively analyzed by confocal laser scanning microscopy (CLSM; **section 2.3.4**). GFP alone was expressed as a negative control for nuclear accumulation, while the NLSs from SV40 and MPyV fused to GFP were used as positive controls. As expected, GFP equally distributed between nucleus and cytoplasm ( $F_n/c \approx 1$ ; **Figure 5.4**) whereas GFP fused to the NLSs of SV40 and MPyV accumulated in the nucleus,

although to different extents (SV40-NLS;  $F_n/c \approx 8$ , MPyV-NLSnt;  $F_n/c \approx 6$ ; MPyV-NLSct;  $F_n/c \approx 1.5$ ). Our analysis (**Figure 5.4**) further revealed that six LTAs possess one highly functional monopartite cNLS (HPyV12, HPyV6, HPyV9, TSPyV, in addition to BK and JC whose NLS is identical to that of SV40). On the other hand, STLPyV and KIPyV LTAs do not possess any active monopartite cNLS, while the cNLS from HPyV7 is extremely weak ( $F_n/c < 2$ ). Intriguingly, MWPyV, MCPyV and WUPyV LTAs possess two functional monopartite cNLSs whose activity is extremely variable. Indeed, while both MWPyV appeared strongly functional ( $F_n/c > 5$  and  $2 < F_n/c < 5$ ), only cNLSm from MCPyV exhibited a good activity ( $F_n/c > 7$ ) while its NLSct had low activity ( $F_n/c < 2$ ), similar to both WUPyV NLSs ( $F_n/c < 2$ ).

Our results thus raised the question of how KI and STLPyV LTAs could reach the cell nucleus. By carefully analyzing their sequences, we noticed that STLPyV and KIPyVs LTAs present basic residues upstream of the non-functional monopartite cNLS, which could potentially form bipartite cNLSs. A similar scenario could be observed for other LTAs, including that from WUPyV, where its two weak cNLSs are only 16 amino acids apart (**Figure 5.5A**).

### **3.2.4. Large T antigens in STL, KI and WUPyV have a potential bipartite cNLS.**

To analyze the possibility of bipartite cNLS mediating nuclear import of certain HPyV LTAs, we generated mammalian expression vectors encoding for bipartite versions of STL, WU, and KIPyV cNLS. Which were transfected into HEK 293 A cells, at 48 hours post-transfection, cells were stained by DRACQ5 to visualize nuclei, before being fixed. Subsequently, the level of nuclear accumulation of transiently expressed GFP fusions were quantified by CLSM as previously. Our analysis indicated that the levels of nuclear accumulation of the bipartite cNLS were significantly higher compared to their monopartite counterparts (**Figure 5.5**). Our results suggested that nuclear import of STL, WU, and KIPyV LTAs might be dependent on bipartite cNLSs rather than monopartite one as previously described for SV40, and MCPyV. We went further in this hypothesis by comparing the activity of such bipartite cNLSs with mutated versions, whereby upstream basic residues were replaced by Alanine. Recombinant plasmids expressing mutant bipartite cNLS of STL, KI, WU LTAs were used to transfect HEK293A cells, the level of nuclear accumulation of transiently expressed GFP fusions were quantified by CLSM as previously (**Figure 5.5**). In all cases substitution of upstream basic residues with alanine strongly impaired nuclear accumulation of all GFP-fusions. These results suggested that cNLS in STL, KI, WUPyV LTAs are bipartite and upstream basic residues are required to bind to the minor binding site of IMPa.

### **3.2.5. HPyV7 LTA accumulates into the nucleus in a concentration dependent manner and contains a functional cNLS, which its activity is modulated by upstream residues**

Having identified strong bipartite cNLSs in STL, KI, WUPyV LTAs, we focused on cNLS in HPyV7, which is the weakest cNLS among all HPyV LTAs. As mentioned earlier (**Figure 5.3**), all identified functional cNLSs from LTAs contain upstream putative CK2 phosphorylation sites. Negative charge upstream residues of cNLSs have been shown to increase the cNLS activity, and upstream sequences has been shown in some cases increase binding to IMP $\alpha$ / $\beta$ . We decided to study the contribution of HPyV7 cNLS upstream sequences and CK2 mediated phosphorylation in nuclear accumulation of HPyV7 LTA. Therefore, we generated plasmids mediating the expression of HPyV7-cNLS including upstream residues either (115-153) or CK2 phosphorylation sites (115-153, S116A, S117A, S118A) see (**Figure 4.6A**). We transfect HEK293A cells to transiently expressed such fusion proteins or minimal cNLS and compared their levels of nuclear accumulation CLSM. Importantly GFP – HPyV7 (115-153) accumulated into the nucleus significantly higher than GFP – HPyV7 (145-152) (Fn/c ~ 2.5 vs ~1.6 respectively). These results showed upstream sequences increase HPyV7 cNLS activity. Surprisingly, substitution of CK2 phosphorylation sites with Alanine increase the nuclear accumulation (Fn/c ~ 3.2), indication that CK2 mediated phosphorylation reduce the HPyV7 LTA nuclear accumulation rather than increasing it (**Figure 5.6**). Overall, our results imply that upstream sequences are involved in modulating the activity of HPyV7 weak cNLS. We therefore noticed that if HPyV7 LTA subcellular localization is influenced by expression levels. To this end, we transfect different concentration of plasmids expressed HPyV7-GFP in HEK293A cells. We observed reduction of plasmid amount, increased accumulation of HPyV7 LTA-GFP into the nucleus and also the frequency (68% vs 20% of cases) of HPyV7 LTA nuclear accumulation. Importantly, nuclear accumulation was entirely dependent on its cNLS (PPKQKKPN-152), because substitutions K147A/K149A (**Figure 5.6D**) abrogated nuclear import (**Figure 5.6E**) in cells transfected with either low or high amounts of plasmid DNA (Fn/c=0.1, **Figure 5.6F**), resulting in cytoplasmic localization in 100% of transfected cells (**Figure 5.6G**).

### **3.2.6. Nuclear accumulation of HPyV LTAs is dependent on the IMP $\alpha$ / $\beta$ heterodimer.**

To further evaluate the role of these newly identified HPyV LTAs cNLSs, and the role of other extra regulating factors in nuclear import, we had synthesized plasmids expressing full-length LTAs from STLPyV, KIPyV, MWPyV, HPyV7, and MCPyV fused to the N-terminal GFP. Each plasmid was transfected to HEK293A cells, and localization of GFP fusions was quantitatively analyzed as

previously. We observed substantial nuclear accumulation for STL, KI, MW, and MCPyV LTAs while the localization of HPyV-7 LTA, was extremely heterogenous. Indeed, HPyV7 LTA accumulated in the nucleus in only 20% of analyzed cells, whereas it was mainly retained in the cytoplasm in 80% of them (**Figure 5.7**). We then decided to investigate if LTAs nuclear import is dependent on IMP $\alpha$ / $\beta$ , as expected due to the presence of functional cNLS. To this end, GFP-full-length proteins were co-transfected with a plasmid expressing mCherry-Bimax2, well-known inhibitor of IMP $\alpha$ -mediated nuclear transport. At 48h post-transfection, cells were processed as previously, and the subcellular localization of fusion proteins analyzed by quantitative CLSM. GFP-H1E, and GFP-UL44 were used as negative and positive controls, respectively. As expected, both GFP-UL44 and GFP-H1E strongly accumulated into the nucleus in the absence of mCherry-Bimax2 (**Figure 5.7**). However, upon the co-expression with the latter GFP-UL44 remained in the cytoplasm, in accordance with its well-known IMP $\alpha$   $\beta$  dependent nuclear import, and its ability to form homodimers incapable to passively diffuse into the nucleus in the absence of active transport. On the other hand, co-expression of mCherry-Bimax2 had no effect on subcellular localization of GFP-H1E, 45 KDa fusion protein capable to bind to dsDNA, and being imported into the nucleus by multiple importing pathways. Importantly, STL, KI, MW, and MCPyV LTA retained in the cytoplasm upon co-transfection with bimax2-mcherry, confirming they are imported to the nucleus by IMP $\alpha$   $\beta$  (**Figure 5.7**).

### 3.2.7. cNLS in MWPYV and MCPyV LTA works as a bipartite.

Next, we decided to investigate if bipartite cNLSs are present in additional LTAs. Indeed, as mentioned earlier (Figure 4.5) both MCPyV and MWPYV LTAs possess two closely located monopartite cNLSs, whereby the upstream cNLS is stronger than the C-terminal one (Figure 4.4). In the case of MWPYV LTA such signals are separated by 11 amino acids and for MCPyV LTA by 20 amino acids. we hypothesized that these two NLSs might work synergically or as a bipartite cNLS. In MWPYV, we compared the generated potential bipartite cNLS with monopartite cNLSs, and our analysis showed a higher Fn/c value for bipartite rather than monopartite cNLSs. These results suggesting that these two basic clusters in MWPYV LTA work together as a bipartite cNLS (**Figure 5.8**). For MCPyV LTA we not only compared the activity of monopartite and bipartite cNLSs, but we also generated a number of substitution derivatives therefore (**Figure 5.9**). Indeed, MCPyV LTA cNLS has been previously extensively characterized and is considered to be only dependent on nuclear import. The k278T substitution has been shown to completely abolish MCPyV LTA nuclear

localization. On the other hand, K280T substitution did not affect its nuclear targeting. We introduced such substitution in the context of monopartite and bipartite cNLSs and compared the activity of such cNLSs by fusion to GFP (for monopartite cNLSs) or YFP (for bipartite cNLSs). Our results indicated that the bipartite cNLS is stronger than cNLS<sub>m</sub> (Fn/c ~ 8 and ~ 7 respectively). when we introduced the K278T substitution, it strongly impaired activity of both cNLS<sub>m</sub> and cNLS<sub>bip</sub> (Fn/c ~ 1.79 and ~ 0.9 respectively). However, the K280T substitution strongly impaired activity of cNLS<sub>m</sub> (Fn/c ~ 1.26), but not of cNLS<sub>bip</sub> (Fn/c ~ 5.01). Similar effects were obtained in the context of full length MCPyV LTA. These results suggested the K280T bipartite substitution (PFSRKRtFGGSRSSASSASSASFTSTPPKPKKNRE-307) is functionally as a bipartite and not as a monopartite cNLS as previously proposed. However, substitutions K303A/K304A within MCPyV NLS<sub>ct</sub> did not completely abolish nuclear import of MCPyV LTA (**Figure 5.9**), which mainly localized to the nucleus (Fn/c of 13.8) in almost 100% of analyzed cells (**Figure 5.9G**). Therefore, upon inactivation of the C-terminal stretch of basic amino acids within its bipartite NLS, MCPyV LTA cNLS<sub>m</sub> is sufficient to mediate nuclear transport. Finally, when either the K278T or the K280T substitution were tested in combination with the K303A/K304A substitutions, a complete loss of nuclear import was observed, both in the context of MCPyV LTA NLS<sub>bip</sub> (Fn/c=1, **Figure 5.9B-C**) and of full length MCPyV LTA. Taken together our results suggest that MCPyV LTA possess two closely located stretches of basic amino acids which function as a bipartite cNLS. However, in the absence of downstream basic residues, cNLS<sub>m</sub> from MCPyV is still functional.

### **3.3. Twenty-six previously undescribed viral cNLSs mediate active transport of GFP to the cell nucleus.**

As mentioned in the *Introduction* section, our previous bioinformatic analysis identified several viral proteins which are not known as nuclear but contain putative cNLSs according to two prediction software. Such proteins therefore represent potential novel nuclear localizing proteins. In order to begin the functional characterization of such hypothetical nuclear proteins, we focused on those bearing the strongest putative cNLSs according to the software cNLS mapper. We focused on 26 viral proteins bearing a putative cNLS with a cNLS mapper score higher than 9. Surprisingly, most of such proteins belong to the *Poxviridae* (11 proteins) and the *Herpesviridae* (10 proteins) families, while two proteins belong to *Anelloviridae* family (**Figure 5.10A**). Three remaining proteins belong to *Phaboviridae*, *Phenuiviridae*, and *Orthomyxoviridae* family members. We decided to test the activity of such putative cNLS as we did for Polyomaviruses (**Figure 5.10B**). To this end, plasmids expressing such sequences fused to GFP were transfected in HEK293A cells and the subcellular localization of fusion proteins quantitative analyzed by CLSM 48 hours later. GFP alone was

expressed as a negative control for nuclear accumulation, while the cNLS from SV40 LTA fused to GFP was used as a positive control. Interestingly, all GFP-fusion proteins accumulated in the nucleus to higher levels as compared to GFP alone, suggesting cNLS functionality (**Figure 5.10C**). However, the levels of nuclear accumulation were extremely variable for among the GFP fusion tested (Fn/c ranging between around 1.12 to 20) implying important differences in cNLS activity. Furthermore, several GFP fusions also accumulated to different extents in the nucleoli of transfected cells (Q98187-NLS, Q6TVJ0-NLS, Q98291-NLS, Q69514-NLS, Q6TVM4-NLS, and Q9DUB7-NLS). This observation raised the possibility that nuclear accumulation of such proteins could be due to the passive diffusion across the NPC into the nucleus, followed by intracellular binding to nuclear structures such as RNA, rather than due to active nuclear transport. To ensure that nuclear translocation of GFP fusions is due to the active transport, we analyzed their subcellular localization upon ATP depletion (section 2.3.3). GFP alone and GFP-SV40LTA-NLS were used as negative and positive controls, respectively. Our analysis showed that ATP depletion impaired nuclear accumulation of all fusion proteins tested, demonstrating the identified cNLS are capable of conferring active nuclear transport properties to heterologous proteins (**Figure 5.11**).

#### **3.4. A functional bipartite cNLS is conserved across Poxviridae N2 proteins.**

As mentioned above, 11 of the 26 potential novel nuclear localizing viral proteins identified, belong to the Poxviridae family (**Figure 5.11A**). Intriguingly four of them are orthologs of the N2 protein from Vaccinia virus (Figure 4.11B). These include orthologues from Cowpox (U5TCT3), Horsepox (Q0GP58), Monkeypox (Q8V556), and Variola virus (P0DSQ1). Previous studies showed that the Vaccinia virus N2 protein localizes in the nucleus of infected cells and inhibits interferon regulatory factor (IRF3). However, its NLS and its nuclear transport mechanism have not been characterized so far. We hypothesized that nuclear targeting and modulation of host immune response during infection could be conserved among Poxviridae N2 orthologues. However, the cNLS from such orthologues tested in our previous experiment exhibited very different activity. Indeed, while the cNLS of Cowpox U5TCT3 was highly active (Fn/c  $\approx$ 8.2), those from Horsepox, Monkeypox, and Variola virus orthologues were not (Fn/c  $\approx$ 1.5). Intriguingly the sequence identified by cNLS mapper for Cowpox U5TCT3 is a bipartite cNLS, while those identified for the other orthologues are monopartite (Figure 4.11C). Detailed sequence analysis of such proteins revealed that, all coding sequences are highly similar for all orthologues, and all contain partially overlapping monopartite and a bipartite cNLSs, including Vaccinia Virus N2 (**Figure 5.12D**). This suggests the presence of potential bipartite cNLSs for VACCW N2 and all its orthologues. To verify if functional bipartite

cNLSs are present also in N2 from Monkeypox, Vaccinia, Horsepox and Variola viruses, we generated mammalian expression vectors encoding such putative bipartite cNLSs fused to GFP and compared their subcellular localization to that of their monopartite counterparts fused to GFP after transfection of HEK293A cells by quantitative CLSM (**Figure 5.12E**). Our results indicated that in all cases bipartite cNLSs [Q8V556/P14357: **IRKRPNQHHTIDLFKRIKRTRYDTF**, Fn/c ~ 11.25; Q0GP58/ P0DSQ1: **IRKRPNQHHTIDLFKKIKRT**, Fn/c ~7.49], were significantly more efficient in mediating nuclear transport compared to their monopartite counterparts [Q8V556/P14357: LFKRIKRTRYDTF, Fn/c ~ 1.64; Q0GP58/ P0DSQ1: LFKKIKRTRYDTF, Fn/c ~ 1.51] (**Figure 5.12FG**). To further verify the bipartite nature of such newly identified cNLSs, we compared the subcellular localization of GFP fusions with wild-type bipartite NLSs of cowpox, monkeypox and vaccinia virus, to that of their substitution derivatives whereby N or C terminal stretches of basic amino acids are replaced with alanine (**Figure 5.13A**). Our results indicated that mutation of either stretch was sufficient to abolish nuclear targeting completely (**Figure 5.13B-C**). Thus, our result clearly shows that Vaccinia virus N2 protein contains a bipartite cNLS which is conserved across other Poxviridae members.

### **3.5. The bipartite cNLS of Vaccinia virus N2 is responsible for IMP $\alpha/\beta$ mediated nuclear import.**

Sequence analysis revealed that such bipartite cNLSs are conserved across all known N2 orthologues (**Figure 5.13A-B**). We therefore reasoned that such sequences must play an important role during virus life cycle by mediating nuclear import of such proteins. We decided to test this hypothesis by characterizing the nuclear import process of vaccinia virus N2 protein. To this end, we investigated if its nuclear localization is dependent on the identified cNLS and IMP $\alpha/\beta$  (**Figure 5.14C**). To this end, we compared the nuclear accumulation of full-length N2 in the presence or absence of bimax2, ATP depletion, and substitution of its basic residues to alanine (IaaaPNQHHTIDLFaaIaa). As expected, YFP-N2 accumulated to the cell nucleus (Fn/c 2.5), confirming previous observations from other groups for the untagged protein, and implying the YFP tag does not alter its subcellular localization. (**Figure 5.13D-E**). Importantly, nuclear accumulation was prevented upon co-expression with Bimax2 (Fn/c 1.5). This observation demonstrates for the first time that N2 is actively transported into the nucleus by IMP $\alpha/\beta$ . Similar results were obtained by substitution of both basic stretches of amino acids with alanine (Fn/c 1.5). This observation proves that the bipartite cNLS we identified is absolutely required IMP $\alpha/\beta$  dependent nuclear import. Intriguingly, ATP depletion further increased YFP-N2 nuclear accumulation (**Figure 5.14**). Given its estimated apparent MW of 47 kDa, this result might be partially explained by its ability to form homodimers

after translocation into the nucleus, or by the presence of undescribed nuclear export sequences. Overall, our data indicated that VACCW N2 proteins could be actively transported in the nucleus of infected cells in the presence of bipartite cNLS despite having a small molecular weight. This evidence increases the idea that N2 competes with host transcription factors for available IMP $\alpha$ 2 to modulate nuclear transport, thus promoting virulence.

### **3.6 A19 is actively transported into the nucleus relying on IMP $\alpha$ / $\beta$ 1 complex.**

Among, 11 potential novel nuclear localizing viral proteins identified, three of them are orthologs of Vaccinia A19 viral protein [Molluscum Contagiosum, Orf virus, and Yaba monkey tumor virus] (**Figure 5.15B**). Vaccinia virus A19 protein is a small protein (9-kDa) that is synthesized after viral DNA replication. Previous sequence analysis of A19 orthologs indicated the presence of conserved NLS in A19 protein and its orthologous (**Figure 5.15C**), however the activity of their NLS and the exact role of A19 in the nucleus of infected cells is still unclear.

The cNLS from such orthologues tested in our previous experiment exhibited active cNLS ( $F_n/c \approx 2$ ). To investigate the role active cNLS in nuclear transport of A19 proteins and characterize the mechanism of their nuclear import. To this end we fused the full-length protein sequence of A19, or substitution derivatives whereby basic residues were replaced with alanine with YFP (**Figure 5.15D**) and transfect HEK293A cells with plasmids encoding such YFP-fusion proteins. After 48h, fusion proteins analyzed by quantitative CLSM. GFP alone used as a negative control. GFP-A19 accumulated in the nucleus to higher levels as compared to GFP alone ( $F_n/c \approx 2.29$ ), and substitution of its cNLS basic residues with alanine (MKSR $^{aaa}$ PaTT-20) completely impaired nuclear accumulation ( $F_n/c \approx 1.08$ ). Thus, our result clearly shows that Vaccinia virus A19 protein contains a monopartite cNLS which is conserved across other Poxviridae members, and it is essential for nuclear import (**Figure 5.15E**). To characterize the nuclear import process of vaccinia virus A19 protein, we investigated if its nuclear localization is due to active transport dependent IMP $\alpha$ / $\beta$ . To this end, we compared the nuclear accumulation of full-length A19 in the presence or absence of bimax2, and ATP depletion. Importantly upon Bimax2 co-expression nuclear accumulation of fusion proteins were abolished ( $F_n/c \approx 0.95$ ). A similar reduction in nuclear accumulation was obtained after ATP depletion ( $F_n/c \approx 1.25$ ), indicating active import. This observation demonstrates for the first time that the nuclear localization reported previously for A19 is due to active transport into the nucleus by IMP  $\alpha$ / $\beta$  rather than due to passive diffusion due to its small size (**Figure 5.15E- F**).

## 4. Discussion

In eukaryotes the nucleus is separated from the cytoplasm by a double membrane called nuclear envelope. Exchange of molecules between the nucleus and cytoplasm occurs through nuclear pore complexes (NPCs), large multimeric structures that are embedded in the nuclear envelope and act as permeability barriers between the cytoplasm and nucleoplasm. Small molecules up to 240 kDa pass through the NPC via passive diffusion, while large molecules exploit nuclear transporters which necessitates the presence of a nuclear localization signal (NLS) on cargo. To this end, NLSs are recognized by the nuclear transporters belonging to the importin (IMP) superfamily, which can interact with hydrophobic residues within nucleoporins to facilitate the nuclear transportation of NLS-containing proteins. There are different kinds of NLSs according to their residue composition and IMPs recognizing them. The best characterized NLSs are the classic (c)NLSs that are highly basic sequences recognized by  $IMP\alpha$   $\beta$ . another type of NLS is represented by Arginine-rich NLSs (R-rich NLSs) which are directly recognized by  $IMP\beta$ . In contrast, PY-NLSs have diverse sequence and larger structure compared to cNLSs and are directly recognized by  $IMP\beta 2$  by multiple interactions between the weak NLS and  $IMP\beta 2$  (51).

cNLS can either be monopartite or bipartite. Monopartite cNLSs have a single cluster of basic amino acids, whereas bipartite cNLSs have two clusters of basic amino acids separated by 10-12 amino acids linker (51). The consensus for Monopartite NLS is usually defined as  $K(K/R)X(K/R)$ , where X can be any residue. By contrast, in bipartite cNLS, the consensus sequence can be expressed as  $R/K(X)-10/12-KRXX$ . several studies addressed the recognition of cNLS and  $IMP\alpha$  from a biochemical, functional, and structural point of view.  $IMP\alpha$  contains a major and a minor NLS binding site. Interaction of monopartite NLSs with the major site is sufficient for nuclear translocation, while bipartite NLSs occupy both sites with their basic clusters (11).

As discussed before (section 1.1.5), many viruses exploit the host cell nuclear trafficking machinery, in particular, the pathway mediated by the host  $IMP\alpha/\beta 1$  heterodimer, to ensure of viral encoded proteins access to the nucleus to facilitate viral replication or to antagonize the host antiviral response (13). In recent years, nuclear transport is becoming an attractive target for cancer and viral infection treatment, and researchers paid more attention to identify novel NLSs and the import nucleoporins that recognize and bind them.

Therefore, our lab previously developed an automated pipeline allowing to combine UniProt annotations for subcellular localization and the primary sequence of every human viral protein, with the two most popular cNLS mapping software (PSORTII and cNLS mapper), to identify novel potential viral cNLSs. This analysis led to classification of all human viral proteins into three main

categories: Confirmed Nuclear, Putative Nuclear, and Cytosolic proteins. Consequently, identification of about 200 novel viral proteins potentially translocated into the nucleus of infected cells via IMP  $\alpha/\beta$  pathway due to their putative cNLSs. In this work, we began to functionally validate such hypothetical nuclear proteins, with two different approaches. First, we focused on viral proteins, which were already known to be nuclear in some viral species. However, they were not characterized as nuclear protein in their orthologous belonging to the same viral family. As a second approach we functionally validated the 26 top ranking newly identified viral cNLS based on their strength, according to cNLS mapper.

### **HPyVs LTAs are transported to the nucleus by the IMP $\alpha/\beta$ heterodimer**

As an example of the first approach, we focused on Human polyomaviruses (HPyVs) large T antigens (LTAs), which were classified as “confirmed nuclear with cNLSs” in BK and KI *Polyomavirus* while those for WU, JC, and MCP were classified as “hypothetical nuclear proteins with cNLS.” The crucial role of HPyVs LTAs in viral genome expression and replication suggested missannotation on Uniprot for these proteins. Moreover, studies from several groups have shown that nuclear localization of SV40 LTA and protein-protein interactions between LTA and the cellular replication factors in the host nucleus is critical for virus replication (23), since K127T substitution, which is sufficient to inactivate LTA nuclear transport, completely abrogates viral replication. Therefore, we decided to extend our analysis to all HPyVs. By combining the phylogenetic and bioinformatic analysis from LTAs of all the 14 known HPyVs, we showed each LTA bears at least one putative cNLSs (Figure 4.3). By quantitative laser scanning microscopy (CLSM), we quantified the level of nuclear accumulation of all HPyV LTAs’ cNLSs fused with GFP. Our results showed that the LTA of BKP, JC, TS, MP, and NUPyVs, as well as HPyV6, 9, and HPyV12, have a strong functionally active monopartite cNLS (**Figure 5.4**). We could also show the absence of a functional monopartite cNLS on KIPyV and STLPyV LTAs, and presence of additional basic amino acids upstream of their non-functional monopartite cNLSs, suggesting that nuclear import of STLPyV and KIPyV LTA could be mediated by bipartite cNLSs (**Figure 5.5**). Furthermore, MWPyV LTA contains two active cNLSs which are separated by a 11 amino acids linker. This raised the possibility that, in the context of full-length LTA, both cNLSs could bind to the major IMP $\alpha$  binding site. However, they could also function as a bipartite cNLS. By quantitative CLSM of cells transiently expressing either monopartite or bipartite NLSs fused to GFP and their substitution derivatives which basic residues were replaced with Alanine and introduction of the same substitutions in the context of full length STLPyV, KIPyV, and MWPyV LTA, we showed that STLPyV, KIPyV, and MWPyV LTA nuclear import is dependent on bipartite cNLSs.

Merkel cell polyomavirus (MCPyV)-positive Merkel cell carcinoma (MCC) tumor cell growth is dependent on the expression of LTA with an intact retinoblastoma protein (RB)-binding site. Previous studies from other scientists demonstrated that nuclear localization is essential for LTA function, MCPyV LTA cNLS has been previously characterized (cNLS<sub>m</sub>, PFSRKRK-280) and it is conserved in most MCCs. In some cases, MCPyV genome integration results in LTA truncation upstream of the cNLS<sub>m</sub>. In such cases LTA is still capable of reaching the cell nucleus due to its small size (24). For such reason MCPyV LTA nuclear transport has been to object of intense research in the past. The K278T substitution in MCPyV LTA cNLS<sub>m</sub> completely abrogates nuclear import, whereas the K280T substitution has no effect (8). This is surprising since the sequence PFSRKRtFGGS bearing the K280T substitution almost completely lost its activity as a monopartite cNLS (**Figure 5.5A-C**). The presence of an additional monopartite cNLS located 20 amino acids downstream (**Figure 5.5A**) suggests the possibility that the two sequences could operate as a bipartite cNLS. By introducing K280T and K278T substitution in the context of monopartite and bipartite cNLSs we compared the activity of such cNLSs fused to GFP. Indeed, the K280T substitution had no effect on nuclear translocation of fusion GFP and MCPyV NLS bip. Similar effects were obtained in the context of full length MCPyV LTA (**Figure 5.9 D-G**). Our results support the idea that previously reported nuclear accumulation of MCPyV LTA K280T, could not depend on its cNLS<sub>m</sub> and MCPyV LTA is imported into the nucleus by a bipartite cNLS. However, substitutions K303A/K304A within MCPyV NLS<sub>ct</sub> did not completely abolish nuclear import of MCPyV LTA. Therefore, uncharged mutation in C-terminal stretch of basic amino acids within its bipartite cNLS, demonstrates MCPyV LTA cNLS<sub>m</sub> is sufficient to mediate nuclear transport. Finally, when either the K278T or the K280T substitution were tested in combination with the K303A/K304A substitutions, a complete hamper of nuclear import was observed. These observations suggest us that MCPyV LTA has a bipartite cNLS while, in the absence of C-terminal basic residues, cNLS<sub>m</sub> can bind to the major binding site of IMP  $\alpha$  and functionally transfer LTA to the cell nucleus. This is quite surprising since the two basic clusters forming MCPyV cNLS bip are located 20 aa apart. However, our results were consistent with the recent study which demonstrated in yeasts aa linker between two basic clusters in bipartite cNLS can be extended up to 29 amino acids in length (52). Thus, we showed for the first time that, MCPyV LTA possess two stretches of basic residues located 20 aa far from each other which work as a bipartite CNLS to import LTA into the nucleus.

This is the first study to show heterogenicity in cNLSs of LTAs from polyomaviruses members. Indeed, bipartite NLSs were not described so far in any LTA. Our observations let us to investigate nuclear import more deeply.

The weakest cNLS was found in HPyV7, which indeed showed a variable localization pattern, with nuclear accumulation inversely proportional to protein expression levels. We do not yet know the reason behind this localization pattern. We assumed that since the LTA can bind to different transcription factors such as Rb (retinoblastoma) and P53 proteins or heat shock proteins such as Hsc70 (21), the HPyV7 protein nuclear importation pathway might require other functional transcription factors to enhance nuclear transportation. As we know from other studies, interaction with other proteins might modulate its nuclear transport. For example, it has been shown that SV40 LTA nuclear import is negatively regulated by binding of p110Rb to the retinoblastoma binding site (RbBS) in cNLS of SV40 LTA (53). This finding implies that complexation of LTA with other proteins in the cytoplasm has an important effect on recognition of LTA by IMP $\alpha$   $\beta$ . The most persuasive argument is, the nuclear accumulation of HPyV7 LTA is entirely dependent on its cNLS, because we demonstrated that substitutions K147A/K149A abrogates nuclear import (Figure 4.6 A-E). On the other hand, Phosphorylation has been shown to be able to regulate nuclear protein import through modulation of NLS-IMP interaction either positively or negatively (54). A well characterized example is SV40 LTA, which has several phosphorylation residues N-terminal to its NLS, and they are phosphorylated during infection, including the CK2 site at Ser111/112, which enhances nuclear uptake by enhancing NLS recognition by IMP  $\alpha/\beta$  (55), and the cyclin-dependent kinase site at Thr124, which effects cytoplasmic retention. We hypothesized that the heterogeneous subcellular distribution of HPyV7 LTA-GFP could be influenced by upstream phosphorylation events. According to this idea, GFP-HPyV7 LTA CcN, containing HPyV7 LTA minimal cNLS in addition to upstream CK2 phosphorylation sites, localized to the nucleus significantly higher than GFP-HPyV NLS (Fn/c 2.7 vs 1.5, **Figure 5.6 H-J**), indicating that upstream sequences are involved in IMP $\alpha$  binding and contribute to nuclear import.

Similar to mutagenesis experiment for full-length LTAs from KIPyV, STLPyV, MWPyV, and MCPyV that suggests their nuclear transport is dependent on a functional bipartite cNLS, we expressed such full-length LTAs in the presence or in the absence of mcherry-Bimax2 to make sure their nuclear transport is relying on IMP $\alpha$   $\beta$ . Bimax2 is a well-known inhibitor of IMP $\alpha$   $\beta$  which binds tightly to full-length versions of four mammalian importin  $\alpha$  members (56). Importantly, the presence of Bimax2, strongly inhibited nuclear import of all tested LTAs (**figure 5.7**).

## FUTURE DIRECTIONS

Since nuclear import of HPyV LTA is dependent on the  $IMP\alpha/\beta$  pathway, we can study the effect of Ivermectin, a well-known  $IMP\alpha/\beta$  inhibitor(ref), on HPyVs replication to develop new therapeutic approach for combat HPyVs virulence which is in progress by the group of prof. Serena Delbue (UniMi). To further verify of bipartite nature of cNLSs, these sequences are being crystallized in complex with  $IMP\alpha$  in addition, their binding affinity to different  $IMP\alpha$  isoform will be tested by the group of prof. Jade Forwood (Wagga Wagga University). As mentioned above we observed that upstream sequences in HPyV7 are involved in  $IMP\alpha$  binding affinity and contribute to nuclear import. By site specific mutagenesis of cdk1 phosphorylation site adjacent to the HPyV7 LTA NLS (T144A, T144E) we can investigate if HPyV7 LTA nuclear transport is dependent on the cdk1 phosphorylation as well. Furthermore, we can exploit BRET assay to evaluate LTAs interaction with RB and P53. We will also bioinformatically analyze the sequence of LTAs from all known PyVs by clustalW, to generate phylogenetic trees to identify other putative cNLSs in other species and characterize their nuclear import pathway.

As a second approach, we functionally validated the 26 top ranking newly identified viral cNLSs based on their strength according to cNLS mapper. To this end, we measured their nuclear targeting properties as GFP fusions in the presence or absence of IMP $\alpha$   $\beta$  inhibitors and ATP. Interestingly, 11 of such cNLSs belong to members of the Poxviridae family. Intriguingly, three of them are orthologs of the A19 and four of the N2 Vaccinia virus (VACV) proteins.

### **Vaccinia N2 protein contains bipartite cNLS which is conserved among poxviruses members**

Bioinformatics analysis of the poxvirus proteome identified VACV N2 is highly conserved in the *Orthopoxvirus* genus, and they contain potential bipartite cNLS, which is highly conserved in the full-length proteins. The presence of the cNLS suggests that N2 orthologous may typically localize into the nucleus relying on IMP $\alpha$   $\beta$  heterodimer. By comparing subcellular localization of GFP fused bipartite NLSs in Cowpox, Monkeypox, and Vaccinia virus with their monopartite counterparts fused to GFP, and their substitution derivatives whereby N or C terminal stretches of basic amino acids replaced with alanine, we confirmed that N2 and its orthologous: Cowpox (U5TCT3), Horsepox (Q0GP58), Monkeypox (Q8V556), and Variola virus (P0DSQ1), have an active bipartite cNLS (**Figure 5.13**). Proteins bearing an NLS are known to translocate to the nucleus through the host IMP proteins. To determine if N2 protein translocate into the nucleus relying on IMP $\alpha$   $\beta$ , we compared the nuclear accumulation of full-length N2 in the presence or absence of bimax2, ATP depletion, and substitution of its basic residues to alanine. Indeed, bimax2 co-expression or NLS mutation strongly reduced nuclear accumulation of YFP-N2 (Fn/c 3.80 vs 1.49 or 3.80 vs 1). Our data from bimax2 co-expression and uncharged mutations, confirms N2 proteins could actively transport in the nucleus of infected cells in the presence of bipartite cNLS despite having a small molecular weight. The N2 protein fused with YFP (47 kDa) still is below the ~240kDa cut-off for spontaneous diffusion through the nuclear pore. Alternatively, it may indicate that N2 binding to IMPs plays some role in modulating or interfering with the activities of protein nuclear transport systems, a function that would have obvious advantages for VACV. One hypothesis is that the N2 retard nuclear transport of host nuclear transcription factors and competes with them for available IMP $\alpha$  to inhibit the host antiviral response and facilitate viral replication. By contrast, Upon ATP depletion, the N2 protein was trapped in the nucleus, while we expected to see a ubiquitous distribution pattern due to its small molecular weight. This evidence raised the probability of this hypothesis that N2 proteins like vaccinia N1 protein, an inhibitor of nuclear factor NF $\kappa$ B and apoptosis that contributes to virulence, after translocation into the nucleus, forms the homodimers. Or it can be due to the presence of active

nuclear export signal (NES) in N2 protein, and N2 protein can shuttle continuously between the nucleus and the cytoplasm. Nevertheless, in our study we could not identify nuclear transport mechanism in N2 and its orthologous, still we do not know the exact effect of N2 protein on the nuclear transport mechanism.

## **FUTURE DIRECTIONS**

Previous studies from other research groups demonstrated that VACV N2 is a nuclear inhibitor of IRF3. However, the mechanism of action of N2 on IRF3 remains unknown. But also, they showed that N2 had no effect on IRF3 phosphorylation or translocation into the nucleus (47). Therefore, further work is needed to determine if N2 is disrupting the assembly or function of the transcriptional complex inside the nucleus. In this context, having an identified N2 functional NLS will allow characterization of N2 nuclear import in immune system evasion. Moreover, we can investigate if N2 protein has a NES for nuclear export. To this end we can compare subcellular localization of VACV N2 wild type-YFP and its substitution derivative in the presence or absence of LMB, an inhibitor of the nuclear export receptor CRM1, which leads to the further understanding of the function of N2 upon exit from the nucleus. Moreover, the identification of VACV N2 substitution derivative in impairing the active nuclear transport will allow us to perform functional assay to evaluate the role of protein nuclear transport in inhibition of host innate immune response. Crystallization assays are underway by the group of prof. Jade Forwood (Wagga Wagga University), aimed at resolving the 3D structure of N2 NLS bip (**IRKRPNQHHTIDLFKKIKRT**) in complex with  $\text{IMP}\alpha$  as well as binding assay to calculate  $K_d$  with fluorescence polarization.

## **Vaccinia virus A19 protein translocate to the nucleus of infected cells relying on $\text{IMP}\alpha$ $\beta$**

A19 is a small protein (9-kDa) which is conserved among *chordopoxviruses*, Amino acid sequences of A19 orthologs from each representative genus of *chordopoxviruses* were compared using the clustalW program and indicated conservation of cNLSs among all of them (50). Previous Immunofluorescence experiment on replication competent recombinant VACV A19 with a HA tag, determined A19 was distributed in the nucleus, while the activity of their NLS and nuclear transport mode were not characterized. By generating plasmids encoding full-length A19 proteins fused with YFP or substitution derivatives whereby basic residues were replaced with Alanine, following transfecting HEK293A cells with such plasmids and then quantitative analysis by CLSM, we confirmed previous observations of other groups regarding nuclear distribution of A19, and implying the YFP tag does not alter its subcellular localization. Importantly, our results under different conditions such as substitution of basic residues in cNLS with Alanine, ATP depletion and co-

expression with bimax2 inhibitor, clearly shows that Vaccinia virus A19 actively translocate into the nucleus relying on IMP $\alpha$   $\beta$ . Since substitution of its cNLS basic residues with Alanine, ATP depletion and presence of bimax2 completely impaired nuclear accumulation. Here we showed for the first time while A19 protein with a predicted mass 9-kDa from cytoplasmic compartment could passively enter the nucleus contains the active monopartite cNLS that has a crucial role for nuclear transport of A19 during the infection.

## **FUTURE DIRECTIONS**

The conserve active cNLSs in A19 protein must have an important role during the infection. Alternatively, Crystallization assays are underway by the group of prof. Jade Forwood (Wagga Wagga University), confirmed that A19 interact with IMP $\alpha$ . It may indicate that A19 binding to IMPs plays some role in modulating or interfering with the activities of protein nuclear transport systems. In future, we can examine if A19 can alter hosts nuclear transport system. To this end, we can investigate the effects of A19 on nuclear translocation of reporter proteins such as host polymerase II.

Overall, our research let us to identify new functional, highly conserved NLS in viral proteins which can target nuclear import pathways, subsequently manipulating innate immune response, in two vaccinia virus proteins, A19 and N2, and their orthologues, or to exploit host nuclear transport to gain access to the nucleus in order to guarantee their life cycle in all HPyVs LTA.

## 5- Figures and Legends

Figure 5.1

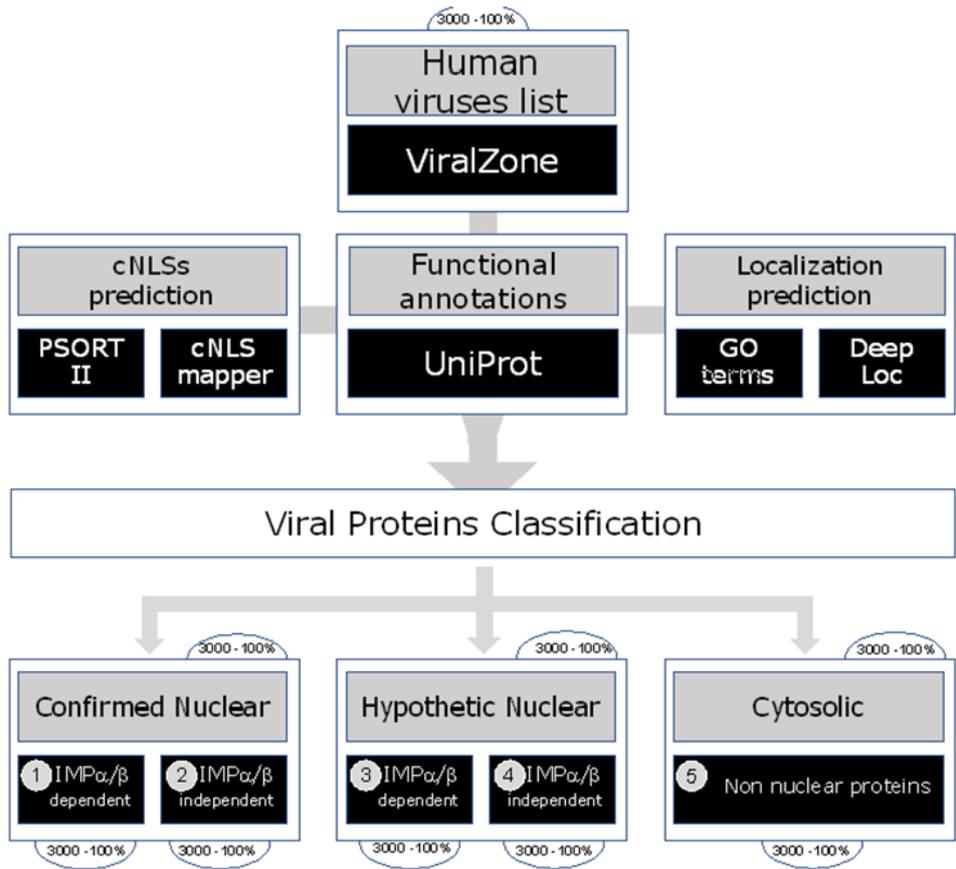


Figure 5.2

Protein	BKPyV	JCPyV	KIPyV	MCPyV	WUPyV
LT	confirmed nuclear, cNLS	hypothetic nuclear, cNLS	confirmed nuclear, cNLS	hypothetic nuclear, cNLS	hypothetic nuclear, cNLS
sT	confirmed nuclear, cNLS	cytosolic	confirmed nuclear, no cNLS	cytosolic	cytosolic
Ag	cytosolic	confirmed nuclear, cNLS	protein not found	protein not found	protein not found
VP1	confirmed nuclear, cNLS	confirmed nuclear, cNLS	confirmed nuclear, no cNLS	hypothetic nuclear, cNLS	cytosolic
VP2	confirmed nuclear, cNLS	hypothetic nuclear, cNLS	confirmed nuclear, no cNLS	cytosolic	cytosolic
VP3	protein not found	protein not found	protein not found	protein not found	cytosolic

Figure 5.3

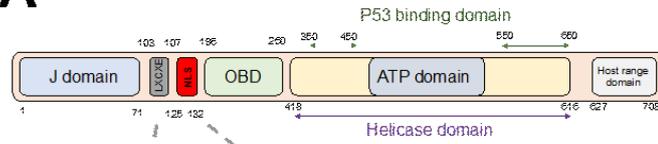
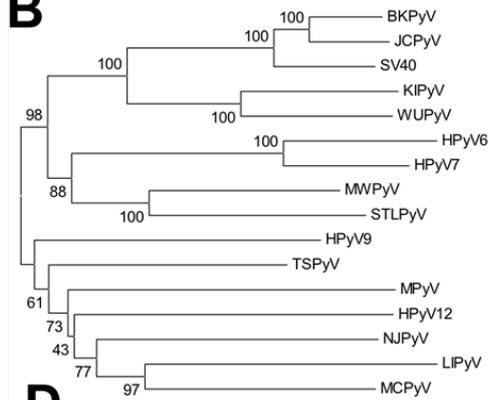
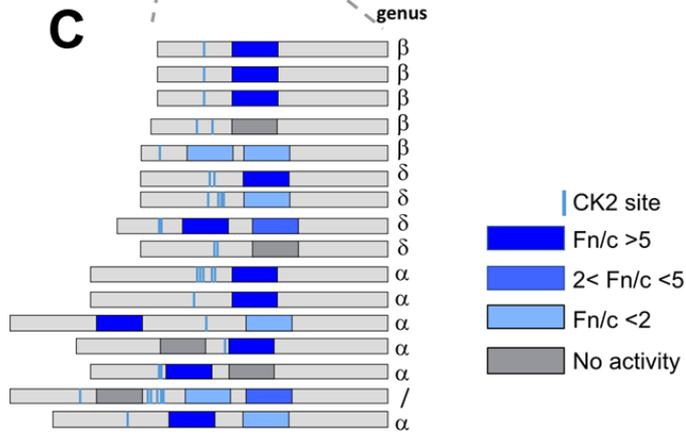
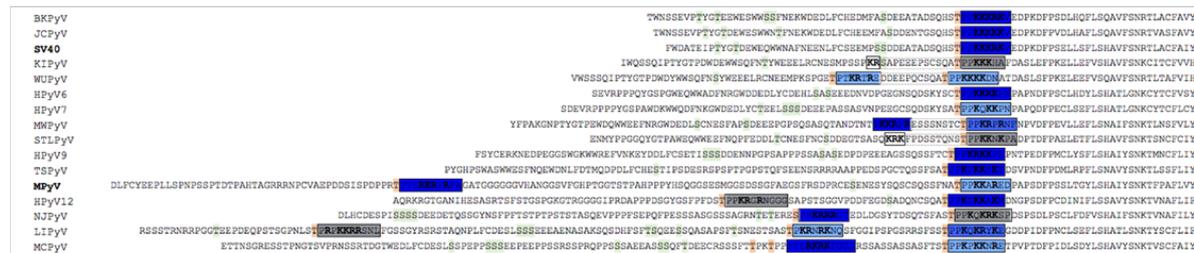
**A****B****C****D**

Figure 5.4

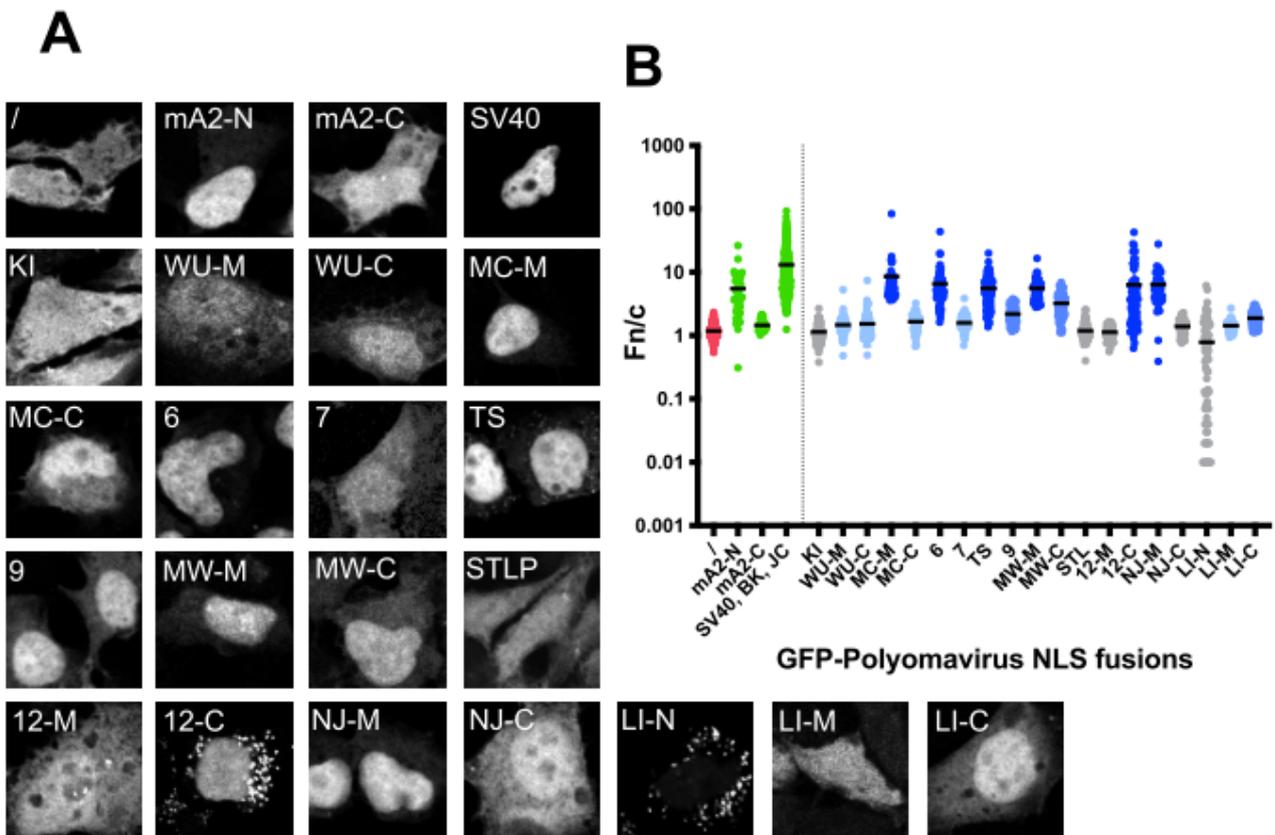


Figure 5.5

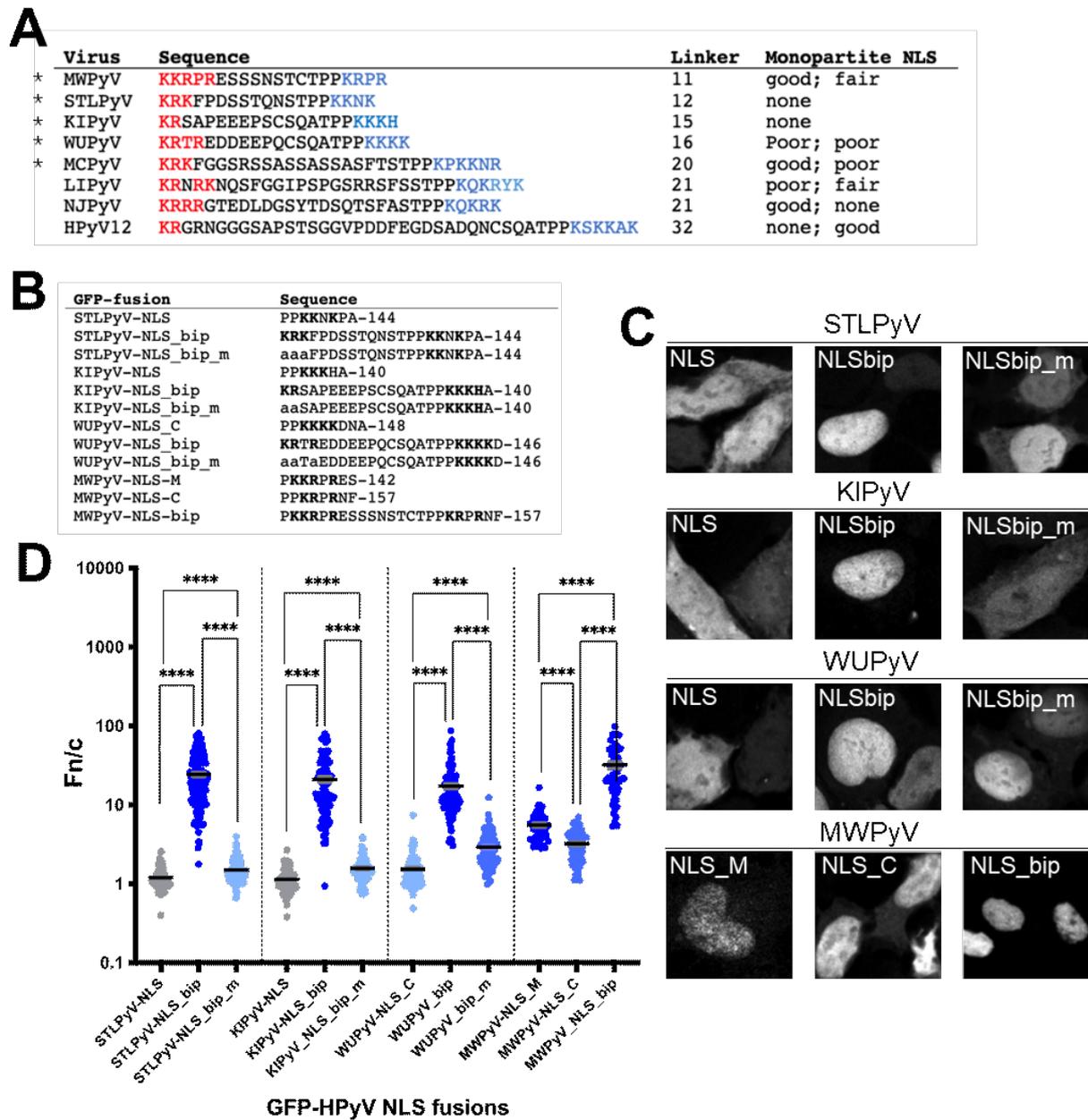


Figure 5.6

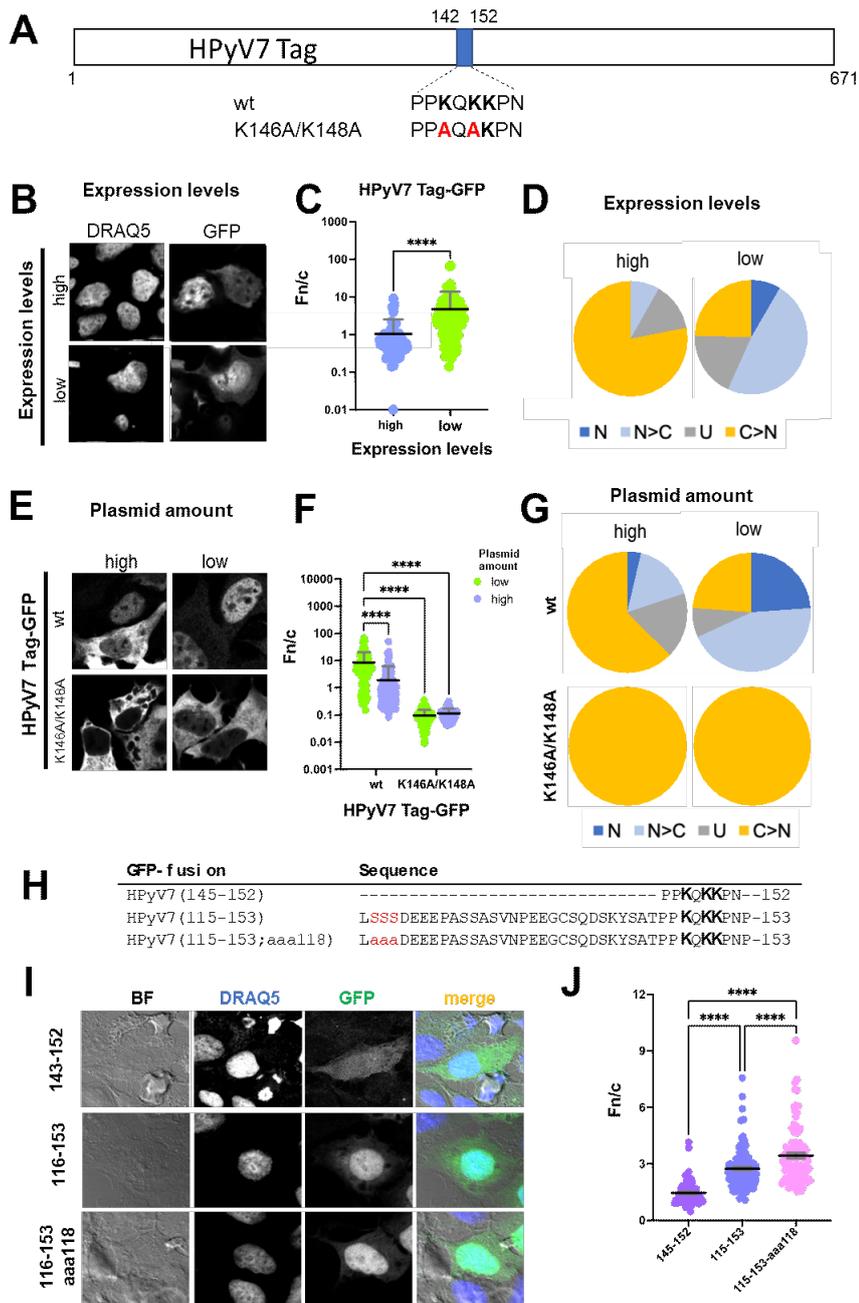


Figure 5.7

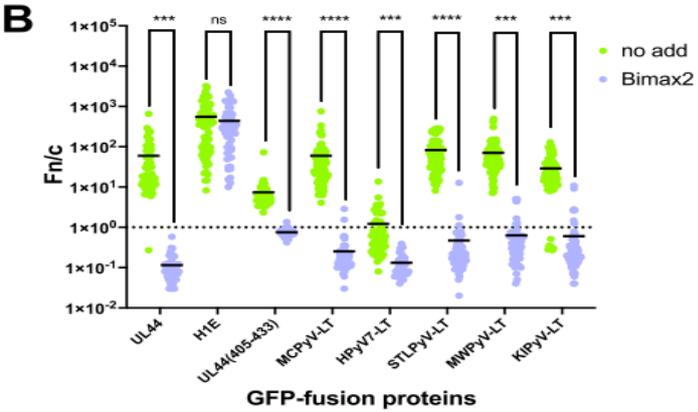
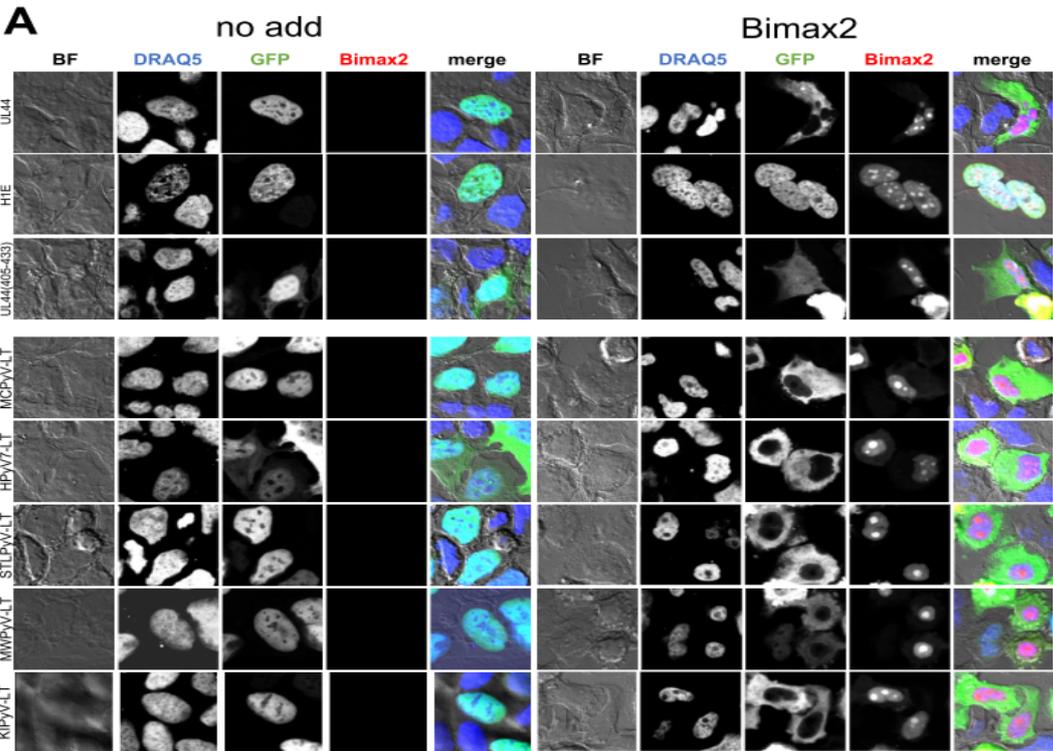
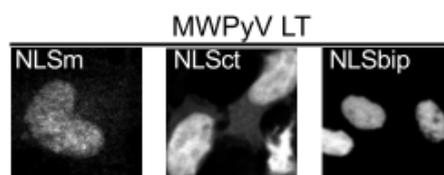


Figure 5.8

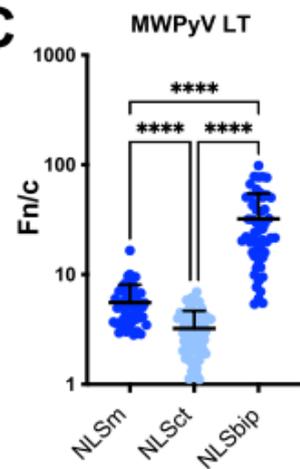
**A**

GFP-HPyVs LT NLS fusions		Sequence
MWPyV	NLSm	--PKKRPRES-----142
MWPyV	NLSct	-----PPKRPRNF-157
MWPyV	NLSbip	--PKKRPRESSSNSTCTPPKRPRNF-157

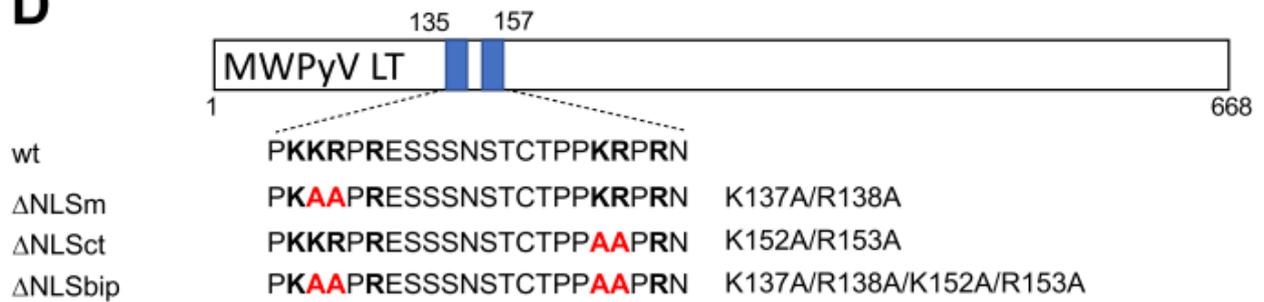
**B**



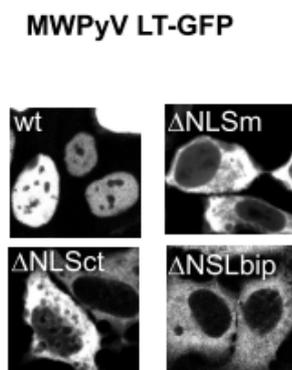
**C**



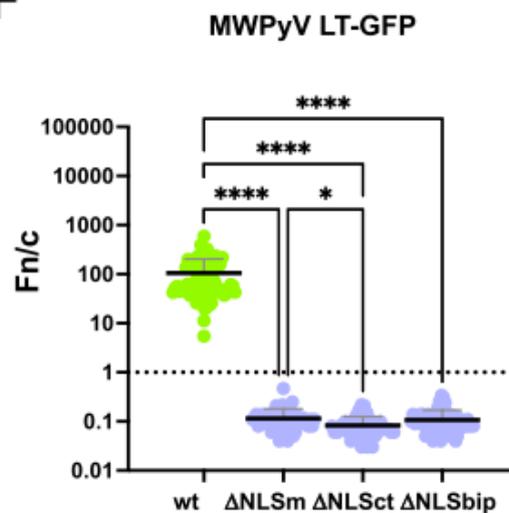
**D**



**E**



**F**



**G**

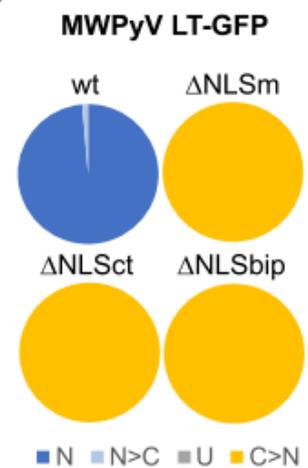
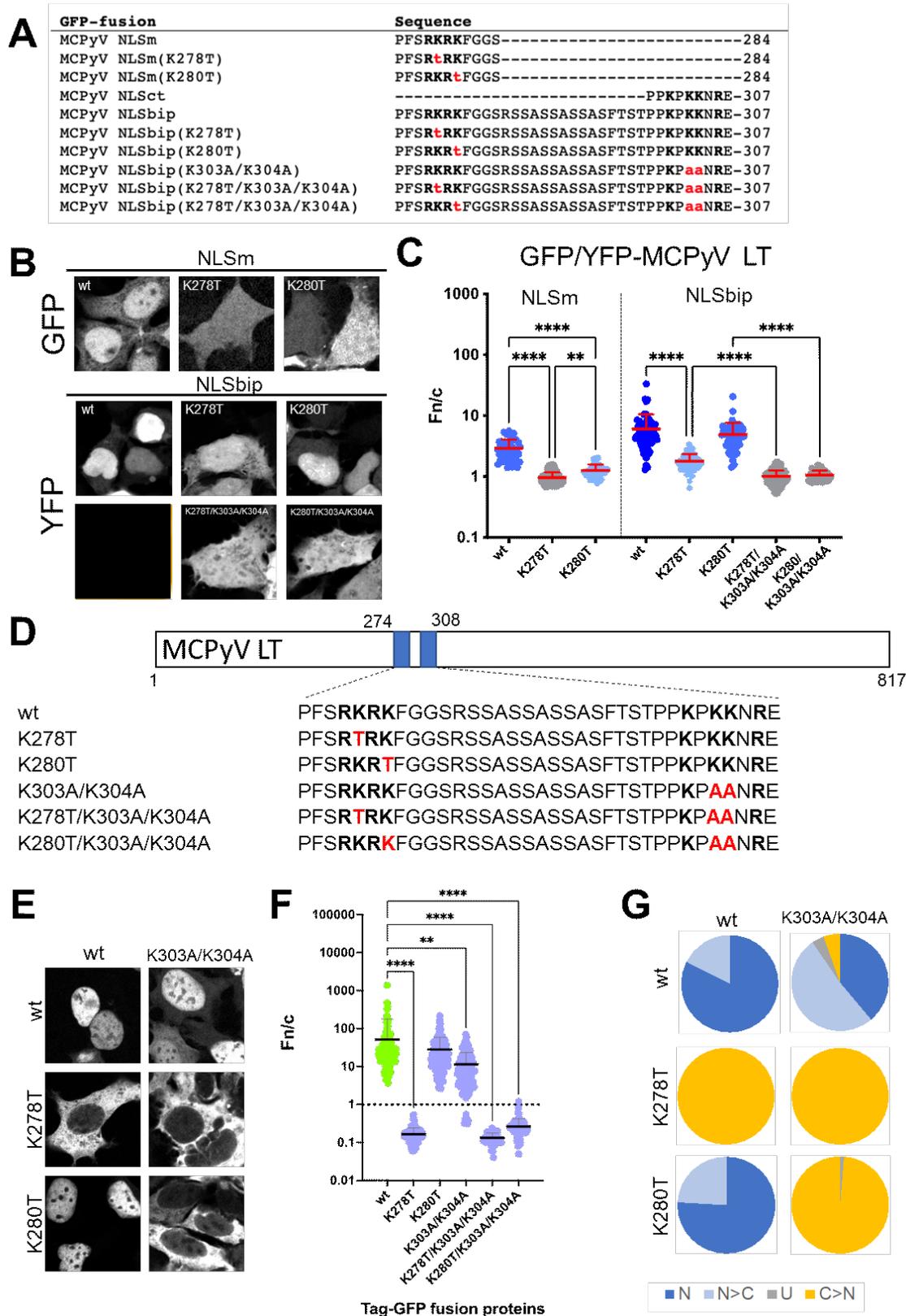
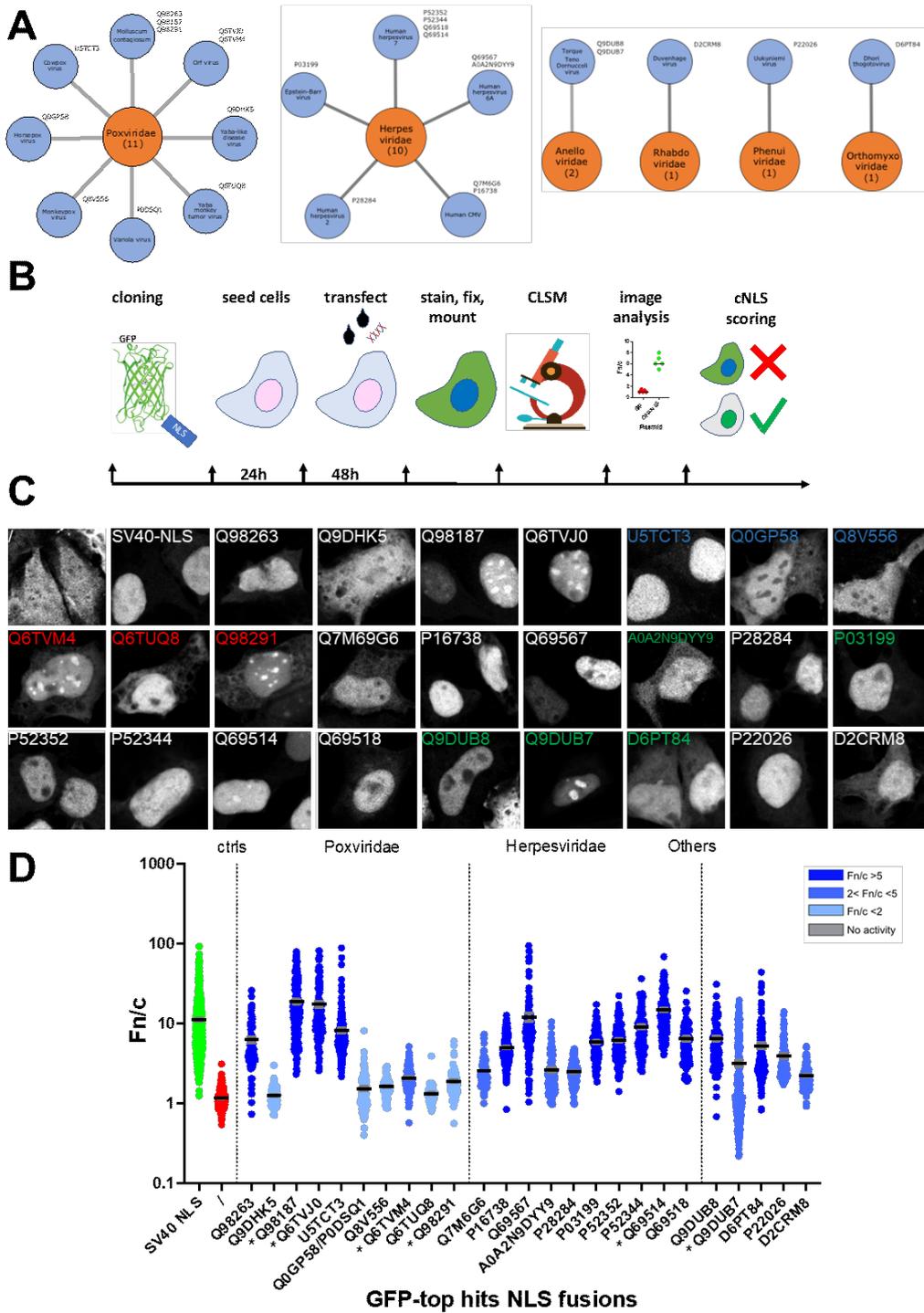


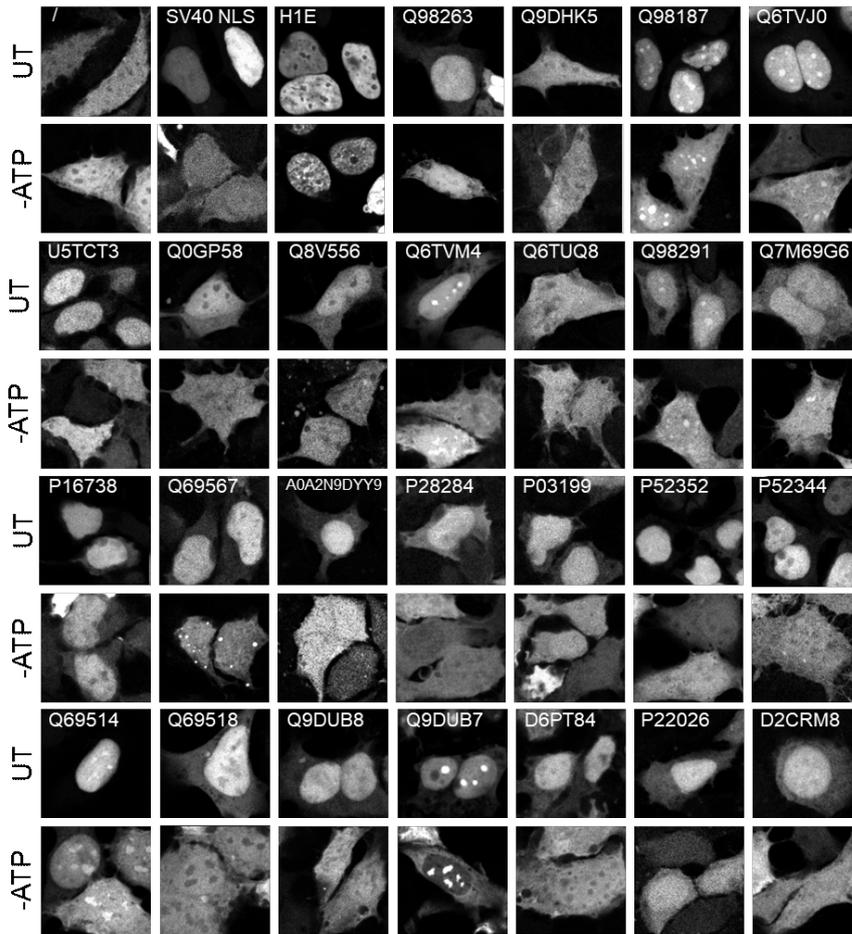
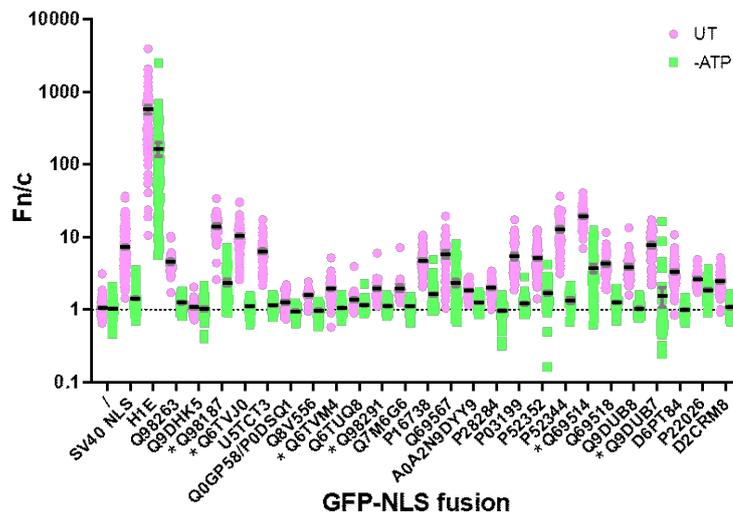
Figure 5.9



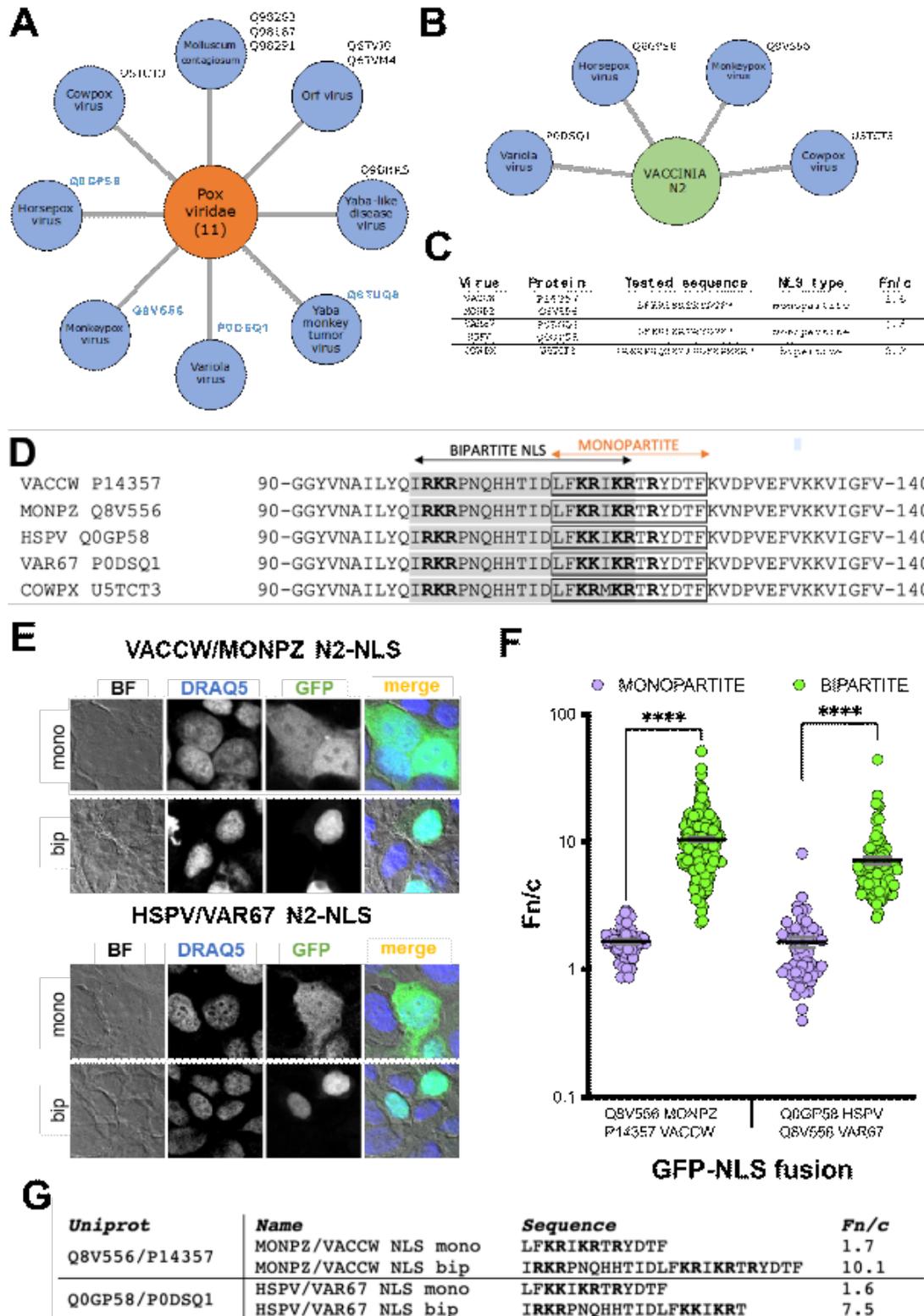
**Figure 5.10**



**Figure 5.11**

**A****B**

**Figure 5.12**



**Figure 5.13**

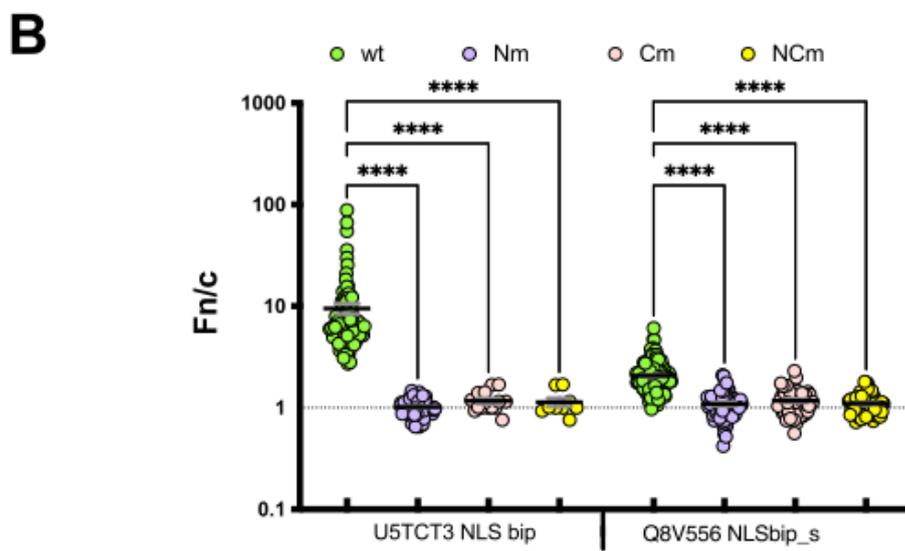
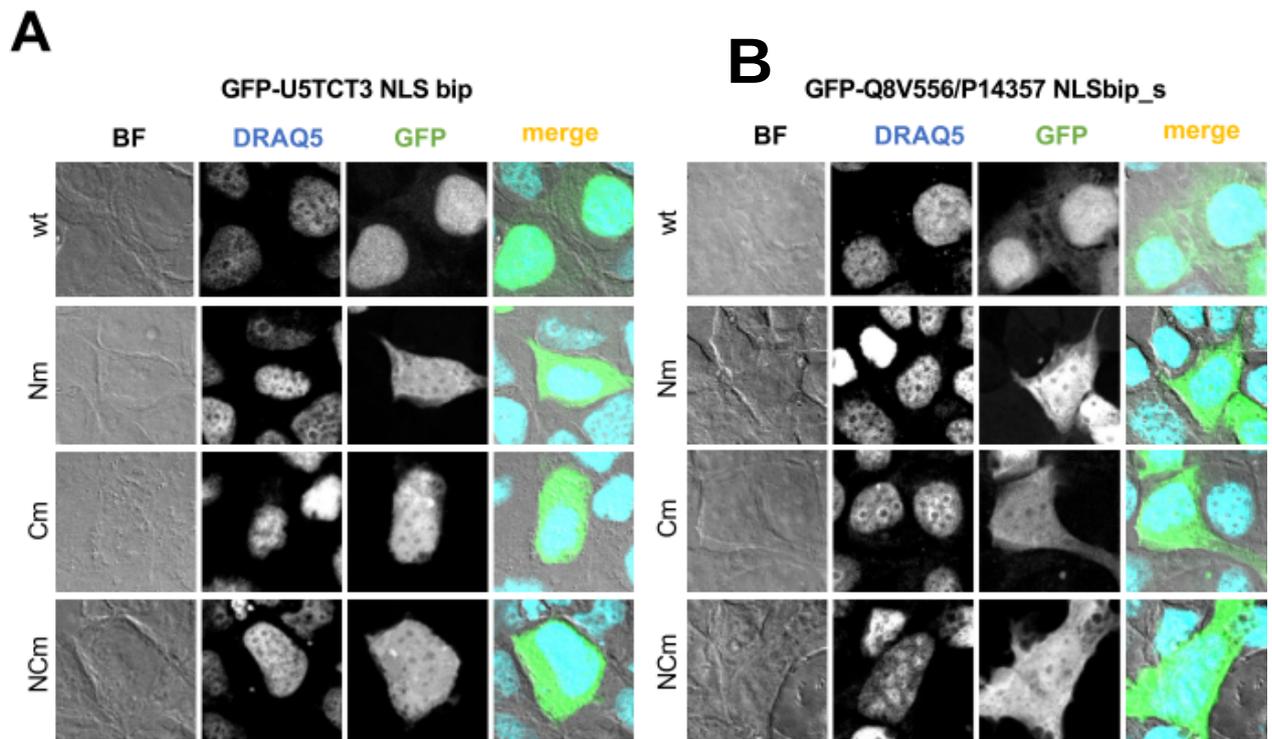


Figure 5.14

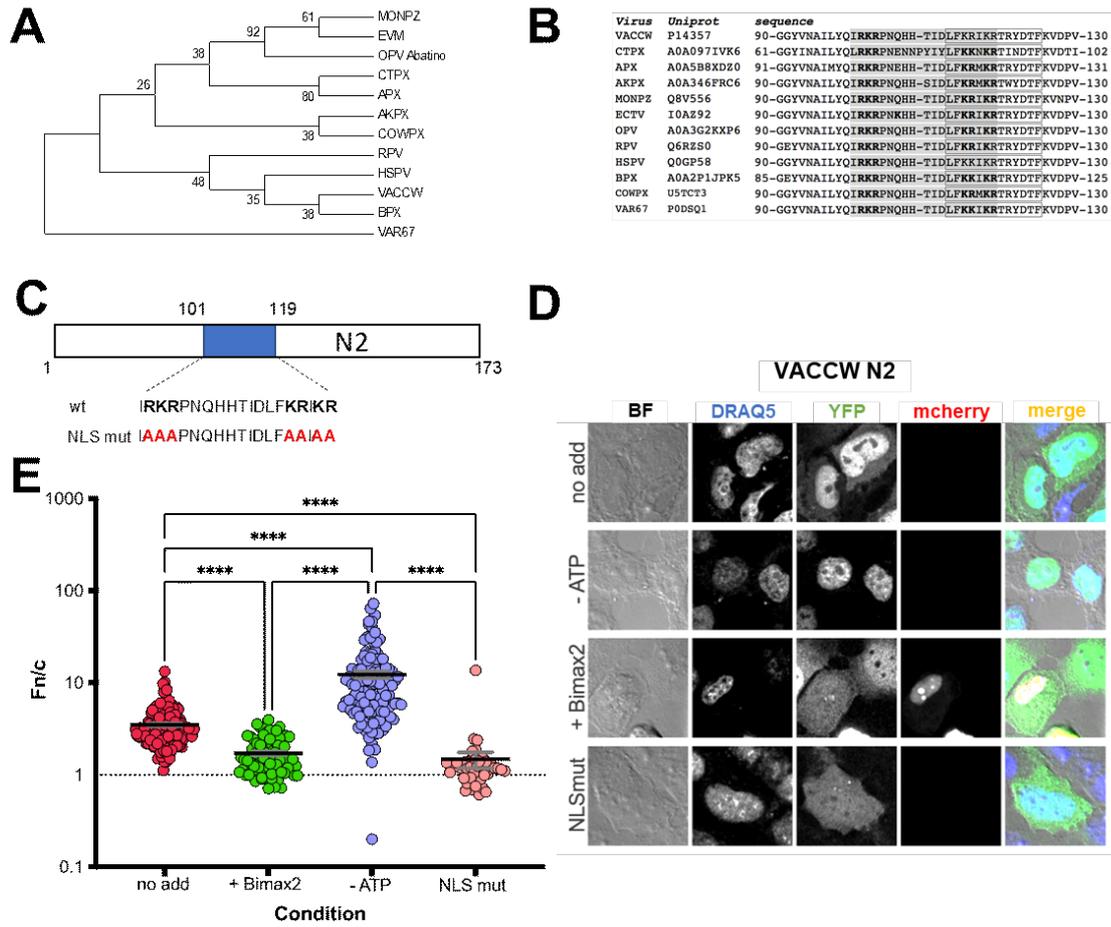
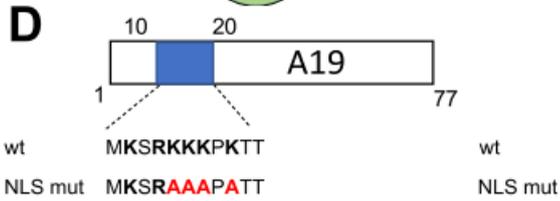
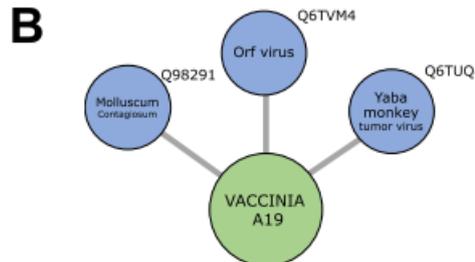
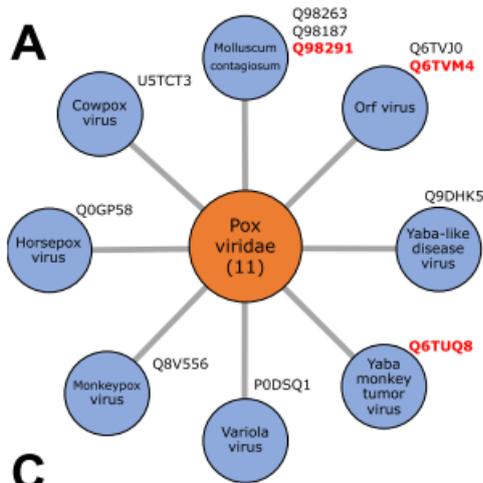


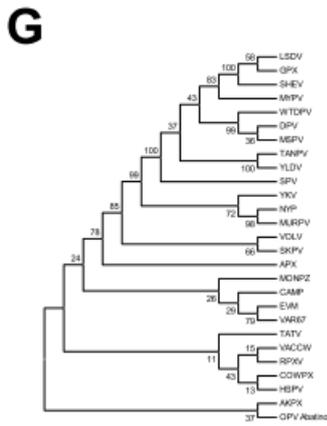
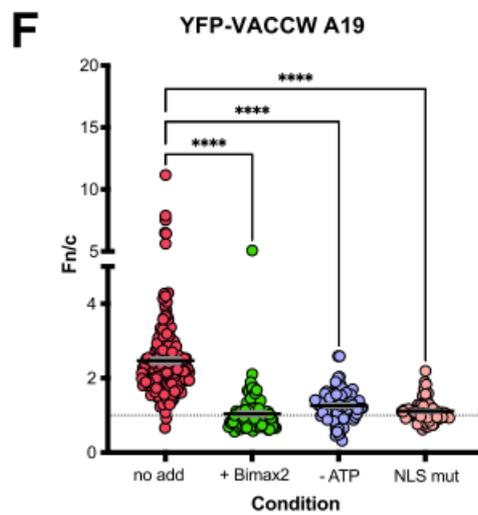
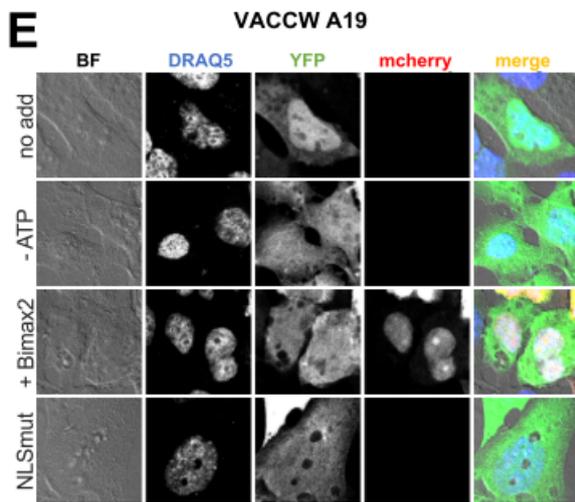
Figure 5.15



**C**

```

VACCW P68714 -----MDSTNVRSGMKSRKKKPKTTVIDDDDCMTCACQSKLVKISDITKVELDYINTMR-GNTLACAACGSSLLKLLNDFAS 77
YMTV5 Q6TUQ8 -----MDDMGAAKRRKKRPK-IATQNNNDVCVTCSSCHSKLVKVS DITKISLDELKLVAGKGNVLTCSACGSELRLLSGFVS 74
ORFSA Q6TVM4 MSAVKAKAGAGGSGGVVDVAAGNKKRRRRRTVVVEDGD--PVCSSNSRLVSIKDVRDLSTLSGLAC--SSTLGAACGSGALTPLRDLAR 90
MVCV1 Q98291 -----MDVSGAKQRRKKRPRTTVEEPPADSCCTCSVCQSRLAFFSDVSKLQASLRMATPGPDTLHCAACGSAALCPLSEFAR 78
  
```



**H**

Virus	UniProt code	sequence
VACCW	P68714	MDSINVRSGHRSKKKKKKTVIDDDDCMTCACQSKLVKISDITKVELDYINTMRGNTLACAACGSSLLKLLNDFAS 77
YEV	G32115	--NDYKAGHKIKKKKKTVID--DCMTCVVCQSKLVKISDITKVELDYIQPTKRVNLTSCSSCGSTLRLNDFVH 73
NYP	A0A223FHH3	--NDSTKAGHKIKKKKKTVID--DCMTCVVCQSKLVKISDITKVELDYIQ--MGNLTLSGSAACGSELRLNDFVR 73
MUREV	A0A223FKK5	--NDSTKAGHKIKKKKKTVID--DCMTCVVCQSKLVKISDITKVELDYIQ--MGNLTLSGSAACGSELRLNDFVR 74
VOLV	A0A1C9K09	MEHTITRSGHRSKKKKKKTVID--DCMTCVVCQSKLVKISDITKVELDYINT--NRGTLTCAACGSSLLKLLDFVR 76
SKPV	A0A1C9K04	--NRVVRSGHRSKKKKKKTVID--DCMTCVVCQSKLVKISDITKVELDYINT--NRGTLTCAACGSSLLKLLDFVS 75
SYM	Q8JL96	MDSINVRSGHRSKKKKKKTVIDDDDCMTCVVCQSKLVKISDITKVELDYINT--NRGTLTCAACGSSLLKLLDFAS 77
NONP2	Q8V4VB	MDSINVRSGHRSKKKKKKTVIDDDDCMTCVVCQSKLVKISDITKVELDYINT--NRGTLTCAACGSSLLKLLDFAS 77
CONF2	U5TRK4	MDSINVRSGHRSKKKKKKTVIDDDDCMTCVVCQSKLVKISDITKVELDYINT--NRGTLTCAACGSSLLKLLDFAS 77
BBPV	Q0GDU5	MDSINVRSGHRSKKKKKKTVIDDDDCMTCVVCQSKLVKISDITKVELDYINT--NRGTLTCAACGSSLLKLLDFAS 77
RFXV	Q6K2G3	MDSINVRSGHRSKKKKKKTVIDDDDCMTCVVCQSKLVKISDITKVELDYINT--NRGTLTCAACGSSLLKLLDFAS 77
AKFX	A0A346F8B1	MDSINVRSGHRSKKKKKKTVIDDDDCMTCVVCQSKLVKISDITKVELDYINT--NRGTLTCAACGSSLLKLLDFAS 77
GPV	A0A585XA02	MDSINVRSGHRSKKKKKKTVIDDDDCMTCVVCQSKLVKISDITKVELDYINT--NRGTLTCAACGSSLLKLLDFAS 77
APX	A0A585XA02	MDSINVRSGHRSKKKKKKTVID--DCMTCVVCQSKLVKISDITKVELDYINT--NRGTLTCAACGSSLLKLLDFAS 76
VAK67	P33842	MDSINVRSGHRSKKKKKKTVID--DCMTCVVCQSKLVKISDITKVELDYINT--NRGTLTCAACGSSLLKLLDFAS 76
TATV	Q0NP70	MDSINVRSGHRSKKKKKKTVID--DCMTCVVCQSKLVKISDITKVELDYINT--NRGTLTCAACGSSLLKLLDFAS 76
MCNP2	A0A0K1LE09	MDSINVRSGHRSKKKKKKTVID--DCMTCVVCQSKLVKISDITKVELDYINT--NRGTLTCAACGSSLLKLLDFAS 76
CAMP	Q8V319	--MDTA--GSKKKKKTVID--DCMTCVVCQSKLVKISDITKVELDYINT--NRGTLTCAACGSSLLKLLDFAS 72
SPV	A7XCP2	--MDNLGSSKKKKKKTVID--DCMTCVVCQSKLVKISDITKVELDYINT--NRGTLTCAACGSSLLKLLDFAS 74
TANPV	A7XCP2	--MDNLGSSKKKKKKTVID--DCMTCVVCQSKLVKISDITKVELDYINT--NRGTLTCAACGSSLLKLLDFAS 74
YL5V	Q9D6K2	--MDNLGSSKKKKKKTVID--DCMTCVVCQSKLVKISDITKVELDYINT--NRGTLTCAACGSSLLKLLDFAS 74
TATV	Q08F62	--MDTGGSKIKKKKKTVID--DCMTCVVCQSKLVKISDITKVELDYINT--NRGTLTCAACGSSLLKLLDFAS 75
DPV	Q08F62	--MDTGGSKIKKKKKTVID--DCMTCVVCQSKLVKISDITKVELDYINT--NRGTLTCAACGSSLLKLLDFAS 75
MSPV	A0A357SW40	--MDTGGSKIKKKKKTVID--DCMTCVVCQSKLVKISDITKVELDYINT--NRGTLTCAACGSSLLKLLDFAS 75
WDFV	A0A216BQ45	--MDTGGSKIKKKKKTVID--DCMTCVVCQSKLVKISDITKVELDYINT--NRGTLTCAACGSSLLKLLDFAS 75
NPV	22A046	--MDTGGSKIKKKKKTVID--DCMTCVVCQSKLVKISDITKVELDYINT--NRGTLTCAACGSSLLKLLDFAS 73
HPV	A0A1C9LH04	--MDTGGSKIKKKKKTVID--DCMTCVVCQSKLVKISDITKVELDYINT--NRGTLTCAACGSSLLKLLDFAS 74
GPX	A0A1B2LPR1	--MDTGGSKIKKKKKTVID--DCMTCVVCQSKLVKISDITKVELDYINT--NRGTLTCAACGSSLLKLLDFAS 74
AKFX	A0A2H4EYB6	--MDTGGSKIKKKKKTVID--DCMTCVVCQSKLVKISDITKVELDYINT--NRGTLTCAACGSSLLKLLDFAS 74
SHEV	A0A2H4EYB6	--MDTGGSKIKKKKKTVID--DCMTCVVCQSKLVKISDITKVELDYINT--NRGTLTCAACGSSLLKLLDFAS 74

## Legends

**Figure 5.1. Schematic workflow of bioinformatics analysis performed in this study.** The proteomes of all human viruses present reported on ViralZone were downloaded from on Uniprot along with individual proteins annotations, their sequence was scanned with both cNLSs prediction softwares (PSORTII and cNLS mappers) and protein subcellular localization algorithms (DeepLoc and an ad hoc developed GO terms-based approach) in order to assign each protein to one of the following classes, based on their predicted subcellular localization and pathway used for nuclear import. (1) Confirmed nuclear viral proteins relying on the IMP $\alpha$   $\beta$  nuclear import pathway: annotated as nuclear in Uniprot and with a cNLS predicted by at least one algorithm. (2) Confirmed nuclear viral proteins not relying on the IMP $\alpha$   $\beta$  nuclear import pathway: annotated as nuclear on Uniprot but without a cNLS predicted by any algorithm. (3) Hypothetic nuclear viral proteins relying on the IMP $\alpha$   $\beta$  nuclear import pathway: not annotated as nuclear in Uniprot but with either at least one cNLS predicted by both algorithms or at least one cNLS predicted by one algorithms and nuclear localization predicted by at least one algorithm. (4) Hypothetic nuclear viral proteins not relying on the IMP $\alpha$   $\beta$  nuclear import pathway: not annotated as nuclear in Uniprot, with no cNLS predicted by any algorithm and nuclear localization predicted by at least one algorithm. (5) Non-nuclear proteins: not annotated as nuclear in Uniprot, cNLS predicted by one or less algorithms and no nuclear localization predicted by any algorithms.

**Figure 5.2. Comparative results of our bioinformatics analysis for LTA from different five HPyVs.** Large Tumor Antigens (LTA) from BK and KI *Polyomavirus* were classified as “confirmed nuclear with cNLSs”, while those for WU, JC and MCP were classified as “hypothetical nuclear proteins with cNLS”.

**Figure 5.3. Phylogenetic analysis of LTA from all HPyV performed using clustal W, along with the relative distribution of putative cNLSs.** (A) representation of HPyV LTs domains, using SV40 LTA amino acid positions as a reference. (B) The sequences from all known HPyV LTAs were retrained from uniprot and analysis using clustal W. (C) The protein sequence is represented as pale gray boxes, and cNLS are shown as rectangles colored according to the Fn/c relative to the quantitative analysis. Putative CK2 phosphorylation sites are represented as blue vertical lines. (D) The single letter amino acid code is used, with NLSs highlighted in different colors as in C according to their activity, and shorter stretches of basic amino acids boxed. Basic amino acids are in boldface. Cdc2 putative phosphorylation sites are highlighted in pink, whereas CK2 phosphorylation sites in green.

**Figure 5.4. Subcellular localization of all cNLSs present on Polyomaviruses LTAs.** Putative cNLS sequences of our interest were fused the C-terminal of GFP in Mammalian expression plasmids and transfected in HEK293-A cells for 48 h before processing cells for IF and quantitative CLSM.

Cell micrographs was used to calculate the Fn/c relative to single cells expressing the indicated GFP-fusion proteins. Each filled circle corresponds to quantification of a single cell and is colored as in **Figure 5.5. STLPyV and KIPyV LT NLS are bipartite cNLSs and monopartite cNLS in WUPyV can potentially synergize as bipartite cNLS.** (c) 48h p.t. subcellular localization of GFP fusion proteins was analyzed by quantitative CLSM. (D) cell micrographs also were used to quantify the Fn/c relative to single cells expressing the indicated fusion proteins. Data shown are the mean  $\pm$  standard error of the mean relative to singles cells, with indicated the significance scores from Brown–Forsythe and Welch Anova test.

**Figure 5.6. upstream sequences contain conserved CK2 phosphorylation sites has a role to enhance IMP $\alpha$   $\beta$  interaction and therefore nuclear import.** (A) HPyV-NLS upstream sequences contain conserved CK2 phosphorylation sites. (B) The indicated GFP fusions to HPyV7 residues were used to transfect HEK293A cells. 48 h p.t. cells were fixed, and the subcellular localization analysed by quantitative CLSM. (C) Micrographs were used to calculate the Fn/c relative to single cells. Data shown are single measurements (coloured circles), along with mean (black horizontal bars), standard error of the mean (grey vertical bars) and test of significance performed by one way ANOVA.

**Figure 5.7. Nuclear accumulation of HPyV LTA antigens is dependent on the IMP $\alpha$   $\beta$  heterodimer.** (A) HEK293A cells were transfected with plasmids encoding the indicated GFP-fusion proteins in the absence (left panels) or presence (right panels) of a plasmid encoding for mCherry-Bimax2. At 48 h p.t. cells were processed for IF and subcellular localization of fusion proteins analyzed by quantitative CLSM. The bright field (BF), GFP (GFP) and mcherry (Bimax2) channels are shown, along with a merged image (merge) relative to representative phenotypes. (B) Micrographs such as those shown in (A) were quantitatively analyzed to calculate the Fn/c relative to individual cells. Data shown are mean  $\pm$  standard error of the mean relative to each GFP-fusion are shown with the results of the t-Student test for significance between expression in the absence (green circles) or in the presence (purple circles) of mcherry-Bimax2

**Figure 5.8. Monopartite cNLSs in MWPyV LTA is a potentially bipartite cNLS.** (A) The sequence of the indicated HPyV Tags (first column) forming putative bipartite cNLS is shown, with upstream residues potentially binding to the minor IMP $\alpha$  site shown in red and downstream residues potentially binding to the major IMP $\alpha$  site shown in blue. The single letter amino acid code is used (second column). The length of the linker separating the two basic stretches is indicated (third column) along with the activity of the monopartite NLSs (fourth column). (B) The indicated sequences from HPyVs corresponding to monopartite or bipartite putative cNLSs were cloned to the C-terminus of GFP expressing plasmids and used to transfect HEK293A cells. (C) 48h p.t. cells were

processed for IF and the subcellular localization of GFP fusion proteins was analyzed by quantitative CLSM. (D) cell micrographs such as those shown in (C) were used to quantify the Fn/c relative to single cells expressing the indicated fusion proteins. Data shown are the mean and standard deviation of the mean relative to single cells, with indicated the significance scores from Brown–Forsythe and Welch Anova test

**Figure 5.9. Monopartite cNLSs in MCPYV LTA can potentially synergize as bipartite cNLSs.**

The indicated sequences from MCPyV corresponding to monopartite or bipartite putative cNLSs were cloned to the C-terminus of GFP expressing plasmids and used to transfect HEK293A cells. 48h p.t. cells were processed for IF and the subcellular localization of GFP fusion proteins was analyzed by quantitative CLSM. Micrographs such as those shown in (B) were quantitatively analyzed to calculate the Fn/c relative to individual cells. Data shown are mean  $\pm$  standard error of the mean relative to each GFP-fusion are shown with the results of the t-Student test for significance.

**Figure 5.10. Functional validation of the 26 top ranking viral cNLS newly identified in our study.** (A) Representative images of the GFP channels are shown, along with the Uniprot code of each protein bearing the cNLS analysed. (B) Micrograms such as those shown in (A) were quantitatively analyzed to calculate the Fn/c value relative to single cells expressing the indicated GFP-fusion proteins. Each filled circle corresponds to quantification of a single cell and colored according to mean the Fn/c of each fusion protein. GFP alone (red circles) was also expressed as a negative control.

**Figure 5.11. ATP depletion confirmed Viral cNLS active nuclear import abilities.** (A) Plasmids encoding the indicated GFP fusion proteins were transfected in HEK293A cells, and 48h later were either left untreated (UT) or incubated for 1h in ATP depletion media (-ATP) before being processed for IF and CLSM to quantitatively analyse the subcellular localization of each GFP fusion protein at the single cell level. Representative images of the GFP channel relative to each of the indicated GFP fusion protein are shown. (B) Micrographs such as those shown in (A) were quantitatively analysed to calculate the Fn/c value relative to the GFP channel of single cells under the indicated conditions. Data shown are single values, along with the mean (black horizontal bars)  $\pm$  standard error of the mean (grey vertical bars) relative to single cells measurements ( $n \geq 12$ ).

**Figure 5.12. Poxviridea N2 protein contains bipartite cNLS.** (A) First analysis indicated strong nuclear accumulation for COWPX-cNLS N2 protein while for Horse Pox virus (HSPV), Monkey Pox Virus (MNPZ), and VAR67 the signal was very weak. (B) Sequence analysis of such cNLSs revealed some basic residues upstream of mono partite cNLS. (C) The indicated sequences from MONPZ, and HSPV corresponding to bipartite putative cNLSs were cloned to the C-terminus of GFP expressing plasmids and used to transfect HEK293A cells. 48h p.t. cells were processed for IF and the

subcellular localization of GFP fusion proteins was analyzed by quantitative CLSM, and the ratio of F n/c for bipartite putative cNLSs were compared with monopartite. (D) Micrographs such as those shown in (C) were quantitatively analyzed to calculate the Fn/c relative to individual cells.

**Figure 5.13. Each basic cluster in U5TCT3 and Q8V556 NLSs has a crucial role in protein nuclear transport.** (A) The indicated GFP fusions to (Nm, Cm, NCm) residues in U5TCT3 cNLS, and Q8V556 NLS were used to transfect HEK293A cells. 48 h p.t. cells were fixed and processed for IF, and the subcellular localization analysed by quantitative CLSM. (B) Micrographs were used to calculate the Fn/c relative to single cells.

**Figure 5.14. Nuclear accumulation of VACCW N2 protein is due to active transport dependent on the IMP $\alpha$   $\beta$  heterodimer.** (B) HEK293A cells were transfected with plasmids encoding the indicated GFP-fusion proteins in the absence or presence of a plasmid encoding for mCherry-Bimax2. At 48 h p.t. cells were processed for IF and subcellular localization of fusion proteins analyzed by quantitative CLSM, and for ATP depletion experiment Plasmids encoding the indicated GFP fusion proteins were transfected in HEK293A cells, and 48h later were either left untreated (UT) or incubated for 1h in ATP depletion media (-ATP) before being processed for IF and CLSM. (C) Micrographs were used to calculate the Fn/c relative to single cells.

**Figure 5.15. nuclear location of VACCW A19 protein might be a function the putative cNLS rather than relatively small size.** HEK293A cells were transfected with plasmids encoding the indicated GFP-fusion proteins in the absence or presence of a plasmid encoding for mCherry-Bimax2. At 48 h p.t. cells were processed for subcellular localization of fusion proteins analyzed by quantitative CLSM, and for ATP depletion experiment Plasmids encoding the indicated GFP fusion proteins were transfected in HEK293A cells, and 48h later were either left untreated (UT) or incubated for 1h in ATP depletion media (-ATP) before being processed for IF and CLSM. Micrographs were used to calculate the Fn/c relative to single cells.

**6-Appendix**

**Table 6.1. List of primers Utilized for PCR Gateway® Technology cloning**

Primer	Primer sequence (5' 3')	Tm (°C)
NLS 123 HP7_Tag_C_ EattB1	GGGGACAAGTTTGTACAAAAAAGCAGGCTTACTGAGCAG CTCCGACGAGGAAGAAC	86.8
NLS 124 HP7_Tag_C_ EattB2	GGGGACCACTTTGTACAAGAAAGCTGGGTCCTAAGGATTG GGCTTCTTCTGCTTAGG	86.4
NLS 125 HP7_Tag_C_ EattB1m	GGGGACAAGTTTGTACAAAAAAGCAGGCTTACTGGCCGCC GCCGACGAGGAAGAAC	90.7
NLS 128 MCPyV_ bip_ FWD	GGGGACAAGTTTGTACAAAAAAGCAGGCTCTCCCCATTCT CAAGAAAGCGAAAAAT	85.0
NLS 129 MCPyV_ bip_ REV	GGGGACCACTTTGTACAAGAAAGCTGGGTCCTATTCTCTG TTCTTTTTTGGCTTTGG	85.3
NLS 138 MCPyV_ bip_ F_K278T	GGGGACAAGTTTGTACAAAAAAGCAGGCTCTCCTCCATTCTC AAGAACCCGAAAAAT	85.8
NLS 139 MCPyV_ bip_ F_K280T	GGGGACAAGTTTGTACAAAAAAGCCAGGCTCTCCTCCATCTC AAGAAAGCGAACCTTT	86.0
NLS 140 MCPyV_ bip_ cmut_rev	GGGGACCACTTTGTACAAGAAAGCTGGGTCCTATTCTCTGTT GGCGGCTGCTTTGG	89.9

**Table 6.2. List of primers Utilized for Stratagene's QuikChange site-directed mutagenesis**

PRIMER	PRIMER SEQUENCE (5' 3')	Tm (°C)
NLS 143 STPyVN-FWD	GGTGCTGCTGTTCGGG GAACGCTG CGCCTGGCTGGCGCT TGTGCC	84.1
NLS 144 STPyVN-REV	GGGCACAAGCGCCAGCCAGGCGGCAGCGTTCCTCCGACAG CAGCACC	84.1
NLS 145 KIPyVN-FWD	CTTCCTCAGGGGCGCTTGCCCCAGGGCTGCTAGGCATGC	82.3
NLS 146 KIPyVN-REV	GCATGCCTAGCAGCCCTGGGGCAAGCGCCCTGAGGAAG	82.3
NLS 149 MWPyVN-FWD	GAGCTGCTCTCTCTGGGCGCCGCCTGGGTGTGTTGGTATC	81.3
NLS 150 MWPyVN-REV	GATACCAACACACCCAAGGCGGCGCCAGAGAGAGCAGCTC	81.3
NLS 151 MWPyVC-FWD	AAGTTCGGGGCGCCGCTGGAGGTGTACAGGTGCTGTTGG	81.3
NLS 152 MWPyVC-REV	CCAACAGCACCTGTACACCTCCAGCGGCGCCCCGGAAGTT	81.3
NLS 153 MCPyV-K278T-FWD	CCAAATTTTCGCGTTCTTGAGAATGGAGGAGGGG	79.2
NLS 154 MCPyV-K278T-REV	CCCCTCCTCCATTCTCAAGAACGCGAAAATTTGG	79.2

NLS 155 MCPyV-K280T-FWD	TTCGGGACCCCCCAAATGTTTCGCTTTCTTGAGAAT	79.3
NLS 156 MCPyV-K280T-REV	ATTCTCAAGAAAGCGAACATTTGGGGGGTCCCGAA	79.3
NLS 157 MCPyV-C-FWD	GTAGGAACAGGAGTTTCTCTGTTTCGCTGCTGGCTTTGGTGGA GTGCTTGTA	81.3
NLS 158 MCPyV-C-REV	TTTACAAGCACTCCACCAAAGCCAGCAGCGAACAGAGAA ACTCCTGTTCTAC	81.3
NLS 159 HPyV7-FWD	GAGCAGGATTGGGCTTCGCCTGCGCAGGAGGTGTGGCGCTGT	83.3
NLS 160 HPyV7-REV	ACAGCGCCACACCTCCTGCGCAGGCGAAGCCCAATCCTGCTC	83.3

\*\*Tm calculated using OligoAnalyzer™ Tool IDT Tm calculator

**Table 6.3 List of primers Utilized for TOPO cloning**

PRIMER	Primer sequence (5' 3')
NLS 1/2 SV40_tag_C-FWD/REV	FWD: CCCCCCAAAAAGAAAAGGAAGGTCTAAGTCGACA REV: GTCGACTTAGACCTTCCTTTTCTTTTGGGGGA
NLS 3/4 STLP_tag_C-FWD/REV	CCCCCAAGAAGAACAAGCCCGCCTAAAGATCTA AGATCTTTAGGCGGGCTTGTCTTCTTGGGGGA
NLS 5/6 MWP_tag_C_FWD/REV	CCCCCAAGAGACCCAGGAATTCTAAAGATCTA AGATCTTTAGAAGTTCCTGGGTCTCTTGGGGGA
NLS 7/8 HP6_tag_C-FWD/REV	CCCCCAAGAAGAGGAAGCCCAACTAAGTCGACA GTCGACTTAGTTGGGCTTCCTCTTCTTGGGGGA
NLS 9/10 NJP_tag_C-FWD/REV	CCCCCAAGCAGAAGAGGAAGAGCCCTAAAAGCTTA AAGCTTTTAGGGGCTTCCTCTTCTGCTTGGGGGA
NLS 11/12 HP9_tag_C-FWD/REV	CCCCCAAGAGGAAGAAGCCCGAGTAAAGATCTA AGATCTTTACTCGGGCTTCTCCTCTTGGGGGA
NLS 13/14 MP_tag_N-FWD/REV	CCCGTGAGCAGGAAGAGACCCAGACCCGCCTAAAGATCTA AGATCTTTAGGCGGGTCTGGGTCTTCTCCTGCTCACGGGA
NLS 15/16	CCCAAGAGGAACAGGAAGAACCAGTAAAAGCTTA

LIP_tag_M-FwD/REV	AAGCTTTTACTGGTTCTTCTGTTCTTCTTGGGA
NLS 17/18 STLP_tag_Cb-FwD/REV	AAGAGGAAGTTCCCCGACAGCAGCACCCAGAACAGCACCCCCCAAGAAGAACAAGCCCGCCTAAA TTAGGCGGGCTTGTCTTCTTGGGGGGGTGCTGTTCTGGGTGCTGCTGTCGGGGAACTTCTCTTA
NLS 19/20 MWP_tag_M-FwD/REV	CCCAAGAAGAGACCCAGGGAGAGTTAAGTCGACA GTCGACTTAACTCTCCCTGGGTCTTCTTGGGA
NLS 21/22 NJP_tag_M-FwD/REV	CCCCCAAGAGGAGGAGAGGCACCTAAAAGCTTA AAGCTTTTAGGTGCCTCTCCTCTTGGGGGGA
NLS 23/24 LIP_tag_N-FwD/REV	CCCAGACCCAAGAAGAGGAGGAGCAACCTGTAAAAGCTTA AAGCTTTTACAGGTTGCTCCTCTTCTTGGGTCTGGGA
NLS 25/26 LIP_tag_C_FwD/REV	CCCCCAAGCAGAAGAGGTACAAGGAGTAAAAGCTTA AAGCTTTTACTCCTTGTACCTCTTCTGCTTGGGGGGA
NLS 27/28 MP_tag_C-FwD/REV	CCCCCAAGAAGGCTAGAGAAGACTAAAGATCTA AGATCTTTAGTCTTCTCTAGCCTTCTTGGGGGGA
NLS 29/30 KIP_tag_C-FwD/REV	CCCCCAAGAAGAAACATGCTTAAGTCGACA GTCGACTTAAGCATGTTTCTTCTTGGGGGA
NLS 31/32 TSP_tag_C FwD/REV	CCCCCAAGCCCAAGAAGAGCAAGTACTAAAGATCTA AGATCTTTAGTACTTGCTCTTCTTGGGCTTGGGGGGA
NLS 33/34 WUP_tag_M-FwD/REV	CCCACCAAGAGGACCAGGGAGTAAGTCGACA GTCGACTTACTCCCTGGTCCTTCTTGGTGGGA
NLS 35/36 WUP_tag_C-FwD/REV	CCCCCAAGAAGAAGAAGGACAACGCCTAAGTCGACA GTCGACTTAGGCGTTGTCTTCTTCTTCTTGGGGGGA
NLS 37/38 HP7_tag_C-FwD/REV	CCCCCAAGCAGAAGAAGCCCAACTAAGTCGACA GTCGACTTAGTTGGGCTTCTTCTGCTTGGGGGGA
NLS 39/40 HP12_tag_M_FwD/REV	CCACCTAAGAGGGGCAGGAACGGCGGCGGCTAAAGATCTA AGATCTTTAGCCGCCCGTTCCTGCCCCCTTAGGTGGA
NLS 41/42 HP12_tag_C_FwD/REV	CCCCCAAGAGCAAGAAGGCTAAGATGTAAAAGCTTA AAGCTTTTACATCTTAGCCTTCTTCTGCTTGGGGGGA

nls 43/44 US34A_HCMVA_Fwd/rev	AAGTTCCGGAAGCGGAGAAGGCCTGTGGTGGTGTAAAGTCGACA GTCGACTTACACCACCACAGGCCTTCTCCGCTTCCGGAAGCTTA
NLS 45/46 IE1_HHV6U NLS Fwd/rev	AAGAGAGTGGCCAAGCGGAAGCACGTGTCCAGCAAGAGCCCCAAGAACAAGAAGATCAAGACCGACTGAA TCAGTCGGTCTTGATCTTCTTGTCTTGGGGCTCTTGCTGGACACGTGCTTCCGCTTGGCCACTCTCTTA
nls 47/48 ORF2_TTVZ1_Fwd/rev	CCTCCTCCACCTAAGAAAAGACGGCCCTGGTGTGAAAGCTTA AAGCTTTCAGCACCCAGGGCCGCTTTTCTTAGGTGGAGGAGGA
nls 49/50 MWP_tag_MC-Fwd/rev	CCCAAGAAGAGGCCTAGAGAGAGCAGCTCCAACAGCACATGCACCCCTCAAAGCGGCCAGAACTTCTGAA TCAGAAGTTTCTGGGCCGCTTTGGAGGGGTGCATGTGCTGTTGGAGCTGCTCTCTAGGCCTTCTTGGGA
nls 51/nls 52 KIP_tag_Cb-FWD/rev	AAGAGGAGCGCCCCGAGGAGGAGCCAGCTGCAGCCAGGCCACCCCCCAAGAAGAAGCACGCCTAAA TTAGGCGTGCTTCTTCTTGGGGGGGGTGGCCTGGCTGCAGCTGGGCTCCTCCTCGGGGGCGCTCCTCTTA
nls 53/54 MCP_tag_M-FWD/rev	CCCTTCAGCCGGAAGAGAAAGTTTGGCGGCAGCTGAA TCAGCTGCCGCCAAACTTTCTTCTCCGGCTGAAGGGA
nls 55/56 MCP_tag_C-Fwd/rev	CCTCCGAAGCCGAAGAAAAACAGAGAGTGAA TCACTCTGTTTTTCTTCGGCTTCGGAGGA
nls 57/58 MCV1-NLS-Fwd/rev	AGACCCTCTGCCAAGCGGAGAAGATGCAGCAGATGAAAGCTTA AAGCTTTCATCTGCTGCATCTTCTCCGCTTGGCAGAGGGTCTA
nls 59/60 HHV7J-NLS-fwd/rev	AAGAGAAGCTTCCCCGAGATCTGCCCCGAGCACTTCAAGAAGCGGCGTTTCTCTGAA TCAGATGAACCGCCGCTTCTTGAAGTGCTCGGGGCAGATCTCGGGGAAGCTTCTCTTA
nls 61/62 YLDV-NLS-Fwd/rev	CTGAAGTGGCTGCGGAAGAAGAGAAAGATCGCCCTGCAGACCTACTGAAAGCTTA AAGCTTTCAGTAGGTCTGCAGGGCGATCTTCTTCTTCTTCCGCAGCCACTTCAGA
nls 63/64 MCV1-NLS-Fwd/rev	CGGACCCACAAGAGAAAGGGAACACCTCTGCCTCTGCGGCCAGAAAGCAAGCGCGTTAGAGCTAGATGAA TCATCTAGCTCTAACGCGCTTGCTTCTGGGCCGAGAGGCAGAGGTGTTCCCTTCTTCTTGTGGGTCCGA
nls 65/66 BLLF2_EBV9-NLS-Fwd/rev	AGACCTCCTGTGGCCAAGCGGAGAAGATTCCCCAGATGAGTCGACA GTCGACTCATCTGGGGAATCTTCTCCGCTTGGCCACAGGAGGTCTA
nls 67/68	CGGTTCCGGAAGCGGAGAAGAGCAAGCCTTGAAGCTTA

DUVV-NLS-FWd/rev	AAGCTTTCAAGGCTTGCTCTTTCTCCGCTTCCGGAACCGA
nls 69/70 ICP0_HHV2H-NLS-FWd/rev	AGACCCAGAAAGAGAAGAGGCAGCGACAGCTGAGTCGACA GTCGACTCAGCTGTCGCTGCCTCTTCTTTCTGGGTCTA
nls 71/72 GB_HHV7J-NLS-FWd/rev	GCCTCCAGAAAGAGAAGAAAGCGCGAGCTGTGAGTCGACA GTCGACTCACAGCTCGCGCTTTCTTCTTTCTGGAGGCA
nls 73/74 ORFSA-NLS-FWd/rev	CGGCGGAAGCGGAAGAGAAAGACCCCTAACTGCTGAGTCGACA GTCGACTCAGCAGTTAGGGTCTTTCTTCCGCTTCCGCCGA
nls 75/76 YMTV5-NLS-FWd/rev	GGCGGAGCCAAGCGGAAGAAAAGAAAGCCCAAGTGAAGTCGACA GTCGACTCACTGGGCTTTCTTTCTTCCGCTTGGCTCCGCCAA
nls 77/78 COWPX-NLS-FWd/rev	ATCAGAAAGCGGCCCAACCAGCACACCATCGACCTGTTCAAGCGGATGAAGCGGTGAA TCACCGCTTCATCCGCTTGAACAGGTGATGGTGTGGTGTGGTGGGCCGCTTTCTGATA
nls 79/80 9ORTO-NLS-FWd/rev	CGGAGAGCCAACAAGCGGAGACTGGAAGAACTGTGAAAGCTTA AAGCTTTCACAGTTCTTCCAGTCTCCGCTTGTGGCTCTCCGA
nls 83/84 GCVK_HHV7J-NLS-	AGACCCTGCAAAGTGAAGCGGAAGCTGTTCCGCGAGCGAGAACATCAGACCCAACAAGAAAATCCCTCTGTGAA TCACAGAGGGATTTTCTTGTGGGTCTGATGTTCTCGCTGCCGAACAGCTTCCGCTTCACTTTCAGGGTCTA
nls 85/86 Q98291_MCV1-NL	CGGCGGAGAAAGAGAAAGCCAGAACACCTGAAAGCTTA AAGCTTTCAGGTGGTTCTGGGCTTTCTTTCTCCGCCGA
nls 87/88 HSPV-NLS-FWd/rev	CTGTTCAAGAAGATCAAGCGGACCCGCTACGACACCTTCTGAAAGCTTA AAGCTTTCAGAAGGTGTCGTAGCGGGTCCGCTTGATCTTCTTGAACAGA
nls 89/90 HHV7NLS FWD/rev	AAGAAGCAGCTGAAGCGGAAGTCCGAGAGCAAGCTGAAAACCAGCAAGGCCAAGAAGAAGCTGATCTGAA TCAGATCAGCTTCTTCTTCTTGGCCTTGCTGGTTTTAGCTTCTCGGACTTCCGCTTCCGCTTCTTCTTA
nls91/92 WUP_tag_MC-FWd/rev	AAGAGAACCAGAGAGGACGACGAGGAACCCAGTGCTCTCAGGCCACACCTCCTAAGAAGAAGAAGGACTGAA TCAGTCTTCTTCTTCTTAGGAGGTGTGGCCTGAGAGCACTGGGGTTCCTCGTCTCCTCTCTGTTCTCTTA
nls 93/94 A0A2N9DYY9_HHV6 FWd/rev	GAGTACACCAAGAAGCGGCGGAGACACAGAGTGTGAA TCACACTCTGTGTCTCCGCCGCTTCTTGGTGTACTCA

nls 95/96 Q8V556_MONPZ-NLS- fwd/rev	ATCAGAAAGCGGCCCAACCAGCACCACACCATCGACCTGTTCAAGCGGATCAAGAGAACCCGCTACGACACCTTCTGAA TCAGAAGGTGTCGTAGCGGGTTCTTTGATCCGCTTGAACAGGTCGATGGTGTGGTGTGTTGGGCCGCTTTCTGATA
nls 97/98 Q6TVM4_ORFSA-NLS- fwd/rev	GGCAACAAGAAGCGGAGAAGGCGGAGAGTGTGAA TCACACTCTCCGCTTCTCCGCTTCTTGTGCCA
nls 99/100 NSS_UUKS-NLS fwd/rev	CGGCTGAGAAGAAAGAAGCGGTCCAGAGTCAGCTGAA TCAGCTGACTCTGGACCGCTTCTTTCTTCTCAGCCGA
nls 101/102 Q8V556_MONPZ-NLS sh. FWd/rev	ATCAGAAAGCGGCCCAACCAGCACCACACCATCGACCTGTTCAAGCGGATCAAGTGAA CACTTGATCCGCTTGAACAGGTCGATGGTGTGGTGTGTTGGGCCGCTTTCTGATA
nls 103/104 Q0GP58_HSPV-NLS bip FWD/rev	ATCAGAAAGCGGCCCAACCAGCACCACACCATCGACCTGTTCAAGAAGATCAAGCGGACCTGAA TCAGGTCCGCTTGATCTTCTTGAACAGGTCGATGGTGTGGTGTGTTGGGCCGCTTTCTGATA
nls 105/106 Q8V556_MONPZ- NLSshNm.FWd/rev	ATTGCCGCCGCTCCTAATCAGCACCACACCATCGACCTGTTCAAGCGGATCAAGTGAA TCACTTGATCCGCTTGAACAGGTCGATGGTGTGGTGTGATTAGGAGCGGCGGCAATA
nls 107/108 Q8V556_MONPZ- NLSshCm.FWd/rev	ATCAGAAAGCGGCCCAACCAGCACCACACCATCGATCTGTTTGCCGCCATTGCCTGAA TCAGGCAATGGCGGCAACAGATCGATGGTGTGGTGTGTTGGGCCGCTTTCTGATA
nls 109/110 Q8V556_MONPZ- NLSshNCm.FWd/rev	ATTGCCGCCGCTCCTAATCAGCACCACACCATCGATCTGTTGCGGCCATTGCCTGAA TCAGGCAATGGCGGCAACAGATCGATGGTGTGGTGTGATTAGGAGCGGCGGCAATA
nls 111/112 U5TCT3_COWPX-NLS_Nm FWD/rev	ATTGCCGCCGCTCCTAATCAGCACCACACCATCGACCTGTTCAAGCGGATGAAGCGGTGAA TCACCGTTCATCCGCTTGAACAGGTCGATGGTGTGGTGTGATTAGGAGCGGCGGCAATA
nls 113/114 U5TCT3_COWPX-NLS_Cm FWD/rev	ATCAGAAAGCGGCCCAACCAGCACCACACCATCGATCTGTTTGCCGCCATGGCTGCTTGAA TCAAGCAGCCATGGCGGCAACAGATCGATGGTGTGGTGTGTTGGGCCGCTTTCTGATA
nls 115/116 U5TCT3_COWPX- NLS_NCm FWD/REV	ATTGCCGCCGCTCCTAATCAGCACCACACCATCGATCTGTTGCGGCCATGGCTGCTTGAA TCAAGCAGCCATGGCGGCAACAGATCGATGGTGTGGTGTGATTAGGAGCGGCGGCAATA

NLS 117/118 KIP_tag_Cb_mut FWD/REV	GCTGCTTCTGCCCCTGAGGAAGAACCTAGCTGCTCTCAGGCCACACCTCCTAAGAAGAAGCACGCCTTCGACGCCTGAA TCAGGCGTGAAGGCGTGCTTCTTCTTAGGAGGTGTGGCCTGAGAGCAGCTAGTTCTTCTCAGGGGCAGAAGCAGCA
NLS 119/120 STLP_tag_Cb_mut FWD/REV	GCCGCTGCCTTTCCTGATAGCAGCACCCAGAATAGCACCCCTCCAAAGAAGAACAAGCCCGCCTGAA TCAGGCGGGCTTGTCTTCTTTGGAGGGGTGCTATTCTGGGTGCTGCTATCAGGAAAGGCAGCGGCA
NLS 121/122 WUP_tag_MC_mut FWD/REV	CTGCCACAGCCGAGGATGATGAGGAACCTCAGTGCAGCCAGGCCACACCTCCTAAGAAGAAGAAGGACTGAA TCAGTCCTTCTTCTTCTTAGGAGGTGTGGCCTGGCTGCACTGAGGTTCTCATCATCCTCGGCTGTGGCAGCA
NLS 132/133 Q9DUB7ctNLS- FWD/REV	GACGAGGCCAGACGGAAGAGACTGAAGCGCGTGTGCCTGATGTGAA TCACATCAGGCACACGCGCTTCAGTCTCTCCGTCTGGCCTCGTCA
NLS 134/135 MCPyV-NLS-m-278- FWD/REV	CCCTTCAGCCGGACCAGAAAGTTTGGCGGCTCTTGAA TCAAGAGCCGCCAAACTTTCTGGTCCGGCTGAAGGGA
NLS 136/137 MCPyV-NLS-m-280- FWD/REV	CCCTTCAGCCGGAAGAGAACATTTGGCGGCAGCTGAA TCAGCTGCCGCCAAATGTTCTTCTCCGGCTGAAGGGA

**Table 6.4 Antibiotics used in this study**

Antibiotic	Stock solution	Working solution	Dilution
Ampicillin (SIGMA ALDRICH, #A9518-25G)	50 mg/ml	50 g/ml	1:1000
kanamycin (Applichem, #A1493,0005)	10 mg/ml	10 g/ml	1:500

Gentamycin  (SIGMA ALDRICH, #G1272-10mL)	10 mg/ml	10 g/ml	1:500
------------------------------------------------	----------	---------	-------

**Table 6.5 List of Cell lines utilized for Bacteria transformation in this study**

Plasmid DNAs	<i>E. COLI</i> STRAINS	Amount of <i>E. coli</i> strains in reaction L	Amount of Plasmid DNAs in reaction 1
BP products	DH5 $\alpha$ <i>E. coli</i> (Life Technologies, #C404010)	45	2.5
LR products	DH5 $\alpha$ <i>E. coli</i> (Life Technologies, #C404010)	45	2.5
site-specific mutagenesis reactions	<i>XL10-Gold</i> (Agilent, #C200315)	90	6
TOPO reaction products	TOP10 Chemically Competent <i>E. coli</i> (Thermofisher Scientific #C404003)	16	1.6

**Table 6.6 List of primers and templates used for PCR Gateway® Technology**

PCR product		Template		primers
code	size (bp)	code	size (bp)	
[(139+140) *804]	183	804	7143	NLS 139/140
[(139+129) *804]	183	804	7143	NLS 139/129
[(138+129) *804]	183	804	7143	NLS 138/129
[(128+129) *722]	183	722	7143	NLS 128/129
[(123+124) *724]	172	724	5748	NLS 123/124
[(125+124) *779]	172	779	5748	NLS 124/125

[(138+140) *822]	183	822	7143	NLS 138/140
[(139+140) *823]	183	823	7143	NLS 139/140

**Table 6.7 List of pDN207 clones which were generated in this study**

<i>code</i>	<i>Insert name</i>	<i>sequence</i>	<i>reaction</i>
637	P68714	Full length protein 1-77, containing the cNLS 10-MKSRKKKPKTT-20	BP 561xGW5
638	P14357	Full length protein 1-175, containing NLSbip 101-IRKRPNQHHTIDLFKRIKR-119	BP 562xGW5
751	P14357 (R102A/K103A/R104A K115A/R116A/K118A/R119A)	Full length protein 1-175, containing NLSbip 101- <b>Iaaa</b> PNQHHTIDL <b>Faa</b> 119	BP 749xGW5
752	P68714 (K15A/K16A/K17A/K19A)	Full length protein 1-77, containing the cNLS 10-MKSR <b>aaa</b> PaTT-20	BP 750xGW5
763	HPyV7-Tag-CcN	116-SSSDEEEPASSASVNPEEGCSQDSKYSATPPK <b>QKK</b> PNP-153	BP GW5xPCR [(123+124) *724]
764	HPyV7-Tag-CcN (S116A/S117A/S118A)	116- <b>aaa</b> DEEEPASSASVNPEEGCSQDSKYSATPPK <b>QKK</b> PNP-153	BP GW5xPCR [(125+124) *779]
797	MCPyV-Tag-NLSbip	274-PFS <b>RKRK</b> FGGSRSSASSASSASFTSTPPK <b>PKKNRE</b> - 307	BP GW5xPCR [(128+129) *772]
820	MCPyV-Tag-NLSbip (K278T)	274-PFS <b>RtRK</b> FGGSRSSASSASSASFTSTPPK <b>PKKNRE</b> - 307	BP GW5xPCR [(129+138) *804]
821	MCPyV-Tag-NLSbip (K280T)	274-PFS <b>RKRt</b> FGGSRSSASSASSASFTSTPPK <b>PKKNRE</b> - 307	BP GW5xPCR [(129+139) *804]
825	MCPyV-NLSbip (K278T/K303A/K304A)	274-PFS <b>RtRK</b> FGGSRSSASSASSASFTSTPPK <b>Paa</b> NRE- 307	BP GW5xPCR [(138+140) *822]

826	MCPyV-NLSbip (K280T/K303A/K304A)	274-PFS <b>RRK</b> RtFGGSRSSASSASSASFTSTPP <b>KPaa</b> NRE- 307	BP GW5xPCR [(139+140) *823]
-----	-------------------------------------	--------------------------------------------------------------------	--------------------------------------

**Table 6.8 List of pDESTntYFP clones which were generated in this study**

<i>code</i>	<i>Insert name</i>	<i>sequence</i>	<i>reaction</i>
697	P68714	Full length protein 1-77, containing the cNLS 10-MKSRKKKPKTT-20	LR 637 x GW22
715	P14357	Full length protein 1-175, containing NLSbip 101-IRKRPNQHHTIDLFKRIKR-119	LR 638 x GW22
753	P14357 (R102A/K103A/R104A K115A/R116A/K118A/R119A)	Full length protein 1-175, containing NLSbip 101-IaaaPNQHHTIDLFa <del>al</del> aa-119	LR 751 x GW22
754	P68714 (K15A/K16A/K17A/K19A)	Full length protein 1-77, containing the cNLS 10-MKSRaaaPaTT-20	LR 752 x GW22
779	HPyV7-Tag-CcN	116-SSSDEEEPASSASVNPEEGCSQDSKYSATPP <b>KQKK</b> PNP-153	LR 763 x GW22
786	HPyV7-Tag-CcN (S116A/S117A/S118A)	116-aaaDEEEPASSASVNPEEGCSQDSKYSATPP <b>KQKK</b> PNP-153	LR 764 x GW22
804	MCPyV-Tag-NLSbip	274-PFS <b>RRK</b> RKFGGSRSSASSASSASFTSTPP <b>KPKKN</b> NRE- 307	LR 797 x GW22
822	MCPyV-Tag-NLSbip (K278T)	274-PFS <b>R</b> tRKFGGSRSSASSASSASFTSTPP <b>KPKKN</b> NRE- 307	LR 820 x GW22
823	MCPyV-Tag-NLSbip (K280T)	274-PFS <b>RRK</b> RtFGGSRSSASSASSASFTSTPP <b>KPKKN</b> NRE- 307	LR 821 x GW22
828	MCPyV-NLSbip (K278T/K303A/K304A)	274-PFS <b>R</b> tRKFGGSRSSASSASSASFTSTPP <b>KPaa</b> NRE- 307	LR 825 x GW22
844	MCPyV-NLSbip (K280T/K303A/K304A)	274-PFS <b>RRK</b> RtFGGSRSSASSASSASFTSTPP <b>KPaa</b> NRE- 307	LR 826 x

**Table 6.9 List of pGFP-NT TOPO clones which were generated in this study****a) Polyomaviridae Large Tumor Antigens NLSs fusions**

<i>Code</i>	<i>Insert name</i>	<i>Sequence</i>	<i>Reaction</i>
606	KIPyV Tag-NLSct	134-PPKKKHA-140	TOPO (29+30)
645	KIPyV Tag-NLSbip	119-KRSAPEEEPPSCSQATPPKKKHA-140	TOPO (51+52)
608	WUPyV Tag-NLSm	122-PTKRTRE-128	TOPO (33+34)
609	WUPyV Tag-NLSct	140-PPKKKKDNA-148	TOPO (35+36)
695	WUPyV Tag-NLSbip	124-KRTREDDEEPQCSQATPPKKKD-146	TOPO (91+92)
576	STLPyV Tag-NLSct	137-PPKKNKPA-144	TOPO (3+4)
593	STLPyV Tag-NLSbip	124-KRKFPDSSTQNSTPPKKNKPA-144	TOPO (17+18)
610	HPyV7 Tag-NLSct	145-PPKQKKPN-152	TOPO (37+38)
592	LIPyV Tag-NLSm	211-PKRNRKNQ-218	TOPO (15+16)
600	LIPyV Tag-NLSnt	134-PRPKKRRSNL-143	TOPO (23+24)
601	LIPyV Tag-NLSct	236-PPKQKRYKE-244	TOPO (25+26)
647	MCPyV Tag-NLSm	274-PFSRKRKFGGS-284	TOPO (53+54)
648	MCPyV Tag-NLSct	299-PPKPKKNRE-307	TOPO (55+56)
579	NJPyV Tag-NLSct	206-PPKQKRKSP-214	TOPO (9+10)
595	NJPyV Tag-NLSm	181-PPKRRRGT-188	TOPO (21+22)
607	TSPyV Tag-NLSct	175-PPKPKKSKY-183	TOPO (31+32)
611	HPyV12 Tag-NLSm	169-PKRGRNGGG-177	

			TOPO (39+40)
612	HPyV12 Tag-NLSct	204-PPKSKKAKM-212	TOPO (41+42)
578	HPyV6 Tag-NLSct	140-PPKKRKPN-147	TOPO (7+8)
590	HPyV9 Tag-NLSct	171-PPKRKKPE-178	TOPO (11+12)
591	MPyV Tag-NLSnt	188-PVSRKRPRPA-197	TOPO (13+14)
602	MPyV Tag-NLSct	279-PPKKARED-286	TOPO (27+28)
577	MWPyV Tag-NLSct	150-PPKRPRNF-157	TOPO (5+6)
594	MWPyV Tag-NLSm	135-PKKRPRES-142	TOPO (19+20)
646	MWPyV Tag-NLSbip	135-PKKRPRESSNSTCTPPKRPRNF-157	TOPO (49+50)
755	KIPyV Tag-NLSbip (K119A/R120A)	119-aaSAPEEEEPSCSQATPPKKKHA-140	TOPO (117+118)
757	WUPyV Tag-NLSbip (K124A/R125A/K127A)	124-aaTaEDDEEPQCSQATPPKKKKD-146	TOPO (121+122)
756	STLyVP Tag-NLSbip (K124A/R125A/K126A)	124-aaaFPDSSTQNSTPPKKNKPA-144	TOPO (119+120)
807	MCPyV Tag-NLSm (K278T)	274-PFSRtRKFGGS-284	TOPO (134+135)
808	MCPyV Tag-NLSm (K280T)	274-PFSRKRtFGGS-284	TOPO (136+137)

## b) Top hits NLS fusions

code	Insert name	Sequence	Reaction
613	Q7M6G6-NLS	106-KFRKRRRPVVV-116	TOPO (43+44)
614	Q69567-NLS	609-KRVAKRKHVSSKSPKNKKIK TD-630	TOPO (45+46)
657	Q98263-NLS	137-RPSAKRRRCR-147	TOPO (57+58)
615	Q9DUB8-NLS	68-PPPPKRRPWC-79	TOPO (47+48)
659	Q9DHK5-NLS	187-LKWLRRKRKIALQTY-201	

			TOPO (61+62)
660	Q98187-NLS	262- <b>RTHKRK</b> GTPLPLRPR <b>SKRVRA</b> <b>R-283</b>	TOPO (63+64)
676	D2CRM8-NLS	482- <b>RFRKRRK</b> SKP-491	TOPO (67+68)
677	P28284-NLS	511- <b>RPRKRRG</b> SDS-520	TOPO (69+70)
661	P03199-NLS	47- <b>RPPVAKRRRF</b> PR-58	TOPO (65+66)
678	P52352-NLS	391- <b>ASRKRRK</b> REL-400	TOPO (71+72)
679	Q6TVJ0-NLS	28- <b>RRKRKRK</b> TPN-38	TOPO (73+74)
681	U5TCT3-NLS	101- <b>IRKRPNQHHTIDLFKRMKR-</b> 119	TOPO (77+78)
685	Q98291-NLS	9- <b>RRRKRPRTT</b> -18	TOPO (85+86)
682	D6PT84-NLS	581- <b>RRANKRR</b> LEEL-591	TOPO (79+80)
683	P16738-NLS	8- <b>RRGKRRK</b> L-16	TOPO (81+82)
684	P52344-NLS	29- <b>RPCKVKRK</b> LFGSENIRPN <b>KK</b> IPL-51	TOPO (83+84)
658	Q69518-NLS	483- <b>KRSFPEICPEHF</b> <b>KKRRFI</b> -500	TOPO (59+60)
687	Q69514-NLS	820- <b>KKQLKRK</b> SES <b>KLKTSKA</b> <b>KKKKLI</b> -842	TOPO (89+90)
686	Q0GP58/P0DSQ1-NLS	113- <b>LFKKIKR</b> TRYDTF-125	TOPO (87+88)
772	Q8V556-NLS	113- <b>LFKRIKR</b> TRYDTF-125	TOPO (126+127)
719	Q6TVM4-NLS	24- <b>GNKKRRRRR</b> V-33	TOPO (97+98)
717	A0A2N9DYY9-NLS	129- <b>EYTKRRR</b> HRV-139	TOPO (93+94)
720	P22026-NLS	133- <b>RLRRKRS</b> RVS-143	TOPO (99+100)
680	Q6TUQ8-NLS	5- <b>GGAKRKR</b> KPK-15	TOPO (99+100)

718	Q8V556-NLSbip_long	101- <b>IRKRP</b> NQHHTIDLF <b>KRIK</b> RTR YDTF-125	TOPO (75+76)
731	Q8V556-NLSbip_short	101- <b>IRKRP</b> NQHHTIDLF <b>KRIK</b> -118	TOPO (101+102)
733	Q8V556-NLSbip_short (R102A/K103A/R104A)	101- <b>laaa</b> PNQHHTIDLF <b>KRIK</b> -118	TOPO (105+106)
734	Q8V556-NLSbip_short (K115A/R116A/K118A)	101- <b>IRKRP</b> NQHHTIDLF <b>Faala</b> -118	TOPO (107+108)
735	Q8V556-NLSbip_short (R102A/K103A/R104A/K115A/R116A/K118A)	101- <b>laaa</b> PNQHHTIDLF <b>Faala</b> -118	TOPO (109+110)
732	Q0GP58-NLSbip	101- <b>IRKRP</b> NQHHTIDLF <b>KKIK</b> R- 120	TOPO (103+104)
736	5UTCT3-NLSbip (R102A/K103A/R104A)	101- <b>laaa</b> PNQHHTIDLF <b>KRMKR</b> -119	TOPO (111+112)
737	5UTCT3-NLSbip (K115A/R116A/K118A/R119A)	101- <b>IRKRP</b> NQHHTIDLF <b>FaaMaa</b> -119	TOPO (113+114)
738	5UTCT3-NLSbip (R102A/K103A/R104A/K115A/R116A/K118A/R119A)	101- <b>laaa</b> PNQHHTIDLF <b>FaaMaa</b> -119	TOPO (115+116)
805	D6PT84-cNLSct	698- <b>DEARRKRLKR</b> VCLM-712	TOPO (130+131)
806	Q9DUB7-cNLSct	632- <b>LPPPEKRAR</b> WGFP-644	TOPO (132+133)

## Reference

- ADDIN Mendeley Bibliography CSL\_BIBLIOGRAPHY 1. Lopez-Denman AJ, Russo A, Wagstaff KM, White PA, Jans DA, Mackenzie JM. Nucleocytoplasmic shuttling of the West Nile virus RNA-dependent RNA polymerase NS5 is critical to infection. *Cell Microbiol.* 2018;20(8):1–13.
2. Lopez-Denman AJ, Mackenzie JM. The IMPORTance of the nucleus during flavivirus replication. *Viruses.* 2017;9(1).
  3. Wente SR, Rout MP. The nuclear pore complex and nuclear transport. *Cold Spring Harb Perspect Biol.* 2010;2(10):1–19.
  4. Chook YM, Süel KE. Nuclear import by karyopherin- $\beta$ s: Recognition and inhibition. *Biochim Biophys Acta - Mol Cell Res.* 2011;1813(9):1593–606.
  5. Kosugi S, Hasebe M, Tomita M, Yanagawa H. Systematic identification of cell cycle-dependent yeast nucleocytoplasmic shuttling proteins by prediction of composite motifs. *Proc Natl Acad Sci U S A.* 2009;106(25):10171–6.
  6. Daniel E Shumer NJNNPS. Classical nuclear localization signals: definition, function, and interaction with importin alpha. *Physiol Behav.* 2017;176(12):139–48.
  7. Shunichi Kosugi†§1 MH, Matsumura§ N, Takashima§ H, Miyamoto-Sato§ E, Tomita† M, and Hiroshi Yanagawa†§. Six Classes of Nuclear Localization Signals Specific to Different Binding Grooves of Importin a. 148:148–62.
  8. Kalderon D, Roberts BL, Richardson WD, Smith AE. A short amino acid sequence able to specify nuclear location. *Cell.* 1984;39(3 PART 2):499–509.
  9. Pemberton LF, Paschal BM. Mechanisms of receptor-mediated nuclear import and nuclear export. *Traffic.* 2005;6(3):187–98.
  10. Azuma Y, Tabb MM, Vu L, Nomura M. Isolation of a yeast protein kinase that is activated by the protein encoded by SRP1 (Srp1p) and phosphorylates Srp1p complexed with nuclear localization signal peptides. *Proc Natl Acad Sci U S A.* 1995;92(11):5159–63.
  11. Roman N, Christie M, Swarbrick CMD, Kobe B, Forwood JK. Structural characterisation of the nuclear import receptor importin alpha in complex with the bipartite NLS of Prp20. *PLoS One.* 2013;8(12).

12. News CC. Regulating post-mitotic nuclear access : Cdk1-phosphorylation of NLSs. 2015;695–6.
13. Fontoura BMA, Faria PA, Nussenzveig DR. Viral interactions with the nuclear transport machinery: Discovering and disrupting pathways. *IUBMB Life*. 2005;57(2):65–72.
14. Gustin KE, Sarnow P. Effects of poliovirus infection on nucleo-cytoplasmic trafficking and nuclear pore complex composition. *EMBO J*. 2001;20(1–2):240–9.
15. Shabman RS, Gulcicek EE, Stone KL, Basler CF. The Ebola virus VP24 protein prevents hnRNP C1/C2 binding to karyopherin  $\alpha$ 1 and partially alters its nuclear import. *J Infect Dis*. 2011;204(SUPPL. 3).
16. Petty R V., Basta HA, Bacot-Davis VR, Brown BA, Palmenberg AC. Binding Interactions between the Encephalomyocarditis Virus Leader and Protein 2A. *J Virol*. 2014;88(22):13503–9.
17. Malik P, Tabarraei A, Kehlenbach RH, Korfali N, Iwasawa R, Graham S V., et al. Herpes simplex virus ICP27 protein directly interacts with the nuclear pore complex through Nup62, inhibiting host nucleocytoplasmic transport pathways. *J Biol Chem*. 2012;287(15):12277–92.
18. Jans DA, Martin AJ, Wagstaff KM. Inhibitors of nuclear transport. *Curr Opin Cell Biol* [Internet]. 2019;58:50–60. Available from: <https://doi.org/10.1016/j.ceb.2019.01.001>
19. Canga AG, Prieto AMS, Diez Liébana MJ, Martínez NF, Sierra Vega M, García Vieitez JJ. The pharmacokinetics and interactions of ivermectin in humans - A mini-review. *AAPS J*. 2008;10(1):42–6.
20. Van Der Watt PJ, Chi A, Stelma T, Stowell C, Strydom E, Carden S, et al. Targeting the nuclear import receptor Kpn $\beta$ 1 as an anticancer therapeutic. *Mol Cancer Ther*. 2016;15(4):560–73.
21. cargo inhibitors.pdf.
22. Veach RA, Liu Y, Zienkiewicz J, Wylezinski LS, Boyd KL, Wynn JL, et al. Survival, bacterial clearance and thrombocytopenia are improved in polymicrobial sepsis by targeting nuclear transport shuttles. *PLoS One*. 2017;12(6):1–17.
23. Ali SH, DeCaprio JA. Cellular transformation by SV40 large T antigen: Interaction with host proteins. *Semin Cancer Biol*. 2001;11(1):15–22.
24. Houben R, Angermeyer S, Haferkamp S, Aue A, Goebeler M, Schrama D, et al. Characterization of functional domains in the Merkel cell polyoma virus Large T antigen. *Int J Cancer*. 2015;136(5):E290–300.
25. An P, Robles MTS, Pipas JM. Large T antigens of polyomaviruses: Amazing molecular machines. *Annu Rev Microbiol*. 2012;66:213–36.

26. Ahuja D, Rathi A V., Greer AE, Chen XS, Pipas JM. A Structure-Guided Mutational Analysis of Simian Virus 40 Large T Antigen: Identification of Surface Residues Required for Viral Replication and Transformation. *J Virol.* 2009;83(17):8781–8.
27. Cyrklaff M, Risco C, Fernández JJ, Jiménez MV, Estéban M, Baumeister W, et al. Cryo-electron tomography of vaccinia virus. *Proc Natl Acad Sci U S A.* 2005;102(8):2772–7.
28. Slifka MK, Hanifin JM. Smallpox: The basics. *Dermatol Clin.* 2004;22(3):263–74.
29. Simmons BJ, Falto-Aizpurua LA, Griffith RD, Nouri K. Smallpox: 12000 years from plagues to eradication: A dermatologic ailment shaping the face of society. *JAMA Dermatology.* 2015;151(5):482.
30. Seitz R. Orthopox viruses: Infections in humans. *Transfus Med Hemotherapy.* 2010;37(6):351–64.
31. Möstl K, Addie D, Belák S, Boucraut-Baralon C, Egberink H, Frymus T, et al. Cowpox Virus Infection in Cats: ABCD guidelines on prevention and management. *J Feline Med Surg.* 2013;15(7):557–9.
32. Lu Y, Zhang L. DNA-Sensing Antiviral Innate Immunity in Poxvirus Infection. *Front Immunol.* 2020;11(August):1–7.
33. Manuscript A. STING SPECIFIES IRF3 phosphorylation by TKB1. 2013;5(214).
34. Unterholzner L, Keating SE, Baran M, Horan KA, Jensen SB, Sharma S, et al. IFI16 is an innate immune sensor for intracellular DNA. *Nat Immunol.* 11, 997–1004 (2010).e immune sensor for intracellular DNA. *Nat Immunol* [Internet]. 2010;11(11):997–1004. Available from: <http://dx.doi.org/10.1038/ni.1932>
35. Nagaraj Kerur, Mohanan Valiya Veetil, Neelam Sharma-Walia, Virginie Bottero S, Sadagopan, Pushpalatha Otageri and BC. IFI16 acts as a nuclear pathogen sensor to induce the inflammasome in response to Kaposi sarcoma associated herpesvirus infection. *Bone.* 2013;23(1):1–7.
36. Smith GL, Benfield CTO, Maluquer de Motes C, Mazzon M, Ember SWJ, Ferguson BJ, et al. Vaccinia virus immune evasion: Mechanisms, virulence and immunogenicity. *J Gen Virol.* 2013;94(PART 11):2367–92.
37. Jones JO, Arvin AM. Inhibition of the NF-κB Pathway by Varicella-Zoster Virus In Vitro and in Human Epidermal Cells In Vivo. *J Virol.* 2006;80(11):5113–24.
38. Stack J, Haga IR, Schröder M, Bartlett NW, Maloney G, Reading PC, et al. Vaccinia virus protein A46R targets multiple Toll-like-interleukin-1 receptor adaptors and contributes to virulence. *J Exp Med.* 2005;201(6):1007–18.

39. de Motes CM, Cooray S, Ren H, Almeida GMF, McGourty K, Bahar MW, et al. Inhibition of apoptosis and NF- $\kappa$ B activation by vaccinia protein N1 occur via distinct binding surfaces and make different contributions to virulence. *PLoS Pathog.* 2011;7(12).
40. Chen RAJ, Ryzhakov G, Cooray S, Randow F, Smith GL. Inhibition of I $\kappa$ B kinase by vaccinia virus virulence factor B14. *PLoS Pathog.* 2008;4(2).
41. Mansur DS, Maluquer de Motes C, Unterholzner L, Sumner RP, Ferguson BJ, Ren H, et al. Poxvirus Targeting of E3 Ligase  $\beta$ -TrCP by Molecular Mimicry: A Mechanism to Inhibit NF- $\kappa$ B Activation and Promote Immune Evasion and Virulence. *PLoS Pathog.* 2013;9(2).
42. Ferguson BJ, Benfield CTO, Ren H, Lee VH, Frazer GL, Strnadova P, et al. Vaccinia virus protein N2 is a nuclear IRF3 inhibitor that promotes virulence. *J Gen Virol.* 2013;94(PART9):2070–81.
43. Graham SC, Bahar MW, Cooray S, Chen RAJ, Whalen DM, Abrescia NGA, et al. Vaccinia virus proteins A52 and B14 share a Bcl-2-like fold but have evolved to inhibit Nf- $\kappa$ B rather than apoptosis. *PLoS Pathog.* 2008;4(8).
44. Smith GL. Vaccinia Virus. *Encycl Virol.* 2008;378:243–50.
45. Kvensakul M, Yang H, Fairlie WD, Czabotar PE, Fischer SF, Perugini MA, et al. Vaccinia virus anti-apoptotic F1L is a novel Bcl-2-like domain-swapped dimer that binds a highly selective subset of BH3-containing death ligands. *Cell Death Differ.* 2008;15(10):1564–71.
46. García-Arriaza J, Gómez CE, Sorzano CÓS, Esteban M. Deletion of the Vaccinia Virus N2L Gene Encoding an Inhibitor of IRF3 Improves the Immunogenicity of Modified Vaccinia Virus Ankara Expressing HIV-1 Antigens. *J Virol.* 2014;88(6):3392–410.
47. Bartlett N, Symons JA, Tschärke DC, Smith GL. The vaccinia virus N1L protein is an intracellular homodimer that promotes virulence. *J Gen Virol.* 2002;83(8):1965–76.
48. Satheshkumar PS, Weisberg AS, Moss B. Vaccinia Virus A19 Protein Participates in the Transformation of Spherical Immature Particles to Barrel-Shaped Infectious Virions. *J Virol.* 2013;87(19):10700–9.
49. Satheshkumar PS, Olano LR, Hammer CH, Zhao M, Moss B. Interactions of the Vaccinia Virus A19 Protein. *J Virol.* 2013;87(19):10710–20.
50. Szajner P, Weisberg AS, Lebowitz J, Heuser J, Moss B. External scaffold of spherical immature poxvirus particles is made of protein trimers, forming a honeycomb lattice. *J Cell Biol.* 2005;170(6):971–81.
51. Lu J, Wu T, Zhang B, Liu S, Song W, Qiao J, et al. Types of nuclear localization signals and mechanisms of protein import into the nucleus. *Cell Commun Signal [Internet].* 2021;19(1):1–10. Available from: <https://doi.org/10.1186/s12964-021-00741-y>

52. Allison Lange\*, 1, Laura M. McLane\*, 2, Ryan E. Mills<sup>3</sup>, Scott E. Devine<sup>4</sup>, and Anita H. Corbett<sup>1</sup> <sup>1</sup>Department of Biochemistry, Emory University School of Medicine, 1510 Clifton Road NE, Atlanta G 30322, Abstract. NIH Public Access. NIH Public Access Author [Internet]. 2010;23(1):1–7. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3624763/pdf/nihms412728.pdf>
53. Fulcher AJ, Dias MM, Jans DA. Binding of p110 retinoblastoma protein inhibits nuclear import of simian virus SV40 large tumor antigen. *J Biol Chem.* 2010;285(23):17744–53.
54. Jans DA, Ackermann MJ, Bischoff JR, Beach DH, Peters R. p34cdc2-mediated phosphorylation at T124 inhibits nuclear import of SV-40 T antigen proteins. *J Cell Biol.* 1991;115(5):1203–12.
55. Róna G, Borsos M, Kobe B, Vértessy BG. Factors influencing nucleo-cytoplasmic trafficking: Which matter Response to Alvisi & Jans' comment on Phosphorylation adjacent to the nuclear localization signal of human dUTPase abolishes nuclear import: Structural and mechanistic insights. *Acta Crystallogr Sect D Biol Crystallogr.* 2014;70(10):2777–8.
56. Kosugi S, Hasebe M, Entani T, Takayama S, Tomita M, Yanagawa H. Design of Peptide Inhibitors for the Importin  $\alpha/\beta$  Nuclear Import Pathway by Activity-Based Profiling. *Chem Biol.* 2008;15(9):940–9.

