The human polyomavirus KI:

A study on cell permissivity, sub-cellular localization of VP1 and presence in cerebrospinal fluid and urine

Master Thesis in Medical Biology

MB-3910

By

Maria A. Ludvigsen

Host-Microbe Interactions Research group

Department of Medical Biology

University of Tromsø

November 2011

Acknowledgements

The work presented in this thesis was carried out during the period August 2010 to November 2011 at the Host-Microbe Interactions Research group, Department of Medical Biology, University of Tromsø (UIT) under the guidance of supervisor Professor Ugo Moens and co-supervisor senior scientist Marijeke van Ghelue.

I would like to extend my sincere thanks to my supervisor Professor Ugo Moens for welcoming me into the group. Thank you for always keeping the door open and for taking time to answer my questions. You have provided me with excellent support, encouragement and guidance both in the lab and while writing my thesis. Thank you to my co-supervisor Marijeke van Ghelue for your warm welcome, guidance and help in finding my way around the lab on the 6.th floor.

Thank you Gianina and Sergiy for taking the time to help me in the lab and for sharing your knowledge. I would also like to thank everyone else who's made the research group such a pleasure to be a part of!

Thank you to my fellow students Hanne, Maria, Ann Iren and Kristin for your friendship and for making the office such a lovely place to be in, mornings, evenings and week-ends too. Your knowledge, pep-talks and gossip-breaks have been most welcome ☺

Finally I would like to thank my friends and family for your support and welcome distractions. Thank you Vegard for your patience, love and encouragement.

Tromsø, November 2011

Maria

Abstract

KI polyomavirus (KIPyV) is a relatively newly discovered human polyomavirus originally identified in respiratory tract samples. Little is known about the route of infection or transmission, and the pathogenic properties of this virus are virtually unknown. The life cycle of KIPyV has not been studied as a permissive cell system has not been identified so far. The bona fide sites of viral replication in the human body remain unknown, but indications of a respiratory or oral route of infection led us to investigate whether the A549 lung cell line was permissive to KIPyV.

In this thesis we have investigated if the A549 cell line allows KIPyV propagation and if viral protein is detectable by antibodies directed against KIPyV VP1. This was performed by transfecting with KIPyV genome and analyzing the cells for mRNA expression of LT-ag and antibody detection of VP1 by western blotting. LT-ag mRNA was successfully detected by PCR but detection of KIPyV VP1 by the use of our antibody was unsuccessful.

The sub-cellular localization of KIPyV VP1 protein has been investigated by confocal microscopy using EGFP-KI VP1 fusion protein. Both A549 and Vero cell line were studied and the EGFP signal localized differently in the two cell lines. In A549 cells the localization was mainly in the nucleus while in Vero cells cytoplasm localization was mostly observed.

We have also examined cerebrospinal fluid from patients with suspicion of neurological diseases and urine specimens from immunocompromised patients with systemic lupus erythematosus for KIPyV DNA by nested PCR. We have indicated the presence of KIPyV VP1 DNA by nested PCR and sequencing of the PCR products.

Abbreviations

aa	Amino acid
Amp	Ampicillin
bp	Base-pairs
BKV	BK polyomavirus, used in plasmid names
BKPyV	Derives from initials of the patient in which the virus was discovered
CNS	Central nervous system
CSF	Cerebrospinal fluid
CT DNA	Calf thymus DNA
c/w	Cells per well
ctr	Control
dH_2O	Distilled and sterilized water
ddNTP	Dideoxynucleotidetri-phosphates/ dideoxynucleotide
dNTP	Deoxynucleotidetri-phosphates/ deoxynucleotide
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
GST	Glutathione S-transferase
HPyV	Human polyomavirus
HPyV6/7/9	Human polyomavirus 6/7/9
IPTG	Isopropyl β-D-1-thiogalactopyranoside
JCPyV	Derives from initials of the patient in which the virus was discovered
K	Lysine
Kan	Kanamycin
kbp	kilobase-pairs
KI	Karolinska Institutt polyomavirus, used in naming of primers and plasmids
KIPyV	Karolinska Institutt polyomavirus
LB	Lauria Bertani
LPyV	Lymphotropic polyomavirus
LT-ag	Large tumor antigen
MCC	Merkel cell carcinoma
MCPyV	Merkel cell polyomavirus
miRNA	Micro ribonucleic acid
mRNA	Messenger ribonucleic acid
NCCR	Non-coding control region
NLS	Nuclear localization signal
NPA	Nasopharyngeal aspirates
nPCR	Nested PCR
ON	Over night
ORF	Open reading frame
PBS	Phosphate Buffered Saline
PCR	Polymerase chain reaction
PFA	Paraformaldehyde
PHFG	Primary human fetal glial cells
PML	Progressive multifocal leukoencephalopathy
PP2A	Protein phosphataste 2A
pRb	Retinoblastoma protein
PyV	Polyomavirus
R	Arginine
RNA	Ribonucleic acid
RT	Room temperature
SLE	Systemic lupus erythematosus
ST-ag	Small tumor antigen
SV40	Simian virus 40
TS	Trichodysplasia spinulosa
TSPyV	Trichodysplasia spinulosa-associated polyomavirus
UNN	University hospital of Northern Norway
UV	Ultra violet
VP1-4	PyV capsid proteins
WB	Western blotting
WUPyV	Washington university polyomavirus
^{R}	Resistance gene

Table of contents

1 Introduction 1

1.1 Human polyomaviruses 1
1.2 Viral structure and genome of polyomaviruses 2
1.3 Life cycle of HPyV 5
1.4 Tropism and permissive cell cultures of HPyV 6
1.5 Seroprevalence of HPyV 7
1.6 Pathogenesis of HPyV 8
1.7 KIPyV 9
1.8 Aims of the study 12

2 Materials 13

3 Methods 19

3.3 Purification of Nucleic acid 19
3.4 Evaluation of Nucleic Acids 23
3.5 PCR 25
3.6 ExoSAP 28
3.7 DNA sequencing by capillary electrophoresis 28
3.8 Cloning in plasmid vectors 29
3.9 Mammalian cell culture techniques 37
3.10 Transfection of mammalian cells 40
3.11 SDS PAGE 42
3.12 Western Blotting 43
3.13 GST-protein purification 44

4 Results 47

4.1 Study of the permissivity of A549 cell line to KIPyV DNA 47
4.2 Study of the sub-cellular localization of KIPyV VP1 protein 63
4.3 Detection of KIPyV DNA in urine and CSF 65

5 Discussion 69

5.1 Study of the permissivity of A549 cell line to KIPyV DNA 69
5.2 Study of the sub-cellular localization of KIPyV VP1 protein 74
5.3 Detection of KIPyV DNA in urine and CSF 74

6 Conclusion and future perspective 76

1 Introduction

Polyomaviruses (PyV) are small non enveloped DNA viruses whose name is derived from the Greek words *poly* meaning many and *-oma* referring to cancer, and refers to the ability of the first known polyomavirus to induce multiple ranged tumors in mice (1). PyV have been isolated from a variety of birds and mammals, including humans. Their host range is however rather restricted and an infection in other species does not generally result in a productive viral replication (2).

1.1 Human polyomaviruses

In 1971 the first two human polyomaviruses (HPyV) were discovered, BKPyV and JCPyV. Gardner et al. detected BKPyV (patients initials B.K.) in the urine of a kidney transplant recipient, while Padgett and colleagues described the discovery of JCPyV (patients initials J.C.) in the brain of a Hodgkin lymphoma patient who suffered from progressive multifocal leukoencephalopathy (PML) (3, 4).

In 2007 two additional HPyV were discovered in nasopharyngeal aspirates (NPA), KIPyV and WUPyV (5, 6). WUPyV was discovered in a sample from a patient with acute respiratory disease while KIPyV was discovered in randomly selected samples from volunteers with respiratory tract infections. DNA from the samples were randomly amplified by polymerase chain reaction (PCR), cloned and sequenced looking for homology with other known species. The two viruses are more closely related with each other than with previously described polyomaviruses and form a new clade in the PyV family.

In 2008 Feng and colleagues identified a new PyV in Merkel cell carcinoma (MCC) tissue, called Merkel cell polyomavirus (MCPyV) (7). Subsequently the development of an improved rolling circle amplification technique led to the discovery of three additional HPyV. In 2010 Schowalter et al. described the discovery of HPyV6 and 7 in skin swabs (8) and in the same year van der Meijden and colleagues reported the isolation of trichodysplasia spinulosa (TS) associated PyV (TSPyV) in a case of the rare skin disease TS (9). A report by Scuda et al. from 2011 describes the identification of a 9th HPyV in serum from a kidney transplant patient, using degenerated primers against conserved regions in the VP1 genes of known HPyV(10).

In addition to the above described HPyV there are some polyomaviruses like simian virus 40 (SV40) and lymphotropic polyomavirus (LPyV), that can infect humans but whose natural hosts are monkeys. SV40 was transferred to humans in a contaminated polio vaccine, but has also been detected in people that have not received this vaccine (2). LPyV antibodies and DNA have also been detected in humans (11).

1.2 Viral structure and genome of polyomaviruses

PyVs are small non-enveloped viruses of 40-45 nm with icosahedral capsids. The capsid is composed of 72 pentamers consisting of five VP1 molecules and one VP2 or VP3 molecule. The viral DNA is super coiled, circular and double-stranded, and consists of approximately 5 kilo base pairs (kbp). The organization of the PyV genome is conserved and can be organized into three functional regions (**Figure 1.1**). The early region contains the genes that are expressed before DNA replication begins and codes for at least two regulatory proteins, large tumor antigen (LT-ag) and small tumor antigen (ST-ag). The late region contains the genes that are expressed after the DNA replication has started and codes for the structural proteins of the capsid (VP1-3). BKPyV, JCPyV and SV40 also have an agnoprotein coding gene in the late region. The regulatory region or non-coding control region (NCCR) contains the single origin of replication, the promoters and regulatory regions for the early and late genes. The early and late regions are non-overlapping in their coding sequences and give rise to primary transcripts from opposite strands of the DNA (2).

In addition to the early and late proteins miRNA has been described for SV40, BKPyV, JCPyV and MCPyV. Studies of SV40 have reported that a miRNA down-regulates LT-ag expression levels and may therefore aid in avoiding detection by the hosts immune system. Cells expressing SV40 miRNA displayed lower susceptibility to cytotoxic T-cells and lowered the cytokine expression compared to cells infected with a SV40 mutant lacking miRNA (12).

Several phylogenetic analyses based on DNA or amino acid sequences have shown that the PyVs are divided into genetically related clades. BKPyV and JCPyV are grouped together with the monkey PyV SV40, while KIPyV forms a clade with WUPyV and is also more distantly related to HPyV6 and HPyV7. HPyV9 is closely related to the monkey LPyV and more distantly related to the TSPyV and MCPyV. This has most recently been described by Scuda et al in 2011 (10).

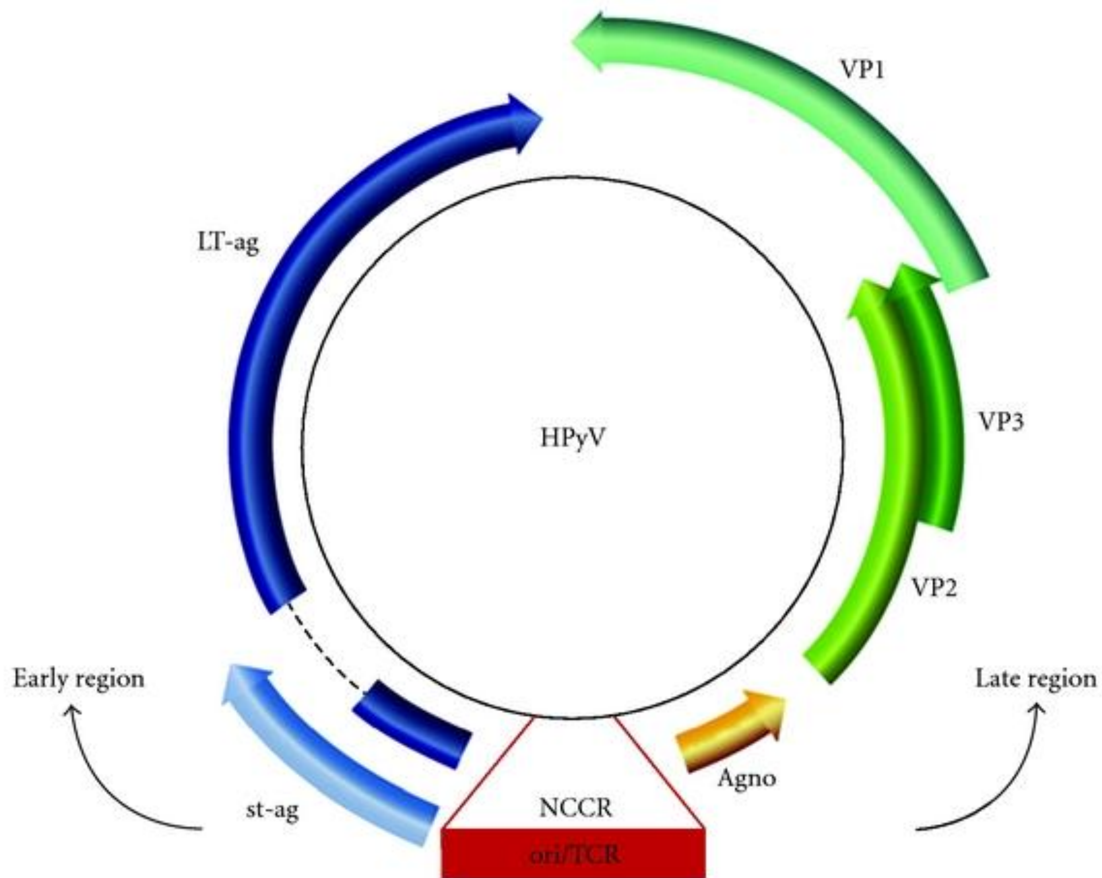


Figure 1.1: Schematic presentation of the functional organization of the HPyV genome. The viral genome is divided into a non-coding control region (NCCR), early region and late region. The NCCR contains the origin of replication, the early and late promoters and regulatory regions. The early region encodes the large tumor antigen (LT-ag) and small tumor antigen (st-ag) and some HPyV encode additional early proteins. The late region codes for the capsid proteins VP1, VP2, VP3 and some HPyV also encode an agnoprotein. The early and late regions are transcribed from opposite strands of the DNA. The figure is retrieved from Moens 2011 (13).

1.2.1 The early proteins

The early genes of the HPyV encode two major proteins, the LT-ag and ST-ag. The proteins are expressed by alternative splicing of a single primary transcript. Some PyV also encode additional early proteins, e.g. mouse PyV encodes a middle T antigen and SV40 has the 17k T antigen that shares the first 131 amino acids of the large T antigen and terminates with 4 unique amino acids (2). JCPyV and BKPyV have additional early proteins (14) and MCPyV expresses the 57kT protein (15). Other HPyV have putative ORFs in their early region, although the expression of these potential proteins remains to be proven.

The LT-ag contains ~700 amino acids. It plays a major role in the regulation of the viral life cycle. In the early phase of the infection LT-ag induces the cell to enter a phase of high DNA replication and gene expression. In this way the virus exploits the cell's DNA replication machinery to produce viral DNA and express viral proteins. LT-ag is composed of

several functional domains including the DnaJ domain, the retinoblastoma (pRb) binding domain, the p53 binding domain, a nuclear localization signal (NLS), a DNA binding domain, and the helicase domain. The LT-ag pushes the targeted cell into this high activity phase through binding and inactivating the retinoblastoma susceptibility protein pRb and two other retinoblastoma protein family members, p107 and p130. This inactivation results in the release of the transcription factor E2F which in turn stimulates the transcription of genes involved in the phase switching and DNA synthesis. The DnaJ domain enables transformation of cells through binding of HSc70, which affects the cell cycle and gene expression. The LT-ag plays a more direct role in the DNA replication by binding to specific sequences (i.e. GAGGC) in the origin of replication in the viral DNA and recruiting the host DNA synthesis machinery. The LT-ag also possesses helicase activity and is involved in the unwinding of the DNA at the origin (16). The LT-ag therefore has several oncogenic potentials. It can inhibit apoptosis, stimulate telomerase activity, modulate protein turn-over, affect signaling pathways and gene expression, disturb chromosome fidelity, and induce angiogenesis (14).

After the replication has been initiated LT-ag stimulates the transcription of the late genes and represses transcription of the early genes LT-ag and ST-ag. The ST-ag consists of ~175 amino acids, where the first 80 are shared with the LT-ag. The ST-ag protein appears to have a supplementary role in stimulating cell growth and lytic activity, while its major contribution in the transformation may be through inactivation of protein phosphatase 2A (PP2A) (2, 16).

1.2.2 The late proteins

The late region consists of genes coding for the capsid proteins VP1, VP2 and VP3 and some PyVs also have genes coding for the additional proteins agno or VP4. VP4 has been described for SV40 and is a protein that enhances lysis of the host cell and facilitates release of mature virions (17). The recently described human polyomaviruses KIPyV, WUPyV, MCPyV, TSPyV, HPyV6, HPyV7 and HPyV9 all lack an open reading frame (ORF) corresponding to VP4 or agnoprotein, while BKPyV and JCPyV contain a putative VP4 ORF although the expression of this protein has not been confirmed. The late region of BKPyV, JCPyV and SV40 encodes the agnoprotein consisting of ~70 amino acids. The function of this protein is not completely understood, but it is believed that agnoprotein facilitates the release of newly formed virions from infected cells and may contribute to viral genome transcription, translation of the late genes and viral assembly (18-20). The reason why some HPyV have the

genes coding for agnoprotein or VP4 while others do not is currently not known, but may be the result of different cell tropism and differences in the viral life cycle such as release from the infected host cell.

1.3 Life cycle of HPyV

The full infectious viral life cycle of the HPyVs has only been studied in BKPyV and JCPyV due to a lack of cell systems supporting viral replication for the more recently discovered ones. The general notion is that the PyV VP1 protein makes contact with the host cell by attaching to cell surface receptors and then enters the cell by endocytosis. The studied PyVs do not share receptor specificity or endocytic pathways when entering the cell. BKPyV have been reported to use both gangliosides (GT1b and GD1b) and glycoproteins, which both have the presence of sialic acid. JCPyV also attaches to a sialic acid receptor and a serotonin receptor. SV40 and BKV enter the cell through caveolae-mediated endocytosis while JCV enters through clathrin-dependent endocytosis (2, 21).

The receptors or endocytic pathways are not known for the recently discovered HPyV, but indications are that ganglioside GT1b may be a putative receptor for MCPyV (13). However, it is believed that a co-receptor providing optimal binding and further cell specificity is needed (13). Structural studies have shown that the core structure of the VP1 proteins of KIPyV and WUPyV have a high identity with the VP1 proteins in other PyVs while the surface loops have profoundly different conformations. None of the oligosaccharide binding residues or sialic-acid binding residues found in other PyVs have been identified in the KIPyV VP1 surface structure; although it is possible that the virus might still bind these types of receptors in a different manner or orientation. Other possibilities are that the VP1 surface binds to other carbohydrates or proteinaceous receptors (22).

The virion gains entry first through the hosts cellular transport pathways and is subsequently transported to the nucleus through the nuclear pores. Replication of the viral genome and assembly of mature viral particles occur in the nucleus. The N-terminal region of VP1 of most HPyV contains a nuclear localization signal (NLS), while VP2 and VP3 contain such a signal in their C-terminal region. This NLS allows nuclear transport of the capsid proteins that are synthesized in the cytoplasm and are masked before the mature virion leaves the nucleus. Because PyV VP2 and VP3 proteins also contain NLS motifs and VP1 binding domains they can interact with VP1 and support each other in the transportation into the nucleus (23, 24). The NLS signal is generally made up by a stretch of the basic amino acids

lysine (K) and arginine (R) arranged in one or two clusters (24, 25). The method by which the PyV leaves the cell is not completely understood. Studies have both indicated that the PyV can cause cell lysis and that the virus can be shed from intact cells (2).

1.4 Tropism and permissive cell cultures of HPyV

HPyV show a distinct host cell restriction, while some avian, rodent and monkey PyV can productively infect different cell types (2). JCPyV is considered a neurotropic virus, but it can also establish a persistent infection in the urinary tract and can infect blood cells. BKPyV is a nephrotropic virus, but also resides in blood cells. DNA of the HPyV has been detected in a variety of tested tissues and fluids which means they spread efficiently in their natural hosts (2). Although MCPyV, HPyV6, HPyV7 and TSPyV are chronically shed from the skin and may therefore be dermatotropic viruses (13), the natural host cells for these viruses remain to be determined. The replication site and thereby the genuine host cells for KIPyV, WUPyV and HPyV9 are not known.

PCR studies have been performed to identify where in the human body the PyVs reside and what cell types may be permissive. This information can be used to find a suitable cell culture for the viruses to propagate in for further studies. KIPyV DNA has been detected with various frequencies in blood, feces, urine, NPA, lymphoid tissue (e.g. tonsils), lung tissue, brain tissue, but not yet in cerebrospinal fluid (CSF). A detailed overview of the prevalence for KIPyV in different biological specimens is provided in **Supplementary Table 1**. WUPyV DNA has been detected in the same types of samples as KIPyV, but has in addition been detected in CSF (26). The highest prevalence of KIPyV and WUPyV DNA has been shown in samples from stool, tonsils and the respiratory tract, suggesting that the lung or gastro intestinal tract may be host tissue for these two viruses (11, 13).

MCPyV was detected by Feng and coworkers in tissue of MCC, a rare and aggressive form of skin cancer (7). It is probable that MCPyV has other cell tropisms as Merkel cells make up less than 1 % of the epidermis while the virions are shed continuously from the skin at high levels (8, 13). In the healthy population MCPyV DNA is frequently detected in skin, tonsils and NPA. TSPyV is another PyV that appears to have a skin tropism. It was first detected by van der Meijden and colleagues in the spicules obtained from a transplant patient with TS (9). TS is a rare skin disease which only affects immunocompromised patients and is characterized by the development of follicular papules and spicules mainly on the nose and

eyebrows (27). The same group revealed that TSPyV was also detectable in a low copy number in plucked eyebrows from immunocompetent transplant patients without TS.

BKPyV can propagate with high viral yields in human or monkey epithelial or fibroblast cell cultures. The production of viral plaques in cultures can take weeks, but viral proteins can be detected one to two days post infection. JCPyV can be cultured in primary human fetal glial cells (PHFG) as the most sensitive cell culture, but tonsillar stromal cells, urothelial cells and HEK cells amongst others can also be used (2, 21). Cell cultures in which any of the new PyVs can propagate have not been identified yet.

1.5 Seroprevalence of HPyV

Serological studies have been performed on HPyV to better understand the spread in the human population. By looking at the seroprevalence throughout the lifetime it is possible to suggest when the initial infection occurs. BKPyV and JCPyV are closely related and show a prevalence of up to 90 % and 60%, respectively, in the human population (13). Both BKPyV and JCPyV have been suggested to cause a primary infection during early childhood with a near adult level of seroprevalence reached before adolescence (2). The seroprevalence for the monkey PyV SV40 and LPyV in the human population is 2 % and 15 %, respectively (28). However, the seroprevalence of LPyV may be overestimated (see further).

KIPyV and WUPyV are common in the human population with seropositivity between 55 % and 90 % (28-30). In a pediatric test group of 721 persons younger than 21 years old 56% were seropositive for KIPyV and 54 % for WUPyV, and from the children less than five years old 44 % were seropositive for KIPyV and/or WUPyV (28). These results indicate that initial infection with these viruses happens at an early age.

HPyV 6 and 7 were first detected in skin swabs by Schowalter and colleagues, and have a seroprevalence of 69 % and 35 %, respectively (8). The seroprevalence values for MCPyV in the healthy human population are between 25 and 42 %, and studies have indicated that initial exposure to MCPyV happens during childhood (13). A TSPyV seroprevalence study in Finland has shown a prevalence of 70 % for an adult population, and that the primary exposure occurs at early childhood with a seroprevalence of 30-48 % before the age of 10 (31). HPyV 9 was first isolated from the blood of a kidney transplant patient by Scuda et al. VP1 sequencing analysis revealed a high deduced amino acid similarity to the LPyV VP1 (87%), which Scuda and colleagues suggest might explain the high seropositivity of LPyV in the human population. The amino acid similarity provides the theoretical possibility of a cross

reaction with antibodies against LPyV (10). As most of the HPyVs are newly discovered the studies of their seroprevalence are still in their infancy and there is much yet to discover.

1.6 Pathogenesis of HPyV

In permissive host cells the virus generally establishes a lytic infection resulting in cell death, while in non-permissive hosts the viral replication is blocked leading to aborted infection or transformation and oncogenicity of the infected cells (16). PyVs are capable of establishing latent or persistent infections and can be detected in healthy individuals. In fact, most HPyVs seem to co-exist harmlessly in immunocompetent individuals.

1.6.1 Disease

Primary infection with BKPyV and JCPyV is rarely linked to disease in healthy children, but when symptoms are present they are usually of a respiratory character (21). The primary infection route for the BKPyV and JCPyV has not been established, but a fecal-oral route is suspected (21). After the primary infection with BKPyV or JCPyV a state of viremia follows where the virus is transported to permissive cells. BKPyV and JCPyV enter a latent or persistent phase after primary infection, usually in the kidneys (21). Suppression of the immune system can lead to reactivation of the virus and consequently lead to diseases, including malignancy. Reactivation of BKPyV and JCPyV has been known to cause viruria not only in immunocompromised but also immunocompetent individuals, although very low virus loads are registered (32). The rare detections of KIPyV and WUPyV in urine may imply that they do not reside in the kidneys like BKPyV or JCPyV.

Of the HPyV only BKPyV, JCPyV and TSPyV have been linked to non-malignant diseases. The role of KIPyV, WUPyV, HPyV6, HPyV7 and HPyV9 in human disease has not been sufficiently studied to associate any of these viruses with pathogenicity, and in most findings they are co-infections (33, 34). Common for all HPyV is that induction of disease is regularly linked to immunodeficiency. BKPyV and JCPyV are associated with persistent infection and disease of the urogenital tract. BKPyV can cause nephropathy or hemorrhagic cystitis in transplant patients. JCPyV is the etiological agent of PML a rare and usually fatal disease that is characterized by progressive inflammation at multiple locations of the white matter of the brain which can affect AIDS patients (2). As mentioned earlier, TSPyV may be the etiological agent in trichodysplasia spinulosa (27), but before a definite link between

TSPyV and TS can be made additional studies on the occurrence and causal role of TSPyV needs to be performed (13).

1.6.2 Cancer

Under normal circumstances a cells response to forced phase change is to turn on a p53 response that leads to cell arrest or apoptosis. The PyV overcomes this response with the help of LT-ag which binds and inhibits p53. This way the cell is induced to proliferate without inhibition. Under a lytic infection the cells are kept alive to produce high viral yields before the host cell is killed (16). Oncogenic transformation occurs when the virus infects a cell and the early genes are expressed, resulting in cell proliferation. However, the viral replication is not completed and the cell is not killed, causing tumor formation. This can be an effect from the host cell not being permissive to the virus or of an infection by a viral mutant incapable of completing a lytic replication (2). The agnoprotein has been shown to inhibit p53 and apoptosis and to interfere with DNA repair making the cell more prone to oncogenesis (18), while the additional early proteins expressed by some of the PyV may be involved in oncogenesis by influencing the cell cycle and gene expression (13).

All of the known HPyV have the potential to be oncogenic because they all produce LT-ag and ST-ag. However, the tumorigenic role of most HPyV has not yet been sufficiently studied, and for BKPyV, JCPyV and SV40 the role in human cancers remains controversial. One HPyV that may be an oncovirus is MCPyV, which is believed to be the causal factor in Merkel cell carcinoma (13). Studies have shown that around 80 % of examined MCC tumors contain MCPyV DNA which suggests that MCPyV might be the etiological agent in development of these tumors (11).

1.7 KIPyV

KIPyV has since it was first discovered in 2007 been detected in many studies of samples from immunocompromised and immunocompetent individuals around the world. Seroprevalence studies have demonstrated that exposure to KIPyV is common in the human population with seropositivity up to 90 % (28-30). Like other known HPyV the primary infection with KIPyV is indicated to occur during early childhood as adult seropositivity levels are reached before the age of 12 (29). PyVs generally infect humans asymptomatically, but when symptoms have been reported they have been of an upper respiratory character (34). This might suggest a respiratory route of infection for KIPyV. Therefore most studies aimed

at detection of KIPyV DNA have been performed on samples from the respiratory system (NPA, swabs, lung tissue) in patients with respiratory symptoms. The prevalence in respiratory samples is quite low with 5 % and infections with KIPyV are in most cases accompanied by co-infections with known pathogens, which might exclude KIPyV as the primary cause of these symptoms. KIPyV is indeed regularly detected in such samples (**Supplementary Table 1**); however KIPyV DNA was also found in feces, lymphoid tissue (e.g. tonsils), CNS, blood and urine. The detection of KIPyV DNA in tonsils (35, 36) and feces (5, 36-39) opens up the possibility of an oral route of infection as is suggested with BKPyV and JCPyV. KIPyV DNA has been detected in urban sewage which supports an oral route of transmission (40). The site of replication remains elusive, as well as the pathogenic properties of this virus. KIPyV DNA has also been amplified in different cancer tissues (**Supplementary Table 1**) but a causal role for KIPyV in these cancers remains to be proven.

The genome of the KIPyV has the same functional organization as the other PyVs, with an early region coding for regulatory proteins, a late region coding for structural proteins and a NCCR (**Table 1.1**). The early region consists of two putative ORF's for the regulatory proteins, ST-ag and LT-ag. The first 82 amino acids in the N-terminal are shared between the LT- and ST-ag. LT-ag contains the J-domain, a putative pRb binding domain and a p53 binding domain. In other HPyV it was shown that these domains contribute to the oncogenic properties of LT-ag (14). LT-ag also possesses a nuclear localization signal, a DNA binding domain, a zinc finger region and an ATPase domain, but the functionality of these domains has not yet been tested. The ST-ag has a cystine rich C-terminal which is common for all HPyV and is involved in binding PP2A. It is however not known whether KIPyV ST-ag binds PP2A.

The late region contains three putative ORFs for the capsid proteins VP1-3. The VP3 protein is encoded within the same ORF as the VP2 protein, and the VP1 N-terminus overlaps the C-terminus of the VP2/3, both are structural properties common for other PyV. VP2 and VP3 have extremely low identity with other PyV, and are only indicated by the positioning in the genome (5). The protein sequences contain a putative VP1 binding domain with homology to those of other HPyVs (**Supplementary Figure 1**). As mentioned earlier the N-terminal region of VP1 of most HPyV contains a NLS motif. Sequence analysis shows that KIPyV VP1 also has a NLS like motif (**Supplementary Figure 2**), although compared to other HPyV the motif consists of only two basic amino acids with a few more in near vicinity which might provide a weaker signal. The NCCR does not show any sequence homology with the

other HPyV, except for the presence of LT-ag binding motifs and stretches of AT-rich sequences (5).

The amino acid sequences of the structural proteins are more divergent than the sequences of the regulatory proteins compared to those of other HPyV (**Table 1.1**). KIPyV lacks the agno gene and so far does not have any other putative genes than the two common T antigens and three capsid proteins. Phylogenetic analyses where the PyVs are divided into genetically related clades describe KIPyV in a clade with the highly similar WUPyV and its more distantly related HPyV6 and HPyV7 (8, 10). KIPyV and WUPyV's early genes most resemble those of SV40, BKPyV and JCPyV, while their late genes are quite different, which is why they form a separate clade.

There is a scarcity of knowledge relating to the life cycle and on the pathogenic properties of KIPyV. The isolation of infectious virus and the discovery of a cell line susceptible to infection have to date not been successful, achieving this would be beneficial in the study of KIPyVs biology.

Table 1.1: Description of the putative proteins of KIPyV. The information is retrieved from (5, 6).

Protein	Putative coding region(s)	No. of aa	Calculated mass (kDa)	% aa identity to:			
				BKPyV	JCPyV	SV40	WUPyV
VP1	1498-2634	378	41.6	29	30	29	65
VP2	441-1643	400	41.8	22	23	22	71
VP3	870-1643	257	28.2	22	24	22	64
LT-ag	4967-4716, 4328-2655	641	74.3	48	47	47	70
ST-ag	4967-4392	191	23.2	37	36	40	68

1.8 Aims of the study

The study of KIPyV biology has been greatly restricted through the absence of a suitable cell culture system which is required for viral propagation. The bona fide sites of viral replication in the human body remain unknown. Moreover, previous studies with BKPyV and JCPyV have shown viral activation and viruria in immunocompromised patients, but studies on KIPyV activity in immunodeficient patients are sparse. The objectives of this study were therefore to try to establish a permissive cell line for KIPyV and to increase our knowledge on the pathological conditions that can lead to activation of KIPyV virus. For this purpose, following studies were undertaken:

- To investigate whether the human lung carcinoma cell line A549 allows replication of KIPyV
- To validate the specificity of an antibody directed against KIPyV VP1 and its use in detection of viral infections
- To examine the sub-cellular localization of KIPyV VP1
- To investigate the presence of KIPyV in urine of samples from immune compromised patients (systemic lupus erythrematosus (SLE) patients) and in cerebrospinal fluid from patients with suspicion of neurological disorders.

2 Materials

The patient samples used in this study were received from the University hospital of Northern Norway (UNN) Tromsø. Urine samples (73 samples) were from 5 patients with SLE. CSF samples (64 samples) were received from the neurological department gathered from patients with the suspicion of or suffering from neurological disorders. No additional information about the samples or patients is available.

Table 2.1: Kits used in this thesis

Kit	Manufacturer	Purpose
Nucleospin® Plasmid	Macherey Nagel	Plasmid purification, small quantities.
Nucleobond® Xtra Midi	Macherey Nagel	Plasmid purification, medium quantities
Nucleospin® RNAII	Macherey Nagel	RNA isolation
QIAamp® MinElute® Virus Spin	Qiagen	Viral DNA purification
GFX™ PCR DNA and Gel band purification	GE Healthcare	Purification of DNA from gel or solution
iScript cDNA Synthesis	Bio-Rad	cDNA synthesis

Table 2.2: Buffers and solutions used in this thesis

Buffers and solutions	Manufacturer/ Contents	Purpose
PBST	PBS with 0.1% Tween 20	Western blot
50 x TAE (Tris Acetat EDTA)	242 g Tris base, 57.1 ml glacial acetic acid, 100 ml 0.5 M EDTA (pH 8.0), dH_2O up to 1 L	Agarose gel electrophoresis
6 x Loading Buffer	0.25 % bromphenol blue, 40% agarose	Agarose gel electrophoresis
SeaKem® LE Agarose	Lonza	Agarose gel electrophoresis
JumpStart™ Taq ReadyMix™	Sigma	PCR
Big Dye® Terminator v3.1 5 x Sequencing buffer	Applied Biosystems	Sequencing
10x TA Buffer	330 mM Tris-acetate (pH 7.5), 660 mM potassium acetate, 100 mM magnesium acetate, and 5 mM DTT	Restriction digestion
NEBuffer (1-4)	New England Biolabs	Restriction digestion
10x T4 DNA ligase buffer	New England Biolabs	Restriction ligation
Trypsin/EDTA	Lonza. 0.25 % Trypsin in PBS and 0.05% Na_2-EDTA	Mammalian cell culture
Dulbecco's PBS without Ca^{2+} and Mg^{2+}	Biochrom. Dissolve the powder content in 5 L of dH_2O (1xPBS with pH 7.4)	Mammalian cell culture, western blot and fixation of cells for confocal microscopy

Buffers and solutions	Manufacturer/ Contents	Purpose
Sample buffer	50 % NuPAGE® LDS Sample Buffer (4X), 40 % H2O, 10 % 1M DTT	Harvesting cells
NuPAGE® LDS Sample Buffer (4X)	Invitrogen	Western blotting
Nu Page 20xRunning buffer	Invitrogen, working dilution 1:20	Western blotting
Blotting buffer	5.8 g Tris base+ 29 g glycin + 200 ml methanol + 800 ml dH_2O	Western blotting
Blocking buffer	150 ml PBS + 7.5 g dry milk + 150 µl Tween 20	Western blotting
10x Washing buffer	100 mM Tris HCl pH9.5, 100 mM NaCl, 10 mM $MgCl_2$ and dH_2O up to 1 L. Working dilution 1:10	Western blotting
Tropix® CDP-Star®	Applied Biosystems	Western blotting
CDP star buffer	10 ml DEA + 850 ml ddH_2O. pH 9.5. dH_20	Western blotting
1xTE	100 mM Tris/10 mM EDTA	DNA storage
PFA (8%)	0.8 g paraformaldehyde dissolved in 9 ml dH2O. 1 M NaOH added drop wise until paraformaldehyde is dissolved, 1 ml 10x PBS	Fixation of cells for confocal microscopy
10x Buffer CIP (*Calf Intestinal Alkaline Phosphatase*)	Finnzymes	Preventing relegation of plasmids
PBT	PBS with 1% Triton X-100	GST protein purification
Glutathione Beads	GE Healthcare. Washed 3x in PBT before resuspension to 50 % in PBT and kept at 4ºC >2 hr prior to use	GST protein purification
Inhibitor cocktail tablets	Roche. 1 tablet dissolved in 2 ml H_2O	GST protein purification
5 mM glutathione	Diluted in 50 mM Tris pH 8.0.	GST protein purification
Protein Assay	Bio-Rad	GST protein quantification
Coomassie Blue solution	0.25 g Coomassie brilliant blue R-250, 250 ml Methanol, 50 ml Acetic acid, 200 ml dH_2O	Coomassie blue staining
Fixation solution	400 ml dH_2O, 500 ml Methanol, 200 ml Acetic acid	Coomassie blue staining
Destaining solution	880 ml dH_2O, 50 ml Methanol, 70 ml Acetic acid	Coomassie blue staining

Table 2.3: Molecular markers used for agarose and acrylamide gel electrophoresis in this thesis

Molecular markers	Manufacturer	Purpose
GelRed™	Biotium	Agarose gel electrophoresis
1 kb Plus DNA ladder	Invitrogen	Agarose gel electrophoresis
SeeBlue® Plus 2 Prestained Standard (1x)	Invitrogen	Western blotting
MagicMarker™ XP Western Standard	Invitrogen	Western blotting

Table 2.4: Primers used in this thesis. The primer sites for all but the EGFP primers are described in **Supplementary Figure 3**, and the EGFP primer sites are included in **Supplementary Figure 4**.

Analysis (product size)	Primer	Sequence	Source
NCCR (514bp)	NCCR F	5' GCA-TTA-GCT-GCT-TTG-CCT-CT 3'	This study
	NCCR R	5' GGT-GAC-CCT-CTA-TAT-CCA-AAG-GT 3'	This study
KI VP1 (324 bp)	KI VP1 39 F	5' AAG-GCC-AAG-AAG-TCA-AGT-TC 3'	Allander et al. 2007 (5)
	KI VP1 363 R	5' ACA-CTC-ACT-AAC-TTG-ATT-TGG 3'	Allander et al. 2007 (5)
Nested KI VP1 (206bp)	KI VP1 118 F	5' CGC-AGT-ACC-ACT-GTC-AGA-AGA-AAC 3'	Allander et al. 2007 (5)
	KI VP1 324 R	5' TTC-TGC-CAG-GCT-GTA-ACA-TAC 3'	Allander et al. 2007 (5)
APRT gene (720bp)	APRT F	5' GGG-GAA-GCT-GCC-AGG-CCC-CAC-T 3'	This study
	APRT R	5' AGC-CTG-GTG-GAG-CTG-ACC-TCG-C 3'	This study
LT-ag (~440 bp)	KI LT-ag F	5' TGG-CAA-TCT-TCT-CAG-ATA-CCT-ACA-TAC-GG 3'	This study
	KI LT-ag R	5' GCA-CTA-ACT-CTA-TGC-TTG-TGA-GGA-G 3'	This study
CREB (358 bp)	h CREB F	5' ATG-GAA-TCT-GGA-GCC-GAG-AAC 3'	This study
	h CREB R	5' TCC-TGT-AGG-AAG-GCC-TCC-TTG 3'	This study
Cloning KI VP1 (~1.2 kb)	VP1 SalI (100 ng/µl)	5' TCT-GCA-GTC-GAC-ATG-AGC-TGC-ACC-CCG-TGT-CGC-CCA 3'	This study
	VP1 SacII (100 ng/µl)	5' CGG-GCC-CGC-GGT-TCA-CTT-TGA-ATT-TTG-TTG-AGT 3'	This study
Seq EGFP (664 bp)	pEGFP-C1F	5' TAT-ATC-ATG-GCC-GAC-AAG-CA 3'	This study
	pEGFP-C1R	5' CGA-TTT-CGG-CCT-ATT-GGT-TA 3'	This study
Seq KI VP1	KI VP1 2071-2093 F	5' GGA-GAC-CCT-AGA-ACA-CTG-CAT-GT 3'	This study

Table 2.5: Bacterial strains used in this thesis

Bacterial strain	Description	Purpose
Escherichia coli DH5α	A recombination-deficient, suppressing, competent strain	Amplification of plasmid vectors
Escherichia coli BL21	A protease deficient, competent strain	Expression of GST fusion proteins

Table 2.6: Plasmid Constructs used in this thesis, * Theoretical mass of the encoded protein or fusion protein

Plasmid construct	Size of plasmid	Mass of protein*	Source	Properties	Purpose
pCR2.1	3.9 kbp	-	Invitrogen	Kan^R, Amp^R	Cloning of KIPyV VP1
pSL301	3.2 kbp	-	Invitrogen	Amp^R	Cloning of KIPyV VP1
pGEX-4T-3	4.9 kbp	26 kDa	GE Healthcare	Amp^R	Construction of GST-KI VP1.
pUC18.KI	7.8 kbp	-	Allander et al. 2007 (5)	Amp^R	Amplification of KIPyV genome
pUC19.BKV	7.9 kbp	-	Moens	Amp^R	Amplification of BKPyV genome
pEGFP-BKV VP1	5.9 kbp	70 kDa	Moens	Kan^R	Control of transfection and WB
pRc-CMV-agno	5.7 kbp	8 kDa	Rinaldo et al. 1998 (41)	Amp^R	Transfection
pEGFP-C1	4.7 kbp	29.4 kDa	Clontech	Kan^R	Control for transfection and WB. Construction of EGFP-KI VP1.
pcDNA™3.1$^{(+)}$	5.4 kbp	-	Invitrogen	Amp^R	Construction of pcDNA™3.1$^{(+)}$-KI VP1
CR2.1-KI VP1	5.1 kbp	-	This study	Kan^R, Amp^R	Intermediate cloning product
EGFP-KI VP1	5.9 kbp	71.4 kDa	This study	Kan^R	Control for transfection and WB
SL301-KI VP1	4.4 kbp	-	This study	Amp^R	Intermediate cloning product
GST-KI VP1	6.1 kbp	67.6 kDa	This study	Amp^R	Production of protein control for WB
pcDNA3.1$^{(+)}$-KI VP1	6.6 kbp	41.6 kDa	This study	Amp^R	Expression control for mammalian cell culture

Table 2.7: Enzymes used in this thesis

Enzyme	Manufacturer	Purpose
T4 DNA ligase	New England Biolabs	DNA ligation
Nde I (20 000 U/ml)	New England Biolabs	Restriction digestion
Bam HI (20 000 U/ml)	New England Biolabs	Restriction digestion
SalI (10 U/µl)	Promega	Restriction digestion
SacII (20 000 U/ml)	New England Biolabs	Restriction digestion
Not I (20 000 U/ml)	New England Biolabs	Restriction digestion
Big Dye® Terminator v3.1	Applied Biosystems	Sequencing
ExoSAP-IT® (*Exonuclease I* and *Shrimp Alkaline Phosphatase*)	USB®	ExoSAP treatment of PCR product
CIP (*Calf Intestinal Alkaline Phosphatase* (10U/µl)	Finnzymes	Dephosphorylation of 5' ends to prevent relegation of plasmids

Table 2.8: Growth media used in this thesis

Growth media	Manufacturer / Contents	Purpose
LB^+	950 ml dH_2O, 10 g bactotryptone, 5 g yeast extract, 10 g NaCl, NaOH to pH 7.0 (~0.2 ml), appropriate antibiotics, dH_2O up to 1 L	Bacterial culture
SOC	950 ml dH_2O, 20 g bactotryptone, 5 g bacto-yeast extract, 0.5 g NaCl, 20 mM glucose, 10 ml 250 mM KCl, NaOH to pH 7.0 (~0.2 ml), dH_2O up to 1 L	Transformation of bacterial cells
LB agar plate	LB medium, 15 g bacto-agar per L	Transformation and cloning of bacterial cells
NZCYM	950 ml dH_2O, 10 g NZ amine, 5 g NaCl, 5 g bacto-yeast extract, 1 g casaminoacids, 2 g $MgSO_4$, NaOH to pH 7.0, dH_2O up to 1 L	GST protein purification
DMEM	Sigma. Standard Dulbecco's Modified Eagle's medium, penicillin (100 U/ml), Streptomycin (100 µl/ml)	Mammalian cell culture
EMEM	Lonza. Eagle's Minimum Essential Medium, penicillin (100 U/ml), Streptomycin (100 µl/ml)	Mammalian cell culture
F-12K	Gibco®. Kaighn's modification 1 x, + L-Glutamine	Mammalian cell culture
FBS	Gibco®. Heat inactivated Fetal Bovine Serum	Mammalian cell culture
Opti-MEM®	Gibco®. GlutaMAX™ I, 2.4 g/L Sodium Bicarbonate, HEPES, Sodium pyruvate, Hypoxanthine, Thymidine, L-glutamine, Trace elements, Growth factors, 1.1 mg/L Phenol Red	Transfection of mammalian cell culture

Table 2.9: Mammalian cell lines used in this thesis

Cell-line	Organism	Organ	Reference number	Purpose
A549	Human	Lung	CCL-185	Transfection
HEK239	Human	Kidney	CRL-1573	Transfection
Vero	African green monkey	Kidney	CCL-81	Transfection

Table 2.10: Transfection reagents used in this thesis

Transfection reagent	Manufacturer	Purpose
Lipofectamine™ 2000	Invitrogen	Transfection of mammalian cell cultures
Metafectene® Pro	Biontex	Transfection of mammalian cell cultures

Table 2.11: Antibodies used in this thesis. * Antibodies produced by immunizing rabbits with a mixture of two peptides derived from KIPyV VP1. These peptide sequences are conserved in WUPyV, but not in MCPyV, BKPyV, JCPyV and SV40. When these antibodies were ordered, the HPyV6, 7, 9 were not known. The amino acid sequences of the peptides are: APPDIPNQVSECDM and VPLSEETEFKVELFV, **non-specific.

Antibody	Manufacturer	Dilution	Purpose
EP 101303 KI VP1 Polyclonal Rabbit *	Eurogentec	1:1000 1:600 1:500	Primary antibody for detection of KIPyV VP1 protein in Western blot
EP 101304 KI VP1 Polyclonal Rabbit *, **	Eurogentec	1:1000	Primary antibody for detection of KIPyV VP1 protein in Western blot
Goat Anti-Rabbit Ig, Human ads-AP	Southern Biotech	1:2000	Secondary antibody in Western blot

Polyclonal Rabbit Anti-mouse Ig/AP	Dako	1:1000	Secondary antibody in Western blot
Anti-GFP IgG Mouse	Roche	1:1000	Primary antibody for detection of GFP and GFP fusion proteins in Western blot
ERK 2 (C-14) Rabbit polyclonal IgG	Santa Cruz Biotechnology	1:1000	Primary antibody for detection of ERK2, p42 protein in Western blot

Table 2.12: Equipment used in this thesis

Equipment	**Manufacturer**	**Purpose**
Sub Cell System	Bio-Rad	Agarose gel electrophoresis
Gel Doc 2000	Bio-Rad	Agarose gels and Coomasie blue stained SDS-Page Photo documentation
Avanti® J-26 XP	Beckman Coulter™	Centrifugation of ≥ 15 ml tubes
Microfuge® 22R Refrigerated Centrifuge	Beckman Coulter™	Centrifugation of eppendorf tubes
Chamber slide 15 µ-Slide 8 well for Life Cell Analysis	Ibidi	Confocal microscopy
LSM 510 META	Zeiss	Confocal microscopy
KI 260 Basic	IKA®	Flat shaker
Leica Fluorescence microscope DM IRB	Leica	Fluorescent microscopy
AccuBlock™ Digital Dry Bath	Labnet	Heating block
Vortex	VWR	Mixing
ND-1000	Thermo Scientific	Nucleic acid measurement
THERMO$_{MAX}$ Microplate reader	Heigar	Protein measurement
Spectrafuge™ Mini Centrifuge	Labnet	Quick spin
XCell SureLock™ Mini-Cel	Invitrogen	SDS page/Western blotting
Bürker counting chamber		Seeding out cells for transient transfection
GeneAmp® PCR System 9700	Applied Biosystems	Thermal cycling
Rotator SB3	Stuart	Tube rotator
TW8	Julabo	Water bath
Immobilon®-P Transfer Membrane pore size 0.45 µm	Millipore®	Western blotting
Chromatography paper 3 mm	Whatman/ GE Healthcare	Western blotting
NuPAGE® 4-12 % Bis-Tris gel	Invitrogen	Western blotting
LAS-3000	Fujifilm	Western blotting Luminescent Image Analyzer

3 Methods

3.3 Purification of Nucleic acid

In this thesis several different protocols for purification of nucleic acids were used according to the amount, source and type of nucleic acid desired. All the protocols are column based and depend on releasing the nucleic acid from its source prior to loading. The nucleic acid binds to a silica-based membrane in high salt and pH conditions, and after various washing steps the nucleic acid is eluted. A brief description of the kits used for purification of nucleic acids are presented in **Table 3.1** followed by an in depth description of the protocols.

Table 3.1: Description of the purification kits used in this thesis

Kit	Nucleic acid	Source of nucleic acid	Method	Specifications
Nucleobond® Xtra Midi	Plasmid DNA	Medium sized bacterial cultures	Gravity flow	Enlarged for high flow rate and DNA binding capacity, removable filter for clarification and loading of lysate
Nucleospin® Plasmid	Plasmid DNA	Small bacterial cultures	Centrifugal force	-
QIAamp® MinElute® Virus Spin	Viral DNA (and RNA)	CSF and Urine	Centrifugal force	Uses protease in the lysis and carrier RNA to improve nucleic acid binding.
Nucleospin® RNAII	RNA	Mammalian cell culture	Centrifugal force	RNase inactivation and removal of DNA by rDNases.
GFX™ PCR DNA and Gel band purification	DNA	PCR solution or agarose gel	Centrifugal force	Chatatropic agent which denaturizes proteins and dissolves agarose

3.3.1 Protocol for plasmid purification with the Nucleobond®Xtra Midi kit

To purify high-copy plasmids from a medium bacterial culture (**Table 2.5**) the Nucleobond®Xtra Midi kit from Macherey-Nagel was used (42). All steps were performed at room temperature (RT) unless otherwise specified. Bacteria were grown overnight in 100 ml LB containing the appropriate antibiotics (**Table 2.6**) at 37ºC and 230 rpm. The cells were then harvested by centrifuging the overnight culture at 6500 rpm for 10 min at RT to pellet the bacteria and remove the supernatant. The pellet was resuspended in 8 ml RES buffer and the cells lysed by adding 8 ml LYS buffer and mixing by gentle inversion ~6 times before incubating for 5 min. While incubating the column and filter was prepared and equilibrated by applying 12 ml EQU buffer to the rim of the column filter allowing the buffer to empty through the filter and column by gravity flow. The previous content needed to be emptied

from the column before loading anything new. The lysate was then neutralized by adding 8 ml NEU buffer and immediately mixing by gentle inversion ~15 times before being loading the lysate onto the rim of the filter. The filter was washed to clear out any remaining lysate by adding 5 ml of EQU buffer onto the rim. When the column had emptied the filter was discarded and the column washed by adding 8 ml of WASH buffer. To elute the bound DNA 5 ml of ELU buffer was added. The eluted DNA was precipitated by adding 3.5 ml of room temperature isopropanol and mixing by pipetting before centrifugation at 20,000 g for 30 min at 4ºC. After centrifugation the supernatant was discarded and the DNA pellet washed with 2 ml of 70 % ethanol and centrifuged at 20,000 g for 5 min. The ethanol was discarded and the pellet dried until the ethanol had evaporated. The DNA was then resuspended in 100-200 µl 1xTE buffer.

3.3.2 Protocol for plasmid purification with the Nucleospin®Plasmid kit

To purify plasmids from a small bacterial culture the Nucleospin®Plasmid kit from Macherey-Nagel was used (43). All steps were performed at RT and all centrifugations were performed at 11,000 g unless otherwise mentioned. Bacteria were grown overnight in 1-5 ml LB containing the appropriate antibiotics at 37ºC and 230 rpm. The cells were then harvested by centrifuging the overnight culture for 30 s to pellet the cells. The supernatant was discarded and the cells were resuspended in 250 µl buffer A1 and mixed by vortexing until there were no visible cell clumps. To lyse the cells 250 µl of lysis buffer A2 was added and the content was mixed by gentle inversion ~8 times to avoid shearing of genomic DNA. The lysate was incubated for up to 5 min until the lysate appeared clear before being neutralized by adding 300 µl neutralization buffer A3 and mixing by inversion ~8 times. The lysate was clarified by centrifuging for 5 min. A Nucleospin® column was placed into a collection tube and a maximum of 750 µl of the clarified supernatant was loaded onto the column. The column was centrifuged for 1 min and the flow through discarded. The column was washed by adding 600 µl buffer A4 supplemented with ethanol, and centrifuged for 1 min. The flow through was discarded and the column dried by centrifuging for another 2 min. The DNA was eluted in a 1.5 ml eppendorf tube by adding ~ 50 µl buffer AE, incubating for up to 3 min and centrifuging for 1 min. The eluted plasmids concentration and purity was evaluated and were then stored at -20ºC.

3.3.3 Protocol for purification of total RNA

To isolate RNA from mammalian cell culture the Nucleospin®RNA II kit from Macherey-Nagel wad used (44). All centrifugation steps were performed at 11,000 g and all steps were performed at RT. Cells harvested from mammalian cell culture were pelleted and the supernatant removed. To the cells 350 µl buffer RA1 and DTT in a final concentration of 10 mM was added and the mixture was vortexed vigorously to lyse the cells. The cell lysate was loaded onto a NucleoSpin®Filter column placed in a collection tube and centrifuged for 1 min to clear the lysate. The filter was discarded and 350 µl of 70% ethanol was added to the flow through and mixed by pipetting. For each lysate one NucleoSpin®RNA II column was placed in a collection tube and 750 µl lysate loaded onto the column. The column was emptied by centrifugation for 30 s and placed in a new collection tube and the loading procedure was repeated if necessary. To desalt the membrane 350 µl MDB was added and the tube was centrifuged for 1 min. A DNase reaction mixture was made for each sample by adding 10 µl reconstituted rDNase to 90 µl reaction buffer for rDNase and applying 95 µl of this to the center of the column. The column was incubated at RT for 15 min. After incubation the membrane was washed by adding 200 µl buffer RA2 and centrifuged for 30 s. Subsequently the column was transferred to a new collection tube and washed a second time by adding 600 µl buffer RA3 and centrifuged for 30 s. The flow through was discarded and the column placed back in the same collection tube before adding 250 µl buffer RA3 for a third washing step and centrifuged for 2 min to completely dry the membrane. Elution was performed by adding 55 µl RNase-free H_2O and centrifuging for 1 min after 1 min incubation. The RNA is immediately placed on ice or at -70ºC.

3.3.4 Protocol for viral DNA purification

To purify viral DNA from CSF and urine samples the QIAamp® MinElute® Virus Spin kit from Qiagen was used (45). All centrifugation steps were performed at RT and 8,000 rpm unless otherwise specified. The concentration of ethanol used was 96-100 %. Before starting the heating block was set to 56ºC; the buffer AVE and samples equilibrated to RT, the sample added the appropriate amount of 0.9 % sodium chloride until a total volume of 200 µl was reached; and the buffers, carrier RNA and protease solution was prepared with an addition of a one-time only amount of specified solution. Twenty-five ml and 30 ml of ethanol were added to the buffer AW1 and AW2, respectively. The QIAGEN protease was dissolved in 1.4 ml of buffer AVE. The carrier RNA was dissolved by adding 310 µl buffer AVE and stored at

-20ºC until use. The AL buffer had to be prepared freshly for each batch of samples by adding carrier RNA to the buffer. The volume of Buffer AL and carrier RNA needed for the purification was calculated as followed:

N x 0.22 ml = Y ml

Y ml x 28 µl/ml = Z µl

- N = number of samples
- Y = calculated volume of AL buffer
- Z = volume of carrier RNA/AVE buffer to be added to the AL buffer

Twenty-five µl of QIAGEN Protease was prepared in each 1.5 ml eppendorf tube and 200 µl of sample or sample/0.9% sodium chloride mix and 200 µl of AL buffer containing carrier RNA was added. The content was then mixed by pulse-vortexing for 15 sec or until the solution was homogenous, and then incubated at 56 ºC for 15 min. After incubation the tube was quickly spun and 250 µl of ethanol was added. The content was mixed thoroughly by pulse-vortexing for 15 s and incubated for 5 min. The tube was quickly spun after incubation and the lysate loaded onto the QIAamp MinElute column. The column was centrifuged for 1 min and then moved into a clean 2 ml collection tube while the flow through was discarded. Subsequently, 500 µl of AW1 buffer was added to the column and the column centrifuged for 1 min. The flow through was discarded, the column placed into a clean collection tube, and 500 µl of AW2 buffer was added to the column which was then centrifuge for 1 min. Again the flow through was discarded and the column placed into a clean collection tube. Five hundred µl of ethanol was added to the column and centrifuged for 1 min. The column was placed into a clean collection tube and centrifuged at 14,000 rpm for 3 min before placing the column into a new collection tube and incubating with the lid open at 56ºC for 3 min to dry the membrane completely. Finally, the column was moved into a 1.5 ml eppendorf tube and 105 µl dH_2O was applied to the membrane. The column was incubated for 5 min before centrifuging at 14,000 rpm for 1 min to elute the DNA. The concentration of the eluate is measured before storage at -20 ºC.

3.3.5 Illustra GFX^{TM} PCR DNA and Gel band purification

To purify DNA from solutions or gels the illustra GFX^{TM} PCR DNA and Gel band purification kit from GE Healthcare was used (46).

3.3.5.1 Protocol for purification of DNA from solution or enzymatic reactions

All centrifugation steps were performed at 16,000 g. The binding solution was prepared by adding 500 µl Capture buffer type 3 to ≤ 100 µl of the sample followed by a thorough mix. The GFX MicroSpin column was placed into a collection tube and the binding solution was loaded. A centrifugation of 30 s was performed and the flow through discarded before placing the column back in the same collection tube. Five hundred µl of Wash buffer type 1 was added and the column was centrifuged for 30 sec. The flow through was discarded before placing the column back in the same collection tube followed by a 1 min centrifugation to dry the membrane. The collection tube was then discarded and the column placed in a clean 1.5 ml eppendorf tube. Finally, 10-50 µl of elution buffer type 4 was added to the membrane which was then incubated at RT for 1 min before eluting by centrifugation for 1 min.

3.3.5.2 Protocol for purification of DNA from TAE agarose gels

All centrifugation steps were performed at 16,000 g. After agarose gel electrophoresis the band of interest was excised from the gel under UV light and placed in a pre-weighed eppendorf tube. The tube was weighed and the amount of gel was used to calculate the amount of Capture buffer type 3 which was to be added. Ten µl Capture buffer was added for each 10 mg of gel, with a minimum of 300 µl regardless the amount of gel. The tube was incubated at 60ºC for 15-30 min until the gel had dissolved. A GFX MicroSpin column was placed into a collection tube and binding solution added. The tube was left to incubate at RT for 1 min before centrifuging for 30 sec. The flow through was discarded and the column placed back in the same collection tube. Then 500 µl of Wash buffer type 1 was added and the column centrifuged for 30 sec. The flow through was discarded before placing the column back in the same collection tube and centrifuged for 1 min to dry the membrane. Finally the column was placed in a clean 1.5 ml eppendorf tube and 10-50 µl of elution buffer type 4 was added to the membrane and incubated for 1 min at RT before eluting the DNA by centrifugation for 1 min.

3.4 Evaluation of Nucleic Acids

In this study two general methods were used to evaluate the concentration and purity of the nucleic acids; UV-spectrophotometry and Agarose gel electrophoresis. In this study the method of UV spectrophotometry was performed with NanoDrop-1000 Spectrophotometer.

3.4.1 UV-spectrophotometry

The aromatic rings in the nucleic acid structure absorb UV light of 230nm to 320nm, and have a mean absorbance peak at 260 nm. The absorbance measured is proportionally correlated to the concentration as described by Beer-Lambert Law. An OD_{260} of 1 corresponds to ~50 ng/µl for double-stranded DNA and ~40 ng/µl for RNA. The most common contaminants like salts and proteins that can influence downstream analysis are also absorbing light in this range. By evaluating the ratio of absorbed light between different wavelengths the purity of the samples can be assessed. If there are significant contamination the quantitation will not be accurate (47).

At 230 nm Guanidium salts which are used to help the DNA bind to silica membrane absorb light strongly, and at 280 nm aromatic amino acids absorb light. A ratio A_{260}/A_{230} between 1.8 and 2.2 and a ratio A_{260}/A_{280} of ~1.8 for DNA and ~2.0 for RNA is generally accepted as pure. A ratio much lower may indicate presence of co-purified contaminants (48, 49). In this thesis ND-1000 was used to measure the nucleic acid concentration.

3.4.2 Gel electrophoresis

Agarose gel electrophoresis is a basic and simple method of visualizing DNA, that allow evaluation of the size and conformation as well as separation and extraction of different DNA fragments. The agarose gel is made by solidifying a boiled liquid containing buffer and agarose. An agarose gel contains a network of polymeric molecules with an average pore size depending on the type and percent of agarose and buffer. The liquid is poured into a suitable retainer with a comb to form wells and allowed to harden. The gel is then placed in an electrophoresis chamber and covered by buffer. DNA is loaded into the wells along with a ladder for size determination and a constant voltage is run through the gel. The negatively charged DNA will migrate from a negatively charged electrode towards a positively charged electrode (50).

The mass and shape of the DNA will affect the time it takes to migrate through the gel. In general smaller molecules will travel faster, and the more coiled a DNA fragment is the faster it will travel through the gel. A plasmid which is super coiled will travel faster than the same plasmid which has an open circular structure. If the same plasmid is made linear it will travel faster than the open circular but slower than the super coiled plasmid (51). To visualize the DNA the gel was prestained with a fluorescent dye and illuminated under UV light the gel was taken a picture of in the Bio-Rad Gel Doc 2000 using the Quantity One software. In this

thesis agarose gel electrophoresis was used after purification of nucleic acids as well as restriction cutting, ligation and PCR, and the agarose gel electrophoresis conditions are described in **Table 3.2**.

Table 3.2: Agarose gel electrophoresis conditions

Reagents and conditions	Small PCR products	Large PCR products or plasmids
Dye	10 µl	10 µl
Agarose	1.4 %	1 %
1xTAE buffer	100 ml	100 ml
Voltage	100-110	90-100
Time	~30 min	40-60 min

3.5 PCR

Polymerase chain reaction is a rapid mean for amplification, identification and analysis of DNA. The PCR amplifies a fragment of a template DNA with a pair of oligonucleotide primers complementary to the flanking regions of the fragment template, a DNA polymerase and free nucleotides in a buffer solution. The basic principle of PCR is divided into three steps: denaturation, annealing and extension. In the first step the dsDNA is denatured into single strands by exposure to temperature above 90ºC. In the annealing step the temperature is lowered until a temperature optimal for the specific primers to hybridize to the DNA template. The last step of extension involves raising the temperature and elongation of the primers from 5' to 3' direction by the DNA polymerase. After one round of denaturation, annealing and extension the number of DNA fragments will be doubled and each DNA molecule will act as template in the repetition of the cycle. A series of cycles will thus provide a rapid exponential increase in copies of the specific target fragment of DNA (50, 52).

In this thesis PCR was used for screening of CSF and urine samples for KIPyV DNA by primers directed against the VP1 and NCCR region, cloning of the VP1 sequence, test medium of transfected cells for the presence of KIPyV DNA and to monitor the presence of viral RNA in cells transfected with KIPyV DNA.

3.5.1 Standard PCR

In this thesis a ready mix including a thermo stable DNA polymerase (Taq-DNA polymerase), nucleotides and a buffer solution optimal for the PCR was used.

3.5.1.1 PCR Protocol

The reaction mix and all reagents were kept on ice until the PCR tubes were ready to put in the thermal cycler. The reaction mix was made as described in **Table 3.3** containing all but the template, or water and template if the amount of sample input is variable. The total volume and amount of water and template DNA varies according to the purpose of the PCR. The reaction mix was distributed in PCR tubes before adding the samples. The PCR tubes were placed in a PCR machine and incubated in the programs described in **Table 3.4** or **Table 3.5**.

Table 3.3: Reaction mix for PCR. * PCR with NCCR, KI VP1, APRT, LT-ag and CREB primers were run with total volume of 25 µl and SalI/SacII with 35 µl. **Standard template volume added was 2 µl or 100 ng/µl, exceptions are medium where 5 µl is added, PCR product 1 µl, and in the case of cloning PCR 10 µl KIPyV genome was added.

Reagents	Amount	Amount
JumpStart™ Taq ReadyMix™	15 µl	15 µl
Forward primer	1.5 µl of [100 ng/µl]	1 µl of [10 µM]
Reverse primer	1.5 µl of [100 ng/µl]	1 µl of [10 µM]
dH_2O	Up to the total volume *	Up to a total volume of 25 µl
Template DNA	Up to10 µl **	1-2 µl**

Table 3.4: Thermal cycler program used for cloning PCR

Number of cycles	Temperature (°C)	Time
1	94	5 min
30	94	30 sec
	55	30 sec
	72	60 sec
1	4	∞

Table 3.5: Thermal cycler program used for primers KI VP1 (40 cycles in PCR1 and 35 cycles in nested PCR2), NCCR (25 cycles in sensitivity test and otherwise 40 cycles), APRT, LT-ag and CREB (40 cycles).

Number of cycles	Temperature (°C)	Time
1	94	5 min
25-40	94	30 sec
	55	30 sec
	72	30 sec
1	72	7 min
	4	∞

3.5.2 Nested PCR

Nested PCR is used to increase the sensitivity and specificity of a primary PCR product. The primary PCR product is used as template for a second primer pair placed internal to the first primer pair. The second PCR is generally performed with fewer cycles and only 1 µl of the first PCR product as template but otherwise conditions are the same as the first PCR.

3.5.3 RT-PCR

To study the gene expression in cells transfected with KIPyV DNA, mRNA was extracted and converted into cDNA. This reverse transcription reaction was performed with the Bio-Rad iScript cDNA synthesis kit, which contains a blend of oligo(dT) and random short primers. The cDNA was then used as template in a PCR with specific primers directed against the large T-antigen sequence of KIPyV. The method is therefore known as reverse transcription PCR or RT-PCR.

3.5.3.1 Reverse transcription protocol

The reaction mix and all reagents were kept on ice while working. The reverse transcription reaction mix was made as described in **Table 3.6** and run as described in **Table 3.7**. The amount of cDNA used in downstream PCR was 2 µl.

Table 3.6: Reaction mix for reverse transcription of RNA

Reagents	Amount in KIPyV tube	Amount in - ctr tube
5 x iScript reaction mix	4 µl	4 µl
iScript reverse transcriptase	1 µl	1 µl
RNA template (1µg)	4 µl	13 µl
Nuclease free H_2O	11 µl	2 µl
Total	20 µl	20 µl

Table 3.7: Thermal cycler program for reverse transcription of RNA

Number of cycles	Temperature (°C)	Time
1	25	5 min
	42	30 min
	85	5 min
1	4	∞

3.6 ExoSAP

PCR products are purified before being used as template in downstream analysis such as sequencing or nested PCR to avoid interference from excess primers and nucleotides. ExoSAP-IT is a reagent that contains *Exonuclease* I and *Shrimp Alkaline Phosphatase*, two hydrolytic enzymes that together remove unincorporated nucleotides and excess primers from a PCR product (47, 53). Treatment with ExoSAP-IT is an alternative to purifying the PCR product with the illustra GFX^{TM} PCR DNA and Gel band purification kit from GE Healthcare. In this thesis ExoSAP-IT treatment is used on PCR product before sequencing.

3.6.1 ExoSAP protocol

One µl ExoSAP-IT was added to each PCR tube containing 3-15 µl of product and 1 extra µl of water was added if the total volume was less than 5 µl. The treatment was performed in a thermal cycler with the program described in **Table 3.8**.

Table 3.8: ExoSAP-IT treatment incubation program in Thermal cycler

Number of cycles	Temperature (ºC)	Time
1	37	1 hr
	85	15 min
1	4	∞

3.7 DNA sequencing by capillary electrophoresis

In this thesis sequencing has been used to verify the structure of recombinant plasmids and to confirm the products obtained after PCR of the VP1 region of KI in the biological and cultivated samples. The principle of the sequencing technique used in this thesis is built upon the inclusion of dideoxynucleotide (ddNTP) when complementary strands of the template are created. A ddNTP lack the 3’ hydroxyl group which is present in deoxynucleotides (dNTP) which enables the linkage of nucleotides. The ddNTPs are labeled with fluorochromes, one type for each of the nucleotides. When the ddNTP is randomly included the synthesis of the new strand ceases and the strand is made detectable by laser light. The randomly inclusion of ddNTP will create strands of all different lengths. The synthesized strands will travel through small capillaries containing a liquid polymer and the fluorochromes are excited by a laser making the different fluorochromes emit light of different wavelengths. The travel time and type of emitted light is registered and interpreted by a computer, producing a readable sequence of nucleotides (51, 52).

3.7.1 Sequencing protocol

The sequencing reaction is made and performed with the reagents and conditions described in **Table 3.9** and **Table 3.10**. The BigDye3.1 reagent contains a ready mix of DNA-polymerase, dNTPs and fluorescence-labeled ddNTPs. The analysis of the sequence products were performed by the sequencing facility at UNN using a 3130*xl* Genetic Analyzer from Applied Biosystems/Hitachi.

Table 3.9: Sequencing reaction mix

Reagent	Amount
Primer 3.2 pmol/µl	1 µl
5x Seq buffer	3 µl
BigDye3.1	0.5 µl
dH_2O	Up to a total volume of 20 µl
Template DNA	~250-500 ng

Table 3.10: Sequencing program in Thermal cycler

Number of cycles	Temperature (°C)	Time
30	96	10 sec
	50	5 sec
	60	4 min
1	4	∞

3.8 Cloning in plasmid vectors

Cloning of DNA fragments in plasmid vectors involves restriction cutting and ligation to produce recombinant plasmids capable of replication, and introduction of the recombinant plasmid into a host such as *E. coli* for amplification. Restriction endonucleases typically used in genetic engineering recognize short palindromic DNA sequences and cleave the double strand at specific sites within these recognition sequences (52). Plasmid vectors contain multiple cloning sites with several single recognition sites for different restriction enzymes. If it is not possible to insert a fragment directly into a vector due to lack of common restriction sites or inappropriate reading frames, other vectors can be used as intermediates or restriction sites can be introduced by performing PCR with partially complementary primers containing the restriction site. After restriction cleavage the fragments of both donor and recipient DNA are separated by gel electrophoresis and selected fragments are cut out and purified. The plasmid vector and the fragment that is to be cloned are joined together by DNA-ligase. The newly ligated plasmid is transformed into competent bacteria, allowing the plasmid to replicate and amplify its copy number. The vector contains genes for antibiotic resistance for

selection of transformed colonies. In this thesis both PCR and restriction cleavage have been used to select for the plasmids with an inserted VP1 gene. After selection the recombinant plasmid is purified and sequenced.

3.8.1 Restriction cutting protocol

The restriction reaction was made out of one or two restriction enzymes, a suitable buffer, water and DNA (see **Table 3.11** and **Table 3.12**). A control reaction using only one restriction enzyme was made to ensure that both enzymes were in fact working (**Table 3.12**). The restriction reaction was incubated overnight at 37ºC and the digested fragments were analyzed on a 1% agarose gel. The fragments of interest were cut out from gel and purified with the illustra GFX^{TM} PCR DNA and Gel band purification kit from GE Healthcare.

Table 3.11: Restriction mix for cutting of KI and BKV out from pUC18 and pUC19 vector

Reagents	Amount	Amount
Nde I buffer (10x)	10 % of total volume	-
Nde I (10 U/µl)	16 µl	-
10x buffer	-	10 % of total volume
*BamH*I (20 U/µl)	-	2 µl
pUC18.KI (1.45 µg/µl)	~100 µg	-
pUC18.BKV (1.29 µg/µl)	-	~40 µg
dH_2O	Until a total volume of 200 µl	Until a total volume of 200 µl

Table 3.12: Restriction mix for cutting of VP1 fragment from donor vector and preparation of acceptor vector.

Reagents	Amount in the Donor reaction	Amount in the Vector reaction	Amount in the Control reaction
Buffer	10 % of total volume	10 % of total volume	10 % of total volume
Restriction enzyme I	0.5 µl	5 µl	0.5 µl of one or the other
Restriction enzyme II	0.5 µl	5 µl	
Donor plasmid	0.5- 2 µg	-	-
Recipient vector	-	0.5- 42 µg	0.5 µg
H_2O	Until a total volume of 20 or 50 µl	Until a total volume of 20 or 200 µl	Until a total volume of 20 or 50 µl

3.8.1.1 CIP treatment

The CIP reagent from Finnzymes contains *Alkaline phosphatase* an enzyme that catalyses the removal of 5'phosphate residues from nucleic acids (47). In this thesis CIP is used to remove 5' phosphates from a linearized plasmid vector to prevent self-ligation.

3.8.1.1.1 CIP protocol

The reaction mix was prepared as described in **Table 3.13** and incubated at 37ºC for 1.5 hrs. The CIP treated *BamH*I and *Not*I cut pcDNA3.1+ was then inactivated by incubation at 75ºC for 10 min.

Table 3.13: CIP reaction mix

Reagents	Amount
H_2O	7 µl
10 x CIP buffer	2 µl
CIP (10 U/µl)	1 µl
Purified pcDNA 3.1 (+) cut with *BamH*I and *Not*I	10 µl
Total	20 µl

3.8.2 Ligation protocol

The ligation mix was made combining T4-Ligase, T4 ligase buffer, water and DNA as described in **Table 3.14** or **Table 3.15,** depending on the kind of ligation desired. A negative ligation mix with water instead of DNA was always included. The ligation mixture was incubated over night at ~16 ºC, and inactivated for 10 min at 70 ºC. To ensure re-ligation of the plasmid or genome the ligation mix was analysed on a 1 % agarose gel.

Table 3.14: Ligation mix for KIPyV and BKPyV genome cut out from vectors, made to a final concentration of 50 ng/µl

Reagents	Amount
T_4 ligation buffer	14.4 µl
T_4 ligase	14.4 µl
Purified genome	~7.5 µg
dH_2O	Until a total volume of 144 µl

Table 3.15: Ligation mix for ligation of VP1 fragments into vectors

Reagents	Amount in ligation mix	Amount in negative ligation mix
T_4 ligation buffer	2 µl	2 µl
T_4 ligase	1 µl	1 µl
Purified KI VP1 cut with restriction enzymes	10 µl	-
Vector cut with restriction enzymes	1 µl	1 µl
dH_2O	Until a total volume of 20 µl	Until a total volume of 20 µl

3.8.3 Transformation of competent bacteria

To select plasmids with a successful recombination and/or to amplify the amount plasmid, the plasmids are used to transform competent *Escherichia coli*. Competent bacteria have been induced to take up DNA from their environment artificially by treatment with $CaCl_2$ or by electroporation (52). In this thesis two different strains of competent *E. coli* were used; DH5α and BL21. The BL21 strain was used when expression of the GST-fusion protein was desired because this strain is protease deficient. After colony selection PCR and restriction cutting was used to identify plasmids with the VP1 insert. In addition, removal of 5´ phosphates from the cut vector by treatment with alkaline phosphatase was occasionally done to prevent self-ligation.

By exposing the competent cells to a heat shock they will take up circular DNA from their surroundings (52). The plasmid vectors usually contain a gene of resistance to certain antibiotics which can be used for selection of transformed bacteria. After the heat shock the cells are incubated with nutrient medium to allow expression of the resistance genes before plating out on medium containing the selective antibiotic. Colonies growing on this media are chosen for further inoculation, purification and identification. It is not certain that colonies growing on the media contain the recombinant plasmid, and further testing is needed to select colonies with the recombinant plasmid from the ones with empty vectors. If the plasmid is one of the recombinated, sequencing will be performed to conclude if the plasmid contains the desired insert and if necessary, if it is in the correct reading frame.

3.8.3.1 Transformation protocol

To transform competent *E. coli* cells purified plasmid or a ligation mix was added to 200 µl competent cells in a falcon tube kept on ice and incubated for 30 min. The cells were then heat shocked for 90 sec at 42 ºC before adding 800 µl SOC medium and incubated for 45 min at 37 ºC with shaking. After incubation 300 µl of the transformation mix was plated out on a LB plate containing the appropriate antibiotic (**Table 2.6**) and incubated over night at 37 ºC.

3.8.3.1.1 Protocol for selection of colonies by PCR

The primers used are specific for the VP1 fragment of KIPyV and used with a concentration of 100 ng/µl. The reaction mix was prepared according to the standard PCR protocol **3.5.1.1** and **Table 3.3** with a total volume of 25 µl disregarding the template volume and distributed into the PCR tubes. The PCR was performed with the program described in **Table 3.4.**

Colonies were selected from a LB plate after transformation and a part of the colony was transferred to the PCR tube while the rest was inoculated in 3 ml LB with appropriate antibiotic (**Table 2.6**) overnight with shaking at 37 ºC. The PCR product was run on a 1 % agarose gel and colonies with PCR products of ~1.2 kb were selected for plasmid purification.

3.8.3.1.2 Protocol for selection of colonies by restriction cutting

Colonies were selected from a LB plate after transformation and inoculated in 3 ml LB with appropriate antibiotic (**Table 2.6**) over night with shaking at 37 ºC for subsequent plasmid purification with the Nucleospin® Plasmid kit from Macherey-Nagel. After purification 1 µg of the plasmids were cut with appropriate restriction enzymes, according to the restriction cutting protocol for donor plasmids described in **Table 3.12**. The restriction reaction was run on a 1 % agarose gel and plasmids producing bands of ~1.2 kb were selected. The next step can either be isolation of the 1.2 kb fragment with the illustra GFX™ PCR DNA and Gel band purification kit from GE Healthcare for further cloning or sequencing of the ready to use plasmid.

3.8.4 Cloning protocol

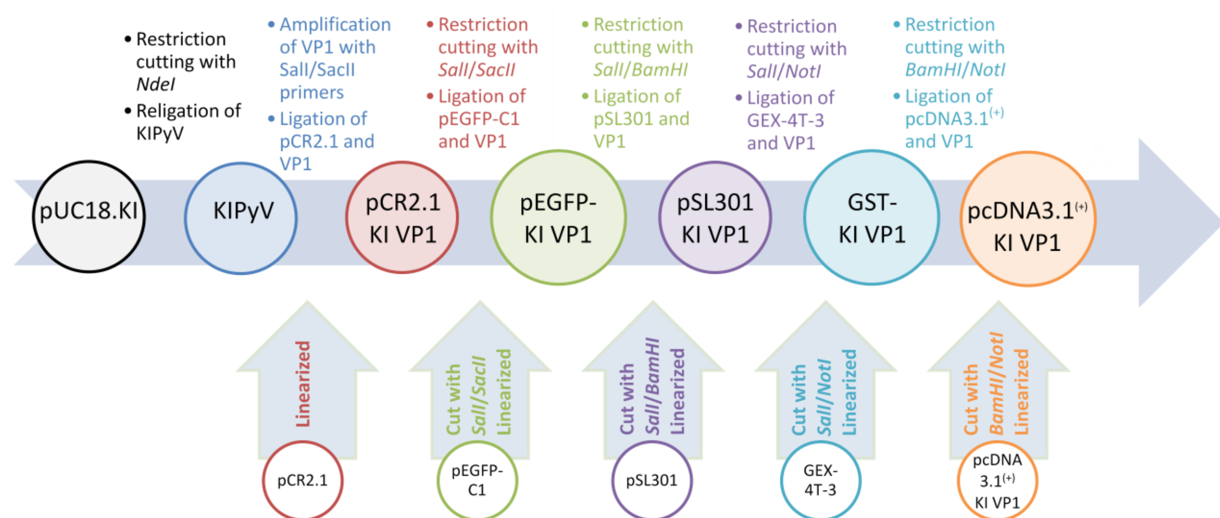


Figure 3.1: Flowchart of the cloning protocol with all the plasmids used in this thesis. The KIPyV genome is used to amplify the VP1 gene which is inserted into various plasmid vectors. The large plasmids are those used as controls in this thesis. The smallest plasmids are vectors in which the KIPyV VP1 gene was inserted, and the intermediate sized were either not used any further or in the case of KIPyV genome used in transfections and as a control.

In this thesis single gene cloning of the KIPyV gene VP1 was performed to produce controls for the transfection of mammalian cells with the KI genome, to study the sub-cellular

localization of the VP1 protein and to produce VP1 protein in fusion with GST and pcDNA3.1$^{(+)}$. The cloning procedure is briefly presented in **Figure 3.1** and in depth in the following sections.

3.8.4.1.1 pUC18.KI DNA

The pUC18.KI plasmid contains the entire KI genome cloned in the *Nde*I site of pUC18. This cloning interrupts the VP1 region. To make the sequence whole again the KI genome was cut out with the help of restriction enzyme *Nde* I and religated after purification from gel. The protocol for restriction cutting is described in section **3.8.1** and **Table 3.11** and the protocol for purification from gel is described in **3.3.5.2** and **Table 3.2**. The ligation mix was made to produce a concentration of 50 ng/µl KI genome as described in section **3.8.2** and **Table 3.14**. Both the cut and religated KI genome was used in downstream analysis.

3.8.4.1.2 Cloning of KI VP1 into pCR2.1 vector

There were no suitable restriction sites for cutting the VP1 gene out of the KI genome, which is why a PCR using primers complementary to the flanking sequences of the VP1 gene with additional non-complementary nucleotides containing known restriction sites in the 5' end was performed. The PCR product contained the whole VP1 gene with restriction sites for *SalI* before the start codon and *SacII* after the stop codon. Ten µl of ligated KI genome was used as template in a PCR described in **Table 3.3** and **Table 3.4** with the primers VP1 SalI and VP1 SacII (100 ng/µl), and a total volume of 35 µl. The PCR product was analyzed on a 1 % agarose gel, cut out and purified as described in section **3.3.5.2** and **Table 3.2**.

The KI VP1 PCR fragment was ligated into the pCR2.1 vector by making the reaction mix as described in **Table 3.16** and incubating for 5 min at RT. The ligation mix was transformed in DH5 cells as described in section **3.8.3.1**. The next day PCR of a selection of colonies was performed to identify the colonies containing the recombinant pCR2.1 KI VP1 plasmid as described in section **3.8.3.1.1** which were further inoculated in LB medium for plasmid purification as described in section **3.3.2**.

Table 3.16: Ligation reaction mix with the pCR2.1 vector

Reagent	Ligation mix	Negative ligation mix
H_2O	Up to a total volume of 6 µl	4 µl
Salt solution	1 µl	1 µl
Purified *Sal*I/*Sac*II PCR fragment	1-4 µl	-
pCR2.1 vector	1 µl	1 µl

3.8.4.1.3 Cloning of KI VP1 from pCR2.1 vector into the EGFP-C1 vector

The VP1 insert was cut out after amplification and purification of the pCR2.1 vector. The VP1 gene was then inserted into the pEGFP-C1 vector by cutting both insert and vector with two different restriction enzymes to ensure correct directionality of the gene upon ligation (directional cloning). Plasmids pCR2.1 KI VP1 (0.6 µg) and pEGFP-C1 (42 µg) were cut with *Sal*I and *Sac*II as described in section **3.8.1** and **Table 3.12** with a total volume of 50 and 200 µl for the insert and vector respectively. The next day the restriction mix was run on a 1 % agarose gel to see if the cutting was complete and to purify the VP1 fragment. The restriction mix containing cut pEGFP-C1 was inactivated at 65ºC for 20 min before being added to the ligation mix. The ligation mix was made as described in section **3.8.2** and **Table 3.15** with 10 µl of the insert and 1 µl of the vector. As previously described the ligation mix was transformed in DH5 α cells.

The following day colonies were selected by PCR to identify colonies containing the recombinant EGFP-KI VP1 plasmid. The colonies were inoculated and the EGFP-KI VP1 plasmid purified and sequenced. This plasmid was used as a control for transfection efficiency as it allows the visualization of cells that have been transfected by emitting green fluorescence light produced by the EGFP fusion protein, and as a control of the antibodies against KI VP1 in WB. EGFP-KI VP1 was also used when the VP1 gene was inserted into a GST vector for expression in *E. coli*. Moreover, the sub-cellular localization of the EGFP-VP1 fusion protein was studied by confocal microscopy.

3.8.4.1.4 Cloning of KI VP1 from EGFP vector into the GEX-4T-3 vector

By cloning the VP1 gene into a GEX vector the protein can be expressed as a fusion protein with the enzyme glutathione S-transferase (GST) as a tag which facilitates purification. The cloning of a GST-KI VP1 recombinant started with cutting out the VP1 gene from the EGFP-KI VP1 plasmid and inserting it into the pSL301 vector by directional cloning. Two µg of the EGFP-KI VP1 plasmid and 1 µg of an intermediate vector pSL301 were cut with restriction

enzymes *Sal*I and *Bam*HI as described in **Table 3.12** with a total reaction volume of 20 µl. The next day the restriction mix was run on a 1 % agarose gel to see if the cutting was complete and to cut out and purify the VP1 fragment and the linear pSL301 as previously described. The ligation mix was made as described in **Table 3.15** with 10 µl of the insert and 1 µl of the vector, and the incubated ligation mix was transformed in DH5 α cells for amplification.

The day after colonies were selected by restriction cutting as described in section **3.8.3.1.2**. One µg of the purified plasmids and 0.5 µg of the GEX-4T-3 vector were cut with restriction enzymes *Sal*I and *Not*I in a total reaction volume of 20 µl as described in **Table 3.12**. After cutting the restriction mix was run on a 1 % agarose gel and the VP1 fragments and linear pGEX-4T-3 were purified. Three ligation mixes were made using 4 µl of the different extracted KI VP1 and 1 µl of the purified linear vector as described in **Table 3.15** and incubated over night at RT. The different ligation mixes were transformed in DH5 α cells and colonies were selected by restriction cutting. The purified plasmids were cut with *Sal*I and *Not*I as described in **Table 3.12** using 0.5 µg of plasmid. The selected colonies were sequenced and used to transform BL21 cells able to express VP1 protein. After purification the protein was used as a control for the KI VP1 antibodies during WB. The GST-KI VP1 plasmid was also used when the VP1 gene was inserted into pcDNA3.1$^{(+)}$ vector for expression in mammalian cells.

3.8.4.1.5 Cloning of KI VP1 from GST vector into the pcDNA3.1$^{(+)}$ vector

The VP1 sequence was cut out from the GST-KI VP1 plasmid and inserted into the pcDNA3.1$^{(+)}$ vector by directional cloning. One µg GST-KI VP1 plasmid and 0.5 µg pcDNA3.1$^{(+)}$ vector were cut with restriction enzymes *Bam*HI and *Not*I in a total reaction volume of 20 µl as described in **Table 3.12**. The following day the restriction mix was run on a 1 % agarose gel to observe if the cutting was complete, to cut out and to purify the VP1 fragment and the linear pcDNA3.1$^{(+)}$. After purification the linear pcDNA3.1$^{(+)}$ was treated with CIP as described in **3.8.1.1.1**. The ligation mix was made as previously described and **Table 3.15** using 10 µl purified VP1 fragment and 2 µl of purified vector in a total volume of 20 µl. The ligation mix was transformed in DH5 α cells. Positive colonies were selected the following day and PCR with the primers KI VP139F (100 ng/ µl) and KI VP1 363R (100 ng/ µl) as described in section **3.8.3.1.1** was performed to identify colonies containing the

recombinant pcDNA3.1$^{(+)}$/KI VP1 plasmid. The pcDNA3.1$^{(+)}$/KI VP1 plasmid was purified and sequenced before being used for expression of VP1 in mammalian cells.

3.9 Mammalian cell culture techniques

Mammalian cell cultures are widely used in the field of molecular biology as experimental models for *in vivo* situations. In this thesis the two cell lines A549 and HEK 293 were used to investigate whether they are permissive to KIPyV. Media and cells from transfected cells were harvested and RNA, DNA or proteins were used for PCR or Western Blotting, respectively (**Figure 3.2**). In addition A549 and Vero cells were used to study the sub-cellular localization of KIPyV VP1 using an EGFP-VP1 fusion protein by confocal microscopy. The cell lines used in this thesis grow adherently in monolayers. All media, PBS, trypsin, and serum were preheated to 37 ºC before use. Cells were kept in a humidified CO_2 incubator at 37ºC.

A549 cells are human alveolar basal epithelial cells derived from an adenocarcinoma first cultivated in 1972 by D.J Giard et al. HEK293 is a cell line derived from human embryonic kidney cells transformed by adenovirus. The HEK cells were established in the early 1970s at the University of Leiden, Holland. The Vero cell line derives from the kidney of a healthy African green monkey and was developed in 1962 by Y. Yasumura and Y. Kawakita at the Chiba University in Japan.

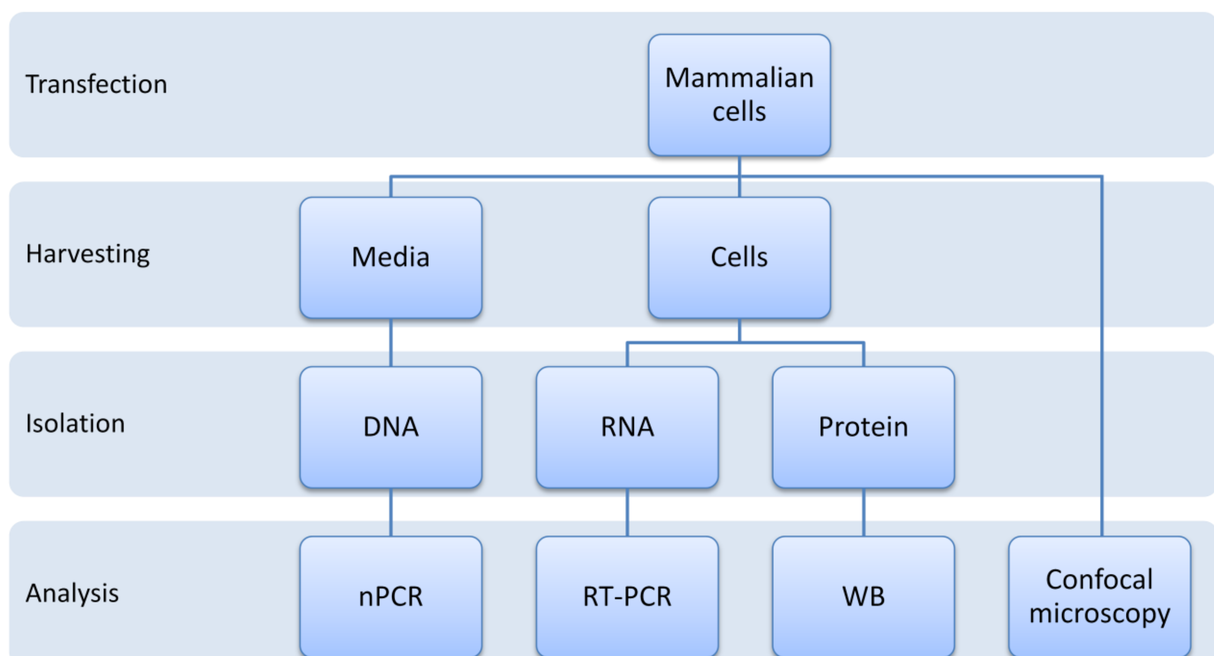


Figure 3.2: Flowchart of the analysis processes all starting with transfection of mammalian cells.

3.9.1 Sub-culturing of cells

Adherent cells that grow in a monolayer will after some time cover the whole surface of the culture vessel and become confluent. In addition their resources will become scarce and need replenishing. To keep the cells healthy and viable they should be sub-cultured or "split" just before they become confluent. In this thesis the cells have been split by enzymatic treatment with Trypsin and shaking the flask to detach them from the surface of the culture flask. The cells were then resuspended and a portion of the cells was transferred into a new flask with fresh growth media (52).

3.9.1.1 Protocol for sub-culturing

The old growth media was aspirated and the cells were washed with 10 ml 1xPBS. One ml of Trypsin was then added and the cells were incubated at 37ºC until the cells detached, between 30 s to 5 minutes depending on the cell line. To aid in the detachment the flask was given a few taps. The cells were then resuspended in 9 ml fresh growth media. One ml of the resuspension was then transferred to a new flask containing fresh growth media and FBS. The amounts of the reagents used in the new flasks are described in **Table 3.17**. The split ratio of 1:10 is commonly used in this thesis, but if required the ratio was changed to 1:5.

Table 3.17: Volumes of media used for splitting of cells in medium and large culture flask

Reagents	Medium culture flask	Large culture flask
FBS	10 or 15 %	10 or 15 %
Resuspension of cells	1 ml	1-2 ml
Fresh growth media	Up to a total volume of 25 ml	Up to a total volume of 40 ml

3.9.2 Seeding out cells for transient transfection

In this thesis cells were seeded out in wells for transfection procedures. The amount of cells in each well varied through the experiments. In this study the number of cells per well varied between 100,000 and 250,000 for seedings in 6 or 12 well trays for analysis by PCR or WB, and between 10,000 and 20,000 for seedings in chamber slides for confocal studies.

3.9.2.1 Protocol for seeding out cells

A drop of the resuspension of trypsinized cells and media, made according to the subculture protocol, was applied onto a Bürker counting chamber and cells counted manually under a light microscope. The average number of cells counted in 5 grids was multiplied by 10^4 to

find the number of cells per ml. The suspension was diluted to lower the density of cells used in the counting chamber if needed, and the number of cells per ml is multiplied by the dilution factor to get the correct concentration of cells. To calculate the amount of cell suspension needed per well the amount of cells wanted in each well was divided by the concentration of cells in the suspension. The amount of cell suspension was diluted in fresh growth media according to the size of the well. In a 12 well plate the total volume for one well is 1 ml, for a 6 well plate the total volume for one well is 2 ml and in a 8 well chamber slide the total volume for one chamber is ~125 µl.

3.9.3 Harvesting cells or media

After transient transfection cells and/or media are harvested to analyze the result of the transfection. The medium needed to be changed 3-6 hours after transfection and the cells were washed twice with 1xPBS to remove any DNA not taken up by the cells before overnight incubation, to be able to harvest the media and use it directly as template in PCR. The cells were harvested for protein analysis or RNA isolation.

3.9.3.1 Protocol for harvesting media

A certain time (24 hours to 4 weeks) after transfection 200 µl of media was transferred to eppendorf tubes. The media was centrifuged at 11,000 rpm for 2 min to remove floating cells, and frozen at -20ºC until used in PCR. Five µl media was used in the PCR mix.

3.9.3.2 Protocol for harvesting cells for protein analysis

The media from the wells was removed and the cells were washed carefully with 1xPBS twice. To harvest the cells the lysis buffers was made as described in **Table 3.18**. Eighty µl of buffer was added to each well and the cell lysate was transferred to eppendorf tubes. The cell lysate was sonicated on ice three times for 3 sec and denaturized at 70ºC for 10 min. The lysate was stored at -20ºC until used in WB.

Table 3.18: Sample buffer content

Reagents	Amount for one well
4x LDS buffer	40 µl
1 M DTT	32 µl
H_2O	8 µl
Total volume	80 µl

3.9.3.3 Procedure for harvesting cells for RNA analysis

After transfection the cells in the KI transfected wells and the negative control were transferred to medium flasks. When the growth was confluent the cells were harvested by adding 1 ml Trypsin and incubating at 37ºC until the cells had detached. The cells were then resuspended in ~3 ml fresh growth media and divided in eppendorf tube. The cells were centrifuged at 11,000 rpm for 2 min to create a cell pellet and the supernatant was removed. The cells were immediately used for RNA purification.

3.10 Transfection of mammalian cells

To study whether a cell line was permissive for KIPyV, the viral genome was transferred into the cells by transfection. Two different liposome-mediated transfections have been used in this study. A liposome-mediated transfection involves uptake of foreign DNA by endocytosis, based on the ionic interaction between the DNA and liposomes and the interaction between the cationic liposomes and the negatively charged cell membrane (50).

3.10.1 Lipofectamine™2000 protocol

The cells are seeded out in a concentration optimal for the experiment as described in the seeding out protocol. The following day Lipofectamine™2000 (**Table 2.10**) was mixed with OptiMem and incubated for 5 min at RT. Meanwhile the DNA was diluted in OptiMem. The two premixes were combined and incubated for 20 min at RT. The old cell media was replaced with fresh media and the combined Lipofectamine/DNA transfection mix was added to the well. The cells were incubated in the 37ºC CO_2 incubator. After 3-6 hrs the media was removed and the cells were washed twice with 1xPBS before adding fresh growth media with FBS. The plates were placed back into the incubator and left until the next day. See **Table 3.19 and Table 3.20** for exact amount of reagents for the different culture vessels and experiments.

Table 3.19: Amounts of reagents for premix 1 of different experiments and culture vessels

Reagents for premix 1	WB or PCR studies		Confocal studies
	12 well plate	6 well plate	8 well chamber slides
Lipofectamine™2000 or Metafectene®Pro	7.5 µl	7.5 µl	1 µl (Lipofectamine™2000)
OptiMem or growth media	250 µl	500 µl	25 µl

Table 3.20: Amounts of reagents for premix 2 of different experiments and culture vessels

Reagents for premix 2	WB or PCR studies		Confocal studies
	12 well plate	6 well plate	8 well chamber slides
DNA	1-3 µg	1-3 µg	0.4 µg
OptiMem or growth media	250 µl	500 µl	50 µl

3.10.2 Metafectene®Pro protocol

The cells were seeded out in a concentration optimal for the experiment as described in the seeding out protocol. The following day Metafectene®Pro (**Table 2.10**) was mixed with growth media and incubated for 5 min at RT. Meanwhile the DNA was diluted in growth media. The two premixes were combined by adding the DNA premix to the premix containing Metafectene and incubated for 20 min at RT. The old cell media was replaced with fresh media and the combined Metafectene/DNA transfection mix was added to the well. The cells were incubated in the 37ºC CO_2 incubator. After 3-6 hrs the media was removed and the cells were washed twice with 1xPBS before adding fresh growth media with FBS. The plates were placed back into the incubator and left until the next day. See **Table 3.19 and Table 3.20** for exact amount of reagents for the different culture vessels and experiments.

3.10.3 Protocol for fixating cells for confocal microscopy

To study the localization of the fluorescence the cells were fixated to preserve the structural integrity, and then observed in a confocal microscope. The fixation was done by using PFA which cross links proteins by creating methylene bridges between reactive groups (54).

The chamber slide was kept on ice through the procedure. Two hundred µl PFA were added to each well and the plate was incubated for 15 minutes. The cells were then washed 4-5 times with 200 µl cold PBS. In the last step 200 µl PBS were added to the wells before covering the plate in aluminum foil and placing it at 4 ºC until microscopy was performed.

3.10.4 Confocal microscopy

The microscopy studies of the chamber slides were done using the Zeiss LSM 510 META confocal microscope. The technique enables obtaining high resolution optical images of sections of one focal plane within a sample, called optical sectioning. The sample is scanned point by point by a laser beam with filters removing unwanted fluorescence from the above and below sections (52, 55).

3.11 SDS PAGE

SDS polyacrylamide gel electrophoresis (SDS-PAGE) is a widely used method for separating proteins based on molecular mass. In this thesis SDS-PAGE has been used to detect proteins in lysates of transfected and harvested cells, and the purified fusion protein GST-KI VP1. Polyacrylamide gels are composed of polymerized acrylamide chains cross-linked in the presence of bis-acrylamide. The density of the gel can be constant or graded, and will cause molecules of different sizes to travel through the gel at different speeds. The proteins in the samples are dissociated into primary polypeptide chains by treatment with SDS or LDS, a reducing agent such as DTT and heat before being loaded into wells in the gel. The denaturation allows separation to be based purely on the molecular mass and not the conformation of the protein. The denatured polypeptides bind SDS and become negatively charged. Markers of known molecular mass are used to compare and estimate the mass (in kDa) of the polypeptide chains in the sample. Once the samples and markers are loaded onto the gel current is passed through and the negatively charged molecules travel through the gel from the cathode towards the anode. The proteins in the polyacrylamide gel can be stained with Coomassie blue or visualized using specific antibodies after transfer to a membrane as in western blotting (47, 50).

3.11.1 Protocol for SDS PAGE

For SDS PAGE gel electrophoresis the NuPage gels (4-12 % gradient) were used (**Table 2.12**). The gel was placed in a XCell SureLock™ Mini-Cell device and 18 µl sample were loaded together with 1.5 µl of the molecular markers SeeBlue® (**Table 2.3**) and MagicMarker™ (**Table 2.3**). The gels were run at 200 V for ~40 min in the NuPage gel-program.

3.12 Western Blotting

Western blotting or immunoblotting is a method where proteins separated by SDS-polyacrylamide gel electrophoresis are transferred from the gel to a membrane and stained by specific antibodies. Before the membrane can be stained with antibodies the proteins in the gel have to be transferred onto a membrane, a process called blotting. The blotting technique is performed by electrophoretic transfer. The gel and membrane is sandwiched in a cassette along with filter papers, sponge pads and buffer, and current is passed through until the proteins have passed through the gel and trapped onto the membrane. After blotting the membrane is soaked in a protein solution of 5 % non-fat dried milk to block the remaining hydrophobic binding sites on the membrane to avoid unnecessary back ground noise from non-specific binding of antibodies. Western blotting uses monoclonal or polyclonal primary antibodies specific to an epitope of the protein, and secondary antibodies with specificity to the constant heavy chain of the primary antibodies and that are conjugated to an enzyme such as alkaline phosphatase or horseradish peroxidase. Substrate is then added which will be converted to product by the enzyme and allows detection of the antigen:primary antibody:secondary antibody complex. Alternatively, secondary antibodies can be labeled with a fluorochrome (47, 50).

3.12.1 Protocol for western blotting

Western blotting was performed at RT unless stated otherwise. All incubations of the membrane in a tub were performed on a shaker at 50-100 rpm, and while handling the membrane inside a centrifuge tube the incubations occur in a rotating wheel. The SDS-PAGE was run as described in the **3.11.1** procedure while the preparations for the blotting procedure was performed. A Immobilon®–P Transfer membrane (**Table 2.12**) was washed for 3 s in methanol, 10 s in dH_2O and for ≥ 5 min in blotting buffer. Sponge pads were also soaked in blotting buffer. After the electrophoresis the gel cassette was opened and the smallest of the sides removed. The wells and the bottom part of the gel were removed. The blotting cassette was assembled by placing presoaked sponge pads and Whatman filter paper (**Table 2.12**), the gel, the membrane, more filter paper and sponge pads in this order in the cassette before sealed tightly with the lid and moving the cassette to the XCell SureLock™ Mini-Cel. The cassette was filled up with blotting buffer while the outer chamber was filled with cold water. The blotting was performed at 30 V for 1 hour.

After blotting the membrane was rinsed for 10 min in PBS in a tub before being incubated in blocking buffer for 1 hour. The membrane was then transferred to a 50 ml centrifuge tube with the protein blotted side inwards and incubated ON at 4 ºC in 3 ml blocking buffer and the primary antibody in a desired dilution. To prevent evaporation the lid was sealed with parafilm.

After incubation with primary antibody the fluid was removed and the membrane washed 3x5 min with 3-5 ml PBST. The PBST was removed and 3 ml blocking buffer and secondary antibody in the desired dilution was added and incubated for 1 hour. After incubation with the secondary antibody the membrane was washed 2x 5 min with 3-5 ml PBST and 2x 5 min with 3-5 ml 1x Washing buffer. Then 5 ml CDP star buffer and 10 µl CDP star was added to the membrane and incubated for 5 min. The membrane was removed from the tube, sealed in plastic and left in the dark for a few minutes before development using Fujifilm Image Analyzer LAS-3000.

3.12.2 Stripping the membrane

The membrane can be stripped of antibodies to allow a new immunostaining to be performed using different antibodies. This is useful in the situations where more than one protein is investigated, for instance the protein of interest and a loading control.

3.12.2.1 Protocol for membrane stripping

The membrane was removed from the plastic and incubated in a tub with 0.2 M NaOH for 5 min with shaking. Afterwards the membrane was washed 3x5 min with PBST and the steps from procedure **3.12.1** was continued onwards from the one hour blocking of the membrane.

3.13 GST-protein purification

The glutathione S-transferases are present in both eukaryotes and prokaryotes and act as detoxication enzymes (56). In this thesis GST-KI VP1 plasmid was transformed into a protease deficient strain BL21. The GST expression vector contains the IPTG-inducible *lac* promoter. The product will be a fusion protein with a GST tag which can be cleaved off after purification. The binding ability between GST proteins and glutathione is used for affinity purification of GST tagged fusion proteins from a solution. The GST tag is purified by

binding to glutathione beads and is eluted by adding a solution of free reduced glutathione which will compete with the glutathione on the beads for the binding of GST protein (57).

The theoretical mass of the GST tag is 26 kDa, which together with the KIPyV VP1 makes up a theoretical mass of 67.6 kDa for the fusion protein GST-KI VP1. After purification the rabbit IgG antibodies vs. EP101303 was used to detect the fusion protein.

3.13.1 Protocol for GST-protein purification

All solutions used during protein isolation should be kept on ice, and all work performed on ice. Gluthatione beads and protease inhibitor cocktail were prepared prior to purification as described in the **Table 2.2**. Before purification was possible, BL21 cells needed to be transformed with the GST-KI VP1 expression plasmid. One colony from the LB plate was divided in two tubes with 5 ml LB w/amp and grown ON in a 37 ºC shaking incubator. The overnight culture was added 100 ml NZCYM w/amp and grown until $OD_{600} = 0.6$. Five ml of the culture was used as a control of expression prior to induction, and kept on ice until it was spun down and frozen with the rest of the cells. IPTG was added to the culture in a final concentration of 1 mM and the culture was incubated at RT for 2 hours. The bacterial culture and the non-induced culture were then transferred into 50 ml centrifuge tubes and spun at 4000 rpm for 20 min at 4 ºC. The bacterial pellet was at this point storable at -20ºC until further use. After thawing the pellets on ice they were resuspended in 5 ml PBT and 200 µl of protease inhibitor cocktail. A sonication was performed on ice 3x10 sec before centrifugation at 9,500 rpm for 10 min at 4 ºC. The supernatants were transferred to sterile 15 ml tubes, 250 µl 50 % glutathione beads were added and the tubes were incubated in a rotor for 30-60 min at 4 ºC. The beads were then washed two times with PBT and once with PBS before another centrifugation at 4,000 rpm for 1 min. After the last washing step the supernatants were removed and the protein eluted from the beads by twice adding 250 µl 5 mM glutathione and separating the beads from the supernatant by quickly spinning the tube.

The purified protein was evaluated by SDS-Page and Coomassie blue staining, protein concentration measurement and WB stained with the antibodies against KI VP1 and the secondary Polyclonal Rabbit Anti-mouse Ig/AP antibodies.

3.13.2 Coomassie blue staining

The result of the GST-purification was evaluated by staining the SDS-Page gel with Coomassie Blue staining. The dye binds non-specifically to most proteins and will stain the

molecular markers and the samples in the gel providing visual comparison and assessment of purity and size of the proteins present. The gel was destained to remove unbound background dye. After GST-protein purification it was expected to find only one product and therefore only one band on the gel (52).

3.13.2.1 Protocol for Coomassie blue staining

All incubations were performed at RT. After SDS-PAGE the gel was soaked in a fixation solution and incubated for 1 hr at 50 rpm on a shaking incubator. The fixation solution was removed and replaced by Coomassie blue solution and incubated for 30 min at 50 rpm. The Coomassie blue solution was removed and the gel rinsed once in fixation solution and then destaining solution was added together with a piece of paper to increase the absorbing the dye). Before incubating the gel overnight the destaining solution was changed and the paper removed in exchange for fresh destaining solution. The gel was photographed the following day in the Bio-Rad Gel Doc 2000 using the Quantity One software.

3.13.3 Determination of protein concentration

To determine the amount of protein in the eluate after the GST-protein purification, the concentration was measured by the use of the Bio-Rad Protein Assay. The assay was made with a dye Coomassie Brilliant Blue G-250 which changes color in response to the concentration of the protein in the sample. The dye binds to primarily basic and aromatic amino acid residues (58).

3.13.3.1 Protocol for protein measurement

The Bio-Rad Protein Assay was diluted 1:5 in dH_2O and 200 µl was added to each well in a microtiter plate. A serial dilution of a globuline standard (1.44 mg/ml, 0.5 mg/ml, 0.25 mg/ml; 0.1 mg/ml, 0.075 mg/ml, 0.05 mg/ml) was made and the purified protein sample and negative control was diluted 1:5 in dH_2O. Ten µl of the standards, both the diluted and undiluted samples and controls were added in parallels to the protein assay. The content was mixed thoroughly by pipetting before incubation for 5 min in RT followed by measuring the absorbance at 590 nm using the $\text{THERMO}_{\text{MAX}}$ Microplate reader.

4 Results

4.1 Study of the permissivity of A549 cell line to KIPyV DNA

To investigate whether the A549 cell line was permissive to KIPyV, cells were transfected with KIPyV DNA. Medium or cells were harvested at different time points and PCR with KIPyV specific primers was performed on medium while RT-PCR was performed on RNA isolated from cells. To evaluate the PCR results a primer sensitivity test for both VP1 and NCCR primer sets were required. Bands of expected lengths were seen for all primer sets (**Table 2.4**). KI DNA can be detected in concentrations of ≤ 100 fg/µl by the use of nested VP1 primers (**Figure 4.1**) and 250 fg/µl by the use of NCCR primers (**Figure 4.2**).

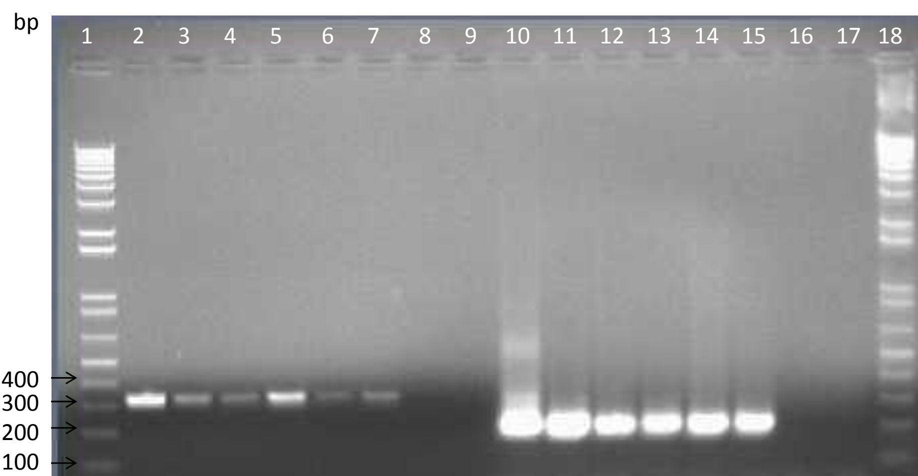


Figure 4.1: Agarose gel electrophoresis of KI VP1 primer sensitivity test using 1:10 serial dilutions of EGFP-KI VP1 plasmid. Well 1 and 18: 1 Kb Plus DNA ladder, wells 8, 9, 16 and 17: Negative water control, wells 2-7: PCR 1 product 10 ng/µl-100 fg/µl, wells 10-15: nested PCR product 10 ng/µl-100 fg/µl.

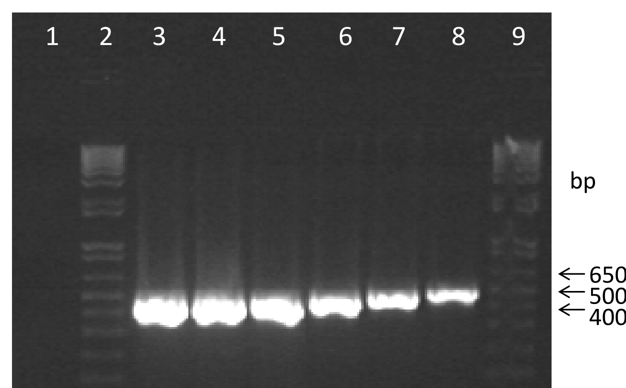


Figure 4.2: Agarose gel electrophoresis of KI NCCR primer sensitivity test using 1:10 serial dilutions of the KIPyV genome. Well 1: negative control, well 2 and 9: 1 Kb Plus DNA ladder, wells 3-8: PCR products of dilutions 25 ng/µl-250 fg/µl.

Due to lack of an infectious virus transfections of A549 cells with KIPyV genome were performed. One µg KIPyV DNA and 3 µg of the controls EGFP-KI VP1 and pEGFP-C1 was used to transfect 300,000 A549 cells per well. The control plasmids emit fluorescent light upon expression by the host cell and were used to estimate transfection efficiency. The media was always changed 3-6 hours after transfection to remove any DNA not taken up by the cells before harvesting. The KIPyV transfected cells were harvested after 24, 48, 72 hours and one and four weeks. KIPyV DNA was detected in the media of A549 cells after 24 hrs and up to 4 weeks with the use of nested VP1 PCR (**Figure 4.3**). PCR with NCCR primers gave no positive transfection samples (**Figure 4.4**).

A co-transfection of 150,000 cells per well with a total of 3 µg DNA was performed. KIPyV DNA was co-transfected with CT-DNA, BKPyV genome and an expression plasmid for the BKPyV agno protein. Agnoprotein has been shown to be necessary for viral maturation and release (20). Because KIPyV does not express agnoprotein we reasoned that provided agnoprotein may help propagation of KIPyV. Medium was harvested after 24 and 48 hours, and one and two weeks. KIPyV DNA was detected in the media of co-transfected A549 cells after 24 hrs for all co-transfections and up to two weeks with the use of nested VP1 PCR (**Figure 4.3**). After PCR with NCCR primers only KIPyV co-transfected with CT-DNA and harvested after 48 hours were positive (**Figure 4.4**). There were generally a high number of cell deaths with each transfection.

All primer sets gave positive KIPyV genome control and negative water controls. EGFP-C1 plasmid were included to display that the primers were specific for KIPyV and did not react with EGFP DNA from the EGFP-KI VP1 plasmid which was used in the sensitivity test of VP1 nested PCR (nPCR). Moreover, expression of EGFP could be easily monitored in the transfected cells and was used to estimate the transfection efficiency.

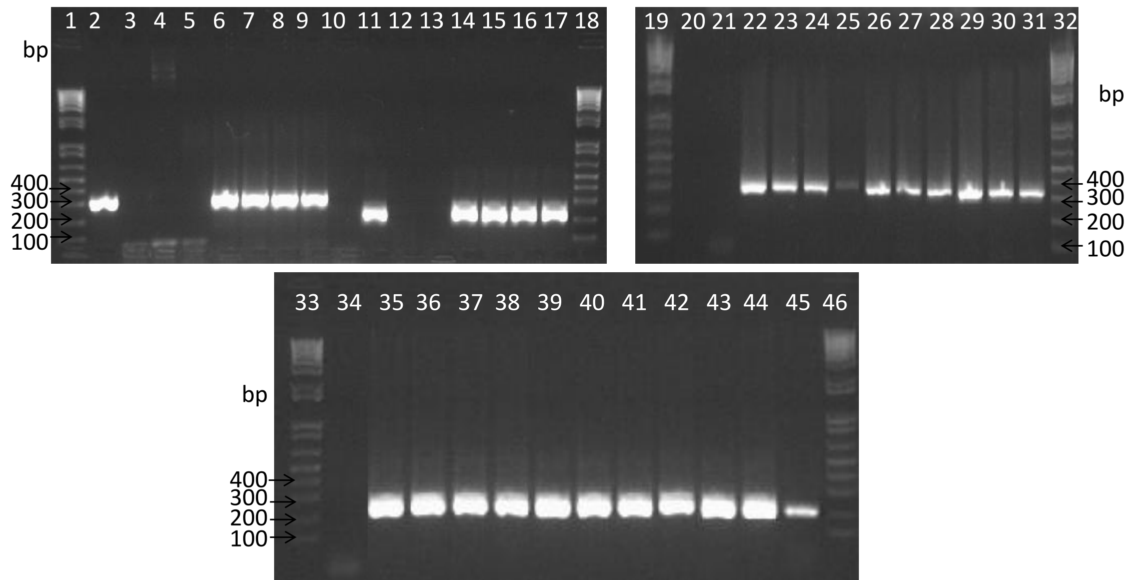


Figure 4.3: Agarose gel electrophoresis of VP1 PCR of media harvested from transfected A549 cells, wells 2-10 and 20-31: PCR1, wells 11-17 and 34-45: PCR2. Well 1, 18, 19, 32, 33, 45 : 1 Kb Plus DNA ladder, well 2 and 11: positive control (KI genome), well 3 and 12: negative control (H_2O), well 4: EGFP-C1, well 5 and 13: negative transfection, well 6 and 14: KIPyV 24 hrs, well 7 and 15: KIPyV 48 hrs, well 8 and 16: KIPyV 72 hr, well 9 and 17: KIPyV 1 week, well 10 and 45: KIPyV 4 weeks, well 20: empty, well 21 and 34: negative transfection, well 22 and 35: KIPyV + CT-DNA 24 hrs, well 23 and 36: KIPyV + CT-DNA 48 hrs, well 24 and 37: KIPyV + CT-DNA 1 week, well 25 and 38: KIPyV + CT-DNA 2 weeks, well 26 and 39: KIPyV + BKPyV 48 hrs, well 27 and 40: KIPyV + BKPyV 1 week, well 28 and 41: KIPyV + BKPyV 2 weeks, well 29 and 42: KIPyV + agno 48 hrs, well 30 and 43: KIPyV + agno 1 week, well 31 and 44: KIPyV + agno 2 weeks.

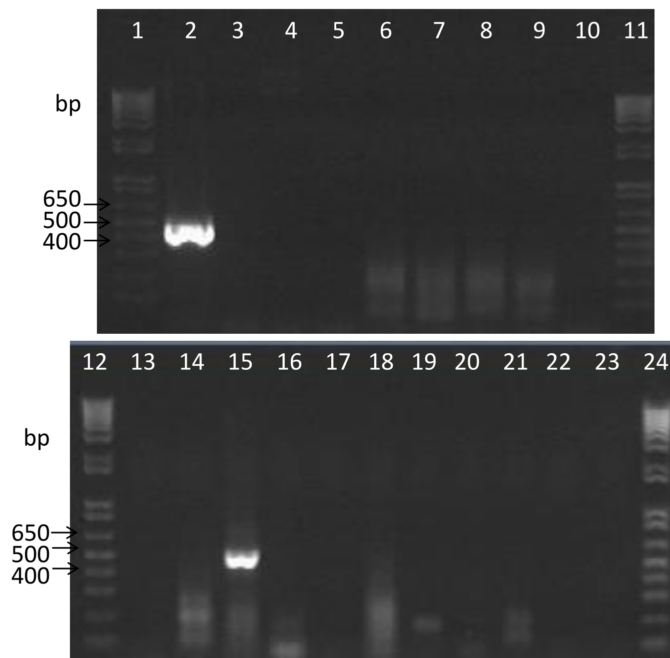


Figure 4.4: Agarose gel electrophoresis of NCCR PCR of media harvested from transfected A549 cells (wells 5-10 and 14-24). Well 1,11,12 and 24: 1 Kb Plus DNA ladder, well 2: positive control (KIPyV), well 3: negative control (H_2O), well 4: EGFP-C1 control, well 5: negative transfection, well 6: KIPyV 24 hrs, well 7: KIPyV 48 hrs, well 8: KIPyV 72 hr, well 9: KIPyV 1 week, well 10: KIPyV 4 weeks, well 13: negative transfection, well 14: KIPyV + CT-DNA 24 hrs, well 15: KIPyV + CT-DNA 48 hrs, well 16: KIPyV + CT-DNA 1 week, well 17: KIPyV + CT-DNA 2 weeks, well 18: KIPyV + BKPyV 48 hrs, well 19: KIPyV + BKPyV 1 week, well 20: KIPyV + BKPyV 2 weeks, well 21: KIPyV + agno 48 hrs, well 22: KIPyV + agno 1 week, well 23: KIPyV + agno 2 weeks.

To further investigate if these results may in fact be caused by replicated DNA and not only DNA released by cells that have died of the transfection; total RNA was isolated from KIPyV transfected cells. A negative control where no DNA had been added to the cells was also included. A549 cells and HEK293 cells were seeded out with 200,000 and 300,000 cells per well, respectively. The cells were transfected and after 48 hours cells from one of each well were transferred to a cultivation flask where they were kept until the cells were almost confluent. The cells were harvested after 3-4 days and RNA was extracted. The RNA was reverse transcribed and PCR of LT-ag cDNA was performed to see if this viral gene was transcribed in transfected cells. As a control for the RT-PCR, cDNA of CREB was amplified. The PCR results are presented in **Figure 4.5**. Strong bands of expected size (**Table 2.4**) were detected by the LT-ag primers in the wells originating from KI transfected cells and bands of expected size (**Table 2.4**) were seen in all samples from the CREB PCR. The negative water sample remained negative after both PCRs. The negative transfection sample from HEK293 cells was negative for LT-ag, but the negative transfection sample from A549 cells produced a weak LT-ag band. These results provide evidence for transcription of viral DNA in both A549 and HEK293 cells. To further investigate whether the transcripts are expressed WB was performed on transfected cell lysate to look for the VP1 protein.

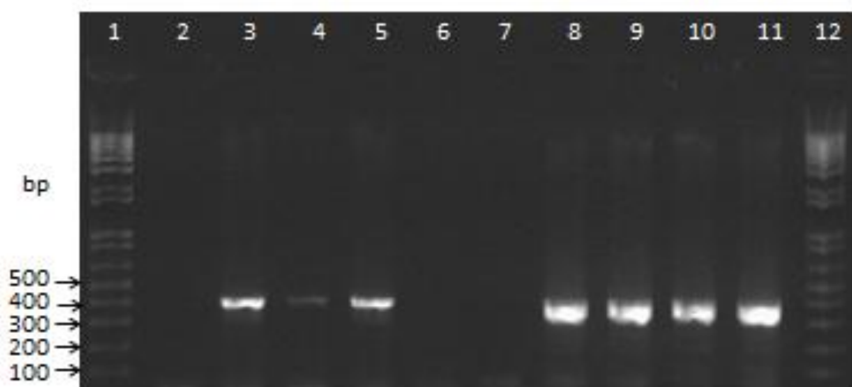


Figure 4.5: Agarose gel electrophoresis of LT-ag (wells 2-6) and CREB (wells 7-11) PCR of cDNA. Well 1 and 12: 1 Kb Plus DNA ladder; wells 2 and 7: Negative control (H_2O), wells 3 and 8: KIPyV transfected A549 cells, well 4 and 9: Negative transfection of A549 cells, wells 5 and 10: KIPyV transfected HEK293 cells, wells 6 and 11: Negative transfection of HEK293 cells.

4.1.1 Antibody detection of KIPyV VP1

To monitor viral protein production in transfected cells, KIPyV VP1 antibodies were generated as described in Material **Table 2.11**. These antibodies were first validated for their reactivity against EGFP-KI VP1 and GST-KI VP1 fusion protein. To ensure that the

transfection process was successful and to provide a fusion protein which could be recognized by the KI VP1 antibody a recombinant EGFP-KI VP1 expression plasmid was made (**Supplementary Figure 4**). A GST-KI VP1 expression vector was also constructed and the fusion protein product purified from bacteria was used as a control for the KI VP1 antibody and in the WB analysis (**Supplementary Figure 5**).

A visual evaluation of transfected cells 24 hours after transfection generally showed fluorescent signals in all wells containing EGFP plasmids; however the EGFP-KI VP1 fusion protein signal was much rarer. This could mean that EGFP-KI VP1 had much lower transfection efficiency than the rest of the EGFP plasmids or that the EGFP signal in these cells is generally weaker resulting in fusion proteins not even being detected. The fluorescent signals were more frequently observed in HEK cells (~80 %) than in A549 cells (below 50 %).

4.1.1.1 Transfection and antibody detection of EGFP plasmids

Before the KI VP1 antibody was tested cells were transfected with different EGFP expression plasmids or EGFP fusion plasmids and proteins detected by antibodies directed against EGFP to evaluate the expected mass of the EGFP-KI VP1 fusion protein.

A549 and HEK293 cells were transfected with EGFP-C1, EGFP-BKV VP1 and EGFP-KI VP1 and a negative control with water instead of DNA. Cells were harvested and lysed before being analyzed by running WB using primary antibody Anti-GFP IgG Mouse (**Table 2.11**) and secondary antibody Polyclonal Rabbit Anti-mouse Ig/AP (**Table 2.11**). The results from the WB are presented in **Figure 4.6**. The negative control did not show any EGFP protein bands as expected. The EGFP-C1 transfected cell lysate provided a clear band of ~30 kDa which is expected (**Table 2.6**). The EGFP-KI VP1 transfected cell lysates display a strong band of ~70 kDa and a week band of ~60 kDa. The strong band corresponds to the expected mass of the EGFP-KI VP1 fusion protein (**Table 2.6**). The EGFP-BKV VP1 transfected cell lysates have a strong band that is a tiny bit larger and a weak band slightly smaller than the EGFP KI VP1 samples strong band. The larger band corresponds to the theoretical mass of the fusion protein EGFP-BKV VP1 (**Table 2.6**). The EGFP-KI VP1 transfected A549 cell lysate has a much weaker band than the HEK293 transfected cell lysate.

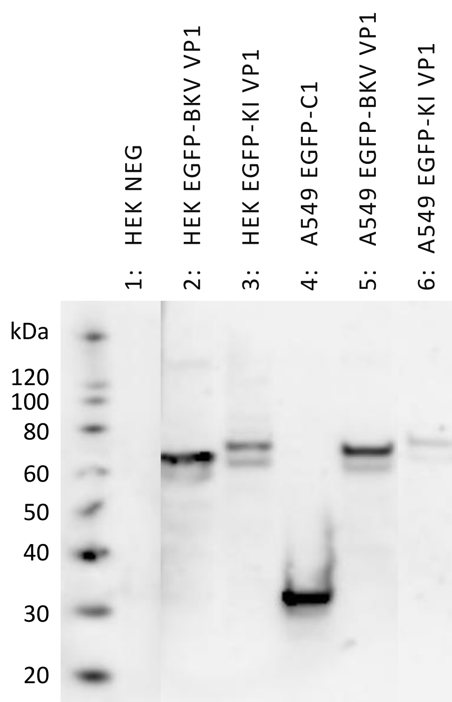


Figure 4.6: WB of transfected HEK293 and A549 cells. The amount of DNA used in the transfections was 3 µg and the amount of cells per well was 250,000 HEK293 cells and 200,000 A549 cells. The GFP antibody dilution was 1:1000

4.1.1.2 Transfection and antibody detection of KIPyV VP1

Two different antibodies directed against KIPyV VP1 were tested on cell lysate from EGFP-C1 and EGFP-KI VP1 transfected HEK293 cells. Cells were harvested and lysed before being analyzed by running WB using primary antibodies Polyclonal Rabbit EP101303 and EP101304 (**Table 2.11**) raised against KIPyV VP1, and secondary antibod**y** Goat Anti-Rabbit Ig human ads-AP (**Table 2.11**).

The EP101303 antibody was able to detect a band of between 60 and 80 kDa, corresponding to the expected ~71.4 kDa size of the EGFP-KI VP1 fusion protein, and did not detect a band of this length in the negative EGFP-C1 control (**Figure 4.7**). This is most likely the EGFP-KI VP1 fusion protein. There is a weak band of ~40 kDa in the EGFP-KI VP1 sample which is not present in the EGFP-C1 sample. The EP101304 antibody did not produce any clear bands in the EGFP-KI VP1 sample or the control and will not be used further. From now on the EP101303 is referred to as KI VP1 antibody.

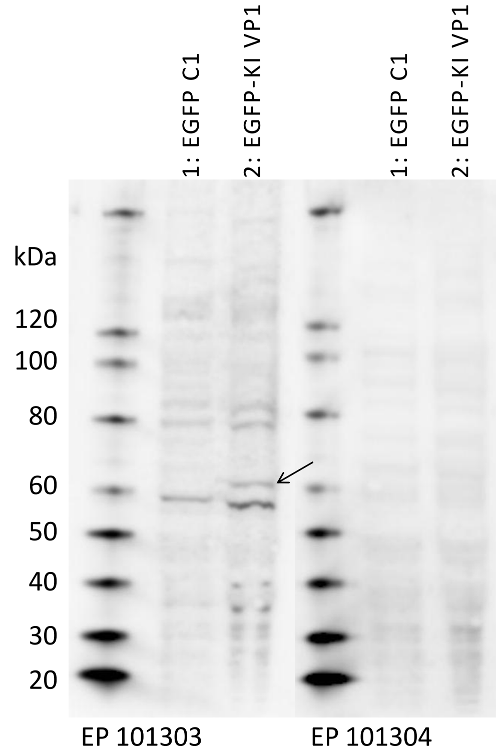


Figure 4.7: WB of transfected HEK293 cells. The arrow points to what is probably the EGFP-KI VP1 fusion protein. The amount of DNA used in the transfections was 3 µg and the amount of cells per well was 250 000. The two KI VP1 antibody dilutions were 1:1000

4.1.1.2.1 Expression of KIPyV VP1 in bacterial cells

The KI VP1 antibody was also tested on purified GST-KI VP1 protein. The purified GST fusion protein was run on a SDS gel and dyed with Coomassie blue, but the result was too weak to gather any information about the protein (result not included), however in WB the protein was detectable. The concentration was also measured by a protein assay (results not included). The KI VP1 antibody was used to detect purified GST-KI VP1 fusion protein and the results from the WB are presented in **Figure 4.8**. The GST-KI VP1 fusion protein has an expected molecular mass of ~67.6 kDa (**Table 2.6**) and a band of this size was seen in both the purified protein and the beads. No bands were seen in the non-induced control as expected. Lane 4 in **Figure 4.8** is a repeated WB of GST-KI VP1 made because of the over exposed first picture.

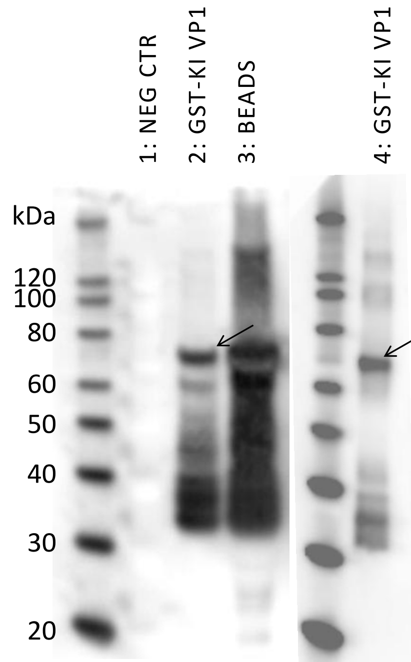


Figure 4.8: WB of purified GST-KI VP1 protein expressed in *E. coli* BL21 cells. The arrows point to what is probably the GST-KI VP1 fusion protein. The KI VP1 antibody dilutions were 1:1000 and 1:500.

4.1.1.3 Detection of VP1 protein in KIPyV transfected A549 cell line

To see if there was viral protein production in A549 cells the KI VP1 antibody was tested on lysates prepared from KIPyV DNA transfected cells. Lysates from EGFP-C1 and EGFP-KI VP1 transfected cells were used as controls. In addition EGFP-KI VP1 transfected HEK293 cell lysate from previous analysis was included as a control of the antibody and the EGFP-C1 as a negative control. The cells were harvested after 48 hours and KIPyV transfected cells were also harvested after one week. The cells were lysed and analyzed by WB. The results of the WB analysis are presented in **Figure 4.9**. There was no visible band of ~70 kDa for the lysate from the EGFP-KI VP1 transfected A549 cells, nor were there visible bands of ~40 kDa observed for the lysate from any of the KI genome transfected cells. The EGFP-KI VP1 control from HEK cell lysate showed a strong band of a little bit less than 60 kDa and a very weak band of between 60 and 80 kDa, however it did not quite resemble the first WB where the band was a little bit bigger than 60 kDa.

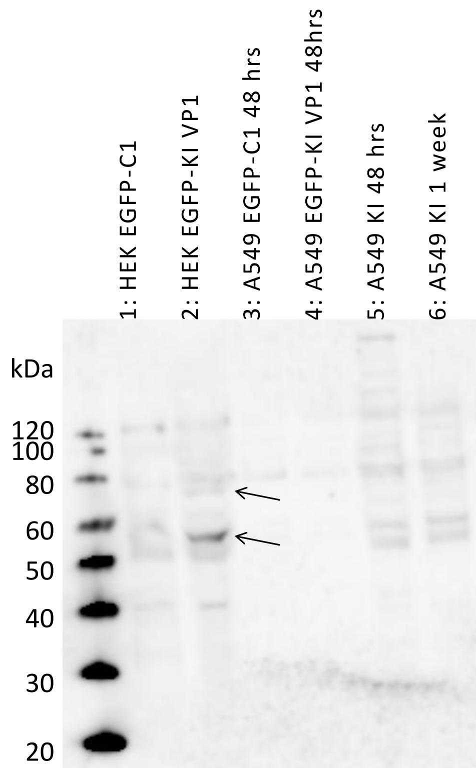


Figure 4.9: WB of transfected HEK293 and A549 cells. The arrows point to what may be the EGFP-KI VP1 fusion protein. One µg of KIPyV DNA and 3 µg of EGFP-C1 and EGFP-KI VP1 control plasmids were transfected per 200 000 A549 cells. The samples from HEK293 cells are the same as in **Figure 4.7**. The KI VP1 antibody dilution was 1:1000.

A549 cells were transfected with expression plasmids for EGFP-KI VP1, EGFP-BKV VP1 and the KIPyV genome which was also co-transfected with BKPyV genome and with an expression plasmid for BKPyV agno protein. The reason for the co-transfection was that we hypothesized that BKPyV proteins may stimulate replication of KIPyV. For comparison a pair of wells was transfected with KIPyV and CT-DNA. A negative control and the EGFP-BKV VP1 were only used for observation of the transfection process which was normal, and cells were not harvested for WB. The cells transfected with the EGFP-KI VP1 were harvested after 24 hours while the cells transfected with KIPyV were harvested after 24 hours and one week. The purified GST-KI VP1 protein was used as a control of the KI VP1 antibody which function was normal.

The results of the WB analysis are presented in **Figure 4.10** and the samples included described in **Table 4.1**. No band of ~70 kDa for the EGFP-KI VP1 control was observed. No specific band of 41.2 kDa was seen only for the KIPyV genome transfected cells or the co-transfections, although a non-specific band of ~ 40 kDa was observed throughout the samples

and stronger in the 1 week samples. There is a lot of noise from non-specific binding of antibodies which was not present in the testing of the antibody at dilution 1:1000. However because there seemed to be weak signals in transfected cell lysates (**Figure 4.7**) the dilution was kept low.

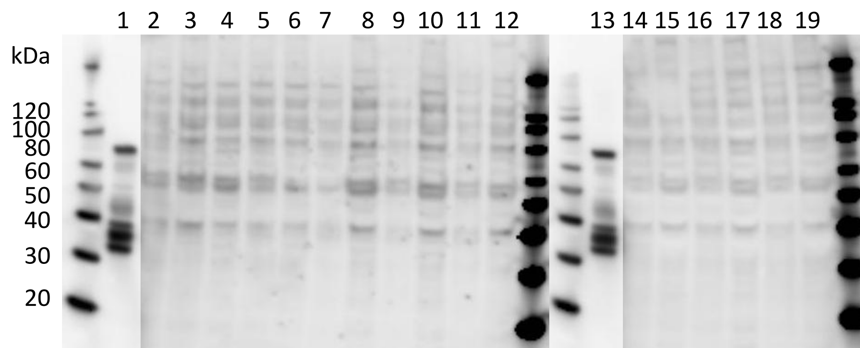


Figure 4.10: WB of transfected A549 cells. Cells were transfected with 1 and 3 µg DNA of the controls EGFP-KI VP1, EGFP-BKV-VP1 and KIPyV genome which was also co-transfected with BKPyV genome and the agno gene where the total amount of DNA was 3 µg. For comparison a pair of wells was transfected with 3 µg of KIPyV and CT-DNA. The samples are presented in **Table 4.1**. KI VP1 antibody dilution was 1:600.

Table 4.1: The samples analyzed by WB in **Figure 4.10** and their individual conditions.

Well	c/w	Amount of plasmid	Plasmid used for transfection	Time after transfection
1 and 13	-	-	GST-KI VP1 control (protein)	-
2	200 000	3 µg	EGFP-KI VP1	24 hrs
3	200 000	1 µg	KIPyV genome	24 hrs
4	200 000	1 µg	KIPyV genome	1 week
5	200 000	3 µg	KIPyV genome	24 hrs
6	200 000	3 µg	KIPyV genome	1 week
7	150 000	3 µg	KIPyV genome + CT-DNA	24 hrs
8	150 000	3 µg	KIPyV genome + CT-DNA	1 week
9	150 000	3 µg	KIPyV genome + BKPyV genome	24 hrs
10	150 000	3 µg	KIPyV genome + BKPyV genome	1 week
11	150 000	3 µg	KIPyV genome + agno gene	24 hrs
12	150 000	3 µg	KIPyV genome + agno gene	1 week
14	100 000	3 µg	KIPyV genome + CT-DNA	24 hrs
15	100 000	3 µg	KIPyV genome + CT-DNA	1 week
16	100 000	3 µg	KIPyV genome + BKPyV genome	24 hrs
17	100 000	3 µg	KIPyV genome + BKPyV genome	1 week
18	100 000	3 µg	KIPyV genome + agno gene	24 hrs
19	200 000	3 µg	KIPyV genome + agno gene	1 week

A549 cells were transfected with expression plasmids for EGFP-KI VP1 or EGFP-BKV-VP1 or with the KIPyV genomic DNA. The cells transfected with the controls and a mock transfection were harvested after 24 hours while the cells transfected with KIPyV DNA were

harvested after 24 hours, 48 hours and one week. The WB results are presented in **Figure 4.11** and the samples are described in **Table 4.2**. No specific band corresponding to KIPyV VP1 was observed in any of the samples. To exclude that there might irregularities with the WB analysis or the samples loaded the membranes were stripped and challenged with primary ERK 2 (C-14) Rabbit polyclonal IgG (**Table 2.11**) and secondary antibody Goat Anti-Rabbit Ig human ads-AP**.** The WB results are presented in **Figure 4.12**. The WB showed a quite similar protein distribution in wells 2-10, while in wells 13-20 the amount of protein loaded was lower and irregular. There was no correlation between the signal strength and the amount of cells used for transfection in these wells. There seemed to be some remaining signal from the first WB which could be seen in wells 1 and 11 containing GST-KI PyV protein. As all samples produced a clear signal with anti-ERK-2 antibodies there did not seem to be anything wrong with the samples that could cause the result in **Figure 4.11**. Therefore the WB with antibodies against KI VP1 was repeated. After repeating the WB of the transfected A549 cells there still seemed to be poor results seen in **Figure 4.13**. No distinct bands were seen for any of the controls but GST-KI VP1, or for the KIPyV transfected cell lysates.

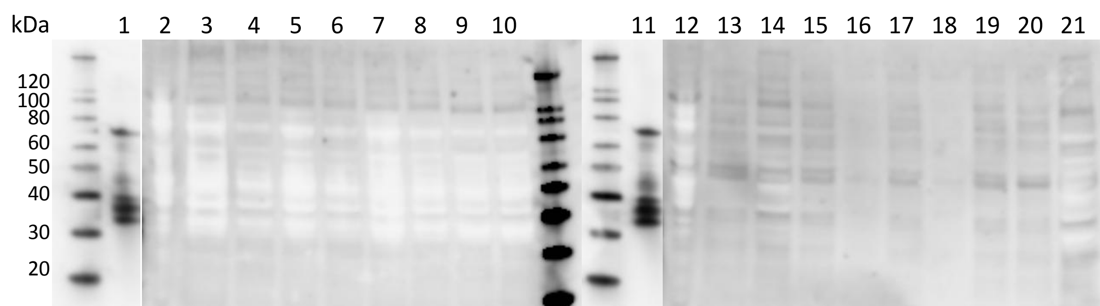


Figure 4.11: WB of transfected A549 cells. The samples are presented in **Table 4.2.** KI VP1 antibody dilution was 1:1000.

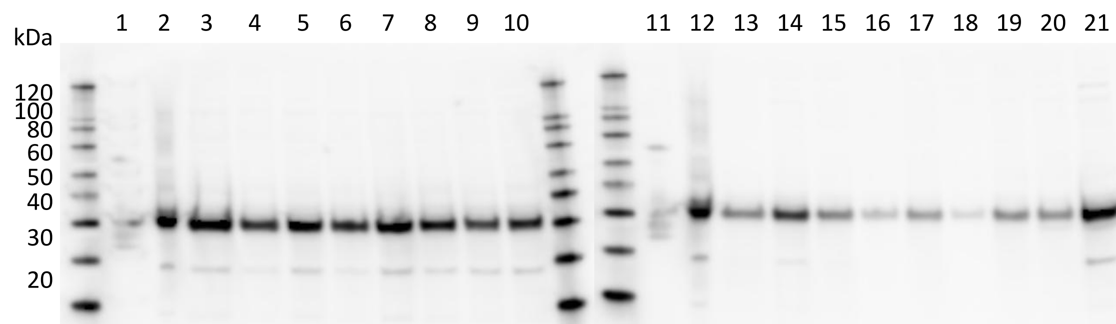


Figure 4.12: WB of transfected A549 cells. After stripping of membrane and using ERK2 primary antibodies to evaluate the sample loading. The samples are presented in **Table 4.2**.

Table 4.2: The samples analyzed by WB in **Figure 4.11** and **Figure 4.12** and their individual conditions.

Well	c/w	Amount of plasmid	Plasmid used for transfection	Time after transfection
1 and 11	-	-	GST-KI VP1 control (protein)	-
2 and 12	250 000 (HEK293)	3 µg	EGFP-KI VP1	24 hrs
3	100 000	-	NEG	24 hrs
4	200 000	3 µg	EGFP- BKV VP1	24 hrs
5	200 000	1 µg	KIPyV genome	24 hrs
6	200 000	3 µg	KIPyV genome	24 hrs
7	200 000	1 µg	KIPyV genome	48 hrs
8	200 000	3 µg	KIPyV genome	48 hrs
9	200 000	1 µg	KIPyV genome	1 week
10	200 000	3 µg	KIPyV genome	1 week
13	150 000	-	NEG	24 hrs
14	100 000	-	NEG	24 hrs
15	100 000	3 µg	EGFP-KI VP1	24 hrs
16	150 000	3 µg	KIPyV genome	24 hrs
17	100 000	3 µg	KIPyV genome	24 hrs
18	150 000	3 µg	KIPyV genome	48 hrs
19	100 000	3 µg	KIPyV genome	48 hrs
20	150 000	3 µg	KIPyV genome	1 week
21	100 000	3 µg	KIPyV genome	1 week

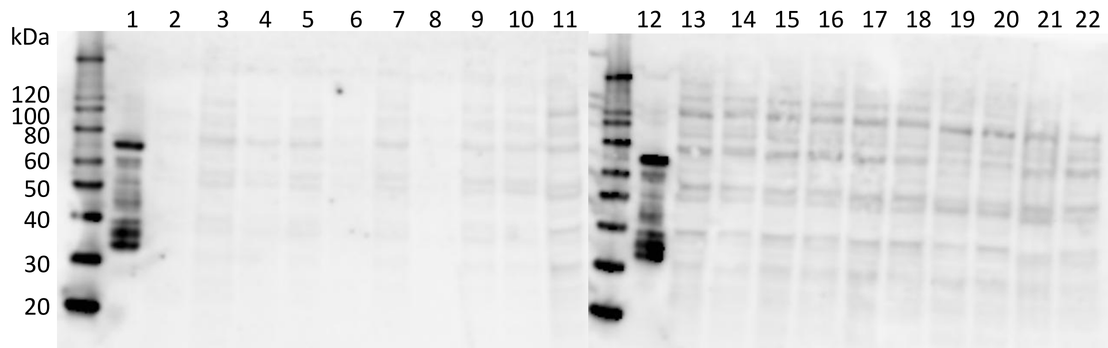


Figure 4.13: Repeated WB of transfected A549 cells. KI VP1 antibody dilution was 1:500. The samples are presented in **Table 4.3**.

Table 4.3: The samples analyzed by WB in **Figure 4.13** and their individual conditions.

Well	c/w	Amount of plasmid	Plasmid used for transfection	Time after transfection
1 and 12	-	-	GST-KI VP1 control (protein)	-
2	150 000	3 µg	EGFP-KI VP1	24 hrs
3	100 000	3 µg	EGFP-KI VP1	24 hrs
4	150 000	3 µg	EGFP- BKV VP1	24 hrs
5	100 000	3 µg	EGFP- BKV VP1	24 hrs
6	150 000	3 µg	KIPyV genome	24 hrs
7	100 000	3 µg	KIPyV genome	24 hrs
8	150 000	3 µg	KIPyV genome	48 hrs
9	100 000	3 µg	KIPyV genome	48 hrs
10	150 000	3 µg	KIPyV genome	1 week
11	150 000	3 µg	KIPyV genome	1 week
13	200 000	3 µg	EGFP-C1	24 hrs
14	200 000	3 µg	EGFP-BKV VP1	24 hrs

15	200 000	1 µg	KIPyV genome	24 hrs
16	200 000	3 µg	KIPyV genome	24 hrs
17	200 000	1 µg	KIPyV genome	48 hrs
18	200 000	3 µg	KIPyV genome	48 hrs
19	200 000	1 µg	KIPyV genome	1 week
20	200 000	3 µg	KIPyV genome	1 week
21	200 000	-	NEG	24 hrs
22	100 000	-	NEG	24 hrs

4.1.1.3.1 Infection of A549 cells with media

Before the results from the primary transfections with KIPyV were analyzed an infection of fresh A549 cells with medium from previous transfections of A549 was attempted. One hundred µl media from previous transfections was added to fresh A549 cells and left to incubate for a week before harvesting the cells and running WB to observe if any VP1 protein had been produced. The sample description and transfection conditions can be found in **Table 4.4** and the results are presented in **Figure 4.14**. No strong bands of 41.6 kDa were observed for any of the samples; in fact all samples looked the same as the negative controls. A positive control is lacking to confirm that the KI VP1 antibody is indeed working, however the results from the primary infection gives no reason to expect positive result after addition of their media to fresh cells.

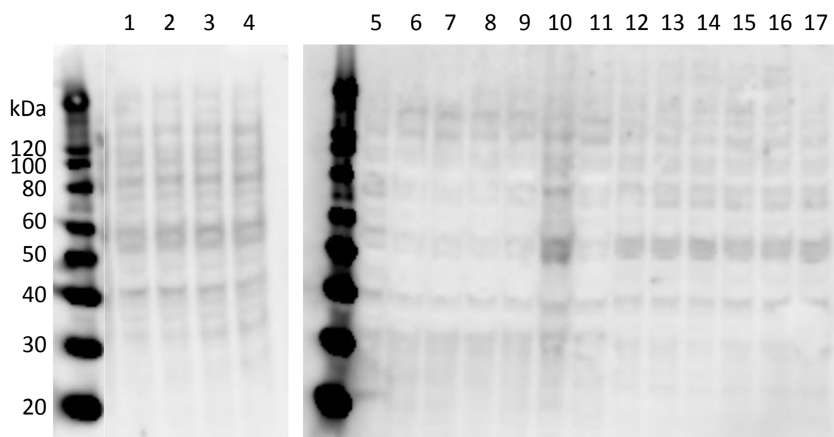


Figure 4.14: WB of lysate from A549 cells that were infected with medium harvested from A549 cells transfected with KIPyV DNA. The samples are presented in **Table 4.4**. KI VP1 antibody dilution was 1:500.

Table 4.4: The samples analyzed by WB in **Figure 4.14** and their individual conditions.* unknown cell count

Well	c/w transfection/ c/w infection	Amount of plasmid in transfection	Plasmid used for transfection	Time after transfection/infection
1	200 000/100 000	3 µg	KIPyV genome	1 week/1 week
2	200 000/100 000	3 µg	KIPyV genome + BKPyV genome	1 week/1 week
3	200 000/100 000	3 µg	KIPyV genome + agno gene	1 week/1 week
4	*/100 000	1 µg	KIPyV genome	4 weeks/1 week

5	200 000/100 000	1 µg	KIPyV genome	1 week/1 week
6	150 000/100 000	3 µg	KIPyV genome	24 hrs/1 week
7	100 000/100 000	3 µg	KIPyV genome	24 hrs/1 week
8	200 000/100 000	3 µg	KIPyV genome	24 hrs/1 week
9	150 000/100 000	3 µg	KIPyV genome 1.	24 hrs/1 week
10	150 000/100 000	3 µg	KIPyV genome 2.	24 hrs/1 week
11	/100 000	-	NEG	/1 week
12	150 000/100 000)	3 µg	KIPyV genome	1 week/1 week
13	100 000/100 000)	3 µg	KIPyV genome	1 week/1 week
14	200 000/100 000	3 µg	KIPyV genome	1 week/1 week
15	150 000/100 000	3 µg	KIPyV genome 1.	1 week/1 week
16	150 000/100 000	3 µg	KIPyV genome 2.	1 week/1 week
17	/100 000	-	NEG	/1 week

4.1.1.4 Detection of VP1 protein in KIPyV transfected HEK293 cell line

HEK293 were transfected with the expression plasmids for EGFP-KI VP1 or EGFP-BKV-VP1 or with the KIPyV genome. Transfection with KIPyV DNA was also done in the presence of BKPyV genome and an expression plasmid for BKPyV agno protein. Description of the samples and the transfection conditions can be seen in **Table 4.5**, and the results are presented in **Figure 4.15**. The EGFP-KI VP1 control does not appear as it should on the WB, but the GST-KI VP1 fusion protein was detected by the VP1 antibody. All cell lysates gave a non-specific band of ~40 kDa. The bands of the one week lysates are in general stronger than the others, however the negative control resembles these samples more than the 48 hour samples. There are two bands of ~70 kDa in the EGFP-KI VP1 transfected cell lysate and the KIPyV and BKPyV co-transfection. This is the expected size of EGFP-KIVP1.

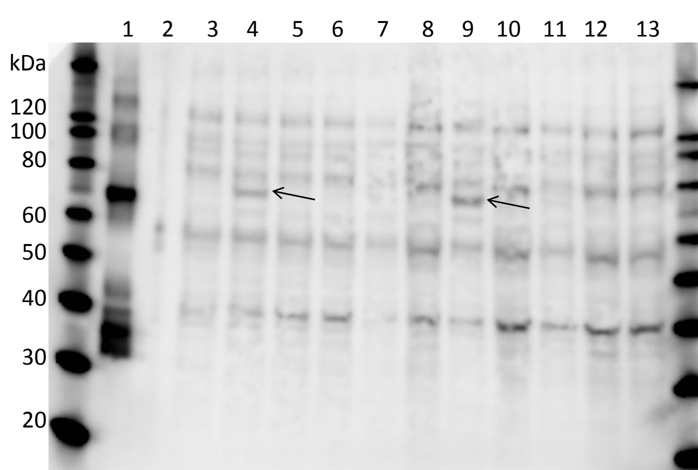


Figure 4.15: WB of transfected HEK293 cells. The arrows point at two bands that are not present in the other lysates. The samples are presented in **Table 4.5.** KI VP1 antibody dilution was 1:500.

Table 4.5: The samples analyzed by WB in **Figure 4.15** and their individual conditions.

Well	c/w	Amount of plasmid	Plasmid used for transfection	Time after transfection
1	-	-	GST-KI VP1 control (protein)	-
2	250 000 (HEK293)	3 µg	EGFP-KI VP1	24 hrs
3	200 000	3 µg	EGFP- BKV VP1	48 hrs
4	200 000	3 µg	EGFP-KI VP1	48 hrs
5	200 000	3 µg	KIPyV genome	48 hrs
6	200 000	3 µg	KIPyV genome	1 week
7	200 000	3 µg	KIPyV genome + CT-DNA	48 hrs
8	200 000	3 µg	KIPyV genome + CT-DNA	1 week
9	200 000	3 µg	KIPyV genome + BKPyV genome	48 hrs
10	200 000	3 µg	KIPyV genome + BKPyV genome	1 week
11	200 000	3 µg	KIPyV genome + agno gene	24 hrs
13	200 000	3 µg	KIPyV genome + agno gene	1 week

4.1.1.4.1 Infection of HEK293 cells with media

Before the results from the primary transfections with KIPyV were analyzed medium from the primary transfection of HEK293 cells was inoculated on fresh HEK293 cells. After one week the cells were harvested and analyzed for VP1 production by WB. The results are presented in **Figure 4.16** and the samples analyzed are described in **Table 4.6**. A positive control to confirm that the primary antibody is in fact working is lacking, but because of time limitations the WB was not repeated. The WB picture looks like all previous transfection lysates and there is no reason to believe the KI VP1 antibody is not working as normal. Some of the samples had a stronger band of ~40 kDa than the others, but a band of corresponding mass can also be observed in lysates of cells not infected (lanes 6 and 12 in **Figure 4.16**).

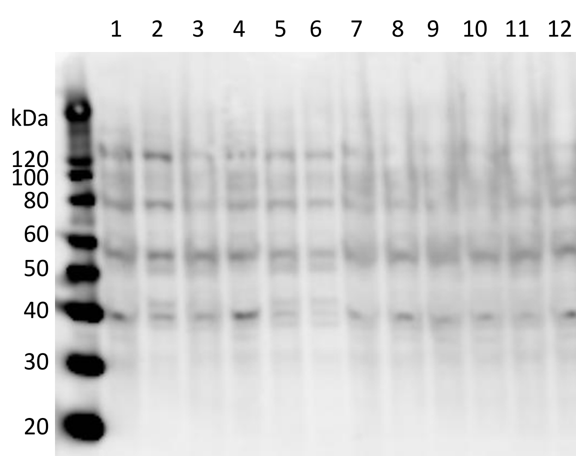


Figure 4.16: WB of lysates prepared from HEK293 cells that were inoculated with 100 µl medium from KIPyV DNA transfected cells. The samples are presented in **Table 4.6**. KI VP1 antibody dilution was 1:500.

Table 4.6: The samples analyzed by WB in **Figure 4.16** and their individual conditions.

Well	c/w	Amount of plasmid	Plasmid used for transfection	Time after transfection
1	100 000	3 µg	KIPyV genome	48 hrs
2	100 000	3 µg	KIPyV genome	1 week
3	100 000	3 µg	KIPyV genome + CT-DNA	1 week
4	100 000	3 µg	KIPyV genome + BKPyV genome	1 week
5	100 000	3 µg	KIPyV genome + agno gene	1 week
6	100 000	-	NEG	48 hrs
7	150 000	3 µg	KIPyV genome	48 hrs
8	150 000	3 µg	KIPyV genome	1 week
9	150 000	3 µg	KIPyV genome + CT-DNA	1 week
10	150 000	3 µg	KIPyV genome + BKPyV genome	1 week
11	150 000	3 µg	KIPyV genome + agno gene	1 week
12	150 000	-	NEG	48 hrs

4.1.1.5 Expression of KIPyV VP1 in mammalian cells

To see if HEK293 cells could express the VP1 protein without the whole KIPyV genome a mammalian expression vector pcDNA3.1$^{(+)}$KI VP1 was made (**Supplementary Figure 6**) and used for transfection. HEK293 cells were transfected with this expression plasmid and the cells were harvested after 48 hrs. VP1 was not detected in any of the wells containing transfected cell lysate, but was detected as normal in the well containing GST-KI VP1 protein (**Figure 4.17**). The antibodies did not detect any EGFP-KI VP1 fusion protein in lysates of cells transfected with an expression plasmid for this fusion protein (**Figure 4.17**, lane 2).

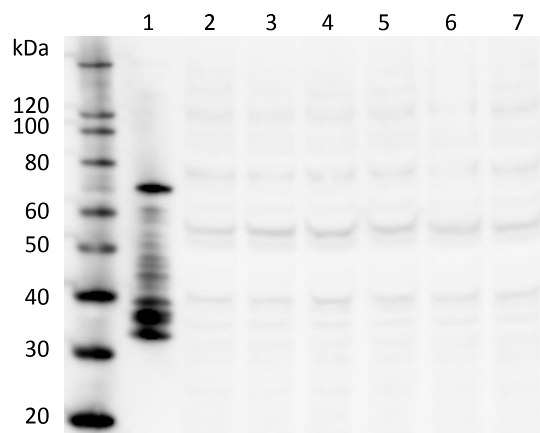


Figure 4.17: WB of lysate from transfected HEK293 cells. Three µg of DNA was used for transfection of 200,000 HEK293 cells. KI VP1 antibody dilution was 1:500. The DNA used for transfection of the cells lysated and analyzed by WB was: 1: GST-KI VP1, 2: EGFP-KI VP1, 3: pcDNA3.1$^{(+)}$, 4: pcDNA3.1$^{(+)}$KI VP1, 5: negative control, 6: KI linear, 7: KI circular.

4.2 Study of the sub-cellular localization of KIPyV VP1 protein

The maturation of viral particles occurs in the nucleus which means that the synthesized capsid proteins are transported from the cytoplasm to the nucleus (2). HPyVs have nuclear localization signals in N-terminal region of VP1 that enable transportation into the nucleus(2). Sequence analysis of the KIPyV shows that KIPyV VP1 protein also has a NLS like motif (**Supplementary Figure 1**). To see whether the VP1 protein does in fact localize in the nucleus cells transfected with EGFP-KI VP1 were fixated and studied by confocal microscopy. The sub-cellular localization of KIPyV VP1 was studied in A549 and Vero cells and for comparison the cells were also transfected with EGFP-C1 and EGFP-BKV VP1.

A selected number of photos which did not represent the frequencies of the observations from the confocal microscopy study are presented in **Figure 4.18** and Figure **4.19**. Previous transfection of Vero cells with EGFP-BKV VP1 expression plasmid showed clear nuclear localization of EGFP-BKV VP1 (Dumitriu and Moens, unpublished results). In A549 cells EGFP-BKV VP1 was not as restricted to the nucleus (**Figure 4.18** nuclear localization in A and throughout the cell in B) compared to transfection in Vero cells (**Figure 4.19** A, B, C). Transfection of Vero cells with EGFP-C1 has also been performed previously and the fluorescent signal is expected to be localized throughout the cell (Dumitriu and Moens, unpublished results). This was the result achieved when transfecting both Vero and A549 cells. The fluorescent signal from EGFP-KI VP1 was expected to be localized in the nucleus in accordance with BKPyV VP1. There was in general lower transfection efficiency in cells transfected with EGFP-KI VP1 than with the control plasmids. The fluorescence in both A549 and Vero cells was granulated and could be seen in the nucleus (**Figure 4.18** D, **Figure 4.19** I), the cytoplasm (**Figure 4.19** D, E, G and H) or both (**Figure 4.18** C). In the A549 cells a few cells where the signal was mostly in the nucleus was observed (**Figure 4.18** D). However in the Vero cells most signals were cytoplasm localized (**Figure 4.19** D, E, G and H) and not in the nucleus. When the signal was in the nucleus the signal seemed more condensed than for the EGFP-BKV VP1 nucleus signals (**Figure 4.19** F and I).

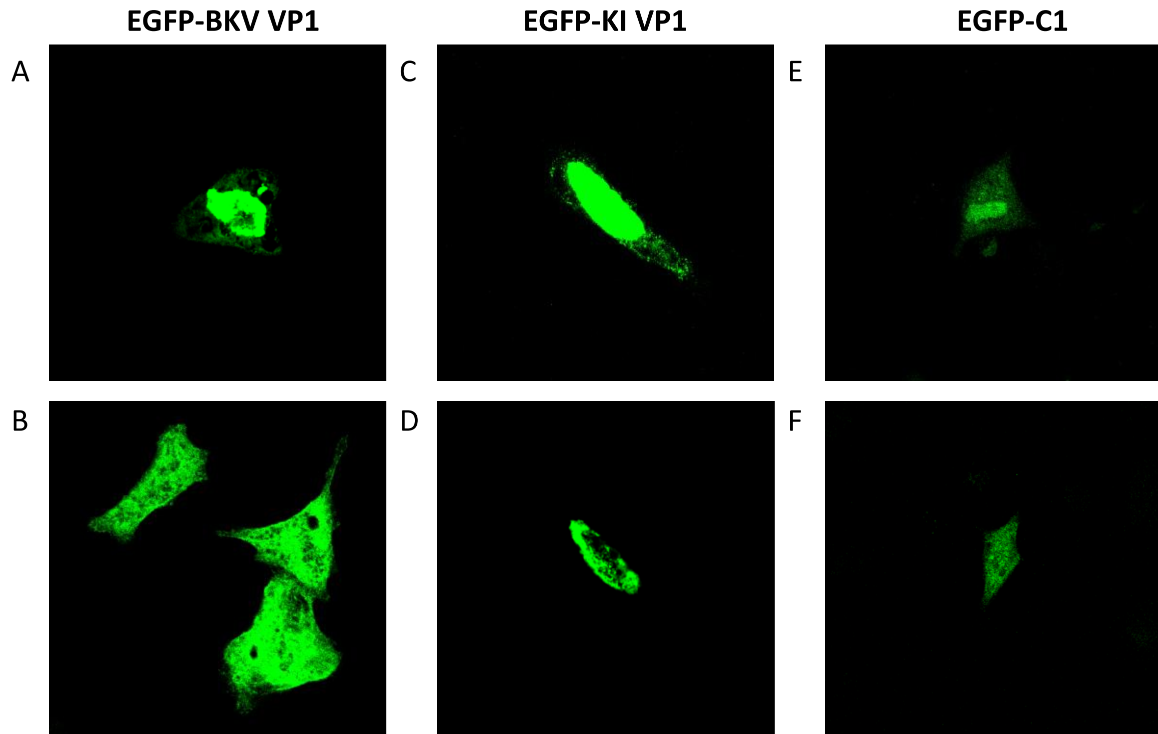


Figure 4.18: Confocal microscopy results from transfection of A549 cells with EGFP-BKV(A and B), EGFP-KI VP1(C and D) and EGFP-C1(E and F).

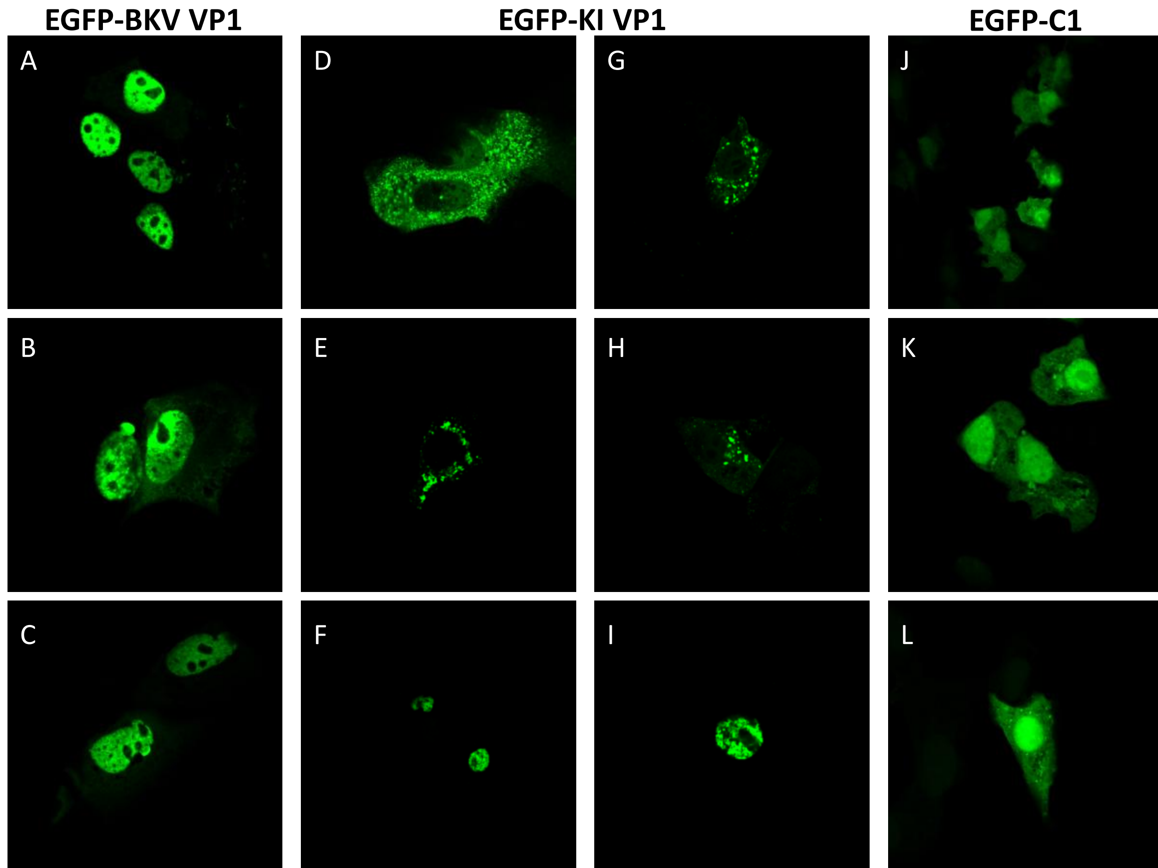


Figure 4.19: Confocal microscopy results from transfection of Vero cells with EGFP-BKV (A- C), EGFP-KI VP1 (D-I) and EGFP-C1 (J-L).

4.3 Detection of KIPyV DNA in urine and CSF

To examine whether KIPyV reactivation is associated with immunodeficient conditions or neurological diseases, urine specimens from SLE patients and CSF samples from patients with suspicion of neurological diseases were screened by nested PCR for the presence of KIPyV DNA. Viral DNA was isolated from 73 urine and 64 CSF samples and APRT PCR was run to test if there was human DNA in the samples. The expected length of the APRT PCR product was 720 bp, and roughly 1/3 of the CSF samples did not produce a product of this size (**Figure 4.20** and **4.21**). Of the urine samples 84 % of 69 tested samples were positive for human DNA (Moens, personal communication).

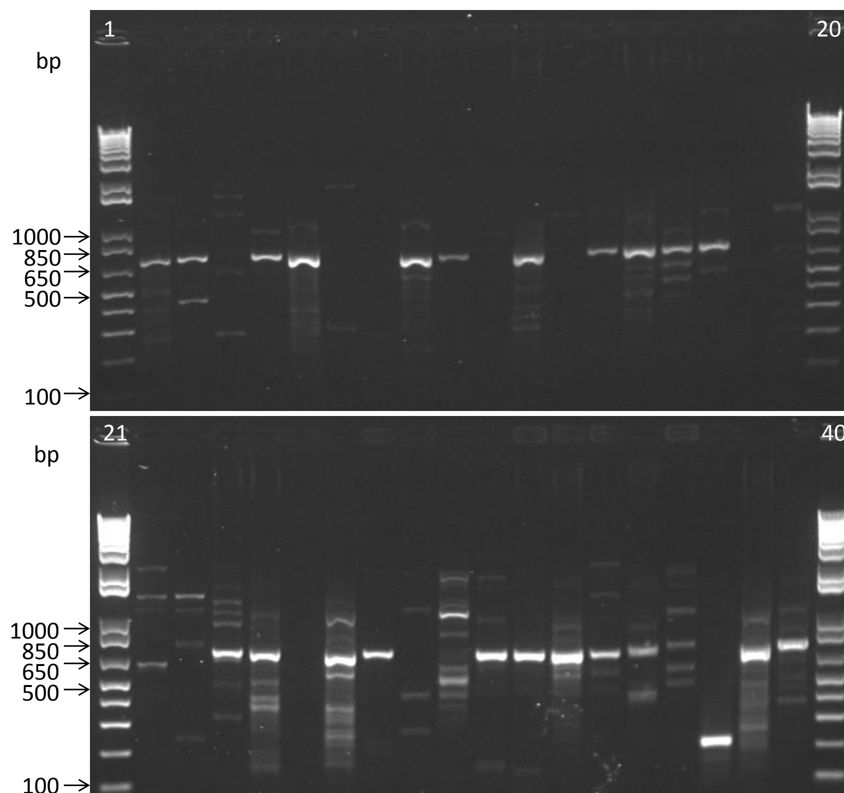


Figure 4.20: Agarose gel electrophoresis of products from APRT PCR of CSF. The PCR was run with halved reaction volume and 1 µl template. Well 1, 20, 21 and 40 contain 1 Kb Plus DNA ladder; the remaining wells contains PCR product from CSF samples 1-36 in this order.

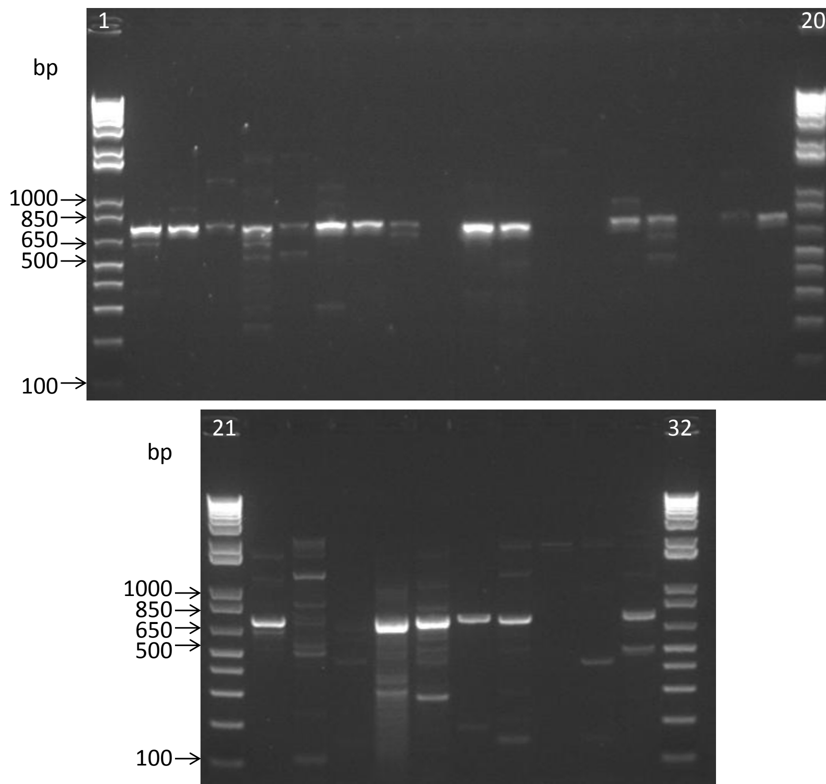


Figure 4.21: Agarose gel electrophoresis of products from APRT PCR of CSF. The PCR was run with halved reaction volume and 1 µl template. Well 1, 20, 21 and 32 contains 1 Kb Plus DNA ladder; the remaining wells contains PCR product from CSF samples 37-65 (no sample 56) in this order.

Nested KIPyV VP1 PCR was performed on all samples. Due to trouble with contamination, the PCR was repeated several times. The results from the first two rounds of PCR are enclosed in the appendix, **Supplementary Figure 7**. Only the last repeated PCR results of all previously positive samples and a selected few negative samples are pictured in **Figure 4.22** and **Figure 4.23**. No bands were seen in the negative water controls and the positive control gave a band of expected length. Six of the CSF samples were positive after nested KI VP1 PCR (9.4 %) and one sample had a very week band. Three of the urine samples were positive after nested PCR (4 %). To confirm these results the VP1 positive samples were sequenced and a new PCR using NCCR primers was performed. The NCCR PCR gave no positive PCR products for any of the tested samples (**Figure 4.24**). The negative control produced no bands and the positive control gave a band of the expected length. Sequencing revealed KIPyV VP1 sequences in all of the CSF and in two of the urine VP1 PCR positive samples. One urine sample did not provide any sequences and the other two urine samples had a very high background noise/sequence of unknown origin their sequences.

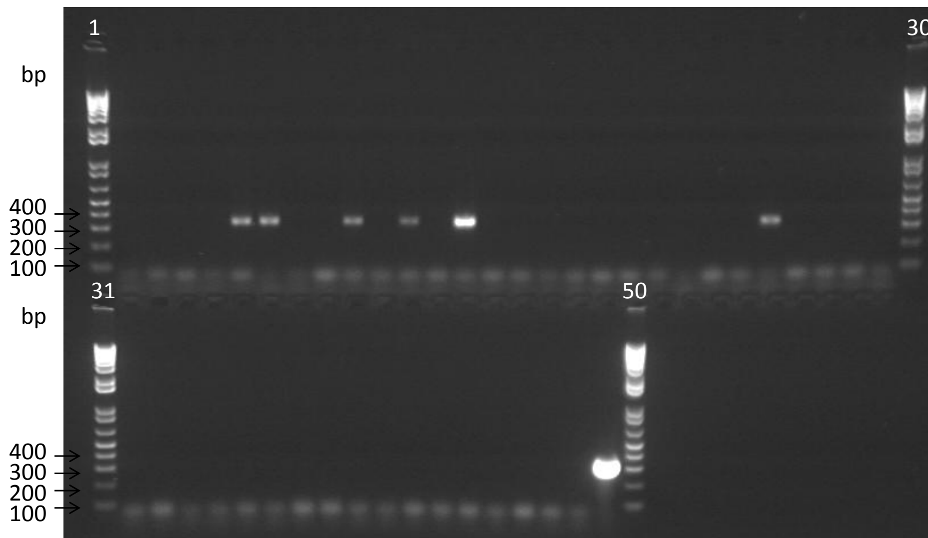


Figure 4.22: Agarose gel electrophoresis of products from VP1 PCR 1 of CSF and urine. The PCR was performed with 2 µl or 100 ng/µl template DNA. Well 1,30,31 and 50: 1 Kb Plus DNA ladder; wells 2-29: PCR products from CSF samples 3-9, 11-15, 18, 23, 30, 34, 37, 38, 43, 45, 46, 48, 52, 53, 55, 62, 63 and 65; wells 32-47: PCR products from urine samples 3, 4, 6, 12, 35, 36, 39, 40, 42-44, 49, 55, 59, 63 and 73; well 48: negative water control; well 49: positive KIPyV control. The previously completely negative samples are CSF sample 3 and 65, and urine samples 3, 4, 5 and 75.

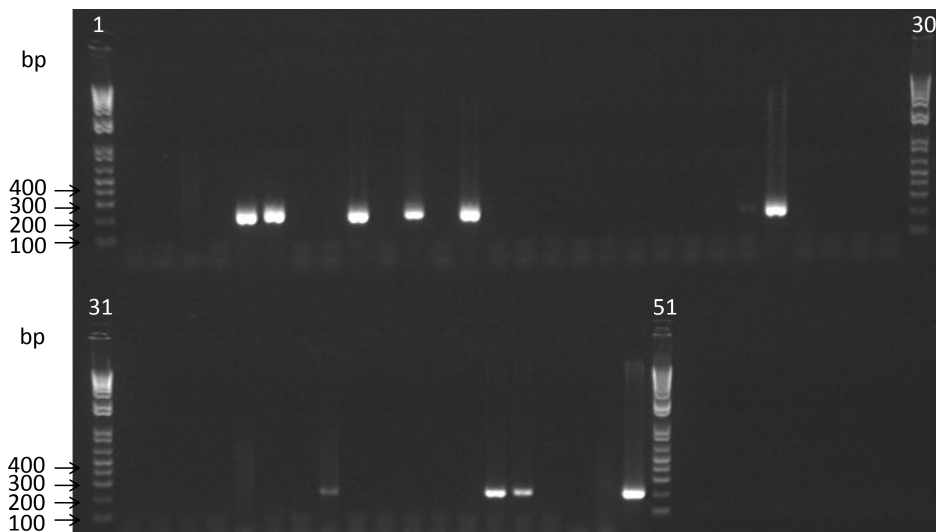


Figure 4.23: Agarose gel electrophoresis of products from nested VP1 PCR of CSF and urine. Well 1,30,31 and 51: 1 Kb Plus DNA ladder; wells 2-29: PCR 2 products from CSF samples 3-9, 11-15, 18, 23, 30, 34, 37, 38, 43, 45, 46, 48, 52, 53, 55, 62, 63 and 65; wells 32-47: PCR 2 products from urine samples 3, 4, 6, 12, 35, 36, 39, 40, 42-44, 49, 55, 59, 63 and 73; well 48 and 49: negative water control and well 50: positive control (KIPyV genome). The previously completely negative samples are CSF sample 3 and 65, and urine samples 3, 4, 5 and 75.

Figure 4.24: Agarose gel electrophoresis of products from NCCR PCR of positive CSF and urine samples. Well 1 and 14: 1 Kb Plus DNA ladder, well 2: CSF 7, well 3: CSF 8, well 4: CSF 12, well 5: CSF 14, well 6: CSF 18, well 7: CSF 52, well 8: CSF 53, well 9: urine 40, well 10: urine 59, well 11: urine 63, well 12: negative water control, well 13: positive KIPyV control.

5 Discussion

5.1 Study of the permissivity of A549 cell line to KIPyV DNA

The permissivity of A549 cell line to KIPyV was investigated by transfecting cells with KIPyV DNA. Viral replication was measured by analyzing the medium of transfected cells for the presence of KIPyV DNA by PCR, examination of total RNA of infected cells for LT-ag transcripts by RT-PCR, and monitoring for expression of VP1 protein by WB.

In medium from KIPyV DNA transfected A549 cells VP1 DNA was detected after 24 hours (**Figure 4.3**). However, KIPyV DNA was only detected in one sample by NCCR PCR, in the co-transfection with KIPyV DNA and CT-DNA after 24 hours (**Figure 4.4**). The positive control also produced a band of the expected length. The reason for why NCCR DNA could not be amplified in these samples might be because the PCR with NCCR is not as sensitive as the nested PCR. Alternatively, mutations in the NCCR during viral replication may prevent the primer(s) to bind. Indeed previous studies with other PyV have shown that the NCCR is prone for mutation after cell passages (59). The discovery of KIPyV DNA in the medium is not by its self a confirmation of viral replication as it is very likely that the DNA originates from transfected KIPyV DNA released by dead cells when taking into account the high number of dead cells after a transfection. Because other PyV have been reported to be present in KIPyV DNA positive samples (11), we reasoned that co-infection of the same host cells by BKPyV and KIPyV may offer advantages on viral replication. No differences in the strength of the PCR product bands when comparing transfections with or without BKPyV genome and the expression plasmid for the BKPyV agno protein were detected. This may indicate that co-presence of BKPyV does not support KIPyV replication, at least in A549 cells.

To further investigate if the positive VP1 PCR results in medium may be caused by viral particles shed after viral replication, isolation of total RNA from KIPyV DNA transfected A549 and HEK293 cells was performed. The RNA was reverse transcribed and the cDNA was used in PCR of LT-ag to see if this viral gene was transcribed in KIPyV transfected cells. LT-ag was detected in both A549 and HEK293 cells, however only the negative transfection in HEK293 cells was negative (**Figure 4.5**). Since the negative water control and the negative transfection from HEK293 cells did not display any bands after LT-ag PCR, the reason for the weak band from the A549 negative transfection is not caused by contamination during the PCR. New cDNA was made for the negative transfection in A549 cells was made with no change in results. The reason for why the negative transfection RNA

provided a PCR product with LT-ag primers might be that there was cross contamination of DNA or cells from the KIPyV DNA transfected well/flasks causing transcription of KIPyV genes, or by RNA from the KIPyV DNA transfected cells upon RNA isolation. The LT-ag RT-PCR results provide evidence for transcription of the early genes of KIPyV in both A549 and HEK293 cells.

To further examine if the transcripts were expressed WB of lysate from KIPyV DNA transfected cells was performed to identify VP1 protein. Two antibodies directed against KIPyV VP1 were tested for their reactivity against EGFP-KI VP1 fusion protein and the negative control EGFP-C1 protein. Before the KI VP1 antibodies were tested antibodies directed against EGFP was used to evaluate the expected mass of the EGFP-KI VP1 fusion protein.

The mass of the strongest band was ~70 kDa and is believed to correspond to the EGFP-KI VP1 fusion protein as the expected theoretical mass is 71.4kDa. The strong band in the sample transfected with EGFP-BKV VP1 DNA was a little bit smaller compared to that of EGFP-KI VP1 when looking at the WB result (**Figure 4.6**). This coheres with the smaller theoretical mass of EGFP-BKV VP1 with 70 kDa. In addition factors like amino acid charges or posttranslational modifications may affect the protein migration. Both the EGFP-KI VP1 and EGFP-BKV VP1 transfected samples have weaker bands that migrated a little bit shorter than the strong bands. This might be a result of degradation products or non-specific bands. The EGFP-KI VP1 bands are a lot weaker in the A549 cell lysate than the HEK293 cell lysate. This corresponds with the general notion that the transfection efficiency of EGFP-KI VP1 plasmid was higher in the HEK293 cell line than in the A549 cell line, although it might mean that the production of EGFP-KI VP1 fusion protein is lower in A549 cells or that more A549 cells died after transfection than HEK293 cells and thereby less cells were harvested.

Out of the two antibodies directed against KIPyV VP1 only the Polyclonal Rabbit EP101303 gave any bands at all in the WB analysis, hence this was now called KI VP1 antibody. In the primary testing with EGFP-KI VP1 and EGFP-C1 transfected cell lysate there was a band of between 60 and 80 kDa in the EGFP-KI VP1 sample. This was not present in the EGFP-C1 negative control sample which is corresponding with the expected mass of the fusion protein. In this sample there was also a weak band of ~40 kDa which was not present in the EGFP-C1 control. This could be a result of degraded fusion protein only including VP1. The KI VP1 antibody was also tested on purified GST-KI VP1 fusion protein where a band of approximately expected theoretical mass (67.6 kDa) was observed (**Figure 4.8**).

5.1.1 Transfection of A549 cell line with KIPyV DNA

To investigate whether KIPyV DNA transfection of A549 cells could support viral protein production the KI VP1 antibody was tested on lysates from these cells. Cells were harvested after 48 hours and a week supported by the fact that BKPyV proteins could be detected in viral plaques in cell cultures after one to two days after infection (21). The WB analysis shown in **Figure 4.9** includes controls such as the previously analyzed EGFP-KI VP1 transfected HEK293 cell lysate and the newly EGFP-KI VP1 transfected A549 cell lysate. The EGFP-KI VP1 fusion protein might be detectable in the HEK293 cell lysate with a band of less than ~60kDa or between 60 and 80 kDa, which suggests that the antibodies are working and the WB analysis was successful. However, the A549 cell lysate showed no sign of any such band; neither did the KIPyV genome transfected cells show signs of ~40 kDa VP1 proteins. A WB with anti-EGFP would have let us know if the failure to detect EGFP-KI VP1 protein was due to problems with the KI VP1 antibody and if the fusion protein was present. The transfection efficiency with EGFP-KI VP1 based on the number of green fluorescence emitting cells was low (<50 %), which might result in an amount of expressed EGFP-KI VP1 protein under our level of detection.

A549 cells were repeatedly transfected with KIPyV genome and control EGFP fusion proteins, in addition to being co-transfected with BKPyV genome and with an expression plasmid for BKPyV agno protein. The WB results from these experiments show no sign of the EGFP-KI VP1 fusion protein while the antibodies do detect the GTS-KI VP1 fusion protein (**Figure 4.10**, **4.11** and **4.13**). This means that the WB processes were successful and might indicate that the transfections did not have the same success. When taken in consideration the very weak EGFP-KI VP1 bands the A549 cells produced in the WB with antibodies against the EGFP protein (**Figure 4.6**) it might not be so strange that they were not detected in subsequent transfections since all these results are very weak too. The high background noise in the pictures also makes it very difficult to separate a ~70 kDa specific band. In addition, no specific band of 41.2 kDa was seen only for the KIPyV genome transfected cells or the co-transfections. However, a non-specific band of ~ 40 kDa was observed throughout the samples and stronger bands were observed in the 1 week samples. Because of strange results in **Figure 4.11** the membrane was stripped and antibodies against ERK 2 were used. The results presented in **Figure 4.12** showed a generally similar protein distribution, however a few samples were more irregular. This might be because fewer cells survived the transfection process and were harvested while the sample buffer used to harvest the cells were the same

for all samples. However, these results did not explain the results in **Figure 4.11**, and a repetition of the WB produced improved results (**Figure 4.13**).

An attempt to infect fresh A549 cells with medium from previous transfections with KIPyV DNA was not successful. A positive control was lacking in this experiment but since the results shown in **Figure 4.14** resemble previous transfections it is not suspected that there was anything wrong with the primary antibodies. Indeed there was not any reason to expect results from an attempted infection with media from previous transfections to be successful when the primary transfections did not produce detectable VP1 protein. This was however attempted before the WB analyses of the primary transfection cell lysates were concluded. There was no viral production after transfection of KIPyV DNA or addition of media from KIPyV DNA transfected A549 cells to fresh A549 cells. To ensure that the KI VP1 antibodies were in fact working a new WB including the GST-KI VP1 fusion protein should have been performed; however this was not done because of time limitations.

5.1.2 Transfection of HEK293 cell line with KIPyV DNA

HEK293 cells were transfected with KIPyV DNA to see if these cells could support viral protein production. The KI VP1 antibody was tested on lysates from cells transfected with KIPyV genome and control EGFP fusion proteins. KIPyV was also co-transfected with KIPyV and BKPyV genomes, or KIPyV and an expression plasmid for BKPyV agno protein. GST-KI VP1 fusion protein was detected by the VP1 antibody and the EGFP-KI VP1 control appears like a thin line with bands being unrecognizable in the WB in **Figure 4.15**. This means that the WB is working as it should, but that there is something wrong with the EGFP control. The reason for this result could if only observed once, be caused by remaining gel in the well after removal of the comb; however this was observed with this sample in all WB after this analysis (Results not included). Another reason could be poor resuspension of the protein pellet. All cell lysates gave a non-specific band of ~40 kDa which was not present in previous analysis (**Figure 4.9**), this might however be due to very weak bands. Due to the non-specific bands it is impossible to evaluate the effect of transfections with KIPyV and co-transfections before a transfection with the expression plasmid pcDNA3.1(+)KI VP1 could be performed to evaluate if the 40 kDa signal is stronger when KIPyV VP1 is expressed. Two bands of ~70 kDa were detected in the EGFP-KI VP1 transfected cell lysate and the KIPyV and BKPyV co-transfection lysate. This is the expected size of EGFP-KIVP1, but the reason why this band appears in the other sample is not known. One possible source of error could be

transfection with the wrong plasmid. This sample was made from the transfection mix together with the 1 week sample which did not present this band. This could again be explained by death of the transfected cells after 1 week, and that the lysate therefore does not originate from transfected cells. Another possibility might be a mix up with other samples, but because only one well of the set up contained the EGFP-KI VP1 transfection and this already is counted for, there is no other sample to switch it with.

Medium from the previous transfections was added to fresh HEK293 cells to see if infectious virus was produced. A positive control to confirm that the primary antibody is working is lacking but as the WB results look like all previous transfected lysates the KI VP1 antibody is probably working as normal. Due to time limitations the WB was not repeated. Some of the samples had a stronger band of ~40 kDa than the others. This could indicate that infectious virus was produced, however, because the negative control also displayed this stronger band it is not likely.

5.1.3 Expression of KIPyV VP1 in mammalian cells

The mammalian expression vector pcDNA3.1$^{(+)}$KI VP1 was used to see if HEK293 cells could express the VP1 protein without the whole KIPyV genome. VP1 was not detected in any of the wells containing transfected cell lysate including the EGFP-KI VP1 transfected control, but was detected the GST-KI VP1 protein as normal. This means that the WB process was successful but that the transfection was not. The experiment was repeated in HEK293 and in addition performed in A375 cells without any change in results. The reason why the EGFP-KI VP1 transfected cells do not produce detectable VP1 protein could be that there is something wrong with the plasmid, either too low concentration or perhaps it has been degraded by DNases. However, green fluorescent cells were observed the following day after transfection which suggests that the plasmid is not degraded. Another possible source of error is the KI VP1 antibody. Because GST-KI VP1 protein was detected successfully the antibody is working. However it is possible that the KI VP1 antibody is not able to detect low levels of protein. Because the EGFP-KI VP1 fusion protein was not detected and the transfection result was uncertain we do not know if the pcDNA3.1$^{(+)}$KI VP1 is able to produce VP1 protein. The results from this transfection could have been used to evaluate if the expression of VP1 would increase the strength of the ~40 kDa non-specific band obtained after WB of transfected cell lysates. Since this did not work another possibility could have been to cleave off the GST-tag

of the GST-KI VP1 fusion protein to see if the VP1 protein does in fact migrate to the same place as the non-specific band.

5.2 Study of the sub-cellular localization of KIPyV VP1 protein

Newly synthesized capsid proteins are transported from the cytoplasm to the nucleus upon HPyV maturation (2). The VP1 of HPyV have nuclear localization signals in the N-terminal region that enable this transportation (2) and KIPyV VP1 protein also seems to have a NLS like motif. To investigate the sub cellular localization of VP1 protein A549 and Vero cells were transfected with EGFP-KI VP1 and studied by confocal microscopy. For comparative reasons cells were also transfected with EGFP-BKV VP1 and EGFP-C1 which were expected to localize in the nucleus and throughout the cell, respectively. This was also observed in this study; however the EGFP-BKV VP1 signals were not as restricted to the nucleus when transfected in A549 cells as in Vero cells. The fluorescent signal from EGFP-KI VP1 protein was expected to localize in the nucleus as with BKV VP1, but did in fact localize in the cytoplasm and throughout the cells as well. A peculiar feature with the EGFP-KI VP1 signals compared to the controls was that the signals were granulated and in Vero cells the nuclear signals seemed condensed than in A549 cells. Nuclear signal localization was mostly observed in the A549 cells, however, in Vero cells most signals were cytoplasm localized and not in the nucleus. Due to the low cell number and the poor transfection efficiency any conclusions about the localization of KIPyV VP1 is uncertain. PyV VP2 and VP3 proteins also contain NLS motifs and they can interact with VP1 (23, 24). It is therefore possible that KIPyV VP2/3 aid with nuclear localization of VP1. This could be tested by studying nuclear localization of VP1 in cells co-transfected with plasmids encoding VP2 or/and VP3.

5.3 Detection of KIPyV DNA in urine and CSF

Little is known about the genuine host cells of KIPyV and whether KIPyV is associated with a specific disease (33, 34). To examine whether KIPyV reactivation is associated with immunodeficient conditions or neurological diseases, 73 urine specimens from patients with SLE and 64 CSF samples from patients with suspicion of neurological diseases were screened for the presence of KIPyV DNA. Viral DNA was isolated and nested KI VP1 PCR was performed on all samples. Of the CSF samples 9.4 % were positive and one sample was weakly positive and 4 % of the urine samples were positive. The weak band could have been

caused by transfer from the neighboring positive sample. These results have been confirmed by sequencing but an attempt of NCCR PCR was not successful. Because the negative control was negative after all PCR performed there is not a contamination in any of the reagents used for PCR. The positive results could be a result of contamination from one or a few of the KIPyV DNA containing samples during DNA isolation. Another source of contamination might be PCR product; however in the very first PCR with the outer VP1 primers performed several samples were already positive. It is not believed that the positive findings are a result of contaminating KIPyV DNA from plasmids or purified genome as these were not handled in the same local as the patient samples. KIPyV DNA has only been detected in the urine of 1 out of 50 renal transplant patients (**Supplementary Table 1**), and not in healthy individuals or other patients. Because we found KIPyV DNA in another group of immunocompromised patients (SLE), KIPyV viruria may occur in patients with immunodeficiencies. So far, KIPyV DNA has not been reported in CSF (**Supplementary Table 1**). Our study is the first to demonstrate the presence of KIPyV in CSF. Unfortunately, we do not know the medical history of our patients, but findings urge us to test more CSF samples.

NCCR PCR gave no positive PCR products for any of the tested samples. Both the negative and positive control behaved as expected. One reason could be that this PCR is less sensitive. Another reason for lacking any positive results after NCCR PCR might be because this region is more disposed to mutations than the protein coding sequenced as mentioned earlier. Sequencing confirmed KIPyV VP1 sequences in all the CSF and two of the urine VP1 PCR positive samples. For one urine samples no sequences were obtained and for the other two urine samples the sequences had a very high background noise/sequence of unknown origin. The lacking sequencing results and the samples with high background might be caused by too little template DNA in the sequencing reaction and the lacking sequences might be caused by errors in the sequencing analysis for instance failure in the capillary injections or not good enough suspension of samples in loading buffer.

6 Conclusion and future perspective

The aims of this study were to investigate the permissivity of A549 cell line to KIPyV DNA and simultaneously test antibodies directed against the KIPyV VP1 protein, to investigate the sub-cellular localization of this protein, and finally to examine if KIPyV DNA was detectable in CSF and urine specimens from patients with neurological disorders and immunocompromised patients, respectively.

Based on our RT-PCR results we have indicated that there is viral activity in A549 cells after transfection with KIPyV genome as we could detect LT-ag transcripts. We also detected KIPyV DNA in medium of transfected cells that may originate from viral replication and shedding of DNA including viral particles. However, due to the failure to detect VP1 protein the conclusion about whether the A549 cell line is permissive to KIPyV is unambiguous. The antibody used for detection of VP1 protein displayed non-specific binding approximately at the same area as where we would expect to find the VP1 protein. We were successful in producing the GST-KI VP1 fusion protein and transfecting cells with EGFP-KI VP1 plasmid providing us with the expression of EGFP-KI VP1 fusion protein, both which we were able to detect by the use of the KI VP1 antibody. To confirm these results a new antibody directed against KIPyV VP1 should be tried out. In addition electron microscopy studies can be performed to find out if virus particles can be detected in the cells transfected with KIPyV DNA or in the medium of these cells.

We were able to study the sub-cellular localization of VP1 protein through transfection with EGFP-KI VP1 fusion protein which we found to be localized mostly in the nucleus in A549 cells. In Vero cells the localization was divided between the cytoplasm and nucleus, although mostly in the cytoplasm. To confirm the sub-cellular localization a transfection with an expression vector for KIPyV VP1 such as pcDNA3.1$^{(+)}$ followed by an immunofluorescent labeling using VP1 specific antibodies could be performed, or immunohistochemistry on KIPyV infected cells if a permissive cell line is available. The functionality of the NLS signal in the KIPyV genome can be tested by creating mutations in the NLS using site-directed mutagenesis.

We have indicated the presence of KIPyV DNA in CSF specimens from patients with suspicion of neurological disorders and urine of immunocompromised patients with SLE. The results were confirmed by sequencing but a secondary PCR of the NCCR region was not successful. These results need to be confirmed by another method such as qPCR, preferably with primers directed against another region of the genome. The genomes or part of the

genomes (e.g. VP1 or NCCR) of strains present in urine and CSF should be sequenced to determine whether specific strains are circulating in these patient groups. Since an oral route of infection has also been suggested and KIPyV DNA can be detected in sewage, cells from the gastrointestinal tract could be tested for their permissivity for KIPyV.

References

1. Eddy BE, Stewart SE, Berkeley W. Cytopathogenicity in tissue culture by a tumor virus from mice. Proc Soc Exp Biol Med. 1958 Aug-Sep;98(4):848-51.
2. Knipe DM, Howley PM, Fields BN. Fields virology. 5th ed ed. Philadelphia: Wolters kluwer/Lippincott Williams & Wilkins; 2007. p: 2263-2298.
3. Gardner SD, Field AM, Coleman DV, Hulme B. New human papovavirus (B.K.) isolated from urine after renal transplantation. Lancet. 1971 Jun 19;1(7712):1253-7.
4. Padgett BL, Walker DL, ZuRhein GM, Eckroade RJ, Dessel BH. Cultivation of papova-like virus from human brain with progressive multifocal leucoencephalopathy. Lancet. 1971 Jun 19;1(7712):1257-60.
5. Allander T, Andreasson K, Gupta S, Bjerkner A, Bogdanovic G, Persson MA, et al. Identification of a third human polyomavirus. J Virol. 2007 Apr;81(8):4130-6.
6. Gaynor AM, Nissen MD, Whiley DM, Mackay IM, Lambert SB, Wu G, et al. Identification of a novel polyomavirus from patients with acute respiratory tract infections. PLoS Pathog. 2007 May 4;3(5):e64.
7. Feng H, Shuda M, Chang Y, Moore PS. Clonal integration of a polyomavirus in human Merkel cell carcinoma. Science. 2008 Feb 22;319(5866):1096-100.
8. Schowalter RM, Pastrana DV, Pumphrey KA, Moyer AL, Buck CB. Merkel cell polyomavirus and two previously unknown polyomaviruses are chronically shed from human skin. Cell Host Microbe. 2010 Jun 25;7(6):509-15.
9. Van der Meijden E, Janssens RW, Lauber C, Bouwes Bavinck JN, Gorbalenya AE, Feltkamp MC. Discovery of a new human polyomavirus associated with trichodysplasia spinulosa in an immunocompromized patient. PLoS Pathog. 2010;6(7):e1001024.
10. Scuda N, Hofmann J, Calvignac-Spencer S, Ruprecht K, Liman P, Kühn J, et al. A novel human polyomavirus closely related to the African green monkey derived lymphotropic polyomavirus (LPV). Journal of virology. 2011.
11. Moens U, Johannessen M, Bárcena-Panero A, Gerits N, Ghelue MV. Emerging polyomaviruses in the human population. Reviews of infection. 2010:59-93.
12. Moens U. Silencing viral microRNA as a novel antiviral therapy? J Biomed Biotechnol. 2009;2009:419539.
13. Moens U, Ludvigsen M, Van Ghelue M. Human polyomaviruses in skin diseases. Patholog Res Int. 2011;2011:123491.
14. Moens U, Van Ghelue M, Johannessen M. Oncogenic potentials of the human polyomavirus regulatory proteins. Cell Mol Life Sci. 2007 Jul;64(13):1656-78.
15. Shuda M, Feng H, Kwun HJ, Rosen ST, Gjoerup O, Moore PS, et al. T antigen mutations are a human tumor-specific signature for Merkel cell polyomavirus. Proc Natl Acad Sci U S A. 2008 Oct 21;105(42):16272-7.
16. Imperiale MJ. The human polyomaviruses: An overview. In: Khalili K, Stoner GL, editors. Human polyomaviruses, Molecular and clinical perspectives. New York: Wiley-Liss, Inc; 2001. p. 53-71.
17. Daniels R, Sadowicz D, Hebert DN. A very late viral protein triggers the lytic release of SV40. PLoS Pathog. 2007 Jul;3(7):e98.
18. Khalili K, White MK, Sawa H, Nagashima K, Safak M. The agnoprotein of polyomaviruses: a multifunctional auxiliary protein. J Cell Physiol. 2005 Jul;204(1):1-7.
19. Suzuki T, Orba Y, Okada Y, Sunden Y, Kimura T, Tanaka S, et al. The human polyoma JC virus agnoprotein acts as a viroporin. PLoS Pathog. 2010 Mar;6(3):e1000801.

20. Del Valle L, White MK, Enam S, Pina Oviedo S, Bromer MQ, Thomas RM, et al. Detection of JC virus DNA sequences and expression of viral T antigen and agnoprotein in esophageal carcinoma. Cancer. 2005 Feb 1;103(3):516-27.
21. Ahsan N, Shah KV. Polyomaviruses and Human Diseases. In: Ahsan N, editor. Polyomaviruses and Human Diseases. New York: Eurekah.com and Springer Science+Business Media; 2006. p. 1-18.
22. Neu U, Wang J, Macejak D, Garcea RL, Stehle T. Structures of the major capsid proteins of the human Karolinska Institutet and Washington University polyomaviruses. J Virol. 2011 Jul;85(14):7384-92.
23. Moens U, Rekvig OP. Molecular biology of BK virus and clinical and basic aspects of BK virus renal infection. In: Khalili K, Stoner GL, editors. Human Polyomavirus: Molecular and clinical perspectives: Wiley-Liss, Inc; 2001. p. 359-408.
24. Garcia-Bustos J, Heitman J, Hall MN. Nuclear protein localization. Biochim Biophys Acta. 1991 Mar 7;1071(1):83-101.
25. Shishido-Hara Y, Nagashima K. Synthesis and assembly of polyomavirus virions. In: Khalili K, Stoner GL, editors. human polyomaviruses: Molecular and clinical Perspectives. New York: Wiley-Liss, Inc; 2001. p. 149-77.
26. Barzon L, Squarzon L, Pacenti M, Scotton PG, Palu G. Detection of WU polyomavirus in cerebrospinal fluid specimen from a patient with AIDS and suspected progressive multifocal leukoencephalopathy. J Infect Dis. 2009 Jul 15;200(2):314-5.
27. Tan BH, Busam KJ. Virus-associated Trichodysplasia Spinulosa. Adv Anat Pathol. 2011 Nov;18(6):450-3.
28. Kean JM, Rao S, Wang M, Garcea RL. Seroepidemiology of human polyomaviruses. PLoS Pathog. 2009 Mar;5(3):e1000363.
29. Nguyen NL, Le BM, Wang D. Serologic evidence of frequent human infection with WU and KI polyomaviruses. Emerg Infect Dis. 2009 Aug;15(8):1199-205.
30. Carter JJ, Paulson KG, Wipf GC, Miranda D, Madeleine MM, Johnson LG, et al. Association of Merkel cell polyomavirus-specific antibodies with Merkel cell carcinoma. J Natl Cancer Inst. 2009 Nov 4;101(21):1510-22.
31. Chen T, Mattila PS, Jartti T, Ruuskanen O, Soderlund-Venermo M, Hedman K. Seroepidemiology of the newly found trichodysplasia spinulosa-associated polyomavirus. J Infect Dis. 2011 Nov;204(10):1523-6.
32. Doerries K. Human polyomavirus JC and BK persistent infection. In: Ahsan N, editor. Polyomaviruses and human disease. New York: Eurekah.com and Springer Science+Business Media; 2001. p. 102-16.
33. Babakir-Mina M, Ciccozzi M, Perno CF, Ciotti M. The novel KI, WU, MC polyomaviruses: possible human pathogens? New Microbiol. 2011 Jan;34(1):1-8.
34. Dalianis T, Ramqvist T, Andreasson K, Kean JM, Garcea RL. KI, WU and Merkel cell polyomaviruses: a new era for human polyomavirus research. Semin Cancer Biol. 2009 Aug;19(4):270-5.
35. Babakir-Mina M, Ciccozzi M, Bonifacio D, Bergallo M, Costa C, Cavallo R, et al. Identification of the novel KI and WU polyomaviruses in human tonsils. J Clin Virol. 2009 Sep;46(1):75-9.
36. Bergallo M, Terlizzi ME, Astegiano S, Ciotti M, Babakir-Mina M, Perno CF, et al. Real time PCR TaqMan assays for detection of polyomaviruses KIV and WUV in clinical samples. J Virol Methods. 2009 Dec;162(1-2):69-74.
37. Babakir-Mina M, Ciccozzi M, Alteri C, Polchi P, Picardi A, Greco F, et al. Excretion of the novel polyomaviruses KI and WU in the stool of patients with hematological disorders. J Med Virol. 2009 Sep;81(9):1668-73.

38. Bialasiewicz S, Whiley DM, Lambert SB, Nissen MD, Sloots TP. Detection of BK, JC, WU, or KI polyomaviruses in faecal, urine, blood, cerebrospinal fluid and respiratory samples. J Clin Virol. 2009 Jul;45(3):249-54.
39. Kantola K, Sadeghi M, Lahtinen A, Koskenvuo M, Aaltonen LM, Mottonen M, et al. Merkel cell polyomavirus DNA in tumor-free tonsillar tissues and upper respiratory tract samples: implications for respiratory transmission and latency. J Clin Virol. 2009 Aug;45(4):292-5.
40. Bofill-Mas S, Rodriguez-Manzano J, Calgua B, Carratala A, Girones R. Newly described human polyomaviruses Merkel cell, KI and WU are present in urban sewage and may represent potential environmental contaminants. Virol J. 2010;7:141.
41. Rinaldo CH, Traavik T, Hey A. The agnogene of the human polyomavirus BK is expressed. J Virol. 1998;72(7):6233-6.
42. Macherey-Nagel. User Manual: Nucleobond®Xtra Midi kit. 2011 [16.11.11]; Available from: http://www.mn-net.com/Portals/8/attachments/Redakteure_Bio/Protocols/Plasmid%20DNA%20Purification/UM_pDNA_NuBoXtra.pdf.
43. Macherey-Nagel. User Manual: NucleoSpin®Plasmid kit 2010 [16.11.11]; Available from: http://www.mn-net.com/Portals/8/attachments/Redakteure_Bio/Protocols/Plasmid%20DNA%20Purification/UM_pDNA_NS.pdf.
44. Macherey-Nagel. User Manual: Nucleospin®RNA II kit 2011 [16.11.11]; Available from: http://www.mn-net.com/Portals/8/attachments/Redakteure_Bio/Protocols/RNA%20and%20mRNA/UM_TotalRNA.pdf.
45. Qiagen. Handbook: QIAamp® MinElute® Virus Spin kit 2010 [16.11.11]; Available from: http://www.qiagen.com/products/rnastabilizationpurification/cellviralrnapurificationsystems/qiaampminelutevirusspin.aspx#Tabs=t2.
46. GEHealthcare. Product booklet: illustra GFX™ PCR DNA and Gel band purification kit. 2008 [16.11.11]; Available from: http://www.gelifesciences.com/aptrix/upp00919.nsf/Content/7D3C39CAF8206AD1C1257628001D5012/$file/28951562AA.pdf.
47. Sambrook J, Russel DW. Molecular Cloning: A laboratory manual. 3 ed. New York: Cold Spring Harbor Laboratory Press; 2001. p: 9.62, 9.92, A 4.37, A.8.40, A9.28.
48. Bitesizebio. Quick reference: Determining DNA Concentration & Purity. 2007 [03.10.11]; Available from: http://bitesizebio.com/articles/dna-concentration-purity/.
49. ThermoFisherScientificInc. User Manual: NanoDrop 1000 Spectrophotometer V3.7. 2008 [03.10.11]; Available from: http://nanodrop.com/Library/nd-1000-v3.7-users-manual-8.5x11.pdf.
50. Primrose SB, Twyman RM. Principles of gene manipulation and genomics. 7th ed. Malden, Mass.: Blackwell; 2006. p: 9-24, 26-30, 218-228.
51. Reece RJ. Analysis of genes and genomes. Chichester, West Sussex, England ; Hoboken, NJ: John Wiley & Sons; 2004. p: 89-95, 300, 301.
52. Wilson K, Walker JM. Principles and techniques of biochemistry and molecular biology. 6th ed. Cambridge ; New York: Cambridge University Press; 2005. p: 84-90, 207-210, 217-219, 252, 235-239.
53. Affymetrix®Inc, USB®Products. Product information: ExoSAP-IT® For PCR Product Clean-Up [03.10.11]; Available from: http://www.affymetrix.com/estore/browse/brand/usb/product.jsp?navMode=34000&productId=131310&navAction=jump&aId=productsNav#1_2.

54. Pawley JB, editor. Handbook of biological confocal microscopy. 3rd ed. New York, NY: Springer; 2006. p: 369.
55. Freshney RI. Culture of animal cells : a manual of basic technique. 5th ed. Hoboken, N.J.: Wiley-Liss; 2005. p: 56.
56. Pickett CB, Lu AY. Glutathione S-transferases: gene structure, regulation, and biological function. Annu Rev Biochem. 1989;58:743-64.
57. Burgess RR, Deutscher MP, editors. Guide to protein purification. 2 ed. Amsterdam: Elsevier Inc; 2009. p: 246, 247.
58. Bio-Rad. User manual: Bio-Rad Protein Assay. [03.10.11]; Available from: http://www3.bio-rad.com/LifeScience/pdf/Bulletin_9004.pdf.
59. Yogo Y, Hara K, Guo J, Taguchi F, Nagashima K, Akatani K, et al. DNA-sequence rearrangement required for the adaptation of JC polyomavirus to growth in a human neuroblastoma cell line (IMR-32). Virology. 1993 Dec;197(2):793-5.

Appendix

Supplementary Figure 1: VP2/VP3 protein sequence

Supplementary Figure 2: VP1 protein sequence

Supplementary Figure 3: Primer sites in the respective genomes

Supplementary Figure 4: EGFP-KI VP1 expression plasmid

Supplementary Figure 5: GST-KI VP1 expression plasmid

Supplementary Figure 6: pcDN3.1$^{(+)}$-KI VP1 expression plasmid

Supplementary Figure 7: Sequencing results of positive VP1 PCR products

Supplementary Table 1: Prevalence of KIPyV

Supplementary Figure 1: VP2/3 protein sequence

```
SV40      MGAALTLLGDLIATVSEAAAATGFS--VAEIAAGEAAAAIEVQLASVATVEGLT-----T 53
BKPyV     MGAALALLGDLVASVSEAAAATGFS--VAEIAAGEAAAAIEVQIASLATVEGITS----T 54
JCPyV     MGAALALLGDLVATVSEAAAATGFS--VAEIAAGEAAATIEVEIASLATVEGITS----T 54
KIPyV     MGIFLAVPEIIAASIAGGAEALSIAGSGAAIATGEGLAALGGITEGAALLGET---IPIS 57
WUPyV     MGILLAVPEIIAASVAGGAEALSIAGSGAAIATGEGLAALGGLTESAALLGET---VEIS 57
MCPyV     MGGIITLLANIGEIATELSATTGVT--LEAILTGEALALEADISSLMTIEGISG----- 53
HPyV6     MGILFSLPEIIAAAAVGGGEALEIAGGLGALVSGEGLATLEALQSAAALSSEATAALAVS 60
HPyV7     MGILLSIPEIIAATAIGGGEALEIAGSVGALVSGEGLATLEALQSAAALSTEATAALAVS 60
TSPyV     MGGVISLFFDIIEISTELSAATGFA--VDAVLAGEAAAAVEVEVMGLMTVE-----GLSA 53
HPyV9     MGGVLTILLDIFELAEQLSLSTGFT--VDAILSGEAIAAATTEAAWLIEIEAVDVAGLGA 58
LPyV      MGGVLSLLFNISEIAAELSLSTGFT--LDAILTGEAFAAVSTEAAWLIEIEAVDLAGLST 58
          **  :::   :        . :   .:      : :**. *:

SV40      SEAIAAIGLTPQAYAVISGAPAAIA---GFAALLQTVTGVSAVAQVGYRFFSDWDHKVST 110
BKPyV     SEAIAAIGLTPQTYAVIAGAPGAIA---GFAALIQTVSGISSLAQVGYKFFDDWDHKVST 111
JCPyV     SEAIAAIGLTPETYAVITGAPGAVA---GFAALVQTVTGGSAIAQLGYRFFADWDHKVST 111
KIPyV     EAATTVLTKVPELVQATQAVTAAVQGGAGLVGGIYTALASDHPGDLPPNTPTGSASGLHP 117
WUPyV     EAAATVLTKVPELVTVTQGVTAAVQGGAGLVGGIYTALAADRPGDLPASTPTGSPSGLHP 117
MCPyV     IEALAQLGFTAEQFSNFSLVASLVNQGLTYGFILQTVSGIGSLITVGVRLSREQVSLVNR 113
HPyV6     NEAAIVLSTVPELSQTLFGAQLLLSSVAGVGGVIYSNYNPGELYKAPE-----GPGGLGP 115
HPyV7     NEAAIVLSTIPELSQTLFGVQTLLSSVAGVGGVVY-NLNPGELYQAPE-----GPGGLGP 114
TSPyV     SEALGTLGLTMENFSLMHALPGMLSEAVGIGTLFQTISGASGLVAAGIR-YGYAREVSIV 112
HPyV9     LEALTLTGLSAEEFSLLSALPTALNNAIGIGIFFQTVTGASAVVAAGVTTFGYSKEVPLV 118
LPyV      LEALSLTGLTTEQFSLLSAIPTALNNAIGIGVFFQTVSGASAVVAAGLTTFGYSKQVPVV 118
            *         :               :         .        .

SV40      VGLYQQPGMAVDLYRPDD-----------------YYDILFPGVQTFVHSVQY--LDPRH 151
BKPyV     VGLYQQSGMALELFNPDE-----------------YYDILFPGVNTFVNNIQY--LDPRH 152
JCPyV     VGLFQQPAMALQLFNPED-----------------YYDILFPGVNAFVNNIHY--LDPRH 152
KIPyV     TSGYNPQGAGLNLQSVHKPIHAPYSGMALVPIPEYQLETGIPGIPDWLFNLVASYLPELP 177
WUPyV     PAGYNPQGGGLNIQSIHKPLHAPYPGMALAPIPEYNLETGIPGVPDWVFNFIASHLPELP 177
MCPyV     DVS-----------------------------------WVGSNEVLRHALMAFSLDPLQ 137
HPyV6     RVGNTTMALQLWLP------------------QVWSWGGAGRGLPDWLINMLR----EVP 153
HPyV7     RIGSTTMALQLWLP------------------QAWPWGGAAGGVPDWLLEVLR----EVP 152
TSPyV     NRNISQ--MALQVWRPWD-----------------YYDILFPGVQTFAHYLN----VLDH 149
HPyV9     S-------MALTPWFP-------------------QVDYLFPGLSSFSYYLN----AALD 148
LPyV      N-------MALVPWFP-------------------QVDYLFPGFTSFSYYLN----AVLD 148
                                                    .       .

SV40      WGPTLFNAISQAFWRVIQNDIPRLT---SQELERRTQRYLRDSLARFLEETTWTVINAPV 208
BKPyV     WGPSLFATISQALWHVIRDDIPSIT---SQELQRRTERFFRDSLARFLEETTWTIVNAPI 209
JCPyV     WGPSLFSTISQAFWNLVRDDLPALT---SQEIQRRTQKLFVESLARFLEETTWAIVNSPA 209
KIPyV     SLQDVFNRIAFGIWSSYYNAGSTVVN---RVLSDEIQRLLRDLEYGFR--ATLASIGESD 232
WUPyV     SLQDVFNRIAYGIWTSYYNTGRTVVN---RAVSEELQRLLGDLEYGFR--TALATIGESD 232
MCPyV     WENSLLHSVGQNIFN---------------SLSPTSRLQIQSNLVNLILNSRWVFQTTAS 182
HPyV6     SPTEILSDIVRGIWTSYYRAGREIIQ---RTASRELGALLSRVRETVIHGAERALEAAPD 210
HPyV7     TPSEILYNIARGIWTSYYRTGRELIQ---RTATRELASLLSRLRQNIINGANRAIEMAPD 209
TSPyV     WASSLIHTVSRYVWDAILHEGRHQIGHASRELMIRGTNHFQDLMARLIENSRWVLTTGPS 209
HPyV9     WGESLFHAVGREIWRNIMRQATQQIGYTSRALAVRGTNEFQHMLAQIAENARWALTNGPI 208
LPyV      WGESLFHAVGTELWRHLMRQATLQIGQATRAVAVRSTNELSHTLAQIAENARWALTSGPV 208
              :: :   .:                          :       .   :   .    .

SV40      NWYNSLQDYYSTLSPIRPTMVRQVANREGLQISFGHTYD----------------NIDE 251
BKPyV     NFYNYIQQYYSDLSPIRPSMVRQVAEREGTRVHFGHTY-----------------SIDD 251
JCPyV     NLYNYISDYYSRLSPVRPSMVRQVAQREGTYISFGHSYTQ---------------SIDD 253
KIPyV     PVNAIATQVRSLATTARERELLQITAGQPLDLSRPTS-----------ALSAAAGALTEA 281
WUPyV     PVNAIVEQVRSFVSGGRERELLQIAAGQPVDISEGVSRGTATISNAVEAVRDATQRLSQA 292
MCPyV     QNQGLLS----------------------------------------------------- 189
HPyV6     PVQGLVNLVNYAVNYNRQWETRALLEGRP---------------------------LFEG 243
HPyV7     PVQGLLNLINQAIAYNRDWETRALLEGRP---------------------------LFEP 242
TSPyV     NIYSHLESYYRDLPGISPPQARDLYRRLQEKIP--------------------------- 242
HPyV9     HIYSSVEEYYRGLPSVNPIQLRQQYRSRGELPPTREQFE--------------------Y 248
LPyV      HIYSTVQDYYRYLPARNPIQLRQEYRNRGEPPPSTADFE--------------------Y 248
```

```
                                                              VP1 binding domain
SV40            ADSIQQVTERWEAQSQSPN----VQSGEFIEKFEAPGGANQRTAPQWMLPLLLGLYGSVT 307
BKPyV           ADSIEEVTQRMDLRNQQS-----VHSGEFIEKTIAPGGANQRTAPQWMLPLLLGLYGTVT 306
JCPyV           ADSIQEVTQRLDLKTPN------VQSGEFIERSIAPGGANQRSAPQWMLPLLLGLYGTVT 307
KIPyV           AYNFIYDASSLPKDGFNALSEGVHRLGQWISFSGPTGGTPHYATPDWILYVLEQLNADTY 341
WUPyV           TYNFVYDASTLPRDGFNALSDGVHRLGQWISMPGATGGTPHYAAPDWILYVLEELNSDIS 352
MCPyV           --------------------------GEAILIPEHIGGTLQQQTPDWLLPLVLGLSGYIS 223
HPyV6           NGVVNYDMQNLPVNGNNDQRGGFHDEGLWVSFSAEQGNTGQYCIPQWLLFVLEELDKEIK 303
HPyV7           GGVVMYDTQNLPLSGNNDQRGGFHDEGTWVSFQGEEGNTPQYTIPQWMLFVLEELQKEVN 302
TSPyV           --------DRYQLEAATD------ESAEVIETYSAPGGAHQRVCPDWMLPLVLGLYGDIT 288
HPyV9           QEQVR---LRREIGGSEP------RSGHYVQHYAAPGGANQRVSQDWMLPLILGLYGDIT 299
LPyV            QENREGQTARRELGYDEP------RSGQYVEHYTAPGGAHQRVTQDWMLPLILGLYGDIT 302
                                          .  :      *.: :    :*:* ::  *

SV40            S-------ALKAYEDGPNKKKRKLSRGSSQK-TKGTSASAKARHKRRNRSSRS------- 352
BKPyV           P-------ALEAYEDGPNQKKRRVSRGSSQK-AKGTRASAKTTNKRRSRSSRS------- 351
JCPyV           P-------ALEAYEDGPNKKKRRK---------EGPRASSKTSYKRRSRSSRS------- 344
KIPyV           KIPTQAVKRK--QDELHPVS--PTKKANKAKKSSSPGTNSGNRSKKRRGRSTSRSTTVRR 397
WUPyV           KIPTQGIKRKLQQNGLHSKASLHSKTRKVTKKSTHKSAKPSKTSQKRRGRRAGRRTTVRR 412
MCPyV           P----------------------------------ELQVIEDGTKKKSIIHL-------- 241
HPyV6           ED---ALSQKR----------------------KWTNSKASQSNKKRRSGGYGNSATF-- 336
HPyV7           KENKHALSKKR----------------------KWTHTETSQSDKKRRR----------- 329
TSPyV           PTFGYYLREEERED-GPQKKRRRMR----------------------------------- 312
HPyV9           PTWEKELSKLEKEEYGPPKKKARVRSMSCKKNLSNTRSRSQTPCQRRGRSSRS------- 352
LPyV            PTSEVELNKLEKEEDGPSKKKARR---SMQKNMPYSRSRPQAPSKRRSRGARSKNRA--- 356

SV40            ---
BKPyV           ---
JCPyV           ---
KIPyV           NRI 400
WUPyV           NRV 415
MCPyV           ---
HPyV6           ---
HPyV7           ---
TSPyV           ---
HPyV9           ---
LPyV            ---
```

Supplementary Figure 2: VP1 protein sequence

```
                          NLS                                       AB-loop
BKPyV           MAPT------KRKGECPGAA-------PKKPKEPVQVPKLLIKGGVEVLEVKTG-VDAIT 46
SV40            MAPT------KRKGSCPGAA-------PKKPKEPVQVPKLVIKGGIEVLGVKTG-VDSFT 46
JCPyV           MAPT------KRKGER---------------KDPVQVPKLLIRGGVEVLEVKTG-VDSIT 38
TSPyV           MAPK------RKGEGCAR----KCPTKT--CPTPKPVPKLIMKGNIEVLNLVTG-PDSIT 47
MCPyV           MAPK------RKASSTCKTPKRQCIPKPGCCPNVASVPKLLVKGGVEVLSVVTG-EDSIT 53
LPyV            MAPQ------RKRQD-GAC--------KKTCPIPAPVPRLLVKGGVEVLEVRTG-PDAIT 44
HPyV9           MAPQ------RKRQECGACP------VKKTCPTPAPVPKLLVKGGVEVLEVRTG-PDAIT 47
HPyV6           MPCH-------RKGNGP----------------IQKLPRVIKKGGVEVMETVPLSEDTIY 37
HPyV7           MPCQ-------RKGNGP----------------TQKLPRVIRKGGVEVLDTVPLTEETQY 37
KIPyV           MSCT--PCRPQKRLTRPRSQ----------VPRVQTLATEVKKGGVEVLAAVPLSEETEF 48
WUPyV           MACTAKPACTAKPGRSPRSQ----------PTRVQSLPKQVRKGGVDVLAAVPLSEETEF 50
                *.         :                        :.  : :*.::*:   .   ::

                          ←--------------BC loop----------------------→
BKV             EVECFLNPEMGDPD---------ENLRGFSLKLSAENDFS----SDSPERKMLPCYSTAR 93
SV40            EVECFLNPQMGNPD---------EHQKGLSKSLAAEKQFT----DDSPDKEQLPCYSVAR 93
JCV             EVECFLTPEMGDPD---------EHLRGFSKSISISDTFE----SDSPNRDMLPCYSVAR 85
TSPyV           TIELYLNTRMGQND-------ESKDNYGYSEKVTVANSSD----QDKPTSGEIPTYSTAR 96
MCPyV           QIELYLNPRMGVNSPDLP---TTSNWYTYTYDLQPKGSS-----PDQPIKENLPAYSVAR 105
LPyV            QIEAYLNPRMGNNI-------PSEDLYGYSNSINTAFSKA----SDTPNKDTLPCYSVAV 93
HPyV9           QIEAYLNPRMGNNN-------PTDELYGYSADINVASSKA----SDNPNATTLPTYSVAV 96
HPyV6           KVEAILLPNFASGSNT-------AVYQSRGAPYT---------FTDTLDAGSSLCYTLAV 81
HPyV7           KVEAVLLPNFGKAATT-------GNFQSRGLPYP---------MSDTLGPGAALCYSVAV 81
KIPyV           KVELFVKPVIGNTTAAQDGREPTPHYWSISSAIHDKESGSSIKVEETPDADTTVCYSLAE 108
WUPyV           KVELFVKPVIGNAEGT------TPHYWSISSPLKTAEAAN-----VTPDADTTVCYSLSQ 99
                 :*  : . :.                                            *: :

                       CD loop                    ←----DE loop-----→
BKV             IPLPNLNEDLTCGNLLMWEAVTVQTEVIGITSMLNLHAGSQKVHEH-GGGKPIQGSNFHF 152
SV40            IPLPNLNEDLTCGNILMWEAVTVKTEVIGVTAMLNLHSGTQKTHEN-GAGKPIQGSNFHF 152
JCV             IPLPNLNEDLTCGNILMWEAVTLKTEVIGVTSLMNVHSNGQATHDN-GAGKPVQGTSFHF 144
TSPyV           INLPMLNEDLTCNTLTMWEAVSVKTEVVGVSSLVNVHMATKRMYDDKGIGFPVEGMNFHM 156
MCPyV           VSLPMLNEDITCDTLQMWEAISVKTEVVGISSLINVHYWDMKRVHDYGAGIPVSGVNYHM 165
LPyV            IKLPLLNEDMTCDTILMWEAVSVKTEVVGISSLVNLHQGGKYIYGSSSGCVPVQGTTYHM 153
HPyV9           IKLPMLNEDMTCDTLLMWEAVSVKTEVMGISSLVNLHQGGKYIYGSSSGTIPVQGTTLHM 156
HPyV6           VNLPEIPEALCDDTLLVWEAFRVETELIFTPQVG-----SAGYIRAQGTPAGVEGSQMYF 136
HPyV7           INLPEIPDAMCEDTMIVWEAYRLETELLFAPQMA-----SSGYQRANGTLAGTEGSQLYF 136
KIPyV           IAPPDIPNQVSECDMKVWELYRMETELLVVPLVN-----ALGNT--NGVVHGLAGTQLYF 161
WUPyV           VAPPDIPNQVSECDMLIWELYRMETEVLVLPVLN-----AGILT--TGGVGGIAGPQLYF 152
                :  * : : :    : :**   ::**::  . :              .      *    ::

                    ←------------------------------EF loop-----------------
BKV             FAVGGEPLEMQGVLMNYRSKYPDGT---------ITPKNPTAQSQVMNTDHKAYLDKNNA 203
SV40            FAVGGEPLELQGVLANYRTKYPAQT---------VTPKNATVDSQQMNTDHKAVLDKDNA 203
JCV             FSVGGEALELQGVLFNYRTKYPDGT---------IFPKNATVQSQVMNTEHKAYLDKNKA 195
TSPyV           FAVGGEPLELQFLTGNYRTDYSAND-------KLVVPPIKHQSTQGLNPHYKQKLTKDGA 209
MCPyV           FAIGGEPLDLQGLVLDYQTEYPKTTNGGPITIETVLGRKMTPKNQGLDPQAKAKLDKDGN 225
LPyV            FAVGGHPLELQGLVASSTATYPDDV---------VAIKNMKPGNQGLDPKAKPLLDKDGN 204
HPyV9           FSVGGEPLELQGLVASSTTTYPTDM---------VTIKNMKPVNQALDPNAKALLDKDGK 207
HPyV6           WACGGSPLDVIGINP-DPERMNVAAG-------LEGPSKENQPSVAGIK-ATRKQVTAAN 187
HPyV7           WACGGGPLDVIGINP-DPERLKVNEA-------LEGPGN---TDVASLQ-ALRKQVNAAN 184
KIPyV           WAVGGQPLDVVGVTPTDKYKGPTTYT-------INPPGDPRTLHVYNSN-TPKAKVTSER 213
WUPyV           WAVGGQPLDVLGLAPTEKYKGPAQYT-------VNPKTNGTVPHVYSSSETPRARVTNEK 205
                :: ** .*::  :                                           .
```

```
                 -------------→                           ←-GH loop-→
BKV              YPVECWVPDPSRNENARYFGTFTGGENVPPVLHVTNTATTVLLDEQGVGPLCKADSLYVS 263
SV40             YPVECWVPDPSKNENTRYFGTYTGGENVPPVLHITNTATTVLLDEQGVGPLCKADSLYVS 263
JCV              YPVECWVPDPTRNENTRYFGTLTGGENVPPVLHITNTATTVLLDEFGVGPLCKGDNLYLS 255
TSPyV            FPVECWCPDPSKNENTRYYGSYTGGQSTPPVLQFTNTVTTVLLDENGVGPLCKGDGLYVS 269
MCPyV            YPIEVWCPDPSKNENSRYYGSIQTGSQTPTVLQFSNTLTTVLLDENGVGPLCKGDGLFIS 285
LPyV             YPVEVWCPDPSKNENTRYYRSFTGGATTPPVMQFTNSVTTVLLDENGVGPLCKGDKLFLS 264
HPyV9            YPVEVWSPDPSKNENTRYYGSFTGGATTPPVMQFTNSVTTVLLDENGVGPLCKGDKLFLS 267
HPyV6            FPIEIWSADPTRNENCRYFGRIVGGSVTPPVVSFGNQSTTPLVDENGVGILCLFGAIYLT 247
HPyV7            FPVELWVADPTKNDNTRYFGRVVGGGVTPPVVSYGNQSTTPLIDENGVGILCTFGSVYLT 244
KIPyV            YSVESWAPDPSRNDNCRYFGRVVGGAATPPVVSYGNNSTIPLLDENGIGILCLQGRLYIT 273
WUPyV            YSIESWVADPSRNDNCRYFGRMVGGAATPPVVSFSNNSTIPLLDENGIGILCLQGRLYIT 265
                 :.:* * .**::*:* **:     *  .*.*:   *  *  *:** *:* **  . ::::

                      ←----HI loop-----→
BKV              AADICGLFTN----------SSGTQQWRGLARYFKIRLRKRSVKNPYPISFLLSDLINRR 313
SV40             AVDICGLFTN----------TSGTQQWKGLPRYFKITLRKRSVKNPYPISFLLSDLINRR 313
JCV              AVDVCGMFTN----------RSGSQQWRGLSRYFKVQLRKRRVKNPYPISFLLTDLINRR 305
TSPyV            CCDIVGFLVG----------KDGDMQYRGLPRYFNILLRKRTVRNPYPVSSLLNNLFTGL 319
MCPyV            CAHIVGFLFK----------TSGKMALHGLPRYFNVTLRKIWVKNPYPVVNLINSLFSNL 335
LPyV             CADIAGVHTN----------YSETQVCTALPRYFNVTLRKRIVKNPYPVSSLLNTFFSGL 314
HPyV9            AVDIVGIHTN----------YSESQNWRGLPRYFNVTLRKRVVKNPYPVSSLLNSLFSGL 317
HPyV6            SADMLGMVGYAGNPTLSDAYSQQRSVQAAFGRFFRVHFRQRRVKHPYTVDMMFRQFLQPQ 307
HPyV7            SADMVGMTGLPGLPTLSADYSNQRTVQAGYGRFFRVHCRQRRIKHPYTVDMMFRQFLQPQ 304
KIPyV            CADMLGTA--------------NSRIHTPMARFFRLHFRQRRVKNPFTMNVLYKQVFN-- 317
WUPyV            CADLLGVN--------------KNRVHTGLSRFFRLHFRQRRVRNPYTINLLYKQVFN-- 309
                 .  .: *                         *:*.:   *:   :::*:.:  :    .:

BKV              TQRVDGQPMY-GMESQVEEVRVFDGTERLPGDPDMIRYIDKQGQLQTKML---------- 362
SV40             TQRVDGQPMI-GMSSQVEEVRVYEDTEELPGDPDMIRYIDEFGQTTTRMQ---------- 362
JCV              TPRVDGQPMY-GMDAQVEEVRVFEGTEELPGDPDMMRYVDKYGQLQTKML---------- 354
TSPyV            MPAVQGQPMDNGLSTQVEEVRVYDGTEGLPGDPDMVRYIDKFGQDKTRPPFPARLY---- 375
MCPyV            MPKVSGQPME-GKDNQVEEVRIYEGSEQLPGNPDIVRFLDKFGQEKTVYPKPSVAPAAVT 394
LPyV             MPQIQRQPIE-RVSGQVEEVRIFQGTEGLPGDPDLNRYVDKFCQHQTVLPVSNDM----- 368
HPyV9            MPQIQGQPME-GTSGQVEEVRIYQGTEGLPGDPDLDRYVDKFCQNQTVPPRSNDQ----- 371
HPyV6            KPQVQGTQPNAVQEVVMEQMQPSILPTTLEGAIGYSPSTKFILQNGELIYPSSTVAAGAA 367
HPyV7            KPQVQGQQPAAVQEVTMEQMQPATIPPTVEGGLGFAPTSKFLIQNGELIYPSSNAAAAAA 364
KIPyV            RPTETVDAQVGVTEVTMVEEIG-PLPPSIQ--TTLPTSVNLTQ------LPRTVTLQSQA 368
WUPyV            KPADDISGQLQVTEVTMTEETG-PLPPTVEGNVGVPTTSNLSH------LPATVTLQATG 362
                             .   :  :        .   :            .

BKV              -----------------------------
SV40             -----------------------------
JCV              -----------------------------
TSPyV            -----------------------------
MCPyV            FQSNQQDKGKAPLKGPQKASQKESQTQQL 423
LPyV             -----------------------------
HPyV9            -----------------------------
HPyV6            NLFGPPVEKQTSKEPSKGEL--------- 387
HPyV7            KISVAPKKNTDNKK----EL--------- 380
KIPyV            PLLNTQQNSK------------------- 378
WUPyV            PILNTQG---------------------- 369
```

Supplementary Figure 3: Primer sites in the respective genome

>gi|124366173|gb|EF127906.1| KI polyomavirus Stockholm 60, complete genome
1
CCTCTGGCCTCCTGTAATATAGAAAAAAAGGGCACAGTGATTGACAGTTGTGTATACAAGCATGTGTGGT
ATGTTTAGTGTGTAAGCCAATAAAGTTAAAGGTCACTACTGTGGGTGGTGACACCTGATACCGGCGGAAC
TAGTTGCTACGTGCCACACAATAGCTTTCACTCTTGGCGTGAAGCCAACTTCCTGGGCCGTGAGCCAGCT
TCCTGCGGCCTTGTGTTTTTTACCAACACACCTGGTGAACTTCTACTGTCCTTGACACAGGTAAGACTGG
GGACCCTTTGTAGGCCAAAGGAGAGTGAAGGGTAACTGAAATGCTAAGACTGTAAGTTCTAATCCTAGTA
TTTCAGTTCGGGGATGTTGGCGCCATCGTCTCGAACCTGGCCTGCATACCTTTGGATATAGAGGGTCACC NCCR R
AATTTTTATTTTGTTTTTAGATGGGAATATTTCTTGCTGTACCTGAAATTATAGCAGCATCTATAGCTGG
AGGAGCAGAGGCCTTGTCAATTGCTGGTTCTGGAGCTGCCATAGCTACTGGTGAAGGCTTAGCTGCCCTT
GGTGGTATAACAGAGGGAGCTGCACTTCTTGGGGAAACAATACCAATATCTGAAGCAGCTACTACTGTGT
TAACTAAAGTTCCTGAACTTGTTCAAGCTACCCAAGCTGTAACAGCAGCTGTGCAAGGAGGTGCAGGCCT
TGTAGGTGGCATTTATACTGCCTTAGCATCAGACCACCCTGGGGACCTGCCTCCGAACACCCCAACAGGA
AGTGCAAGTGGACTCCATCCCACCTCAGGATACAATCCCCAGGGAGCAGGACTCAATCTCCAGTCAGTGC
ACAAACCAATTCACGCCCCCTACTCAGGAATGGCATTGGTACCTATCCCTGAATACCAGTTGGAAACTGG
AATTCCGGGCATCCCAGACTGGCTCTTCAACCTTGTCGCGTCATACTTGCCCGAGTTGCCTAGTTTGCAG
GACGTGTTCAATAGAATTGCCTTTGGAATCTGGTCATCATATTACAATGCGGGGTCAACAGTAGTTAACA
GAGTTCTTAGTGATGAAATTCAAAGACTTTTAAGAGACCTTGAATATGGATTTAGAGCAACTTTAGCCAG
CATAGGAGAATCAGATCCTGTAAATGCTATAGCTACTCAAGTTAGAAGCTTAGCTACCACAGCAAGGGAA
AGGGAGCTGCTACAAATTACTGCAGGTCAGCCACTTGACCTTTCTAGGCCTACCAGCGCCTTGTCAGCAG
CAGCTGGAGCATTAACTGAAGCAGCTTATAATTTTATATATGATGCTTCAAGCCTGCCCAAAGATGGCTT
TAATGCTTTAAGTGAGGGTGTACACAGACTAGGTCAGTGGATTTCCTTTTCAGGGCCTACAGGAGGTACC
CCACATTATGCAACTCCTGACTGGATTTTGTATGTGCTAGAACAACTAAATGCTGATACCTATAAAATTC Part of
CTACACAGGCAGTAAAAAGAAAACAAGATGAGCTGCACCCCGTGTCGCCCACAAAAAAGGCTAACAAGGC VP1 SalI
CAAGAAGTCAAGTTCCCCGGGTACAAACTCTGGCAACAGAAGTAAAAAAAGGAGGGGTAGAAGTACTAGC KI VP1 39F
CGCAGTACCACTGTCAGAAGAAACAGAATTTAAAGTGGAACTATTTGTAAAGCCAGTAATTGGAAATACA KIVP1 118F
ACAGCTGCTCAGGATGGGCGTGAGCCCACCCCTCATTACTGGTCAATTAGCTCTGCCATTCATGACAAGG
AAAGCGGTTCAAGTATCAAAGTTGAAGAAACTCCAGATGCTGACACAACTGTATGTTACAGCCTGGCAGA KIVP1 324R
AATTGCTCCCCCTGATATACCAAATCAAGTTAGTGAGTGTGACATGAAAGTATGGGAGCTGTACAGAATG KIVP1 363R
GAAACAGAGTTGCTTGTTGTACCTCTAGTTAATGCTCTAGGAAACACCAATGGTGTTGTACATGGTCTGG
CTGGAACCCAGTTGTATTTTTGGGCTGTTGGGGGACAGCCACTTGATGTAGTTGGTGTAACACCCACAGA
CAAATATAAAGGCCCAACTACCTATACAATTAATCCACCAGGAGACCCTAGAACACTGCATGTGTACAAT KIVP1 2071
AGTAATACACCCAAAGCAAAGGTTACCAGTGAGAGATATTCTGTTGAATCATGGGCCCCAGACCCCAGTA -2093 F
GAAATGACAATTGTAGATATTTTGGAAGAGTGGTAGGTGGTGCTGCAACACCTCCAGTTGTATCATATGG
TAACAACTCTACTATTCCACTATTGGATGAAAATGGCATTGGTATACTTTGCTTGCAGGGAAGATTGTAC
ATTACTTGTGCAGATATGCTTGGAACAGCTAATAGTAGAATCCATACCCCTATGGCTAGGTTTTTTAGGC
TACATTTTAGACAAAGAAGGGTTAAGAATCCTTTTACAATGAATGTGCTGTATAAACAAGTGTTTAACAG
ACCCACAGAAACTGTTGATGCACAGGTTGGTGTAACAGAGGTGACTATGGTAGAAGAAATAGGCCCACTG
CCCCCCAGTATACAAACTACCCTCCCCACCAGTGTAAATCTTACTCAGCTTCCACGCACTGTAACACTTC Part of
AGTCCCAGGCTCCTTTGTTAAATACTCAACAAAATTCAAAGTGAATGTATTATAATAAAGCTGTTTATTC VP1 SacII
AAACCACTTTTCTAGTATGTTTTTCCATTGGTTACATTCATTTGCATAGTGGCAAACTCAGTAAGGCCT
ATATAATTATCAAGAACTTCTTTCCAATACACAACATTACTTTGAATTTCTTCATCAAAGTCACTTACAG
GTCTATACCATATTAGCAATAACAACATAGCTATTCCACTGTGTAATATTCTTTGTGATAGTAACTGTGG
TGTTTTTTCTAGGCTTTCTCTTAAATTTCTTTTAATAGTGAACATAACAGTTTTTTCAAATCTCACAGCA
ACTGTTTCTGGAATACAATATTCATTCATGGTTACAATCCCTGGGGGAAATATTTGTGACCTCTTATTAA
CATGTTTTTTTTCTAAGTTAACCTCAACACAACCATCAAGGTAGTCTCTTAAGTTATCAAGATTATTTAC
CCCATTACCAGAAGGTAGTTGCTTTTCAATACCTATTTGACCTTTGACATCTTCAAACACAACCATATAT
TGATCAAGAGCAACTCCAAGTTCAAATGGTAGTTTTTCTGAAGGTATGTTAATGTTCAAGGCTTTGCCAC
CACACAAATTTAACAAAGCAGTGGCTAATGTGGTTTTGCCACTGTTAATTGGGCCTTTGAAAAGCCAATA
TCTGTGTTTTGGCTTGTTGTCTACAATTAGCTTTAAAAATCTAACAACTAATTCATCTATTTTTTTATTT
AGTCCAGTATACCAGGCTATTGAAGCCATATAAAGCAGTATTTTTCTTTCTCCAACAACAATTTTATCCA
TTTCATTAAAAAGCTTTTTAAACCTTTTAGTTAGTAGCTTTTTCCTAGTTAAGGTTAAACAGTTATATCT
TCTTTCTGCTATAACTCTATCAATGGCTTGTGTGCAAGCATTTTTTTGGCTTTTACTTTCTTGCAAAAGC
ACAGCATTTTCATGATTTTGCACATGTACCCTTTTATGGGTAGCATCATCATTTTTACTGCATTTTTCAC
ACTCTTCAGGCTTAGTAGCAAACCTTATATATGACCCCAGTAAGTAATACACATCATCTATGTCAAATTC
CAGAGCAAAGTTAGCTATTTGATTCCAGCTTAATGAGGGTTCCTTAGGTTCCCCAAATAAATCTTCTGGA
TTAAATTCATTTTCTTTTAACCCTCCAGGAATATTTTCTTCACACAAATTAAATGGTTGTCTACACATCC
TACTATACAGACCATAAGGGTTATTAACTCCTTTACAGAATAAAAAGCTAACAGTACAATGACCTTTGCA
GAAGTTATTAATAGCACTAACTCTATGCTTGTGAGGAGTTAAGAAAAATATTAAAACAGTATTATAAAAT LT ag R
GCATGCTTACTTATAAAGCTACATTGATACTTTTGAAGTAACTTTTTATAAAGTACTTCCCCTTTTTCTC
TAGTAGTATGTACTACAAAGCACGTTATACACTTATTACTAAATACAGCATGTGAAACAAACTCTAACAA
CTCTTTAGGAAATTCTAAAGAAGCATCAAATGCATGTTTTTTCTTAGGAGGGGTTGCCTGAGAACAGCTA
GGCTCCTCTTCTGGTGCACTTCTTTTTGGAGAAGATGGCATGGATTCATTACATCTTAACTCTTCTTCCC
AGTAGGTGTTGAACTGTGACCACCACTCATCCCAGTCAGGTGTACCGTATGTAGGTATCTATACAACAAA LT-ag F
AATATACATTTAATATTACACAAACAATTATTTAAAAACATTATTTTATGCTTACCGTAATAATGGTTGA 1. part

```
ACTGCTCGTATTTCAGTATTAAAAATAATCACTTTCCAAAAGAAAAATGTTTGCATTGAAATTGGAAAAA
AAAACCACTCCCTGTAGCACTTATAGCAGTAGCACTCTATCCACACAGGTGGTTTTCTATAAATTTTGTA
CTTTTCTTTATGATATTTGCTTAAAATACAGTGAATGCATTGGCATTCGTGATTGTAGCCATGAATGCAT
ACATCCCACTGCTTCATAATATATGCTTCATAGTAATCACCATATATAGCACCAACTAATCTAAACTTAT
TGCAGCACAAATCTTTGCAGTATACCTGAGAAGATTGCCATATATTATCTTCTTCTTCATTTAAATCATG LT-ag F
TACACTGCTAACTGAGTCTTGCAGTTTTAAATATAAAGAATTAAGAAGCTTCATGCTTTCCTCATTTCCT 2. part
CCTTTATCTGGATGATATTCTTTGCACTTTACTAAGTATTGTCTTCTCATTAATGGTAAATTGCCCCAGC
AACTCATGTCCAAACATAAAAGTTGCATTAGCTGCTTTGCCTCTTCTCTAGATAAAGTTTTATCCATTTT NCCR F
GCCTCTCTAGGCCTCTCAAATGCCTCTGCAGGCCCTCTCCTTCTTCTAGAAAAAGCTGGGGCTTTTTTGG 5040
```

VP2
VP3
VP1
LT-ag
ST-ag

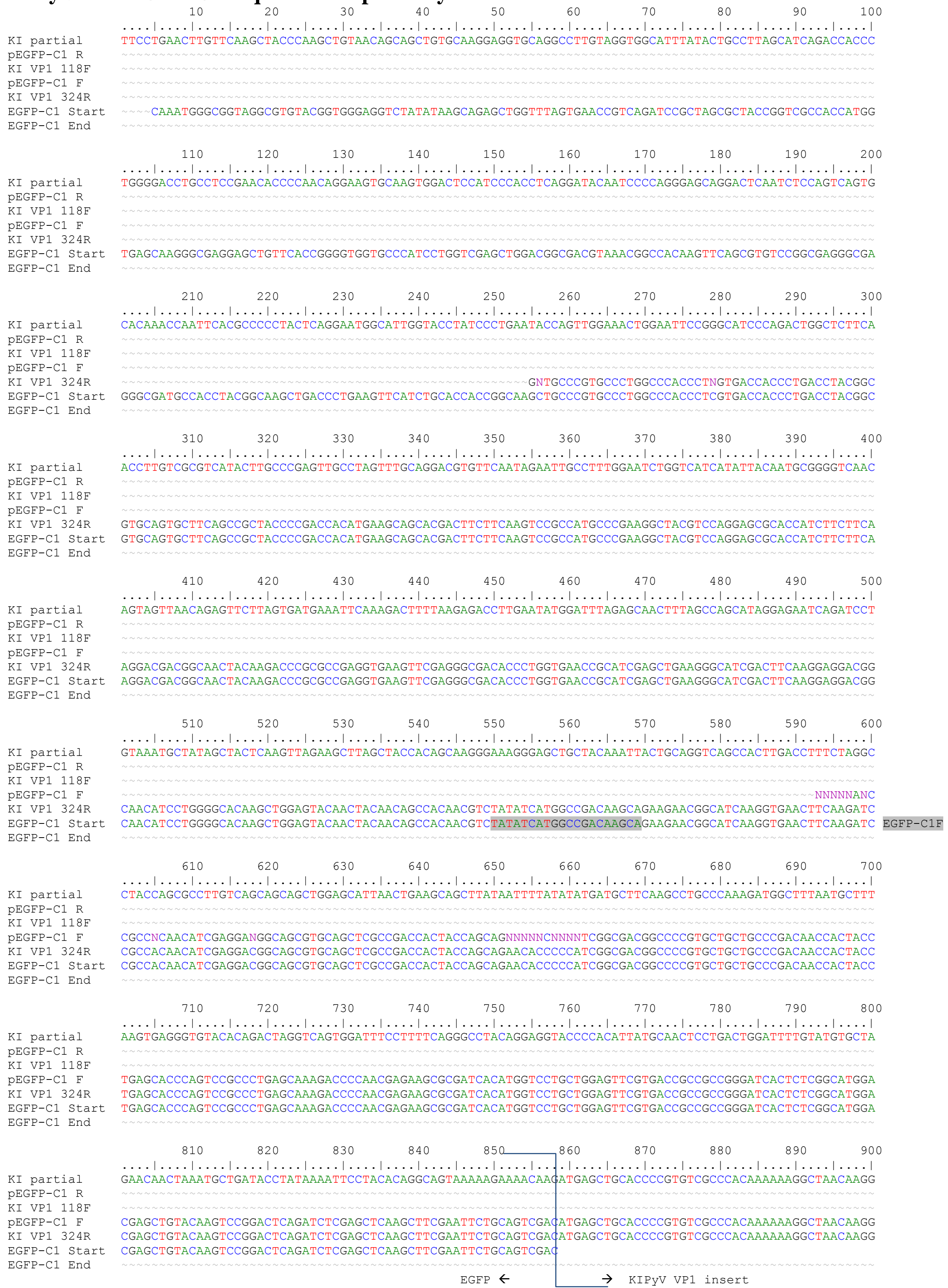
gi|22219459|ref|NM_004379.| Homo sapiens cAMP responsive element binding protein 1 (CREB1), transcript variant A, mRNA

```
   1 agtcggcggc ggctgctgct gcctgtggcc cgggcggctg ggagaagcgg agtgttggtg
  61 agtgacgcgg cggaggtgta gtttgacgcg gtgtgttacg tgggggagag aataaaactc
                                                   Exson 1      Exson 2
 121 cagcgagatc cgggccgtga acgaaagcag tgacggagga gcttgtacca ccggtaacta
                        h CREB F primer
 181 aatgaccatg gaatctggag ccgagaacca gcagagtgga gatgcagctg taacagaagc
                                                             Exson 3
 241 tgaaaaccaa caaatgacag ttcaagccca gccacagatt gccacattag cccaggtatc
 301 tatgccagca gctcatgcaa catcatctgc tcccaccgta actctagtac agctgcccaa
 361 tgggcagaca gttcaagtcc atggagtcat tcaggcggcc cagccatcag ttattcagtc
                          Exson 4
 421 tccacaagtc caaacagttc agatttcaac tattgcagaa agtgaagatt cacaggagtc
                                                   h CREB R primer
 481 agtggatagt gtaactgatt cccaaaagcg aagggaaatt ctttcaagga ggccttccta
          Exson 5
 541 caggaaaatt ttgaatgact tatcttctga tgcaccagga gtgccaagga ttgaagaaga
 601 gaagtctgaa gaggagactt cagcacctgc catcaccact gtaacggtgc caactccaat
 661 ttaccaaact agcagtggac agtatattgc cattacccag ggaggagcaa tacagctggc
 721 taacaatggt accgatgggg tacagggcct gcaaacatta accatgacca atgcagcagc
 781 cactcagccg ggtactacca ttctacagta tgcacagacc actgatggac agcagatctt
 841 agtgcccagc aaccaagttg ttgttcaagc tgcctctgga gacgtacaaa cataccagat
 901 tcgcacagca cccactagca ctattgcccc tggagttgtt atggcatcct ccccagcact
 961 tcctacacag cctgctgaag aagcagcacg aaagagagag gtccgtctaa tgaagaacag
1021 ggaagcagct cgagagtgtc gtagaaagaa gaaagaatat gtgaaatgtt tagaaaacag
1081 agtggcagtg cttgaaaatc aaaacaagac attgattgag gagctaaaag cacttaagga
1141 cctttactgc cacaaatcag attaatttgg gatttaaatt ttcacctgtt aaggtggaaa
1201 atggactggc ttggccacaa cctgaaagac aaaataaaca ttttattttc taaacatttc
1261 tttttttcta tgcgcaaaac tgcctgaaag caactacaga atttcattca tttgtgcttt
1321 tgcattaaac tgtgaatgtt ccaacacctg cctccacttc tcccctcaag aaattttcaa
1381 cgccaggaat catgaagaga cttctgcttt tcaacccccca ccctcctcaa gaagtaataa
1441 tttgtttact tgtaaattga tgggagaaat gaggaaaaga aaatcttttt aaaaatgatt
1501 tcaaggtttg tgctgagctc cttgattgcc ttagggacag aattacccca gcctcttgag
1561 ctgaagtaat gtgtgggccg catgcataaa gtaagtaagg tgcaatgaag aagtgttgat
1621 tgccaaattg acatgttgtc acattctcat tgtgaattat gtaaagttgt taagagacat
1681 accctctaaa aaagaacttt agcatggtat tgaaggaatt agaaatgaat ttggagtgct
1741 ttttatgtat gttgtcttct tcaatactga aaatttgtcc ttggttctta aaagcattct
1801 gtactaatac agctcttcca tagggcagtt gttgcttctt aattcagttc tgtatgtgtt
1861 caacattttt gaatacatta aaagaagtaa ccaactgaac gacaaagcat ggtatttgaa
1921 ttttaaatta aagcaaagta aataaaagta caaagcatat tttagttagt actaaattct
1981 tagtaaaatg ctgatcagta aaccaatccc ttgagttata taacaagatt tttaaataaa
2041 tgttattgtc ctcaccttca aaaatattta tattgtcact catttacgta aaaagatatt
2101 tctaatttac tgttgcccat tgcacttaca taccaccacc aagaaagcct tcaagatgtc
2161 aaataaagca aagtgatata tatttgttta tgaaatgtta catgtagaaa aatactgatt
2221 ttaaatattt tccatattaa caatttaaca gagaatctct agtgaatttt ttaaatgaaa
2281 gaagttgtaa ggatataaaa agtacagtgt tagatgtgca caaggaaagt tattttcaga
2341 catatttgaa tgactgctgt actgcaatat ttggattgtc attcttacaa aacatttttt
```

```
2401 tgttctcttg taaaaagagt agttattagt tctgctttag ctttccaata tgctgtatag
2461 cctttgtcat tttataattt taattcctga ttaaaacagt ctgtatttgt gtatatcata
2521 cattgttttc aataccactt ttaattgtta ctcattttat tcactaagct cgataaatct
2581 aacagttact cttaaaaaaa aaaaaaagac taaggtggat tttaaaaatt ggaaactgac
2641 ataatgttag gttataattt ctcatttgga gccgggcgca gtggctcacg cctgtaatcc
2701 cagcactttg ggaggccaag gtgggtggat cacctgtggt caagagttca agaccagcct
2761 ggccatcatg gtgaaacccc atctctacta aaaatacaaa aattagccag gcgtggtggc
2821 tggcgcctgt aatcccagct actcaggagg ttgaggcagc agaattgctt gaacccagga
2881 ggcagagggt tgcagtgagc cgagatagca ccattgcact ccagcctggg cgactccatc
2941 tcaaaaaata aaaaaaaaaa aaaa
```

```
>gi|224589807:c88878342-88875877 Homo sapiens chromosome 16,
GRCh37.p5 Primary Assembly
GCGCTCGGGCTGCCGCTGGCTCTTCGCACGCGGCCATGGCCGACTCCGAGCTGCAGCTGGTTGAGCAGCG
GATCCGCAGCTTCCCCGACTTCCCCACCCCAGGCGTGGTATTCAGGTGCACGCACAGGCCGCCCTCGTGG
CGCCCCGACCTGCGGGCCTACGGATGGGAGCGCGTGGCCCGCGACCTCCGGGCGGGCGGGGCGGGAACCC
TCGTCTTTCGCCCCCGGGGCCCTGCCCTCCTTCGGCCCCGGCGTCACCAGGCCTGTCCTTGGGTCCAGGG
ACATCTCGCCCGTCCTGAAGGACCCCGCCTCCTTCCGCGCCGCCATCGGCCTCCTGGCGCGACACCTGAA
GGCGACCCACGGGGGCCGCATCGACTACATCGCAGGCGAGTGCCCAGTGGCCGCATCTAGGGCGCTTCCG
CCTCTGCGCGCGCCGAGGGCAGCACGTGGGCTCTGCGCGTCTGCTTGGGGGAGGGCCTTTGGGGTGCTTC
AGGGGGCGCCGGGACGGGCGCCGTGCTTGGGTCGCCCGGGAAGGGTTGTGAGATTGAGCCCCCGAGGCCG
CCGCGCTGTGCAGGCGTCCTTCCCGCAGGTTCCGGGTCCCCAGCCCAGGACAGGCGTGACCGAGTTGCCG
GGTCAGTTGGTCTCCCTGGAGTGCCCAAGCTGAATCCACAGGGCCCAGCTGCCTTGCTTCTTGTTCCTTC
TGCGAGCTGGTATTGAGCGCCTGCCACGAGCCAGGCCTTCCCTGGTGAAGATCACGGAATGCCCACCCAG
GGAAGGGAGGCCTGGAGGCCTCCGGGAGAGCCCAAGAGGTGGCCCAGGGAGAACAGAGTGTTCCTGGCCG
TCTTGCCTCTCCTAGGGTGTGACAGCCCACTCCCTGGACACTGCCCTGAGGAAAGCGCCAGCTCTTGCTG
GAGCCACAACACTGCCAGAGCTCCCTTCTCACCTCCTGCAGGAAGCCCTCCCTGACCTCCTGCCAGGCCG
GGGCAGGGTTTCCCTGAGCGTCCCCCAACCATCACAGCTCAGGCCACCTCGAGAGACTCCCTTTTTAGAC
AGAAGCCCTGGTGCAGAGCTGCCTTTGAGAGTAAGCTGAGGCCTGTCAGGTTTCTACCAGCCCAGTTACA
GATGGGCTGCTCAGCTCAGAGAGAGGGGTGGTGACTCCCCTAGGAACACACAGCTAAGAAGTGGTCCCTT
AAAAGACAGACCCAGGTCTGCACTCTGACCTGGAAGCAGCTCCGGGTAGGTGATGGGTAACATTCCTTAA
ATGGTGCATGTCACTGGCCTTTCAGCTGGGAGCCAACCAGGTACCCCTTGCCACCGGCCAACCCTGGCCC
CTGGGGATTCCCATGCTGCCGAGTCACTCCTGTCACTTACCCTGACAGGCCTAGACTCCCGAGGCTTCCT
CTTTGGCCCCTCCCTGGCCCAGGAGCTTGGACTGGGCTGCGTGCTCATCCGAAAGCGGGGGAAGCTGCCA APRT F
GGCCCCACTCTGTGGGCCTCCTATTCCCTGGAGTACGGGAAGGTAAGAGGGCTGGGGTGGCCAGAGGAAG
GGCAGGGCCAGGCCACCGTGGCCACTCTCCCCCAGTTCTAAAAGGCCTTCCCAGGCGTGTCAAGTGGAGC
TGCTGTGGTTACAGTGGCCTTGGGAGCTCAGAGAGGTTGAGACATAGGCTGGGCTCACACAGCCAGGTAA
CAGCAAGGTGGGGTTGGAGTCAGGGTCTAGGGTGGCAGCTGCCAAGCTGTGCAACAAAGCTGTTTTCTGC
GGGAGGCTGAGGACCACACACCACTTCCCACTCCAGGCTGAGCTGGAGATTCAGAAAGACGCCCTGGAGC
CAGGACAGAGGGTGGTCGTCGTGGATGATCTGCTGGCCACTGGTGGTAAGGGTCTCCCCGCAGCCAACTG
CTGTGGCTCCAAGGGCCTGGTGGGAGTGGGACAGGACCTCGCTGTGTGACATGGGATGCAGCTTACTGTT
GTCCAGAGGGTGCCTGGTGGCCAGGCCGACACCTTCCTCTCCCCATGCCTTCCCCTCCCCAACCCAGGGG
CTGGCCTGGAGCACCTGCTCTCTGCAGCCCAGGCCAACTGGGGACCTCACCCTCCCATCCCCAGGAACCA
TGAACGCTGCCTGTGAGCTGCTGGGCCGCCTGCAGGCTGAGGTCCTGGAGTGCGTGAGCCTGGTGGAGCT APRT R
GACCTCGCTTAAGGGCAGGGAGAAGCTGGCACCTGTACCCTTCTTCTCTCTCCTGCAGTATGAGTGACCA
CAGGGCCTCCCAGCCCAACATCTCCAGCTGGATCCCAGGGAAATATCAGCCTTGGGCAACTGCAGTGACC
AGGGGCACCGGCTGCCCACAGGGAACACATTCCTTTGCTGGGGTTCAGCGCCTCTCCTGGGGCTGGAAGT
GCCAAAGCCTGGGGCAAAGCTGTGTTTCAGCCACACTGAACCCAATTACACACAGCGGGAGAACGCAGTA
AACAGCTTTCCCACAA
```

Supplementary Figure 4: EGFP-KI VP1 sequence. The top and bottom two sequences are the KIPyV and EGFP-C1 sequences respectively



```
                        910       920       930       940       950       960       970       980       990      1000
                ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
KI partial      CCAAGAAGTCAAGTTCCCCGGGTACAAACTCTGGCAACAGAAGTAAAAAAAGGAGGGGTAGAAGTACTAGCCGCAGTACCACTGTCAGAAGAAACAGAAT 118F
pEGFP-C1 R      ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
KI VP1 118F     ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
pEGFP-C1 F      CCAAGAAGTCAAGTTCCCCGGGTACAAACTCTGGCAACAGAAGTAAAAAAAGGAGGGGTAGAAGTACTAGCCGCAGTACCACTGTCAGAAGAAACAGAAT
KI VP1 324R     CCAAGAAGTCAAGTTCCCCGGGTACAAACTCTGGCAACAGAAGTAAAAAAAGGAGGGGTAGAAGTACTAGCCGCAGTACCACTGTCAGAAGAAACAGAAT
EGFP-C1 Start
EGFP-C1 End     ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

                       1010      1020      1030      1040      1050      1060      1070      1080      1090      1100
                ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
KI partial      TTAAAGTGGAACTATTTGTAAAGCCAGTAATTGGAAATACAACAGCTGCTCAGGATGGGCGTGAGCCCACCCCTCATTACTGGTCAATTAGCTCTGCCAT
pEGFP-C1 R      ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
KI VP1 118F     ~~~~~~~~~~~~~~~~~~~~~~~~CAGTAATTGGAAATACAACAGCTGCTCAGGATGGGCGTGAGCCCACCCCTCATTACTGGTCAATTAGCTCTGCCAT
pEGFP-C1 F      TTAAAGTGGAACTATTTGTAAAGCCAGTAATTGGAAATACAACAGCTGCTCAGGATGGGCGTGAGCCCACCCCTCATTACTGGTCAATTAGCTCTGCCAT
KI VP1 324R     TTAAAGTGGAACTATTTGTAAAGCCAGTAATTGGAAATACAACAGCTGCTCAGGATGGGCGTGAGCCCACCCCTCATTACTGGTCAATTAGCTCTGCCAT
EGFP-C1 Start
EGFP-C1 End     ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

                       1110      1120      1130      1140      1150      1160      1170      1180      1190      1200
                ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
KI partial      TCATGACAAGGAAAGCGGTTCAAGTATCAAAGTTGAAGAAACTCCAGATGCTGACACAACTGTATGTTACAGCCTGGCAGAAATTGCTCCCCCTGATATA 324 R
pEGFP-C1 R      ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
KI VP1 118F     TCATGACAAGGAAAGCGGTTCAAGTATCAAAGTTGAAGAAACTCCAGATGCTGACACAACTGTATGTTACAGCCTGGCAGAAATTGCTCCCCCTGATATA
pEGFP-C1 F      TCATGACAAGGAAAGCGGTTCAAGTATCAAAGTTGAAGAAACTCCAGATGCTGACACAACTGTATGTTACAGCCTGGCAGAAATTGCTCCCCCTGATATA
KI VP1 324R     TCATGACAAGGAAAGCGGTTCAAGTATCAAAGT~GA
EGFP-C1 Start
EGFP-C1 End     ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

                       1210      1220      1230      1240      1250      1260      1270      1280      1290      1300
                ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
KI partial      CCAAATCAAGTTAGTGAGTGTGACATGAAAGTATGGGAGCTGTACAGAATGGAAACAGAGTTGCTTGTTGTACCTCTAGTTAATGCTCTAGGAAACACCA
pEGFP-C1 R      ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
KI VP1 118F     CCAAATCAAGTTAGTGAGTGTGACATGAAAGTATGGGAGCTGTACAGAATGGAAACAGAGTTGCTTGTTGTACCTCTAGTTAATGCTCTAGGAAACACCA
pEGFP-C1 F      CCAAATCAAGTTAGTGAGTGTGACATGAAAGTATGGGAGCTGTACAGAATGGAAACAGAGTTGCTTGTTGTACCTCTAGTTAATGCTCTAGGAAACACCA
KI VP1 324R
EGFP-C1 Start
EGFP-C1 End     ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

                       1310      1320      1330      1340      1350      1360      1370      1380      1390      1400
                ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
KI partial      ATGGTGTTGTACATGGTCTGGCTGGAACCCAGTTGTATTTTTGGGCTGTTGGGGGACAGCCACTTGATGTAGTTGGTGTAACACCCACAGACAAGTATAA
pEGFP-C1 R      ~NNNGNTNNNNCANNGTCTGGCNGNANCCCAGNNNTATTTTTGGGCTGNTGGGGGNCAGCCACTNGATGTAGTNGGTGTAACACCCACAGACAAATATAA
KI VP1 118F     ATGGTGTTGTACATGGTCTGGCTGGAACCCAGTTGTATTTTTGGGCTGTTGGGGGACAGCCACTTGATGTAGTTGGTGTAACACCCACAGACAAATATAA
pEGFP-C1 F      ATGGTGTTGTACATGGTCTGGCTGGAACCCAGTTGTATTTTTGGGCTGTTGGGGGACAGCCACTTGATGTAGTTGGTGTAACACCCACAGACAAATATAA
KI VP1 324R
EGFP-C1 Start
EGFP-C1 End     ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

                       1410      1420      1430      1440      1450      1460      1470      1480      1490      1500
                ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
KI partial      AGGCCCAACTACCTATACAATTAATCCACCAGGAGACCCTAGAACACTGCATGTGTACAATAGTAATACACCCAAAGCAAAGGTTACCAGTGAGAGATAT
pEGFP-C1 R      AGGCCCAACTACCTATACAATTAATCCACCAGGAGACCCTAGAACACTGCATGTGTACAATAGTAATACACCCAAAGCAAAGGTTACCAGTGAGAGATAT
KI VP1 118F     AGGCCCAACTACCTATACAATTAATCCACCAGGAGACCCTAGAACACTGCATGTGTACAATAGTAATACACCCAAAGCAAAGGTTACCAGTGAGAGATAT
pEGFP-C1 F      AGGCCCAACTACCTATACAATTAATCCACCAGGAGACCCTAGAACACTGCATGTGTACAATAGTAATACACCCAAAGCAAANGTTACCAGNGAGAGATAT
KI VP1 324R
EGFP-C1 Start
EGFP-C1 End     ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

                       1510      1520      1530      1540      1550      1560      1570      1580      1590      1600
                ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
KI partial      TCTGTTGAATCATGGGCCCCAGACCCCAGTAGAAATGACAATTGTAGATATTTTGGAAGAGTGGTAGGTGGTGCTGCAACACCTCCAGTTGTATCATATG
pEGFP-C1 R      TCTGTTGAATCATGGGCCCCAGACCCCAGTAGAAATGACAATTGTAGATATTTTGGAAGAGTGGTAGGTGGTGCTGCAACACCTCCAGTTGTATCATATG
KI VP1 118F     TCTGTTGAATCATGGGCCCCAGACCCCAGTAGAAATGACAATTGTAGATATTTTGGAAGAGTGGTAGGTGGTGCTGCAACACCTCCAGTTGTATCATATG
pEGFP-C1 F      TCTGTTGAANCATGGGCCCCAGACCCCAGNAGAAATGACNATTGNAGATATTTTGGNANANNGGNNNNTGGNNCNGNNNN~CCTNCANTNGTATCANANG
KI VP1 324R
EGFP-C1 Start
EGFP-C1 End     ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

                       1610      1620      1630      1640      1650      1660      1670      1680      1690      1700
                ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
KI partial      GTAACAACTCTACTATTCCACTATTGGATGAAAATGGCATTGGTATACTTTGCTTGCAGGGAAGATTGTACATTACTTGTGCAGATATGCTTGGAACAGC
pEGFP-C1 R      GTAACAACTCTACTATTCCACTATTGGATGAAAATGGCATTGGTATACCTTGCTTGCAGGGAAGATTGTACATTACTTGTGCAGATATGCTTGGAACAGC
KI VP1 118F     GTAACAACTCTACTATTCCACTATTGGATGAAAATGGCATTGGTATACCTTGCTTGCAGGGAAGATTGTACATTACTTGTGCAGATATGCTTGGAACAGC
pEGFP-C1 F      GN~ACNNCTCNACNA
KI VP1 324R
EGFP-C1 Start
EGFP-C1 End     ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

                       1710      1720      1730      1740      1750      1760      1770      1780      1790      1800
                ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
KI partial      TAATAGTAGAATCCATACCCCTATGGCTAGGTTTTTTAGGCTACATTTTAGACAAAGAAGGGTTAAGAATCCTTTTACAATGAATGTGCTGTATAAACAA
pEGFP-C1 R      TAATAGTAGAATCCATACCCCTATGGCTAGGTTTTTTAGGCTACATTTTAGACAAAGAAGGGTTAAGAATCCTTTTACAATGAATGTGCTGTATAAACAA
KI VP1 118F     TAATAGTAGAATCCATACCCCTATGGCTAGGTTTTTTAGGCTACATTTTAGACAAAGAAGGGTTAAGAATCCTTTTACAATGAATGTGCTGTATAAACAA
pEGFP-C1 F
KI VP1 324R
EGFP-C1 Start
EGFP-C1 End     ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

                       1810      1820      1830      1840      1850      1860      1870      1880      1890      1900
                ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
KI partial      GTGTTTAACAGACCCACAGAAACTGTTGATGCACAGGTTGGTGTAACAGAGGTGACTATGGTAGAAGAAATAGGCCCACTGCCCCCCAGTATACAAACTA
pEGFP-C1 R      GTGTTTAACAGACCCACAGAAACTGTTGATGCACAGGTTGGTGTAACAGAGGTGACTATGGTAGAAGAAATAGGCCCACTGCCCCCCAGTATACAAACTA
KI VP1 118F     GTGTTTAACAGACCCACAGAAACTGTTGATGCACAGGTTGGTGTAACAGANGTGACTATGGTAGAAGAAATAGGCCCACTGCCCCCCAGTATACAAACTA
pEGFP-C1 F
KI VP1 324R
EGFP-C1 Start
EGFP-C1 End     ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
```

```
                  1910      1920      1930      1940      1950      1960      1970      1980      1990      2000
              ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
KI partial    CCCTCCCCACCAGTGTAAATCTTACTCAGCTTCCACGCACTGTAACACTTCAGTCCCAGGCTCCTTTGTTAAATACTCAACAAAATTCAAAGTGAATGTA
pEGFP-C1 R    CCCTCCCCACCAGTGTAAATCTTACTCAGCTTCCACGCACTGTAACACTTCAGTCCCAGGCTCCTTTGTTAAATACTCAACAAAATTCAAAGTGAACCGC
KI VP1 118F   CCCTCCCCACCAGTGTAAATCTTACTCAGCTTCCACGCACTGTAACACTTCAGTCCCAGGCTCCTTTGTTAAATACTCAACAAAATTCAAAGTGAACCGC
pEGFP-C1 F
KI VP1 324R
EGFP-C1 Start
EGFP-C1 End   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ACCGC

                  2010      2020      2030      2040      2050      2060      2070      2080      2090      2100
              ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
KI partial    TTATAATAAAGCTGTTTATTCAAACCACTTTTCTAGTATGTTTTTTCCATTGGTTACATTCATTTGCATAGTGGCAAACTCAGTAAGGCCTATATAATTA
pEGFP-C1 R    GGGCCCGGGATCCACCGGATCTAGATAACTGATCATAATCAGCCATACCACATTTGTAGAGGTTTTACTTGCTTTAAAAAACCTCCCACACCTCCCCCTG
KI VP1 118F   GGGCCCNGGNNCCACCNGATCTAGANN~NTGATCATAN~CAGNCATACCACATTT
pEGFP-C1 F
KI VP1 324R
EGFP-C1 Start
EGFP-C1 End   GGGCCCGGGATCCACCGGATCTAGATAACTGATCATAATCAGCCATACCACATTTGTAGAGGTTTTACTTGCTTTAAAAAACCTCCCACACCTCCCCCTG

                  2110      2120      2130      2140      2150      2160      2170      2180      2190      2200
              ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
KI partial    TCAAGAACTTCTTTCCAATACACAACATTACTTTGAATTTCTTCATCAAAGTCACTTACAGGTCTATACCATATTAGCAATAACAACATAGCTATTCCAC
pEGFP-C1 R    AACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATA
KI VP1 118F
pEGFP-C1 F
KI VP1 324R
EGFP-C1 Start
EGFP-C1 End   AACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATA

                  2210      2220      2230      2240      2250      2260      2270      2280      2290      2300
              ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
KI partial    TGTGTAATATTCTTTGTGATAGTAACTGTGGTGTTTTTTCTAGGCTTTCTCTTAAATTTCTTTTAATAGTGAACATAACAGTTTTTTCAAATCTCACAGC
pEGFP-C1 R    AAGCATTTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAANTCATCAATGTATCTTAACGCGTAAATTGTAAGCGTTAATATTTTGTTAAAANTCGNNT
KI VP1 118F
pEGFP-C1 F
KI VP1 324R
EGFP-C1 Start
EGFP-C1 End   AAGCATTTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTAACGCGTAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGT

                  2310      2320      2330      2340      2350      2360      2370      2380
              ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
KI partial    AACTGTTTCTGGAATACAATATTCATTCATGGTTACAATCCCTGGGGGAAATATTTGTGA
pEGFP-C1 R    AAA
KI VP1 118F
pEGFP-C1 F
KI VP1 324R
EGFP-C1 Start
EGFP-C1 End   TAAATTTTTGTTAAATCAGCTCATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACC                    EGFP-C1R
```

Supplementary Figure 5: GST-KI VP1 sequence

```
                                      10        20        30        40        50        60        70        80        90       100
                              ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
gi|170280087 KI, VP1 start    ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ATGAGCTGCACCCCGTGTCGCCCACAAAAAAGGCTAACAAGGCCAAGAA
GEX-4T3/KI VP1 2A             CCTCCAAAATCNNATCTGGTTCCGCGTGGATCCCCGAATTCCCGGGTCGACATGAGCTGCACCCCGTGTCGCCCACAAAAAAGGCTAACAAGGCCAAGAA
GEX-4T3/KI VP1                ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
pGEX-4T-3, 4968 bp            CCTCCAAAATCGGATCTGGTTCCGCGTGGATCCCCGAATTCCCGGGTCGAC~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

                                     110       120       130       140       150       160       170       180       190       200
                              ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
gi|170280087 KI, VP1 start    GTCAAGTTCCCCGGGTACAAACTCTGGCAACAGAAGTAAAAAAAGGAGGGGTAGAAGTACTAGCCGCAGTACCACTGTCAGAAGAAACAGAATTTAAAGT
GEX-4T3/KI VP1 2A             GTCAAGTTCCCCGGGTACAAACTCTGGCAACAGAAGTAAAAAAAGGAGGGGTAGAAGTACTAGCCGCAGTACCACTGTCAGAAGAAACAGAATTTAAAGT
GEX-4T3/KI VP1                ~~~~~~~~~~~~~~~~~~~~~~CCCTCAGGCACGAGTAAAAAA~GGAGGGGTAGAAGTACTAGCCGCAGTACCACTGTCAGAAGAAACAGAATTTAAAGT
pGEX-4T-3, 4968 bp            ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

                                     210       220       230       240       250       260       270       280       290       300
                              ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
gi|170280087 KI, VP1 start    GGAACTATTTGTAAAGCCAGTAATTGGAAATACAACAGCTGCTCAGGATGGGCGTGAGCCCACCCCTCATTACTGGTCAATTAGCTCTGCCATTCATGAC
GEX-4T3/KI VP1 2A             GGAACTATTTGTAAAGCCAGTAATTGGAAATACAACAGCTGCTCAGGATGGGCGTGAGCCCACCCCTCATTACTGGTCAATTAGCTCTGCCATTCATGAC
GEX-4T3/KI VP1                GGAACTATTTGTAAAGCCAGTAATTGGAAATACAACAGCTGCTCAGGATGGGCGTGAGCCCACCCCTCATTACTGGTCAATTAGCTCTGCCATTCATGAC
pGEX-4T-3, 4968 bp            ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

                                     310       320       330       340       350       360       370       380       390       400
                              ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
gi|170280087 KI, VP1 start    AAGGAAAGCGGTTCAAGTATCAAAGTTGAAGAAACTCCAGATGCTGACACAACTGTATGTTACAGCCTGGCAGAAATTGCTCCCCCTGATATACCAAATC
GEX-4T3/KI VP1 2A             AAGGAAAGCGGTTCAAGTATCAAAGTTGAAGAAACTCCAGATGCTGACACAACTGTATGTTACAGCCTGGCAGAAATTGCTCCCCCTGATATACCAAATC
GEX-4T3/KI VP1                AAGGAAAGCGGTTCAAGTATCAAAGTTGAAGAAACTCCAGATGCTGACACAACTGTATGTTACAGCCTGGCAGAAATTGCTCCCCCTGATATACCAAATC
pGEX-4T-3, 4968 bp            ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

                                     410       420       430       440       450       460       470       480       490       500
                              ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
gi|170280087 KI, VP1 start    AAGTTAGTGAGTGTGACATGAAAGTATGGGAGCTGTACAGAATGGAAACAGAGTTGCTTGTTGTACCTCTAGTTAATGCTCTAGGAAACACCAATGGTGT
GEX-4T3/KI VP1 2A             AAGTTAGTGAGTGTGACATGAAAGTATGGGAGCTGTACAGAATGGAAACAGAGTTGCTTGTTGTACCTCTAGTTAATGCTCTAGGAAACACCAATGGTGT
GEX-4T3/KI VP1                AAGTTAGTGAGTGTGACATGAAAGTATGGGAGCTGTACAGAATGGAAACAGAGTTGCTTGTTGTACCTCTAGTTAATGCTCTAGGAAACACCAATGGTGT
pGEX-4T-3, 4968 bp            ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

                                     510       520       530       540       550       560       570       580       590       600
                              ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
gi|170280087 KI, VP1 start    TGTACATGGTCTGGCTGGAACCCAGTTGTATTTTTGGGCTGTTGGGGGACAGCCACTTGATGTAGTTGGTGTAACACCCACAGACAAGTATAAAGGCCCA
GEX-4T3/KI VP1 2A             TGTACATGGTCTGGCTGGAACCCAGTTGTATTTTTGGGCTGTTGGGGGACAGCCACTTGATGTAGTTGGTGTAACACCCACAGACAAATATAAAGGCCCA
GEX-4T3/KI VP1                TGTACATGGTCTGGCTGGAACCCAGTTGTATTTTTGGGCTGTTGGGGGACAGCCACTTGATGTAGTTGGTGTAACACCCACAGACAAATATAAAGGCCCA
pGEX-4T-3, 4968 bp            ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

                                     610       620       630       640       650       660       670       680       690       700
                              ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
gi|170280087 KI, VP1 start    ACTACCTATACAATTAATCCACCAGGAGACCCTAGAACACTGCATGTGTACAATAGTAATACACCCAAAGCAAAGGTTACCAGTGAGAGATATTCTGTTG
GEX-4T3/KI VP1 2A             ACTACCTATACAATTAATCCACCAGGAGACCCTAGAACACTGCATGTGTACAATAGTAATACACCCAAAGCAAAGGTTACCAGTGAGAGATATTCTGTTG
GEX-4T3/KI VP1                ACTACCTATACAATTAATCCACCAGGAGACCCTAGAACACTGCATGTGTACAATAGTAATACACCCAAAGCAAAGGTTACCAGTGAGAGATATTCTGTTG
pGEX-4T-3, 4968 bp            ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
```

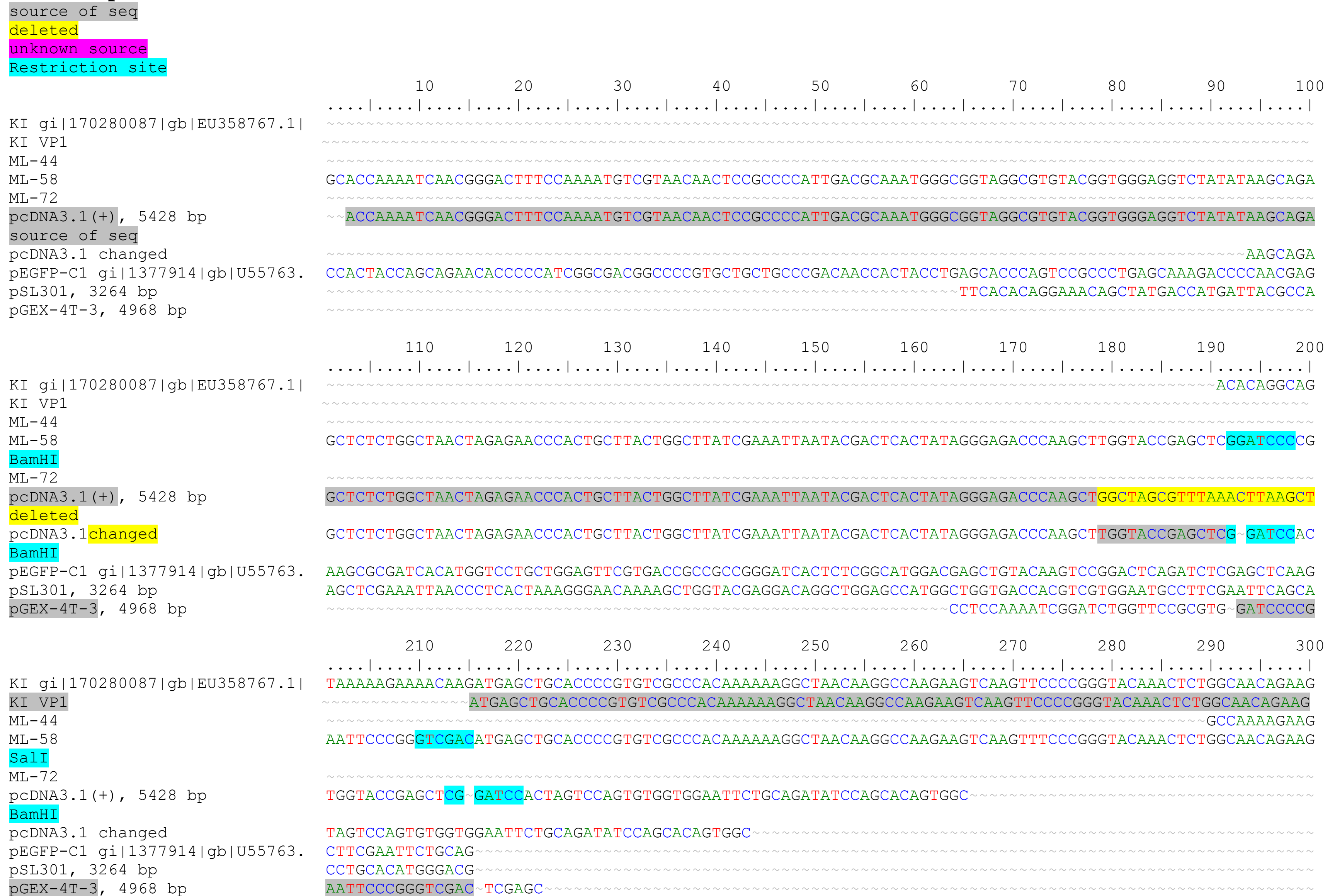
```
                                    710       720       730       740       750       760       770       780       790       800
                            ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
gi|170280087 KI, VP1 start  AATCATGGGCCCCAGACCCCAGTAGAAATGACAATTGTAGATATTTTGGAAGAGTGGTAGGTGGTGCTGCAACACCTCCAGTTGTATCATATGGTAACAA
GEX-4T3/KI VP1 2A           AATCATGGGCCCCAGACCCCAGTAGAAATGACAATTGTAGATATTTTGGAAGAGTGGTAGGTGGTGCTGCAACACCTCCAGTTGTATCATATGGTAACAA
GEX-4T3/KI VP1              AATCATGGGCCCCAGACCCCAGTAGAAATGACAATTGTAGATATTTTGGAAGAGTGGTAGGTGGTGCTGCAACACCTCCAGTTGTATCATATGGTAACAA
pGEX-4T-3, 4968 bp          ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

                                    810       820       830       840       850       860       870       880       890       900
                            ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
gi|170280087 KI, VP1 start  CTCTACTATTCCACTATTGGATGAAAATGGCATTGGTATACTTTGCTTGCAGGGAAGATTGTACATTACTTGTGCAGATATGCTTGGAACAGCTAATAGT
GEX-4T3/KI VP1 2A           CTCTACTATTCCACTATTGGATGAAAATGGCATTGGTATACCTTGCTTGCAGGGAAGATTGTACATTACTTGTGCAGATATGCTTGGAACAGCTAATAGT
GEX-4T3/KI VP1              CTCTACTATTCCACTATTGGATGAAAATGGCATTGGTATACCTTGCTTGCAGGGAAGATTGTACATTACTTGTGCAGATATGCTTGGAACAGCTAATAGT
pGEX-4T-3, 4968 bp          ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

                                    910       920       930       940       950       960       970       980       990      1000
                            ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
gi|170280087 KI, VP1 start  AGAATCCATACCCCTATGGCTAGGTTTTTTAGGCTACATTTTAGACAAAGAAGGGTTAAGAATCCTTTTACAATGAATGTGCTGTATAAACAAGTGTTTA
GEX-4T3/KI VP1 2A           AGAATCCATACCCCTATGGCTAGGTTTTTTAGGCTACATTTTANANNAANAAGGGTTAANAATCCTTTTNNNNTGAATGTGCTGNATAAACAAGTNNTTA
GEX-4T3/KI VP1              AGAATCCATACCCCTATGGCTAGGTTTTTTAGGCTACATTTTAGACAAAGAAGGGTTAAGAATCCTTTTACAATGAATGTGCTGTATAAACAAGTGTTTA
pGEX-4T-3, 4968 bp          ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

                                   1010      1020      1030      1040
                            ....|....|....|....|....|....|....|....|
gi|170280087 KI, VP1 start  ACAGACCCACAGAAACTGTTGATGC
GEX-4T3/KI VP1 2A           ACANACCCNCN~NANCNGTTGATGC
GEX-4T3/KI VP1              ACAGACCCACAGAAACTGTTGATGCACAGGTTGGTGTAAC
pGEX-4T-3, 4968 bp          ~~~~~~~~~~~~~~~~~~~~~~~~~
```

Supplementary Figure 6: pcDNA3.1(+)KI VP1 sequence. ML-44, 58 and 72 are sequences of three different primers used on the same plasmid



```
                                        310       320       330       340       350       360       370       380       390       400
                                ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
KI gi|170280087|gb|EU358767.1|  TAAAAAAAGGAGGGGTAGAAGTACTAGCCGCAGTACCACTGTCAGAAGAAACAGAATTTAAAGTGGAACTATTTGTAAAGCCAGTAATTGGAAATACAAC
KI VP1                          TAAAAAAAGGAGGGGTAGAAGTACTAGCCGCAGTACCACTGTCAGAAGAAACAGAATTTAAAGTGGAACTATTTGTAAAGCCAGTAATTGGAAATACAAC
ML-44                           TAAAAAA~GGAGGGGTAGAAGTACTAGCCGCAGTACCACTGTCAGAAGAAACAGAATTTAAAGTGGAACTATTTGTAAAGCCAGTAATTGGAAATACAAC
ML-58                           TAAAAAAAGGAGGGGTAGAAGTACTAGCCGCAGTACCACTGTCAGAAGAAACAGAATTTAAAGTGGAACTATTTGTAAAGCCAGTAATTGGAAATACAAC
ML-72                           ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
pcDNA3.1(+), 5428 bp            ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
pcDNA3.1 changed                ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
pEGFP-C1 gi|1377914|gb|U55763.  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
pSL301, 3264 bp                 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
pGEX-4T-3, 4968 bp              ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

                                        410       420       430       440       450       460       470       480       490       500
                                ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
KI gi|170280087|gb|EU358767.1|  AGCTGCTCAGGATGGGCGTGAGCCCACCCCTCATTACTGGTCAATTAGCTCTGCCATTCATGACAAGGAAAGCGGTTCAAGTATCAAAGTTGAAGAAACT
KI VP1                          AGCTGCTCAGGATGGGCGTGAGCCCACCCCTCATTACTGGTCAATTAGCTCTGCCATTCATGACAAGGAAAGCGGTTCAAGTATCAAAGTTGAAGAAACT
ML-44                           AGCTGCTCAGGATGGGCGTGAGCCCACCCCTCATTACTGGTCAATTAGCTCTGCCATTCATGACAAGGAAAGCGGTTCAAGTATCAAAGTTGAAGAAACT
ML-58                           AGCTGCTCAGGATGGGCGTGAGCCCACCCCTCATTACTGGTCAATTAGCTCTGCCATTCATGACAAGGAAAGCGGTTCAAGTATCAAAGTTGAAGAAACT
ML-72                           ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
pcDNA3.1(+), 5428 bp            ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
pcDNA3.1 changed                ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
pEGFP-C1 gi|1377914|gb|U55763.  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
pSL301, 3264 bp                 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
pGEX-4T-3, 4968 bp              ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

                                        510       520       530       540       550       560       570       580       590       600
                                ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
KI gi|170280087|gb|EU358767.1|  CCAGATGCTGACACAACTGTATGTTACAGCCTGGCAGAAATTGCTCCCCCTGATATACCAAATCAAGTTAGTGAGTGTGACATGAAAGTATGGGAGCTGT
KI VP1                          CCAGATGCTGACACAACTGTATGTTACAGCCTGGCAGAAATTGCTCCCCCTGATATACCAAATCAAGTTAGTGAGTGTGACATGAAAGTATGGGAGCTGT
ML-44                           CCAGATGCTGACACAACTGTATGTTACAGCCTGGCAGAAATTGCTCCCCCTGATATACCAAATCAAGTTAGTGAGTGTGACATGAAAGTATGGGAGCTGT
ML-58                           CCAGATGCTGACACAACTGTATGTTACAGCCGGCAGAATGCTCCC
ML-72                           ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
pcDNA3.1(+), 5428 bp            ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
pcDNA3.1 changed                ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
pEGFP-C1 gi|1377914|gb|U55763.  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
pSL301, 3264 bp                 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
pGEX-4T-3, 4968 bp              ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

                                        610       620       630       640       650       660       670       680       690       700
                                ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
KI gi|170280087|gb|EU358767.1|  ACAGAATGGAAACAGAGTTGCTTGTTGTACCTCTAGTTAATGCTCTAGGAAACACCAATGGTGTTGTACATGGTCTGGCTGGAACCCAGTTGTATTTTTG
KI VP1                          ACAGAATGGAAACAGAGTTGCTTGTTGTACCTCTAGTTAATGCTCTAGGAAACACCAATGGTGTTGTACATGGTCTGGCTGGAACCCAGTTGTATTTTTG
ML-44                           ACAGAATGGAAACAGAGTTGCTTGTTGTACCTCTAGTTAATGCTCTAGGAAACACCAATGGTGTTGTACATGGTCTGGCTGGAACCCAGTTGTATTTTTG
ML-58
ML-72                           ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
pcDNA3.1(+), 5428 bp            ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
pcDNA3.1 changed                ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
pEGFP-C1 gi|1377914|gb|U55763.  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
pSL301, 3264 bp                 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
pGEX-4T-3, 4968 bp              ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
```

```
                                         710       720       730       740       750       760       770       780       790       800
                                 ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
KI gi|170280087|gb|EU358767.1|   GGCTGTTGGGGGACAGCCACTTGATGTAGTTGGTGTAACACCCACAGACAAGTATAAAGGCCCAACTACCTATACAATTAATCCACCAGGAGACCCTAGA
KI VP1                           GGCTGTTGGGGGACAGCCACTTGATGTAGTTGGTGTAACACCCACAGACAAGTATAAAGGCCCAACTACCTATACAATTAATCCACCAGGAGACCCTAGA
ML-44                            GGCTGTTGGGGGACAGCCACTTGATGTAGTTGGTGTAACACCCACAGACAAATATAAAGGCCCAACTACCTATACAATTAATCCACCAGGAGACCCTAGA
ML-58
ML-72                            ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
pcDNA3.1(+), 5428 bp             ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
pcDNA3.1 changed                 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
pEGFP-C1 gi|1377914|gb|U55763.   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
pSL301, 3264 bp                  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
pGEX-4T-3, 4968 bp               ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

                                         810       820       830       840       850       860       870       880       890       900
                                 ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
KI gi|170280087|gb|EU358767.1|   ACACTGCATGTGTACAATAGTAATACACCCAAAGCAAAGGTTACCAGTGAGAGATATTCTGTTGAATCATGGGCCCCAGACCCCAGTAGAAATGACAATT
KI VP1                           ACACTGCATGTGTACAATAGTAATACACCCAAAGCAAAGGTTACCAGTGAGAGATATTCTGTTGAATCATGGGCCCCAGACCCCAGTAGAAATGACAATT
ML-44                            ACACTGCATGTGTACAATAGTAATACACCCAAAGCAAAGGTTACCAGTGAGAGATATTCTGTTGAATCATGGGCCCCAGACCCCAGTAGAAATGACAATT
ML-58
ML-72                            ~~~~~~~~~~~~~~~~~~~~~AATACACCCAAAGCAAAGGTTACCAGTGAGAGATATTCTGTTGAATCATGGGCCCCAGACCCCAGTAGAAATGACAATT
pcDNA3.1(+), 5428 bp             ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
pcDNA3.1 changed                 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
pEGFP-C1 gi|1377914|gb|U55763.   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
pSL301, 3264 bp                  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
pGEX-4T-3, 4968 bp               ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

                                         910       920       930       940       950       960       970       980       990      1000
                                 ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
KI gi|170280087|gb|EU358767.1|   GTAGATATTTTGGAAGAGTGGTAGGTGGTGCTGCAACACCTCCAGTTGTATCATATGGTAACAACTCTACTATTCCACTATTGGATGAAAATGGCATTGG
KI VP1                           GTAGATATTTTGGAAGAGTGGTAGGTGGTGCTGCAACACCTCCAGTTGTATCATATGGTAACAACTCTACTATTCCACTATTGGATGAAAATGGCATTGG
ML-44                            GTAGATATTTTGGAAGAGTGGTAGGTGGTGCTGCAACACCTCCAGTTGTATCATATGGTAACAACTCTACTATTCCACTATTGGATGAAAATGGCATTGG
ML-58
ML-72                            GTAGATATTTTGGAAGAGTGGTAGGTGGTGCTGCAACACCTCCAGTTGTATCATATGGTAACAACTCTACTATTCCACTATTGGATGAAAATGGCATTGG
pcDNA3.1(+), 5428 bp             ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
pcDNA3.1 changed                 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
pEGFP-C1 gi|1377914|gb|U55763.   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
pSL301, 3264 bp                  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
pGEX-4T-3, 4968 bp               ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

                                        1010      1020      1030      1040      1050      1060      1070      1080      1090      1100
                                 ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
KI gi|170280087|gb|EU358767.1|   TATACTTTGCTTGCAGGGAAGATTGTACATTACTTGTGCAGATATGCTTGGAACAGCTAATAGTAGAATCCATACCCCTATGGCTAGGTTTTTTAGGCTA
KI VP1                           TATACTTTGCTTGCAGGGAAGATTGTACATTACTTGTGCAGATATGCTTGGAACAGCTAATAGTAGAATCCATACCCCTATGGCTAGGTTTTTTAGGCTA
ML-44                            TATACCTTGCTTGCAGGGAAGATTGTACATTACTTGTGCAGATATGCTTGGAACAGCTAATAGTAGAATCCATACCCCTATGGCTAGGTTTTTTAGGCTA
ML-58
ML-72                            TATACCTTGCTTGCAAGGAACATTGTACATTACTTGTGCATATATGCTTGGAACAGCTAATAATAGAATCCATACCCCTATGGCTAGGTTTTTTAGGCTA
pcDNA3.1(+), 5428 bp             ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
pcDNA3.1 changed                 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
pEGFP-C1 gi|1377914|gb|U55763.   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
pSL301, 3264 bp                  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
pGEX-4T-3, 4968 bp               ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
```

```
                                         1110      1120      1130      1140      1150      1160      1170      1180      1190      1200
                                 ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
KI gi|170280087|gb|EU358767.1|   CATTTTAGACAAAGAAGGGTTAAGAATCCTTTTACAATGAATGTGCTGTATAAACAAGTGTTTAACAGACCCACAGAAACTGTTGATGCACAGGTTGGTG
KI VP1                           CATTTTAGACAAAGAAGGGTTAAGAATCCTTTTACAATGAATGTGCTGTATAAACAAGTGTTTAACAGACCCACAGAAACTGTTGATGCACAGGTTGGTG
ML-44                            CATTTTAGACAAAGAAGGGTTAAGAATCCTTTTACAATGAATGTGCTGTATAAACAAGTGTTTAACAGACCCACAGAAACTGTTGATGCACAGGTTGGTG
ML-58
ML-72                            CATTTTACACAAAGAAGGGTTAAGAATCCTTTTACAATGAATGTGCTGTATAAACAAGTGTTTAACAGACCCACATAAACTGTTGATGCACAGGTTGGTG
pcDNA3.1(+), 5428 bp             ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
pcDNA3.1 changed                 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
pEGFP-C1 gi|1377914|gb|U55763.   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
pSL301, 3264 bp                  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
pGEX-4T-3, 4968 bp               ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

                                         1210      1220      1230      1240      1250      1260      1270      1280      1290      1300
                                 ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
KI gi|170280087|gb|EU358767.1|   TAACAGAGGTGACTATGGTAGAAGAAATAGGCCCACTGCCCCCCAGTATACAAACTACCCTCCCCACCAGTGTAAATCTTACTCAGCTTCCACGCACTGT
KI VP1                           TAACAGAGGTGACTATGGTAGAAGAAATAGGCCCACTGCCCCCCAGTATACAAACTACCCTCCCCACCAGTGTAAATCTTACTCAGCTTCCACGCACTGT
ML-44                            TAACAGAGGTGACTATGGTAGAAGAAATAGGCCCACTGCCCCCCAGTATACAAACTACCCTCCCCACCAGTGTAAATCTTACTCAGCTTCCACGCACTGT
ML-58
ML-72                            TAACAGAGGTGACTATGGTATAAGAAATAGGCCCACTGCCCCCCAGTATACAAACTACCCTCCCCACCAGTGTAAATCTTACTCAGCTTCCACGCACTGT
pcDNA3.1(+), 5428 bp             ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
pcDNA3.1 changed                 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
pEGFP-C1 gi|1377914|gb|U55763.   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
pSL301, 3264 bp                  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
pGEX-4T-3, 4968 bp               ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
                                         1310      1320      1330      1340      1350      1360      1370      1380      1390      1400
                                 ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
KI gi|170280087|gb|EU358767.1|   AACACTTCAGTCCCAGGCTCCTTTGTTAAATACTCAACAAAATTCAAAGTGAATGTATTATAATAAAGCTGTTTATTCAAACCACTTTTCTAGTATGTTT
KI VP1                           AACACTTCAGTCCCAGGCTCCTTTGTTAAATACTCAACAAAATTCAAAGTGA
SacII
ML-44                            AACACTTCAGTCCCAGGCTCCTTTGTTAAATACTCAACAAAATTCAAAGTGAACCGCGGGCCCGGGATCCAATTGCCATTGATCATCATGAACGATCTGC
BamHI
ML-58
ML-72                            AACACTTCAGTCCCAGGCTCCTTTGTTAAATACTCAACAAAATTCAAAGTGAACCGCGGGCCCGGGATCCAATTGCAATGATCATCATGACAGATCTGCG
pcDNA3.1(+), 5428 bp             ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
pcDNA3.1 changed                 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
pEGFP-C1 gi|1377914|gb|U55763.   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~TCGACGGT~ACCGCGGGCCCGGGATCCACCGGATCTAGATAACTGATCATAATCAGC
pSL301, 3264 bp                  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~GATCCAATTGCAATGATCATCATGACAGATCTGCG
pGEX-4T-3, 4968 bp               ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
                                         1410      1420      1430      1440      1450      1460      1470      1480      1490      1500
                                 ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
KI gi|170280087|gb|EU358767.1|   TTTCCATTGGTTACATTCATTTGCATAGTGGCAAACTCAGTAAGGCCTATATAATTATCAAGAACTTCTTTCCAATACACAACATTACTT
KI VP1
ML-44                            GCCGCGAATCGATATCGACGCTTAATTGCCCAATGGCTTAGCTCTTATATG
ML-58
ML-72                            CGCGATCGATATCAGCGCTTTAAATTTGCGCATGCTAGCTATAGTTCTAGAGGTACCGGTTGTTAACGTTAGCCGGCTACGTATACTCCGGAATATTAAT
pcDNA3.1(+), 5428 bp             ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
pcDNA3.1 changed                 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
pEGFP-C1 gi|1377914|gb|U55763.   CATACCACATTTGTAGAGGTTTTACTTGCTTTAAAAAACCTCCCACACCTCCCCCTGAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTGT
pSL301, 3264 bp                  CGCGATCGATATCAGCGCTTTAAATTTGCGCATGCTAGCTATAGTTCTAGAGGTACCGGTTGTTAACGTTAGCCGGCTACGTATACTCCGGAATATTAAT
pGEX-4T-3, 4968 bp               ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
```

```
                                                 1510      1520      1530      1540      1550      1560      1570      1580      1590      1600
                                        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
KI gi|170280087|gb|EU358767.1|
KI VP1
ML-44
ML-58
NotI
ML-72                                   AGGCCTAGGATGCATATGGCGGCCGCTCGAGCATGCATCTAGAGGGCCCTATTCTATAGTGTCACCTAAATGCTAGAGCTCGCTGATCAGCCTCGACTGT
unknown source
pcDNA3.1(+), 5428 bp                    ~~~~~~~~~~~~~~~~~~~~GGCCGCTCGAGTCTAGAGGGCCCGTTTAAACCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTT
deleted
pcDNA3.1 changed                         ~~~~~~~~~~~~~~~~~~~~GGCCGCTCGAG                                                CGCTGATCAGCCTCGACTGT
pEGFP-C1 gi|1377914|gb|U55763.          TTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACT
pSL301, 3264 bp                         AGGCCTAGGATGCATATGGCGGCCGCCTGCAGCTGGCGCCATCGATACGCGTACGTCGCGACCGCGGACATGTACAGAGCTCGAGAAGTACTAGTGGCCA
pGEX-4T-3, 4968 bp                      ~~~~~~~~~~~~~~~~~~~~GGCCGCATCGTGACTGACTGACGATCTGCCTCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCG

                                                 1610      1620      1630      1640      1650      1660      1670      1680      1690      1700
                                        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
KI gi|170280087|gb|EU358767.1|
KI VP1
ML-44
ML-58
ML-72                                   GCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAA
pcDNA3.1(+), 5428 bp                    TGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTC
pcDNA3.1 changed                        GCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAA
pEGFP-C1 gi|1377914|gb|U55763.          CATCAATGTATCTTAACGCGTAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTTTTAACCAATAGGCCGA
pSL301, 3264 bp                         CGTGGGCCGTGCACCTTAAGCTTGCCTAATCGGACGAAAAAATGACCATGATTACGCCAAGCTCCAATTCGCCCTATAGTGAGTCGTATTACAATTCACT
pGEX-4T-3, 4968 bp                      GAGACGGTCACAGCTTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCGCAGCCATGACCC

                                                 1710      1720      1730      1740      1750      1760      1770
                                        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
KI gi|170280087|gb|EU358767.1|
KI VP1
ML-44
ML-58
ML-72                                   ATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGG
pcDNA3.1(+), 5428 bp                    ATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGC
pcDNA3.1 changed                        ATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGG
pEGFP-C1 gi|1377914|gb|U55763.          AATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGTTCCAGTTTGGAAC
pSL301, 3264 bp                         GGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACAT
pGEX-4T-3, 4968 bp                      AGTCACGTAGCGATAGCGGAGTGTATAATTCTTGAAGACGAAAGGGCCTCGTGATACGCCTATTTTTATA
```

Supplementary Figure 7: Sequencing results of positive VP1 PCR products

Positive CSF samples after VP1 PCR1

```
                                        310       320       330       340       350       360       370       380
                               ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
KI VP1 gi|170280087|gb|EU35876 ACTGTATGTTACAGCCTGGCAGAAATTGCTCCCCCTGATATACCAAATCAAGTTAGTGAGTGTGACATGAAAGTATGGGA      KI VP1 363R
ML-83                          ACTGTATGTTACAGCCTGGCAGAAATTGCTCCCCCTGATATACCAAATCAAGTTAGTGAGTGT
ML-84                          ACTGTATGTTACAGCCTGGCAGAAATTGCTCCCCCTGATATACCAAATCAAGTTAGTGAGTGT
ML-85                          ACTGTATGTTACAGCCTGGCAGAAATTGCTCCCCCTGATATACCAAATCAAGTTAGTGAGTGT
ML-86                          ACTGTATGTTACAGCCTGGCAGAAATTGCTCCCCCTGATATACCAAATCAAGTTAGTGAGTGT
ML-87                          ACTGTATGTTACAGCCTGGCAGAAATTGCTCCCCCTGATATACCAAATCAAGTTAGTGAGTGT
ML-89                          ACTGTATGTTACAGCCTGGCAGAAATTGCTCCCCCTGATATACCAAATCAAGTTAGTGAGTGT
ML-93                          ACTGTATGTTACAGCCTGGCAG
ML-94                          ACTGTATGTTACAGCCTGGC
ML-95                          ACTGTATGTTACAGCCTGGC
ML-96                          ACTGTATGTTACAGCC
ML-97                          ACTGTATGTTACAGCCTGGCAGA
ML-99                          ACTGTATGTTACCGCC
```

Sequence alignment of one CSF and two urine positive nested VP1 PCR products:

```
                                        10        20        30        40        50        60        70        80        90       100
                               ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
KI partial VP1 gi|170280087|   ATGAGCTGCACCCCGTGTCGCCCACAAAAAAGGCTAACAAGGCCAAGAAGTCAAGTTCCCCGGGTACAAACTCTGGCAACAGAAGTAAAAAAAGGAGGGG
MAL-117                        ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
MAL-121                        ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
MAL-118                        ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
MAL-122                        ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
MAL-119                        ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
MAL-123                        ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

                                       110       120       130       140       150       160       170       180       190       200
                               ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
KI partial VP1gi|170280087|    TAGAAGTACTAGCCGCAGTACCACTGTCAGAAGAAACAGAATTTAAAGTGGAACTATTTGTAAAGCCAGTAATTGGAAATACAACAGCTGCTCAGGATGG
MAL-117                        ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~AGGAAACTGAATTGTAAGCCAGTAATTTAGAATACAACAGCTGCTCAGGATGG
MAL-121                        ~~~~~~~~~~~~TCGCAGTACCAGTGTCAGAAGAAACAGAATTTAAAGTGGAAGTATTTGTAAAGCCAGTAATTGGAAATACAACAGCTGCTCAGGATGG
MAL-118                        ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~GGAA~CTATTTGTAAAGCCAGTAATTGGAAATACAACAGCTGCTCAGGATGG
MAL-122                        ~~~~~~~~~~~~~CGCAGTACCACTGTCAGAAGAAACAGAATTTAAAGTGAA~CTATTTGTAAAGCCAGTAATTTGAAATCCAACAGCTGGTCAGGATGG
MAL-119                        ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~TGCGTACTATTTGTAAAGCCAGTAATTGGAAATACAACAGCTGCTCAGGATGG
MAL-123                        ~~~~~~~~~~~~~~GCAGTACCACTGTCAGAAGAAACAGAATTTAAAGTGGAAGTATTTGTAAAGCCAGTAATTGGAAATACAACAGCTGGTCAGGATGG

                                       210       220       230       240       250       260       270       280       290       300
                               ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
KI partial VP1gi|170280087|    GCGTGAGCCCACCCCTCATTACTGGTCAATTAGCTCTGCCATTCATGACAAGGAAAGCGGTTCAAGTATCAAAGTTGAAGAAACTCCAGATGCTGACACA
MAL-117                        GCGTGAGCCCACCCCTCATTACTGGTCAATTAGCTCTGCCATTCATGACAAGGAAAGCGGTTCAAGTATCAAAGTTGAAGAAACTCCAGATGCTGACACA
MAL-121                        GCGTGAGCCCACCCCTCATTACTGGTCAATTAGCTATGCCATTCATGACAAGGAAAGCGGTTCATAAGGCCAAGAAGTCAAGTCCAATGCAATC
MAL-118                        GCGTGAGCCCACCCCTCATTACTGGTCAATTAGCTCTGCCATTCATGACAAGGAAAGCGGTTCAAGTATCAAAGTTGAAGAAACTCCAGATGCTGACACA
MAL-122                        GCGTGAGCCCACCCCTCATTAATGGTCAATTAGCTA~GCCAATCAC~ACAAGGAAAGCGGACATAAG~TCAAAGAGGTGAAGTCCAATTCAACG
MAL-119                        GCGTGAGCCCACCCCTCATTACTGGTCAATTAGCTCTGCCATTCATGACAAGGAAAGCGGTTCAAGTATCAAAGTTGAAGAAACTCCAGATGCTGACACA
MAL-123                        GCGTGAGCCCACCCCTCATTACTGGTCAATTAGCTCTGCCATTCATGACAAGGAAAGCGGATCATGTATCAAAGTGAAGAATTCCAATTAAAAT
```

```
                                        310       320
                                 ....|....|....|....|....|....
KI partial VP1gi|170280087|      ACTGTATGTTACAGCCTGGCAGAAATTGC
MAL-117                          ACTGTATGTTACAGCCTGGCAGAAA
MAL-121
MAL-118                          ACTGTATGTTACAGCCTGGCAGAAA
MAL-122
MAL-119                          ACTGTATGTTACAGCCTGGCAGAAA
MAL-123
```

Supplementary Table 1: Prevalence of KIPyV. AdV: adenovirus, ARI: acute respiratory infection, BAL: Bronchoalveolar lavage, BM: Bone marrow, CAP: community acquired pneumonia, ELISA: Enzyme-linked immunoabsorbent assay, HBoV: human bocavirus, hCoV: human coronavirus, hMPV: human metapneumovirus, HSCT: hematopoietic stem cell transplant, hRSV: human respiratory syncytial virus, hRV: human rhinovirus, IFA: Immunofluorescens assay, MCC: Merkel cell carcinoma, IV A/B: Influenzavirus A or B, N: Total number, NPA: Nasopharyngeal aspirate, nPCR: nested PCR, MNC: mononuclear cells, MS: Multiple sclerosis, ONS: other neurological syndromes, PB: Peripheral blood, PIV 1-3: parainfluenzavirus type 1-3, RTI: respiratory tract infection, RTS: Respiratory tract secretions, SLE: Systemic lupus erythematosus, STD: sexually transmitted diseases, UTI: Urinary tract infection, qPCR: quantitative real-time PCR, * samples from Edinburgh respiratory specimen archive, ** Suspected chronic viral encephalitis including 6 HIV patients.

Sample type	Diagnosis	Age	N of samples (individuals)	Prevalence	Target	Co-infection	Method	Country	References
NPA	ARI	0-90 years	637	6/637 0.9 %	VP1		nPCR	Sweeden	Allander et al., 2007
Stool	Gastroenteritis	0-17 years	192	1/192 0.5 %	VP1		nPCR	Sweeden	Allander et al., 2007
Urine	HSCT recipients		150	0/150 0 %	VP1		nPCR	Sweeden	Allander et al., 2007
Serum	HSCT recipients		33 (17)	0/33 0 %	VP1		nPCR	Sweeden	Allander et al., 2007
Whole blood	Healthy		192	0/192 0 %	VP1		nPCR	Sweeden	Allander et al., 2007
Leukocytes	Mainly immunosuppressed		96	0/96 0 %	VP1		nPCR	Sweeden	Allander et al., 2007
NPA/ BAL	ARI	1 month-95 years	951	24/951 2.5 %	VP1	25 % RSV, IV A, hMPV	nPCR	Australia	Bialasiewicz et al., 2007a
NPA	ARI		200	13/200 6.5 %	VP1		RT-PCR	Australia	Bialasiewicz et al., 2007b
NPA	ARI	1 month-5 years	486	5/486 1 %	VP1	60% hRSV, RV, HBoV, PIV	PCR	South Korea	Han et al., 2007
NPA	Asymptomatic	1 month-6 years	72	0/72 0 %	VP1		PCR	South Korea	Han et al., 2007
RTS	*	0.3 – 34 years	983	14/983 1.4 %	VP1	40 % RSV, AdV, HBoV	nPCR	Scotland	Norja et al., 2007
Throat swabs	ARI	Pediatric	222	1/222 0.5 %	VP1	RSV, PIV 1-3, AdV, IV A+B, hCoV	PCR	Italy	Babakir-Mina et al., 2008
NPA	ARI	3 days-95 years	2866	75/2866 2.6 %	VP1	74.7 % HRV, HBoV	PCR	Australia	Bialasiewicz et al., 2008

Sample type	Diagnosis	Age	N of samples (individuals)	Prevalence	Target	Co-infection	Method	Country	References
Urine	Variable		215	0/215 0 %	VP1		PCR	Australia	Bialasiewicz et al., 2008
Whole blood	Immunocompromised	Pediatric	102	0/102 0 %	VP1		PCR	Australia	Bialasiewicz et al., 2008
Melanoma tissue	Cancer		36	0/36 0 %	VP1	No BKV, JCV or SV49	PCR	Sweeden	Giraud et al., 2008
NPA	ARI	<5 years	537	3/537 0.6 %	VP1	33 % RSV, HMPV	PCR	France	Foulongne et al., 2008
NPA	ARI	7 days-79 years	371	10/371 2.7 %	VP1	75 % hMPV, RSV, HBoV, PIV 1	qPCR	UK	Kiasari et al., 2008
NPA	ARI	<1 year	224	6/224 2.7 %	VP1	75 % hMPV, RSV, HBoV, PIV 1	qPCR	UK	Kiasari et al., 2008
NPA	ARI	1-5 years	58	1/58 1.7 %	VP1	75 % hMPV, RSV, HBoV, PIV 1	qPCR	UK	Kiasari et al., 2008
NPA	ARI	6-14 years	16	0/16 0 %	VP1	75 % hMPV, RSV, HBoV, PIV 1	qPCR	UK	Kiasari et al., 2008
NPA	ARI	15-29 years	11	0/11 0 %	VP1	75 % hMPV, RSV, HBoV, PIV 1	qPCR	UK	Kiasari et al., 2008
NPA	ARI	30-44 years	15	0/15 0 %	VP1	75 % hMPV, RSV, HBoV, PIV 1	qPCR	UK	Kiasari et al., 2008
NPA	ARI	45-60 years	26	1/26 3.8 %	VP1	75 % hMPV, RSV, HBoV, PIV 1	qPCR	UK	Kiasari et al., 2008
NPA	ARI	>60 years	21	2/21 9.5 %	VP1	75 % hMPV, RSV, HBoV, PIV 1	qPCR	UK	Kiasari et al., 2008
NPA	ARI	< 3 years	98	0/98 0 %	Unspecified	48 % RSV	Unspecified	Gernamy	Kleines et al., 2008
NPA	Unspecified	5 days-14 years	302	6/302 2 %	VP1	33 % MPV, HBoV	qPCR, nPCR	Thailand	Payungporn et al., 2008a and 2008b
NPA	ARI	1 month-14 years	415	2/415 0.5 %	VP1	100 % WU	PCR	China	Ren et al., 2008
NPA	ARI	15-97 years	297	0/297 0 %	VP1		PCR	China	Ren et al., 2008
Respiratory specimens	ARI	<2 years	367	8/367 2.2 %	VP1	25 % HBoV, MPV	nPCR	USA	Wattier et al., 2008
Respiratory specimens	Asymptomatic	<2 years	96	0/96 0 %	VP1		nPCR	USA	Wattier et al., 2008

Sample type	Diagnosis	Age	N of samples (individuals)	Prevalence	Target	Co-infection	Method	Country	References
NPA	ARI		406	11/406 2.7 %	VP1	72.7 % PIV, IV A, bocavirus, hMPV, RSV	nPCR	China	Yuan et al., 2008
Plasma	HIV without respiratory symptoms	37- 54 years	62	2/62 3.2 %	VP1, LT-ag, ST-ag		PCR	Italy	Babakir-Mina et al., 2009a
Stool	HSCT recipients	1-75 years	25	12/25 48 %	ST-ag	BKV, AdV, CMV	PCR	Italy	Babakir-Mina et al., 2009b
Stool	Non HSCT Hematological patients	1-75 years	6	0/6 0 %	ST-ag	BKV, AdV, CMV	PCR	Italy	Babakir-Mina et al., 2009b
Tonsils	Tonsillectomy patients	10-88 years	91	11/91 12 %	VP1		qPCR	Italy	Babakir-Mina et al., 2009c
Tonsils	Tonsillectomy patients with chronic tonsillitis	10-88 years	48	6/48 12.5 %	VP1		qPCR	Italy	Babakir-Mina et al., 2009c
Tonsils	Tonsillectomy patients with tonsil hyperplasia	10-88 years	26	3/26 11.5 %	VP1		qPCR	Italy	Babakir-Mina et al., 2009c
Tonsils	Tonsillectomy patients with hypotrophic tonsil	10-88 years	1	1/1 100 %	VP1		qPCR	Italy	Babakir-Mina et al., 2009c
Tonsils	Tonsillectomy patients with tonsil carcinoma	10-88 years	8	0/8 0 %	VP1		qPCR	Italy	Babakir-Mina et al., 2009c
Tonsils	Tonsillectomy patients with lymphoma	10-88 years	5	1/5 20 %	VP1		qPCR	Italy	Babakir-Mina et al., 2009c
Tonsils	Tonsillectomy patients with papilloma	10-88 years	3	0/3 0 %	VP1		qPCR	Italy	Babakir-Mina et al., 2009c
Lung cancer tissue	Cancer	40-85 years	29	9/20 45 %	VP1, LT-ag, ST-ag	SV40, BKV, JCV. HPV	PCR	Italy	Babakir-Mina et al., 2009d
Normal surrounding tissue	Cancer	40-85 years	20	1/20 5 %	VP1, LT-ag, ST-ag	SV40, BKV, JCV. HPV	PCR	Italy	Babakir-Mina et al., 2009d
Paranasal tussue	Transplanted thalassemic patient	13 years	1	1/1 100 %	VP1, LT-ag, ST-ag		PCR	Italy	Babakir-Mina et al., 2009d
Lung tissue	Transplanted thalassemic patient	3 years	1	1/1 100 %	VP1		PCR	Italy	Babakir-Mina et al., 2009d
CSF	Viral encephalitis **	4-88 years	60	0/60 0 %	VP1	HIV	qPCR	Italy	Barzon et al., 2009a
Brain	HIV with PML	44-6 years	4	1/4 25 %	VP1	100 % JCV, WU	PCR	Italy	Barzon et al., 2009b
Brain	HIV with PML	21-37 years	10	3/10 33 %	VP1	67 % WU	PCR	Italy	Barzon et al., 2009b

Sample type	Diagnosis	Age	N of samples (individuals)	Prevalence	Target	Co-infection	Method	Country	References
Brain	HIV negative drug users	22-30 years	8	0/8 0 %	VP1		PCR	Italy	Barzon et al., 2009b
PB	HIV		100	1/100 1 %	VP1		nPCR	Italy	Barzon et al., 2009c
PB	HSCT and solid organ transplant recipients		100	0/100 0 %	VP1		nPCR	Italy	Barzon et al., 2009c
PB	Healthy		100	0/100 0 %	VP1		nPCR	Italy	Barzon et al., 2009c
Stool	Unspecified		84	17/84 20.2 %	NCCR	WU?	qPCR	Italy	Bergallo et al., 2009
Stool	Unspecified		84	18/84 21.4 %	ST-ag	WU?	qPCR	Italy	Bergallo et al., 2009
Stool	Unspecified		31	26/31 31 %	VP1	WU?	qPCR	Italy	Bergallo et al., 2009
Tonsils	Unspecified		91	0/91 0 %	NCCR		qPCR	Italy	Bergallo et al., 2009
Tonsils	Unspecified		91	12/91 13.2 %	ST-ag		qPCR	Italy	Bergallo et al., 2009
Tonsils	Unspecified		91	12/91 13.2 %	VP1		qPCR	Italy	Bergallo et al., 2009
NPA	Healthy	15-85 years	99	0/99 0 %	NCCR, ST-ag	PIV 3, HBoV, rotavirus, AdV	qPCR	Australia	Bialasiewicz et al., 2009
NPA	Immunocompromised	16-79 years	22	0/22 0 %	NCCR, ST-ag	PIV 3, HBoV, rotavirus, AdV	qPCR	Australia	Bialasiewicz et al., 2009
NPA	Healthy	11 days-9 years	100	1/100 1 %	NCCR, ST-ag	PIV 3, HBoV, rotavirus, AdV	qPCR	Australia	Bialasiewicz et al., 2009
NPA	Immunocompromised	2 months -13 years	38	2/38 5.2 %	NCCR, ST-ag	PIV 3, HBoV, rotavirus, AdV	qPCR	Australia	Bialasiewicz et al., 2009
BAL	ARI	2 months -82 years	98	3/98 3.1 %	NCCR, ST-ag	PIV 3, HBoV, rotavirus, AdV	qPCR	Australia	Bialasiewicz et al., 2009
Blood	Immunocompromised	1 month-70 years	100	0/100 0 %	NCCR, ST-ag	PIV 3, HBoV, rotavirus, AdV	qPCR	Australia	Bialasiewicz et al., 2009
Blood	Immunocompromised	1 day-77 years	100	0/100 0 %	NCCR, ST-ag	PIV 3, HBoV, rotavirus, AdV	qPCR	Australia	Bialasiewicz et al., 2009
CSF	Suspected neurological disorder	1 day-82 years	100	0/100 0 %	NCCR, ST-ag	PIV 3, HBoV, rotavirus, AdV	qPCR	Australia	Bialasiewicz et al., 2009

Sample type	Diagnosis	Age	N of samples (individuals)	Prevalence	Target	Co-infection	Method	Country	References
Urine	Also tested for STD and UVI	16-60 years	100	0/100 0 %	NCCR, ST-ag	PIV 3, HBoV, rotavirus, AdV	qPCR	Australia	Bialasiewicz et al., 2009
Stool	Acute gastroenteritis	1 day- 11 years	193	1/193 0.5 %	NCCR, ST-ag	PIV 3, HBoV, rotavirus, AdV	qPCR	Australia	Bialasiewicz et al., 2009
Stool	Undiagnosed acute gastroenteritis	1 month- 97 years	221	0/221 0 %	NCCR, ST-ag	PIV 3, HBoV, rotavirus, AdV	qPCR	Australia	Bialasiewicz et al., 2009
Serum and plasma	MCC	42- 86 years	41	33/41 80 %	VP1		ELISA	USA	Carter et al., 2009
Serum and plasma	Healthy	42- 86 years	76	58/76 75 %	VP1		ELISA	USA	Carter et al., 2009
Serum and plasma	Healthy	24- 77 years	451	406/451 90 %	VP1		ELISA	USA	Carter et al., 2009
Lung biopsy	Cancer		32	0/32 0 %	LT-ag, ST-ag		PCR	USA	Duncavage et al., 2009
Gastrointestinal tract biopsy	Cancer		16	0/16 0 %	LT-ag, ST-ag		PCR	USA	Duncavage et al., 2009
Gynecologic biopsy	Cancer		20	0/20 0 %	LT-ag, ST-ag		PCR	USA	Duncavage et al., 2009
Skin or soft tissue biopsy	Cancer		3	0/3 0 %	LT-ag, ST-ag		PCR	USA	Duncavage et al., 2009
Head and neck	Cancer		2	0/2 0 %	LT-ag, ST-ag		PCR	USA	Duncavage et al., 2009
Bladder biopsy	Cancer		1	0/1 0 %	LT-ag, ST-ag		PCR	USA	Duncavage et al., 2009
MCC	Cancer		1	0/1 0 %	LT-ag, ST-ag		PCR	USA	Duncavage et al., 2009
Brain biopsy	PML		4	0/4 0 %	VP1	JCV	PCR	Italy	Focosi et al., 2009
CSF	PML		3	0/3 0 %	VP1	JCV	PCR	Italy	Focosi et al., 2009
Peripheral blood	PML		2	0/2 0 %	VP1	JCV	PCR	Italy	Focosi et al., 2009
Neuroblastoma	Cancer	0-11.5 years	30	0/30 0 %	VP1		PCR	Sweeden	Giraud et al., 2009
CNS tumors	Cancer	0-18 years	25	0/25 0 %	VP1		PCR	Sweeden	Giraud et al., 2009

Sample type	Diagnosis	Age	N of samples (individuals)	Prevalence	Target	Co-infection	Method	Country	References
Nasal swabs	ALL	2-16 years	106 (51)	4/106 3.8 %	VP1, VP2	RV, RSV, HBoV, IV A	PCR		Kantola et al., 2009
Serum	ALL	2-16 years	115 (51)	0/115 0 %	VP1	RV, RSV, HBoV, IV A	PCR		Kantola et al., 2009
Stool	ALL	2-16 years	75 (51)	2/75 2.7 %	VP1	RV, RSV, HBoV, IV A	PCR		Kantola et al., 2009
Tonsil biopsy	Tonsillectomy patients	1-72 years	229	0/229 0 %	VP1	RV, RSV, HBoV, IV A	PCR		Kantola et al., 2009
Serum	Tonsillectomy patients	1-72 years	229	0/229 0 %	VP1	RV, RSV, HBoV, IV A	PCR		Kantola et al., 2009
Serum	Wheezing children	0.2-15 years	496	0/496 0 %	VP1	RV, RSV, HBoV, IV A	PCR		Kantola et al., 2009
Serum	Healthy	1-21 years	721	406/721 56.3 %	VP1	WU, MCV	ELISA	USA	Kean et al., 2009
Serum	Healthy	21-70 years	1501	818/1501 54.5 %	VP1	WU, MCV	ELISA	USA	Kean et al., 2009
NPA	ARI	0-90 years	637	9/637 1.4 %	VP1	59 % IV A, RSV, MPV, PIV 3, RV	qPCR	Sweden	Lindau et al., 2009
Colon tissue	Colorectal cancer and adjacent tissue	32-89 years	144 (72)	0/144 0%	VP1		qPCR	Italy	Militello et al., 2009
Colon tissue	Colorectal cancer and adjacent tissue	41-92 years	150 (50)	0/150 0 %	VP1		qPCR	Italy	Militello et al., 2009
Colon tissue	Healthy	36-82 years	45 (15)	0/45 0 %	VP1		qPCR	Italy	Militello et al., 2009
Plasma	HIV		120	0/120 0 %	VP2	HCV	nPCR	USA	Miller et al., 2009
Plasma	HCV		80	0/80 0 %	VP2	HCV	nPCR	USA	Miller et al., 2009
NPA	ARI (89 % immunosuppressed)	3-85 years	265 (200)	17/265 6.5 %	VP1	37.5 % respiratory viruses	qPCR, nPCR	France	Mourez et al., 2009
Respiratory specimens	ARI	Pediatric	229	2/229 0.9 %	VP1	No respiratory viruses	PCR	Germany	Mueller et al., 2009
Serum	Unspecified	<0.5 years	30	13/30 43.3 %	VP1	100 % WU	ELISA	USA	Nguyen et al., 2009
Serum	Unspecified	0.5-1 year	29	7/29 24.1 %	VP1	100 % WU	ELISA	USA	Nguyen et al., 2009

Sample type	Diagnosis	Age	N of samples (individuals)	Prevalence	Target	Co-infection	Method	Country	References
Serum	Unspecified	1 year	30	12/30 40 %	VP1	83.3 % WU	ELISA	USA	Nguyen et al., 2009
Serum	Unspecified	2 years	30	13/30 43.3 %	VP1	76.9 % WU	ELISA	USA	Nguyen et al., 2009
Serum	Unspecified	3 years	30	15/30 50 %	VP1	66.6 % WU	ELISA	USA	Nguyen et al., 2009
Serum	Unspecified	4 years	30	22/30 73.3 %	VP1	72.7 % WU	ELISA	USA	Nguyen et al., 2009
Serum	Unspecified	5 years	30	28/30 93.3 %	VP1	85.7 % WU	ELISA	USA	Nguyen et al., 2009
Serum	Unspecified	6-8 years	30	26/30 86.7 %	VP1	92.3 % WU	ELISA	USA	Nguyen et al., 2009
Serum	Unspecified	9-12 years	30	30/30 100 %	VP1	90 % WU	ELISA	USA	Nguyen et al., 2009
Serum	Unspecified	13-19 years	30	28/30 93.3 %	VP1	96.4 % WU	ELISA	USA	Nguyen et al., 2009
Serum	Unspecified	20-34 years	30	21/30 70 %	VP1	100 % WU	ELISA	USA	Nguyen et al., 2009
Serum	Unspecified	35-49 years	30	22/30 73.3 %	VP1	100 % WU	ELISA	USA	Nguyen et al., 2009
Serum	Unspecified	50-64 years	30	19/30 63.3 %	VP1	100 % WU	ELISA	USA	Nguyen et al., 2009
Serum	Unspecified	65-79 years	30	22/30 73.3 %	VP1	100 % WU	ELISA	USA	Nguyen et al., 2009
Lymphoid tissue	HIV		42	3/42 7.1 %	VP2, ST-ag		nPCR	Scotland	Sharp et al., 2009
Lymphoid tissue	Variable (6 HIV)		55	1/55 1.8 %	VP2, ST-ag		nPCR	Scotland	Sharp et al., 2009
NPA	Healthy		727 (499)	0/727 0 %	ST-ag		nPCR	Scotland	Sharp et al., 2009
NPA	ARI	<5 years	78	0/78 0 %	NCCR, ST-ag	RSV (50 %), IV (50 %)	qPCR	Netherlands	Van de Pol et al., 2009
NPA and throat swabs	Healthy	<18 years	83	4/83 4.8 %	NCCR, ST-ag	50 % HBoV	qPCR	Netherlands	Van de Pol et al., 2009

Sample type	Diagnosis	Age	N of samples (individuals)	Prevalence	Target	Co-infection	Method	Country	References
Swabs	ARI	1-7 years	230 (18)	6/230 3 %	NCCR, ST-ag	RV, enterovirus, RSV, hCoV, IV A+B, MPV, AdV, Mycoplasma pneumonia, Chlamydophila pneumonia	qPCR	Netherlands	Van der Zalm et al., 2009
NPA	ARI		300	3/300 1 %	VP1	RV, HBoV, AdV, RSV, PIV, IV A, hCoV, hMPV	PCR	South-Africa	Venter et al., 2009
NPA	Healthy		50	0/50 0 %	VP1		PCR	South-Africa	Venter et al., 2009
Plasma	HIV	Adults	153	4/153 2.6 %	VP1		RT-PCR	Italy	Babakir-Mina et al., 2010
Plasma	Healthy	Adults	130	4/130 3.1 %	VP1		RT-PCR	Italy	Babakir-Mina et al., 2010
NPA	HSCT patient		126 (31)	1/126 0.8 %	VP2, LT-ag		nPCR	Italy	Debaggi et al., 2010
NPA	ARI	2-9 months	486	1/486 0.2 %	VP2, LT-ag	100 % hMPV	nPCR	Italy	Debaggi et al., 2010
NPA	Healthy		47	0/47 0 %	VP2, LT-ag		nPCR	Italy	Debaggi et al., 2010
NPA	Influenza like illness	1 month-54 years	465	2/465 0.5 %	VP2	11.6 % IV, AdV, rhinovirus, RSV	PCR	Philippines	Furuse et al., 2010
NPA and sawbs	ARI	1 day-88 years	2599 (162)	72/2599 2.8 %	VP1	71 % RV, RSV, PIV, AdV, HBoV	qPCR, nPCR	USA	Hormozdi et al., 2010
Throat swab	Suspected CAP	20-95 years	567	3/567 0.5 %	VP1, NCCR, ST-ag	RSV, IV B, Haemophilus influenzae	qPCR	Netherlands	Huijskens et al., 2010
Plasma	Healthy	20-66 years	100	67/100 67 %	VP1		IFA	Germany	Neske et al., 2010
Plasma	Renal transplant recipients		195	2/195 1 %	VP1, VP2	Not tested	qPCR, nPCR	Hungary	Csoma et al., 2011
Plasma	Healthy		200	0/200 0 %	VP1, VP2	Not tested	qPCR, nPCR	Hungary	Csoma et al., 2011
Urine	Renal transplant recipients		50	1/50 2 %	VP1, VP2	Not tested	qPCR, nPCR	Hungary	Csoma et al., 2011
Urine	Healthy		36	0/36 0 %	VP1, VP2	Not tested	qPCR, nPCR	Hungary	Csoma et al., 2011
Resperatory specimens	Renal transplant recipients		90	6/90 6.7 %	VP1, VP2	Not tested	qPCR, nPCR	Hungary	Csoma et al., 2011

Sample type	Diagnosis	Age	N of samples (individuals)	Prevalence	Target	Co-infection	Method	Country	References
Brain biopsy, CSF, PB, BM	PML with and without HIV		80 (61)	0/80 0 %	VP1, ST-ag, NCCR		qPCR, PCR	USA	Dang et al., 2011
CSF, PB, urine	MS		115 (21)	0/115 0 %	VP1, ST-ag, NCCR		qPCR, PCR	USA	Dang et al., 2011
Brain biopsy, CSF, PB, BM, urine	Immunsuppressed		113 (90)	0/113 0 %	VP1, ST-ag, NCCR		qPCR, PCR	USA	Dang et al., 2011
CSF, PB, BM, urine	Immunocompetent		156 (33)	0/156 0 %	VP1, ST-ag, NCCR		qPCR, PCR	USA	Dang et al., 2011
Brain biopsy, CSF, PB, BM, urine	Non-PML, Non-MS patients		74 (41)	0/74 0 %	VP1, ST-ag, NCCR		qPCR, PCR	USA	Dang et al., 2011
Respiratory specimens	Immunocompromised most with ARI		161 (102)	9/161 5.6 %	VP1	67 % respiratory viruses: enterovirus/rhinovirus, AdV, RSV,	qPCR	USA	Rao et al., 2011
Respiratory specimens	Immunocompetent most with ARI		295	7/295 2.3 %	VP1	86 % respiratory viruses: PIV, hMPV, RSV, IV A, enterovirus/rhinovirus	qPCR	USA	Rao et al., 2011
NPS	RTI	1 month – 7 years	232 (219)	7/232 3 %	VP1, NCCR	42 % hMPV, HBoV	qPCR, nPCR	Japan	Teramoto et al., 2011
Lung tissue	Lung adenocarsinoma		30	0/30 0 %	VP1, NCCR		qPCR, nPCR	Japan	Teramoto et al., 2011
Lung tissue	Normal lung tissue		30	1/30 3.3 %	VP1, NCCR		qPCR, nPCR	Japan	Teramoto et al., 2011
CSF	Neurological complications after HSCT	1-60 years	20	0/20 0 %	VP1		PCR	Sweden	Rubin et al., 2011
Urine	SLE		72 (5)	3/72	VP1		nPCR	Norway	Unconfirmed results
CSF	Suspected or confirmed neurological disorders		64	7/64	VP1		nPCR	Norway	Unconfirmed results

KIPyV-DNA has been detected in NPA, stool, lung tissue, urine, blood, lymphoid tissue, brain tissue, tonsils but not in CSF.

Allander et al (2007) J Virol 81:4130-4136
Bialasiewicz et al (2007a) J Clin Virol 40:15-18
Bialasiewicz et al (2007b) J Clin Virol 40:9-14
Han et al (2007) Emerg Infect Dis 13:1766-1768
Noreja et al (2007) J Clin Virol 40:307-311
Babakir-Mina et al (2008) J Med Virol 80:2012-2014
Bialasiewicz et al (2008) J Clin Virol 41:63-68
Giraud et al (2008) J Clin Microbiol 46:3595-3598
Foulongne et al (2008) Emerg Infect Dis 14:523-525
Kiasari et al (2008) J Clin Virol 43:123-125
Kleines et al (2008) Intervirology 51:444-446
Payungporn (2008a) Virus Res 135:230-236
Payungporn (2008b) J Virol Methods 153:70-73
Ren et al (2008) J Clin Virol 43:330-333
Wattier et al (2008) Emerg Inf Dis 14:1766-1768
Yuan et al (2008) J Clin Virol 46:3522-3525
Babakir-Mina et al (2009a) Emerg Inf Dis 15:1323-1325
Babakir-Mina et al (2009b) J Med Virol 81:1668-1673
Babakir-Mina et al (2009c) J Clin Virol 46:75-79
Babakir-Mina et al (2009d) J Med virol 81:558-561
Barzon et al (2009a) J Infect Dis 200:314-315
Barzon et al (2009b) J Inf Dis 200: 1755-1758
Barzon et al (2009c) J Clin Virol 45:370
Bergallo et al (2009) J Virol Methods 162:69-74
Bialasiewicz et al (2009) J Clin Virol 45:249-254
Carter et al (2009) J Natl Cancer Inst 1001:1510-1522
Duncavage et al (2009) Am J Surg Pathol 33:1771-1777
Focosi et al (2009) J Clin Virol 45:161-162
Kantola et al (2009) J Clin Virol 45: 292-295
Kean et al (2009) PloS Pathog 5:e10000363
Lindau et al (2009) J Clin Virol 44:24-26
Militello et al (2009) Int J Cancer 124:2501-2503
Miller et al (2009) Emerg Infect Dis 15:1095-1097
Mourez et al (2009) Emerg Infect Dis 15:107-109
Mueller et al (2009) Arch Virol 154:1605-1608
Nguyen et al (2009) Emerg Infect Dis 15:1199-1205
Sharp et al (2009) J Inf Dis 199:398-404
van de Pol et al (2009) Emerg Infect Dis 15:454-457
van der Zalm et al (2009) Emerg Infect dis 14:1787-1789
Venter et al (2009) J Clin Virol 44:230-234
Babakir-Mina et al (2010) Emerg Infect Dis 16:1482-1484
Furuse et al (2010) J Med Virol 82:1071-1074
Huijskens et el (2010) J Clin Virol 49: 306-307
Neske et al (2010) BMC Infectious Diseases 10:215
Csoma et al (2011) J Med Virol 83:1275-1278
Dang et al (2011) PloS ONE 6:e16736
Rao et al. (2011) J Clin Virol 52:28-32
Teramoto et al. (2011) Microbiol Immunol 55:525-530
Rubin et al. (2011) Anticancer Res 10:3489-3492