Università degli Studi di Cagliari

PHD DEGREE

Molecular and Translational Medicine

Cycle XXXI

TITLE OF THE PHD THESIS

A Genetic Barcode Approach to Elucidate the Mechanisms of Polyomavirus Persistent Infection *in vivo*

Scientific Disciplinary Sector(s)

BIO/13 - Applied Biology

PhD Student: Sylvain Blois

Coordinator of the PhD Programme Prof. Sebastiano Banni

Supervisor Prof. Patrizia Zavattari

Final exam. Academic Year 2017 – 2018

Thesis defense: May - June 2019 Session

Table of content

Abstract

Human polyomaviruses are highly prevalent in the general population and establish a long-term asymptomatic infection in healthy individuals. However, a change in the immune status or hormone levels can lead to life-threatening diseases, including tumors. These diseases typically require reactivation from a persistent infection. Currently, a long-standing question in the field of virology is understanding how polyomaviruses undergo a life-long persistent infection in their natural host. Indeed, despite the high seroprevalence of polyomaviruses, little is understood about the mechanisms allowing the establishment, maintenance or reactivation from a long-term persistent infection. The objective of my PhD thesis was to demonstrate whether polyomaviruses undergo a "smoldering" infection, as commonly suggested in literature, or a cryptic true latency, akin to Herpesviruses. In order to solve this enigma, I engineered an innovative genetic barcoded murine polyomavirus to tract single genomes *in vivo* and to define the early and late genes expression patterns associated to each genome involved in the persistent infection. This barcoded virus approach combined with a non-invasive urine-monitoring system that allows the longitudinal study of PyV persistence, will consent to elucidate for the first time the mode of polyomavirus persistence *in vivo* using a limited number of mice. Deciphering the mechanisms of persistence may lead to new therapies and preventive strategies against serious Human diseases.

1. Introduction

Most DNA viruses, including polyomavirus (PyV), Human papillomavirus, adenovirus, adeno-associated virus, Torque Teno virus, and Herpesviruses establish a life-long asymptomatic persistent infection in the general population and can cause life-threatening diseases when changes occur in their microenvironment.

Despite the high prevalence of the PyV in the population, little is understood about the mechanisms used by the viruses to establish, maintain, or reactivate from the persistent state of infection in their natural reservoirs. Many PyV-associated diseases, including cancers, are a consequence of persistent infection, thus, deciphering the mechanisms of viral persistence is a key toward new therapies and preventative strategies against serious human diseases.

1.1 Polyomaviruses

PyVs are ubiquitous viruses belonging to the *Polyomaviridae* family that infect a wide range of hosts including birds, rodents, cattle, fish and humans with a narrow specificity. Surprisingly, despite a highly conserved structural organization, PyVs have a restricted tropism[1]. Though, a common characteristic of Human polyomaviruses (HPyVs) investigated so far is that primary infections occur in early childhood and they establish an asymptomatic life-long persistent infection in their host[2,3]. However, in the context of immune suppression a reactivation from persistency can lead to life-threatening diseases. Serological studies showed that the HPyVs prevalence increases with age and can reach from 40% to 90% in the general adult population. These studies also showed an average co-exposure rate of 6 to 7 HPyVs, suggesting that co-infections with multiple HPyVs are common[3–8]. HPyVs are believed to be transmitted through the respiratory route[9,10], gastrointestinal tract[11], saliva and/or skin contact[1213], blood transfusion, organ transplantation[14] and sexual intercourse[15].

MuPyV was the first PyV identified in 1953 in filtrated extracts from mice with leukemia. The name of the *Polyomaviridae* derive from the observation that these filtrated extracts induced multiple (from Greek poly) tumors (from Greek oma) when inoculated into newborn mice[16,17]. Afterward, Simian Virus 40 (SV40) was the first primate PyV discovered as a contaminant in African green monkey cells used for the production of the polio vaccine[18]. Like MuPyV, SV40 is oncogenic in newborn and immunocompromised adult mice and rodents and these observations in animals leaded to the concern that SV40-contaminated vaccines could have induced tumors to almost 100 million vaccinated individuals during the 1960s. In 1971, the first two HPyVs BK polyomavirus (BKPyV) and JC polyomavirus (JCPyV) were identified in the urine of a renal transplant recipient who developed urethral stenosis and from the brain of a patient with a progressive multifocal leukoencephalopathy (PML), respectively. These viruses were named with the initials of the patients from where they were identified[19,20]. Both BKPyV and JCPyV are ubiquitous in Humans, they establish an asymptomatic and persistent infection following a primary exposure during early childhood, and can be reactivated under various immunosuppressed conditions that ultimately may lead to severe diseases[2]. Although the route of transmission of these viruses is uncertain, the prominent sites of persistence for both viruses are the urinary tract, the genital tract, the central nervous system, and the hematopoietic system[21]. For both BKPyV and JCPyV, two forms of viral genomes exist based on the non-coding control region (NCCR) structure: the archetype and rearranged forms. The archetype virus is the

predominant form detected in the urine of healthy and diseased people and is thought to be the persistent and transmissible form of this virus[22,23]. On the other hand, the rearranged variants are usually considered disease-associated derived forms of the archetype virus [24]. The rearranged viruses are characterized by deletions and duplications in the NCCR compared with the archetype virus that probably occurred during or after reactivation from persistence. As SV40, BKPyV and JCPyV are oncogenic in newborn hamsters and able to transform mammalian cells *in vitro*, however, to date, no conclusive prospective studies support a causal association between JCPyV and BKPyV infections and Human cancer development[25,26].

From 2007, 11 more HPyVs have been identified. HPyV3 (Karolinska Institute polyomavirus (KIPyV))[27], HPyV4 (Washington University polyomavirus (WUPyV))[28], HPyV10 (Malawi polyomavirus (MWPyV))[29], HPyV11 (St. Louis polyomavirus (STLPyV))[30] and HPyV13 (New Jersey polyomavirus (NJPyV))[31] were named accordingly to the first geographical location where they have been found. HPyV5 (Merkel Cell polyomavirus (MCPyV)) was named due to its detection in Merkel cell carcinoma[32] and HPyV8 (Trichodysplasia Spinulosa polyomavirus (TSPyV)), was named accordingly to its identification in trichodysplasia spinulosa, a rare skin disease[33]. The following HPyVs, namely HPyV6, HPyV7, HPyV9, HPyV10 and HPyV12, were named following the order in which they were discovered[34–37]. In 2017, a fourteenth HPyV was discovered, the Lyon IARC polyomavirus (LIPyV)[38].

Currently, the International Committee on Taxonomy of Viruses (ICTV) has divided the *Polyomaviridae* family into 5 genera: *Alphapolyomavirus,* with 37 species that infect animals (including *Mus musculus* polyomavirus 1 formerly named MuPyV) and Humans, including HPyVs 5, 8, 9, 12 and 13; *Betapolyomavirus*, with 29 species that infect animals (including *Macaca mulatta* polyomavirus 1, formerly named SV40) and Humans, including HPyVs 1 (JC and BK) and 4; *Deltapolyomavirus* with only HPyVs 6, 7, 10 and 11; *Gammapolyomavirus*, with seven animal virus species, including PyVs that infect birds; and, finally, an unassigned genus that contain three species of animal PyVs[1,39]. I am aware about the new nomenclature of the ICTV but in this dissertation I will keep the former name of the PyVs discussed here to avoid confusion, as most of the current literature cited here does.

1.2 Virion structure

PyVs are *T=7* icosahedral, nonenveloped viruses that are approximately 45 nm in size. Capsids contain three structural proteins, VP1, VP2, and VP3. The bulk of the capsid is constructed from 360 molecules of VP1 (arranged at 72 pentameric capsomers). The capsid surrounds a single molecule of closed circular, double-stranded viral DNA genome that is approximately 5 kbp in size[40]. In the mature virion, the DNA genome exists as a supercoiled, chromatin-like structure[41] in association with host cell histone proteins H2A, H2B, H3 and H4, forming the viral minichromosome. In the cell, the minichromosome is associated with histone H1[42] (Figure 1).

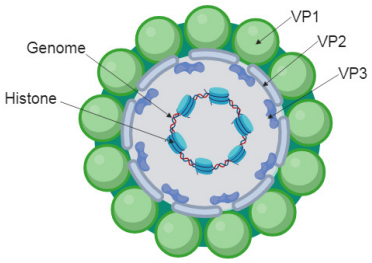

Figure 1 : PyV virion structure. Cartoon not in scale created with BioRender.com.

1.3 Genome organization

While apparently simply organized, PyVs have an unusual genome structure that provides a platform for several cellular gene regulatory mechanisms. Although the general genome organization among PyVs is similar, the number of encoded proteins varies (Figure 2). PyVs genome is arbitrarily divided into a regulatory region named the non-coding control region (NCCR), which is situated between the early and late regions (Figure 2).

The NCCR consists of only several hundred nucleotides but is a complex region that serves multiple functions during infection[43,44]. In particular, NCCR contains the origin of replication, the enhancers and the bidirectional promoter controlling the early and late gene transcription. Each regulatory domain of the NCCR is impacted by distinct molecular machineries that competes for the overlapping sequence elements. The early promoter is a typical RNA polymerase II promoter, including a TATA-box that specifies the early transcription start sites and an upstream enhancer region. The late promoter is TATA-less and generates transcripts with a multitude of 5'-ends spanning more than 100 nucleotides. The NCCR sequence diversity is high between PyV genomes, even if some motifs are conserved[45–49]. NCCR rearrangements can occur and consequently may affect viral DNA replication, promoter activity, virus production and the pathogenic properties of the virus[24,50–52]. Unlike MuPyV, establishment of SV40 persistence was largely independent of the structure of the regulatory region[53].

The early genes region encodes the large T (LT) (100 kDa), middle T (MT) (56 kDa) and small T (ST) (22 kDa) antigens (Figure 2), as well as various truncated variants. A fourth T antigen called tiny T antigen was also described in MuPyV[54]. T antigens are produced by alternative splicing of the virus-encoded early genes transcripts[55]. All T antigens share a DNAJ domain at the N-terminal amino acid sequence that has been shown to be a regulatory domain of the tumor suppressor function of the Rb family[56,57]. The polyfunctional LT antigen has been found in all PyVs and is involved in both viral replication and virus-induced cell transformation. MT antigen has been identified in MuPyV[58,59], hamster polyomavirus (HaPyV)[60] and TSPyV[61] , which was described as the principal transforming protein in MuPyV[62,63]. MT and another T antigen called the Alternative T (ALT) are both encoded by the Alternate T-antigen open reading frame (ALTO), also known as ORF5. TSPyV ORF5 was shown to express both MT and ALT[61], while MCPyV ALTO only expresses ALT[64]. ST antigen has also been described to interact with several cellular proteins and play an important role in DNA replication and the control of the early-to-late transcriptional switch[65–68].

The third region of the PyV genome encoding for the late genes which are transcribed from the opposite strand respected to the early genes. Late primary transcripts accumulate after the onset of viral genome replication and are alternatively spliced into mRNAs coding for three capsid proteins, namely VP1, VP2 and VP3. VP4 has so far been detected in SV40[69], avian polyomavirus[70], potentially BKPyV[71]. In addition, JCPyV, BKPyV, and SV40 express from the late transcript a small non-structural protein called agnoprotein[72–74] (Figure 2). This protein is well conserved between the two HPyVs and SV40, particularly within the N-terminal region, suggesting a conserved function. The agnoprotein is involved in viral transcription, DNA replication, virion biogenesis and release (reviewed in Gerits and Moens 2012[75]).

The 3' end of the early and late regions is separated by a short segment that contains the two bi-directional polyadenylation (poly A) signals separated by one or few base pairs (Figure 2).

Many PyVs, including BKPyV[76], JCPyV[76], MCPyV[77], SV40[78], raccoon polyomavirus (RacPyV)[79] and MuPyV[80] encode small regulatory RNAs (~22 nucleotides), called microRNAs (miRNAs), in the late transcripts orientation, in different region of the genome opposite to the T antigen mRNAs. SV40, BKPyV, and JCPyV miRNA are encoded just downstream of the late poly(A) signal and the target site is near the 5' end of the LT antigen open reading frame (ORF), while MuPyV and MCPyV miRNAs are encoded further downstream of the late poly(A) signal (~1.7 kb) and the target site is close to the 3' end of the LT antigen ORF. The different position of the miRNAs target in the LT antigen region among PyVs correlates with the evolution of these viruses[81] (Figure 2). The perfect complementary of miRNAs to the LT antigen transcripts allows the cleavage of these mRNAs by Argonaute 2 (Ago2). As a result, the LT protein expression will be decreased during the late phase of infection. The autoregulatory activity on early gene products is a conserved function among PyV miRNAs[76–80,82].

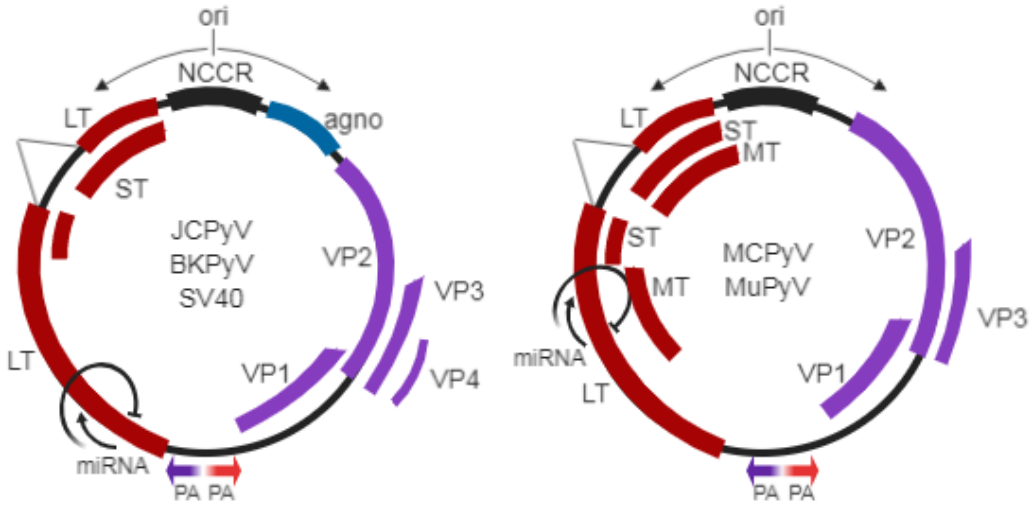

Figure 2 : Generic PyVs genome organization. Early genes are in red and late genes are in blue. Thicker regions represent open reading frames. VP4 has been described only for SV40 and avian PyV (not shown here). The Tiny T antigen is not represented in the MuPyV genome for simplification. The replication origin (ori), non-coding control region (NCCR) and poly (A) signal (PA) for the late genes (blue) and early genes (red) are shown. The miRNAs are represented by a thin black arrow and the target sequence is indicated by the curved arrow. Cartoon not in scale created with BioRender.com.

1.4 PyV life cycle

1.4.1 Attachment and entry

PyVs have a narrow host range and limited cell type tropism, thus, the mechanism of viral entry into the host cell depends on the virus and the cell type. Attachment of PyVs to target cells is mediated by VP1 and cellular receptors. Sialylated lipids called gangliosides are the common primary receptors for HPyVs. Because gangliosides are widely distributed on a variety of cell types they are unlikely to be the major determinants of host or tissue tropism. BKPyV uses gangliosides GD1b and GT1b[83], while JCPyV requires GT1b, GD1b, and GD2 for the attachment[84]. JCPyV makes also contact with the pentasaccharide α2,6-linked lactoseries tetrasaccharide c (LSTc) and the serotonin receptor 5-hydroxytryptamine receptors 5HT2A acts as a possible co-receptor[85–87]. MCPyV binds GT1b too and uses glycosaminoglycans as possible co-receptors[88], while TSPyV forms complexes with GM1 glycan and α2,3- and α2,6-sialyllactose[89]. However, HPyV6 and HPyV7 may employ non-ganglioside receptors as suggested by high resolution X-ray studies[90]. MuPyV binds to cell surface gangliosides GD1a and GT1b and the α4ß1-integrin receptor[91,92]. Upon binding to mouse embryonic fibroblasts, MuPyV activates the

phosphatidylinositol 3-kinase (PI3K), FAK/SRC, and mitogen-activated protein kinase (MAPK) pathways[93]. SV40 uses the ganglioside GM1 as a receptor and MHC class I acts as a co-receptor[94].

Following the binding step to the receptor, PyVs enter the cells mostly by caveolar/lipid raft-mediated endocytosis, except JCPyV that is internalized by clathrin-mediated endocytosis[95]. Caveolin-dependent and -independent entry has been both described for MuPyV[96], while caveolin- and clathrin-independent processes has been described in different cell types for SV40 and BKPyV[97,98].

Following the internalization step, PyVs particles are transferred into the endoplasmic reticulum (ER) where the uncoating step can start[99]. Then, the partially uncoated virus particle is translocated across the ER membrane[100–102]. Upon reaching the cytoplasm, disassembly of polyomavirus particles proceeds and the viral genome free of VP2/VP3 enters the nucleus[103]. The importation of the PyVs genome to the nucleus is not well understood. Microtubules and microfilaments have been shown to play important functions in the transport of MuPyV genome to the nucleus[104].

1.4.2 Gene regulation in PyV infection

PyVs genome is divided into “early” and “late” regions, which are expressed and regulated differently as infection proceeds. The transcription of the early and late genes and the replication of the viral genome are governed by the NCCR, where the origin of replication and transcriptional control elements (promoter/enhancer) are located. The early and late transcription units extend in opposite directions around the circular genome from start sites close to the bi-directional origin of DNA replication[105,106]. Once in the nucleus, the viral genome is free of capsid proteins and maintained in the episomal form. The transcription of PyVs genes is mediated by the cellular RNA polymerase II and host factors. The replication of the viral genome is mediated by the LT antigen, the cellular DNA polymerase and hosts factors. The transcription of the early region initiates prior to the onset of the viral DNA replication while the late transcription starts after the viral genome replication[107]. Bethge T *et al.* showed that BKPyV bidirectional gene expression is controlled by the host cell transcription factor Sp1 and involves the DNA strand directionality and the binding affinity of Sp1[108]. After the early genes expression, the BKPyV LT antigen mediates genome replication and then late genes expression independently of Sp1 and TATA-like elements[108].

1.4.3 Early transcription

The early promoter in the NCCR contain a TATA-box like sequence upstream from the start site for the early transcription region, and multiple binding sites for the LT antigen and several cellular transcription factors[43,44]. Early transcription generate a single early precursor mRNA from which different transcripts are alternatively spliced. The late products generated are the regulatory T antigens.

At low concentrations, such as during the early phase of infection, LT antigen binds to the high-affinity LT antigen-binding motifs in the NCCR and interacts with cellular proteins to stimulate early transcription by the RNA polymerase II. Later in infection, when the LT antigen concentration increases, LT antigen interacts with low-affinity binding motifs located downstream of the TATA-box promoter in the NCCR. As a result, LT antigen impedes the early transcription by the RNA polymerase II complex and autoregulates the synthesis of LT antigen.

LT antigen in complex with recruited host transcription factors interact with the TATA-less late promoter and thus promotes late transcription[109–111]. These events may also be regulated by epigenetic mechanisms due to the association of viral DNA with cellular histones as described in 1.4.5 Early-late switch. MT antigen has been reported to be involved in the regulation of viral DNA replication and transcription[67,112]. HPyVs early transcription produce other early proteins (reviewed in Moens *et al.* 2017[113]) but their role in the transcription remains mostly unknown.

1.4.4 Viral DNA replication

LT antigen has a SF3 helicase domain that is required for the unwinding of the double-stranded DNA of PyV prior to viral DNA replication[107]. The helicase domain is near to an ORI binding domain (OBD) that recognizes the origin of replication (ORI) on the viral genome (Figure 3). ORI site is composed of 3 cores required for unwinding and replication: a central palindrome containing 4 repeats of 5'-GAGGC-3' (5'-CC(W)6GG-3' in case of bird PyVs), an AT-rich element and an imperfect palindrome[114]. Binding of ATP at the ATP binding domain is required for the hexameric conformation of the LT antigen to form a "ring" around the origin of replication[115]. The hydrolysis of ATP also produces the energy required for DNA melting and the replication fork unwinding. LT antigen's ATPase activity is also required for DNA elongation[116]. Two head-to-head "rings" are needed to form the active helicase and unwind the double-stranded DNA genome in opposite directions (bidirectional replication), before the duplication of the viral genome by cellular DNA polymerase. Thus, the LT antigen is indirectly involved in the initiation of DNA replication by recruiting the cellular DNA replication machinery at the origin of replication[117,118]. SV40 ST antigen has also been shown to play a role in the stabilization of the LT antigen by preventing its dephosphorylation by the cellular protein phosphatase 2A (PP2A)[119]. JCPyV agnoprotein enhances LT antigen binding to the origin of replication and stimulates viral DNA replication[120].

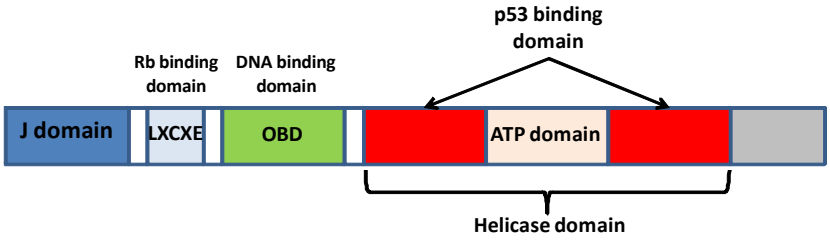

Figure 3 : Functional domains of the large T antigen. Cartoon not in scale.

1.4.5 Early-late switch

MuPyV genes expression during lytic infection of permissive mouse cells is temporally well-defined and finely regulated[121–123]. Immediately after infection, the early genes transcription begins. Between 12 and 15 hours p.i. the viral DNA replication starts and the late transcripts begin to increase, almost exponentially, while early gene products (RNA and proteins) accumulate much slowly up to the end of the lytic cycle to reach the same amount of early transcripts observed at 12 and 24 hours p.i. When the cell enters the S-phase and the viral genome replication occurs, the virus undergoes an early-late switch of gene expression. The early-late switch depends on viral DNA replication and appears to be the result from changes in the transcription elongation and/or RNA stability[124–127]. During the early infection, before the beginning of the viral DNA replication, the late genes are transcribed with a low accumulation of transcripts due to an efficient polyadenylation of the primary

RNAs that favors the alternative splicing into VP1 and VP3 mRNAs (VP2 mRNA is unspliced). These mRNAs are inefficiently exported from the nucleus to the cytoplasm, by a mechanism that remains unclear, and are degraded[128].

In the MuPyV (and likely other PyVs), the inefficient termination of late genes transcription leads to the overlapping of the 3' ends of the early and late transcripts. This transcript overlap that may lead by complementary to the formation double-stranded RNAs has been shown to be essential to enter the late phase of infection[129]. The polyadenylation signals (AAUAAA) overlap results in an A-rich double-stranded RNA region. Interestingly, the onset of viral DNA replication also correlates with an increase in A-to-I editing of viral RNA by an Adenosine Deaminase Acting on RNAs (ADARs)[123,130,131]. ADARs binds to double-stranded RNAs with a preference for A-rich regions and catalyzes the deamination of the adenosine to inosines (which act biochemically and genetically like guanosines) on both strand[132]. Thus, the editing of the polyadenylation signals is likely to be responsible for the loss of polyadenylation efficiency that leads to giant transcripts during the late phase of infection[129,133]. In turns, multigenomic late transcripts can form RNA–RNA duplexes with early transcripts that become a substrate for the ADARs. As a consequence, edited RNAs are retained in the nucleus by a quality control system involving the p54nrb/NONO protein[134] and paraspeckles[135,136] to prevent them to be translated into aberrant proteins[134]. Because late primary transcripts accumulate to high levels in the nucleus, the inhibition of early transcripts by editing is strong.

In addition to the RNA editing, another mechanism of altered early-strand start sites at late times of infection has been reported for MuPyV[123,137,138], SV40[139] and JCPyV[140] to regulate the early transcripts expression at the transcription level. Indeed, after the beginning of the viral DNA replication, the early transcriptional start sites downstream the early TATA element is progressively shifted upstream. Thus, during late infection, many early-strand mRNAs have 5'-UTRs longer hundreds of nucleotides respected to mRNAs at earlier times. As a consequence, the RNA stability is altered and the levels of early mRNAs are reduced. This early transcripts start sites shift at late times is dependent on DNA replication[123].

The PyVs genome exists as a typical eukaryotic chromatin in complex with cellular histones both within the virion and in infected cells (intracellular minichromosome). Several studies showed that SV40 undertakes most of epigenetic regulations found in cellular chromatin during its life cycle, mostly mediated by histone modifications and nucleosome location[141–143]. MuPyV also undergoes histone modifications[144]. However, PyVs do not appear to use DNA methylation as a form of regulation of the virus replication[145,146], but PyVs infections may lead to cellular DNA hypermethylation[147–150]. Interestingly, Kumar MA *et al.* recently described a novel mechanism of viral transcriptional control in which the location of nucleosomes in the regulatory region can change in order to expose or block transcription factor binding sites required for the activation or repression of SV40 early and late genes expression[151,152].

1.4.6 Late transcription

The late genes are transcribed from the strand in opposite direction respected to the early genes after the beginning of viral genome replication. LT antigen stimulates the late transcription by binding to transcription cofactors such as TATA-binding protein (TBP) and several TBP-associated factors like, for example, Sp1 and Tef-1

in the NCCR[110]. In BKPyV, however, it has been recently shown that the LT antigen can activate late genes expression independently of Sp1 and TATA-like elements[108]. The late region encodes for the capsid proteins and the agnoprotein, which are translated from differently spliced mRNAs[69,72–74,153–155].

At early times, the polyadenylation of late transcripts is efficient. However, after the beginning of the DNA replication, the late transcripts polyadenylation becomes inefficient (as described above). As a consequence, the transcription does not terminate and the RNA polymerase II can continue to transcribe many time the circular genome, leading to the generation of not polyadenylated giant multigenomic transcripts of different size[133,156–159]. The MuPyV genome encodes a 57-nucleotide non-coding exon close to the 5'-end of the late genes, called late leader. Late multigenomic transcripts that contain a tandem of introns and exons are first processed into pre-mRNAs by the splicing of genome-length introns between two late leader exons[160]. As a consequence, most late pre-mRNAs contain multiple tandems of late leader exons at the 5' end[159], corresponding to the number of times the RNA polymerase II crossed over the polyadenylation site[125]. Then, the "leader-to-body" splicing can occur between the 5'-splice site of the late leader exon and the last 3'-splice site of the VP1 and VP3 coding body . No splicing occurs for VP2[159,160]. The late mature mRNAs are polyadenylated and efficiently exported to the cytoplasm[161].

Interestingly, the complementarity of the multiple tandem of late leaders with the 3'-end of the mouse 18S rRNA could increase the recruitment of ribosomes and enhance the translation of capsid proteins [162].

1.4.7 Assembly and release

The virus particles assembly occurs in the nucleus[163,164], however, the mechanism by which mature progeny virus particles are released is poorly understood. PyVs can cause lysis of the infected host cells[165,166] or shed from intact cells[167,168]. The agnoprotein seems to participate in the release of progeny virus[169–171]. The JCPyV agnoprotein possesses viroporin activity[172,173], while BKPyV agnoprotein may interfere with the exocytotic pathway[174]. SV40 VP4 also possesses viroporin activity but it is not required for progeny release[71].

1.5 PyV-associated diseases in Humans

Most HPyVs infections result in a life-long persistent infection without clinical symptoms. However, under immunocompromised states a virus reactivation can occur and lead to life-threatening diseases. HPyV-associated diseases range from inflammatory, to hyperplastic and neoplastic disorders[175,176]. The components of the immune system that keep persistent viruses clinically silent and the perturbations in immunity that enable these HPyVs to cause disease are mostly unknown[177].

1.5.1 HPyV-associated non-malignant diseases

The viral genome of BKPyV persists for life in a number of organs, including the kidney epithelial cells[178] and lower urinary tract[179], lymphoid tissue[9], leukocytes[180] and brain[181], without any apparent symptoms. However, BKPyV becomes a major issue since reactivation of BKPyV in immunocompromised patients receiving renal or hematopoietic stem cell transplantation may cause serious complications that include BKPyV-associated nephropathy (BKVAN) in about 10% of kidney recipients, hemorrhagic cystitis in 10–25% of hematopoietic-cell transplant recipients and less commonly urethral stenosis[182,183]. BKVAN is significantly associated with renal

transplant dysfunction and allograft loss in 90% of these patients[184]. Currently, no specific anti-BKPyV diseases drug has been approved and the only treatment available is based on the immunosuppression reduction, though it may enhance the risk of transplant rejection.

On the other hand, JCPyV also causes an asymptomatic infection in the kidney in the majority of the population. In immunosuppressed individuals however, JCPyV may reactivate and spread to the brain, causing the progressive multifocal leukoencephalopathy (PML), a fatal progressing subacute demyelinating disease of the central nervous system[185].PML is associated with a JCPyV lytic infection in the oligodendrocytes and astrocytes[186,187] that leads to the oligodendrocyte death and demyelination. JCPyV variants isolated from the brain of PML patients derive from the "archetype" found in the urine of asymptomatic people with extensive deletions and duplications within the promoter/enhancer region of their genome[50,188]. These mutations may alter the tropism and promote the lytic infection. So far, no specific antiviral therapy of PML has been approved.

Considering the severity of the BKPyV- and JCPyV-associated diseases, there is an urgent need for the research and development of specific antiviral treatments, either for prophylaxis and treatment of these diseases[189].

In addition to BK and JC PyVs, TSPyV was also found associated with a non-malignant human disease called Trichodysplasia spinulosa, a rare proliferative skin disease[33]. Other HPyVs were found in patient sample, however, the pathogenic role for most of them remain to be demonstrated. For example, KIPyV and WUPyV have been found in patients with respiratory symptoms[190]. HPyV6 and HPyV7 may be associated with pruritic and dyskeratotic dermatoses in immunosuppressed patients[34,191,192] and HPyV7 was also found in lung transplant recipients with epithelial hyperplasia and viruria[191].

1.5.2 HPyV-associated malignant diseases

In their natural host, PyVs are rarely associated with tumorigenesis. The potential oncogenic ability of PyVs has been mainly attributed to the pleiotropic regulatory T antigens[193]. Early studies have shown that SV40 and MuPyV T antigens contribute to cellular transformation in culture and to tumors development in laboratory animals[194–196].

Interestingly, in addition to the regulatory function of the viral transcription and replication mediated by the DnaJ domain, T antigens also have the capability to disrupt cellular signaling pathways and the cell cycle regulation. Indeed, the DNAJ domain of the LT antigen is a heat shock protein 70 (Hsp70) binding domain and the conserved LXCXE motif of the LT antigen can bind to the pRb proteins[197]. As described for SV40, BKPyV, JCPyV and MCPyV, the interaction between Hsp70 and the DNAJ domain activates the ATPase activity of Hsp70[198] and disrupt complexes formed by two tumor suppressor proteins of the pRb family, p107 and p130, and the E2F transcription factor. This inactivation of the E2F pathway results in the entry of the cells into S-phase, thus in the induction of cell proliferation[56,199–205].

In response to DNA damage, the helicase domain of the LT antigen of SV40[206], and in a similar manner of JCPyV[204,207] and BKPyV[203], binds to the tumor suppressor protein p53, thus prevents the cell cycle arrest and apoptosis (Figure 3). Interestingly, the MuPyV LT antigen does not interact with p53, indeed, most of the MuPyV contribution to cellular transformation is due to its transmembrane MT antigen[62,208]. In addition to the N-terminal

region of the LT antigen, a conserved threonine residue in the C-terminal domain of the SV40 LT antigen impedes, when phosphorylated, the degradation of the G1–S-specific cyclin E1 and MYC by competing with FBXW7 thus, contributes to the cell growth and proliferation[209]. Moreover, SV40[210,211], MuPyV[212], JCPyV[213] and MCPyV[214] T antigens inactivate the cellular protein phosphatase A2 (PP2A), a critical regulator of the mitogen-activated protein kinase (MAPK) signaling cascade involved in crucial cellular functions, such as metabolism, cell cycle progression, and apoptosis[210,211,215–219]. Interestingly, other studies demonstrated that late gene auxiliary agnoprotein of BKPyV and JCPyV may also be involved in tumorigenesis[220–223].

Current evidence only supports the role of MCPyV as a carcinogen to Humans[32]. In the general population MCPyV is considered a ubiquitous skin-tropic virus, its seroprevalence is high. MCPyV DNA is also found in more than 80% of Merkel Cell Carcinoma (MCC), a rare and fatal aggressive skin tumor associated with elderly or immunosuppressed individuals[224]. Immunosuppression has been associated to the development of MCPyV-associated MCC[225]. Current therapeutic strategies, i.e. surgery, radiotherapy and chemotherapy, reach only a limited effect in presence of metastases[226].

The genome of MCPyV is clonally integrated into the MCC cell genome[32], that results in LT antigen expression without virus replication and lysis of the cell. MCPyV LT antigens-associated MCC are C-terminal truncated but (i) conserved the DnaJ and LXCXE motifs that can promote cellular growth by interaction with the retinoblastoma protein family members, and (ii) lost of the origin-binding sites and the helicase domains that are essential for the virus replication and the binding to p53[227–229]. MCPyV ST antigen-associated MCC can interact with several cellular processes and promote cell motility, migration and invasion *in vitro* and induce tumors *in vivo*[193,230–232]. To induce tumors and allow the infected cells to progress, MCPyV needs to escape the immune system[233]. In order to do it, MCPyV LT and ST antigens downregulate the Toll-like receptor 9 (TLR9), a receptor in the host innate immune response that senses viral dsDNA in the cell[234]. As BKPyV, MCPyV LT antigen inhibits TLR9 expression by decreasing the expression of the transcription factor C/EBPβ. C/EBPβ has an important role in the regulation of cytokines expresssion[235] and in controlling cell proliferation and differentiation[236]. Thus, by decreasing the expression of C/EBPβ, MCPyV LT antigen may both act on the immune response and induce cell proliferation. Another mechanism by which MCPyV evades the immune system is (i) by downregulating cell surface markers such as MHC-I[237] and simultaneously the natural killer group 2, member D (NKG2D) to avoid the elimination of the tumor cells that express low MHC-I by Natural Killer (NK) cells[238] or (ii) by overexpressing the immune-inhibitory ligand programmed death-ligand 1 (PD-L1) in the tumor microenvironment to inhibit the T-cell responses[239]. MCPyV also evades the immune system by ST antigen-mediated downregulation of the NF-κB pathway[240]. A complete review of the cell surface markers and genes modulated by MCPyV in MCC have been reviewed in Moens U *et al.*[241]. Another mechanism by which HPyVs may affect the immune system and the host genes expression is the miRNAs, however, MCPyV miRNA has not been shown to be involved in MCC[242].

More recently, studies[243,244] of the MCPyV-positive MCC miRNAome showed that MCPyV regulates cellular miRNAs, including a tumor suppressor (miR-203)[243], suggesting that MCPyV infection is interfering with the cell growth and cell cycle arrest comparable to LT antigen inhibition of p53 and pRb[243].

Though JCPyV and BKPyV DNA have been shown to be present in tumors and initially proposed to cause tumors, their role in oncogenesis remains inconclusive and controversial[25,26]. The presence of BKPyV may increase the risk of the development of bladder, renal and prostate cancer for example, while JCPyV may be associated with colorectal cancer and CNS tumors[193,245–248]. However, the fact that up to 90% of the Human population is infected with JCPyV and BKPyV and the association with tumors is rare, suggests that there is an insufficient evidence to conclude that JCPyV and BKPyV 'drivers' in associated-tumors. As opposed to MCPyV that is classified as a group 2A carcinogenic biological agent (i.e., probably carcinogenic to Humans), BKPyV and JCPyV are classified as a group 2B agents (i.e., possibly carcinogenic to humans).

As a conclusion on this chapter, HPyVs can cause devastating diseases in immunocompromised individuals and because the use of immunosuppressant drugs increases, the incidence of HPyV-associated diseases is likely to boost. Many of HPyV-associated diseases are likely to manifest from persistent infection[175], therefore, understanding the mechanisms mediated by HPyVs to maintain and re-activate from a persistent infection is crucial to prevent and treat these pathologies.

1.6 Roles of PyV-encoded miRNAs at the host–virus interface

1.6.1 Biogenesis and function of miRNAs

miRNAs are small, non-coding RNAs (~22 nucleotides) that regulate gene expression at the post-transcriptional level. miRNAs are produced from long precursor primary transcripts (pri-miRNAs) typically transcribed by RNA polymerase II. The pri-miRNAs contain an imperfect stem-loop hairpin structure that is processed in the nucleus by the RNAse III-like endonuclease Drosha (in the canonical miRNA maturation pathway)[249]. Drosha with its essential co-factor DGCR8, forms the core Microprocessor to bind to the pri-miRNAs and cleave towards the base of the hairpin stem. The newly produced ~70 nt hairpin, called pre-miRNA, is exported to the cytosol via the RAN-GTPase Exportin-5[250,251] and cleaved by the RNAse III-like endonuclease Dicer into a transient ~22 nt duplex RNA, usually with short (~2 nucleotides) 3′ overhangs[252]. One of the two strands of the RNA duplex, called the miRNA or "guide" strand is loaded into the multiprotein RNA-induced silencing complex (RISC) while the other strand, called the "star" (*) or "passenger" strand is likely to be degraded. A key component of RISC, the Argonaute (Ago) protein, associates with the guide strand that directs RISC to the target sequence through Watson-Crick base pairing. The mature guide sequence from a pre-miRNA may shift from one arm to the other in different tissues[253]. Mature miRNAs inhibit gene expression at the posttranscriptional level by targeting complementary sequences, typically in the 3' untranslated region (3' UTR) of mRNA targets, which prevents their translation or induces their degradation[254]. In animal miRNAs, the target site recognition is primarily guided by perfect Watson-Crick pairing of the miRNA seed sequence (nucleotides 2-8), whereas the distal sequences typically have poor complementarity[255]. This partial pairing leads mainly to translational inhibition of the mRNA. Although rarely described for animal miRNAs, plant miRNAs and some virus encoded-miRNAs can bind to the target with perfect complementarity, which results in RISC-mediated endonucleolytic cleavage of the mRNA.

miRNAs can be released from cells encapsulated in extracellular vesicles, such as exosomes or in apoptotic bodies[256–258] and 90% of the secreted miRNAs are found in complex with proteins (e.g., protein argonaute 2)[259]. Circulating miRNAs can reach recipient cells and play a role in intercellular communication[260–264]. miRNAs can regulate numerous cellular processes, including cell proliferation, differentiation, development, apoptosis, angiogenesis, metabolism, and immune responses[265–269]. Interestingly, miRNA activity can be attenuated or even inverted in some circumstances, such as stress[270–272]. Aberrant expression of miRNAs has been shown to be frequently involved in pathogenic processes, including cancer[273,274].

Most viruses that encode miRNAs have a step of their life cycle into the nucleus, a DNA component and the ability to establish a persistent mode of infection. These viruses belong to diverse families, most with a DNA genome, for example Herpesviruses, PyVs, and Adenoviruses, and some with an RNA genome, like Retroviruses. Overall, viruses encode more than 500 miRNAs[275,276]. Viral miRNAs can be classified into two groups. The first group clusters viral miRNAs called ‘‘analogs’’ because they mimic host miRNAs, and the second group represent most viral miRNAs that do not share target sites with host miRNAs. Virus-encoded miRNAs can be subgrouped into different categories based on their biological function, i.e. those that (i) prevent the cell death, (ii) evade the immune response, and (iii) limit the infection by targeting host or viral genes[277].

The molecular mechanisms of PyV miRNAs expression have not been thoroughly explored yet. Studies from our lab on SV40 and MuPyV suggest that giant transcripts resulting from the inefficient termination of the late genes transcription may serve as precursors for the processing of viral miRNAs[76,78,80]. More recently, a transcript that originates immediately upstream of the MCPyV miRNAs has been identified, suggesting that the PyV miRNA expression may be driven by a cryptic promoter (i.e. a NCCR-independent expression)[278].

1.6.2 Role of the PyV miRNA locus in the viral infection

Non-rearranged BKPyV strains closely related to the circulating BKPyV in humans and characterized by a lower levels of early transcripts express miRNAs able to down-regulate LT antigen levels to sufficiently reduce the viral replication in primary renal proximal tubule epithelial cells[279]. Consistent with these results, MCPyV miRNA promotes long-term persistence by inhibiting DNA replication in a cell culture model[278], and the SV40 miRNAs reduce persistent SV40 DNA loads in Syrian golden hamster tissues, B lymphocytes and myeloid cells[53,280]. These studies show that the PyVs miRNAs-mediated reduction of virus replication by down-regulating LT antigen expression in late infection is a conserved function among the PyVs family miRNAs[76–80,82]. Recently, Burke *et al.*[281] confirmed *in vivo* that the MuPyV miRNA is not required for infection and replication in tissues during the acute phase of infection and showed that the miRNA expression correlates with a reduction of genomic DNA in select tissues during the persistent phase of infection. Interestingly, the authors showed that the MuPyV miRNA locus enhances the viral DNA shedding in urine during the acute phase of infection but limits the number of shedding events during the persistent infection[281]. Overall, these findings suggest a role of PyVs miRNAs in the prevention of lytic replication, the establishment and maintenance of PyV persistence, as well as the dissemination/shedding of the virus.

1.6.3 Role of the PyV miRNA locus in reducing the immune responses

Cell-mediated immunity involves both antigen-specific cytotoxic T lymphocytes (CTLs) and non-specific natural killer (NK) cells to eliminate infected cells through processes that induce apoptosis and the subsequent cell lysis. The clearance of virally infected cells by CTLs requires the presentation of viral antigens by infected target cells to specific CTL receptors. In the murine model this response can be induced by the LT antigen[282].

Interestingly, several studies showed that PyV miRNAs can inhibit the host immune response. Indeed, Chen *et al.*[283] identified natural miRNA-null strains of SV40 and JCPyV from severely immunocompromised hosts that can replicate similarly to the wild-type counterparts in cell cultures. This study suggests that the immune system provides a selective pressure on the miRNA expression and it is likely that PyV miRNAs are required to evade the immune response. Interestingly, Sullivan *et al.*[78] showed that the SV40 miRNA-mediated down-regulation of T antigens decreases the susceptibility of SV40-infected cells to the cytotoxic CD8 T-cell-mediated lysis and IFN-γ release in cell culture. Moreover, the same group suggested that a SV40 strain miRNA may hinder the inflammatory processes[82]. On the other hand, BKPyV and JCPyV 3p arm miRNAs have been proposed to down-regulate the host transcript of ULBP3, a stress-induced ligand recognized by the killer receptor NKG2D, thus leading to the reduction of NK cells-mediated killing of BKPyV- and JCPyV-infected cells[284]. Recently, Bauman *et al.*[285] showed that SV40 also evades NK cells killing by down-regulating ULBP1, another stress-induced ligand of NKG2D. These studies suggest a central role of the NK cells-mediated immune response against PyVs and the important function of PyVs miRNAs to enhance the life of infected cells, either by limiting or preventing the lytic replication, as well as by modulating the innate and adaptive immune responses, thus promoting the viral persistence.

Interestingly, Burke *et al.*[281] showed that the viral DNA shedding defect observed in urine of C57BL/6 mice infected with a MuPyV miRNA-mutant was restored in C57BL/6 *Rag2*-/- mice lacking mature T and B cells during the acute phase of infection (10-13 days p.i.). These results suggest that the PyV miRNAs support viruria during the acute phase of infection in immunocompetent hosts, probably by helping to evade the host adaptive immune response[281]. This function may participate to the mode of transmission of PyVs. Burke *et al.* suggested that PyVs miRNAs would have a efficient role on the adaptive immune response during a limited period of time because of a change from the CD8+ T cell-mediated response against LT antigen to a CD8+ T cell-mediated response against VP2 peptide[286].

1.6.4 Role of the PyV-encoded miRNAs in tumorigenesis

Although some viral miRNAs have been suggested to play a role in cancer[287], a direct implication of HPyVs miRNAs in cancer is lacking[242]. So far, RacPyV and MCPyV are the only PyVs to be associated with cancer in their natural host[227,288]. The fact that RacPyV genome is predominantly episomal in RacPyV-associated tumors[289,290] while the MCPyV genome is clonally integrated into the MCCs cell genome[32] suggests distinct modes of tumorigenesis associated with PyVs. This hypothesis allowed our lab to identify for the first time a highly expressed PyV-encoded miRNA in tumors[79]. Indeed, RacPyV miRNA was among the most abundant miRNAs detectable in RacPyV-associated tumors, but not in RacPyV-negative tissues[79]. This is in contrast with MCPyV-positive MCCs in which MCPyV miRNA is detectable in less than half of the MCCs at a level <0.025% of total

miRNAs[242,291]. This suggests that MCPyV miRNA is not involved in MCC. However, MCPyV miRNA can target the host cell protein AMBRA1, which is involved in autophagy and apoptosis[292], and MuPyV miRNA downregulates the pro-apoptotic factor Smad2 during the late phase of infection, thus suppressing apoptosis *in vivo*[293]. Taken together, these studies suggest that PyVs miRNAs may participate to the neoplastic progression by interfering with some crucial cellular genes expression during the late phase of infection.

Interestingly, sequences alignment of the virally encoded SV40-miR-S1-5p and SV40-miR-S1-3p showed similarities with human miRNAs (i.e. miR-423-5p[294] and miR-1266-3p[295], respectively) known to modulate signaling pathways involved in Human cancers[296,297]. Even if the biological relevance of SV40-encoded miRNAs in tumorigenesis remains to be investigated, these studies support the hypothesis that PyVs miRNAs may have a role in Human cancers.

1.7 MuPyV persistent infection

MuPyV infection in mice is characterized by a rapid systemic infection with a high virus titer followed by an efficient immune response (both humoral and cell-mediated) that rapidly clears both virus and infected cells[298,299]. However, like other PyVs, MuPyV establishes a long-time persistent infection in some tissues[300] and can induce later in life and in some models, many tumors in a variety of organs , probably after virus reactivation from the persistent state of infection[301]. The persistent mode of infection can be affected by the route of infection[298,302,303], the virus genotype[304], and by the genetic background and the age of the mouse[305,306]. These factors also affect the level and phase of viral replication, which ultimately may determines the sites of persistence[303]. A change in the microenvironment (e.g., tissue damage), immune status (e.g., immunosuppression), or hormone levels (e.g., pregnancy or stress), can reactivate the replication of MuPyV from persistent reservoirs[307,308]. Like BK and JC PyVs, MuPyV is excreted in the urine of mice at high titers during the acute phase of infection (13 to 27 days p.i.). Then, the virus level in urine decreases and episodic shedding events can be observed during the persistent phase of infection (28 to 87 days p.i.)[281,309]. The periodic shedding of MuPyV in the urine is likely to be a primary route of transmission[309], although the recent identification of BKPyV, JCPyV, MCPyV and SV40 miRNAs in saliva of HIV-infected and healthy patients reveals a persistent infection in the oral cavity, thus suggesting another potential route of transmission[310].

The infection via the respiratory tract is likely the natural route of infection, as higher doses of virus are required to infect through the stomach[299,309]. Moreover, it has been shown that virus infection through the intragastric route is delayed and the virus load in these organs was lower [311]. In support to this hypothesis, MuPyV is highly oncogenic when inoculated through intraperitoneal (i.p.) injection into newborn mice[17] but rarely when mice are inoculated intranasally (i.n.) or when the infection is naturally acquired[309,312]. Other studies showed that i.n. infection of MuPyV in newborn mice result in persistently infected mice that in turn were able to transmit MuPyV to other mice[313,314]. Thus, it is assumed that a persistent infection naturally occurs following an initial respiratory infection, probably by inhalation of infected urines or saliva.

Early studies showed that the experimental route of inoculation determines the level of virus replication, the phases of infection and the sites of persistency in mice. Dubensky and Villarreal[303] showed that i.n. infection is characterized by four phases of infection, (i) an initial primary replication phase characterized by a viral DNA

replication in the lungs reaching about 1 copy/cell at day 3 p.i., followed by (ii) a systemic phase of infection until days 10 to 12 p.i. that involves other organs like liver, spleen and kidneys. The highest level of viral DNA was detected at day 6 p.i. in lungs (1000 copies/cell) and in kidneys (10 copies/cells). The widespread of virus DNA during the systemic phase of infection is concomitant with an increase in virus titers in whole blood and serum, that also reached a maximum at day 6 p.i.[303], as well as in viral shedding in the urine and saliva[309]. Thereafter, (iii) a phase of partial clearance of viral DNA arises in all organs at 12 days p.i. with around a 10-fold decrease of viral DNA in most organs. In lungs, the DNA level was 200 copies/cells at that time. After the clearance phase, (iv) a persistent phase of infection occurs in lungs and kidneys by 22 days p.i., with viral DNA titers ranging from 1 to 10 copies/cell in lungs. Then, the level of DNA slowly decreases during the next months mostly because of the host immune response[299,303]. At 84 days p.i., MuPyV DNA persists in lungs at about 1 copy/cell and less than one copy/cell in kidneys. During the systemic and persistent infections, MuPyV DNA was detected as a free, supercoiled molecule, most likely in the cell nucleus. No integration of viral DNA was found into the host DNA (the limit of detection was 1 copy/100 cells). The clearance phase of viral DNA and the decrease in virus titer in blood and serum coincide with the appearance of MuPyV-neutralizing antibodies, that started at day 3 p.i. and rapidly increased from day 6 p.i.[303]. This study suggests that the initial respiratory infection may naturally leads to a persistent infection in lungs and kidneys of newborn mice.

On the other hand, Dubensky and Villarreal[303] showed that the i.p. infection is characterized by only three phases of infection, (i) an initial systemic phase of infection, similar to the secondary replication described with an i.n. infection, that involves the same organs (i.e. lungs, liver, spleen, and kidneys) with different levels of DNA replication. At day 3 p.i., viral DNA levels reached only 1-10 copies/cell in lungs and in the liver and 10-100 copies/cell in the spleen and kidneys. At 8 days p.i., levels of viral DNA were increased in all organs, mostly in kidneys with 100-1000 copies/cell, but not in the spleen (about 1 copy/cell). Virus DNA was then cleared from all organs (including lungs) during (ii) the clearance phase, except from kidneys, where viral DNA persists during (iii) the persistent phase of infection with 1-10 copies/cell at day 30 p.i.[303]. These results are in agreement with other studies using i.p. inoculation[299,300]. Dubensky and Villarreal hypothesized that the different virus replication in lungs following an i.n. or i.p. infection may be due to the different cell types reached by the virus infection or immune mechanisms[303]. Moreover, the authors hypothesized that the low replication of MuPyV in lungs after an i.p. inoculation may explain the absence of persistent infection in lungs. Thus, these studies suggest that the persistent infection in the urinary tract is independent on the route of inoculation, whereas the persistent infection in lung is likely to be dependent on i.n. infection.

In situ hybridization study showed that the acute phase of infection in mouse kidneys is characterized by high MuPyV DNA replication in proximal collecting tubules, while in persistently infected kidneys some cortical (but not proximal) tubule epithelial cells were positive for MuPyV DNA, thus suggesting that the cortical tubule epithelial cells, a minority cell type in kidneys, may be the site of MuPyV persistency in kidneys[304]. In a more recent study, Gottlieb and Villarreal[314] demonstrated in newborn mice that a subset of the Clara cells is the initial site of MuPyV DNA replication after i.n. The fact that Clara cells are actively maturing and differentiating in newborn mice is in line with the hypothesis that MuPyV replication may require host cell differentiation *in*

vivo[307].The authors also observed an early innate immune response consisting of infiltrating neutrophils starting before the virus DNA replication and followed by an adaptive immune response consisting of a lymphocytic infiltration at day 8-10 p.i. The authors also observed that other lung cells, although infected with MuPyV, are not permissive for replication[314].

Surprisingly, Gottlieb and Villarreal[314] showed that the first cycle of MuPyV replication in lungs after i.n. infection is not followed by a second round of lytic infection before the peak of infection in kidneys at 8-12 days p.i., i.e. the time of the clearance phase in lungs. The authors suggested that a local immune response, the target cells differentiation or the exfoliation of numerous permissive cells may limit the infection in lungs[314].

Early studies showed that CD4+ and CD8+ T cells are the primary immune cells required for the protection against MuPyV-induced tumors[315]. More recently, Wilson *et al.*[316] demonstrated that naïve anti-MuPyV CD8 T cells recruited after the acute phase of infection are primed *de novo* during persistent infection and become stably maintained and functionally competent against a smoldering viral infection. These specific CD8 T cells contribute to the maintenance of a low-level of persistent infections[316].

The tissue tropism of PyVs may be controlled by (i) the ability to penetrate into the target cell and by (ii) intracellular restrictions on the virus replication or gene expression. Indeed, as observed previously, an efficient acute infection of an organ does not necessarily result in an persistent infection in this organ[304], suggesting the presence of distinct tissue-specific mechanisms of virus replication and genes expression for acute versus persistent infections. Early studies from Rochford *et al.*[304] showed that organs can select for distinct enhancer-specific viral DNA replication during acute and persistent infections. Moreover, studies on HPyVs showed that JCPyV in the urine of healthy people has an archetype NCCR structure associated with persistent infection, while in the brain and cerebrospinal fluid of PML patients, JCPyV has a rearranged NCCR structure characterized by an increased early genes expression and virus replication compared to the archetype NCCR. Sequencing of these variants revealed that the rearranged NCCR structure results from deletion and duplication of promoter/enhancer sequences of the archetype NCCR and these rearrangements can be different in distinct sites of persistency in the same patient and have different ability to direct gene expression and viral replication[52,317–319]. Similarly to JCPyV, BKPyV persists life-long in the urinary tract of healthy immunocompetent individuals, with intermittent asymptomatic shedding into urine BKPyV strains containing an archetype NCCR[320,321]. Sequencing of BKPyV variants isolated from kidney transplant patients revealed the presence of rearranged NCCRs with partial duplications and deletions respected to the archetype NCCR sequence, that alter the composition of transcription factor binding sites[24,322]. Early phenotypic analysis shown that rearranged NCCR variants constitutively activate the early genes expression compared to the archetype NCCR[24,323]. Recently, Bethge *et al.* demonstrated that the regulatory function of the genes expression and virus replication of the BKPyV archetype NCCR is controlled by the presence, localization and composition of specific host cell transcription factor binding sites (TFBS), that may serve as viral sensors of host cell signals (i.e. Sp1, Ets1 and Nf1), as well as LT antigen binding sites[324,325]. The same authors[108] showed that the bidirectional early and late genes expression is directed and finely modulated by two TFBS for Sp1, one located in the early genes TATA-box promoter and one in the late genes TATA-like promoter. The Sp1 TFBS involve different regulations on the genes expression in function of the DNA strand directionality

and binding affinity. Taken together, these studies demonstrated that the organization of PyVs NCCRs governs persistency as well as activation/reactivation of the viral infection.

The host immune response also plays a important role in the control of PyVs infection, as described for BKPyV[24] and JCPyV[326]-mediated diseases, for example, that typically occur in individuals with altered immune functions. Overall, the final pattern of a persistent infection by PyVs is therefore governed by a dynamic relationship between virus tissue tropism and host immune response.

1.8 Mechanisms of virus latency

Virus persistence is defined as a state following an initial period of productive infection and a subsequent antiviral host response, in which viruses maintain the capacity for continued low-level replication and episodic reactivation over a life-long period. During the persistent phase of infection, viruses are not fully cleared by the host immune response and they are maintained by the modulation of both viral and cellular genes expression. The persistent modes of infection include (i) a state of latency and (2) a “smoldering” infection. Latency represents a cryptic infection characterized by a highly restricted genes expression pattern associated with a genomic persistence that does not lead to the production of viruses over a long period. A fundamental characteristic of latency is reversibility as a consequence of appropriate signals (either systemic or local) that allow productive (or lytic) infection associated with the full viral gene expression and the production of virions[327–329]. Cryptic persistent infections without reversibility are considered abortive infections and typically occur in nonpermissive cells. The restricted viral genes expression and no (or low) viral genome replication during latency represent a successful immune evasion strategy. In addition to targeting the host immune system to establish latency, latent viruses maintain latency by regulating their own genes expression and several cellular pathways to express and present a limited number of viral proteins and no viral particles, thus providing for a further reduced antigenicity. These mechanisms can be direct or indirect consequences of virus- and/or host-associated factors. Because latency is characterized by cryptic infections of limited host reservoirs, it is likely that under reactivation events the viruses are released at high titers from very limited sites of latent infection. On the contrary, a “smoldering” infection is defined as a long-term chronic infection characterized by a low-level of continuous viral genome expression and replication. Under reactivation events, it is likely that the shed viruses released come from more sites of persistent infection.

Few cases of true latency have been deeply described in literature, and despite important progress in the field, there is still a very limited understanding of the mechanisms that control the establishment and maintenance of latency, as well as the reactivation of lytic infections in the natural host reservoirs.

Currently, there is no clear evidence that PyVs can establish a latent infection. However, Herpesviruses, a huge family of highly prevalent DNA viruses in the Human population, have been reported to undergo true latency from which they can episodically reactivate and cause serious pathologies, including cancers. Herpesviruses are enveloped linear dsDNA viruses classified into three subfamilies called α, β, γ. α-Herpesviruses include the closely related Herpes Simplex Virus types 1 and 2 (HSV-1 and HSV-2), and varicella-zoster virus, they achieve latent infections in neurons, β-Herpesviruses have a variable tropism, for example, the pathogenic cytomegalovirus establishes latency in monocytes, while γ-Herpesviruses which comprise Epstein–Barr virus

(EBV), and Kaposi's Sarcoma-associated Herpesvirus (KSHV), commonly establish latency in lymphocytes and can transform latently infected cells.

Herpes Simplex virus 1 (HSV-1), a member of the α-Herpesviruses subfamily, represents one of the most studied authentic latent virus in the up-to-date literature. HSV-1 primarily infects epithelial cells of the oro-nasal mucosa, and less frequently, the genital mucosa. In these regions, HSV-1 initiates productive/lytic infections characterized by an abundant viral genes expression following by a temporal coordinated program of genes expression. First, the host RNA polymerase II transcribes the viral DNA into immediate early (IE) genes mRNA, IE genes products regulate the viral replication and the transcription early (E) genes. E genes protein are involved in DNA synthesis and packaging. After viral DNA synthesis, late (L) genes express the components of the capsid and tegument, and the envelope formation can occur before the release of new virions[330]. Then, viruses can spread and access to the peripheral nervous systems via the nerve terminal end of neurons of sensory ganglia innervating the site of primary infection. Upon entry, viral particles move in axons toward the neuronal cell body by retrograde axonal transport to reach the nucleus and deliver their genome[331]. The linear DNA genome found in lytic infection is rapidly circularized, chromatinized, and genes expression repressed. HSV-1 proteins recruit host epigenetic factors that regulate lytic and latent infections[332,333]. Latent genomes are mostly extrachromosomal and organized into nucleosomes with cellular histones[334,335]. During latency, infectious virus and surface proteins are undetectable, which allows the virus to evade the host immune system. In sensory ganglia, HSV-1 lytic genes are epigenetically silenced in latent infections, and the only viral gene products expressed abundantly are non-coding RNAs called latency-associated transcripts (LATs) and miRNAs[336–338]. LATs suppress viral replication and promote lytic gene silencing[339,340] through the formation of heterochromatin on lytic promoters[339–343] and protect participate to the neurons survival[344]. Kwiatkowski *et al.*[345] also demonstrated that HSV-1 exploits the mechanism of genes repression mediated by the polycomb group protein Bmi1 to establish latency. Deep sequencing of human trigeminal ganglia confirmed that LATs serve as a primary precursor for six miRNAs highly expressed only in latent infections[337,338,346]. One of them, miR-H2-3p has been described to downregulate the expression of ICP90, and miR-H6, another miRNA found just upstream of the LAT, downregulates the expression of ICP4[338]. ICP90 and ICP4 are two viral immediate-early proteins that function as transcriptional activators of productive HSV-1 infection[347,348]. It has been suggested that miRNAs can have a role in facilitating the establishment and maintenance of HSV-1 latency. Other studies showed that some cellular miRNAs maintain the HSV-1 latency by attenuating the viral replication and evading the host immune system. For example, miR-138 downregulates ICP0[349]. miR-101 expression induced by ICP4, an early transactivator of early and late viral genes, downregulates ATP5B, a mitochondrial ATP synthase, as well as GRSF1, the RNA-binding protein G-rich sequence factor 1, both known to promote HSV-1 replication[350,351]. Moreover, miR-23a and miR-649 are upregulated following HSV-1 infections to help the virus evasion from both innate and adaptive immune responses[352–354]. Another example of immune system evasion mediated by cell encoded-miRNA is miR-H8 that inhibits the PIGT-mediated presentation of surface proteins like NK-cell ligands and tetherin[355]. However, HSV-1 may reactivate following immunosuppression or other stimuli like stress, trauma to sensory nerves and ultraviolet light[356,357]. Not surprisingly, upon reactivation the pattern of latency is reversed, i.e. lytic genes are associated with markers of

euchromatin and accumulation of transcripts, while the *LATs* gene is associated with a decrease in active chromatin and transcripts level[358,359] that overall lead to the production of virions. After reactivation, progeny viral particles travel back from the ganglia to nerve terminals, resulting in re-infection of the dermis or other tissues mostly at or near to the primary sites of infection. The reactivation can manifest in localized vesicular eruption (i.e. herpes labialis) that is likely the main route of virus transmission.

In summary, HSV-1 establishes and maintains latent infections in neurons of sensory ganglia by epigenetic silencing of lytic genes expression mediated by LATs through the accumulation of heterochromatin on lytic genes promoters, and by viral and cellular miRNA-mediated inhibition of viral genes involved in the productive infection as well as the cell survival and the immune response. The chromatin state and both cellular and viral miRNAs are important regulators of the establishment, maintenance and reactivation from HSV-1 latent infection.

Human Cytomegalovirus (HCMV), a member of the β-Herpesviruses subfamily, can establish latent, "smoldering" or productive infections, even concomitantly, depending on the host cell types and/or status. Indeed, during the primary infection, HCMV can infect several cell types where the virus can establish a productive infection and then can be eventually cleared by the immune system. In other cells, like endothelial and epithelial cells, the virus can produce a long-term chronic infection resulting in low-level virus replication and shedding (i.e. a "smoldering" infection). On the other hand, hematopoietic progenitor cells and cells of the myeloid lineage provide the reservoir for latent infections that are characterized by the absence of active viral replication[360,361]. A fundamental characteristic of HCMV latency is the suppression of immediate early genes expression. The immediate early genes expression is controlled by the major immediate early promoter and enhancers that are likely epigenetically silenced soon after the entry into the nucleus since the virus genome is rapidly chromatinized[362–364]. These epigenetic modifications can be mediated by both direct or indirect binding of cellular[365] and viral[366] factors (i.e. long non-coding RNA) to the viral regulatory sequences. Other cellular factors, such as cell type-specific miRNAs inhibit the viral gene transcription to support latency[367]. Some viral proteins, like UL138, have also been shown to inhibit the viral replication, thus promoting the viral latency[368]. Like other Herpesviruses, HCMV latency requires the maintenance of the episomal viral genome in daughter cells during the cell proliferation. It has been shown that IE1 is involved in the interaction with the cell chromatin and may tether the host chromosome to the viral episome[369]. Other viral proteins like LAcmvIL-10 have immunomodulatory activity and enhance cell survival[370,371], thus promoting the maintenance of latently infected cells. In addition, the presence of host transcription factor binding sites in the viral genome act as sensors that allow the virus to regulate its replication based on the host signaling[372–375]. HCMV also encodes several miRNAs that target mostly cellular gene transcripts to favor the immune evasion. For example, HCMV-miR-UL112, miR-US25-3p and miR-UL148D help to evade the NK cells recognition[376,377]. Few HCMV-encoded miRNAs directly target viral transcripts to reduce the virus replication in order to enhance latency, for example, miR-UL112-1 targets the major immediate early transactivator IE72[378].

Another example of latent viruses well described in literature is Epstein-Barr virus (EBV), the first γ-herpesvirus identified and currently the best characterized of its sub-family. EBV primarily infects human epithelial and B lymphocytes and establishes a lifelong persistent and asymptomatic infection in memory B cells in

the peripheral blood of approximately 90% of the population worldwide[379]. Like other members of the herpesviruses family, EBV can switch between latency and lytic state. Reactivation from latency can occur following different kind of stimuli but EBV infections can remain under the immune control of healthy individuals[379,380]. During latency, the viral genome is not fully silenced and EBV can express few RNAs and proteins[380]. EBV genome includes six latency promoters named Wp, Cp, Qp, latent membrane protein (LMP)1, LMP2A and LMP2B promoters that regulate the viral gene expression and the cell survival during distinct periods of infection. These promoters are themselves controlled by viral and cellular transcription factors, as well as epigenetic markers. EBV uses host DNA methylation to specifically silence viral gene expression during the progression of the infection. As a result, and depending on the cellular context, EBV can undergo limited and distinct patterns of gene expression during latency, named latency types 0, I, II or III[379,381,382]. Importantly, Cp plays a fundamental role in the expression of six Epstein Barr nuclear antigen (EBNA) 1 to 6 that are essential for the establishment of latency. EBV is also known to induce lymphoproliferative and epithelial malignant diseases through different gene expression programs. In particular, it has been shown that EBV can transform B cells during the latency type III via activation of cellular pathways, including the NF-kB pathway, by the latent membrane protein 1 (LMP1)[383–386]. In proliferating cells, EBV uses the cellular replication machinery and a latent antigen, called EBV nuclear antigen 1 (EBNA1), which binds to the origin of replication (oriP), to promote the episomal DNA replication. Moreover, EBNA1 tethers the viral episomal genome to host chromosomes in order to passively transmit the virus genome to daughter cells during mitosis[387]. In addition, EBNA-1 has been shown to be involved in the shift from promoters regulated by cellular transcription factors to promoters controlled by viral factors. EBV also encodes abundant miRNAs targeting both viral[388,389] and cellular genes expression, to contributing to the tumorigenesis[390], cell survival[391] and to evade the immune response[392,393]. Overall, these non-coding RNAs are likely to be involved in the establishment and maintenance of viral latency.

In summary, latent EBV is established and maintained in proliferating cells by using the cellular replication machinery, cellular pathways, non-coding RNAs and viral encoded latent antigens. During EBV latency, epigenetics markers are important mechanisms to control both viral and cellular genes expression.

2. Hypothesis and objective

The working hypothesis of my PhD study is that PyVs can undergo long-term persistent infections through a "smoldering" infection in specific organs or a true latency in cryptic reservoirs akin to Herpesviruses. The choice of the mode of persistent infection may be controlled by PyV miRNAs.

Currently, a mechanistic understanding for how PyVs establish, maintain and reactivate from life-long persistent infections in their natural host reservoirs is missing. However, deciphering these mechanisms may lead to new therapies and preventative strategies against serious PyV-mediated diseases in Human. The objective of my PhD thesis was to demonstrate if PyVs undergo a "smoldering" infection, as commonly suggested in literature, or a true latent infection. In order to define the mode of persistent infection, I used the MuPyV, a prototypic model to study PyV- and host-associated mechanisms involved in persistent infections *in vivo*. The MuPyV is an attractive model because it recapitulates much of the biology of pathological HPyVs such as the genome organization, the genes and miRNA expression in the context of its natural host. Previous attempts to elucidate the mechanisms of viral persistence *in vivo* have been challenging because they rely on harvesting tissues overtime. Indeed, even using many experimental animals, the number of biological samples at different time p.i. per single animal remains limited, thus making in-depth longitudinal studies incomplete. MuPyV, like JC and BK PyVs for example, establish long-term persistent infections in specific organs, such as the urogenital tract, and sporadically shed in urine during persistent infections in healthy hosts[281,321,394]. Preliminary experiments in our lab demonstrated a partial correlation between virus titers shed in the urine and titers in kidneys. Based on these observations, we reasoned that monitoring viruses shed in the urine may allows us to non-invasively study the state of infection and reactivation overtime in the kidneys and possibly in other organs. Another major concern in previous studies on persistent infections was to consider the average of a population of viruses produced and released from different cell types or tissues instead of single viruses from a specific location. In order to overcome this issue, I engineered a library of genetically barcoded MuPyVs to tract single genomes *in vivo* and to define the early and late genes expression patterns associated to each genome involved in persistent infections. This innovative approach combined with a non-invasive urine-monitoring system that enables the longitudinal study of PyV persistence in the kidney/urogenital tract will allow to elucidate the mode of PyV persistence *in vivo* using a limited number of mice.

Based on this model, subsequently to virus reactivation from true latent infections, I expected to see high titers of few barcoded virus genomes only sporadically shed in urines during the persistent phase of infection. Moreover, I expected to see latent genomes in a limited number of specific organs during the late phase of persistent infection. At (or close to) the period of virus reactivation, these genomes will be associated with a full genome expression pattern, and likely with high titers, while at the time of the viral latency, these genomes will be associated with highly restricted early-late gene expression patterns, and likely with low titers (I expect that few cells are infected during latency). On the other hand, in the case of "smoldering" infections, I expected to see low-to-medium titers with a high diversity of barcoded virus genomes continuously shed in urines, and higher titers of a huge variety of barcoded virus genomes sporadically shed in urine during virus "reactivations" from long-term chronic infections (e.g., "reactivations" mediated by systemic cues). In the case of "smoldering"

infections, I expected to see a high diversity of barcoded genomes in different organs that will be associated with a full genome expression pattern, and likely with a high titer. In this model I also expected to see both "smoldering" and latent infections, akin to some Herpesviruses (Figure 4).

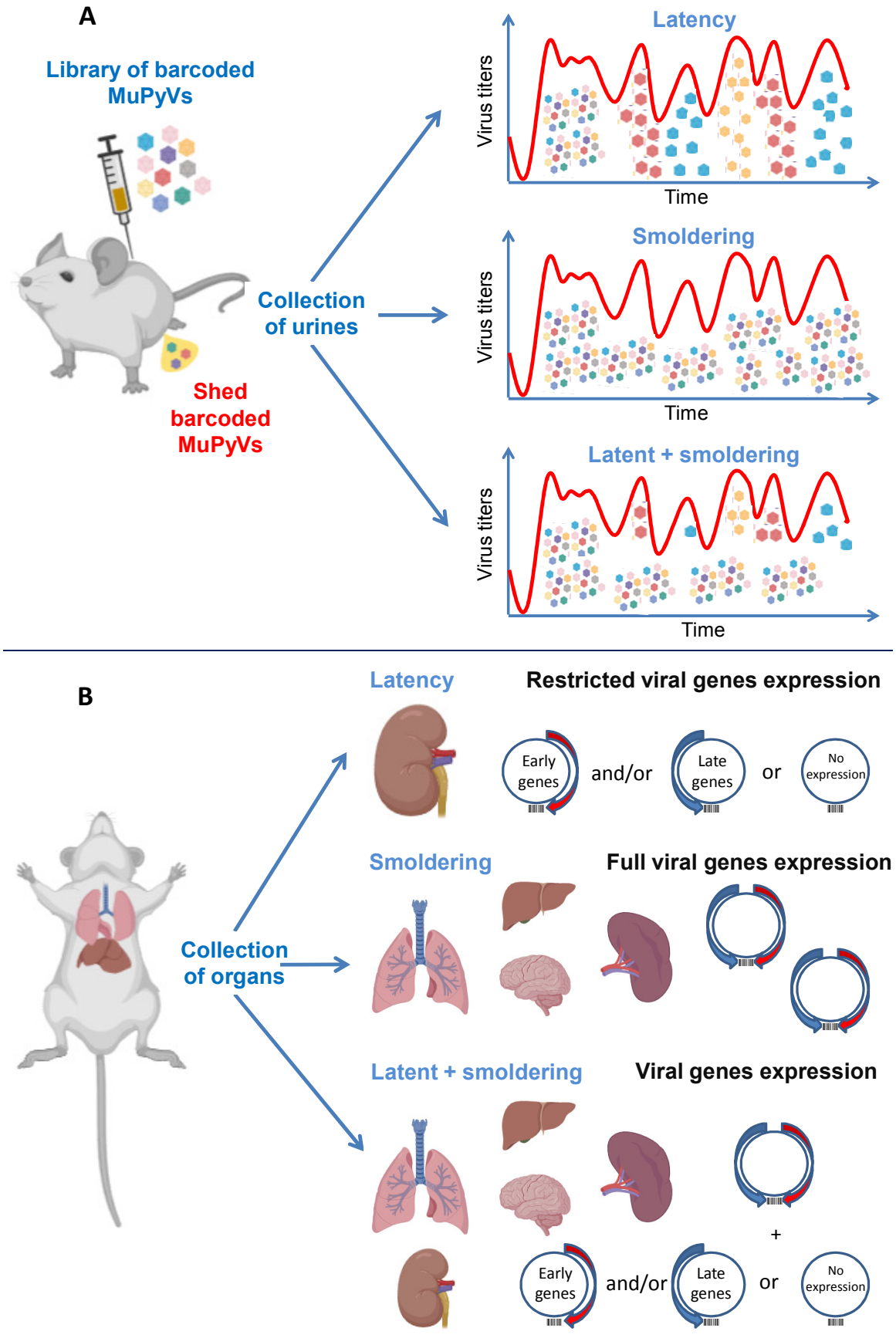

Figure 4 : Model of the hypothetical MuPyV modes of persistent infection. A-Barcoded MuPyVs shed in urines; B-Barcoded MuPyV gene expression patterns in organs post-mortem. Cartoon not in scale partially created with BioRender.com.

3. Materials and methods

3.1 Preparation of the barcoded MuPyVs library

The naturally occurring MuPyV PTA strain (GenBank accession No U27812) has been used as a background for the construction of the barcoded MuPyVs library. The pBluescript-sk+PTA vector containing the full PTA genome at the BamHI site was purified using PureLink™ HiPure Plasmid Filter Maxiprep Kit (Invitrogen). A 12-random nucleotide barcode, which potentially allows for the production of 1.68E+07 different barcodes, as well as the MfeI restriction site (for the detection of barcoded genomes) have been cloned between the early and late poly A sites in the MuPyV genome by inverted PCR using the 5' phosphorylated forward primer BCFSB7 5'-NNNNNNCAATTGAATAAACTGTGTATTCAGCTATATTC-3' and the 5' phosphorylated reverse primer BCRSB7 5'-NNNNNNGAATAAACATTAATTTCCAGGAAATAC-3' (Figure 5). Briefly, 50µl of the PCR reaction contained 100pmol of the purified pBluescript-sk+PTA vector, 0.5µM of each primer, 200µM of each dNTP, 1U of Phusion® High-Fidelity DNA polymerase (Thermo Scientific) in 1X of Phusion® HF reaction buffer. After an initial denaturation step for 1 min at 98°C, the amplification was performed by 35 cycles of 10 sec at 98°C, 15 sec at 56°C, 4 min 30 sec at 72°C followed by a final extension step for 5 min at 72°C. The PCR products obtained have been digested by DpnI (NEB) for 2 hours at 37°C and gel purified using the QIAquick Gel Extraction Kit (Qiagen). Then, 500ng of PCR product was self-ligated overnight at 16°C using the T4 DNA ligase (NEB), following the manufacturer's recommendations. The ligation product was purified and concentrated using the DNA Clean and concentrator-5 kit (Zymo Research) and 4.72ng (equivalent to 4.23E+08 plasmids) was used to transform 100µl of the MAX Efficiency® DH5α™ competent cells (Invitrogen). Transformed bacteria were cultured overnight in 200mL of LB medium + 100µg/mL of ampicillin and the pool of plasmids was purified using the PureLink™ HiPure Plasmid Filter Maxiprep Kit (Invitrogen). The number of transformed bacteria has been evaluated on LB+100µg/mL of ampicillin plates. 22 colonies randomly chosen were cultured for plasmid MiniPrep purification and submitted to diagnostic digest by MfeI-HF (NEB) to investigate the presence of the barcode and 10/22 plasmids were processed for Sanger DNA sequencing to estimate the diversity of barcodes. Then, the pool of plasmids was digested by BamHI-HF (NEB) at 37°C and gel purified using the QIAquick Gel Extraction Kit (Qiagen) to separate the barcoded MuPyV genomes from the pBluescript-sk+ vector. The barcoded MuPyV genomes were then self-ligated overnight at 16°C using the T4 DNA ligase (NEB). The ligation products were purified and concentrated using the DNA Clean and concentrator-5 kit (Zymo Research) before transfection. The Normal Murine Mammary Gland (NMuMG) epithelial cells (ATCC® n°CRL-1636™) were plated at a density of 8.5 x 10^5 cells/mL per 10 cm dish and cultured overnight at 37°C 5% CO_2 in DMEM plus 10% FBS. At 80-90% confluency the cells were transfected with 6.84µg of purified self-ligated barcoded genomes and 7.29µg of purified self-ligated wild-type PTA genomes using the Lipofectamin2000 reagent (Invitrogen), following the manufacturer's recommendations. The transfected cells were then cultured until 90% cytopathic effect was reached (5 days post-transfection). The cells and the supernatant were collected and subjected to three quick freeze/thaw cycles of 10 min each before centrifugation at 4°C for 30 min at 6,000 rpm to remove cell debris. Before infection, fresh NMuMG cells were plated at a density of 4 x 10^5 cells/mL per 10 cm dish and cultured overnight at 37°C 5% CO_2 in DMEM plus 10% FBS. At 50-60% confluency the cells were washed once with pre-warmed DMEM medium without FBS and infected with 2.5mL of crude lysate for 2 hours

at 37°C with gentle rock each 15 min. After infection, the supernatant was discarded and the cells washed once with 10mL of pre-warmed complete medium. The infected cells were then cultured until confluency (2 days p.i.) and split with the supernatant into 15 cm dishes. The infected cells were cultured again until 90% cytopathic effect was reached (day 6 p.i.). The cells and the supernatant were collected and subjected to three quick freeze/thaw cycles of 10 min each. The pH was adjusted at 8.5 and the crude lysate was incubated at 42°C for 30 min, vortexed thoroughly and centrifuged at 2,000 rpm for 10 min at room temperature to remove cell debris. The purified stocks were aliquoted and stored at -80°C until use.

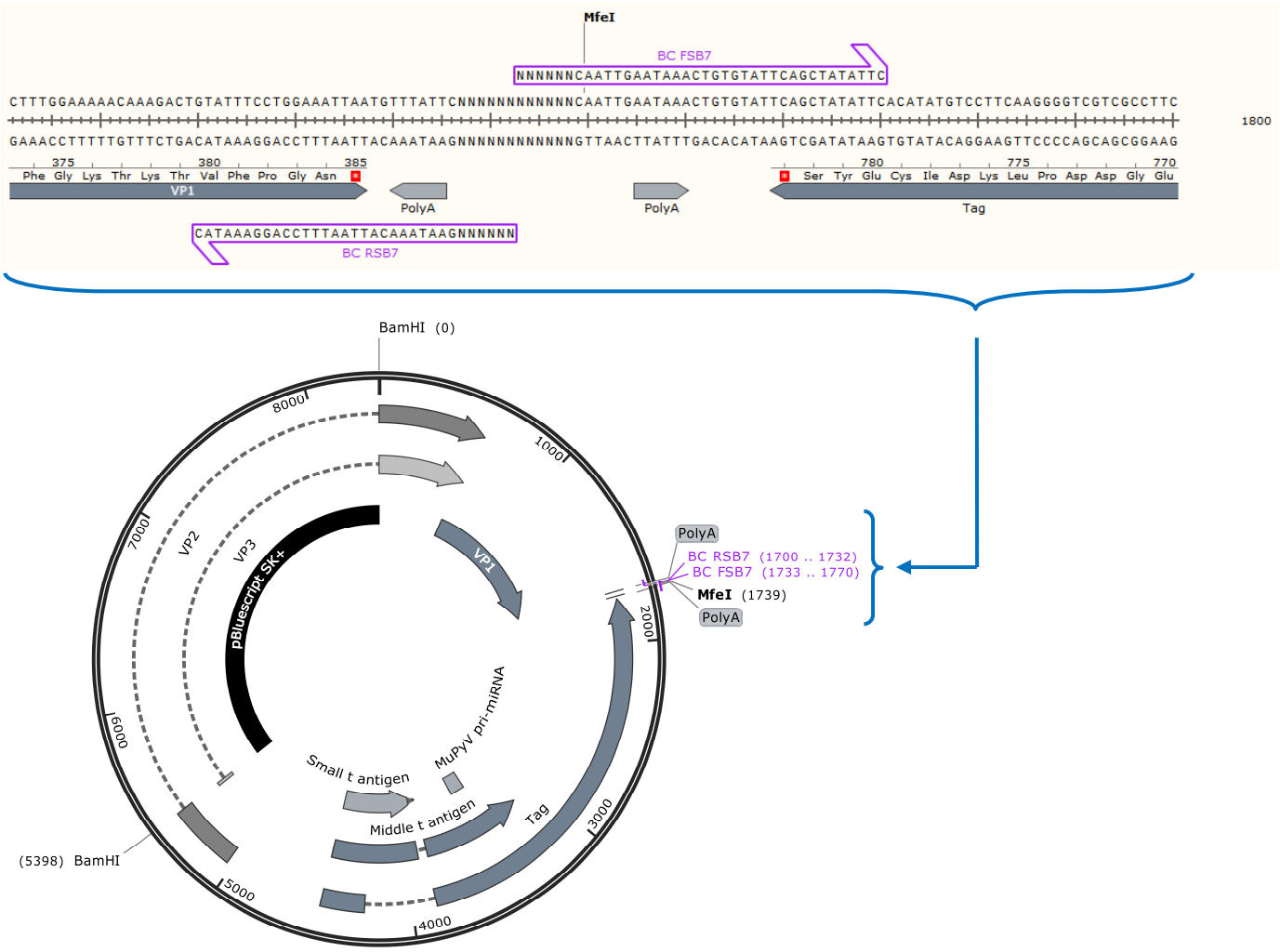

Figure 5 : Location of the primers used to clone the barcode and MfeI restriction site in the pBluescript-sk+PTA vector. Created with SnapGene®.

3.2 Virus titration by immunofluorescence assay

Virus titers of stocks and samples from the life cycle study were determined by immunofluorescence. NMuMG cells were seeded at 6 x 10^5 cells/mL per well in 24-well plates. At 80-90% confluency the cells were washed once with pre-warmed DMEM medium without FBS and infected in duplicate with 100µl of 10^{-1}, 10^{-2} and 10^{-3} dilutions of the virus sample for 1 hour. Then, the supernatant was removed and replaced by fresh pre-warmed complete media and the infected cells were cultured at 37°C. At 30-40 hours p.i., the cells were washed once with PBS and fixed/permeabilized with a solution at 4% paraformaldehyde in PBS for 10 min at room temperature. Cells were washed three times and blocked with 1% goat serum in PBS overnight. Then, the cells

were washed and incubated with a rabbit anti-PyV VP1 antibody (a gift from Richard Consigi) at a 1:50 dilution in PBS for 30 min. After three washes with PBS, the cells were stained with an Alexa Fluor 488 goat anti-rabbit IgG (ThermoFisher Scientific) for 10 min and washed four times with PBS. The number of stained cells per field was counted under an inverted fluorescence microscope (Leica). The infectious titers (IU/mL) were calculated using the formula: (average number of positive nuclei per field of view x the number of fields of view per well x the dilution factor) / volume of inoculum (mL).

3.3 Animal infections and samples collection

Two males and two females FVB/NJ mice between 9 and 10 weeks of age were inoculated with 10^6 IU of barcoded MuPyVs via i.p. injection. To collect urine, mice were placed on separated wire-bottomed cages for 2 to 4 hours. Urine samples were stored at -80°C until use. At the end of the longitudinal study (i.e. 59 and 99 days p.i.), the following samples were harvested and split in 2 aliquots for DNA and RNA purification: bladder, brain, gut, heart, kidney, liver, lung, muscle, salivary gland, spleen, testicle, whole blood (or plasma) and 1 lung tumor. Samples for DNA purification were snap frozen in liquid nitrogen while samples for RNA purification were harvested in 1mL TRIzol™ Reagent (Invitrogen) and quickly homogenized with a bead beater (Bead Mill4, Fisher Scientific) before storage at -80°C until use. Animals were cared for in accordance with protocols approved by the Institutional Animal Care and Use Committee (AUP-2017-00037) and consistent with National Institutes of Health guidelines for the care and use of animals.

3.4 DNA extraction

In order to determine the virus titers and evaluate the complexity of the barcoded MuPyVs library by NGS, 2 x 200µl (equivalent to 10^6 IU of barcoded MuPyVs) of the virus stock (named BC7 thereafter) was first digested with 20U of DNase I at 37°C for one hour. Then, DNA was purified from all the DNaseI digestion mix using the QIAamp DNA Mini kit (Qiagen) and following the manufacturer's protocol. DNA was eluted in 200µl of AE buffer and stored at -20°C until use. DNA was purified from 140µl of whole blood and 40µl of plasma using the same kit and protocol. An additional RNase A digestion step was performed for whole blood samples.

Contrarily to the DNA purified from the barcoded MuPyVs library for the NGS library preparation, the DNA used in the preliminary estimation of the diversity of the barcoded MuPyV library was purified from 2x200 µl of the virus stock by using the QIAamp DNA Mini kit (Qiagen) following the manufacturer's protocol for blood or body fluids without DNase I digestion. DNA was eluted in 2 x 60µl of water and further concentrated in 10µl of water using the DNA Clean and concentrator-5 kit (Zymo Research).

DNA was purified from 50µl of urine samples using the QIAamp viral RNA mini kit (Qiagen), following the manufacturer's protocol. DNA was eluted in 60µl of Tris-EDTA (TE) buffer and stored at -20°C until use.

DNA from 25mg of organs (or the total if less than 25mg and 10mg of spleen) was extracted using with a bead beater (Bead Mill4, Fisher Scientific) and the QIAamp Fast DNA Tissue kit (Qiagen), following the manufacturer's protocol. The DNA concentration was measured by the NanoDrop spectrophotometer and the integrity of each sample was checked on a 0.8% agarose gel. Samples were stored at -20°C until use.

3.5 PCR, cloning and Sanger sequencing to detect the barcode in the barcoded MuPyVs library and urine

In order to have a preliminary estimation about the complexity of the barcoded MuPyVs library (named BC7 thereafter) and to monitor shed barcoded viruses in some urine samples during the study, the region including the barcode was amplified by PCR, cloned and submitted for Sanger sequencing. The 50µl PCR reaction contained 10ng of purified DNA, 0.2µM of each primer NGS_Fwd 5'-CATGGCCTCCCTCATAAGTT-3' and NGS_Rev 5'-GAATATAGCTGAATACACAGTTTATTC-3', 200µM dNTPs, 1X ThermoPol reaction buffer and 1.25U of Taq DNA polymerase (NEB). After an initial denaturation step for 30 sec at 95°C, the amplification was performed by 35 cycles of 15 sec at 95°C, 20 sec at 55°C, 30 sec at 68°C followed by a final extension step for 2 min at 68°C. The specificity of the PCR assay was checked on a 2% TAE-agarose gel stained with ethidium bromide. The presence of barcodes was detected by comparison with the wild-type non-barcoded genome and a single-barcoded MuPyV genome control (previously sequenced). The specific PCR products were purified using the DNA Clean and concentrator-5 kit (Zymo Research), cloned using the TOPO® TA Cloning kit (Invitrogen), and subsequently transformed into the DH5α™ competent cells (Invitrogen), following the manufacturers' recommendations. Single colonies were picked-up and cultured overnight in LB+ampicillin medium. The plasmids were purified using the GenElute™ Plasmid Miniprep kit (Sigma-Aldrich) and sent for Sanger sequencing.

3.6 Life cycle study of the barcoded MuPyVs library

NMuMG cells were seeded at 2 x 10^5 cells/mL per well in 12-well plates and cultured overnight at 37°C in complete growth medium. When the cells reached 50% confluency, the medium was removed and the cells were washed once with pre-warmed DMEM medium without FBS. The cells were then infected in duplicate with 500µl of barcoded MuPyVs or PTA virus stock at a multiplicity of infection (M.O.I.) 5 for 1 hour with a gentle rock each 15 min. Care was taken to avoid any cross-infection of the cells between wild-type and barcoded viruses. The supernatant was then removed and the unabsorbed viruses were washed out with 1mL of pre-warmed complete growth medium. Infected cells were cultured in 1mL of DMEM plus 10% FBS at 37°C. At 4, 20, 24, 28, 40, and 48 h p.i. the supernatant was discarded and the cells washed once with 1mL pre-warmed DMEM medium without FBS. 200µl of trypsin was added and the cells were incubated at 37°C for 1-2 min. 500µl of complete growth medium was added twice to harvest all the cells from the well. The cells were then centrifuged at 7,000rpm for 20 min at RT and 100µl of PBS was added to the cell pellets. The cell pellets were then quickly freezed/thawed three times and centrifuged again at 7,000rpm for 20 min at RT to remove cell debris. The supernatants were collected and the virus titers were determined by immunofluorescence, as described above.

3.7 Real time-qPCR assay for the MuPyV DNA quantification

A specific real time-qPCR assay was performed to quantify the viral DNA load in the barcoded MuPyVs library, urine, whole blood, plasma and tissue samples. DNA from tissues was freshly diluted to 100ng/ul or 10ng/ul in water before qPCR. The 20µl qPCR mixture contained 10µl of PerfeCTa® SYBR® Green FastMix®, ROX™ (Quantabio), 8pmol of each primer PTA/PTA-dl1013 sense 5'-GATGAGCTGGGGTACTTGT-3' and PTA/PTA-dl1013 antisense 5'-TGTATCCAGAAAGCGACCAAG-3' and 2µl of DNA template per reaction. qPCR reactions consisted of

an initial denaturation step of 10 min at 95°C, followed by 40 cycles of 15 sec at 95°C, 30 sec at 60°C. Fluorescence was measured during each extension step. The specificity of each PCR was checked by melting curve analysis. Real time-qPCR was performed on a StepOnePlus real-time PCR system (Applied Biosystems) and analyzed using the StepOne software v2.2.2. The copy number of MuPyV DNA was determined in duplicate in reference to a standard curve ranging from 10 up to 10^8 copies per reaction also prepared in duplicate by ten-fold serial dilution of a single pBluescript-sk+PTA barcoded vector (fully sequenced previously). The copy number of MuPyV DNA was normalized per µl for urine, whole blood, plasma and virus stock, and per µg of total DNA in tissues. The limit of detection was 10 copies per reaction.

3.8 Illumina NextSeq library preparation

I used a customized multiplexing protocol for the ultra-deep sequencing of 121 PCR amplicons each containing potentially thousands of barcodes. The first step of the protocol was to enrich my sample and then to add adapters and indexes to my libraries by PCR.

3.8.1 Enrichment PCR

The enrichment PCR amplified a 360 bp fragment encompassing the barcode region of MuPyVs genome using primers NGS_Fwd 5'-CATGGCCTCCCTCATAAGTT-3' and ReVOUT1 5'-CAGGGTCTTGTGAAGGAGGT-3' (Figure 6). This protocol has been carefully optimized to limit as much as possible PCR bias; i.e. optimization of the annealing temperature allowing the best efficiency of amplification with no primer-dimers formation. The ramp of the thermocycler was also reduced by 50%, as reported previously[395].

To avoid the saturation phase of PCR, I sorted the urine samples based on the Ct value obtained in qPCR and varied the number of cycle to reach (at the end of the PCR reaction) a maximum theoretical copy number comprised between 2.46 x 10^{12} and 2.35 x 10^{14}. Briefly, 50µl of the enrichment PCR reaction contained 20µl of DNA purified from urine (≤4.77 x 10^5 copies per PCR reaction), 0.5µM of each primer, 200µM of each dNTP, 1U of Phusion® High-Fidelity DNA polymerase (Thermo Scientific) in 1X of Phusion® HF reaction buffer. After an initial denaturation step for 1 min at 98°C, the amplification was performed by 27-42 cycles of 10 sec at 98°C (ramp 2°C/sec), 15 sec at 55°C(ramp 2°C/sec), 30 sec at 72°C (ramp 2°C/sec) followed by a final extension step for 5 min at 72°C. PCR reactions were stored immediately at -20°C until use. The presence of a specific amplification product was checked on a 2% agarose gel.

In the case of whole blood, plasma and virus stock samples, I used the same protocol starting the PCR with a maximum theoretical copy number of 4.77 x 10^5 per PCR reaction (calculated from the virus titer obtained by qPCR) and I varied the number of PCR cycles from 22 to 35 to stay below 2 x 10^{12} copies per reaction at the end of the PCR.

In the case of organs, I started the PCR with a maximum theoretical copy number of 4.77 x 10^5 per PCR reaction (calculated from the virus titer obtained by qPCR) within a maximum of 3.1µg of total DNA per reaction (i.e. the maximum quantity of DNA tested with no inhibitory effect on the PCR) and I varied the number of PCR cycles to stay below 2 x 10^{12} copies per reaction at the end of the PCR. Briefly, 50µl of the PCR reaction contained 22µl of DNA purified from organs (≤4.77 x 10^5 copies and ≤3.1µg of total DNA per PCR reaction), 0.3µM of each

primer, 1X of KAPA HiFi HotStart ReadyMix (Kapa Biosystems). After an initial denaturation step for 3 min at 95°C, the amplification was performed by 22-35 cycles of 20 sec at 98°C (ramp 2°C/sec), 15 sec at 56°C(ramp 2°C/sec), 15 sec at 72°C (ramp 2°C/sec) followed by a final extension step for 2 min at 72°C. PCR reactions were stored immediately at -20°C until use .The presence of a specific amplification product was checked on a 2% agarose gel.

3.8.2 Indexing PCR

The indexing PCR amplified a 75 bp fragment encompassing the barcode region of MuPyVs genome using staggered indexing primers (Table 1) generating amplicons sizes ranging from 211 to 216bp. The annealing sites of the staggered indexing primers are shown in Figure 6 (see NGS_Rev and NGS_FwdIN2 primers).

This protocol has been carefully optimized to limit as much as possible PCR bias; i.e. optimization of the annealing temperature allowing the best efficiency of amplification with no primer-dimers formation. The ramp of the thermocycler was also reduced by 50%, as reported previously[395].

30µl of enrichment PCR product have been digested with 20U of ExoI (NEB) at 37°C for 30 minutes to remove residual enrichment PCR primers and heat inactivated at 80°C for 20 min before proceeding to the indexing PCR. Two indexing PCR reactions have been made for each sample.

Briefly, 50µl of the indexing PCR reaction contained 1µl of the enrichment PCR products, 0.3µM of indexing primers following the strategy shown in Table 2, 200µM of each dNTP, 1U of Phusion® High-Fidelity DNA polymerase (Thermo Scientific) in 1X of Phusion® HF reaction buffer. After an initial denaturation step for 1 min at 98°C, the amplification was performed by 22 cycles of 10 sec at 98°C (ramp 2°C/sec), 15 sec at 54°C(ramp 2°C/sec), 30 sec at 72°C (ramp 2°C/sec) followed by a final extension step for 5 min at 72°C. PCR reactions were stored immediately at -20°C until use. The presence of a specific amplification product was checked on a 2% agarose gel.

	Name	Index adapter	i5 or i7 index	Index adapter	Spacer
i5 index forward primers	>NGS_FwdIN2-i5-U6-F3-D501	AATGATACGGCGACCACCGAGATCTACAC	TATAGCCT	ACACTCTTTCCCTACACGACGCTCTTCCGATCT	CAG
	>NGS_FwdIN2-i5-U6-F3-D502	AATGATACGGCGACCACCGAGATCTACAC	ATAGAGGC	ACACTCTTTCCCTACACGACGCTCTTCCGATCT	CAG
	>NGS_FwdIN2-i5-U6-F2-D503	AATGATACGGCGACCACCGAGATCTACAC	CCTATCCT	ACACTCTTTCCCTACACGACGCTCTTCCGATCT	G
	>NGS_FwdIN2-i5-U6-F2-D504	AATGATACGGCGACCACCGAGATCTACAC	GGCTCTGA	ACACTCTTTCCCTACACGACGCTCTTCCGATCT	G
	>NGS_FwdIN2-i5-U6-F1-D505	AATGATACGGCGACCACCGAGATCTACAC	AGGCGAAG	ACACTCTTTCCCTACACGACGCTCTTCCGATCT	
	>NGS_FwdIN2-i5-U6-F1-D506	AATGATACGGCGACCACCGAGATCTACAC	TAATCTTA	ACACTCTTTCCCTACACGACGCTCTTCCGATCT	
	>NGS_FwdIN2-i5-U6-F4-D507	AATGATACGGCGACCACCGAGATCTACAC	CAGGACGT	ACACTCTTTCCCTACACGACGCTCTTCCGATCT	GCAC
	>NGS_FwdIN2-i5-U6-F4-D508	AATGATACGGCGACCACCGAGATCTACAC	GTACTGAC	ACACTCTTTCCCTACACGACGCTCTTCCGATCT	GCAC
	>NGS_FwdIN2-i5-U6-F5-D509	AATGATACGGCGACCACCGAGATCTACAC	TTCGGATG	ACACTCTTTCCCTACACGACGCTCTTCCGATCT	AGCAC
	>NGS_FwdIN2-i5-U6-F5-D510	AATGATACGGCGACCACCGAGATCTACAC	ACTCATAA	ACACTCTTTCCCTACACGACGCTCTTCCGATCT	AGCAC
i7 index reverse primers	>NGS_Rev-i7-U6-R-D701	CAAGCAGAAGACGGCATACGAGAT	ATTACTCG	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	
	>NGS_Rev-i7-U6-R-D702	CAAGCAGAAGACGGCATACGAGAT	TCCGGAGA	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	
	>NGS_Rev-i7-U6-R-D703	CAAGCAGAAGACGGCATACGAGAT	CGCTCATT	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	
	>NGS_Rev-i7-U6-R-D704	CAAGCAGAAGACGGCATACGAGAT	GAGATTCC	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	
	>NGS_Rev-i7-U6-R-D705	CAAGCAGAAGACGGCATACGAGAT	ATTCAGAA	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	
	>NGS_Rev-i7-U6-R-D706	CAAGCAGAAGACGGCATACGAGAT	GAATTCGT	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	
	>NGS_Rev-i7-U6-R-D707	CAAGCAGAAGACGGCATACGAGAT	CTGAAGCT	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	
	>NGS_Rev-i7-U6-R-D708	CAAGCAGAAGACGGCATACGAGAT	TAATGCGC	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	
	>NGS_Rev-i7-U6-R-D709	CAAGCAGAAGACGGCATACGAGAT	CGGCTATG	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	
	>NGS_Rev-i7-U6-R-D710	CAAGCAGAAGACGGCATACGAGAT	TCCGCGAA	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	
	>NGS_Rev-i7-U6-R-D711	CAAGCAGAAGACGGCATACGAGAT	TCTCGCGC	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	
	>NGS_Rev-i7-U6-R-D712	CAAGCAGAAGACGGCATACGAGAT	AGCGATAG	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	
i5 index reverse primer	>NGS_Rev-i5-tracr-F2-D512	AATGATACGGCGACCACCGAGATCTACAC	CGCGGCTA	ACACTCTTTCCCTACACGACGCTCTTCCGATCT	GCAC
i7 index forward primer	>NGS_FwdIN2-i7-tracr-R-D712	CAAGCAGAAGACGGCATACGAGAT	AGCGATAG	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	

Table 1: staggered indexing primers

3.8.3 Amplicons pooling and quality control

The two indexing PCR reactions were merged together and concentrated using the Monarch® PCR & DNA clean up kit (5µg) (NEB). The PCR products were then purified from gel using the Monarch® DNA gel extraction kit (NEB). The DNA concentration of each sample was determined using the QUBIT™ dsDNA BR assay kit (ThermoFisher Scientific). The 121 samples were normalized and pooled together. The pool of libraries was checked on a 2% agarose gel and sent to the Genomic Sequencing and Analysis Facility of the University of Texas at Austin for a Bioanalyzer (Agilent) quality control and Illumina NextSeq SR75 sequencing. DNA samples were stored at -20°C.

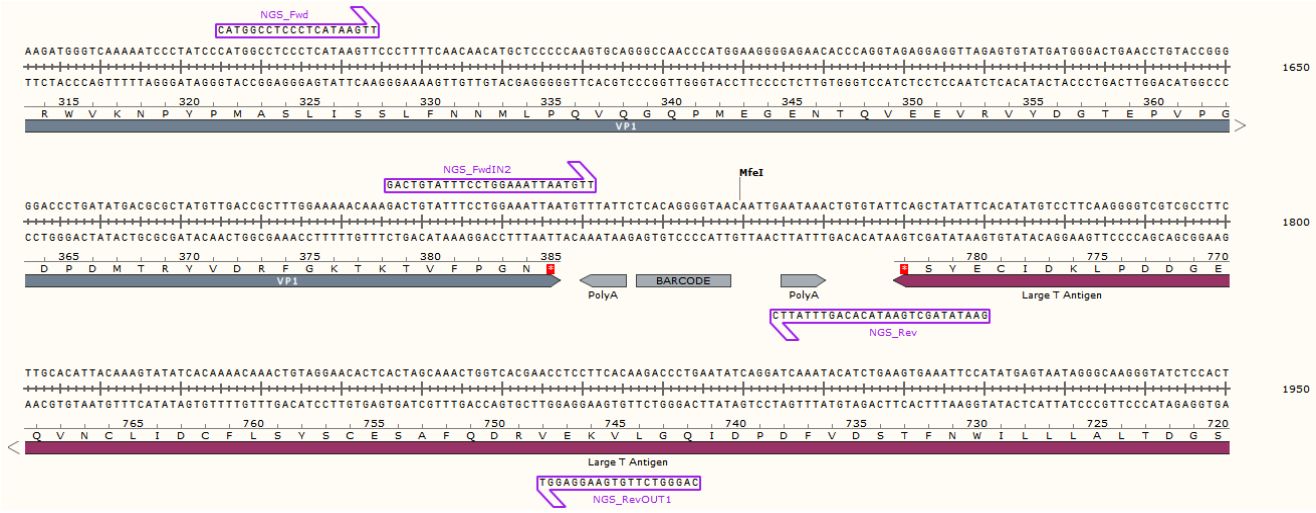

Figure 6 : Map of the primers used in the NextSeq library preparation. The nucleotide position is calculated from the BamHI restriction site in the sequence of a single-barcoded (named plasmid_1 in Figure 7). Created with SnapGene®.

		i5 index primers										
		F3		F2		F1		F4		F5		Rev
		D501	D502	D503	D504	D505	D506	D507	D508	D509	D510	D512
i7 index primers – Reverse	D701	D7FL	D9MR	D30ML	D32ML	D22FL	D1FL	BLFR	KIMR	SGML	MUMR	
	D702	D7ML	D20MR	D34ML	D5FR	D26FL	D11FL	HEFR	SGFR	LIFL	SGFL	
	D703	D8FR	D21MR	D35ML	D20FR	D32FL	D15FL	HEFL	HEMR	BRML	BRFL	
	D704	D9FR	D22MR	D24FR	D21FR	D33FL	D18FL	LIFR	SPMR	BLFL	WBML	
	D705	D11FR	D23MR	D34FR	D23FR	D35FL	D29FL	KIFL	LUML	GOFR	MUML	
	D706	D11ML	D8ML	D35FR	D27FR	D2ML	D1MR	GOFL	GU2MR	GUML	LIMR	
	D707	D11MR	D20ML	D30FL	D28FR	D2MR	D4MR	SPML	HEML	TSML	LUFL	
	D708	D12FR	D21ML	D15MR	D29FR	D1ML	D4ML	KIML	MUFL	GUFR	LTFL	
	D709	D12ML	D22ML	D14ML	D32FR	D1FR	D2FR	KIFR	TSMR	GUFL	LUFR	
	D710	D12MR	D24ML	D15ML	D14FL	D14FR	D2FL	MUFR	LIML	LUMR	WBFL	
	D711	D13ML	D26ML	D16ML	D16FL	D16FR	D23FL	SPFL	BLML	BLMR	WBFR	
	D712	D5MR	D28ML	D19ML	D19FL	D26FR	D5ML	SPFR	BR2MR	BRFR	BC7	
Forward	D712											PLMR

Table 2: Strategy of indexing primers multiplexing used.

3.8.4 Bioinformatics analysis

Adapter and flanking sequences of index primers sequence high-throughput sequencing reads were removed with Cutadapt[396] using the following parameters : cutadapt -j 0 -n 2 -e 0.25 -g GACTGTATTTCCTGGAAATTAATGTTATTC -a CAATTGAATAAACTGTGTATT. The remaining sequences (barcodes) were counted for each sample.

Barcodes were sorted by maximum read count across all samples and then clustered based on having absolute edit distance (Levenshtein) of 2 or less and are named with the most abundant barcode of the cluster. Counts for each barcode were collapsed by summation with other members of the cluster. Barcode cluster counts were standardized per million barcode reads in the sample and filtered for minimum of at least 1 bpm in any sample.

Each barcode cluster was converted to the fraction of each sample by dividing by the amount of total remaining barcode reads for each sample. Titers for barcode clusters were approximated by multiplying by titer determined from qPCR and filtered for at least one sample having titer of 10 per unit or greater. Titers are in the units specific to the sample type (i.e. barcode per µl of urines, barcode per µl of whole blood, barcode per µg of total DNA in tissue).

Urine samples at the time of shedding events during persistent infections were selected, as well as the organs for each mouse. Barcodes with high titers in selected urine samples investigated in shedding events and organs from the same animal.

4. Results and discussion

4.1 Estimation of the barcoded MuPyVs library complexity

I constructed a library of barcoded MuPyVs by cloning a 12-random nucleotides barcode in the MuPyV PTA strain genome between the early and late poly A sites. The combination of 12 nucleotides potentially allows for the generation of 1.68×10^7 different barcoded MuPyV genomes. I estimated the complexity of the pool of plasmids obtained after bacterial transformation by counting the number of colonies on a LB+ampicillin plate cultured in parallel with the pool. Assuming that one colony forming unit was transformed with one plasmid containing one barcoded MuPyV genome, I found 5,678 barcoded MuPyV genomes in the pool of plasmids. Then, I screened the presence of the barcode in 22 randomly selected plasmids by enzymatic digestion using the restriction site MfeI inserted close to the barcode. 21 out of 22 plasmids had a barcode between the early and late poly A signals. To further confirm the presence of the barcode and to estimate the diversity of the pool of barcoded plasmids, I sequenced the region including the barcode in 9 plasmids (with a barcode) and 1 plasmid without barcode (i.e., plasmid_7) by Sanger DNA sequencing. I found that these 9 positive plasmids contained a different barcode in the MuPyV genome (Figure 7). Even if the number of plasmids sequenced was limited, these results suggested a diversity of barcodes in the pool of plasmids and allowed me to start the production of the barcoded viruses library (named BC7 thereafter), as described in the paragraph 3.1. After infection of NMuMG cells, the BC7 library complexity has been evaluated by Sanger sequencing, as detailed in paragraph 3.5. I found that 20 out of 21 sequenced plasmids contained a different barcode. The reason I obtained 2 similar barcodes

could be explained either by the saturation conditions reached in the PCR assay (i.e. limited Taq polymerase units versus high template copies before the end of PCR) or by the complexity of the BC7 library. Of note, one of the plasmid had a point mutation in the early genes poly A site. This mutation has been probably introduced during the PCR step preceding the cloning for sequencing (Taq DNA polymerase without proofreading activity) (Figure 7). These encouraging results suggested that the complexity of the BC7 library was high enough to start the animal experimentation. The complexity of the barcoded MuPyVs library was precisely evaluated using the Illumina NextSeq, as described below.

Figure 7 : Multiple sequences alignment using Clustal Omega of the barcode region. The sequences shown are from the pool of barcoded plasmids (Plasmid_1 to 10), the BC7 library (21 BC7 TOPO® plasmids), and one urine sample (D6MR) collected at day 14 p.i. (6 TOPO® plasmids) and the PTA reference sequence. Identical barcodes: samples BC7_35 and BC7_43). In yellow: point mutations.

4.2 Life cycle study of the barcoded MuPyVs library

Following the preparation of the BC7 library from the supernatant of infected NMuMG cells and the estimation of the library complexity by Sanger DNA sequencing, I determined the virus titer in both the wild-type MuPyV (PTA) and the BC7 library stocks by using the immunofluorescence assay described in paragraph 3.2. Surprisingly, the PTA and the BC7 library stocks titers were very similar: 7.14 x 10^6 IU/mL and 2.50 x 10^6 IU/mL, respectively. This result suggested that the insertion of the genetic barcode did not seem to negatively interfere with the replication of barcoded viruses respected to the wild-type strain. In order to evaluate more precisely the

impact of the barcode on the viral fitness, I studied the life cycle of the BC7 library over two cycles of infection[397] in highly infected NMuMG cells (M.O.I. 5), compared to the wild-type strain, as detailed in paragraph 3.6. I evaluated the production of intracellular live virus progeny (i.e., cell-associated viruses) at 4, 20, 24, 28, 40, and 48 hours p.i. using the immunofluorescence assay. As expected, I observed a linear increase in the PTA virus titers up to 28 hours p.i., followed by a slower increase up to the end of the experiment (at 48 hours p.i.). At 48 hours p.i., I started to see a slight cytopathic effect in the cell culture. Interestingly, no difference was observed between the wild-type PTA strain and the BC7 library replication (Figure 8). This study confirmed that the insertion of 18 nucleotides (12-nucleotide barcode + the MfeI restriction site) between the early and late poly A sites did not influence the virus fitness *in vitro* and suggested that the barcoded viruses could replicate similarly to the wild-type virus *in vivo*. This result also suggested the presence of a potential neutral genomic region between the two poly A sites. This is the first evidence, since my current knowledge, that a PyV has been successfully barcoded with no evident impact on the viral fitness *in vitro*, respected to the wild-type strain.

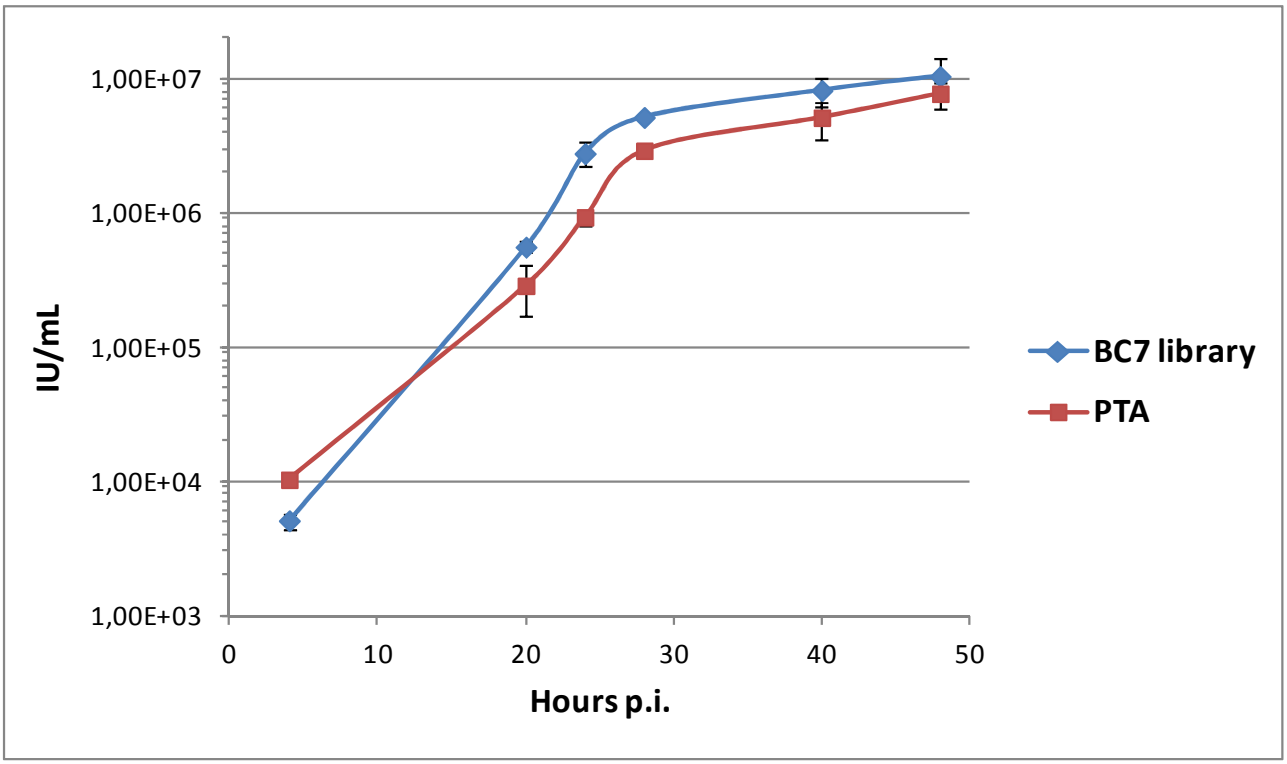

Figure 8 : Life cycle study of the barcoded MuPyVs library (BC7 library) and the wild-type MuPyV strain (PTA) in NMuMG cells infected at M.O.I. 5. The cell-associated virus progeny titers were evaluated at 4, 20, 24, 28, 40, and 48 hours p.i. by immunofluorescence assay, as described in Materials and Methods.

4.3 Barcoded MuPyVs shedding in mice urine

Despite significant advances in understanding the mechanisms involved in the establishment, maintenance or reactivation from persistent infections by DNA viruses, the mode of long-term persistence mediated by PyVs remains poorly defined. During my PhD project, I used an innovative longitudinal approach combined with a library of genetically barcoded MuPyVs to investigate, at the level of single barcoded viruses, the persistent phase of infection in a natural host reservoir.

In this study, two adult males and two adult females FVB/NJ mice were inoculated with 10^6 IU of the BC7 library via i.p. injection. Previous studies showed that high titers of viruses are excreted in the urine of mice

during the acute phase of infection and then episodically shed during the persistent infection[281,309]. In order to validate our approach *in vivo*, urine samples from infected mice were collected three or two times per week. DNA was purified from each sample and the genome copy number equivalent per µl of urine was quantified by real time qPCR, as described in paragraph 3.7. Moreover, a sample harvested at day 14 p.i. was sequenced to confirm the diversity of barcoded viruses shed in urine during the acute phase of infection.

Under experimental conditions described in paragraph 3.3, all the infected mice shed MuPyV DNA in urine over the full period of this study. Due to the substantial variations in the genome equivalent copies per µl of urine, I represented the results obtained for each mice (Figure 9). Overall, I observed high viral DNA titers shed in urines from males than females comparable to those published in a previous study using the wild-type PTA strain in a closely related mouse strain[281]. Interestingly, between days 1 and 9 p.i., the virus DNA titers shed by females decreased rapidly below the limit of detection (12 copies per µl of urine), while the MuPyV DNA titers shed by males increased between days 1 and 5 p.i. before decreasing below the limit of detection at day 9 p.i. These shedding events are concomitant with the initial systemic phase of infection (previously described in paragraph 1.7) lasting at least up to the day 9 p.i. in my experimental conditions. Previous studies shown that during this early phase of infection, the virus DNA spreads in many organs, mainly in kidneys and lungs[303], suggesting the beginning of widespread productive infections. The increased virus titers observed at day 5 p.i. in males could be associated with the virus replication in kidneys, as this organ is part of the urinary tract and described as a site for early MuPyV infection[303]. Between days 9 and 12 p.i., I observed a quick increase in virus DNA titers ranging from 5 to 6 x 10^2 and 1.29 x 10^4 to 2.63 x 10^3 MuPyV genomes equivalent/µl of urine in females and males, respectively. This increase of virus DNA shed in urines is concomitant with the previously described widespread of productive infections in many organs[303]. These high levels of virus DNA titers lasted up to the beginning of the clearance phase that started at day 19 p.i. in females and day 23 p.i. in males (Figure 9). As previously reported, the clearance phase is characterized by the onset of MuPyV-neutralizing antibodies that started at day 6 p.i.[299,300,303]. MuPyV DNA titers at the end of the clearance phase (day 27 p.i.) ranged from 1.13 x 10^2 to 2.93 x 10^2 and 3.94 x 10^2 to 1.96 x 10^3 genomes equivalent/µl of urine in females and males, respectively The end of the clearance phase delimits the end of the acute phase of infection (days 12 to 27 p.i.) and the beginning of the persistent phase of infection that is characterized by a lower basal DNA titer (on average), respected to the acute phase of infection, and episodic shedding events of high viral DNA titers in urines up to the end of my study. I note that during the persistent infections, the titers in males were much higher in males than in females but the amplitude of the episodic shedding was lower (Figure 9). This difference in virus replication could be due to different hormonal status between males and females[308]. Indeed, it has been shown previously that the shedding of In a previous study[281], Buke *et al.* demonstrated in a similar model that the miRNA locus of MuPyV supports the viral DNA shedding in urine during the acute phase of infection and limits the number of these events during the virus persistency. Overall, the occurrence over time as well as the level viral DNA titers in urines during the different phases of infection are consistent with a the model used by Burke *et al.*[281]. The determination of the MuPyV DNA titers in urines was a fundamental prerequisite to set the beginning of the persistent phase of infection and to construct the Illumina library.

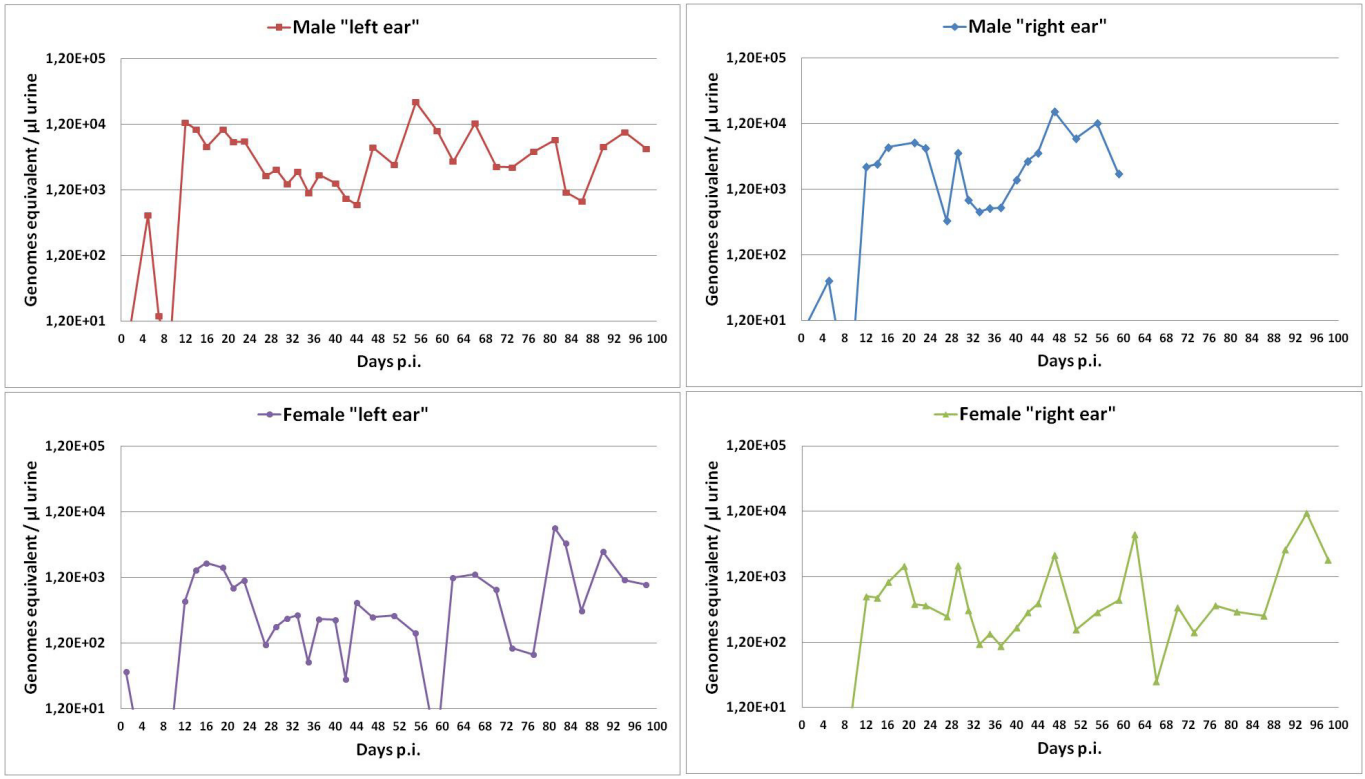

Figure 9 : MuPyV viruria in FVB/NJ mice i.p. infected with 106 IU of the barcoded MuPyVs library.
The copy number of virus genomes equivalent per µl of urine was obtained by real time qPCR analysis at various times p.i., as described in Materials and Methods. The limit of detection (12 copies/µl of urine) was set up at the interception of the x-axis.

During the early phase of this study, I also tested the presence and the diversity of barcoded MuPyV genomes shed in urine to ensure that the barcode was "maintained" in the virus genome *in vivo*. This was not obvious due to the location of the barcode in a non-coding region of the MuPyV genome. In order to do this, the DNA from one urine sample collected two weeks p.i. was purified, the genomic region containing the barcode was amplified by PCR, cloned and sequenced, as detailed in paragraph 3.5. The 6 sequences of barcodes obtained were different between them and from the sequences obtained from the BC7 library and the pool of plasmids (Figure 7), suggesting the presence of a heterogeneous barcoded MuPyV genome composition shed in urine. Of note, I observed in one of the plasmid (D6MR_12) that the single nucleotide naturally present between the two poly A sites in the wild-type strain was mutated. This point mutation has been probably introduced during the PCR amplification before cloning and sequencing (the Taq polymerase used has no proofreading activity). The precise composition of barcodes shed in urine was assayed by Illumina NextSeq at the end of the *in vivo* experimentation (as shown below). The detection of a heterogeneous composition of barcoded MuPyV genomes in urine during the acute phase of infection is the first proof-of-concept reported so far for the use of genetically barcoded DNA viruses. Moreover, the similarity between the shedding events of barcoded MuPyVs in urine and those previously reported with the wild-type strain and a closly related mouse strain[281], suggests that the insertion of a short barcode in a "neutral" region of the MuPyV genome does not impair the virus infection *in vivo*. I am proposing this approach combined to the non-invasive urine-monitoring system as an innovative strategy for the longitudinal study of the mode of PyVs persistence *in vivo*.

4.4 Persistent MuPyV infections in mice organs

Previous studies described a huge variety of PyVs infection sites; here are important examples that guided my choice to investigate the persistent mode of infection in organs. The major sites of persistent infections in healthy individuals reported for BKPyV and JCPyV are kidney and the urinary tract, from which the virus can reactivate under immunosuppression and lead to serious diseases[398]. Epithelial cells of kidney, ureter and bladder were described as predominant sites of BKPyV persistent infection[399]. BKPyV persistent infection has also been associated to bladder cancer[400]. BKPyV DNA has been detected in cervix, prostate and semen, as well as in brain[401–404]. JCPyV DNA was found in the brain, cardiac muscle, liver, lung, lymph nodes and spleen of patients with PML[319,405]. Interestingly, lymphocytes was shown to be of importance in both BKPyV and JCPyV persistence and dissemination[180]. HPyVs have also been found in the gut[406]. SV40 was also found in brain, cervical tissue, sperm fluids and PBCs[402,407]. Interestingly, MuPyV has been previously shown to infect many tissues during the early infection, including for example, salivary gland[309]. Due to the high variety of organs infected by PyVs, I was interested in investigating the persistent mode of infection in most of them.

At the end of the animal experimentation, organs and plasma from one of the two males, named male "left ear", were harvested at day 59 p.i. while organs and whole blood from the other mice were harvested at day 99 p.i. DNA was purified from bladder, brain, gut, heart, kidney, liver, lung, muscle, salivary gland, spleen, testicle, whole blood (or plasma) and 1 lung tumor. The integrity of DNA was checked and the virus DNA titers were determined by qPCR, as described in Materials and Methods.

In the male named "left ear", all organs harvested were infected with DNA titers ranging from 2.58 x 10^2 genomes equivalent/µg of total DNA in salivary gland to 8.46 x 10^4 genomes equivalent/µg of total DNA in spleen (Figure 10). The titers in bladder and liver (4.27 and 4.37 x 10^2 genomes equivalent/µg, respectively) were as low as in salivary gland, while in kidney (8.43 x 10^4 genomes equivalent/µg) the titer was as high as in spleen. In brain, gut, heart, lung and testicle titers were high and in the same range (1.39 to 3.33 x 10^4 genomes equivalent/µg) while in muscle the titer was intermediate (4.62 x 10^3 genomes equivalent/µg) and in whole blood the titer was very low (2.84 x 10^2 genomes equivalent/µl) (Figure 11).

In the male named "right ear", all organs harvested were infected with DNA titers ranging from 2.67 x 10^2 genomes equivalent/µg of total DNA in brain to 6.09 x 10^4 genomes equivalent/µg of total DNA in kidney (Figure 10). The titers in liver and testicle (9.78 and 7.68 x 10^2 genomes equivalent/µg, respectively) were slightly higher than in brain, while in gut, heart, lung, muscle and spleen titers were roughly as high as in kidney (from 1.02 to 4.72 x 10^4 genomes equivalent/µg). The titer in bladder was intermediate (8.73 x 10^3 genomes equivalent/µg, respectively). In plasma, the virus titer was below the limit of quantification (Figure 11), thus suggesting that MuPyV is rarely found in blood as a circulatory cell-free virus.

In both male, kidney was the most infected organ. Spleen, gut, heart, and lung were also highly infected, suggesting a high level of viral replication or the presence of many persistently infected cells in these organs[281,303]. The main difference in titers between the two males were observed in bladder and testicle (more than 1$\log_{10}$ of difference between males), and particularly in brain (almost 2$\log_{10}$ of difference between males). These differences may be due to the different time p.i. in which organs have been harvested.

In the female named “left ear”, all organs harvested were infected with DNA titers ranging from 4.78 x 10^3 genomes equivalent/µg of total DNA in brain to 3.69 x 10^8 genomes equivalent/µg of total DNA in lung (Figure 10). Titers in heart, kidney, spleen and muscle were the highest after the lung titer (9.06 x 10^4 to 1.53 x 10^5 genomes equivalent/µg). Titers in salivary gland (5.81 x 10^3 genomes equivalent/µg) were nearly as low as in brain. In bladder, gut, and liver titers were similar (1.02 to 1.82 x 10^4 genomes equivalent/µg) and slightly lower than in gonads (6.55 x 10^4 genomes equivalent/µg). The titer in whole blood was high (8.03 x 10^4 genomes equivalent/µl) (Figure 11), probably because of virus dissemination from highly infected organs and/or associated to white blood cells (e.g., in lymphocytes).

In the female names “right ear”, all organs harvested were infected with DNA titers ranging from 6.83 x 10^3 genomes equivalent/µg of total DNA in brain to 3.07 x 10^8 genomes equivalent/µg of total DNA in lung (Figure 10). The titers in gut and gonad (1.09 and 1.54 x 10^4 genomes equivalent/µg, respectively) were slightly lower than in kidney, salivary gland and spleen titers (from 5.91 to 7.79 x 10^4 genomes equivalent/µg). Titers in bladder, heart, liver and muscle were high and in the same range (1.05 x 10^5 to 3.55 x 10^5 genomes equivalent/µg). In whole blood the virus titer was high (1.3 x 10^5 genomes equivalent/µl) (Figure 11), thus suggesting that MuPyV is associated to white blood cells virus and/or released from highly infected organs.

In both female, lung was the most infected organ. Many other organs like heart, kidney, spleen, gonads, and muscle were also highly infected. The main titers difference between females was observed in bladder, liver and salivary gland (about 1$\log_{10}$ of difference between females).

Interestingly, virus titers in liver, muscle, salivary gland and particularly in lung from females were higher than in males (Figure 10). Unexpectedly, the female “left ear” developed a huge lung tumor with a very high titer (2.42 x 10^8 genomes equivalent/µg), which is comparable to the titer found in lung (Figure 10). In literature, MuPyV-induced tumors have been found in a wide spectrum of adult mice organs with high levels of virus genomes during persistent infections. Wirth *et al.* reported that persistent genome replication/reactivation correlates with oncogenesis and hypothesized that tumorigenesis may require a continued viral replication during persistency. However, not all persistently infected organs develop tumors (e.g., kidney), indicating that persistence is necessary but not sufficient for oncogenesis[408].

Virus titers in some organs were higher than in previous studies[281,303]. This could be due to different mouse strain, gender and age of animals used in different studies, as well as the quantity of virus inoculated. Surprisingly, I note a difference in titers as well as in the type of most infected organs between mouse genders. Another unexpected result was the very high titers found in lung of females and the development of a lung tumor in one of the two females with a titer close to the “normal” lung.

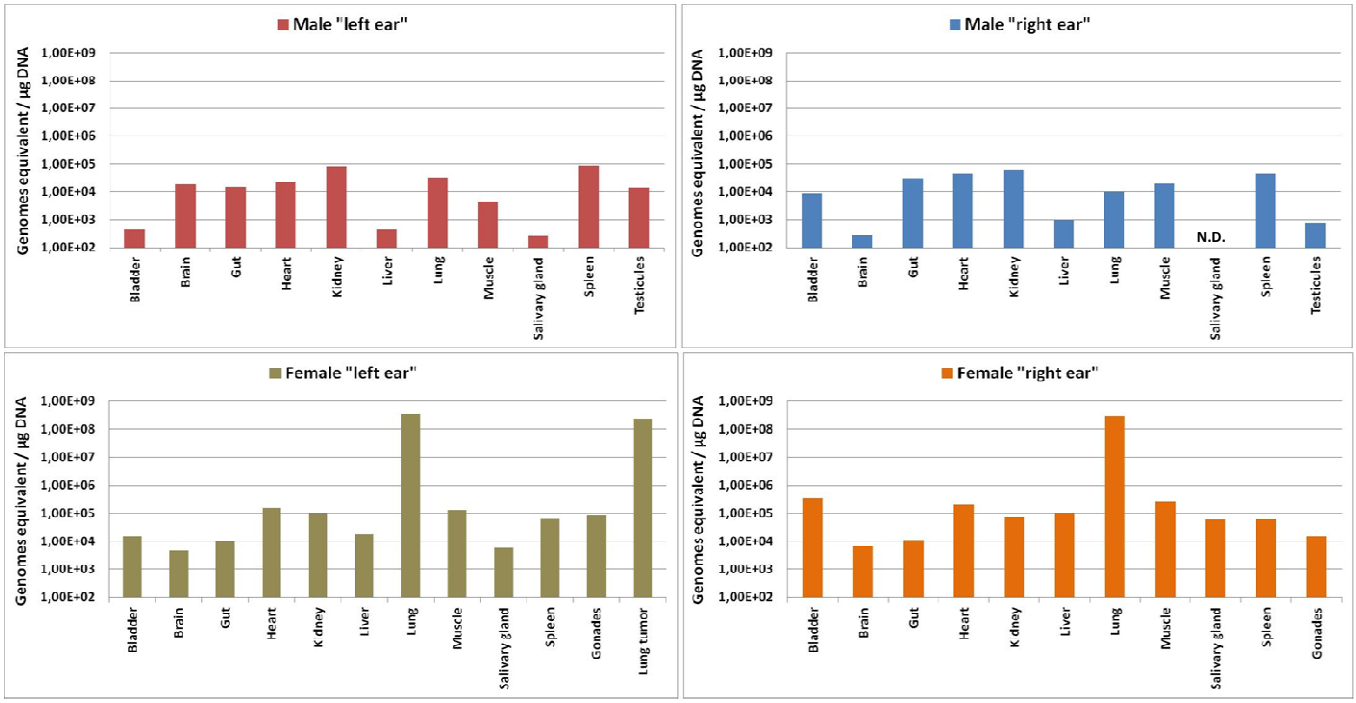

Figure 10 : MuPyV titers in organs of FVB/NJ mice i.p. infected with 10^6 IU of the barcoded MuPyVs library.
The copy number of virus genomes equivalent per µg of total DNA was obtained by real time qPCR analysis at days 59 p.i. (male "right ear") and 99 p.i., as described in Materials and Methods. The limit of detection (100 copies/µg of total DNA) was set up at the interception of the x-axis. N.D.: Not determined.

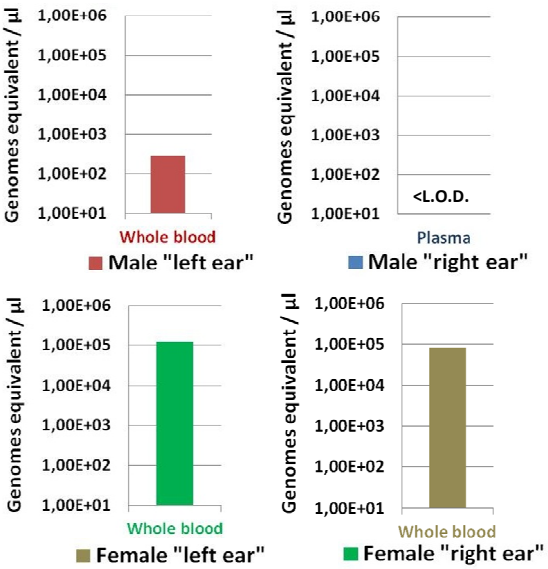

Figure 11 : MuPyV titers in whole blood and plasma of FVB/NJ mice i.p. infected with 10^6 IU of the barcoded MuPyVs library. The copy number of virus genomes equivalent per µl of whole blood or plasma was obtained by real time qPCR analysis at days 59 p.i. (male "right ear") and 99 p.i., as described in Materials and Methods. The limit of detection (10 copies/µl) was set up at the interception of the x-axis. <L.O.D.: below the limit of detection.

4.5 Investigation of barcoded MuPyV genomes during persistent infections

My preliminary analysis of the NGS data was to compare barcodes with high titers in urines collected during persistent infections at different shedding times, and in organs from the same animal (Tables 3 to 6). It is important to state that a direct correlation between barcodes titers in urines and organs is not possible because samples have been harvested at different time points. However, virus titers in urine samples harvested the day before organs were collected can suggest a trend of the overall virus replication at the end of the study (i.e. start, end or no virus reactivation) (Figure 9). For simplification, I named barcodes with four letter starting with BC

(stands for barcode) and ML (for male “left ear”) or MR (for male “right ear”) or FL (for female “left ear”) or FR (for female “right ear”) and a number. In this study, I selected all high titers barcodes in the above urine samples and investigated them at different shedding events during the persistent phase of infection and organs from the same animal.

In the male named “left ear”, I considered shedding events at 55, 66, 81 and 94 days p.i. that correspond to the periods of reactivation of viral replication during persistent infections in this animal (Figure 9). BCML1, 2, 3 and 5 were shed with high titers at 55 days p.i. and found with very low titers in the next shedding events. At day 99 p.i., BCML1 and 5 were present in kidney with higher titers, while BCML3 was present in both kidney and spleen with higher titers. Interestingly, BCML4 was shed at high titers during all shedding events (i.e., at 55, 66, 81 and 94 days p.i.). At day 99 p.i., BCML4 was present in the kidney with a higher titer, suggesting that this barcode may have shed in urine from kidney at least during the last shedding event (94 days p.i.). BCML6, 8, 9 were shed with high titers at 66 days p.i. BCML8 was present in both gut and spleen at day 99 p.i. with a higher titer, while BCML9 was mainly found in kidney. BCML7 was shed later, at 81 days p.i., with a high titer and was mainly found in heart (cardiac muscle) at day 99 p.i. BCML10 shed at the latest time point tested (i.e., at 94 days p.i.) with a high titer. Interestingly, BCML11 shed with high titers during two shedding events, at 81 and 94 days p.i. (Table 3). BCML10 and BCML11 were not found at high titers in organs collected at day 99 p.i., suggesting that these barcodes may have shed from somewhere else or barcoded viruses were probably not replicating at that time, i.e., five days after the shedding event.

		Urine days p.i.				Organs											
Barcodes		55	66	81	94	Bladder	Brain	Gut	Heart	Kidney	Liver	Lung	Muscle	Salivary gland	Spleen	Testicle	Whole blood
BCML1	TGGCACTTGAAT	1,53E+04	8,25E+00	6,57E+00	5,18E-03	3,36E-01	5,02E-01	6,47E-01	1,79E+00	4,06E+01	6,08E-04	5,00E-02	2,38E-03	9,57E-02	1,42E-01	4,99E-02	
BCML2	CAAACCCACCGT	2,92E+03	4,17E+01	2,24E+00	6,35E+01	1,06E-01	7,46E-01			4,29E-01	3,04E-04		8,36E-01	1,98E-04	6,64E-01	8,32E-03	
BCML3	CTAT	2,28E+03	3,00E+01	1,59E-01	1,50E-01	8,08E-01	1,43E-02	4,43E+00	2,65E+00	1,12E+01	1,02E+00	1,70E+00	4,53E+00	5,57E-02	1,41E+01	3,33E-01	
BCML4	CCGCTTCAATGA	1,63E+03	1,65E+02	5,73E+02	4,22E+02	6,48E-01	8,61E-02	1,85E-01	5,52E+00	1,15E+01	6,95E+00	3,00E-01	2,11E+00	5,55E-03	7,12E-01		
BCML5	CCTGCTGCGAAT	1,38E+03	8,14E+00	5,86E+00	1,74E+01	4,07E+00		2,05E-02	3,22E+00	1,67E+01	6,08E-04	1,00E+00		3,96E-04	3,32E-01	8,32E-03	2,83E-04
BCML6	CGTAGATAATGT	2,09E+01	1,22E+03		4,92E+00	2,42E-01		1,96E+00		2,86E-01	3,04E-04		4,76E-03		2,85E-01	8,32E-03	2,26E-03
BCML7	GCCCACAAATAC	4,88E+00	1,17E+00	2,42E+03	4,01E+01	4,27E-01		1,94E+00	2,50E+01	4,77E-02		2,95E+00	2,38E-03	6,68E-02	3,27E+00		
BCML8	AGTATTGGTTGA	1,19E+00	5,51E+03			2,33E-03		3,04E+02	9,55E-02	3,58E+00	6,08E-04	5,50E-01	2,38E-03	3,57E-03	8,01E+02		
BCML9	ACAGATAGTGAG	1,19E-01	1,17E+03	9,78E+01	2,07E-02	4,21E-01		1,13E+00	2,79E+00	1,02E+01	3,04E-03	5,40E+00	8,00E+00	7,93E-04	9,02E-01		
BCML10	TACAAGACT	6,50E-02	1,15E-02	1,38E-01	1,19E+03	1,40E-02	1,43E-02	6,16E-02	2,63E-01	9,54E-02	1,19E-02	5,00E-02	2,88E-01	9,87E-02	1,90E-01	1,66E-02	1,22E-02
BCML11	CACCACCCATGC	1,08E-02	9,32E-01	1,04E+03	2,07E+03	1,12E-01	1,43E-02	4,25E+00		1,10E+00	3,04E-04			3,96E-04	6,17E-01		

Table 12 : Barcodes with high titers in urines collected at 55, 66, 81 and 94 days p.i. and in organs harvested at 99 days p.i. from the mouse named male “left ear”. On the left part are high titers barcodes from urines and on the right part are titers of these barcodes in organs from the same animal. Titers in urines are in “barcodes/µl of urine”, titers in organs are in “barcodes/µg of total DNA”, titers in whole blood are in “barcodes/µl of whole blood”.

In the male named “right ear”, I considered shedding events at 47 and 55 days p.i. that correspond to the main periods of reactivation of viral replication in this animal that was sacrificed at day 59 p.i. (Figure 9). BCMR1, 3 and 4 were shed with high titers at 47 days p.i. and were found with low titers at 55 days p.i. At day 59 p.i., BCMR1 was mainly present in spleen, BCMR3 was found at higher titers in both spleen and kidney, while BCMR4 had very low titers in all tested organs. BCMR2 shed at high titers at both 47 and 55 days p.i. and was found at higher titers in spleen, kidney and heart at day 59 p.i. BCMR5 shed at 55 and 47 days p.i., with a high titer at 55 days p.i. BCMR5 was mainly present in bladder at day 59 p.i. (Table 4).

Barcodes		Urine days p.i.		Organs									
		47	55	Bladder	Brain	Gut	Heart	Kidney	Liver	Lung	Muscle	Spleen	Testicle
BCMR1	TAGGCTA	5,50E+03	2,78E+01	1,32E+00	1,43E-04		4,67E+00	1,73E-01	6,22E-03	1,41E-02	3,18E-01	2,31E+01	1,26E-03
BCMR2	GTTTATGAGG	3,54E+03	4,49E+03	9,64E-01	1,57E-03	1,01E-01	4,93E+01	6,95E+02	1,08E+00	3,32E+00	1,06E+00	5,41E+01	
BCMR3	TTCATTCTACAC	2,34E+03	4,49E+01	1,12E+00				4,84E+01			2,89E-02	1,16E+01	
BCMR4	CAACTAAAGGTG	1,49E+03	3,93E+00	1,36E-02			5,37E-02			2,35E-02		3,76E-01	2,52E-03
BCMR5	CCCCTTGTGAAG	2,59E+02	3,85E+03	1,87E+03	5,71E-04	7,04E-01	5,37E-02	4,29E+00		4,22E-02		6,94E+00	3,15E-03

Table 4 : Barcodes with high titers in urines collected at 47 and 55 days p.i. and in organs harvested at 59 days p.i. from the mouse named male "right ear". On the left part are high titers barcodes from urines and on the right part are titers of these barcodes in organs from the same animal. Titers in urines are in "barcodes/µl of urine", titers in organs are in "barcodes/µg of total DNA".

In the female named "left ear", I considered shedding events at 66, 81 and 90 days p.i. that correspond to the periods of reactivation of viral replication during persistent infections in this animal (Figure 9). Overall, barcodes titers in females were lower than in males. BCFL1, 2 and 3 shed at high titers at 66 days p.i. and were found at very low titers during the next shedding events. At day 99 p.i., these three barcodes were highly present in lung tumor, however, these barcodes were not the most abundant in lung tumor. BCFL1 and 3 were present in muscle with higher titers and BCFL3 was highly present in lung. BCFL4, 5 and 7 shed at high titers at 90 days p.i. At day 99 p.i., BCFL4 was highly present in lung tumor, but it was not the most abundant in lung tumor, while BCFL5 and 7 were not found at high titers in organs, suggesting that these barcodes may have shed from somewhere else or they were probably not replicating at that time, i.e. nine days after the shedding event. BCFL6 shed at high titers at 81 days p.i. and 90 days p.i. and was also mostly present in lung at day 99 p.i. BCFL8, 9: shed later at 90 days p.i. and were mostly present in lung too (Table 5). Interestingly, lung tumor was infected mostly with one barcode (CCACCCTATAAA) at a very high titer 2.34 x 10^8 barcodes/µg that is closed to the overall MuPyV DNA titer (2.42 x 10^8 genomes/µg, as shown in Figure 10). This suggested that the development of lung tumor may be associated to the high virus replication in lung, however, further investigations are needed to confirm this hypothesis.

Barcode		Urine days p.i.			Organs												
		66	81	90	Bladder	Brain	Gut	Heart	Kidney	Liver	Lung	Muscle	Salivary gland	Spleen	Gonades	Whole blood	Lung tumor
BCFL1	CGACTTG	5,09E+02	2,46E-03	1,28E-02	8,36E-01	4,04E-03	1,65E-02	1,91E+00	5,09E+00	1,86E+00		1,34E+01	1,24E-01	9,11E+00		3,15E-02	1,32E+02
BCFL2	GGACTCAAAAAC	3,19E+02		2,41E-02				1,45E+00		1,08E+00			6,22E-02			3,15E-02	1,32E+02
BCFL3	CATGCAGGTAGT	1,71E+02		8,51E-03	1,53E-01	7,08E-01	2,75E-02		1,65E+00	7,77E-01	1,76E+03	1,34E+01	3,46E-02	9,49E-02		9,46E-02	1,32E+02
BCFL4	AGCTGTCTGGTA	8,16E+01	2,07E+00	1,31E+02	1,68E+00	1,62E-02	2,20E-02	1,32E-01	3,30E-01	3,37E+00		2,53E-01		3,16E-02			1,05E+03
BCFL5	ATCATCATTTTG	6,78E+00	8,54E-01	1,46E+02	1,77E-01	8,09E-03			1,32E-01	2,88E-02		1,27E-01		6,33E-02			
BCFL6	CCATATGTCCGA	1,17E-02	2,20E+03	2,23E+02	1,27E+00	8,09E-03	4,46E-01	5,92E-01	7,27E+00		1,26E+03			2,34E+00		9,46E-02	
BCFL7	GCCCTGCAGCCG	8,03E-03		2,21E+03	2,54E+00		5,51E-03	9,01E+00	1,92E+00	1,44E-02				3,16E-01	1,23E-01	6,30E-02	
BCFL8	ACAAGTACGATT		1,42E+03		1,18E-02	4,04E-03	5,51E-03	3,29E-01	2,97E+00	1,44E-02	7,53E+02						
BCFL9	TAAAAAATTACC		3,04E+03	6,38E+00	4,71E-01		5,51E-03	5,92E-01	9,91E-01	4,32E-02	1,76E+03	1,27E-01		1,14E+00	6,13E-02	6,30E-02	

Table 5 : Barcodes with high titers in urines collected at 66, 81 and 90 days p.i. and in organs harvested at 99 days p.i. from the mouse named female "left ear". On the left part are high titers barcodes from urines and on the right part are titers of these barcodes in organs from the same animal. Titers in urines are in "barcodes/µl of urine", titers in organs are in "barcodes/µg of total DNA", titers in whole blood are in "barcodes/µl of whole blood".

In the female named "right ear", I considered shedding events at 47, 62 and 94 days p.i. (Figure 9). As previously described in the female named "left ear", barcodes titers were lower than in males (Figure 9). BCFR1, 2, 3, 4 and 6 shed at 47 days p.i. with high titers. At 99 days p.i., BCFR1 was highly present in both lung tumor and

salivary gland, BCFR2 was found with a low titer in salivary gland, while BCFR3, 4 and 6 were not found at high titer in the selected organs at day 99 p.i. BCFR5 shed both at 47 and 62 days p.i. and was mostly found in lung and salivary gland and at lower titers in bladder, heart (cardiac muscle) and kidney at day 99 p.i. BCFR7, 11, 12 and 15 shed at 62 days p.i. At day 99 p.i., BCFR7 was mainly found in lung and salivary gland as well as in bladder at a lower titer, BCFR11 was mainly found in lung and in salivary gland and muscle at lower titers, while BCFR12 was mainly present in salivary gland and BCFR15 was highly present in lung and in salivary gland at a lower titer. BCFR8, 9, 10 and 14 shed during the last shedding event investigated (94 days p.i.). At day 99 p.i., BCFR8 was mostly present in lung, BCFR9 was found in salivary gland while BCFR14 was present at a high titer in kidney. BCFR10 was not found at high titer in the selected organs at 99 days p.i., suggesting that this barcoded virus may have shed from somewhere else or was probably not replicating five days after the shedding event (Table 6).

		Urine day p.i.			Organs											
Barcode		47	62	94	Bladder	Brain	Gut	Heart	Kidney	Liver	Lung	Muscle	Salivary gland	Spleen	Gonades	Whole blood
BCFR1	CGTAAGT	7,44E+02	3,66E-02	1,21E+00		1,06E-02	7,23E-01	5,48E+00	6,76E+00	4,35E-01	1,82E+03	3,15E+00	6,13E+01	6,08E+00	2,18E+00	2,91E-01
BCFR2	CGCAAGAGTTTA	5,27E+02	1,89E+00	2,31E+01	8,50E-01	2,64E-03	4,76E-01	1,75E+00	3,18E+00	9,66E-02		2,74E-01	1,71E+01	2,75E+00	4,99E-01	
BCFR3	TTAACCGTAAGT	2,94E+02		4,28E-01			1,18E-02	4,39E-01						2,87E-02	4,99E-02	
BCFR4	AGGACTTTA	2,52E+02	2,93E-02	1,99E-01			1,76E-02	2,19E+00	3,98E-02				4,23E+00			
BCFR5	CCAATCCTGAGT	1,53E+02	3,29E+02	2,86E+00	1,11E+01	2,64E-03	2,94E-02	2,57E+01	1,79E+01	5,36E+00	6,62E+02	8,22E-01	1,99E+02	6,69E+00	2,31E+00	1,46E-01
BCFR6	AGCACGGTATAG	1,47E+02		1,76E-02		5,57E-01	2,72E+00		4,77E-01				5,13E+00	2,87E-02		
BCFR7	AAACTCCAGTCT	5,54E+00	3,02E+02	1,76E-02	1,11E+01	5,28E-03	8,11E-01	4,60E+00	1,31E+00	9,66E-02	3,31E+02	1,37E-01	3,06E+01	4,30E-01	3,32E-02	2,91E-01
BCFR8	TCTCCCTGGATG	2,66E-02	1,28E+00	7,48E+02		4,22E-02		8,77E-01			1,66E+02	2,05E+00	3,10E+00	9,18E-01	1,66E-02	2,19E-01
BCFR9	ACGCTCGGGACA	3,80E-03	2,44E-03	6,67E+02		2,90E-02	5,88E-03	2,85E+00	2,47E+00			3,29E+00	2,19E+01	9,18E-01	4,99E-02	
BCFR10	AGGCCCGAGATG	3,80E-03	1,46E-02	2,29E+03	4,25E-01	4,06E-01		4,39E-01	3,98E-02			1,37E+00	4,68E+00	1,43E-01	9,97E-02	7,29E-02
BCFR11	CAGCTATTACAT	3,80E-03	3,17E+03	3,98E+00	5,95E+00	7,92E-02	8,58E-01	5,48E+00	9,55E-01	5,31E-01	8,28E+02	1,03E+01	1,56E+01	5,74E-02		
BCFR12	CCAAAACTCCAA	3,80E-03	1,08E+03	1,24E+01		6,07E-02	5,88E-03	1,75E+00	3,98E-02	3,38E-01			1,08E+01	8,61E-02	1,31E+00	7,29E-02
BCFR13	GAGACC	3,80E-03	3,66E-02	1,28E+02	1,28E+00	5,28E-03	1,18E-02		2,78E-01	9,66E-02	1,66E+02		5,32E+00	9,24E+02		8,74E-01
BCFR14	CTGAAAAAGGAG		4,39E-02	6,54E+03	8,50E-01	2,64E-03		2,19E-01	2,96E+03	3,38E-01		2,74E-01	2,00E+00	1,15E-01	1,66E-01	
BCFR15	GAATGATAGAC		1,05E+02	3,87E+00		2,64E-03		4,39E-01	3,98E-02	4,83E-02	1,66E+02		1,46E+01		1,66E-02	1,46E-01

Table 6 : Barcodes with high titers in urines collected at 47, 62 and 94 days p.i. and in organs harvested at 99 days p.i. from the mouse named female "right ear". On the left part are high titers barcodes from urines and on the right part are titers of these barcodes in organs from the same animal. Titers in urines are in "barcodes/µl of urine", titers in organs are in "barcodes/µg of total DNA", titers in whole blood are in "barcodes/µl of whole blood".

Overall, this preliminary analysis suggested that barcoded MuPyVs with high titers found in urines from males during the last shedding events (i.e., close to the end of the study) may have been shed mostly from kidney, and in some cases from spleen, while in urines from females they may have been shed mainly from lung, and in some cases from salivary gland and kidney. Organs with high titers of barcodes are not necessarily considered sites of latency, however, this cannot be ruled out, especially in the case of virus reactivation observed in urine close to the time when organs were collected and/or in the case of a huge number of latently infected cells in these tissues (perhaps less likely). In organs with low titers of barcoded viruses, the study of gene expression patterns associated to episodically shed viruses (with high titers in urine) will help to define the site of latency.

5. Conclusion and future directions

The originality of this study is first to show that genetically barcoded MuPyVs can replicate *in vivo* and shed in urine similarly to the wild-type parental strain, as compared with a previous study[281]. This is, since to our knowledge, the first time a DNA virus has been successfully barcoded. The second novelty of this longitudinal study of the PyVs persistent infection is the use of a genetically barcoded viruses approach combined with a non-invasive urine-monitoring system. The barcoded viruses approach associated with the Illumina high-throughput sequencing technology permitted to evaluate the virus diversity in urine during episodic reactivations, as well as in a huge variety of organs. The preliminary NGS data analysis showed the presence of a limited number of barcoded virus genomes with high titers episodically shed in urine during the persistent phase of infection, thus suggesting virus reactivation events from latent infections. To confirm latency, in-depth analysis of all urine samples should be performed to ensure that these few barcoded viruses are not continuously shed at lower titers. The study of genes expression for each latent barcoded genome in organs will also support the hypothesis of latent infections. The presence of numerous barcodes at low titers has been observed in some samples, in-depth analysis is currently in progress to confirm the eventual presence of both latent and “smoldering” infections.

Further investigations are planned to evaluate the potential role of PyV miRNA in the choice of the persistent mode of infection and in the regulation of genes expression during persistency. Moreover, as persistent infections are often cell type-specific, further experiments are needed to define it more precisely. Finally, previous studies showed the importance of the regulatory region in persistent infections, such investigations would also be highly informative to understand the regulation of virus genes expression during persistent infections.

Acknowledgements

I would like to thank my supervisor Prof.ssa Patrizia Zavattari for her support and the University of Cagliari for my 3-year PhD fellowship and the opportunity to spend 1 year in two high ranked US Universities.

I would like to thank Prof. Michael Imperiale who invited me in his lab at the University of Michigan, Ann Arbor (MI) and teached me all about BKPyV and many usefull techniques of virology.

I would like to strongly thank Prof. Christopher Sullivan who heartily invited me in his laboratory at the University of Texas at Austin. I would like to thank him for offering me the best opportunity of my young carrieer and for allowing me to reach high impact results to conclude proudly my PhD thesis. I thank him warmly for his exciting scientific talks, his insightful comments, his incredible motivation and enthusiasm during my stay in his lab.

I would like to thank all my colleagues from the University of Cagliari, from Prof Imperiale lab and a special thank to my colleagues from Prof Sullivan lab for their assistance and the good time spent together.

Finally, I would like to thanks everbody who supported my decision to undertake this PhD.

References

1. Moens, U. *et al.* ICTV Virus Taxonomy Profile: Polyomaviridae. *J. Gen. Virol.* **98**, 1159–1160 (2017).
2. Knowles, W. A. Discovery and Epidemiology of the Human Polyomaviruses BK Virus (BKV) and JC Virus (JCV). in *Polyomaviruses and Human Diseases* **577**, 19–45 (Springer New York, 2006).
3. Kean, J. M., Rao, S., Wang, M. & Garcea, R. L. Seroepidemiology of Human Polyomaviruses. *PLoS Pathog.* **5**, e1000363 (2009).
4. Chen, T. *et al.* Seroepidemiology of the Newly Found Trichodysplasia Spinulosa–Associated Polyomavirus. *J. Infect. Dis.* **204**, 1523–1526 (2011).
5. Gossai, A. *et al.* Seroepidemiology of Human Polyomaviruses in a US Population. *Am. J. Epidemiol.* **183**, 61–69 (2016).
6. Nicol, J. T. J. *et al.* Age-specific seroprevalences of merkel cell polyomavirus, human polyomaviruses 6, 7, and 9, and trichodysplasia spinulosa-associated polyomavirus. *Clin. Vaccine Immunol.* **20**, 363–8 (2013).
7. Nicol, J. T. J. *et al.* Seroprevalence of human Malawi polyomavirus. *J. Clin. Microbiol.* **52**, 321–3 (2014).
8. Viscidi, R. P. *et al.* Age-specific seroprevalence of Merkel cell polyomavirus, BK virus, and JC virus. *Clin. Vaccine Immunol.* **18**, 1737–43 (2011).
9. Goudsmit, J., Wertheim-van Dillen, P., van Strien, A. & van der Noordaa, J. The role of BK virus in acute respiratory tract disease and the presence of BKV DNA in tonsils. *J. Med. Virol.* **10**, 91–9 (1982).
10. Monaco, M. C., Jensen, P. N., Hou, J., Durham, L. C. & Major, E. O. Detection of JC virus DNA in human tonsil tissue: evidence for site of initial viral infection. *J. Virol.* **72**, 9918–23 (1998).
11. Bofill-Mas, S., Formiga-Cruz, M., Clemente-Casares, P., Calafell, F. & Girones, R. Potential Transmission of Human Polyomaviruses through the Gastrointestinal Tract after Exposure to Virions or Viral DNA. *J. Virol.* **75**, 10290–10299 (2001).
12. Martel-Jantin, C. *et al.* Merkel cell polyomavirus infection occurs during early childhood and is transmitted between siblings. *J. Clin. Virol.* **58**, 288–291 (2013).
13. Moens, U., Ludvigsen, M. & Van Ghelue, M. Human polyomaviruses in skin diseases. *Patholog. Res. Int.* **2011**, 123491 (2011).
14. Schwarz, A. *et al.* Viral Origin, Clinical Course, and Renal Outcomes in Patients With BK Virus Infection After Living-Donor Renal Transplantation. *Transplantation* **100**, 844–853 (2016).
15. Zanotta, N. *et al.* Merkel Cell Polyomavirus Is Associated with Anal Infections in Men Who Have Sex with Men. *Microorganisms* **7**, 54 (2019).
16. Gross, L. A filterable agent, recovered from Ak leukemic extracts, causing salivary gland carcinomas in C3H mice. *Proc. Soc. Exp. Biol. Med.* **83**, 414–21 (1953).
17. Stewart, S. E., Eddy, B. E. & Borgese, N. Neoplasms in mice inoculated with a tumor agent carried in tissue culture. *J. Natl. Cancer Inst.* **20**, 1223–43 (1958).
18. Sweet, B. H. & Hilleman, M. R. The vacuolating virus, S.V. 40. *Proc. Soc. Exp. Biol. Med.* **105**, 420–7 (1960).
19. Gardner, S. D., Field, A. M., Coleman, D. V & Hulme, B. New human papovavirus (B.K.) isolated from urine after renal transplantation. *Lancet (London, England)* **1**, 1253–7 (1971).
20. Padgett, B. L., Walker, D. L., ZuRhein, G. M., Eckroade, R. J. & Dessel, B. H. Cultivation of papova-like virus from human brain with progressive multifocal leucoencephalopathy. *Lancet (London, England)* **1**, 1257–60 (1971).
21. Doerries, K. Human Polyomavirus JC and BK Persistent Infection. in *Polyomaviruses and Human Diseases* **577**, 102–116 (Springer New York, 2006).
22. Egli, A. *et al.* Prevalence of polyomavirus BK and JC infection and replication in 400 healthy blood donors. *J. Infect. Dis.* **199**, 837–46 (2009).
23. Loeber, G. & Dörries, K. DNA rearrangements in organ-specific variants of polyomavirus JC strain GS. *J. Virol.* **62**, 1730–5 (1988).
24. Gosert, R. *et al.* Polyomavirus BK with rearranged noncoding control region emerge in vivo in renal transplant patients and increase viral replication and cytopathology. *J. Exp. Med.* **205**, 841–52 (2008).
25. Abend, J. R., Jiang, M. & Imperiale, M. J. BK virus and human cancer: innocent until proven guilty. *Semin. Cancer Biol.* **19**, 252–60 (2009).
26. Levican, J., Acevedo, M., León, O., Gaggero, A. & Aguayo, F. Role of BK human polyomavirus in cancer. *Infect. Agent. Cancer* **13**, 12 (2018).
27. Allander, T. *et al.* Identification of a Third Human Polyomavirus. *J. Virol.* **81**, 4130–4136 (2007).
28. Gaynor, A. M. *et al.* Identification of a Novel Polyomavirus from Patients with Acute Respiratory Tract Infections. *PLoS Pathog.* **3**, e64 (2007).
29. Siebrasse, E. A. *et al.* Identification of MW polyomavirus, a novel polyomavirus in human stool. *J. Virol.* **86**, 10321–6 (2012).
30. Lim, E. S. *et al.* Discovery of STL polyomavirus, a polyomavirus of ancestral recombinant origin that encodes a unique T antigen by alternative splicing. *Virology* **436**, 295–303 (2013).
31. Mishra, N. *et al.* Identification of a Novel Polyomavirus in a Pancreatic Transplant Recipient With Retinal Blindness and Vasculitic Myopathy. *J. Infect. Dis.* **210**, 1595–1599 (2014).

32. Feng, H., Shuda, M., Chang, Y. & Moore, P. S. Clonal Integration of a Polyomavirus in Human Merkel Cell Carcinoma. *Science (80-.).* **319**, 1096–1100 (2008).
33. van der Meijden, E. *et al.* Discovery of a New Human Polyomavirus Associated with Trichodysplasia Spinulosa in an Immunocompromized Patient. *PLoS Pathog.* **6**, e1001024 (2010).
34. Schowalter, R. M., Pastrana, D. V., Pumphrey, K. A., Moyer, A. L. & Buck, C. B. Merkel Cell Polyomavirus and Two Previously Unknown Polyomaviruses Are Chronically Shed from Human Skin. *Cell Host Microbe* **7**, 509–515 (2010).
35. Scuda, N. *et al.* A novel human polyomavirus closely related to the african green monkey-derived lymphotropic polyomavirus. *J. Virol.* **85**, 4586–90 (2011).
36. Buck, C. B. *et al.* Complete genome sequence of a tenth human polyomavirus. *J. Virol.* **86**, 10887 (2012).
37. Korup, S. *et al.* Identification of a Novel Human Polyomavirus in Organs of the Gastrointestinal Tract. *PLoS One* **8**, e58021 (2013).
38. Gheit, T. *et al.* Isolation and characterization of a novel putative human polyomavirus. *Virology* **506**, 45–54 (2017).
39. Calvignac-Spencer, S. *et al.* A taxonomy update for the family Polyomaviridae. *Arch. Virol.* **161**, 1739–1750 (2016).
40. Vinograd, J., Lebowitz, J., Radloff, R., Watson, R. & Laipis, P. The twisted circular form of polyoma viral DNA. *Proc. Natl. Acad. Sci. U. S. A.* **53**, 1104–11 (1965).
41. Cremisi, C., Pignatti, P. F., Croissant, O. & Yaniv, M. Chromatin-like structures in polyoma virus and simian virus 10 lytic cycle. *J. Virol.* **17**, 204–11 (1975).
42. Ferenczy, M. W. *et al.* Molecular Biology, Epidemiology, and Pathogenesis of Progressive Multifocal Leukoencephalopathy, the JC Virus-Induced Demyelinating Disease of the Human Brain. *Clin. Microbiol. Rev.* **25**, 471–506 (2012).
43. Van Ghelue, M., Khan, M. T. H., Ehlers, B. & Moens, U. Genome analysis of the new human polyomaviruses. *Rev. Med. Virol.* **22**, 354–377 (2012).
44. Dailey, L. & Basilico, C. Sequences in the polyomavirus DNA regulatory region involved in viral DNA replication and early gene expression. *J. Virol.* **54**, 739–49 (1985).
45. Ehlers, B. & Moens, U. Genome analysis of non-human primate polyomaviruses. *Infect. Genet. Evol.* **26**, 283–294 (2014).
46. Bialasiewicz, S. *et al.* Whole-genome characterization and genotyping of global WU polyomavirus strains. *J. Virol.* **84**, 6229–34 (2010).
47. Lednicky, J. A., Butel, J. S., Luetke, M. C. & Loeb, J. C. Complete genomic sequence of a new Human polyomavirus 9 strain with an altered noncoding control region. *Virus Genes* **49**, 490–2 (2014).
48. Moens, U. & Van Ghelue, M. Polymorphism in the genome of non-passaged human polyomavirus BK: implications for cell tropism and the pathological role of the virus. *Virology* **331**, 209–31 (2005).
49. Song, X. *et al.* Characterization of the non-coding control region of polyomavirus KI isolated from nasopharyngeal samples from patients with respiratory symptoms or infection and from blood from healthy blood donors in Norway. *J. Gen. Virol.* **97**, 1647–1657 (2016).
50. Ault, G. S. & Stoner, G. L. Human polyomavirus JC promoter/enhancer rearrangement patterns from progressive multifocal leukoencephalopathy brain are unique derivatives of a single archetypal structure. *J. Gen. Virol.* **74 (Pt 8)**, 1499–507 (1993).
51. Delbue, S. *et al.* JC virus load in cerebrospinal fluid and transcriptional control region rearrangements may predict the clinical course of progressive multifocal leukoencephalopathy. *J. Cell. Physiol.* **227**, 3511–7 (2012).
52. Gosert, R., Kardas, P., Major, E. O. & Hirsch, H. H. Rearranged JC virus noncoding control regions found in progressive multifocal leukoencephalopathy patient samples increase virus early gene expression and replication rate. *J. Virol.* **84**, 10448–56 (2010).
53. Zhang, S. *et al.* Viral microRNA effects on pathogenesis of polyomavirus SV40 infections in syrian golden hamsters. *PLoS Pathog.* **10**, e1003912 (2014).
54. Riley, M. I., Yoo, W., Mda, N. Y. & Folk, W. R. Tiny T antigen: an autonomous polyomavirus T antigen amino-terminal domain. *J. Virol.* **71**, 6068–74 (1997).
55. Treisman, R., Cowie, A., Favaloro, J., Jat, P. & Kamen, R. The structures of the spliced mRNAs encoding polyoma virus early region proteins. *J. Mol. Appl. Genet.* **1**, 83–92 (1981).
56. Sheng, Q. *et al.* The DnaJ domain of polyomavirus large T antigen is required to regulate Rb family tumor suppressor function. *J. Virol.* **71**, 9410–6 (1997).
57. Sullivan, C. S. & Pipas, J. M. T antigens of simian virus 40: molecular chaperones for viral replication and tumorigenesis. *Microbiol. Mol. Biol. Rev.* **66**, 179–202 (2002).
58. Ito, Y., Brocklehurst, J. R. & Dulbecco, R. Virus-specific proteins in the plasma membrane of cells lytically infected or transformed by pol-oma virus. *Proc. Natl. Acad. Sci. U. S. A.* **74**, 4666–70 (1977).
59. Simmons, D. T., Chang, C. & Martin, M. A. Multiple forms of polyoma virus tumor antigens from infected and transformed cells. *J. Virol.* **29**, 881–7 (1979).
60. Courtneidge, S. A. *et al.* Identification and characterization of the hamster polyomavirus middle T antigen. *J. Virol.* **65**, 3301–8 (1991).
61. van der Meijden, E. *et al.* Characterization of T Antigens, Including Middle T and Alternative T, Expressed by the Human Polyomavirus Associated with Trichodysplasia Spinulosa. *J. Virol.* **89**, 9427–9439 (2015).

62. Fluck, M. M. & Schaffhausen, B. S. Lessons in Signaling and Tumorigenesis from Polyomavirus Middle T Antigen. *Microbiol. Mol. Biol. Rev.* **73**, 542–563 (2009).
63. Mes, A. M. & Hassell, J. A. Polyoma viral middle T-antigen is required for transformation. *J. Virol.* **42**, 621–9 (1982).
64. Carter, J. J. *et al.* Identification of an overprinting gene in Merkel cell polyomavirus provides evolutionary insight into the birth of viral genes. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 12744–9 (2013).
65. Berger, H. & Wintersberger, E. Polyomavirus small T antigen enhances replication of viral genomes in 3T6 mouse fibroblasts. *J. Virol.* **60**, 768–70 (1986).
66. Bikel, I. & Loeken, M. R. Involvement of simian virus 40 (SV40) small t antigen in trans activation of SV40 early and late promoters. *J. Virol.* **66**, 1489–94 (1992).
67. Chen, L. & Fluck, M. M. Role of middle T-small T in the lytic cycle of polyomavirus: control of the early-to-late transcriptional switch and viral DNA replication. *J. Virol.* **75**, 8380–9 (2001).
68. Cicala, C. *et al.* Simian virus 40 small-t antigen stimulates viral DNA replication in permissive monkey cells. *J. Virol.* **68**, 3138–44 (1994).
69. Daniels, R., Sadowicz, D. & Hebert, D. N. A Very Late Viral Protein Triggers the Lytic Release of SV40. *PLoS Pathog.* **3**, e98 (2007).
70. Johne, R. & Müller, H. Avian polyomavirus agnoprotein 1a is incorporated into the virus particle as a fourth structural protein, VP4. *J. Gen. Virol.* **82**, 909–18 (2001).
71. Henriksen, S., Hansen, T., Bruun, J.-A. & Rinaldo, C. H. The Presumed Polyomavirus Viroporin VP4 of Simian Virus 40 or Human BK Polyomavirus Is Not Required for Viral Progeny Release. *J. Virol.* **90**, 10398–10413 (2016).
72. Jay, G., Nomura, S., Anderson, C. W. & Khoury, G. Identification of the SV40 agnogene product: a DNA binding protein. *Nature* **291**, 346–9 (1981).
73. Okada, Shuichi Endo, Hidehiro Takah, Y. *et al.* Distribution and function of JCV agnoprotein. *J. Neurovirol.* **7**, 302–306 (2001).
74. Rinaldo, C. H., Traavik, T. & Hey, A. The agnogene of the human polyomavirus BK is expressed. *J. Virol.* **72**, 6233–6 (1998).
75. Gerits, N. & Moens, U. Agnoprotein of mammalian polyomaviruses. *Virology* **432**, 316–26 (2012).
76. Seo, G. J., Fink, L. H. L., O'Hara, B., Atwood, W. J. & Sullivan, C. S. Evolutionarily Conserved Function of a Viral MicroRNA. *J. Virol.* **82**, 9823–9828 (2008).
77. Seo, G. J., Chen, C. J. & Sullivan, C. S. Merkel cell polyomavirus encodes a microRNA with the ability to autoregulate viral gene expression. *Virology* **383**, 183–7 (2009).
78. Sullivan, C. S., Grundhoff, A. T., Tevethia, S., Pipas, J. M. & Ganem, D. SV40-encoded microRNAs regulate viral gene expression and reduce susceptibility to cytotoxic T cells. *Nature* **435**, 682–6 (2005).
79. Chen, C. J. *et al.* Identification of a polyomavirus microRNA highly expressed in tumors. *Virology* **476**, 43–53 (2015).
80. Sullivan, C. S. *et al.* Murine Polyomavirus encodes a microRNA that cleaves early RNA transcripts but is not essential for experimental infection. *Virology* **387**, 157–67 (2009).
81. Johne, R. *et al.* Taxonomical developments in the family Polyomaviridae. *Arch. Virol.* **156**, 1627–1634 (2011).
82. Chen, C. J., Cox, J. E., Kincaid, R. P., Martinez, A. & Sullivan, C. S. Divergent MicroRNA Targetomes of Closely Related Circulating Strains of a Polyomavirus. *J. Virol.* **87**, 11135–11147 (2013).
83. Low, J. A., Magnuson, B., Tsai, B. & Imperiale, M. J. Identification of gangliosides GD1b and GT1b as receptors for BK virus. *J. Virol.* **80**, 1361–6 (2006).
84. O'Hara, S. D., Stehle, T. & Garcea, R. Glycan receptors of the Polyomaviridae: structure, function, and pathogenesis. *Curr. Opin. Virol.* **7**, 73–78 (2014).
85. Elphick, G. F. *et al.* The human polyomavirus, JCV, uses serotonin receptors to infect cells. *Science* **306**, 1380–3 (2004).
86. Maginnis, M. S., Haley, S. A., Gee, G. V & Atwood, W. J. Role of N-linked glycosylation of the 5-HT2A receptor in JC virus infection. *J. Virol.* **84**, 9677–84 (2010).
87. Mayberry, C. L., Nelson, C. D. S. & Maginnis, M. S. JC Polyomavirus Attachment and Entry: Potential Sites for PML Therapeutics. *Curr. Clin. Microbiol. Reports* **4**, 132–141 (2017).
88. Erickson, K. D., Garcea, R. L. & Tsai, B. Ganglioside GT1b is a putative host cell receptor for the Merkel cell polyomavirus. *J. Virol.* **83**, 10275–9 (2009).
89. Ströh, L. J. *et al.* Trichodysplasia spinulosa-Associated Polyomavirus Uses a Displaced Binding Site on VP1 to Engage Sialylated Glycolipids. *PLoS Pathog.* **11**, e1005112 (2015).
90. Ströh, L. J. *et al.* Structure analysis of the major capsid proteins of human polyomaviruses 6 and 7 reveals an obstructed sialic acid binding site. *J. Virol.* **88**, 10831–9 (2014).
91. Caruso, M., Belloni, L., Sthandier, O., Amati, P. & Garcia, M.-I. Alpha4beta1 integrin acts as a cell receptor for murine polyomavirus at the postattachment level. *J. Virol.* **77**, 3913–21 (2003).
92. Tsai, B. *et al.* Gangliosides are receptors for murine polyoma virus and SV40. *EMBO J.* **22**, 4346–55 (2003).
93. O'Hara, S. D. & Garcea, R. L. Murine Polyomavirus Cell Surface Receptors Activate Distinct Signaling Pathways Required for Infection. *MBio* **7**, e01836-16 (2016).
94. Campanero-Rhodes, M. A. *et al.* N-Glycolyl GM1 Ganglioside as a Receptor for Simian Virus 40. *J. Virol.* **81**, 12846–12858 (2007).

95. Pho, M. T., Ashok, A. & Atwood, W. J. JC virus enters human glial cells by clathrin-dependent receptor-mediated endocytosis. *J. Virol.* **74**, 2288–92 (2000).
96. Richterová, Z. *et al.* Caveolae are involved in the trafficking of mouse polyomavirus virions and artificial VP1 pseudocapsids toward cell nuclei. *J. Virol.* **75**, 10880–91 (2001).
97. Damm, E.-M. *et al.* Clathrin- and caveolin-1-independent endocytosis: entry of simian virus 40 into cells devoid of caveolae. *J. Cell Biol.* **168**, 477–88 (2005).
98. Zhao, L., Marciano, A. T., Rivet, C. R. & Imperiale, M. J. Caveolin- and clathrin-independent entry of BKPyV into primary human proximal tubule epithelial cells. *Virology* **492**, 66–72 (2016).
99. Schelhaas, M. *et al.* Simian Virus 40 Depends on ER Protein Folding and Quality Control Factors for Entry into Host Cells. *Cell* **131**, 516–529 (2007).
100. Geiger, R. *et al.* BAP31 and BiP are essential for dislocation of SV40 from the endoplasmic reticulum to the cytosol. *Nat. Cell Biol.* **13**, 1305–14 (2011).
101. Giorda, K. M., Raghava, S., Zhang, M. W. & Hebert, D. N. The viroporin activity of the minor structural proteins VP2 and VP3 is required for SV40 propagation. *J. Biol. Chem.* **288**, 2510–20 (2013).
102. Walczak, C. P., Ravindran, M. S., Inoue, T. & Tsai, B. A cytosolic chaperone complexes with dynamic membrane J-proteins and mobilizes a nonenveloped virus out of the endoplasmic reticulum. *PLoS Pathog.* **10**, e1004007 (2014).
103. Kuksin, D. & Norkin, L. C. Disassociation of the SV40 genome from capsid proteins prior to nuclear entry. *Virol. J.* **9**, 158 (2012).
104. Gilbert, J. M., Goldberg, I. G. & Benjamin, T. L. Cell penetration and trafficking of polyomavirus. *J. Virol.* **77**, 2615–22 (2003).
105. Griffin, B. E. & Fried, M. Amplification of a specific region of the polyoma virus genome. *Nature* **256**, 175–9 (1975).
106. Crawford, L. V., Robbins, A. K. & Nicklin, P. M. Location of the Origin and Terminus of Replication in Polyoma Virus DNA. *J. Gen. Virol.* **25**, 133–142 (1974).
107. DeCaprio, J. A. & Garcea, R. L. A cornucopia of human polyomaviruses. *Nat. Rev. Microbiol.* **11**, 264–76 (2013).
108. Bethge, T., Ajuh, E. & Hirsch, H. H. Imperfect Symmetry of Sp1 and Core Promoter Sequences Regulates Early and Late Virus Gene Expression of the Bidirectional BK Polyomavirus Noncoding Control Region. *J. Virol.* **90**, 10083–10101 (2016).
109. Alwine, J. C., Reed, S. I. & Stark, G. R. Characterization of the autoregulation of simian virus 40 gene A. *J. Virol.* **24**, 22–7 (1977).
110. Gilinger, G. & Alwine, J. C. Transcriptional activation by simian virus 40 large T antigen: requirements for simple promoter structures containing either TATA or initiator elements with variable upstream factor binding sites. *J. Virol.* **67**, 6682–8 (1993).
111. Rio, D. C. & Tjian, R. SV40 T antigen binding site mutations that affect autoregulation. *Cell* **32**, 1227–40 (1983).
112. Chen, M. C., Redenius, D., Osati-Ashtiani, F. & Fluck, M. M. Enhancer-mediated role for polyomavirus middle T/small T in DNA replication. *J. Virol.* **69**, 326–33 (1995).
113. Moens, U. *et al.* Biology, evolution, and medical importance of polyomaviruses: An update. *Infect. Genet. Evol.* **54**, 18–38 (2017).
114. Dean, F. B. *et al.* Simian virus 40 large tumor antigen requires three core replication origin domains for DNA unwinding and replication in vitro. *Proc. Natl. Acad. Sci.* **84**, 8267–8271 (1987).
115. Borowiec, J. A. & Hurwitz, J. ATP stimulates the binding of simian virus 40 (SV40) large tumor antigen to the SV40 origin of replication. *Proc. Natl. Acad. Sci.* **85**, 64–68 (1988).
116. An, P., Sáenz Robles, M. T. & Pipas, J. M. Large T Antigens of Polyomaviruses: Amazing Molecular Machines. *Annu. Rev. Microbiol.* **66**, 213–236 (2012).
117. Valle, M., Chen, X. S., Donate, L. E., Fanning, E. & Carazo, J. M. Structural basis for the cooperative assembly of large T antigen on the origin of replication. *J. Mol. Biol.* **357**, 1295–305 (2006).
118. Farmerie, W. G. & Folk, W. R. Regulation of polyomavirus transcription by large tumor antigen. *Proc. Natl. Acad. Sci. U. S. A.* **81**, 6919–23 (1984).
119. Scheidtmann, K. H., Mumby, M. C., Rundell, K. & Walter, G. Dephosphorylation of simian virus 40 large-T antigen and p53 protein by protein phosphatase 2A: inhibition by small-t antigen. *Mol. Cell. Biol.* **11**, 1996–2003 (1991).
120. Saribas, A. S., White, M. K. & Safak, M. JC virus agnoprotein enhances large T antigen binding to the origin of viral DNA replication: evidence for its involvement in viral DNA replication. *Virology* **433**, 12–26 (2012).
121. Piper, P. W. Polyoma virus transcription early during productive infection of mouse 3T6 cells. *J. Mol. Biol.* **131**, 399–407 (1979).
122. Beard, P., Acheson, N. H. & Maxwell, I. H. Strand-specific transcription of polyoma virus DNA-early in productive infection and in transformed cells. *J. Virol.* **17**, 20–6 (1975).
123. Garren, S. B., Kondaveeti, Y., Duff, M. O. & Carmichael, G. G. Global Analysis of Mouse Polyomavirus Infection Reveals Dynamic Regulation of Viral and Host Gene Expression and Promiscuous Viral RNA Editing. *PLoS Pathog.* **11**, e1005166 (2015).
124. Liu, Z. & Carmichael, G. G. Polyoma virus early-late switch: regulation of late RNA accumulation by DNA replication. *Proc. Natl. Acad. Sci. U. S. A.* **90**, 8494–8 (1993).
125. Hyde-DeRuyscher, R. & Carmichael, G. G. Polyomavirus early-late switch is not regulated at the level of transcription

initiation and is associated with changes in RNA processing. *Proc. Natl. Acad. Sci. U. S. A.* **85**, 8993–7 (1988).
126. Hyde-DeRuyscher, R. P. & Carmichael, G. G. Polyomavirus late pre-mRNA processing: DNA replication-associated changes in leader exon multiplicity suggest a role for leader-to-leader splicing in the early-late switch. *J. Virol.* **64**, 5823–32 (1990).
127. Liu, Z., Batt, D. B. & Carmichael, G. G. Targeted nuclear antisense RNA mimics natural antisense-induced degradation of polyoma virus early RNA. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 4258–62 (1994).
128. Adami, G. R., Marlor, C. W., Barrett, N. L. & Carmichael, G. G. Leader-to-leader splicing is required for efficient production and accumulation of polyomavirus late mRNAs. *J. Virol.* **63**, 85–93 (1989).
129. Gu, R., Zhang, Z., DeCerbo, J. N. & Carmichael, G. G. Gene regulation by sense-antisense overlap of polyadenylation signals. *RNA* **15**, 1154–1163 (2009).
130. Kumar, M. & Carmichael, G. G. Nuclear antisense RNA induces extensive adenosine modifications and nuclear retention of target transcripts. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 3542–7 (1997).
131. Gu, R., Zhang, Z. & Carmichael, G. G. How a Small DNA Virus Uses dsRNA but Not RNAi to Regulate Its Life Cycle. *Cold Spring Harb. Symp. Quant. Biol.* **71**, 293–299 (2006).
132. Bass, B. L. RNA Editing by Adenosine Deaminases That Act on RNA. *Annu. Rev. Biochem.* **71**, 817–846 (2002).
133. Acheson, N. H., Buetti, E., Scherrer, K. & Weil, R. Transcription of the polyoma virus genome: synthesis and cleavage of giant late polyoma-specific RNA. *Proc. Natl. Acad. Sci. U. S. A.* **68**, 2231–5 (1971).
134. Zhang, Z. & Carmichael, G. G. The fate of dsRNA in the nucleus: a p54(nrb)-containing complex mediates the nuclear retention of promiscuously A-to-I edited RNAs. *Cell* **106**, 465–75 (2001).
135. Chen, L.-L. & Carmichael, G. G. Gene regulation by SINES and inosines: biological consequences of A-to-I editing of Alu element inverted repeats. *Cell Cycle* **7**, 3294–3301 (2008).
136. Chen, L.-L. & Carmichael, G. G. Altered Nuclear Retention of mRNAs Containing Inverted Repeats in Human Embryonic Stem Cells: Functional Role of a Nuclear Noncoding RNA. *Mol. Cell* **35**, 467–478 (2009).
137. Kamen, R., Jat, P., Treisman, R., Favaloro, J. & Folk, W. R. 5' termini of polyoma virus early region transcripts synthesized in vivo by wild-type virus and viable deletion mutants. *J. Mol. Biol.* **159**, 189–224 (1982).
138. Fenton, R. G. & Basilico, C. Changes in the topography of early region transcription during polyoma virus lytic infection. *Proc. Natl. Acad. Sci. U. S. A.* **79**, 7142–6 (1982).
139. Ghosh, P. K. & Lebowitz, P. Simian virus 40 early mRNA's contain multiple 5' termini upstream and downstream from a Hogness-Goldberg sequence; a shift in 5' termini during the lytic cycle is mediated by large T antigen. *J. Virol.* **40**, 224–40 (1981).
140. Khalili, K., Feigenbaum, L. & Khoury, G. Evidence for a shift in 5'-termini of early viral RNA during the lytic cycle of JC virus. *Virology* **158**, 469–72 (1987).
141. Milavetz, B. Hyperacetylation and differential deacetylation of histones H4 and H3 define two distinct classes of acetylated SV40 chromosomes early in infection. *Virology* **319**, 324–336 (2004).
142. Milavetz, B. *et al.* Virion-mediated transfer of SV40 epigenetic information. *Epigenetics* **7**, 528–534 (2012).
143. Balakrishnan, L. & Milavetz, B. Histone hyperacetylation in the coding region of chromatin undergoing transcription in SV40 minichromosomes is a dynamic process regulated directly by the presence of RNA polymerase II. *J. Mol. Biol.* **365**, 18–30 (2007).
144. Schaffhausen, B. S. & Benjamin, T. L. Deficiency in histone acetylation in nontransforming host range mutants of polyoma virus. *Proc. Natl. Acad. Sci. U. S. A.* **73**, 1092–6 (1976).
145. Wollebo, H. S., Woldemichaele, B., Khalili, K., Safak, M. & White, M. K. Epigenetic regulation of polyomavirus JC. *Virol. J.* **10**, 264 (2013).
146. Chang, C.-F. *et al.* Analysis of DNA methylation in human BK virus. *Virus Genes* **43**, 201–207 (2011).
147. Toyooka, S. *et al.* Progressive aberrant methylation of the RASSF1A gene in simian virus 40 infected human mesothelial cells. *Oncogene* **21**, 4340–4344 (2002).
148. Shivapurkar, N. *et al.* Presence of Simian Virus 40 DNA Sequences in Human Lymphoid and Hematopoietic Malignancies and Their Relationship to Aberrant Promoter Methylation of Multiple Genes. *Cancer Res.* **64**, 3757–3760 (2004).
149. Helmbold, P. *et al.* Frequent occurrence of *RASSF1A* promoter hypermethylation and merkel cell polyomavirus in merkel cell carcinoma. *Mol. Carcinog.* **48**, 903–909 (2009).
150. Goel, A. *et al.* Association of JC Virus T-Antigen Expression With the Methylator Phenotype in Sporadic Colorectal Cancers. *Gastroenterology* **130**, 1950–1961 (2006).
151. Kumar, M. A. *et al.* Nucleosome positioning in the regulatory region of SV40 chromatin correlates with the activation and repression of early and late transcription during infection. *Virology* **503**, 62–69 (2017).
152. Kumar, M. A., Kasti, K., Balakrishnan, L. & Milavetz, B. Directed Nucleosome Sliding during the Formation of the Simian Virus 40 Particle Exposes DNA Sequences Required for Early Transcription. *J. Virol.* **93**, (2018).
153. Good, P. J., Welch, R. C., Barkan, A., Somasekhar, M. B. & Mertz, J. E. Both VP2 and VP3 are synthesized from each of the alternative spliced late 19S RNA species of simian virus 40. *J. Virol.* **62**, 944–53 (1988).
154. Smith, A. E., Kamen, R., Mangel, W. F., Shure, H. & Wheeler, T. Location of the sequences coding for capsid proteins VP1 and VP2 on polyoma virus DNA. *Cell* **9**, 481–7 (1976).
155. Schowalter, R. M. & Buck, C. B. The Merkel Cell Polyomavirus Minor Capsid Protein. *PLoS Pathog.* **9**, e1003558 (2013).

156. Acheson, N. H. Transcription during productive infection with polyoma virus and Simian virus 40. *Cell* **8**, 1–12 (1976).
157. Birg, F., Favaloro, J. & Kamen, R. Analysis of polyoma virus nuclear RNA by mini-blot hybridization. *Proc. Natl. Acad. Sci. U. S. A.* **74**, 3138–42 (1977).
158. Acheson, N. H. Polyoma virus giant RNAs contain tandem repeats of the nucleotide sequence of the entire viral genome. *Proc. Natl. Acad. Sci. U. S. A.* **75**, 4754–8 (1978).
159. Acheson, N. H. Kinetics and efficiency of polyadenylation of late polyomavirus nuclear RNA: generation of oligomeric polyadenylated RNAs and their processing into mRNA. *Mol. Cell. Biol.* **4**, 722–9 (1984).
160. Luo, Y. & Carmichael, G. G. Splice site skipping in polyomavirus late pre-mRNA processing. *J. Virol.* **65**, 6637–44 (1991).
161. Huang, Y. & Carmichael, G. G. A suboptimal 5' splice site is a cis-acting determinant of nuclear export of polyomavirus late mRNAs. *Mol. Cell. Biol.* **16**, 6046–54 (1996).
162. Carmichael, G. G. Gene Regulation and Quality Control in Murine Polyomavirus Infection. *Viruses* **8**, (2016).
163. Erickson, K. D. *et al.* Virion Assembly Factories in the Nucleus of Polyomavirus-Infected Cells. *PLoS Pathog.* **8**, e1002630 (2012).
164. Roitman-Shemer, V., Stokrova, J., Forstova, J. & Oppenheim, A. Assemblages of simian virus 40 capsid proteins and viral DNA visualized by electron microscopy. *Biochem. Biophys. Res. Commun.* **353**, 424–430 (2007).
165. Allison, A. C. & Black, P. H. Lysosomal changes in lytic and nonlytic infections with the simian vacuolating virus (SV40). *J. Natl. Cancer Inst.* **39**, 775–780, 782–7 (1967).
166. Norkin, L. C. & Ouellette, J. Cell killing by simian virus 40: variation in the pattern of lysosomal enzyme release, cellular enzyme release, and cell death during productive infection of normal and simian virus 40-transformed simian cell lines. *J. Virol.* **18**, 48–57 (1976).
167. Clayson, E. T., Brando, L. V & Compans, R. W. Release of simian virus 40 virions from epithelial cells is polarized and occurs without cell lysis. *J. Virol.* **63**, 2278–88 (1989).
168. Evans, G. L., Caller, L. G., Foster, V. & Crump, C. M. Anion homeostasis is important for non-lytic release of BK polyomavirus from infected cells. *Open Biol.* **5**, 150041 (2015).
169. Ellis, L. C., Norton, E., Dang, X. & Koralnik, I. J. Agnogene Deletion in a Novel Pathogenic JC Virus Isolate Impairs VP1 Expression and Virion Production. *PLoS One* **8**, e80840 (2013).
170. Sariyer, I. K., Saribas, A. S., White, M. K. & Safak, M. Infection by agnoprotein-negative mutants of polyomavirus JC and SV40 results in the release of virions that are mostly deficient in DNA content. *Virol. J.* **8**, 255 (2011).
171. Suzuki, T. *et al.* Role of JC virus agnoprotein in virion formation. *Microbiol. Immunol.* **56**, 639–646 (2012).
172. Suzuki, T. *et al.* The Human Polyoma JC Virus Agnoprotein Acts as a Viroporin. *PLoS Pathog.* **6**, e1000801 (2010).
173. Suzuki, T. *et al.* Viroporin activity of the JC polyomavirus is regulated by interactions with the adaptor protein complex 3. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 18668–73 (2013).
174. Johannessen, M. *et al.* BKV Agnoprotein Interacts with α-Soluble N-Ethylmaleimide-Sensitive Fusion Attachment Protein, and Negatively Influences Transport of VSVG-EGFP. *PLoS One* **6**, e24489 (2011).
175. Eash, S., Manley, K., Gasparovic, M., Querbes, W. & Atwood, W. J. The human polyomaviruses. *Cell. Mol. Life Sci.* **63**, 865–876 (2006).
176. Dalianis, T. & Hirsch, H. H. Human polyomaviruses in disease and cancer. *Virology* **437**, 63–72 (2013).
177. Monaco, M. C. G. & Major, E. O. Immune System Involvement in the Pathogenesis of JC Virus Induced PML: What is Learned from Studies of Patients with Underlying Diseases and Therapies as Risk Factors. *Front. Immunol.* **6**, 159 (2015).
178. Greenlee, J. E., Phelps, R. C. & Stroop, W. G. The major site of murine K papovavirus persistence and reactivation is the renal tubular epithelium. *Microb. Pathog.* **11**, 237–47 (1991).
179. MONINI, P. *et al.* DNA Rearrangements Impairing BK Virus Productive Infection in Urinary Tract Tumors. *Virology* **214**, 273–279 (1995).
180. Dörries, K., Vogel, E., Günther, S. & Czub, S. Infection of Human Polyomaviruses JC and BK in Peripheral Blood Leukocytes from Immunocompetent Individuals. *Virology* **198**, 59–70 (1994).
181. Vago, L. *et al.* JCV-DNA and BKV-DNA in the CNS tissue and CSF of AIDS patients and normal subjects. Study of 41 cases and review of the literature. *J. Acquir. Immune Defic. Syndr. Hum. Retrovirol.* **12**, 139–46 (1996).
182. Hingorani, S. Renal Complications of Hematopoietic-Cell Transplantation. *N. Engl. J. Med.* **374**, 2256–2267 (2016).
183. Rice, S. J., Bishop, J. A., Apperley, J. & Gardner, S. D. BK virus as cause of haemorrhagic cystitis after bone marrow transplantation. *Lancet (London, England)* **2**, 844–5 (1985).
184. Ramos, E., Drachenberg, C. B., Wali, R. & Hirsch, H. H. The decade of polyomavirus BK-associated nephropathy: state of affairs. *Transplantation* **87**, 621–30 (2009).
185. Hirsch, H. H., Kardas, P., Kranz, D. & Leboeuf, C. The human JC polyomavirus (JCPyV): virological background and clinical implications. *APMIS* **121**, 685–727 (2013).
186. Nagashima, K., Yamaguchi, K., Yasui, K. & Ogiwara, H. Progressive multifocal leukoencephalopathy. Neuropathology and virus isolation. *Acta Pathol. Jpn.* **31**, 953–61 (1981).
187. ZURHEIN, G. & CHOU, S. M. PARTICLES RESEMBLING PAPOVA VIRUSES IN HUMAN CEREBRAL DEMYELINATING DISEASE. *Science* **148**, 1477–9 (1965).
188. Ault, G. S. & Stoner, G. L. Brain and kidney of progressive multifocal leukoencephalopathy patients contain identical

rearrangements of the JC virus promoter/enhancer. *J. Med. Virol.* **44**, 298–304 (1994).
189. Barth, H. *et al.* 45 years after the discovery of human polyomaviruses BK and JC: Time to speed up the understanding of associated diseases and treatment approaches. *Crit. Rev. Microbiol.* **43**, 178–195 (2017).
190. Babakir-Mina, M., Ciccozzi, M., Perno, C. F. & Ciotti, M. The human polyomaviruses KI and WU: virological background and clinical implications. *APMIS* **121**, 746–754 (2013).
191. Ho, J. *et al.* Human Polyomavirus 7-Associated Pruritic Rash and Viremia in Transplant Recipients. *J. Infect. Dis.* **211**, 1560–1565 (2015).
192. Nguyen, K. D. *et al.* Human polyomavirus 6 and 7 are associated with pruritic and dyskeratotic dermatoses. *J. Am. Acad. Dermatol.* **76**, 932-940.e3 (2017).
193. Moens, U., Van Ghelue, M. & Johannessen, M. Oncogenic potentials of the human polyomavirus regulatory proteins. *Cell. Mol. Life Sci.* **64**, 1656–1678 (2007).
194. Cheng, J., DeCaprio, J. A., Fluck, M. M. & Schaffhausen, B. S. Cellular transformation by Simian Virus 40 and Murine Polyoma Virus T antigens. *Semin. Cancer Biol.* **19**, 218–228 (2009).
195. Van Dyke, T. A. *et al.* Relationship between simian virus 40 large tumor antigen expression and tumor formation in transgenic mice. *J. Virol.* **61**, 2029–32 (1987).
196. Manfredi, J. J. & Prives, C. The transforming activity of simian virus 40 large tumor antigen. *Biochim. Biophys. Acta* **1198**, 65–83 (1994).
197. Dyson, N. *et al.* Large T antigens of many polyomaviruses are able to form complexes with the retinoblastoma protein. *J. Virol.* **64**, 1353–6 (1990).
198. Tsai, J. & Douglas, M. G. A conserved HPD sequence of the J-domain is necessary for YDJ1 stimulation of Hsp70 ATPase activity at a site distinct from substrate binding. *J. Biol. Chem.* **271**, 9347–54 (1996).
199. Ludlow, J. W. *et al.* SV40 large T antigen binds preferentially to an underphosphorylated member of the retinoblastoma susceptibility gene product family. *Cell* **56**, 57–65 (1989).
200. Houben, R. *et al.* An intact retinoblastoma protein-binding site in Merkel cell polyomavirus large T antigen is required for promoting growth of Merkel cell carcinoma cells. *Int. J. Cancer* **130**, 847–856 (2012).
201. White, M. K. & Khalili, K. Interaction of retinoblastoma protein family members with large T-antigen of primate polyomaviruses. *Oncogene* **25**, 5286–5293 (2006).
202. Sullivan, C. S., Cantalupo, P. & Pipas, J. M. The molecular chaperone activity of simian virus 40 large T antigen is required to disrupt Rb-E2F family complexes by an ATP-dependent mechanism. *Mol. Cell. Biol.* **20**, 6233–43 (2000).
203. Harris, K. F., Christensen, J. B. & Imperiale, M. J. BK virus large T antigen: interactions with the retinoblastoma family of tumor suppressor proteins and effects on cellular growth control. *J. Virol.* **70**, 2378–86 (1996).
204. Krynska, B. *et al.* Role of cell cycle regulators in tumor formation in transgenic mice expressing the human neurotropic virus, JCV, early protein. *J. Cell. Biochem.* **67**, 223–30 (1997).
205. Dyson, N. The regulation of E2F by pRB-family proteins. *Genes Dev.* **12**, 2245–62 (1998).
206. Lilyestrom, W., Klein, M. G., Zhang, R., Joachimiak, A. & Chen, X. S. Crystal structure of SV40 large T-antigen bound to p53: interplay between a viral oncoprotein and a cellular tumor suppressor. *Genes Dev.* **20**, 2373–2382 (2006).
207. STAIB, C. *et al.* p53 Inhibits JC Virus DNA Replicationin Vivoand Interacts with JC Virus Large T-Antigen. *Virology* **219**, 237–246 (1996).
208. Zhou, A. Y. *et al.* Polyomavirus middle T-antigen is a transmembrane protein that binds signaling proteins in discrete subcellular membrane sites. *J. Virol.* **85**, 3046–54 (2011).
209. Welcker, M. & Clurman, B. E. The SV40 Large T Antigen Contains a Decoy Phosphodegron That Mediates Its Interactions with Fbw7/hCdc4. *J. Biol. Chem.* **280**, 7654–7658 (2005).
210. Pallas, D. C. *et al.* Polyoma small and middle T antigens and SV40 small t antigen form stable complexes with protein phosphatase 2A. *Cell* **60**, 167–76 (1990).
211. Sablina, A. A. & Hahn, W. C. SV40 small T antigen and PP2A phosphatase in cell transformation. *Cancer Metastasis Rev.* **27**, 137–146 (2008).
212. Ulug, E. T., Cartwright, A. J. & Courtneidge, S. A. Characterization of the interaction of polyomavirus middle T antigen with type 2A protein phosphatase. *J. Virol.* **66**, 1458–67 (1992).
213. Bollag, B., Hofstetter, C. A., Reviriego-Mendoza, M. M. & Frisque, R. J. JC Virus Small t Antigen Binds Phosphatase PP2A and Rb Family Proteins and Is Required for Efficient Viral DNA Replication Activity. *PLoS One* **5**, e10606 (2010).
214. Kwun, H. J. *et al.* Restricted Protein Phosphatase 2A Targeting by Merkel Cell Polyomavirus Small T Antigen. *J. Virol.* **89**, 4191–4200 (2015).
215. Alberts, A. S., Thorburn, A. M., Shenolikar, S., Mumby, M. C. & Feramisco, J. R. Regulation of cell cycle progression and nuclear affinity of the retinoblastoma protein by protein phosphatases. *Proc. Natl. Acad. Sci. U. S. A.* **90**, 388–92 (1993).
216. Wlodarchak, N. & Xing, Y. PP2A as a master regulator of the cell cycle. *Crit. Rev. Biochem. Mol. Biol.* **51**, 162–184 (2016).
217. Rodriguez-Viciana, P., Collins, C. & Fried, M. Polyoma and SV40 proteins differentially regulate PP2A to activate distinct cellular signaling pathways involved in growth control. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 19290–5 (2006).
218. Sontag, E., Sontag, J. M. & Garcia, A. Protein phosphatase 2A is a critical regulator of protein kinase C zeta signaling targeted by SV40 small t to promote cell growth and NF-kappaB activation. *EMBO J.* **16**, 5662–71 (1997).

219. Sontag, E. *et al.* The interaction of SV40 small tumor antigen with protein phosphatase 2A stimulates the map kinase pathway and induces cell proliferation. *Cell* **75**, 887–897 (1993).
220. Del Valle, L. & Khalili, K. Detection of human polyomavirus proteins, T-antigen and agnoprotein, in human tumor tissue arrays. *J. Med. Virol.* **82**, 806–811 (2010).
221. Gerits, N. *et al.* Agnoprotein of polyomavirus BK interacts with proliferating cell nuclear antigen and inhibits DNA replication. *Virol. J.* **12**, 7 (2015).
222. Darbinyan, A. *et al.* Alterations of DNA damage repair pathways resulting from JCV infection. *Virology* **364**, 73–86 (2007).
223. Darbinyan, A. *et al.* Evidence for dysregulation of cell cycle by human polyomavirus, JCV, late auxiliary protein. *Oncogene* **21**, 5574–5581 (2002).
224. Chang, Y. & Moore, P. S. Merkel Cell Carcinoma: A Virus-Induced Human Cancer. *Annu. Rev. Pathol. Mech. Dis.* **7**, 123–144 (2012).
225. Banks, P. D., Sandhu, S., Gyorki, D. E., Johnston, M. L. & Rischin, D. Recent Insights and Advances in the Management of Merkel Cell Carcinoma. *J. Oncol. Pract.* **12**, 637–46 (2016).
226. Cassler, N. M., Merrill, D., Bichakjian, C. K. & Brownell, I. Merkel Cell Carcinoma Therapeutic Update. *Curr. Treat. Options Oncol.* **17**, 36 (2016).
227. Shuda, M. *et al.* T antigen mutations are a human tumor-specific signature for Merkel cell polyomavirus. *Proc. Natl. Acad. Sci.* **105**, 16272–16277 (2008).
228. Liu, W., MacDonald, M. & You, J. Merkel cell polyomavirus infection and Merkel cell carcinoma. *Curr. Opin. Virol.* **20**, 20–27 (2016).
229. Borchert, S. *et al.* High-affinity Rb binding, p53 inhibition, subcellular localization, and transformation by wild-type or tumor-derived shortened Merkel cell polyomavirus large T antigens. *J. Virol.* **88**, 3144–60 (2014).
230. Shuda, M. *et al.* Merkel Cell Polyomavirus Small T Antigen Induces Cancer and Embryonic Merkel Cell Proliferation in a Transgenic Mouse Model. *PLoS One* **10**, e0142329 (2015).
231. Verhaegen, M. E. *et al.* Merkel Cell Polyomavirus Small T Antigen Is Oncogenic in Transgenic Mice. *J. Invest. Dermatol.* **135**, 1415–1424 (2015).
232. Knight, L. M. *et al.* Merkel cell polyomavirus small T antigen mediates microtubule destabilization to promote cell motility and migration. *J. Virol.* **89**, 35–47 (2015).
233. Bhatia, S., Afanasiev, O. & Nghiem, P. Immunobiology of Merkel cell carcinoma: implications for immunotherapy of a polyomavirus-associated cancer. *Curr. Oncol. Rep.* **13**, 488–97 (2011).
234. Shahzad, N. *et al.* The T antigen locus of Merkel cell polyomavirus downregulates human Toll-like receptor 9 expression. *J. Virol.* **87**, 13009–19 (2013).
235. Li, H., Gade, P., Xiao, W. & Kalvakolanu, D. V. The interferon signaling network and transcription factor C/EBP-beta. *Cell. Mol. Immunol.* **4**, 407–18 (2007).
236. Nerlov, C. The C/EBP family of transcription factors: a paradigm for interaction between gene expression and proliferation control. *Trends Cell Biol.* **17**, 318–24 (2007).
237. Paulson, K. G. *et al.* Downregulation of MHC-I expression is prevalent but reversible in Merkel cell carcinoma. *Cancer Immunol. Res.* **2**, 1071–9 (2014).
238. Chretien, A.-S. *et al.* Cancer-Induced Alterations of NK-Mediated Target Recognition: Current and Investigational Pharmacological Strategies Aiming at Restoring NK-Mediated Anti-Tumor Activity. *Front. Immunol.* **5**, 122 (2014).
239. Lipson, E. J. *et al.* PD-L1 expression in the Merkel cell carcinoma microenvironment: association with inflammation, Merkel cell polyomavirus and overall survival. *Cancer Immunol. Res.* **1**, 54–63 (2013).
240. Griffiths, D. A. *et al.* Merkel cell polyomavirus small T antigen targets the NEMO adaptor protein to disrupt inflammatory signaling. *J. Virol.* **87**, 13853–67 (2013).
241. Moens, U., Rasheed, K., Abdulsalam, I. & Sveinbjørnsson, B. The Role of Merkel Cell Polyomavirus and Other Human Polyomaviruses in Emerging Hallmarks of Cancer. *Viruses* **7**, 1871–1901 (2015).
242. Lee, S. *et al.* Identification and validation of a novel mature microRNA encoded by the Merkel cell polyomavirus in human Merkel cell carcinomas. *J. Clin. Virol.* **52**, 272–5 (2011).
243. Xie, H. *et al.* MicroRNA Expression Patterns Related to Merkel Cell Polyomavirus Infection in Human Merkel Cell Carcinoma. *J. Invest. Dermatol.* **134**, 507–517 (2014).
244. Veija, T. *et al.* miRNA-34a underexpressed in Merkel cell polyomavirus-negative Merkel cell carcinoma. *Virchows Arch.* **466**, 289–295 (2015).
245. Enam, S. *et al.* Association of human polyomavirus JCV with colon cancer: evidence for interaction of viral T-antigen and beta-catenin. *Cancer Res.* **62**, 7093–101 (2002).
246. Bulut, Y. *et al.* Potential relationship between BK virus and renal cell carcinoma. *J. Med. Virol.* **85**, 1085–9 (2013).
247. Delbue, S., Ferrante, P. & Provenzano, M. Polyomavirus BK and prostate cancer: an unworthy scientific effort? *Oncoscience* **1**, 296–303 (2014).
248. Yin, W.-Y., Lee, M.-C., Lai, N.-S. & Lu, M.-C. BK virus as a potential oncovirus for bladder cancer in a renal transplant patient. *J. Formos. Med. Assoc.* **114**, 373–4 (2015).
249. Bartel, D. P. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**, 281–97 (2004).
250. Lund, E. Nuclear Export of MicroRNA Precursors. *Science (80-.).* **303**, 95–98 (2004).

251. Bohnsack, M. T., Czaplinski, K. & Görlich, D. Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs. *RNA* **10**, 185 (2004).
252. Tomari, Y. Perspective: machines for RNAi. *Genes Dev.* **19**, 517–529 (2005).
253. Ha, M. & Kim, V. N. Regulation of microRNA biogenesis. *Nat. Rev. Mol. Cell Biol.* **15**, 509–524 (2014).
254. Iwakawa, H. & Tomari, Y. The Functions of MicroRNAs: mRNA Decay and Translational Repression. *Trends Cell Biol.* **25**, 651–665 (2015).
255. Bartel, D. P. MicroRNAs: Target Recognition and Regulatory Functions. *Cell* **136**, 215–233 (2009).
256. Lunavat, T. R. *et al.* Small RNA deep sequencing discriminates subsets of extracellular vesicles released by melanoma cells – Evidence of unique microRNA cargos. *RNA Biol.* **12**, 810–823 (2015).
257. Zhang, J. *et al.* Exosome and Exosomal MicroRNA: Trafficking, Sorting, and Function. *Genomics. Proteomics Bioinformatics* **13**, 17–24 (2015).
258. Turchinovich, A., Tonevitsky, A. G. & Burwinkel, B. Extracellular miRNA: A Collision of Two Paradigms. *Trends Biochem. Sci.* **41**, 883–892 (2016).
259. Turchinovich, A., Weiz, L., Langheinz, A. & Burwinkel, B. Characterization of extracellular circulating microRNA. *Nucleic Acids Res.* **39**, 7223–7233 (2011).
260. Hannafon, B. & Ding, W.-Q. Intercellular Communication by Exosome-Derived microRNAs in Cancer. *Int. J. Mol. Sci.* **14**, 14240–14269 (2013).
261. Valadi, H. *et al.* Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells. *Nat. Cell Biol.* **9**, 654–659 (2007).
262. Mittelbrunn, M. *et al.* Unidirectional transfer of microRNA-loaded exosomes from T cells to antigen-presenting cells. *Nat. Commun.* **2**, 282 (2011).
263. Vickers, K. C., Palmisano, B. T., Shoucri, B. M., Shamburek, R. D. & Remaley, A. T. MicroRNAs are transported in plasma and delivered to recipient cells by high-density lipoproteins. *Nat. Cell Biol.* **13**, 423–433 (2011).
264. Mittelbrunn, M. & Sánchez-Madrid, F. Intercellular communication: diverse structures for exchange of genetic information. *Nat. Rev. Mol. Cell Biol.* **13**, 328–335 (2012).
265. Wilfred, B. R., Wang, W.-X. & Nelson, P. T. Energizing miRNA research: A review of the role of miRNAs in lipid metabolism, with a prediction that miR-103/107 regulates human metabolic pathways. *Mol. Genet. Metab.* **91**, 209–217 (2007).
266. Stefani, G. & Slack, F. J. Small non-coding RNAs in animal development. *Nat. Rev. Mol. Cell Biol.* **9**, 219–230 (2008).
267. O'Connell, R. M., Rao, D. S., Chaudhuri, A. A. & Baltimore, D. Physiological and pathological roles for microRNAs in the immune system. *Nat. Rev. Immunol.* **10**, 111–122 (2010).
268. Subramanian, S. & Steer, C. J. MicroRNAs as gatekeepers of apoptosis. *J. Cell. Physiol.* n/a-n/a (2010). doi:10.1002/jcp.22066
269. Landskroner-Eiger, S., Moneke, I. & Sessa, W. C. miRNAs as Modulators of Angiogenesis. *Cold Spring Harb. Perspect. Med.* **3**, a006643–a006643 (2013).
270. Leung, A. K. L. *et al.* Poly(ADP-Ribose) Regulates Stress Responses and MicroRNA Activity in the Cytoplasm. *Mol. Cell* **42**, 489–499 (2011).
271. Vasudevan, S., Tong, Y. & Steitz, J. A. Switching from Repression to Activation: MicroRNAs Can Up-Regulate Translation. *Science (80-.).* **318**, 1931–1934 (2007).
272. Bhattacharyya, S. N., Habermacher, R., Martine, U., Closs, E. I. & Filipowicz, W. Relief of microRNA-Mediated Translational Repression in Human Cells Subjected to Stress. *Cell* **125**, 1111–1124 (2006).
273. Bracken, C. P., Scott, H. S. & Goodall, G. J. A network-biology perspective of microRNA function and dysfunction in cancer. *Nat. Rev. Genet.* **17**, 719–732 (2016).
274. Lou, W. *et al.* MicroRNAs in cancer metastasis and angiogenesis. *Oncotarget* **8**, (2017).
275. Kozomara, A. & Griffiths-Jones, S. miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res.* **39**, D152-7 (2011).
276. Cardin, S.-E. & Borchert, G. M. Viral MicroRNAs, Host MicroRNAs Regulating Viruses, and Bacterial MicroRNA-Like RNAs. in *Methods in molecular biology (Clifton, N.J.)* **1617**, 39–56 (2017).
277. Kincaid, R. P. & Sullivan, C. S. Virus-Encoded microRNAs: An Overview and a Look to the Future. *PLoS Pathog.* **8**, e1003018 (2012).
278. Theiss, J. M. *et al.* A Comprehensive Analysis of Replicating Merkel Cell Polyomavirus Genomes Delineates the Viral Transcription Program and Suggests a Role for mcv-miR-M1 in Episomal Persistence. *PLoS Pathog.* **11**, e1004974 (2015).
279. Broekema, N. M. & Imperiale, M. J. miRNA regulation of BK polyomavirus replication during early infection. *Proc. Natl. Acad. Sci.* **110**, 8200–8205 (2013).
280. McNees, A. L. *et al.* Viral microRNA effects on persistent infection of human lymphoid cells by polyomavirus SV40. *PLoS One* **13**, e0192799 (2018).
281. Burke, J. M., Bass, C. R., Kincaid, R. P., Ulug, E. T. & Sullivan, C. S. The Murine Polyomavirus MicroRNA Locus Is Required To Promote Viruria during the Acute Phase of Infection. *J. Virol.* **92**, (2018).
282. Mylin, L. M. *et al.* Quantitation of CD8(+) T-lymphocyte responses to multiple epitopes from simian virus 40 (SV40) large T antigen in C57BL/6 mice immunized with SV40, SV40 T-antigen-transformed cells, or vaccinia virus

recombinants expressing full-length T antigen or epitope minigenes. *J. Virol.* **74**, 6922–34 (2000).
283. Chen, C. J. *et al.* Naturally arising strains of polyomaviruses with severely attenuated microRNA expression. *J. Virol.* **88**, 12683–93 (2014).
284. Bauman, Y. *et al.* An Identical miRNA of the Human JC and BK Polyoma Viruses Targets the Stress-Induced Ligand ULBP3 to Escape Immune Elimination. *Cell Host Microbe* **9**, 93–102 (2011).
285. Bauman, Y. *et al.* Downregulation of the stress-induced ligand ULBP1 following SV40 infection confers viral evasion from NK cell cytotoxicity. *Oncotarget* **7**, (2016).
286. Swanson, P. A. *et al.* An MHC class Ib-restricted CD8 T cell response confers antiviral immunity. *J. Exp. Med.* **205**, 1647–57 (2008).
287. Zhu, Y., Haecker, I., Yang, Y., Gao, S.-J. & Renne, R. γ-Herpesvirus-encoded miRNAs and their roles in viral biology and pathogenesis. *Curr. Opin. Virol.* **3**, 266–75 (2013).
288. Arora, R., Chang, Y. & Moore, P. S. MCV and Merkel cell carcinoma: a molecular success story. *Curr. Opin. Virol.* **2**, 489–98 (2012).
289. Dela Cruz, F. N. *et al.* Novel Polyomavirus associated with Brain Tumors in Free-Ranging Raccoons, Western United States. *Emerg. Infect. Dis.* **19**, 77–84 (2013).
290. Brostoff, T., Dela Cruz, F. N., Church, M. E., Woolard, K. D. & Pesavento, P. A. The raccoon polyomavirus genome and tumor antigen transcription are stable and abundant in neuroglial tumors. *J. Virol.* **88**, 12816–24 (2014).
291. Renwick, N. *et al.* Multicolor microRNA FISH effectively differentiates tumor types. *J. Clin. Invest.* **123**, 2694–2702 (2013).
292. Fimia, G. M., Corazzari, M., Antonioli, M. & Piacentini, M. Ambra1 at the crossroad between autophagy and cell death. *Oncogene* **32**, 3311–3318 (2013).
293. Sung, C. K., Yim, H., Andrews, E. & Benjamin, T. L. A mouse polyomavirus-encoded microRNA targets the cellular apoptosis pathway through Smad2 inhibition. *Virology* **468–470**, 57–62 (2014).
294. You, X., Zhang, Z., Fan, J., Cui, Z. & Zhang, X.-E. Functionally Orthologous Viral and Cellular MicroRNAs Studied by a Novel Dual-Fluorescent Reporter System. *PLoS One* **7**, e36157 (2012).
295. Micolucci, L., Akhtar, M. M., Olivieri, F., Rippo, M. R. & Procopio, A. D. Diagnostic value of microRNAs in asbestos exposure and malignant mesothelioma: systematic review and qualitative meta-analysis. *Oncotarget* **7**, 58606–58637 (2016).
296. Li, S. *et al.* miR-423-5p contributes to a malignant phenotype and temozolomide chemoresistance in glioblastomas. *Neuro. Oncol.* **19**, 55–65 (2017).
297. Zhang, X. *et al.* miR-1266 Contributes to Pancreatic Cancer Progression and Chemoresistance by the STAT3 and NF-κB Signaling Pathways. *Mol. Ther. - Nucleic Acids* **11**, 142–158 (2018).
298. Dubensky, T. W., Murphy, F. A. & Villarreal, L. P. Detection of DNA and RNA virus genomes in organ systems of whole mice: patterns of mouse organ infection by polyomavirus. *J. Virol.* **50**, 779–83 (1984).
299. Rowe, W. P., Hartley, J. W., Estes, J. D. & Huebner, R. J. Growth curves of polyoma virus in mice and hamsters. *Natl. Cancer Inst. Monogr.* **4**, 189–209 (1960).
300. McCance, D. J. Growth and persistence of polyoma early region deletion mutants in mice. *J. Virol.* **39**, 958–62 (1981).
301. Dawe, C. J. *et al.* Variations in polyoma virus genotype in relation to tumor induction in mice. Characterization of wild type strains with widely differing tumor profiles. *Am. J. Pathol.* **127**, 243–61 (1987).
302. Dubensky, T. W., Freund, R., Dawe, C. J. & Benjamin, T. L. *Polyomavirus Replication in Mice: Influences of VP1 Type and Route of Inoculation*. *J. Virol* **65**, (1991).
303. Dubensky, T. W. & Villarreal, L. P. The primary site of replication alters the eventual site of persistent infection by polyomavirus in mice. *J. Virol.* **50**, 541–6 (1984).
304. Rochford, R., Moreno,', J. P., Peake, M. L. & Villarreal", L. P. *Enhancer Dependence of Polyomavirus Persistence in Mouse Kidneys*. *J. Virol* (1992).
305. Berke, Z. & Dalianis, T. Persistence of polyomavirus in mice infected as adults differs from that observed in mice infected as newborns. *J. Virol.* **67**, 4369–71 (1993).
306. Wirth, J. J., Amalfitano, A., Gross, R., Oldstone, M. B. & Fluck, M. M. Organ- and age-specific replication of polyomavirus in mice. *J. Virol.* **66**, 3278–86 (1992).
307. Atencio, I. A., Shadan, F. F., Zhou, X. J., Vaziri, N. D. & Villarreal, L. P. Adult mouse kidneys become permissive to acute polyomavirus infection and reactivate persistent infections in response to cellular damage and regeneration. *J. Virol.* **67**, 1424–32 (1993).
308. Mccance, D. J. & Mims, C. A. *Reactivation of Polyoma Virus in Kidneys of Persistently Infected Mice During Pregnancy*. *Infection and immunity* **25**, (1979).
309. Rowe, W. P. The epidemiology of mouse polyoma virus infection. *Bacteriol. Rev.* **25**, 18–31 (1961).
310. Martelli, F. *et al.* Polyomavirus microRNA in saliva reveals persistent infectious status in the oral cavity. *Virus Res.* **249**, 1–7 (2018).
311. Huang, G., Zeng, G., Huang, Y., Ramaswami, B. & Randhawa, P. Evaluation of the Gastrointestinal Tract as Potential Route of Primary Polyomavirus Infection in Mice. (2016). doi:10.1371/journal.pone.0150786
312. Law, L. W. Studies of the significance of tumor antigens in induction and repression of neoplastic diseases: presidential address. *Cancer Res.* **29**, 1–21 (1969).

313. Rowe, W. P., Hartley, J. W., Brodsky, I., Huebner, R. J. & Law, L. W. Observations on the spread of mouse polyoma virus infection. *Nature* **182**, 1617 (1958).
314. Gottlieb, K. & Villarreal, L. P. The Distribution and Kinetics of Polyomavirus in Lungs of Intranasally Infected Newborn Mice. *Virology* **266**, 52–65 (2000).
315. Liunggren, G., Liunggren, H. G. & Dalianis, T. T cell subsets involved in immunity against polyoma virus-induced tumors. *Virology* **198**, 714–6 (1994).
316. Wilson, J. J. *et al.* CD8 T Cells Recruited Early in Mouse Polyomavirus Infection Undergo Exhaustion. *J. Immunol.* **188**, 4340–4348 (2012).
317. Martin, J. D. & Foster, G. C. Multiple JC Virus Genomes from One Patient. *J. Gen. Virol.* **65**, 1405–1411 (1984).
318. Flaegstad, T. *et al.* Amplification and sequencing of the control regions of BK and JC virus from human urine by polymerase chain reaction. *Virology* **180**, 553–60 (1991).
319. Newman, J. T. & Frisque, R. J. Detection of archetype and rearranged variants of JC virus in multiple tissues from a pediatric PML patient. *J. Med. Virol.* **52**, 243–252 (1997).
320. Polo, C. *et al.* Prevalence and patterns of polyomavirus urinary excretion in immunocompetent adults and children. *Clin. Microbiol. Infect.* **10**, 640–644 (2004).
321. Markowitz, R. B. *et al.* BK virus and JC virus shed during pregnancy have predominantly archetypal regulatory regions. *J. Virol.* **65**, 4515–9 (1991).
322. Olsen, G.-H. *et al.* Genetic variability in BK Virus regulatory regions in urine and kidney biopsies from renal-transplant patients. *J. Med. Virol.* **78**, 384–93 (2006).
323. Olsen, G.-H., Hirsch, H. H. & Rinaldo, C. H. Functional analysis of polyomavirus BK non-coding control region quasispecies from kidney transplant recipients. *J. Med. Virol.* **81**, 1959–1967 (2009).
324. Bethge, T. *et al.* Sp1 sites in the noncoding control region of BK polyomavirus are key regulators of bidirectional viral early and late gene expression. *J. Virol.* **89**, 3396–411 (2015).
325. Ajuh, E. T. *et al.* Novel Human Polyomavirus Noncoding Control Regions Differ in Bidirectional Gene Expression according to Host Cell, Large T-Antigen Expression, and Clinically Occurring Rearrangements. *J. Virol.* **92**, (2018).
326. Hou, J. & Major, E. Management of infections by the human polyomavirus JC: past, present and future. *Expert Rev. Anti. Infect. Ther.* **3**, 629–640 (2005).
327. Murphy, E., Vanicek, J., Robins, H., Shenk, T. & Levine, A. J. Suppression of immediate-early viral gene expression by herpesvirus-coded microRNAs: Implications for latency. *Proc. Natl. Acad. Sci.* **105**, 5453–5458 (2008).
328. Forte, E. & Luftig, M. A. The role of microRNAs in Epstein-Barr virus latency and lytic reactivation. *Microbes Infect.* **13**, 1156–1167 (2011).
329. Speck, S. H. & Ganem, D. Viral latency and its regulation: lessons from the gamma-herpesviruses. *Cell Host Microbe* **8**, 100–15 (2010).
330. Whitley, R., Kimberlin, D., Diseases, B. R.-C. I. & 1998, U. *Herpes simplex viruses*. *Howley (Eds.), Fields' virology (5th edn.), Lippincott, Philadelphia* (2007).
331. Antinone, S. E. & Smith, G. A. Retrograde axon transport of herpes simplex virus and pseudorabies virus: a live-cell comparative analysis. *J. Virol.* **84**, 1504–12 (2010).
332. Knipe, D. M. & Cliffe, A. Chromatin control of herpes simplex virus lytic and latent infection. *Nat. Rev. Microbiol.* **6**, 211–221 (2008).
333. Knipe, D. M. *et al.* Snapshots: Chromatin control of viral infection. *Virology* **435**, 141–156 (2013).
334. Deshmane, S. L. & Fraser, N. W. During latency, herpes simplex virus type 1 DNA is associated with nucleosomes in a chromatin structure. *J. Virol.* **63**, 943–7 (1989).
335. Mellerick, D. M. & Fraser, N. W. Physical state of the latent herpes simplex virus genome in a mouse model system: evidence suggesting an episomal state. *Virology* **158**, 265–75 (1987).
336. Stevens, J. G., Wagner, E. K., Devi-Rao, G. B., Cook, M. L. & Feldman, L. T. RNA complementary to a herpesvirus alpha gene mRNA is prominent in latently infected neurons. *Science* **235**, 1056–9 (1987).
337. Umbach, J. L., Nagel, M. A., Cohrs, R. J., Gilden, D. H. & Cullen, B. R. Analysis of human alphaherpesvirus microRNA expression in latently infected human trigeminal ganglia. *J. Virol.* **83**, 10677–83 (2009).
338. Umbach, J. L. *et al.* MicroRNAs expressed by herpes simplex virus 1 during latent infection regulate viral mRNAs. *Nature* **454**, 780–783 (2008).
339. Garber, D. A., Schaffer, P. A. & Knipe, D. M. A LAT-associated function reduces productive-cycle gene expression during acute infection of murine sensory neurons with herpes simplex virus type 1. *J. Virol.* **71**, 5885–93 (1997).
340. Mador, N., Goldenberg, D., Cohen, O., Panet, A. & Steiner, I. Herpes simplex virus type 1 latency-associated transcripts suppress viral replication and reduce immediate-early gene mRNA levels in a neuronal cell line. *J. Virol.* **72**, 5067–75 (1998).
341. Cliffe, A. R., Garber, D. A. & Knipe, D. M. Transcription of the herpes simplex virus latency-associated transcript promotes the formation of facultative heterochromatin on lytic promoters. *J. Virol.* **83**, 8182–90 (2009).
342. Wang, Q.-Y. *et al.* Herpesviral latency-associated transcript gene promotes assembly of heterochromatin on viral lytic-gene promoters in latent infection. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 16055–9 (2005).
343. Cliffe, A. R., Coen, D. M. & Knipe, D. M. Kinetics of facultative heterochromatin and polycomb group protein association with the herpes simplex viral genome during establishment of latent infection. *MBio* **4**, e00590-12 (2013).

344. Thompson, R. L. & Sawtell, N. M. Herpes Simplex Virus Type 1 Latency-Associated Transcript Gene Promotes Neuronal Survival. *J. Virol.* **75**, 6660–6675 (2001).
345. Kwiatkowski, D. L., Thompson, H. W. & Bloom, D. C. The polycomb group protein Bmi1 binds to the herpes simplex virus 1 latent genome and maintains repressive histone marks during latency. *J. Virol.* **83**, 8173–81 (2009).
346. Flores, O. *et al.* Mutational inactivation of herpes simplex virus 1 microRNAs identifies viral mRNA targets and reveals phenotypic effects in culture. *J. Virol.* **87**, 6589–603 (2013).
347. Cai, W. *et al.* The herpes simplex virus type 1 regulatory protein ICP0 enhances virus replication during acute infection and reactivation from latency. *J. Virol.* **67**, 7501–12 (1993).
348. Halford, W. P., Kemp, C. D., Isler, J. A., Davido, D. J. & Schaffer, P. A. ICP0, ICP4, or VP16 Expressed from Adenovirus Vectors Induces Reactivation of Latent Herpes Simplex Virus Type 1 in Primary Cultures of Latently Infected Trigeminal Ganglion Cells. *J. Virol.* **75**, 6143–6153 (2001).
349. Pan, D. *et al.* A Neuron-Specific Host MicroRNA Targets Herpes Simplex Virus-1 ICP0 Expression and Promotes Latency. *Cell Host Microbe* **15**, 446–456 (2014).
350. Zheng, S., Li, Y., Zhang, Y., Li, X. & Tang, H. MiR-101 regulates HSV-1 replication by targeting ATP5B. *Antiviral Res.* **89**, 219–226 (2011).
351. Wang, X. *et al.* ICP4-induced miR-101 attenuates HSV-1 replication. *Sci. Rep.* **6**, 23205 (2016).
352. Ru, J. *et al.* MiR-23a Facilitates the Replication of HSV-1 through the Suppression of Interferon Regulatory Factor 1. *PLoS One* **9**, e114021 (2014).
353. Zhang, Y., Dai, J., Tang, J., Zhou, L. & Zhou, M. MicroRNA-649 promotes HSV-1 replication by directly targeting MALT1. *J. Med. Virol.* **89**, 1069–1079 (2017).
354. Yu, J. W. *et al.* MALT1 Protease Activity Is Required for Innate and Adaptive Immune Responses. *PLoS One* **10**, e0127083 (2015).
355. Enk, J. *et al.* HSV1 MicroRNA Modulation of GPI Anchoring and Downstream Immune Evasion. *Cell Rep.* **17**, 949–956 (2016).
356. Perna, J. J., Mannix, M. L., Rooney, J. F., Notkins, A. L. & Straus, S. E. Reactivation of latent herpes simplex virus infection by ultraviolet light: a human model. *J. Am. Acad. Dermatol.* **17**, 473–8 (1987).
357. Pazin, G. J. *et al.* Prevention of reactivated herpes simplex infection by human leukocyte interferon after operation on the trigeminal root. *N. Engl. J. Med.* **301**, 225–30 (1979).
358. Spivack, J. G. & Fraser, N. W. Expression of herpes simplex virus type 1 latency-associated transcripts in the trigeminal ganglia of mice during acute infection and reactivation of latent infection. *J. Virol.* **62**, 1479–85 (1988).
359. Amelio, A. L., Giordani, N. V, Kubat, N. J., O'neil, J. E. & Bloom, D. C. Deacetylation of the herpes simplex virus type 1 latency-associated transcript (LAT) enhancer and a decrease in LAT abundance precede an increase in ICP0 transcriptional permissiveness at early times postexplant. *J. Virol.* **80**, 2063–8 (2006).
360. von Laer, D. *et al.* Detection of cytomegalovirus DNA in CD34+ cells from blood and bone marrow. *Blood* **86**, 4086–90 (1995).
361. Mendelson, M., Monard, S., Sissons, P. & Sinclair, J. Detection of endogenous human cytomegalovirus in CD34+ bone marrow progenitors. *J. Gen. Virol.* **77 (Pt 12)**, 3099–102 (1996).
362. Groves, I. J., Reeves, M. B. & Sinclair, J. H. Lytic infection of permissive cells with human cytomegalovirus is regulated by an intrinsic "pre-immediate-early" repression of viral gene expression mediated by histone post-translational modification. *J. Gen. Virol.* **90**, 2364–74 (2009).
363. Reeves, M. B., MacAry, P. A., Lehner, P. J., Sissons, J. G. P. & Sinclair, J. H. Latency, chromatin remodeling, and reactivation of human cytomegalovirus in the dendritic cells of healthy carriers. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 4140–5 (2005).
364. Murphy, J. C., Fischle, W., Verdin, E. & Sinclair, J. H. Control of cytomegalovirus lytic gene expression by histone acetylation. *EMBO J.* **21**, 1112–1120 (2002).
365. Wright, E., Bain, M., Teague, L., Murphy, J. & Sinclair, J. Ets-2 repressor factor recruits histone deacetylase to silence human cytomegalovirus immediate-early gene expression in non-permissive cells. *J. Gen. Virol.* **86**, 535–44 (2005).
366. Rossetto, C. C., Tarrant-Elorza, M. & Pari, G. S. Cis and trans acting factors involved in human cytomegalovirus experimental and natural latent infection of CD14 (+) monocytes and CD34 (+) cells. *PLoS Pathog.* **9**, e1003366 (2013).
367. O'Connor, C. M., Vanicek, J. & Murphy, E. A. Host microRNA regulation of human cytomegalovirus immediate early protein translation promotes viral latency. *J. Virol.* **88**, 5524–32 (2014).
368. Goodrum, F., Reeves, M., Sinclair, J., High, K. & Shenk, T. Human cytomegalovirus sequences expressed in latently infected individuals promote a latent infection in vitro. *Blood* **110**, 937–45 (2007).
369. Mucke, K. *et al.* Human Cytomegalovirus Major Immediate Early 1 Protein Targets Host Chromosomes by Docking to the Acidic Pocket on the Nucleosome Surface. *J. Virol.* **88**, 1228–1248 (2014).
370. Poole, E. *et al.* Latency-Associated Viral Interleukin-10 (IL-10) Encoded by Human Cytomegalovirus Modulates Cellular IL-10 and CCL8 Secretion during Latent Infection through Changes in the Cellular MicroRNA hsa-miR-92a. *J. Virol.* **88**, 13947–13955 (2014).
371. Jenkins, C. *et al.* Immunomodulatory properties of a viral homolog of human interleukin-10 expressed by human cytomegalovirus during the latent phase of infection. *J. Virol.* **82**, 3736–50 (2008).

372. Umashankar, M. *et al.* A Novel Human Cytomegalovirus Locus Modulates Cell Type-Specific Outcomes of Infection. *PLoS Pathog.* **7**, e1002444 (2011).
373. Wang, X., Huong, S.-M., Chiu, M. L., Raab-Traub, N. & Huang, E.-S. Epidermal growth factor receptor is a cellular receptor for human cytomegalovirus. *Nature* **424**, 456–461 (2003).
374. Chan, G., Nogalski, M. T. & Yurochko, A. D. Activation of EGFR on monocytes is required for human cytomegalovirus entry and mediates cellular motility. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 22369–74 (2009).
375. Reeves, M. B., Breidenstein, A. & Compton, T. Human cytomegalovirus activation of ERK and myeloid cell leukemia-1 protein correlates with survival of latently infected cells. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 588–93 (2012).
376. Stern-Ginossar, N. *et al.* Host Immune System Gene Targeting by a Viral miRNA. *Science (80-.).* **317**, 376–381 (2007).
377. Kim, Y. *et al.* Human cytomegalovirus clinical strain-specific microRNA miR-UL148D targets the human chemokine RANTES during infection. *PLoS Pathog.* **8**, e1002577 (2012).
378. Grey, F., Meyers, H., White, E. A., Spector, D. H. & Nelson, J. A Human Cytomegalovirus-Encoded microRNA Regulates Expression of Multiple Viral Genes Involved in Replication. *PLoS Pathog.* **3**, e163 (2007).
379. Thorley-Lawson, D. A., Hawkins, J. B., Tracy, S. I. & Shapiro, M. The pathogenesis of Epstein-Barr virus persistent infection. *Curr. Opin. Virol.* **3**, 227–32 (2013).
380. Amon, W. & Farrell, P. J. Reactivation of Epstein-Barr virus from latency. *Rev. Med. Virol.* **15**, 149–156 (2005).
381. Rickinson, A. & Kieff, E. Epstein-Barr virus. (2001).
382. Kieff, E. & Rickinson, A. Epstein-Barr virus and its replication. (2001).
383. Saha, A. & Robertson, E. S. Epstein-Barr virus-associated B-cell lymphomas: pathogenesis and clinical outcomes. *Clin. Cancer Res.* **17**, 3056–63 (2011).
384. Izumi, K. M., Kaye, K. M. & Kieff, E. D. The Epstein-Barr virus LMP1 amino acid sequence that engages tumor necrosis factor receptor associated factors is critical for primary B lymphocyte growth transformation. *Proc. Natl. Acad. Sci.* **94**, 1447–1452 (1997).
385. Luftig, M. *et al.* Epstein-Barr virus latent infection membrane protein 1 TRAF-binding site induces NIK/IKK alpha-dependent noncanonical NF-kappaB activation. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 141–6 (2004).
386. Izumi, K. M. & Kieff, E. D. The Epstein-Barr virus oncogene product latent membrane protein 1 engages the tumor necrosis factor receptor-associated death domain protein to mediate B lymphocyte growth transformation and activate NF-kappaB. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 12592–7 (1997).
387. Frappier, L. EBNA1 and host factors in Epstein–Barr virus latent DNA replication. *Curr. Opin. Virol.* **2**, 733–739 (2012).
388. Lo, A. K. F. *et al.* Modulation of LMP1 protein expression by EBV-encoded microRNAs. *Proc. Natl. Acad. Sci.* **104**, 16164–16169 (2007).
389. Lung, R. W.-M. *et al.* Modulation of LMP2A expression by a newly identified Epstein-Barr virus-encoded microRNA miR-BART22. *Neoplasia* **11**, 1174–84 (2009).
390. Navari, M., Etebari, M., Ibrahimi, M., Leoncini, L. & Piccaluga, P. P. Pathobiologic Roles of Epstein-Barr Virus-Encoded MicroRNAs in Human Lymphomas. *Int. J. Mol. Sci.* **19**, 1168 (2018).
391. Choy, E. Y.-W. *et al.* An Epstein-Barr virus–encoded microRNA targets PUMA to promote host cell survival. *J. Exp. Med.* **205**, 2551–2560 (2008).
392. Albanese, M., Tagawa, T., Buschle, A. & Hammerschmidt, W. MicroRNAs of Epstein-Barr Virus Control Innate and Adaptive Antiviral Immunity. *J. Virol.* **91**, (2017).
393. Lu, Y. *et al.* Epstein-Barr Virus miR-BART6-3p Inhibits the RIG-I Pathway. *J. Innate Immun.* **9**, 574–586 (2017).
394. Arthur, R. R. & Shah, K. V. Occurrence and significance of papovaviruses BK and JC in the urine. *Prog. Med. Virol.* **36**, 42–61 (1989).
395. Aird, D. *et al.* Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biol.* **12**, R18 (2011).
396. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10 (2011).
397. Chen, L. & Fluck, M. Kinetic analysis of the steps of the polyomavirus lytic cycle. *J. Virol.* **75**, 8368–79 (2001).
398. Chesters, P. M., Heritage, J. & McCance, D. J. Persistence of DNA Sequences of BK Virus and JC Virus in Normal Human Tissues and in Diseased Tissues. *J. Infect. Dis.* **147**, 676–684 (1983).
399. Shinohara, T. *et al.* BK virus infection of the human urinary tract. *J. Med. Virol.* **41**, 301–5 (1993).
400. Jiang, M., Abend, J. R., Johnson, S. F. & Imperiale, M. J. The role of polyomaviruses in human disease. *Virology* **384**, 266–273 (2009).
401. Elsner, C. & Dörries, K. Evidence of human polyomavirus BK and JC infection in normal brain tissue. *Virology* **191**, 72–80 (1992).
402. Martini, F. *et al.* Papilloma and polyoma DNA tumor virus sequences in female genital tumors. *Cancer Invest.* **22**, 697–705 (2004).
403. Pietropaolo, V. *et al.* Detection and Sequence Analysis of Human Polyomaviruses DNA from Autoptic Samples of HIV-1 Positive and Negative Subjects. *Int. J. Immunopathol. Pharmacol.* **16**, 269–276 (2003).
404. Das, D., Shah, R. B. & Imperiale, M. J. Detection and expression of human BK virus sequences in neoplastic prostate tissues. *Oncogene* **23**, 7031–7046 (2004).
405. Grinnell, B. W., Padgett, B. L. & Walker, D. L. Distribution of Nonintegrated DNA from JC Papovavirus in Organs of

Patients with Progressive Multifocal Leukoencephalopathy. *J. Infect. Dis.* **147**, 669–675 (1983).
406. Rampelli, S. *et al.* Characterization of the human DNA gut virome across populations with different subsistence strategies and geographical origin. *Environ. Microbiol.* **19**, 4728–4735 (2017).
407. Martini, F. *et al.* SV40 early region and large T antigen in human brain tumors, peripheral blood cells, and sperm fluids from healthy individuals. *Cancer Res.* **56**, 4820–5 (1996).
408. Rowe, W. P., Hartley, J. W., Estes, J. D. & Huebner, R. J. *Oncogenesis of Mammary Glands, Skin, and Bones by Polyomavirus Correlates with Viral Persistence and Prolonged Genome Replication Potential*. *J. Virol* **71**, (1997).