Characterization of the Merkel cell polyomavirus encoded miRNA mcv-miR-M1

DISSERTATION

zur Erlangung des Doktorgrades (Dr. rer. nat.)
an der Fakultät für Mathematik, Informatik und Naturwissenschaften
Fachbereich Biologie
der Universität Hamburg

vorgelegt von
Juliane Marie Elisabeth Theiß
geb. Kiermeier

Hamburg, 2016

Tag der Disputation: 05. Februar 2016

Gutachter: Prof. Dr. Nicole Fischer
Prof. Dr. Thomas Dobner
Prof. Dr. Jürgen Becker

Prüfungsvorsitzende: Prof. Dr. Julia Kehr

Dedicated to my family

Abstract

Merkel cell polyomavirus (MCPyV) persists asymptomatically in the majority of the healthy human population and is considered the etiological agent in the majority of Merkel cell carcinoma (MCC) cases, a rare but highly aggressive skin cancer. Due to the lack of appropriate model systems, the natural biology of the virus, including the mechanisms that lead to the establishment and maintenance of lifelong persistence, are only poorly understood.

MCPyV belongs to a subgroup of polyomaviruses that express microRNAs (miRNAs). miRNAs are small non-coding RNAs produced by eukaryotes and certain viruses to post transcriptionally regulate gene expression. Viral miRNAs can target different viral and cellular mRNAs and are often associated with regulation of the host's immune response and maintenance of viral persistence.

Due to its perfect complementarity to mRNAs encoding for viral Tumor-Antigens (T-Ag), the MCPyV miRNA mcv-miR-M1 has the potential to downregulate T-Ag expression. However, if and how mcv-miR-M1 influences the viral life cycle remains elusive. The aim of this work was to elucidate the role of mcv-miR-M1 during viral replication using a recently developed semi-permissive MCPyV model system with a synthetic, replication-competent MCPyV genome (MCVSyn). Using this system, the present study provides the first comprehensive transcriptional analysis of authentically replicating MCVSyn and provides evidence for the involvement of a polyomavirus miRNA in viral persistence.

Analysis of a mutant virus which is deficient for mcv-miR-M1 expression revealed that mcv-miR-M1 causes a reduction in viral genome replication via downregulation of LT-Ag expression. ChIP-Seq and RNA-Seq analyses of replicating MCVSyn indicated that, in addition to mechanisms that depend on inefficient termination of late strand transcription, mcv-miR-M1 can also be expressed independently from NCCR-initiated transcription from an autonomous RNA-pol II promoter that is embedded within the early coding region.

Unexpectedly, MCVSyn was found to establish episomal persistence in a small subset of cells for several months, providing the unique opportunity to study the role of mcv-miR-M1 in viral persistence. Strikingly, in the absence of mcv-miR-M1 expression, MCVSyn was severely limited in its ability to persist over long periods of time, suggesting that mcv-miR-M1 is an essential factor for the maintenance of persistence.

The observation that LT-Ag negatively affects cellular proliferation, but is required for maintenance of viral replication suggests a model in which mcv-miR-M1 plays an important role for the successful maintenance of persistence of viral episomes via balancing LT-Ag expression levels to allow viral DNA replication while minimizing detrimental effects on host cell survival.

This study represents the first comprehensive transcriptomic analysis of a replicating human polyomavirus and by characterizing the role of the viral miRNA in short-term and long-term replication significantly contributes to our understanding of viral miRNAs and their importance for viral persistence.

Zusammenfassung

Das humane Tumorvirus Merkelzellpolyomavirus (MCPyV) persistiert asymptomatisch im überwiegenden Teil der erwachsenen Bevölkerung und ist die Hauptursache der Entstehung von Merkelzellkarzinomen (MCC), einer seltenen aber höchst aggressiven Form von Hautkrebs. Der Lebenszyklus von MCPyV, inklusive der Mechanismen, die die Etablierung und Aufrechterhaltung der Persistenz ermöglichen, ist aufgrund fehlender Modellsysteme bislang kaum erforscht. Einige Polyomaviren, darunter MCPyV, exprimieren microRNAs (miRNAs), kurze, nicht-kodierende RNAs, die in Eukaryoten und einigen Viren die Genexpression auf mRNA-Ebene regulieren können. Virale miRNAs können sowohl die virale als auch die zelluläre Genexpression beeinflussen und dabei unter anderem eine Rolle in der Etablierung viraler Persistenz spielen. Die MCPyV miRNA mcv-miR-M1 besitzt perfekte Komplementarität zu den frühen viralen Transkripten und hat daher das Potential die Expression der Tumor-Antigene (T-Ag) zu verringern. Ob mcv-mir-M1 jedoch tatsächlich die Expression der T-Antigene während der viralen Replikation reguliert und ob dies den viralen Lebenszyklus beeinflusst, ist bislang nicht untersucht worden.

Das Ziel dieser Doktorarbeit war es daher, die Rolle von mcv-miR-M1 während der viralen Replikation aufzuklären. Dazu wurde ein synthetisches MCPyV-Genom verwendet, MCVSyn, das in einem kürzlich etablierten semi-permissives Replikationssystem authentisch repliziert. Mit Hilfe dieses Systems wurde in der vorliegenden Arbeit zum ersten Mal eine umfassende Charakterisierung des MCPyV Transkriptoms durchgeführt. Darüber hinaus konnte in dieser Arbeit erstmalig beobachtet werden, dass eine Polyomavirus kodierte miRNA die Etablierung viraler Persistenz beeinflusst.

Die Charakterisierung einer MCVSyn-Mutante, die keine reife miRNA produzieren kann zeigte, dass mcv-miR-M1 die Expression der T-Antigene verringert und dadurch einen negativen Effekt auf die Virusreplikation besitzt. In RNA-Seq und ChIP-Seq Analysen konnte beobachtet werden, dass mcv-miR-M1 nicht nur aus späten Transkripten, die in der NCCR initiiert werden und über das gesamte virale Genom fortlaufen, prozessiert werden kann, sondern auch in Abwesenheit von NCCR-abhängiger Transkription von einem eigenen RNA-pol II abhängigen Promoter.

Im Laufe dieser Arbeit wurde die unerwartete Beobachtung gemacht, dass MCVSyn für einige Monate in wenigen Zellen persistieren kann, was die bisher einzigartige Möglichkeit eröffnet, den Einfluss einer Polyomavirus kodierten miRNA auf die virale Persistenz zu untersuchen. Interessanterweise stellte sich dabei heraus, dass die Anwesenheit von mcv-miR-M1 essentiell für die Persistenz viraler Episome ist. Die Beobachtung, dass LT-Ag die zelluläre Proliferation negativ beeinflusst, jedoch unabdingbar für die Virusreplikation ist, impliziert ein Model, in dem ein Minimum von LT-Ag für die Aufrechterhaltung viraler Persistenz benötigt wird, jedoch zu

hohe LT-Ag Level einen negativen Effekt auf das Überleben persistent infizierter Zellen besitzen. Die Fähigkeit der viralen miRNA die LT-Ag Expression zu balancieren, ist in diesem Szenario daher ein entscheidender Faktor für die Etablierung und Aufrechterhaltung viraler Persistenz.
Mit diesen Ergebnissen trägt die vorliegende Arbeit maßgeblich zum Verständnis der Funktionen viraler miRNAs bei und bestätigt die bisherigen Vermutungen, dass Polyomavirus kodierte miRNAs eine Rolle in viraler Persistenz spielen können.

Table of Contents

1 Introduction

1.1 Identification of a polyomavirus causally linked to Merkel Cell Carcinoma

Merkel Cell Carcinoma (MCC) is a rare but aggressive human skin cancer. The malignancy was first described in 1972 as "trabecular carcinoma of the skin" [1] and was later on re-named when further characterization of tumor cells suggested that the cancer arises from Merkel cells [2]. Merkel cells are neuroendocrine mechanoreceptor cells located in the stratum basale of the epidermis [3]. More recent investigations question the origin of MCC and suggest that the malignancy rather originates from cutaneous progenitor cells [4] or early B-cell stages [5].
About 1500 cases of MCC are annually reported in the United States [6,7]. The incidence of MCC is rising which can be explained by improved diagnosis, increased exposure to known risk factors such as UV light and steadily increasing life expectancy with concomitant loss of immune competence in elderly people [8-10]. MCC patients have a poor prognosis with a 5-year overall survival rate of approximately 50%, which is primarily caused by rapid metastatic spread [11].
The predominant occurrence of MCC in immunosuppressed and elderly individuals inferred the involvement of an infectious agent in MCC tumorigenesis. In 2008, by applying high throughput sequencing of cDNA from MCC tissue, a single non-human transcript was identified with high sequence identity to Tumor-Antigen (T-Ag) transcripts of African green monkey lymphotropic polyomavirus (LPyV) and human BK polyomavirus (BKPyV) [12]. Based on this discovery, the sequence of a novel polyomavirus (PyV) could be assembled. The genome of this virus was found to be monoclonally integrated into the cellular genome of 80% of examined MCC samples implicating that the integration of the viral genome precedes and probably causes tumorigenesis in the majority of MCC cases [12]. Due to its association with MCC, the novel polyomavirus was named Merkel cell polyomavirus (MCPyV). The current state of knowledge about MCPyV is summarized in chapter 1.4.

1.2 Polyomaviruses in animals and humans

The family *Polyomaviridae* comprises a rising number of avian and mammal viruses with small, non-enveloped, icosahedral virions, which surround a double-stranded (ds) DNA genome [13].
The first polyomavirus, murine polyomavirus (MuPyV), was identified in 1953 as a filterable agent with a high transforming potential in newborn mice [14]. The name polyomavirus refers to the ability of this novel virus to cause numerous tumors in infected animals (*poly* = many and -*oma* = tumor). A very similar virus was identified a few years later as a contaminant in

poliovirus vaccines produced in monkey kidney cells [15-17]. This so-called Simian Vacuolating Virus 40 (SV40) is naturally found in monkeys and exhibits an equally high transforming potential in newborn hamsters as MuPyV but does not cause transformation of its natural host cells [15-17]. Research on SV40 and MuPyV has significantly contributed to our current knowledge about DNA replication, transcription, splicing, tumorigenesis and other fundamental cellular processes [13,18,19].

The first human polyomaviruses were discovered in 1971 and named after the initials of the patients from which they were isolated. JC polyomavirus (JCPyV) was discovered in brain tissue from a patient suffering from progressive multifocal leukoencephalopathy (PML) [20]. BK polyomavirus (BKPyV) was isolated from the urine of a nephropathic kidney transplant recipient [21] and this disease was subsequently termed polyomavirus-associated nephropathy (PVAN). Although both viruses are phylogenetically closely related to SV40 (Figure 1) and can cause fatal diseases by lytic infections that result in tissue necrosis, they are not oncogenic in humans.

Although several novel animal polyomaviruses were described in the following years, for more than 30 years no additional polyomaviruses were found in humans. Since 2007, primarily owing to innovations in the field of sequencing technologies, the number of known human polyomaviruses significantly increased. To date, the human polyomavirus (HPyV) family consists of 13 members. Figure 1 depicts a phylogenetic tree of HPyVs and their closest relatives based on their LT-Ag amino acid (aa) sequences.

Besides BKPyV and JCPyV the list of HPyV comprises the polyomaviruses HPyV6, 7, 9, 10, 12 [22-25] which were named after the order of their discovery, KIPyV [26] and WUPyV [27], which carry the names of the institutions where they were first described, Malawi polyomavirus (MWPyV) [28], St Louis polyomavirus (STLPyV) [29] and New Jersey polyomavirus (NJPyV-2013) [30] which were named after the source of the original virus isolate and finally the disease associated Merkel cell polyomavirus (MCPyV) [12] and Trichodysplasia Spinulosa associated polyomavirus (TSPyV) [31].

The seroprevalence of HPyVs is high in the general human population and it is hypothesized that after a primary asymptomatic infection in early childhood most HPyVs establish a lifelong persistence by so far unknown mechanisms [32-35]. Although these observations imply that most individuals are persistently infected with several different HPyVs, polyomavirus associated disorders are rare and can be considered as exceptional complication upon loss of immunocompetence [36]. JCPyV, BKPyV, TSPyV and MCPyV are so far the only HPyVs that are causally linked to diseases in humans and it remains to be investigated if other HPyVs have the ability to cause medical conditions in their host, as well.

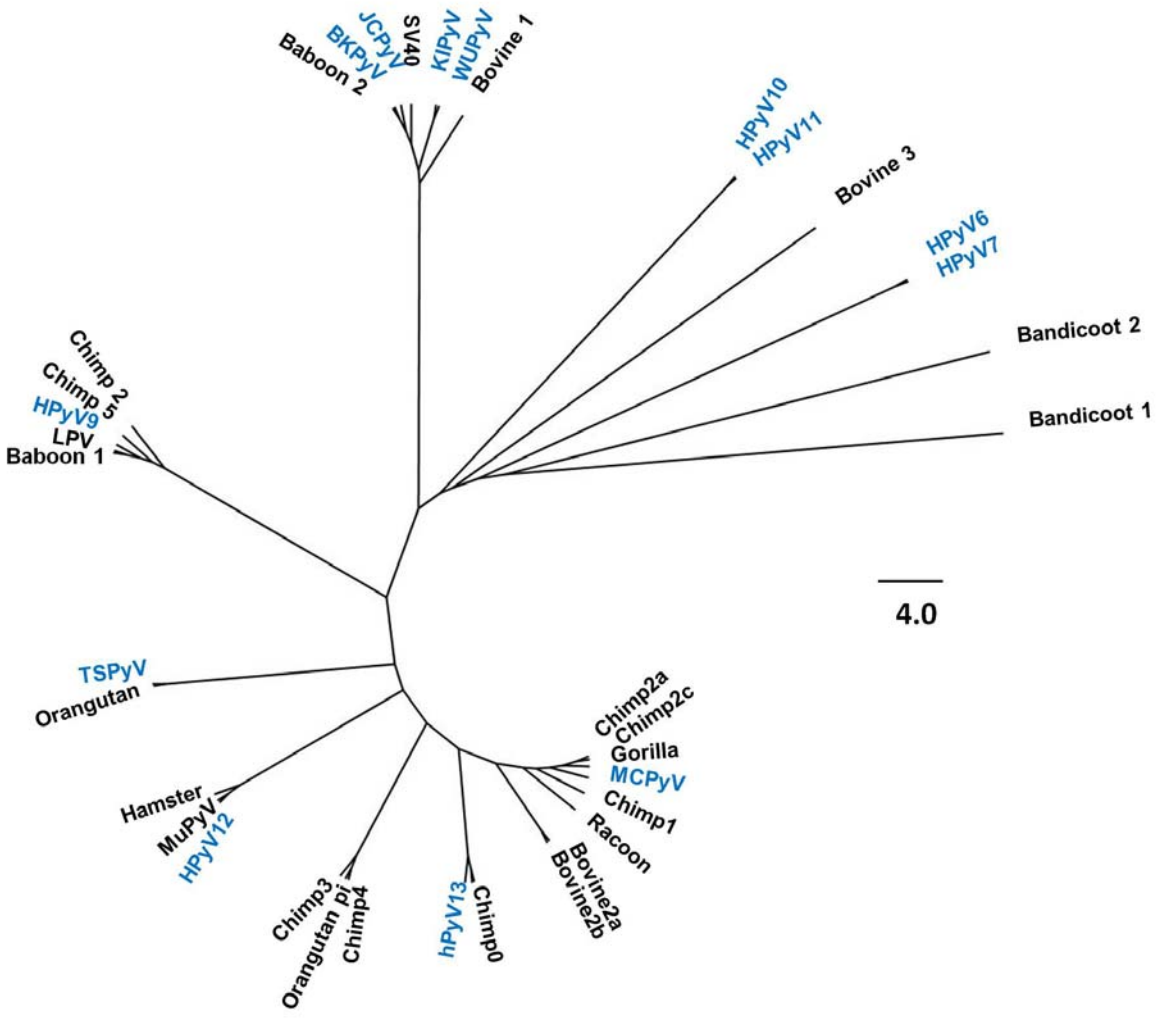

Figure 1: Phylogenetic tree of human polyomaviruses and their closest relatives based on Large T-Antigen amino acid (aa) sequences

Large T-Antigen aa sequences from the indicated viruses were aligned using MUSCLE 3.7 [37], subsequently the phylogenetic tree was generated using PhyML 3.0 [38] and Figtree, Tree Figure Drawing Tool Version 1.4.2. [39] Human polyomaviruses are highlighted in blue.
Specific viruses included were BK virus (BKPyV), NC_001538.1; JC virus (JCPyV), NC_001699.1; WUPyV, NC_009539.1; KIPyV, NC_009238.1; hPyV6, NC_014406.1; hPyV7, NC_014407.1; MCPyV, JN707599.1; TSPyV, NC_014361.1; hPyV9, HQ696593.1; hPyV10, JX262162.1; hPyV11, KF525270.1; hPyV12, NC_020890.1; hPyV13, NC_024118.1; MuPyV, NC_001515.1; SV40, NC_001669.1.; Gorilla PyV, NC_025380.1; Bovine PyV1, KM496323.1; Bovine PyV2, NC_025811.1; Bovine PyV3, NC_025800.1; Hamster PyV, M26281.1; Orangutan PyV, NC_013439.1; Racoon PyV, NC_023845.1; Baboon PyV1, NC_025894.1; Baboon PyV 2, NC_025897.1; Chimpanzee PyV, NC_014743.1; Chimpanzee PyV1, NC_025368.1; Chimpanzee PyV2a, NC_025370.1; Chimpanzee PyV3, NC_019855.1; Chimpanzee PyV4 NC_019856.1; Chimpanzee PyV5, NC_019857.1 and LPV, M30540.1). (Figure: Nicole Fischer and Adam Grundhoff, unpublished).

Strikingly, MCPyV is the only human polyomavirus that can induce tumors in its natural host. Interestingly, a MCPyV-related animal polyomavirus, Raccoon polyomavirus (RacPyV), was shown to cause cancer under biological conditions in its natural adult host [40]. No RacPyV

particles are detectable in these raccoon brain tumor cells, but, in contrast to MCPyV in MCC, the RacPyV is found episomally and not integrated into the host cell genome [40], indicating that MCPyV and RacPyV associated tumors arise via different mechanisms.

1.3 Genomic organization and life cycle of polyomaviruses

Polyomaviruses have a dsDNA genome of ~5 kb in size which can be divided into three functionally distinct segments, as depicted in Figure 2A for MCPyV: the early coding region, the late coding region and the non-coding control region (NCCR) which is located between the early and late region and contains the early and late promoters as well as the origin of replication [13].

Alternative splicing allows the expression of a variety of different viral proteins from the limited coding capacity. From the early coding region at least two proteins, the small and Large Tumor-Antigens (sT-Ag and LT-Ag) are produced and some PyVs express additional T-Ag isoforms from the same coding region [13]. The T-Ags of polyomaviruses are, as their name indicates, potentially tumorigenic, although the majority of known polyomaviruses does not cause transformation of their natural host cell. The transforming potential of T-Ags is based on their interaction with cellular proteins that function in cell cycle regulation and tumor suppression [41,42]. Among them, LT-Ags of a group of PyVs closely related to SV40 bind p53 to block p53-dependent gene expression in response to DNA damage signals [43].

All PyVs LT-Ags have the capacity to bind the Retinoblastoma protein (Rb) via a LXCXE motif [44]. The LT-Ag mediated disruption of the interaction between Rb and transcription factors of the E2F family is required to promote cell cycle entry and progression [45]. Since PyVs are dependent on the cellular DNA replication machinery for replication of the viral genome, sequestering of Rb ensures activation of the DNA replication machinery by driving resting cells into the S-phase. To induce viral replication, LT-Ag assembles in multimeric complexes and binds to GRGGC motifs at the origin of replication with its origin binding domain (OBD) [46]. The helicase-ATPase domain of LT-Ag unwinds the viral DNA to allow assembly of the cellular DNA replication-complex at the origin of replication [2].

The late region of polyomaviruses encodes the structural viral proteins VP1 and VP2 [13] and in most cases a third capsid protein, VP3 [47]. Some PyVs additionally express the non-structural VP4 which functions as a vipoporin by forming hydrophilic pores in the host cell membrane [42]. Another late non-structural protein, Agnoprotein, which is encoded by a group of PyV, plays a role in virus assembly and egress [48].

Moreover, a fraction of polyomaviruses encode for miRNAs, which are located in antisense orientation to the early gene region and have the ability to regulate the expression of early

proteins [49-53]. Distribution and functions of polyomavirus encoded miRNA are described in detail in chapter 1.7.

Polyomaviruses generally have a very narrow host cell tropism, regarding the species as well as the cell type [13], which hampers the establishment of *in vitro* and *in vivo* model systems. As a result, although functions of polyomavirus proteins are well understood, many steps of the viral life cycle remain elusive. Given that the majority of the human population is persistently infected with different HPyVs after a primary infection early in childhood [32-35], polyomaviruses must have developed strategies to enter a persistent state and avoid the host's immune response. However, the underlying mechanisms for these strategies remain in most parts elusive.

Presumably, HPyVs persist at different body sites (reviewed in [54]) but identifying the cell tropism of HPyVs has proved to be extremely challenging. Moreover, if persistence of polyomaviruses requires continuous or episodic low-level virus shedding, or whether the viral genome persists in the absence of particle production (comparable to herpesvirus latency), is still not known. Only JCPyV, BKPyV and TSPyV are known to reactivate from their persistent state upon immune suppression of the host and cause diseases by lytic replication [31,36,55]. Reactivation of other than the mentioned polyomaviruses has so far not been observed, although MCPyV was hypothesized to undergo enhanced replication before the onset of MCC [56] (see chapter 1.4.4 for the mechanisms of MCPyV induced tumorigenesis).

1.4 Merkel cell polyomavirus (MCPyV)

1.4.1 Epidemiology of MCPyV

Like other human polyomaviruses, MCPyV is endemic in the human population. Up to 80% of healthy human adults have been shown to produce serum antibodies against the MCPyV major capsid protein VP1 [32,33,57]. While antibodies for VP1 are undetectable in prenatal blood [58], seropositivity for MCPyV dramatically increases in young children, suggesting that infection with MCPyV occurs early during childhood, followed by the establishment of life-long viral persistence [22,59]. Infection with MCPyV is mostly asymptomatic [60] and, besides MCC, no clinical symptoms have been described in persistent MCPyV infection of immunocompetent hosts.

MCPyV DNA can be found in various body fluids (reviewed in [61]) as well as in sewage [62] and on environmental surfaces [63]. Encapsidated virions of MCPyV are chronically shed from the human skin [22], indicating that virus transmission occurs via smear infection or the fecal-oral-route.

1.4.2 Genomic organization and gene products of MCPyV

Despite strong similarities between MCPyV and other human and animal polyomaviruses, the genome and gene products of MCPyV comprise several unique features.

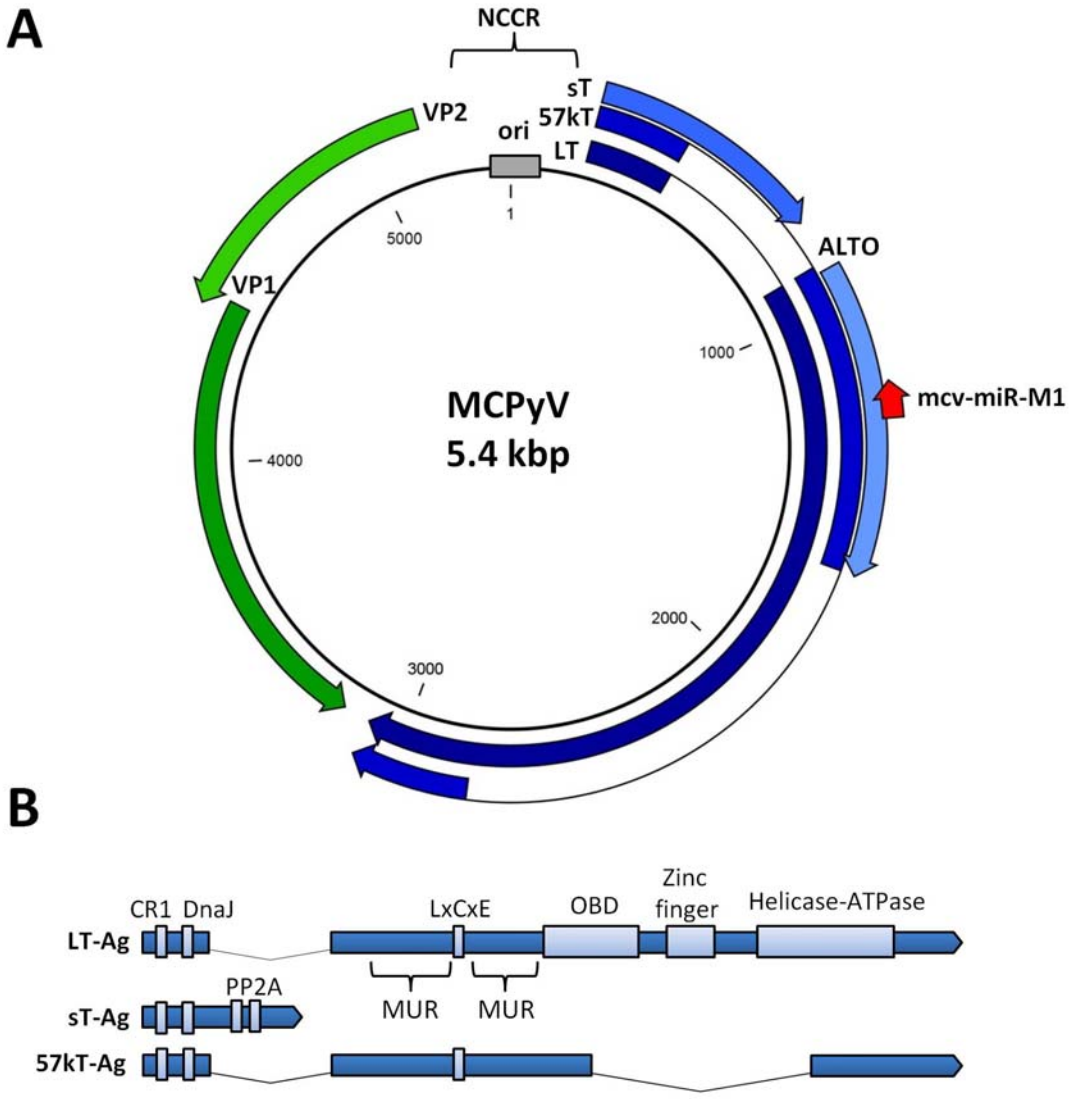

Figure 2: Characteristic features of the MCPyV genome and MCPyV gene products
(A) Schematic depiction of the MCPyV genome with the so far annotated splice variants of early and late genes. The origin of replication (ori), shown as a grey box, is located in the non-coding control region (NCCR) together with predicted bidirectional promoters for early and late gene expression. The early region encodes for the Tumor-Antigens LT-Ag, sT-Ag, 57kT-Ag and ALTO (blue arrows). The late region, which is transcribed in the opposite direction, encodes the viral structural proteins VP1 and VP2 (green arrows). In the orientation of late gene expression the viral miRNA mcv-miR-M1 lies encoded in the early gene region (red arrow). **(B)** Transcripts and functional domains of MCPyV T-Antigens. LT-Ag, sT-Ag and 57kT-Ag share the N-terminal CR1 and DnaJ domains. sT-Ag contain a PP2A binding motif while LT-Ag and 57kT-Ag share the Rb-binding motif (LXCXE). LT-Ag additionally contains an origin binding domain (OBD), a Zinc finger motif and a Helicase-ATPase domain for the induction of viral replication. The LXCXE motif is flanked by MCPyV unique regions (MUR). No functional domains are so far annotated for ALTO. (Figure: Nicole Fischer and Adam Grundhoff, modified).

The MCPyV genome has a size of 5.4 kb and strongly resembles other PyV genomes in its general organization (Figure 2A) [12]. The NCCR consists of a 71-base pair origin of replication that, comparable to other polyomaviruses, has an AT-rich region, a segment harboring LT-binding sites and an early enhancer region [61]. However, GRGGC motifs in the origin of replication that help direct assembly of LT-Ag on the DNA by specifically associating with the OBD are more frequent and located in closer proximity to each other in comparison to other PyVs [64]. It has therefore been proposed that binding of LT-Ag to the NCCR and initiation of replication may represent a more complex mechanism than in other polyomaviruses [61].

The MCPyV early region was described to express four T-Antigens: LT-Ag, sT-Ag, 57kT-Ag [12,65,66] and, by usage of an alternative reading frame, ALTO (Alternate frame of the Large T open reading frame) [67]. LT-Ag, sT-Ag and 57kT-Ag are generated by alternative splicing and share a short N-terminus but have distinct C-terminal regions [12,65,66] (Figure 2B). MCPyV LT-Ag, in comparison to other PyV LT-Ags, displays conserved features, including CR1 and DnaJ domains, an Rb-binding motif, an origin-binding domain and a helicase-ATPase domain highly suggestive of conserved functions with other polyomavirus LT-Ags. However, MCPyV LT-Ag possesses only 30% calculated amino acid identity to other HPyV LT-Ags [12,61]. A unique region of MCPyV LT-Ag (MUR) flanking the Rb-binding site binds to and re-localizes the human protein Vam6p into the nucleus although the implications of this process are not yet fully solved [68]. Additionally, the MUR seems to play a role in recruitment of Bromodomain Protein Brd4 to the site of viral replication which is critical for efficient MCPyV replication [69].

The MCPyV sT-Ag shares the first 78 amino acids with LT-Ag, including the CR1 and DNAJ domain. In its unique C-terminus, sT-Ag harbors a protein phosphatase 2A (PP2A) binding site, which is conserved among different polyomavirus sT-Ags [12]. The ability to bind PP2A has been shown to play an important role in polyomavirus-induced transformation [70]. Moreover, the PP2A binding site of MCPyV sT-Ag is required for optimal viral replication [64]. In contrast to LT-Ag and sT-Ag, the functions of 57kT-Ag and ALTO during viral replication are only incompletely understood [67].

MCPyV capsids contain the two capsid proteins VP1 and VP2, which are encoded by the late gene region [47]. Notably, MCPyV is the only human polyomavirus, which belongs to a group of polyomaviruses that lack the conserved VP3 N-terminal motif. In accordance with this fact, MCPyV was shown not to express the minor capsid protein VP3 [47]. No open reading frames (ORFs) are discernible for late non-structural proteins, like VP4 or Agnoprotein [47]. In the orientation of the late gene expression but in the region of the early genes, a pre-miRNA is encoded which leads to the expression of two mature miRNAs mcv-miR-M1 5p and 3p [51]. Characteristics of polyomavirus encoded miRNAs, including mcv-miR-M1, are further described in chapter 1.7.

1.4.3 Life cycle of MCPyV

Polyomaviruses in general are highly selective with regard to cell tropism [13] and although viral DNA can be isolated from the human skin suggestive of skin cells being the reservoir of infection, it remains elusive which cells are naturally infected by MCPyV [22,71]. The lack of knowledge about the natural reservoir of MCPyV complicates the establishment of fully permissive replication systems. However, semi-permissive model systems using recombinant consensus MCPyV genomes were recently established [72,73]. These systems allow the investigation of replicating MCPyV genomes but do not efficiently produce mature viral particles [72-74]. Further information about MCPyV model systems is provided in chapter 1.4.5.

Due to the lack of fully permissive replication systems, many aspects of the viral life cycle have not yet been elucidated. Figure 3 depicts a model of MCPyV replication and MCPyV induced tumorigenesis, which integrates observations of MCPyV protein functions in immortalized cell lines and the current knowledge about life cycles of other polyomaviruses.

The infectious lifecycle of polyomaviruses is divided into an early and late phase, with the early phase beginning with the adsorption of virions to the cell surface [75]. Attachment and entry of MCPyV likely differs from mechanisms observed in other polyomaviruses. While most of the studied polyomaviruses bind to specific sialic acid bearing glycans at the cell surface, MCPyV instead requires glycosaminoglycans, such as heparan sulfate, for initial attachment and sialylated glycans as co-receptor [74,76]. The usage of different receptors may in part explain the narrow host cell tropism of polyomaviruses. Following attachment, polyomaviruses are internalized into the cytosol by endocytosis, albeit via different pathways [13], and most likely MCPyV also takes advantage of the host's endocytic machinery in a still elusive manner. Once in the cytoplasm, polyomaviruses have to be transported to the nucleus for viral replication, a process which was shown to be supported by the microtubule network in case of JCPyV and BKPyV [77,78]. In the nucleus, polyomaviruses are uncoated and early viral gene expression, the production of T-Antigens, is immediately initiated from chromatinized viral genomes [13]. MCPyV LT-Ag, by binding to Rb, drives infected cells into the S-phase to induce activation of the cellular DNA replication machinery [45]. Moreover, LT-Ag assembles in double hexamers and, supported by sT-Ag, binds to GRGGC motifs at the origin of replication to induce viral replication by unwinding the viral DNA via its helicase-ATPase domain [64,79].

After the onset of viral DNA replication, polyomaviruses usually induce the expression of late genes to produce the capsid proteins. The switch from early to late gene expression in different polymaviruses can be dependent on different cellular transcription factors, binding of LT-Ag at the NCCR, changes in RNA processing, as well as rearrangements of the NCCR. [13,80-84]. So far no factors have been identified that efficiently induce late gene expression in MCPyV. All MCPyV replication systems are stalled at this phase and thus it is likely that the available system lacks

one or more cellular factors of natural host cells, which contribute to the switch from early to late gene expression.

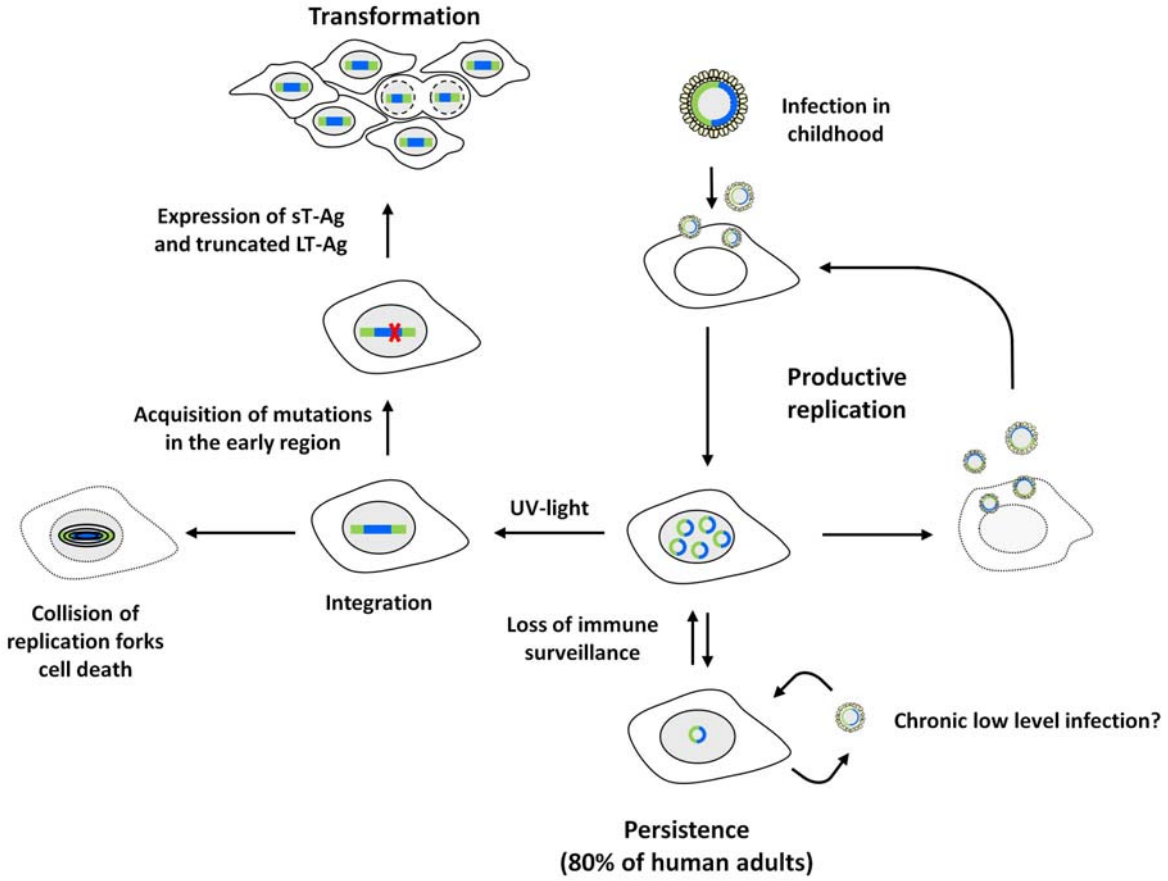

Figure 3: Model of the MCPyV life cycle and MCPyV induced tumorigenesis
Schematic representation of the MCPyV life cycle and a simplified model of MCPyV induced tumorigenesis. Primary infection with MCPyV occurs early in childhood and develops to an asymptomatic persistent infection in approx 80% of the human population. Whether or not virus is shed from cells persistently infected with MCPyV is currently not known. Likewise, if and under which conditions a lytic replication takes place remains elusive.
In the current model of MCPyV induced tumorigenesis, loss of immunocompetence leads to enhanced viral replication, which, probably together with additional risk factors (e.g. UV exposure), leads to integration of the viral genome. Only if mutations are acquired before or after the integration event, which render the virus incapable of replication, transformation can progress. The continuous expression of viral T-Antigens leads to uncontrolled proliferation and eventually to transformation and development of MCC. (Figure: Adam Grundhoff, modified).

Virus assembly occurs at the nuclear membrane, supported by Agnoprotein in some polyomaviruses [13,85], and finally, virions are released by cell lysis or endosomal egress [13]. MCPyV particles are shed from the skin [22], but tissue damage resulting from uncontrolled lytic replication as in BKPyV and JCPyV has so far not been observed for MCPyV.

As stated above, polyomaviruses usually establish a persistent infection in their natural host, meaning a prolonged presence of the virus without clearance by the immune system. Due to the

lack of appropriate model systems neither the nature of polyomavirus persistence (low-level infection versus strictly non-productive persistence of viral genomes) nor the processes that lead to the establishment and maintenance of persistence could be investigated so far.

1.4.4 MCPyV induced tumorigenesis

Polyomavirus induced transformation is associated with loss of viral replication capacity, which can occur by infection of non-permissive cells, and constitutive de-regulation of the cell cycle and cellular proliferation upon expression of viral T-Antigens [19,86-89].

Approximately 80% of MCC tumors harbor MCPyV genomes [7,12,90,91]. Strikingly, in all MCPyV positive MCC tumor cells, the viral genome is found to be monoclonally integrated into the host cell genome, which prevents viral replication [12,92]. Often multiple copies of MCPyV DNA are detectable per tumor cell in a head-to-tail orientation [92-94]. Integration sites of MCPyV are found randomly distributed over the human genome and have thus far not been associated with the transformation process. Notably, the integration sites between primary tumor and subsequent metastases are identical [12,92]. The monoclonal integration pattern of MCPyV implies that integration occurs before the tumor cell precursor undergoes clonal expansion. However, integration of the viral DNA is only the first step towards transformation.

All MCPyV sequences retrieved from tumor cells carry mutations that prevent viral replication and progeny production [64,73,90]. Most importantly, mutations in the second exon of the LT-Ag coding region lead to the expression of a truncated LT-Ag in MCC cells. Truncated LT-Ag lacks the OBD and helicase-ATPase domains rendering it incapable to induce viral replication [12,66,95]. Rarely, mutations in late region or NCCR of integrated MCPyV are observed [64,72,90].

Mutations in the early region of integrated MCPyV are thought to be the result of a selective pressure to prevent initiation of DNA replication at integrated viral origins, possibly leading to collision of replication forks and subsequent cell death [66]. Strikingly, although LT-Ag truncating mutations differ between tumor isolates, an additional positive selective pressure seems to exist to preserve the sT open reading frame and the LT-Ag N-terminus including the Rb-binding domain and thus to retain the ability to stimulate cellular proliferation [66,96]. The tight binding that was observed between truncated LT-Ag and Rb and a statistical analysis of the positions of mutations in a large number of MCC isolates strongly argues for a positive selection to preserve the Rb-binding motif in MCC [96].

While global reduction of T-Ag expression in MCC significantly reduces the proliferation rate and survival of MCC cells [97], individual knockdown of sT-Ag expression leads to reduced cell growth, but cannot fully recapitulate the phenotype of combined knockdown of LT- and sT-Ag

[98]. sT-Ag is expressed in nearly all MCPyV positive MCC tumors and has been shown to exhibit transforming potential in transformation assays likely through interaction with the translation regulator 4E-BP1, leading to deregulation of cap-dependent translation [98]. These observations suggest a synergistic role for sT- and LT-Ag during MCC tumorigenesis [97,98].

So far, the exact mechanisms that cause integration of MCPyV in the host cell genome and lead to acquisition of LT-Ag truncating mutations remain elusive. However, since loss of immune competence and advanced age are known risk factors for MCC development and MCC patients were found to have significantly higher MCPyV VP1 antibody titers than the average population, it can be assumed that MCPyV induced transformation is preceded by a phase of acute viremia [56]. Shedding of viruses from the persistent reservoir might cause infection of non-permissive cells. The replication block in non-permissive cells might be a prerequisite for transformation.

Ultraviolet radiation (UVR) was identified as an additional risk factor for MCC development by the observation that MCC predominantly arises on sun-exposed areas of fair-skinned individuals [10,99,100] and that LT-Ag truncating mutations exhibit pyrimidine substitutions which occur by UVR-induced mutagenesis [66]. Thus, MCC might also arise from the natural host cell of MCPyV if persistent episomes acquire UVR-induced mutations. However, if MCC arises in the natural reservoir cell of MCPyV or in inadvertently infected non-permissive cells remains to be elucidated. Moreover, the temporal order of viral integration and occurrence of mutations is still not known, although LT-Ag mediated unlicensed DNA replication after integration of MCPyV might create a strong selective pressure on abrogating the replication promoting functions of LT-Ag.

Taken together, MCPyV induced tumorigenesis requires integration of the viral genome, acquisition of specific mutations that prevent viral replication and the continuous expression of the viral oncogenes sT-Ag and LT-Ag, which promote cellular proliferation (Figure 4). The combined requirement for these events might be one of the reasons why MCPyV induced tumorigenesis is rare, despite the high abundance of MCPyV in the human population. The inability of MCPyV to produce infectious progeny in MCC renders the tumor cell a dead end for the viral life cycle. Thus, formation of MCC can be regarded as a rare complication of MCPyV infection rather than a part of the viral survival strategy.

1.4.5 Model systems for MCPyV

The lack of knowledge about the cell tropism of MCPyV and the generally narrow host range of polyomaviruses have hampered the development of model systems. Although no fully permissive model system is available for MCPyV, semi-permissive cell-culture based replication systems are currently used to investigate early phases of the viral life cycle [67,72-74]. All

established replication systems so far utilize recombinant MCPyV plasmids from which the viral genome has to be excised and re-ciruclarized prior to transfection into immortalized cell lines or primary cells. Typically, only 3-5% of cells transfected with re-circularized MCPyV genomes support viral replication [72]. Since it is difficult to enrich live MCPyV positive cells, it remains elusive why only a minority of cells allows viral replication. Replication of viral DNA is supported in cell lines of different origin, namely PFSK-1, H1299, HEK293 and the HEK293 derived cell line HEK293-4T, which stably expresses the early genes of SV40 and MCPyV [72-74]. Initially, two replication systems were concurrently developed using replication competent consensus MCPyV clones which were originally derived from repaired MCC isolates, but later on were found to be identical to MCPyV field isolates [72,73]. Both MCPyV clones show robust viral replication as early as two days after initial transfection, expression of T-Antigens and VP1 and a limited amount of particle production. However, efficient virus transmission or cell lysis could not be observed in either of the systems [72,73].

Co-transfection of a recombinant MCPyV field isolate with MCPyV LT-Ag and sT-Ag expression plasmids into MCPyV LT-Ag and sT-Ag-adapted HEK293 cells yields significantly higher amounts of viral particles which could be used for infection studies [71]. Nevertheless, in comparison to other viruses, the amount of produced particles is low and virions are not efficiently shed from the cells.

Due to their narrow host range, human polyomaviruses will only replicate in human cells, which complicates the establishment of animal models. With regard to MCPyV induced transformation and tumorigenesis, two animal models have been established, a xenograft mouse model system and transgenic mice expressing the tumor specific early gene locus harboring premature stop-mutations in the LT-Ag region [101-106]. MCPyV positive MCC cell lines were shown to grow as xenotransplants in NOD-SCID mice and to subsequently form macroscopic tumors similar to MCC [101,104-106].

Two transgenic mouse models allow studying the transforming functions of MCPyV T-Antigens *in vivo* [107,108]. Transgenic mice expressing MCPyV sT-Ag showed alterations in epithelial differentiation, proliferation, and apoptosis and sT-Ag expression in preterm embryos was found to be sufficient for transformation of several epithelia including the skin [107]. In a second model, the oncogenic activity of MCC tumor-derived T-Antigens was investigated, using a conditional, tissue-specific mouse model in which T-Antigens are expressed in the stratified epithelium of the skin and in Merkel cells [108]. This study revealed that MCC-derived T-Antigens are oncogenic *in vivo* by causing increased cellular proliferation, unlicensed DNA synthesis, decreased differentiation and DNA damage response [108]. Although transformation is observed in both transgenic mouse models, no MCC-like lesions are formed. This suggests that either other factors in addition to the MCPyV T-Antigens are required for MCC development or

that MCC arises in other than the tested cell types. The described animal models provide a useful tool to study MCC and the transforming potential of MCPyV T-Antigens *in vivo,* but similar models for replicating MCPyV have not yet been established.

While ectopic expression of viral genes has unraveled several aspects of the MCPyV life cycle and transformation, a profound understanding of the virus biology will require a fully permissive model system. Nevertheless, the described semi-permissive MCPyV model system is very useful to study early phases of the viral life cycle, especially viral replication and early gene expression.

The MCPyV semi-permissive replication system developed by Neumann et al. [72] employs a synthetic MCPyV genome, termed MCVSyn, which was designed as a consensus sequence of MCC-derived viral genomes deposited in the NCBI database and later on proved to be identical to MCPyV field isolates. MCVSyn was used in this work to study the role of the MCPyV encoded miRNA during viral replication as described in the following chapters.

1.5 Functions and biogenesis of miRNAs

microRNAs (miRNAs) are short non-coding RNAs which are produced by all eukaryotes and some viruses for the post transcriptional regulation of gene expression. The approximately 22 nt long miRNAs were discovered in 1993 in *C. elegans* as regulators of larval stage development [109]. Since then, the number of discovered miRNAs has exceeded 28,000 (according to miRbase release 21 [110]) and diverse functions could be assigned to some miRNAs, e.g. regulation of cell differentiation, innate and adaptive immune response, apoptosis, cell proliferation and tumorigenesis [111]. However, a large number of miRNAs still await functional characterization.

The biogenesis of miRNAs is strongly conserved among eukaryotes and the canonical pathway of miRNA biogenesis is illustrated in Figure 4. miRNA coding regions can be found distributed over the genome in protein-coding genes as well as genes that produce non-coding RNAs [112,113]. Although some miRNAs are transcribed by RNA-polymerase III [114], the great majority of miRNA precursors are transcribed as capped and poly-adenylated RNAs by RNA-polymerase II (RNA-pol II). These so called primary miRNAs (pri-miRNAs) are typically more than 1 kb in size and, in the region that encodes the mature miRNA sequence, fold into characteristic stem-loop structures [115]. The nuclear microprocessor protein complex recognizes this hairpin structure and its RNase III-like enzyme component Drosha cleaves the pri-miRNA to produce a precursor miRNA (pre-miRNA), which usually has a size of 60-70 nt and possesses a characteristic 3' overhang of 2 nt [116,117]. The RAN-GTPase Exportin 5 transports the pre-miRNA from the nucleus to the cytoplasm [118] where a second RNase III enzyme, the endonuclease Dicer, cleaves off the terminal loop-structure and liberates a partially complementary siRNA-like duplex with a size of approximately 22 nt [119]. Often, one of the two arms of this duplex, the

miRNA*, is degraded and plays, if at all, only a minor role in regulation of gene expression [120]. The other strand, the mature miRNA, becomes stably incorporated into the RNA-induced silencing complex (RISC) [121]. The function of a miRNA inside of this multiprotein-complex can be regarded as a guide, which leads RISC to the complementary mRNA target site, which is often located in the 3'UTR [114].

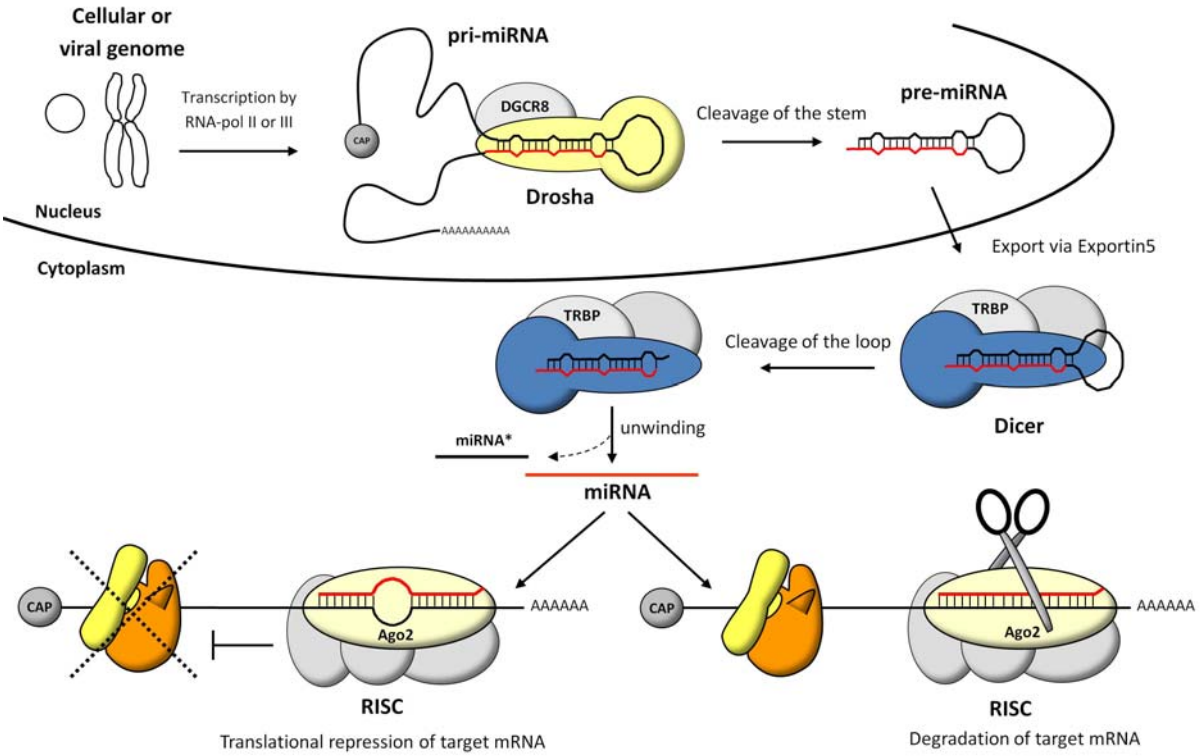

Figure 4: Canonical biogenesis pathway and functional principle of miRNAs
miRNAs are encoded in the genomes of eukaryotes as well as some viruses and are transcribed into a pri-miRNA by RNA-pol II or in rare cases by RNA-pol III. The pri-miRNA folds into a stem-loop structure and undergoes endonucleolytic cleavage of the stem portion by the RNase III-enzyme Drosha. The resulting pre-miRNA is exported into the cytoplasm by Exportin 5 and further processed by the RNase III Dicer which removes the loop region. After unwinding, two single stranded miRNAs of approximately 22 nt are released. Often only one of the two miRNAs becomes stably incorporated into the multiprotein complex RISC. Perfect base-pairing between the miRNA and the target mRNA can lead to Ago2-dependent degradation of the mRNA (right panel) whereas imperfect binding of the miRNA causes translational repression and accelerated turnover of the target mRNA (left panel). (Figure: Adam Grundhoff, modified).

In plants, miRNAs frequently bind with perfect base-pairing to target mRNAs resulting in endonucleolytic cleavage of bound mRNAs by the RISC component Argonaut protein 2 (Ago2) [122]. The cleaved RNA is unstable and subsequently degraded. In animals, miRNA mediated cleavage of mRNAs is a rare event. Usually, vertebrate miRNAs bind with only partial base-pairing to their target sequence [123]. Thereby, the so-called seed-sequence, which comprises nucleotides 2-8 of the mature miRNA, is most important for target recognition [124]. The

imperfect binding of RISC-associated miRNAs leads to translational repression followed by enhanced turnover of the RISC-bound transcripts [125] (Figure 4).
Generally the influence of miRNAs on the expression of mammalian target genes is mild and thus miRNAs are often described as fine-tuners of gene expression [111]. Despite their moderate effects and the observation that numerous miRNAs are not essential for viability [126], evidence exists that most mRNAs are regulated by miRNAs [127].

1.6 Viral miRNAs

In 2004, the first viral miRNAs were discovered in the genome of the human herpesvirus Epstein-Barr Virus (EBV) [128]. To date the number of known viral miRNAs has risen to more than 490 [129]. Most viral miRNAs have been identified and characterized in herpesviruses, accounting for more than 95% of all viral miRNAs [130]. Herpesviruses establish a life-long latency in their hosts during which they express only a very limited number of genes [131]. Interestingly, most herpesvirus miRNAs are expressed during latency, presumably to modulate the cellular or viral gene expression without production of potential immunogenic viral proteins [130,132].
Besides herpesviruses, members of the *Polyomaviridae* [104], *Adenoviridae* [133], *Ascoviridae* [134] and *Baculoviridae* [135] have been shown to produce miRNAs. Since miRNAs were predominantly found in DNA viruses, it was assumed that a DNA genome and nuclear replication were the prerequisite for the presence of virus encoded miRNAs. However, more recently, miRNAs were identified in the retroviruses Bovine Foamy Virus (BFV) and Bovine Leukemia Virus (BLV), among them a miRNA which mimics a B-cell oncomiR and likely contributes to the development of BLV-associated tumors [136,137]. miRNAs were also reported to be produced by HIV but their existence is still controversial [111,138].
miRNAs seem to be perfect tools for viruses since they require little coding capacity, are not immunogenic, can be produced by the conserved cellular miRNA biogenesis machinery and can modulate viral as well as cellular gene expression. Nevertheless, a substantial number of viruses seem not to encode miRNAs. Even some DNA viruses with a nuclear and persistent life cycle, e.g. papillomaviruses [139] thus far have not been shown to express any miRNAs. Moreover, it remains enigmatic why some members of a virus family encode for miRNAs while others do not. For example, the neurotropic alpha-herpesvirus Varizella Zoster Virus (VZV) appears to lack miRNAs whereas the likewise neurotropic alpha-herpesvirus Herpes Simplex Virus (HSV-1) readily expresses miRNAs [140].
Strikingly, all identified miRNA expressing viruses so far share the ability to establish latent or persistent infections, emphasizing a role for miRNAs in the establishment or maintenance of an

asymptomatic and non-immunogenic infection. However, concrete functions and targets have been assigned only to the minority of viral miRNAs. Identification of miRNA targets is generally a challenging task due to the low amount of required sequence identity between a miRNA and its target. Only 6-7 nt at the miRNA 5' end, the seed sequence, have to be complementary to the target mRNA to cause translational inhibition [124,141]. Given the vast amount of cellular and viral 3'UTRs which are potential targets for viral miRNAs, most viral miRNAs could potentially regulate hundreds of targets. Thus, diverse strategies are currently employed to elucidate the roles of viral miRNAs.

A fraction of virus-encoded miRNAs share seeds with host miRNAs and were found to mimic their analogous miRNAs. An example is the KSHV encoded miRNA miR-K12-11 which mimics the human miR-155 to modulate B-cell development during establishment of viral latency, but which may also play a role in KSHV-mediated transformation [140,142].

In numerous studies cDNA expression profiling (e.g. microarray analysis), and RISC pulldown strategies (e.g. PARCLIP), have been employed to identify target mRNAs on a transcriptome wide scale (reviewed in [130,143]). Especially herpesvirus encoded miRNAs were analyzed by these techniques, revealing diverse miRNA functions including the regulation of apoptosis, differentiation and cell cycle progression as well as inhibition of caspase activity (reviewed in [130]). Using *in silico* predictions of miRNA targets and functional screens, the herpesviruses HCMV, EBV and KSHV were found to downregulate MICB expression, a stress induced ligand of natural killer (NK-) cells via non-orthologous miRNAs [144]. Thus, although miRNA sequences are rarely conserved, they can have evolved to target the same pathways.

Some viral miRNAs lie antisense of a coding gene and therefore can bind mRNAs originating from the opposite strand with perfect complementarity, often resulting in degradation of the transcript. The EBV miRNA miR-BART2 which lies encoded on the opposite strand to the BALF5 polymerase transcript is an example for miRNA regulation of antisense transcripts [128]. Another case of autoregulation are the miRNAs of polyomaviruses, all of which are found in antisense orientation to the early gene region and thus can regulate the expression of T-Antigens as described in the following chapter (see references in Table 1).

1.7 Polyomavirus encoded miRNAs

Before the discovery of the first miRNA, the polyomavirus SV40 was shown to produce a short RNA molecule that was associated with early transcripts, hence named SV40-associated small (SAS) RNA [38,145]. After the discovery of viral miRNAs, the SV40 SAS-RNA was described as the first polyomavirus pre-miRNA [52] and additional polyomavirus miRNAs were subsequently identified in nine mammalian polyomaviruses as well as in the polyomavirus–papillomavirus

hybrid Bandicoot Papillomatosis Carcinomatosis Virus (BPCV) (see summary of polyomavirus miRNAs in Table 1).

Table 1: List of polyomavirus encoded miRNAs and putative miRNA targets

Virus	miRNAs	Genomic location	miRNA targets		References
			Viral mRNAs	Cellular mRNAs	
SV40	sv40-miR-S1-5p	3′ end of LT-Ag coding sequence	LT-Ag, sT-Ag	DMWD[2], C20orf27[2]	[52,146]
	sv40-miR-S1-3p				
SA12 (Simian Agent 12)	sa12-miR-S1-5p		LT-Ag[2], sT-Ag[2]		[104]
	sa12-miR-S1-3p				
JCPyV	jcv-miR-J1-5p		LT-Ag, sT-Ag[2]	ULBP3 (jcv-miR-J1-3p)	[49,147]
	jcv-miR-J1-3p				
BKPyV	bkv-miR-B1-5p		LT-Ag, sT-Ag[2]	ULBP3 (bkv-miR-B1-3p)	[49,147]
	bkv-miR-B1-3p				
MuPyV	mpv-mir-M1-5p	5′ end of the second exon of LT-Ag	LT-Ag, MT-Ag, sT-Ag		[53]
	mpv-mir-M1-3p				
MCPyV	mcv-miR-M1-5p		LT-Ag[1]	AMBRA1[2], RBM9[2], MECP2[2], PIK3CD[2], PSME3[2], RUNX1[2]	[51,148]
	mcv-miR-M1-3p				
RacPyV	RacPyV 5p		LT-Ag[1]		[50]
	RacPyV 3p				
PtvPyV2a	PtvPyV2a 5p		LT-Ag[1]		[50]
	PtvPyV2a 3p				
GggPyV1	GggPyV1 5p		LT-Ag[1]		[50]
	GggPyV1 3p				
BPCV1	Bpcv-mir-B1-3p	between the 3′ ends of the T-Antigens and L1/L2	T-Antigens[1]		[149]
BPCV2	Bpcv-mir-B2-3p		T-Antigens[1]		[149]

[1] experimental evidence using luciferase reporter constructs
[2] predicted target without experimental evidence

Although all polyomavirus miRNAs are generally antisense to the early coding region, they can be divided into two groups with respect to their precise genomic location. The closely related

polyomaviruses JCPyV, BKPyV, SV40, and SA12 encode a single pre-miRNA downstream of the late polyadenylation (pA) site, antisense to the 3' terminal region of the early coding region [52,104,147]. JCPyV, BKPyV and SA12 share an identical 3p miRNA but have different 5p miRNAs [147]. Despite being close relatives (Figure 1), the 5p and 3p mature miRNAs of SV40 only have 50% and 77% identity to the JCPyV and BKPyV sequences, respectively [147].

The pre-miRNAs of MCPyV, MuPyV, RacPyV, and the primate polyomaviruses chimpanzee and gorilla polyomavirus (PtvPyV2A and GggPyV1) are encoded in the 5' terminal region of the second exon of LT-mRNA. These miRNAs do not share sequence homology with the miRNAs of JCPyV, BKPyV, SV40, and SA12 but extensive identity exists between the PtvPyV, GggPyV, and MCPyV miRNAs including shared seed sequences [50]. In contrast, the RacPyV and MuPyV miRNAs possess only partial identity to the MCPyV miRNA and do not share a common seed sequence [50,129].

Irrespective of their genomic location and degree of sequence conservation, all polyomavirus encoded miRNAs seem to share the ability to downregulate early viral gene expression [50-53,104,147,149]. Additionally, BPCV, an usual DNA virus with characteristics of the *Polyomaviridae* and the *Papillomaviridae*, has been shown to express miRNAs which are not encoded antisense to the T-Antigen locus. The BPCV miRNAs are located in a non-coding region (NCR) but still mediate downregulation of the expression of T-Antigens by targeting the 3'UTR of T-Antigen transcripts [149].

Although polymorphisms were identified for some polyomavirus miRNAs [147], which partly influence the host targetome [150], the ability to downregulate T-Antigen expression is conserved among all polyomavirus miRNAs. Taken these observations together, in all polyomaviruses that encode miRNAs, downregulation of T-Antigen expression seems to be a conserved function and hence would be expected to play an important role in the viral life cycle, although it remains elusive why some polyomavirus require this mechanism while others don't.

In the current model of the role of polyomavirus miRNAs, expression of miRNAs occurs late in infection and is driven by transcripts from the late viral promoter, which are inefficiently terminated by inefficient late polyadenylation [52,53,147] and function as a pri-miRNA to reduce T-Antigens levels. A reduction of T-Antigen expression might be necessary to avoid the detection of infected cells by cytotoxic T lymphocytes (CTL) to allow establishment of viral persistence [52]. Thus far, however, experimental infections of mice and hamsters with miRNA mutants of SV40 and MuPyV did not show a major influence of the miRNA on virus clearance or production of infectious progeny [39,53].

Nevertheless, further studies on the BKPyV and SV40 miRNAs suggest that polyomavirus miRNAs, by regulating LT-Ag expression, play an indirect role in regulation of virus replication [39,151]. It remains to be investigated if this phenotype is observed for other polyomavirus

miRNAs as well and if so, whether it plays a role in the viral life cycle. The proposed role of polyomavirus miRNAs during viral persistence is challenged by the assumption that miRNA expression is controlled by the late promoter. Most likely, late capsid proteins are expressed, if at all, to very low levels in persistently infected cells, which would be expected to lead to a likewise low expression of the viral miRNAs. However, the mode of expression of polyomavirus miRNAs was never investigated in detail and since BPCV was shown to have an independent miRNA promoter [149], the same might be true for other viral miRNAs.

In addition to downregulation of early gene expression, polyomavirus miRNAs might influence host cell gene expression. The identical 3p miRNA of JCPyV and BKPyV, similar to the herpesviruses EBV, KSHV and HCMV, was shown to downregulate the expression of a stress induced ligand (ULBP3), presumably to avoid eradication of infected cells by NK-cells [49]. Moreover, the 5p miRNA of SV40 was shown to be a seed homolog of the human miRNA hsa-miR423-5p and thus these miRNAs might share common targets [146]. However, no further experimental evidence exists for cellular targets of polyomavirus miRNAs.

The MCPyV encoded miRNA mcv-miR-M1 was discovered in 2010 by *in silico* prediction and overexpression of stem-loop structures [51]. It was shown to be expressed at low abundance in MCPyV positive MCC tissue [50,148] but it remains to be investigated if it is expressed in MCPyV infected cells. Notably, a seed variant of mcv-miR-M1 was described to be produced in MCC tumor cells which would retain its ability to regulate early gene expression but target a different set of host mRNAs (candidates are listed in Table 1), which are hypothesized to play a role in tumorigenesis [148].

2 Aim of the Study

The discovery of miRNAs a decade ago has added a new level to our understanding of gene regulation and revealed that profound cellular pathways, many of clinical importance, are influenced by miRNA-mediated posttranscriptional gene regulation. In only a few of years, a vast number of miRNA genes was identified in the genomes of eukaryotes and, strikingly, also in several viruses. Although great efforts have been made to identify cellular as well as viral miRNA targets, the knowledge about miRNA functions significantly lags behind the discovery of new miRNAs.

Viral miRNAs have been shown to regulate cellular as well as viral gene expression and thereby influence diverse processes e.g. apoptosis, immune response and differentiation [111,130]. Among those viruses that encode miRNAs, a striking number establish life-long infections, suggesting that viral miRNAs have evolved to support the establishment and maintenance of viral persistence. Several members of the *Polymaviridae*, a family of small DNA viruses, which establish persistent and mainly asymptomatic infections in vertebrates, express miRNAs. All polyomavirus miRNAs share the potential to downregulate viral early gene expression [50-53,104,147]. However, the biological role of polyomavirus miRNAs remains elusive.

The human Merkel cell polyomavirus is the causative agent of the majority of Merkel cell carcinoma cases, a rare but highly aggressive human skin cancer. The MCPyV miRNA mcv-miR-M1, like all other polyomavirus miRNAs, has the ability to regulate the expression of early mRNAs. The recently developed MCPyV replication system provides the opportunity to study mcv-miR-M1 during authentic viral replication in cell culture [72]. The aim of this study was to utilize this model system to unravel the role of mcv-miR-M1 in viral replication and gene expression, and to understand how this miRNA, and potentially other polyomavirus miRNAs, influence the viral life cycle. To this end, MCPyV mutant viruses were to be characterized in terms of replication, gene expression and epigenetic modifications by using High Throughput Sequencing technology. Additionally, to allow conclusions about putative functions at different stages of the viral life cycle, the regulation of miRNA expression was to be investigated. A detailed understanding of the function of mcv-miR-M1 in MCPyV replication is required to gain a comprehensive picture of the viral life cycle, including the mechanisms that lead to viral persistence and tumorigenesis of MCC. Moreover, studying mcv-miR-M1 might help to understand the principles of polyomavirus miRNA-mediated gene regulation in general, since their potential to regulate early viral gene expression is highly conserved. Finally, detailed insights into the functions of mcv-miR-M1 and the regulation of mcv-miR-M1 expression are expected to add significantly to our current understanding of viral miRNAs and their implications in persistent infections.

3 Material and Methods

3.1 Material

3.1.1 Chemicals, commercial systems and expendable materials

If not otherwise stated, all chemicals and expendable materials were purchased from the following companies: 5 Prime, Applied Biosystems, BD Biosciences, Bio Rad, Biozym, Eppendorf, GE Healthcare, Gibco, Gilson, Greiner Bio One, Kimberly Clark, Millipore, NEB, Nunc, Hartenstein, Heraeus, Invitrogen, Kodak, Lonza, LTF Labortechnik GmbH, Merck, Mettler, New Brunswick, New England Biolabs, PAA, PeqLab, Promega, Qiagen, Roche, Roth, Santa Cruz, Sarstedt, Schleicher & Schuell, Schott, Sigma, Stratagene, Thermo Scientific, Whatman and Zymo.

3.1.2 Instruments and equipment

Instruments and equipment from the following manufacturers were used in this study: Agilent, BD Biosciences, BioTek, Eppendorf, Fujifilm, Hartenstein, Heraeus, IKA, Illumina, Kendro, Kodak, Leica, Nikon, Raytest, Sarstedt, Sorvall, Stratagene and Qiagen.

3.1.3 Software and online resources

For data analysis and visualization, the following tools were used: Adobe Photoshop CS5.1 (Adobe Systems Inc.), Advanced Image Data Analyzer (AIDA, Raytest), Bedtools [152], blastp NCBI, Bowtie [153], CEAS 1.0.2 [154,155], CLC Main Workbench 7 (Qiagen), CLC Genomics Workbench 6 (Qiagen), DAVID Bioinformatics Resources 6.7 [156,157], ExPASy [145], FACSDiva (BD), Figtree Tree Figure Drawing Tool Version 1.4.2. [39], FIMO 4.1 [158], GeneSpring GX (Agilent), IGV [159], JASPAR 5.0_ALPHA [134], MACS 1.4.2 [160], MUSCLE 3.7 [37], Gel Doc XR (Bio Rad), MEME 4.1 [161], Multigauge (fujifilm), PhyML 3.0 [38], Prism 5 (GraphPad Software), RNAhybrid 2.2 [162], Rotor Gene Q Software (Qiagen), Tophat [124], Tophat 2 [163].

3.1.4 Plasmids

Plasmids used and generated in this study are summarized in Table 2.

Table 2: Recombinant plasmids

Plasmid name	Backbone	Insert
pMCVSyn	pMK (GeneART)	MCPyV consensus sequence [72]
pMCVSyn-ltb	pMK (GeneART)	MCPyV consensus sequence miRNA LTB mutations
pMCVSyn-pmt	pMK (GeneART)	MCPyV consensus sequence miRNA promoter mutant
pMCVSyn-hpko	pMK (GeneART)	MCPyV consensus sequence miRNA hairpin ko mutant
pMCVSyn-seed	pMK (GeneART)	MCPyV consensus sequence miRNA seed mutant
pMCVSyn-VP1-GFP	pMK (GeneART)	2A-GFP
pMCVSyn-LT-GFP	pMK (GeneART)	2A-GFP
pMCVSyn-VP1-GFP-hpko	pMK (GeneART)	2A-GFP
pMCVSyn-VP1-GFP-seed	pMK (GeneART)	2A-GFP
pUC18	pUC18 (LifeTechnologies)	no insert
LeGO G mcv-miR-M1	LeGO G [164]	mcv-miR-M1
LeGO G miR-K12-11	LeGO G [164]	miR-K12-11
LeGO G Ba2	LeGO G [164]	miR Bart2
LeGO G miR-K12-7	LeGO G [164]	miR K7
pHCMV VSV env R861	phCMV	VSV env
Lenti gag-pol	phCMV	gag-pol
Lenti rev	phCMV	rev
pCDNA3.1 GFP	pCDNA3.1(+) (LifeTechnologies)	GFP
pCDNA3.1 GFP mcv-miR-M1 300	pCDNA3.1(+) (LifeTechnologies)	mcv-miR-M1 300 nt fragment nt 1350-1051
pCDNA3.1 SV40 LT	pCDNA3.1(+) (LifeTechnologies)	SV40 early coding region
pER-S	pCR2.1 (LifeTechnologies)	MCPyV full length early region from MCVSyn
pCMV	pCMV-tag2B (Stratagene)	no insert
pCMV:ER-S	pCMV-tag2B (Stratagene)	MCPyV full length early region from MCVSyn
pCMV:ER-AS	pCMV-tag2B (Stratagene)	MCPyV full length early region from MCVSyn antisense
pCMV:ER-trunc	pCMV-tag2B (Stratagene)	MCPyV truncated early region from MCCL-12

3.1.5 Oligonucleotides

Oligonucleotides were purchased from Invitrogen and Sigma. The sequences of primers and probes used in this study are listed in Table 3-7.

Table 3: Primers used for qPCR and RT-PCR

Name	Sequence	Application
MCPyV VP1 fw	AAAACACCCAAAAGGCAATG	qPCR Replication Assay, transcripts, ChIP
MCPyV VP1 rev	GCAGAGACACTCTTGCCACA	qPCR Replication Assay, transcripts, ChIP
MCPyV LT BSP fw	TGATGAGGTTGACGAGGCCCCTA	qPCR transcripts
MCPyV Dpn R	AGGTATATCGGGTCCTCTG	qPCR transcripts
SL mcv-miR-M1	GTCGTATCCAGTGCAGGGTCCGAGG TATTCGCACTGGATACGACTGTACC	cDNA synthesis mcv-miR-M1
mcv-miR-M1 fw	GCATCTGGAAGAATTTCTA	qPCR mcv-miR-M1
Universal rev	GTGCAGGGTCCGAGGT	qPCR mcv-miR-M1
GAPDH fw DNA	TGTGTCCCTCAATATGGTCCTGTC	qPCR Replication Assay
GAPDH BSP fw	GGTCGGAGTCAACGGATTTG	qPCR mcv-miR-M1, transcripts
GAPDH RT rev	ATGGTGGTGAAGACGCCAGT	qPCR Replication Assay, mcv-miR-M1, transcripts
tRNA-met_fw	AGCAGAGTGGCGCAGCGG	qPCR transcripts
tRNA-met_rev	TAGCAGAGGATGGTTTCGATC	qPCR transcripts
p21 F	CTTGTACCCTTGTGCCTCGCT	qPCR transcripts, [165]
p21 R	CGGATTAGGGCTTCCTCTTGG	qPCR transcripts, [165]
GADD45 F	TGCGTGCTGGTGACGAATCC	qPCR transcripts, [165]
GADD45 R	CAGATGCCATCACCGTTCAGG	qPCR transcripts, [165]
MDM2 F	GTGTATCAGGCAGGGGAGAGTG	qPCR transcripts, [165]
MDM2 R	CTTCAGGAAGCCAATTCTCACG	qPCR transcripts, [165]
GAPDH F	GTGAAGGTCGGAGTCAACGGA	qPCR transcripts, [165]
GAPDH R	CCATGGGTGGAATCATATTGGAAC	qPCR transcripts, [165]
ADH5 fw	GCATAATTGAGCCTACGCC	qPCR ChIP
ADH5 rev	GCAGAGGTGTTTGTTACGTG	qPCR ChIP
MCPyV RT ori fw	CCGTCTCCCTCCCAAACAGA	qPCR ChIP
MCPyV RT ori rev	AGCTACCTCACTAAGGAGTGGT	qPCR ChIP
d5335-a861 BSP fw	TTAGTGAGGTTGACGAGGCC	RT-PCR
d5335-a861 rev	AGGTATATCGGGTCCTCTG	RT-PCR
d141-a861 BSP fw	CTTGGCTGCCTAGGTTGAC	RT-PCR
d141-a861 rev	GACGCTGAGAAGGACCCATA	RT-PCR
d1142-a5308 BSP fw	AGAGGATGAGGGTTTCTGGC	RT-PCR
d1142-a5308 rev	TCCTGTGGTGGCACTTAGTT	RT-PCR

Name	Sequence	Application
d5145-a5308 fw	CTTTCTGTTTGGGAGGGAG	RT-PCR
d5145-a5308 BSP rev	CCAGAAACCTCTTTTAATGTC	RT-PCR

Table 4: Primers used for 3'RACE

Name	Sequence 5'-3'
3'RACE QT [166]	CCAGTGAGCAGAGTGACGAGGACTCGAGCTCAAGCTTTTTTTTTTTTTTTTVN
3'RACE Qo [166]	CCAGTGAGCAGAGTGACG
3'RACE Qi [166]	GAGGACTCGAGCTCAAGC
3'RACE LTo	GGTGCTTGCCTGATACAACC
3'RACE LTi	GTCTAGAGGATCCTTGCTTAC
3'RACE VP1o	GGGGTCAGAACAATTACCTGG
3'RACE VP1i	GTCTAGAGGATCCTGATATTG

Table 5: Primers used for 5'RACE

Name	Sequence 5'-3'
5'RACE RNA Adapter	GUUCAGAGUUCUACAGUCCGA
5'RACE Adapter Primer	GTTCAGAGTTCTACAGTCCGA
5'RACE Adapter Primer BamHI	AGGATCCGTTCAGAGTTCTACAGTC
5' RACE late region rev	AGTGACACTTGCTCGCGTGACAACC
5' RACE late rev nested XhoI	ACTCGAGCCACAGTTATTAGAGAGC
5' RACE early region rev	ACGCTGAGAAGGACCCATACCCAGA
5' RACE early rev nested XhoI	ACTCGAGAAGGCTTTCTGGATCTTGAG

Table 6: Primers used for cloning of MCVSyn-GFP constructs

Name	Sequence 5'-3'
Oligo_FMDV_2A+5aa sense	GCTAGCGCACCGGTGAAACAGACTTTGAATTTTGACCTTCTCAAGTTGGCAG GAGACGTTGAGTCCAACCCTGGGCCC
Oligo_FMDV_2A+5aa antisense	GGGCCCAGGGTTGGACTCAACGTCTCCTGCCAACTTGAGAAGGTCAAAATTC AAAGTCTGTTTCACCGGTGCGCTAGC
LT NdeI fw	AGTGAAATATCATATGGTA
LT NheI rev	TGGCTAGCTTGAGAAAAAGTA
VP1 NdeI rev	ATTACCATATGTTTGCCATT
5p 2A 3p GFP	AACCCTGGGCCCATGGTGAGCAA
5p GFP 3p 2A	TTGCTCACCATGGGCCCAGGGTT
2A fw	GCTAGCGCACCGGTGAAACAGA
GFP rev	GGATCCTTACTTGTACAGCT
VP1 BamHI fw	GGATCCCATCTGCTTCATAT
VP1 NheI fw	AGCTAGCTAATTCTTGTGTT

Name	Sequence 5'-3'
LT BamHI rev	GGATCCGCATTCCTATTCAGT

Table 7: Oligonucleotide probes and primers for probe generation

Name	Sequence 5'-3'	Application
mcv-miR-M1 probe	AGTGTACCTAGAAATTCTTCCAGAACGTATGGCAC	small RNA Northern blot probe
Taqman probe mcv-miR-M1	VIC™-CTGGATACGACAGTGTACC-MGB	Taqman probe for qPCR
Taqman probe GAPDH	ROX™-GTGGCGCTGAGTACGTCGTGGAGTC-MGB	Taqman probe for qPCR
MCPyV-LT-s	ATGGATTTAGTCCTAAATAGGAAAG	Generation of MCVSyn Southern blot probes by PCR
MCPyV-LT-as	CTCATCAAACATAGAGAAGTCAC	
SV40-LT-s	GCAGTGCAGCTTTTTCCTTTG	Generation of SV40 Southern blot probes by PCR
SV40-LT-as	ATGGATAAAGTTTTAAACAGAG	

3.1.6 Antibodies

Antibodies used in this study are listed in Table 8-9.

Table 8: Primary antibodies

Name	Description	Application	Company
CM2B4	Mouse monoclonal antibody against LT/57kT exon 2	WB, IFA, ChIP	Santa Cruz
hMICB PE-conjugated antibody	Mouse monoclonal antibody against human MICB, PE conjugated	Flow cytometry	R&D Systems
α-H3K4-me3	rabbit monoclonal antibody against H3K4-me4, clone MC315	ChIP	Merck Millipore
normal mouse IgG	mouse polyclonal antibody control	ChIP	Merck Millipore
normal rabbit IgG	rabbit polyclonal antibody control	ChIP	Merck Millipore

Table 9: Secondary antibodies

Name	Application	Company
α-mouse HRP	WB	Santa Cruz
α-mouse Texas Red	IFA	Santa Cruz
sheep-anti-Digoxigenin-FITC-antibody	FISH	Roche

3.2 Methods of eukaryotic cell culture and cell biology

The eukaryotic cell lines listed in Table 10 were used in this study.

Table 10: Human cell lines

Name	Characteristics	Reference
PFSK-1	Adherent, neuroectodermal tumor cell line	Fults et al. 1992 [160]
HEK293	Adherent, embryonic kidney cell line transformed by AdV5	Graham et al. 1977 [167]
HEK293T	Adherent, HEK293 with constitutive expression of SV40 large T-Antigen	DuBrigde et al. 1987 [168]
RKO	Adherent, colon carcinoma cell line	Boyd et al. 1988 [169]
MKL-1	Suspension, MCPyV positive Merkel cell carcinoma cell line	Rosen et al. 1987 [170]
WaGa	Suspension, MCPyV positive Merkel cell carcinoma cell line	Houben et al. 2010 [97]

3.2.1 Culture of eukaryotic cell lines

Generally, all cells were cultured in polystyrene cell culture flasks or dishes at 37°C in a 5% CO_2 atmosphere with a relative humidity of 95%. The cell lines PFSK-1, H1299, MKL-1 and WaGa were maintained in RPMI medium (Gibco). HEK293, HEK293T and RKO cells were grown in DMEM medium (Gibco). All medium was supplemented with 10% fetal calf serum (FCS, PAA), 1% Penicillin/Streptomycin (PAA) and 1% sodium pyruvate (PAA). Cells were split when they reached a density of 90% in a ratio of 1:10 or 1:5 in the case of PFSK-1 cells. To split or harvest adherent cells, culture medium was aspirated and residual medium was removed by washing with PBS (Sigma). Cells were incubated with Trypsin/EDTA (Gibco) at 37°C for 1-3 min until they detached from the bottom of the cell culture vessel. To achieve a homogenous cell suspension and to detach cells from each other, trypsinized cells were re-suspended thoroughly in their appropriate growth medium and 1/10 to 1/5 of cells were reseeded. Remaining cells were either pelleted by centrifugation (2000 x g, 3 min, 4°C) for the extraction of nucleic acids and protein or else discarded. The non-adherent cell lines MKL-1 and WaGa were centrifuged (2000 x g, 3 min, RT) and re-suspended in RPMI medium once a week and fed with fresh medium when necessary.

3.2.2 Storage of eukaryotic cells

For long-term storage, sub confluent cells were pelleted by centrifugation (2000 x g, 3 min, RT) and the culture medium was removed. Cells were re-suspended in FCS with 10%

dimethylsulfoxide (DMSO, Sigma) and aliquoted in CryoPure tubes. To avoid the formation of ice crystals, cells were slowly frozen to -80°C in an isopropanol bath. After 24 h cells were transferred to liquid nitrogen.

To initiate a culture from frozen cells, an aliquot was thawed in a 37°C water bath. Cells were washed once with culture medium to remove DMSO and were subsequently re-suspended in culture medium and seeded in an appropriate cell culture dish.

3.2.3 Transient transfection

For transient introduction of DNA in eukaryotic cells, different transfection reagents were used, depending on the cell type. For all transfections, cells were seeded one day prior to transfection to reach a density of 60-80%. For seeding, cells were trypsinized and re-suspended thoroughly in cell culture medium. 1 volume of trypan blue was added to an aliquot of trypsinized cells for dead cell exclusion. After 5 min of incubation at RT, cells were counted under the light microscope (Leica) using a Neubauer hemocytometer. Cells in all 4 squares of one chamber were counted and the average number per square was multiplied with the factor 10^4 to obtain the cell number per mL. The number of cells that were seeded in different cell culture vessels is shown in Table 11.

Table 11: Number of seeded cells for transfection in different cell culture vessels

Cell culture vessel	Number of cells
96-well plate	$1x10^4$/well
24-well plate	$5x10^4$/well
6-well plate	$2\text{-}5x10^5$/well
10 cm dish	$5x10^6$/dish

3.2.3.1 Transfection with X-tremeGENE HP

PFSK-1 cells were transfected using the liposomal reagent X-tremeGENE HP (Roche). The components for transfection of one well of a 6-well plate or a 10 cm dish are shown in Table 12 and were mixed in a 1.5 mL tube the indicated order. To allow the formation of complexes, the mixture was incubated for 30 min at RT. Meanwhile, the growth medium was replaced with fresh RPMI containing all supplements with the volume indicated in Table 12. The transfection mixture was added dropwise to the cells and the plate was gently swirled. 48 h after transfection, medium was changed or cells were harvested for analysis.

Table 12: Transfection mixture for X-tremeGENE transfection

Plate	OptiMEM	DNA	X-tremeGENE	Amount of RPMI on cells
6-well plate	100 µL/well	1-2 µg/well	2 µL/well	1 mL/well
10 cm dish	1000 µL/dish	5-10 µg/dish	10 µL/dish	6 mL/dish

3.2.3.2 Transfection with Polyethyleneimin (PEI)

HEK293 and HEK293T cells were transfected with Polyethyleneimine (PEI, Polysciences, Inc.), a basic and branched polymer [146]. PEI was dissolved in sterile H_2O at a concentration of 1 mg/mL. The solution was neutralized to pH 7.2 by addition of HCl and sterilized by filtration (0.22 µm). Aliquots of PEI working solution were stored at -80°C.

Prior to transfection, the growth medium was removed from the cells and replaced with the indicated volumes of serum-free OptiMEM. The composition of the transfection mixture is shown in Table 13.

Table 13: Transfection mixture for PEI transfection

Plate	OptiMEM	DNA	PEI	Amount of OptiMEM on cells
96-well plate	25 µL	0.2 µg	2 µL	75 µL
6-well plate	100 µL	1-2 µg	20 µL	1 mL
10 cmdish	1000 µL	5-10 µg	100 µL	6 mL

First, DNA and OptiMEM were assembled in a 1.5 mL tube. PEI was added and the mixture was vortexed. After incubation at RT for 30 min, the transfection mixture was added dropwise to the cells. Medium was replaced 8 h after transfection with growth medium containing all supplements.

3.2.4 Production of lentiviral particles

The colon carcinoma cell line RKO was the only cell line tested which showed constitutive expression of MICB and was therefore used to investigate the influence of viral miRNAs on MICB levels. Since RKO cells cannot be transfected with the above described methods, RKO cells were transduced with lentiviral vectors (LeGO G [164]) for the expression of mcv-miR-M1, miR-K12-11, miR-K12-7 and miR-BART2.

Lentiviral particles were produced by co-transfection of the LeGO G miRNA expression plasmids with packaging plasmids into HEK293T cells. Table 14 shows the amounts of LeGO plasmid and packaging plasmids used for the transfection of a 10 cm dish. The vector DNA was added to 1 mL OptiMEM medium. 150 µL of PEI were added and the solution was mixed well by vortexing.

DNA-PEI complexes were allowed to form for 30 min at RT while the growth medium was replaced with 6 mL OptiMEM per 10 cm dish. The transfection mixture was added dropwise to the cells and the dishes were carefully swirled. After 8 h, the medium was removed and replaced with 7 mL of DMEM with all supplements.

Table 14: Amounts of transfected plasmid DNA per 10 cm dish for lentivirus production

Plasmid	Amount per 10 cm-dish
LeGO plasmid	10 µg
gag-pol	10 µg
rev	5 µg
VSV env	2 µg

24 h after transfection, the first virus containing supernatant was harvested, sterile filtered, aliquotted and stored at -80°C. 7 mL of fresh growth medium was added to the cells and another 24 h later, the second supernatant was harvested as described.

3.2.5 Transduction with lentiviral particles

To obtain comparable transduction efficiencies, the titer of lentivirus containing supernatants was determined.

RKO cells were seeded in 24-well plates (5 x 10^4 cells per well). After attachment of cells, growth medium was removed and replaced with 500 µL total volume of different ratios of viral supernatant and growth medium. 8 µg/mL polybrene were added to facilitate virus attachment. Plates were spinoculated for 1 h at 37°C and 800 g and incubated for 48 h under standard culture conditions. Growth medium was replaced after 24 h.

Transduction efficiency was evaluated by FACS analysis of GFP positive cells as described in chapter 3.2.8. The titer of individual supernatants was determined with the formula:

$$T = N \times P / V$$

T = titer in ffu/µL (fluorescence forming units)

N = number of cells

P = percent of transduced cells

V = volume of viral supernatant

For MICB expression assays, RKO cells were transduced with a multiplicity of infection (MOI) of 10 in the presence of 8 µg/mL polybrene as described above. FACS analysis was carried out 48 h after transduction.

3.2.6 MCVSyn replication assay

Cell culture based model systems for viruses are most important tools for the investigation of the viral life-cycle and virus-cell-interactions. For MCPyV, the development of a fully permissive model system is hampered by the lack of knowledge about the natural reservoir cell type.

With the establishment of a semi-permissive replication system using a synthetic consensus MCPyV genome (MCVSyn) and immortalized human cell lines, viral replication and gene expression can be studied [72]. Here, the MCVSyn replication assay is used to investigate the role of mcv-miR-M1 during viral replication.

The plasmid pMCVSyn contains a consensus MCPyV sequence which is identical to several field isolates [72] and a bacterial backbone for amplification in *E.coli*. MCVSyn artificial genomes are produced by excision of the bacterial backbone and re-circularization. Upon transfection into HEK293, H1299 and PFSK-1 cells, MCVSyn replicates in a small subset of cells (3-5%). Replication of viral DNA is assessed by either Southern blotting or qPCR analysis of *DpnI* digested Hirt extracts or genomic DNA as described below.

To analyze the viral transcriptome and the expression of mcv-miR-M1, total RNA from MCVSyn replication assays was harvested as described in chapter 3.4.1 and qPCR analysis or High Throughput Sequencing was performed (see chapters 3.4 and 3.8).

3.2.6.1 Re-circularization of MCVSyn plasmid DNA

MCVSyn artificial genomes were produced from the plasmid pMCVSyn which contains a bacterial backbone in the viral early coding region. To excise the bacterial backbone, 8 µg of pMCVSyn were digested with 3 µL of *SacI* Fast Digest restriction enzyme (Thermo Scientific) for 1 h at 37°C. The MCVSyn DNA was separated from the bacterial backbone by gel electrophoresis and purified with the Zymo DNA Gel Extraction kit. The linear DNA was re-ligated in a 500 µL reaction with 2 µL of T4 DNA Ligase (Thermo Scientific) for at least 8 h at 16°C.

The re-ligated DNA was purified with the DNA Clean and Concentrator kit (Zymo) by adding 1000 µL of DNA binding buffer to the religation reaction and loading the sample on one column by successive centrifugation steps. The column-bound DNA was washed once with washing buffer and then eluted in 22 µL H_2O. The religation efficiency was assessed by agarose gel electrophoresis of 2 µL of re-ligated DNA. DNA concentration was measured with the NanoDrop-

1000. Re-ligated DNA was stored in aliquots at -20°C. MCVSyn mutants and MCVSyn-GFP constructs were re-ligated likewise.

3.2.6.2 Transfection of recircularized MCVSyn and subculturing of long-term replication assays

To initiate a MCVSyn replication assay, PFSK-1 cells were seeded in 6-well plates the day before transfection. To determine the transfection efficiency by FACS analysis, one well was transfected with a GFP expression plasmid and one well was transfected with pUC18 alone (mock control). For each long-term replication assay, three wells were transfected with MCVSyn or MCVSyn mutants or with pUC18 alone (mock control).

The transfection mixture for one well was prepared as shown in Table 15. The components were added to 1.5 mL tubes in the indicated order. If more than one well was transfected with the same MCVSyn construct, a mastermix was prepared.

Table 15: Transfection mixture for MCVSyn replication assay

Components	Amount
OptiMEM	100 µL
re-ligated MCVSyn DNA	200 ng
pUC18	500 ng
X-tremeGENE	2 µL

The transfection mixture was mixed well by vortexing and incubated for 30 min at RT.

Before adding the transfection mixture to the seeded cells, the medium was replaced with 1 mL fresh growth medium. The transfection mixture was added dropwise to the cells and plates were gently swirled. 48h after transfection medium was changed on all wells and transfection efficiency was determined by FACS analysis as described in chapter 3.2.8.

For long-term replication assays, one well of transfected cells was harvested 24 h after transfection (1 d.p.t) and a second well was harvested 48 h (2 d.p.t) after transfection. The third well was trypsinized 4 d.p.t and one half of the cells was harvested whereas the remaining cells were reseeded into a new 6-well. This procedure was repeated at day 8. Subsequently, cells were passaged in 6-well plates and harvested once a week. Remaining cells were frozen for storage as described above. For harvesting, cells were washed with PBS, trypsinized and re-suspended in growth medium. Half of the cells were reseeded and the remaining cells were pelleted in aliquots for extraction of nucleic acids.

3.2.7 Re-infection with MCVSyn derived particles

For re-infection assays, PFSK-1 cells were transfected with MCVSyn or MCVSyn-hpko, respectively. Four days post transfection, cells were harvested and lysed by three freeze-thaw-cycles. Cell debris was removed by centrifugation and lysates were passed through a 0.22 µm filter. Lysates prepared from a 10 cm dish were used to infect one 6-well of freshly seeded PFSK-1 cells. 24 h post infection, medium was changed and cells were incubated for additional 3-7 days prior to gDNA preparation.

3.2.8 Fluorescence activated cell sorting (FACS)

Transfection efficiencies, transduction efficiencies and MICB expression were determined by flow cytometry. Fluorescence activated cell sorting (FACS) was used to sort GFP expressing MCVSyn-GFP transfected cells. Prior to analysis or sorting, cells were trypsinized and washed once in PBS. Cells were re-suspended in FACS buffer (Table 16) and pipetted through a cell strainer cap into round-bottom FACS tubes.

Table 16: FACS buffer

Components	Concentration
PBS	1x
FCS	1%
EDTA	5 mM
Hepes	25 mM

For analysis of MICB expression, a PE-conjugated α-hMICB antibody was used to detect MICB on the cell surface of RKO cells transduced with different LeGO G vectors for expression of viral miRNAs. 48 h after transduction, $1x10^6$ cells were harvested, washed once in PBS and re-suspended in ice cold PBS containing 10% FCS and 1% sodium azide. 5 µg/mL of PE-conjugated α-hMICB antibody were added. The cell suspension was gently mixed and incubated for 1 h at 4°C. To remove unbound antibody, cells were washed 3 times in PBS. Finally, cells were re-suspended 500 µL ice cold PBS with 10% FCS and 1% sodium azide. MICB expression was measured by PE intensity in viral miRNA expressing GFP positive cells.

3.2.9 MTT cell proliferation assay

The MTT assay is an indirect colorimetric method to estimate the number of viable cells by measuring metabolic activity. The tetrazolium dye MTT (3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide) is reduced by living cells to formazan, which accumulates as an

insoluble precipitate. In this study, the MTT assay was used to measure the proliferation of cells dependent on LT-Ag expression over a period of 7 days.

PFSK-1 cells were transfected in 6-well plates with increasing amounts of a LT-Ag expression plasmid. 24 h after transfection, cells were seeded in 96 well plates. Every 24 h, 10 µL MTT solution (Chemicon) were added per well and cells were incubated for 4 h under standard culture conditions. The cell culture supernatant was then carefully removed without disturbing the formazan crystals. Plates were then stored at -20°C until all samples were collected.

For measurement, formazan was dissolved in 200 µL DMSO and within 5 min absorbance was measured in a plate reading spectrophotometer (BioTek) at a wavelength of 540 nm and a reference of 690 nm. For all samples, 8 wells were measured.

3.2.10 FISH analysis

$5x10^4$ cells were cytospun for 5 min at 500 x g on Superfrost-Plus slides (Fisher) and fixed in methanol, followed by digestion (0.01% pepsin, 0.01 M HCl) for 5 min at 37°C and treatment with 100 µg/mL RNase A (Invitrogen) in 2x SSC for 1 h at 37°C. After washing in PBS and refixation (3% formaldehyde/PBS, 50mM $MgCl_2$) slides were passed through a dehydration series of 70%, 85%, and 100% ethanol for 5 min each and air dried. For denaturation slides were incubated in 70% formamide in 2x SSC for 5 min at 73°C and then placed in ice-cold 70% ethanol for 5 min and dehydrated again as described above.

1 µg of MCVSyn DNA was labeled with Dig-Nick Translation Mix (Roche) according to the manufacturer's instructions. Labeled DNA was ethanol precipitated in the presence of excess sonicated salmon sperm DNA (Life Technologies). The final product was re-suspended in hybridization buffer (50% formamide and 10% dextran sulfate in 2x SSC) to a final concentration of 10 ng/µL and stored at -20°C. 50 ng of the probe were heat-denatured for 5 min at 73°C and placed under a coverslip on the appropriate area of the slide. The coverslip was fixed with fixogum (Marabu). Slides were hybridized in a humid chamber overnight at 37°C.

After hybridization, slides were washed three times (2x SSC, 0.2% Tween) for 2 min each, twice at 20°C and in between at 70°C. After blocking with 4% BSA/PBS for 30 min at 37°C, slides were incubated with sheep-anti-Digoxigenin-FITC-antibody (Roche), diluted 1:50 in 4% BSA/PBS with 0.2% Tween, for 2 h at 37°C in the dark. Slides were washed with PBS/0.2% Tween three times for 10 min each at 20°C in the dark, counterstained and mounted with vectashield mounting medium with DAPI (Vector). Images were acquired with a confocal laser-scanning microscope (Nikon C2+).

3.2.11 Immunofluorescence analysis

Cells were seeded on gelatin coated coverslips and transfected with re-ligated MCVSyn as described in chapter 3.2.6. 2 d post transfection, medium was changed and cells were incubated for another 48 h. Cells were fixed for 20 min with 4% paraformaldehyde (PFA), permeabilized for 10 min and blocked for 30 min by addition of respective buffers at RT (Table 17 and Table 18). In between these steps, cells were washed twice with PBS for 10 min. For detection of LT-Ag, the antibody CM2B4 was diluted 1:1000 in blocking buffer and added to the cells for 2 h at RT. After washing the coverslips 3 times in PBS, the TRITC labeled secondary α-mouse antibody (Santa Cruz) was diluted 1:1000 and added to the cells for 2 h at RT in the dark. Residual antibody was removed by washing in PBS and coverslips were mounted with DAPI containing vectashield mounting medium (Vector). Images were acquired with a confocal laser-scanning microscope (Nikon C2+).

Table 17: Permeabilization buffer

Components	Concentration
PBS	1x
Triton X-100	1%
Tri-sodium citrate dihydrate	0.1%

Table 18: Blocking buffer

Components	Concentration
PBS	1x
Triton X-100	1%
Tween 20	0.5%
BSA	3%

3.3 DNA techniques

The *E. coli* strain DH5α was used for amplification of plasmid DNA and blue-white screening for selection of positive clones.

Table 19: Features of the *E.coli* strain DH5α

Name	Genotype	Supplier
Escherichia coli DH5α	F– Φ80*lac*ZΔM15 Δ(*lac*ZYA-*arg*F) U169 *rec*A1 *end*A1 *hsd*R17 (rK–, mK+) *pho*A *sup*E44 λ– *thi*-1 *gyr*A96 *rel*A1	Invitrogen

3.3.1 Culture and storage of bacteria

Bacteria were cultured on Lysogeny Broth (LB) agar plates at 37°C or in liquid culture in LB medium at 37°C or at 30°C (for lentiviral plasmids) in a shaking incubator. Media was sterilized by autoclaving and was supplemented with appropriate antibiotics for selection (ampicillin (1 µg/mL) or kanamycin (50 µg/mL)).

For long term storage, single colonies were picked from agar plates and grown overnight in liquid culture. Bacteria were pelleted and re-suspended in 0.5 mL of LB medium and 0.5 mL of sterile glycerin and stored at -80°C.

3.3.2 Generation of chemically competent *E. coli*

Chemically competent cells were produced using the rubidium-chloride-method. Bacteria were inoculated in LB-medium and grown over night at 37°C under agitation. The bacterial culture was diluted 1:100 in LB+ (Table 20) and continuously grown until an OD_{600} of 0.3-0.5 was reached. Bacteria were cooled on ice for 15 min and subsequently pelleted by centrifugation (300 g, 5min, 4°C). Cells were re-suspended in 150 mL cold TFB1 buffer (Table 21) and incubated on ice for 90 min. TFB1 buffer was removed after renewed pelleting. Subsequently, bacteria were re-suspended in 25 mL of cold TFB2 buffer (Table 22) and immediately frozen in 200 µL aliquots in liquid nitrogen and stored at -80°C.

Table 20: LB+ medium

Components	Concentration
LB medium	500 mL
$MgSO_4$	8 mM
KCl	10 mM

Table 21: TFB1 buffer

Components	Concentration
$RbCl_2$	100 mM
KAc	30 mM
$CaCl_2$	10 mM
$MnCl_2(x4H_2O)$	50 mM
Glycerol	15% (v/v)
Acetic Acid	to pH 5.8

Table 22: TFB2 buffer

Components	Concentration
$RbCl_2$	10 mM
MOPS	10 mM
$CaCl_2$	75 mM
Glycerol	15% (v/v)

3.3.3 Transformation of chemically competent bacteria

Competent bacteria were slowly thawed on ice. 100 µL of competent cells were gently mixed with up to 1 µg plasmid DNA or 20 µL of ligated DNA and incubated on ice for 15 min to allow attachment of the DNA to the bacterial surface. Subsequently, a heat shock was carried out at 42°C for 60 sec and cells were immediately put back on ice for 5 min. 800 µL of LB medium were added and cells were grown for 60 min at 37°C in a shaking incubator. Bacteria were pelleted (2000g, 4 min, RT) and re-suspended in 50 µL LB medium. Selection for transformed bacteria was accomplished by further culturing under selective condition either on LB agar plates or in liquid culture.

3.3.4 Preparation of plasmid DNA from *E. coli*

Plasmid DNA was isolated from transformed *E. coli* that were grown for at least 6 h under selective conditions. For analytical purposes, small amounts of plasmid DNA were extracted from 3 mL of bacterial culture using the peqGOLD Plasmid Miniprep Kit (Peqlab) according to the manufacturer`s instructions. Larger amounts of plasmid DNA for transfection or cloning were isolated with the Qiagen Plasmid Midi Kit as described by the manufacturer.
Plasmid DNA preparations were analyzed by restriction digest and agarose gel electrophoresis.

3.3.5 Quantification of nucleic acids

Concentrations of DNA and RNA samples were measured in a spectrophotometer (Nanodrop-1000, Thermo Scientific) at a wavelength of 260 nm. The purity of nucleic acid preparations was estimated by calculating the ratios of 260/280 nm and 260/230 nm.
More precise measurements of nucleic acid concentrations, e.g. for HTS libraries, were achieved by fluorometric quantification using a Qubit fluorometer (Life technologies). Depending on the type and expected concentration of nucleic acids, dsDNA BR, dsDNA HS, RNA BR or RNA HS assay kits were used as described by the manufacturer.

3.3.6 Agarose gel electrophoresis

Agarose is a polysaccharide polymer and is used as a matrix to separate DNA fragments according to their size in an electric field. Depending on the size of DNA fragments, different agarose concentrations were chosen (Table 23).

Table 23: Agarose concentration for separation of DNA fragments of different sizes

Size of DNA fragment	Agarose concentration
> 5 kb	0.8%
1 kb - 5 kb	1.0%
0.1 kb - 1 kb	1.2% - 2%

Gels were prepared by boiling the desired amount of agarose in 100 mL TAE buffer (Table 24) until agarose was completely dissolved. 3 µL of Ethidium Bromide (10 mg/mL) were added and the gel was cast into a gel tray with an appropriate comb depending on the volume of the sample.

Table 24: TAE-buffer

Components	Concentration
Tris-acetate	400 mM
EDTA	20 mM
Acetic acid	to pH 8.5

After the gel had set up it was transferred to the electrophoresis chamber and completely submerged in TAE buffer. DNA samples were mixed with 6x Loading Dye (Table 25) prior to loading on the gel.

Table 25: 6x DNA loading dye

Components	Amount/Concentration
Glycerol	30% (v/v)
Bromphenol blue	0.25% (w/v)
Xylene cyanol	0.25% (w/v)

Additionally, a DNA ladder (Gene Ruler DNA Ladder Mix, Thermo Scientific) was loaded to determine the size of DNA fragments. The gel was run at 5-10 V/cm until the desired separation of DNA fragments was achieved. Since ethidium bromide in the gel intercalates with nucleic acids, the DNA could be visualized under UV light. Analytical gels were documented using a GelDoc 2000 transilluminator (BioRad) with the appropriate software (QuantityOne, BioRad).

Preparative gels were visualized on a UV-transilluminator with longer wave length (365 nm) to avoid damage of the DNA. Specific bands were excised with a scalpel and DNA was extracted using the Gel DNA recovery kit (Zymo) after the manufacturer's instructions.

3.3.7 Enzymatic digestion of DNA

Sequence specific digestion of DNA was performed with Fast Digest endonucleases (Life Technologies). Generally, up to 1 µg of DNA was digested in a 30 µL reaction with 3 µL of 10x reaction buffer and 1 µL of restriction enzyme. Reactions were incubated for 30-60 min at 37°C. Restriction enzymes were heat inactivated under the conditions provided by the manufacturer. To avoid re-circularization of digested plasmids, 1 U of Calf Intestine Phosphatase (CIP, NEB) was added to the reaction for dephosphorylation of DNA ends.

3.3.8 Ligation of DNA fragments

Ligation of DNA fragments into linearized plasmids was catalyzed by T4 DNA Ligase (Thermo scientific). Reactions were set up as shown in Table 26 and incubated for 20-60 min at 22°C or overnight at 16°C.

Table 26: Ligation reaction

Component	Amount
Linear vector DNA	20 - 100 ng
Insert DNA	1:1 - 5:1 molar ratio over vector
10 x T4 DNA Ligase buffer	2 µL
T4 DNA Ligase	0.2 µL
H_2O	ad 20 µL

3.3.9 Polymerase chain reaction (PCR)

The polymerase chain reaction for sequence specific amplification of DNA was originally described by Mullis et al. [171]. For short PCR products or for diagnostic PCR e.g. for the detection of splice variants by RT-PCR, Taq polymerase (DreamTaq DNA polymerase, Life technologies) was used. Amplicons for cloning and sequencing were produced by a proof-reading polymerase (PfuUltra II HS DNA polymerase, Agilent) to reduce the error rate. If not otherwise stated, PCR was performed according to the polymerase manufacturer's recommendations. PCR products for cloning were purified with the Clean and Concentrator Kit (Zymo) as described by the supplier.

3.3.10 TA-Cloning

TA-cloning employs the property of Taq polymerase to add deoxyadenine at the 3'end of PCR products. Therefore, Taq polymerase derived PCR products can be ligated with high efficiency into linearized plasmids with 3' deoxythymidine overhangs. Due to the high error rate of Taq polymerase, sequences for cloning were generally amplified in a Pfu polymerase reaction as described above. A-overhangs were generated by incubating the completed PCR reaction with 2 U Dreamtaq DNA polymerase and 1 µL 10 mM dATP (Thermo Scientific) for 15 min at 72°C. After analysis by agarose gel electrophoresis and optional gel purification, amplicons were cloned into the TA-plasmid pCR2.1 (Thermo Scientific) using 1 µL of linear vector and 2 U T4 DNA-Ligase (Thermo Scientfic) in a 20 µL ligation reaction. Ligation was allowed to proceed for 1 h at RT. The whole reaction was used to transform chemically competent bacteria.

3.3.11 Blue-white-screening

Blue-white screening of transformed bacteria is an efficient way to identify recombinant bacterial clones after TA-cloning. The *E. coli* strain DH5α carries a mutation in the lacZ gene, which leads to the expression of a truncated, non-functional β-galactosidase enzyme. The activity of β-galactosidase can be trans-complemented by uptake of a plasmid, which encodes for the missing α-region of β-galactosidase (e.g. pCR2.1 (Invitrogen)). Bacteria with a complemented β-galactosidase activity can hydrolyze the chromogenic substrate X-Gal (5-Bromo-4-Chloro-3-indolyl- ß-D-Galactopyranosid) leading to a blue color of bacterial colonies.
Production of recombinant plasmids by insertion of a DNA fragment in the α-region of the plasmid abrogates the complementation of the β-galactosidase activity. Therefore, recombinant bacterial colonies are not able to process X-Gal and appear white.
For blue-white-screening, LB agar plates with the antibiotic of choice were supplemented with 40 µl X-Gal (20 mg/mL) and 2.5 µL isopropyl β-D-1-thiogalactopyranoside (IPTG, 0.1 M), a non-metabolizable analog of galactose that induces the expression of lacZ gene.

3.3.12 Sanger sequencing

Sequence analysis of plasmids and PCR products was performed by Seqlab Sequence Laboratories Göttingen GmbH. Sequencing results were analyzed with CLC Main Workbench (Qiagen).

3.3.13 Rolling circle amplification

For specific detection of circular DNA among genomic DNA, the bacteriophage Phi29 DNA polymerase was used. Owing to its strand displacement activity, this polymerase effectively produces concatameric products from circular DNA templates whereas linear DNA is only inefficiently amplified. The illustra TempliPhi Amplification Kit (GE Healthcare) was used for rolling circle amplification (RCA).

RCA reactions were prepared as recommended by the manufacturer using 25 ng of genomic and additional 0.5 µL 10 mM dNTPs as described by Rector et al. [172]. The reaction was allowed to proceed for 18 h at 30°C. The enzyme mix was inactivated by heating to 65°C for 10 min. 1 µL of the RCA product was digested with *BamHI* to obtain monomeric MCVSyn and analyzed on a 0.8% agarose gel.

3.3.14 Preparation of low molecular weight DNA (HIRT extraction)

The small circular DNA of polyomaviruses can be selectively isolated from cultured cells using the protocol of Hirt [173]. This method is based on the preferential precipitation of genomic DNA in the presence of SDS and NaCl and purification of the remaining low molecular weight DNA by phenol chloroform extraction.

MCVSyn DNA was extracted from trypsinized and pelleted cells. 1 x 10^6 cells were re-suspended in 100 µL HIRT solution (Table 27) and incubated at RT for 10 min. 50 µL of 5 M NaCl were added and the samples were gently rocked for 2 min. Precipitation of high molecular weight DNA was allowed to proceed over night at 4°C. To remove the high molecular weight DNA, samples were centrifuged for 40 min at 20,000 x g and the supernatant was transferred to a new 1.5 mL tube. Proteins were removed by incubating the samples for 1 h at 37°C with 12 µg Proteinase K (Peqlab) followed by two rounds of phenol/chloroform/isoamylalcohol (25:24:1) extraction and one additional round of chloroform extraction. DNA was precipitated by addition of 2 volumes of Ethanol for at least 3 h at -80°C and pelleted by centrifugation (30 min, 20,000 x g, 4°C). The pellet was washed once with 70% ethanol and after complete removal of ethanol was dissolved in 50 µL H_2O.

Table 27: HIRT solution

Component	Concentration
Tris-HCl pH 7.5	10 mM
EDTA	10 mM
SDS	0.6%

3.3.15 Preparation of genomic DNA

The small, circular MCVSyn DNA could also be successfully co-isolated by preparation of genomic DNA (gDNA) allowing the quantification of MCVSyn copies relative to the cellular background. For gDNA preparation, 1-5 x 10^6 cells were harvested and washed once in PBS. Pelleted cells were re-suspended in 500 µL PBS and 500 µL 2x lysis buffer (Table 28) were added. After 5 min incubation on ice, nuclei were centrifuged and re-suspended in 50 µL PBS. 300 µL gDNA lysis buffer (Table 29) supplemented with 100 µg Proteinase K (Peqlab) and 50 µg RNase A (Sigma) were added. RNA was degraded by incubation at 37°C for 30 min followed by Proteinase K digestion at 54°C for 16 h. The DNA was purified by two rounds of phenol/chloroform/isoamylalcohol (25:24:1) extraction and one round of chloroform extraction. 0.8 volumes of isopropanol were added to the aqueous phase to precipitate the DNA for 20 min at RT. The DNA was pelleted by centrifugation (30 min, 20,000 x g, 4°C) and washed once with 70% ethanol. After removal of residual ethanol, gDNA was dissolved in 50 µL H_2O. Prior to qPCR analysis, gDNA was linearized by digestion with *EcoRI* and bacterial input DNA was removed by *DpnI* digestion (FastDigest, Thermo Scientific) for 60 min at 37°C

Table 28: 2x lysis buffer

Components	Concentration
Sucrose	650 mM
Tris-HCl pH 7.8	20 mM
$MgCl_2$	10 mM
Triton X-100	2%

Table 29: gDNA lysis buffer

Components	Concentration
NaCl	100 mM
Tris-HCl pH 8	10 mM
EDTA	25 mM
SDS	0.5%

3.3.16 Southern blot analysis

Southern Blot analysis was performed to detect *de novo* replicated MCVSyn. Either HIRT extracts or genomic DNA was isolated as described above. 2 µg of purified *EcoRI* and *DpnI* digested DNA was mixed with 6x loading dye and loaded on a 0.8% agarose with 0.5 µg/mL ethidium bromide. The gel was run until the bromophenol blue band had reached the lower third of the gel and was then visualized in a UV box. To facilitate the transfer of large DNA molecules by fragmentation

and denaturation of DNA, the gel was soaked for 15 min in depurination buffer (Table 30), 30 min in denaturation buffer (Table 31) and 30 min in neutralisation buffer (Table 32). The gel was rinsed in H_2O between these steps. Afterwards, DNA was transferred to a nylon membrane (Zeta-Probe GT Membrane, Bio-Rad) by semi-dry transfer in 10x SSC (Table 33).

Table 30: Depurination buffer

Components	Concentration
HCl	0.25 M

Table 31: Denaturation buffer

Components	Concentration
NaOH	0.5 M
NaCl	1 M

Table 32: Neutralisation buffer

Components	Concentration
Tris-HCl	1.5 M
NaCl	1.5 M
HCl	to pH 7.4

Table 33: 10x SSC

Components	Concentration
NaCl	1.5 M
$Na_3Citrate \times 2H_2O$	150 mM

For detection of MCVSyn DNA, a fragment was amplified from the early viral region, using the primers MCPyV-LT-s and MCPyV-LT-as, and labeled with ^{32}P dCTP (Rediprime DNA Labeling System, GE Healthcare). Blots were hybridized with the labeled probe for 16 h at 42°C in ULTRAhyb buffer (Ambion).

Blots were washed 2x20 min with 1% SSC, 0.1% SDS and 2x20 min with 0.1x SSC, 0.1% SDS at 50°C. After at least 24 h of exposure, blots were scanned with the Fuji phosphoimager FLA7000 and analyzed with Multigauge software.

3.3.17 Isolation of DNase I resistant DNA from cell culture supernatants

Four days after transfection with re-ligated MCVSyn, cell culture supernatants were collected and sterile filtered. 1 mL of supernatant was supplemented with 10x DNase I reaction buffer and

DNA was digested with 10 µL DNase I (amplification grade, Invitrogen) for 1 h at 25°C. DNase I was heat inactivated for 10 min at 65°C in the presence of 2.5 mM EDTA. Proteins were degraded by addition of 5 µL Proteinase K and 1% SDS at 50°C for 16 h. DNA was retrieved by phenol-chloroform extraction and precipitation. Prior to qPCR analysis, the DNA was digested with *DpnI* and *EcoRI*.

3.3.18 Quantitative Realtime PCR (qPCR)

By quantitative Realtime PCR (qPCR) the abundance of a specific DNA or cDNA sequence can be measured using fluorescent dyes that label PCR products during each cycle. The intensity of fluorescent signal during the elongation phase is proportional to the amount of PCR product. The analysis software computes the threshold cycle (C_t) at which the fluorescence signal exceeds the background fluorescence and calculates the concentration of the input sample in respect to a given standard value.

With the exception of primers used for miRNA quantification (see below), qPCR primers were designed to span a region of 80-300 nt. All primers used in qPCR analysis were first validated for their specificity and efficiency by running a standard curve of at least 6 log ratios and a melting curve analysis. Only primer pairs with a PCR efficiency between 0.9-1.1 and a single peak of fluorescence in melting curve analysis were used. Specificity of primers was further confirmed by agarose gel electrophoresis of qPCR products.

3.3.18.1 qPCR with SYBR green

Except for quantification of mcv-miR-M1, SYBR green was used as fluorescent label. SYBR green specifically binds to double stranded DNA and emits fluorescence only when bound. The composition of a standard SYBR green qPCR reaction is shown in Table 34. For each primer pair in every run, a non-template control as well as a dilution of the previously recorded standard curve was included.

Table 34: SYBR green qPCR reaction

Components	Amount
2x SensiMix SYBR Hi-Rox Mastermix (Bioline)	5 µL
Primer fw (10 µM)	1 µL
Primer rev (10 µM)	1 µL
Template (e.g. cDNA, gDNA, ChIP-ed DNA)	1.5 - 2 µL
H_2O	ad 10 µL

qPCR was performed on a Rotor-Gene Q Realtime machine (Qiagen) under the conditions described in Table 35 and was analyzed with the Rotor-Gene 6000 software. A gain optimization was carried out at the beginning of the run. SYBR green fluorescence was recorded during elongation. After completion of the run, a melting curve analysis was performed.

Table 35: Cycling parameters for SYBR green qPCR

Number of cycles	Temperature	Time
1	95°C	10 min
40	95°C	15 s
	55-60°C	15 s
	72°C	15 s

Special considerations for different applications of SYBR green qPCR are elucidated below.

3.3.18.2 Relative quantification of mRNA expression levels by qPCR

To analyze and compare mRNA levels, total RNA was converted into cDNA using random hexamer primers as described in chapter 3.4.2. Primers for qPCR analysis were designed to span exon-boundaries to avoid amplification of residual gDNA. Standard curves for all primer pairs were produced by serial dilutions of a cDNA sample. One sample of the standard curve was included in all following qPCR analyses. The obtained mRNA expression values were normalized to the expression of the housekeeping gene GAPDH.

3.3.18.3 Quantification of replicated MCVSyn genomes by qPCR

Replicated MCVSyn could be detected by qPCR of *EcoRI/DpnI* digested HIRT extracts and gDNA. Although absolute concentration of MCVSyn was higher in HIRT extracts, gDNA was the preferred template for qPCR since it allowed the simultaneous measurement of cellular DNA and thus enables to calculate the absolute number of MCVSyn copies relative to the cellular background.

A GAPDH standard curve was produced from gDNA based on the calculation that 3.3 pg gDNA correspond to the mass of a haploid human genome and contain one copy of the GAPDH gene (m(haploid genome) = $3x10^9$ bp * $1.096x10^{-21}$ g/bp = 3.3 pg). The standard curve contained serial dilutions from 330 ng/µL (100,000 copies GAPDH/µL) to 33 pg/µL (10 copies GAPDH/µL).

The standard curve for MCVSyn was produced using *SacI* digested and gel purified pMCVSyn. Based on the size of 5388 bp, it was calculated that *SacI* digested pMCVSyn, or replicated

MCVSyn, has a mass of $5.9x10^{-18}$ g (m(pMCVSyn *SacI*) = 5388 bp * $1.096x10^{-21}$ g/bp = $5.9x10^{-18}$ g). Accordingly, a pMCVSyn *SacI* standard curve was generated from 59 ng/µL (10^{10} copies/µL) to 0.59 fg/µL (100 copies/µL).

MCVSyn was amplified using primers in the late region (VP1fw/VP1 rev) spanning 3 *DpnI* restriction sites and GAPDH was amplified using the primer pair GAPDH DNA fw/GAPDH RT rev. 25-50 ng gDNA were used as template for each qPCR reaction and the generated standard curves were imported in each run. The ratio of MCVSyn copy number to GAPDH copy number was divided by two to obtain the value "MCVSyn copy number per cell". Notably, this value is only a relative number as the majority of cells did not contain MCVSyn DNA but was used to compare MCVSyn copy numbers between different experiments.

3.3.18.4 Evaluation of ChIP experiments by qPCR

To validate the success of ChIP experiments, ChIP-ed DNA was quantified at specific loci by qPCR. Since it is known that LT-Ag binds to the MCPyV origin of replication, primers spanning this region (MCPyV RT ori fw/rev) were used as a positive control for LT-Ag bound DNA. As a negative control, the primer set VP1 fw/VP1 rev was used as the late region was considered not to be bound by LT-Ag. For H3K4-me3, the human ADH5 locus served as a positive control for H3K4-trimethylated chromatin (primer set ADH5 fw/ADH5 rev). For all primer pairs, standard curves from input chromatin samples were produced.

The calculated concentration of each sample was used to compute the percent of input (input samples were diluted 1:10 or 1:100 to avoid over-amplification) and the fold enrichment against the background control was determined.

3.3.18.5 Quantification of miRNA expression by qPCR

Analysis of miRNA expression by qPCR, in comparison to Northern blot analysis allows a more accurate and comparable quantification of miRNA expression. However, due to the small size of mature miRNAs (~22 nt), the miRNA sequence needs to be artificially elongated during cDNA synthesis to provide sufficient length for primer binding and amplification. The cDNA-synthesis with a stem-loop primer is described in chapter 3.4.3. Figure 5 shows the entire process of SL-RT qPCR. The stem-loop primer binds specifically to a mature miRNA and serves as reverse primer for cDNA-synthesis. qPCR is performed using a miRNA specific forward primer and a universal reverse primer which binds in the conserved region of the stem-loop primer. Using differently labeled Taqman probes, the expression of mcv-miR-M1 and the housekeeping gene GAPDH used for normalization could be measured in the same reaction.

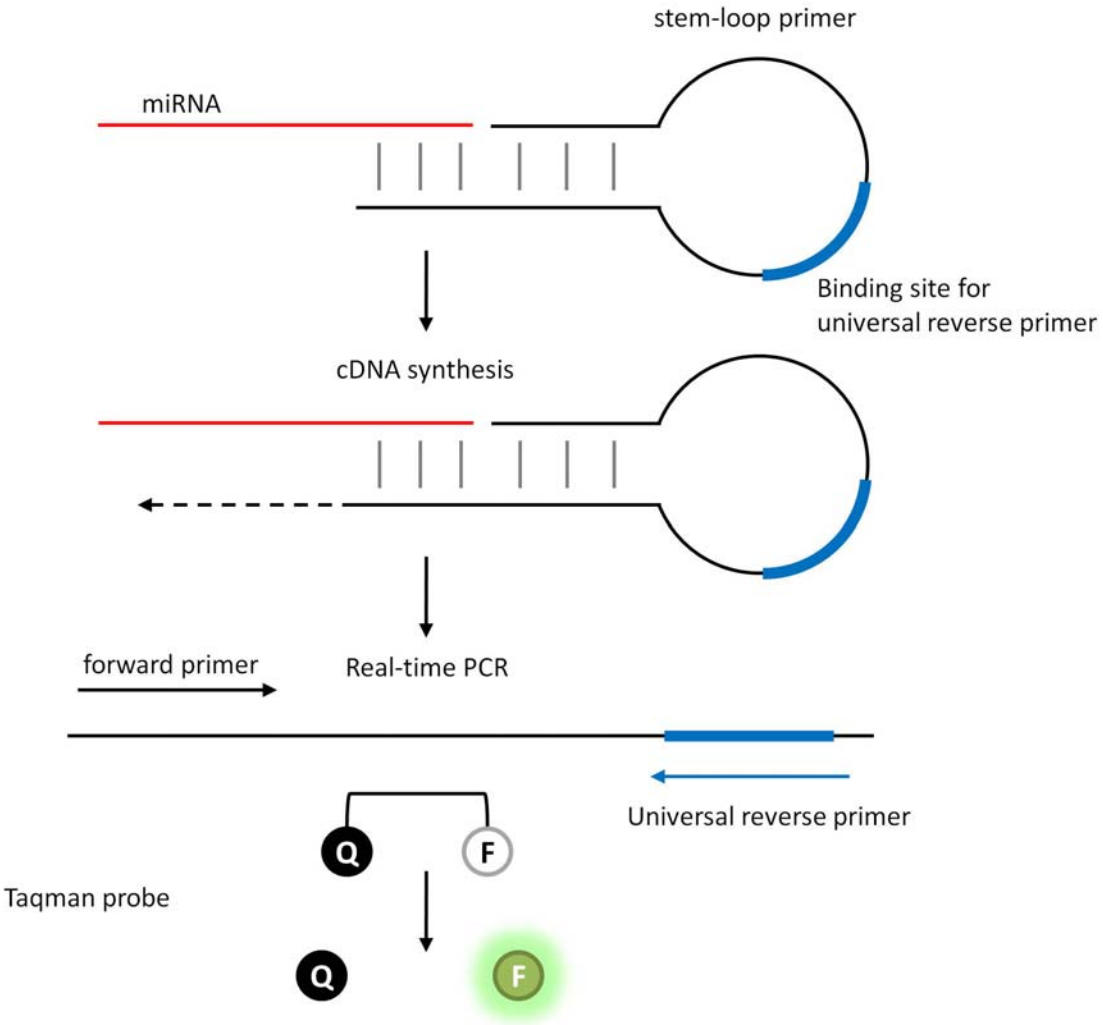

Figure 5: Stem-loop RT-qPCR for quantification of miRNA expression
A stem-loop primer hybridizes to 4-6 nt at the 3'end of a mature miRNA (red) and folds into a hairpin structure. The binding site for the universal reverse primer is depicted in blue. cDNA-synthesis is carried out by elongation of the stem-loop primer. After removal of template RNA by RNase H digestion, qPCR is performed using a miRNA specific forward primer and the universal reverse primer. A sequence specific Taqman probe containing a fluorophore (F) and a quencher (Q) binds to the cDNA and is cleaved by the endonuclease activity of the DNA-polymerase resulting in release and activation of the fluorophore.

Table 36 shows the composition of a qPCR reaction for simultaneous measurement of GAPDH and mcv-miR-M1.

qPCR was performed under the conditions listed in Table 37 and was analyzed with the Rotor-Gene 6000 software. Optimization of the gain for each fluorophore was carried out at the beginning of the run. The fluorescence originating from cleaved Taqman probes was recorded during the combined annealing and elongation step.

Expression of other miRNAs was measured by SYBR green qPCR as described above.

Table 36: Multiplexed Taqman qPCR for mcv-miR-M1 and GAPDH

Components	Amount
2x Rotor Gene Multiplex PCR Mastermix (Qiagen)	5 µL
GAPDH RT rev [10 µM]	0.3 µL
GAPDH BSPIII [10 µM]	0.3 µL
GAPDH Taqman probe (ROX™)	0.2 µL
Merkel miR fw TM2 [10 µM]	0.3 µL
Uni rev [10 µM]	0.3 µL
Merkel Probe TM2 (VIC™)	0.2 µL
cDNA	2 µL
H_2O	ad 10 µL

Table 37: Cycling parameters for Taqman multiplex qPCR

Number of cycles	Temperature	Time
1	95°C	10 min
50	95°C	15 s
	60°C	60 s

3.4 RNA techniques

For all experiments with RNA, nuclease-free or DEPC treated H_2O was used. Prior to all RNA handling, surfaces and pipettes were cleaned with RNase-Zap (Ambion). RNase-free tubes and filter tips were used.

3.4.1 Extraction of total RNA from eukaryotic cells

Total RNA from cultured cells was extracted by acid guanidinium thiocyanate-phenol-chloroform extraction as first described by Chomczynski and Sacchi [174] using RNA Bee reagent (amsbio).

$1x10^6$ cells were lysed in 1 mL of RNA Bee by vigorous pipetting on ice. Adherent cells were lysed directly in the culture vessel after removal of medium. Suspension cells or trypsinized cells were pelleted and medium was removed prior to lysis. Cell homogenates were transferred into 1.5 mL tubes and 200 µL chloroform were added. The aqueous and organic phase were first mixed by shaking the tubes for 30 s and after 5 min incubation on ice, the phases were separated by centrifugation (12,000 g, 15 min, 4°C). While DNA and proteins were located in the lower, organic phase, RNA stayed in the aqueous phase which was transferred to a new tube and precipitated by addition of 500 µL isopropanol. After 10 min precipitation at RT, RNA was pelleted by centrifugation (12,000 g, 10 min, 4°C). The RNA pellet was washed once in 1 mL of

75% ethanol. After centrifugation (12,000 g, 10 min, 4°C) and removal of the wash solution, the RNA pellet was briefly air-dried and reconstituted in 30-50 µL H_2O_{DEPC}. For especially sensitive applications, e.g. for High Throughput Seqeuncing library preparation or RACE analysis, the integrity of RNA was visualized by MOPS-formaldehyd gel electrophoresis or using a Total RNA Bioanalyzer Chip (Agilent). The RNA concentration was determined either with a NanoDrop-1000 Spectrophotometer (Thermo Scientific) or a Qubit Fluorometer (Life Technologies). RNA was stored at -80°C.

3.4.2 cDNA synthesis

RNA was reverse transcribed into cDNA to detect or quantify specific transcripts by PCR. For these applications, only primer pairs were used that span exon boundaries which avoids the necessity of DNase digestion prior to cDNA synthesis. cDNA-synthesis using random hexamer primers was performed as described by the manufacturer of the reverse transcriptase Superscript III (Life Technologies). In brief, 1 µg of RNA in a total volume of 12 µL H_2O_{DEPC} were mixed with 1 µL dNTPs (10 mM) and 250 ng random hexamer primers. To break secondary RNA structures, the samples were heated to 65°C for 5 min and immediately put on ice afterwards. 4 µL 5x First Strand buffer, 1 µL 0.1 M DTT, 0.7 µL RNaseOUT and 0.3 µL Superscript III (all reagents from Life Technologies) were added and the samples were mixed well. The random primers were allowed to anneal for 5 min at 25°C. Reverse transcription proceeded for 60 min at 50°C and was eventually stopped by heating the reaction for 15 min to 70°C. RNA was removed by addition of 0.2 µL RNase H (NEB) and incubation at 37°C for 20 min.

For the detection of splice variants of MCVSyn transcripts, 2 µL of cDNA were subjected to a standard PCR reaction (see chapter 3.3.9). To quantify viral or cellular transcripts, 1-2 µL of cDNA were used in a 10 µL qPCR reaction (see chapter 3.3.18).

3.4.3 miRNA stem-loop cDNA-synthesis

Quantification of miRNAs was accomplished as originally described by Varkonyi-Gasic et al. [175]. Due to their small size, amplification of mature miRNAs is inefficient using standard cDNA-synthesis approaches. A primer with an internal stem-loop structure is used to elongate the short miRNA at its 3' end and to function as a scaffold for a universal reverse primer used in qPCR (Figure 5).

To break secondary structures of RNA and to facilitate binding of the stem-loop primer prior to cDNA synthesis, the components indicated in the upper part of Table 38 were mixed and heated to 65°C for 5 min and afterwards rapidly cooled on ice.

The remaining components were added as indicated in the lower part of Table 38 and cDNA was produced during a pulsed program (Table 39).
RNA was degraded by addition of 0.2 µL RNase H (NEB) and 1.8 µL H_2O and incubation at 37°C for 20 min. Quantitative PCR was carried out as described in chapter 3.3.18.

Table 38: stem-loop cDNA synthesis

Components	Amount
Total RNA	1 µg
SL Merkel miR TM2 primer [1 µM]	0.5 µL
10 mM dNTPs	1 µL
H_2O_{DEPC}	ad 13 µl
5 min 65°C, 2 min ice	
5x First Strand buffer	4 µl
0.1 M DTT	1 µL
GAPDH RT rev primer [2 µM]	1 µL
RNaseOUT	0.5 µl
Superscript III	0.5 µl

Table 39: Pulsed program for miRNA stem-loop cDNA-Synthesis

Number of cycles	Temperature	Time
1	16°C	30 min
60	30°C	30 s
	42°C	30 s
	50°C	1 s
1	85°C	5 min

3.4.4 small RNA Northern blot analysis

For Northern blot analysis, 14 µg of total RNA were separated on a denaturing 15% polyacrylamide urea gel and transferred to a nylon membrane (Zeta-Probe GT membrane, Bio-Rad) by semi-dry electro blotting. An antisense oligonucleotide probe (mcv-miR-M1 probe) was end-labeled with ^{32}P dATP using T4 PNK (NEB).
Blots were pre-hybridized to in ExpressHyb hybridization buffer (Clontech) for 1 h at 37°C. The labeled probe was added and blots were incubated for 16 h at 37°C. Membranes were washed twice in 2x SSC, 0.1% SDS at room temperature and subsequently subjected to autoradiography. Blots were scanned on the BAS-Reader and analyzed with AIDA Software (Raytest).

3.4.5 3' RACE analysis

To obtain the sequences of MCVSyn 3' UTRs, RACE analysis was performed according to the protocol of Scotto-Lavino et al. [166] with minor modifications.

5 µg of total RNA of MCVSyn transfected PFSK-1 cells 4 d post transfection were used for cDNA synthesis with Superscript III (Invitrogen) and an anchored poly T primer (Q_T). Prior to cDNA-synthesis, RNA was heated for 3 min at 80°C and cooled on ice. 1 µL 10 mM dNTPs, 50 ng Q_T primer and H_2O_{DEPC} to a final volume of 20 µL were added to the RNA. The mixture was heated to 65°C for 5 min and cooled on ice prior to addition of 4 µL 5x first strand buffer, 2 µL 0.1 M DTT, 0.5 µL RNaseOUT and 1 µL Superscript III. cDNA synthesis was carried out as shown in Table 40.

Table 40: Program for 3'RACE cDNA-synthesis

Temperature	Time
22°C	5 min
42°C	60 min
50°C	10 min
70°C	15 min
4°C	∞

The input RNA was degraded by addition of 1.5 U RNase H (NEB) and incubation at 37°C for 20 min. The 3'end cDNA pool was filled up to 1 mL with TE buffer pH 8.0 and could be stored at 4°C. 3'ends were amplified using a gene specific primer (LT_o/$VP1_o$) and a primer specific for the sequence of the Q_T primer (Q_o) in a 50 µL Pfu Ultra II PCR reaction under the conditions provided in Table 41.

Nested gene specific reverse primers (LT_i/$VP1_i$) and the forward primer Q_i were used for a second round of amplification to increase the specificity and to add restriction sites at the end of PCR products. 1 µL of 1:20 diluted PCR product was used in a standard 50 µL Pfu Ultra II reaction. Cycling parameters are listed in Table 42.

PCR products were visualized on a 1.2 % agarose gel and discrete bands were cut out and DNA was purified. 3'RACE products were digested with the respective restriction enzymes (Fast Digest, Thermo scientific) and cloned into pCR2.1 plasmid. Randomly picked clones were subjected to Sanger sequencing.

Table 41: Cycling parameters for 3'RACE outer PCR

Number of cycles	Temperature	Time
1	98°C	5 min
1	55°C	2 min
1	72°C	40 min
30	95°C	15 s
	56°C	15 s
	72°C	2 min
1	72°C	15 min
1	22°C	∞

Table 42: Cycling parameters for 3'RACE inner PCR

Number of cycles	Temperature	Time
1	98°C	5 min
30	95°C	15s
	56°C	15 s
	72°C	2 min
1	72°C	15 min
1	22°C	∞

3.4.6 Cap dependent 5' RACE

Mapping of transcriptional start sites (TSS) of MCVSyn transcripts was accomplished by 5' RACE analysis of MCVSyn transfected PFSK-1 cells based on the GeneRacer™ Kit protocol (Invitrogen). 5 µg of total RNA were dephosphorylated in a 10 µL reaction with 1 µL RNaseOUT and 20 U CIP (Invitrogen) for 1 h at 50°C. For clean-up, 90 µL H_2O_{DEPC} were added and two rounds of phenol-chloroform-isoamylalcohol extraction and one round of chloroform extraction was carried out in safe-lock-tubes. The aqueous phase was transferred into a new tube and precipitated by addition of 4 µL Glycogen (5 µg/µL, Invitrogen), 10 µL 3 M Na-Acetate pH 5.2 and 220 µL 100% Ethanol for 10 min on dry ice. Precipitated DNA was washed once with 75% Ethanol and after complete removal of Ethanol was dissolved in 7 µL H_2O_{DEPC}.

RNA was decapped in a 10 µL reaction with 1 µL RNaseOUT and 2 U Tobacco Acid Pyrophosphatase (TAP, Epicentre) for 1 h at 37°C. RNA was cleaned up as described above and dissolved in 6 µL H_2O_{DEPC}. 1 µL 10 mM 5'RACE Adapter was added and secondary structures were removed by heating to 65°C and rapid cooling on ice.

For ligation, 1 µL 10 mM ATP, 1 µL RNaseOUT, 1 µL 10x T4 RNA ligase buffer and 10 U T4 RNA ligase (Fermentas) were added. After 1 h incubation at 37°C, RNA was cleaned up as described above. The whole RNA sample was subjected to a standard cDNA-synthesis with Superscript III

as described in chapter 3.4.2 using 1 µL 10 µM gene specific reverse primer (5‘ RACE late/early region). cDNA synthesis was carried at 55°C for 1 h, followed by inactivation of the reaction at 70°C for 15 min. RNA was degraded by addition of 2.5 U RNaseH (NEB) for 20 min at 37°C.
1 µL of diluted cDNA (1:2 or 1:10 diluted) was amplified in a standard 50 µL Pfu Ultra II reaction with 4 µL 5’RACE adapter primer and 1.5 µL of the above used gene specific primer (5‘ RACE late/early region) using a touchdown PCR program (Table 43).

Table 43: Cycling parameters for 5’RACE touchdown PCR

Number of cycles	Temperature	Time
1	98°C	30 s
5	98°C	10 s
	72°C	30 s
5	98°C	10 s
	70°C	30 s
10	98°C	10 s
	68°C	30 s
	72°C	30 s
10	98°C	10 s
	66°C	30 s
	72°C	30 s
10	98°C	10 s
	64°C	30 s
	72°C	30 s
1	72°C	10 min
1	4°C	∞

1 µL of 1:10 diluted PCR product was amplified in a standard 50 µL Pfu Ultra II reaction with 5‘ RACE late/early region nested rev primer and 5‘ Adapter primer BamHI to add restriction sites at both ends of the amplification products (Table 44).
After nested PCR, amplicons were analyzed on a 1.2% agarose gel and DNA from individual bands was purified. RACE products were digested with *BamHI* and *XhoI* (Fast Digest, Life technologies) and cloned into pCR2.1 plasmids. Individual clones were analyzed by Sanger sequencing.

Table 44: Cycling parameters for 5'RACE nested PCR

Number of cycles	Temperature	Time
1	98°C	30 s
5	98°C	10 s
	72°C	30 s
5	98°C	10 s
	70°C	30 s
15	98°C	10 s
	68°C	30 s
	72°C	30 s
1	72°C	10 min
1	4°C	∞

3.5 SDS PAGE and Western blot analysis

LT-Ag expression in MCVSyn and MCVSyn-hpko transfected cells was investigated by Western blot analysis. For preparation of lysates, cells were trypsinized and washed once in PBS. Cell pellets from approximately $5x10^5$ cells were re-suspended thoroughly in 200 µL lysis buffer (Table 45). For complete disruption of membranes, lysates were further homogenized by passing through a syringe (0.4 mm).

Table 45: Lysis buffer

Components	Concentration
Tris-HCl pH 8.0	50 mM
NaCl	150 mM
NP40	1%
Sodium-deoxycholate	0.5%
EDTA	5 mM
SDS	0.1%
Protease inhibitor cocktail (Roche)	1 tablet per 10 mL lysis buffer

Lysates were incubated on ice for 20 min and subsequently centrifuged for 20 min at 20,000g and 4°C. Supernatants were transferred into new 1.5 mL tubes and stored at -20°C if necessary. Protein concentration was determined photometrically after the method of Bradford et al. [75] using the Bio-Rad Protein Assay kit as described by the manufacturer.
25 µg of protein were supplemented with 4x sample buffer (Table 46) and boiled for 5 min.

Table 46: 4x SDS-PAGE sample buffer

Components	Concentration
SDS	1.2 g
Glycerol	4.65 mL
Tris-HCl pH 7.0	200 mM
Bromophenolblue	0.4 g
β-Mercaptoethanol	0.5 mL
H_2O	ad 10 mL

Proteins were separated on a 10% polyacrylamid gel in Tris-Glycine buffer with a constant current of 35 mA. After completion of electrophoresis, separated proteins were electroblotted on a PVDF-membrane using semi-dry transfer. Blots were blocked in 5% milk-TBST and were probed with MCPyV LT-Ag antibody CM2B4 (Santa Cruz) over night at 4°C. After washing, blots were incubated with the secondary antibody for 1 h at 4°C.

3.6 Chromatin Immunoprecipitation (ChIP)

Chromatin Immunoprecipitation (ChIP) was performed as previously described [176,177]. In brief, 4 d post transfection with MCVSyn wt or MCVSyn mutants or 2 d after transfection with LT-Ag expression plasmids, chromatin of 1×10^6 cells was cross linked by incubation with 1% formaldehyde. The reaction was stopped by the addition of glycine. Chromatin was extracted from isolated nuclei and fragmented by sonication (Bioruptor, Diagenode) to an average length of 200-500 bp. A fraction of the total chromatin sample was set aside for the preparation of input controls. The remaining material was pre-cleared with BSA blocked protein-G sepharose beads (GE Healthcare) to reduce non-specific background. For immunoprecipitation, 2 µg of antibodies specific for the histone modification H3K4-me3 (Millipore) or for MCPyV LT (CM2B4, Santa Cruz Biotechnology, sc-136172) or IgG anti-rabbit (Millipore) antibody were added to the chromatin and incubated for 16 h at 4°C. Chromatin-immunocomplexes were precipitated by the addition of protein-G sepharose beads (GE Healthcare), washed with increasing salt concentrations, eluted and de-crosslinked for 16 h at 65°C. DNA was purified by proteinase K digestion with subsequent phenol-chloroform extraction and ethanol precipitation. ChIP experiments were analyzed by qPCR (chapter 3.3.18.4) and High Throughput Sequencing (chapter 3.8.4).

3.7 Microarrays

To analyze the influence of mcv-miR-M1 on the cellular transcriptome and thereby identifying possible target mRNAs, whole genome expression arrays (Agilent 4x44K) were performed from cells transfected with MCVSyn or with a mcv-miR-M1 expression plasmid.

RNA was isolated from mcv-miR-M1 expressing (sample) and mock transfected (control) cells with RNA Bee as described in chapter 3.4.1. To confirm the integrity of RNA, a Bioanalyzer analysis with the RNA Nano Kit (Agilent) was carried out. 1 µg of high quality sample and control RNA was labeled using the Two Color Quick Amp Labeling Kit (Agilent) as described by the manufacturer. Briefly, using the Two-Color RNA Spike-In Kit (Agilent), spike-in RNA was added to sample and control RNA (spike A for Cy3 and spike B for Cy5) for normalization. RNA including spike-in RNA was transcribed into cDNA using MMLV reverse transcriptase for 2 h at 40°C. cDNA was then converted into cRNA by T7 RNA polymerase. In the same step, the newly generated cRNA was labeled with Cy3 or Cy5 by addition of Cy3-CTP or Cy5-CTP to the sample or control, respectively. cRNA was purified with the RNeasy Kit (Qiagen) and purity and dye incorporation were measured spectrophotometrically with the NanoDrop-1000 using the "microarray measurement" settings.

The specific activity of labeled cRNA was calculated using the following formula:

(*Concentration of Cy3 or Cy5*) / (*Concentration of cRNA*) * 1000 = pmol Cy3 or Cy5 per µg cRNA

cRNA was only used for hybridization if the specific activity was higher than 8.0 pmol Cy3 or Cy5 per µg cRNA and the total yield was more than 825 ng. Equal amounts of sample and control cRNA were used for hybridization (minimum 825 ng). cRNA was fragmented and afterwards hybridized and loaded onto the array. The array was incubated for 17 h at 65 °C and 10 rpm in a hybridization oven. After washing and drying, array slides were scanned using the GenePix reader 4100A. First, a preview scan with 40 µm resolution was performed to automatically determine the optimal gain for Cy3 and Cy5, the scanning was performed with 5 µm resolution with the calculated optimal gain. Further scans were carried out with higher and lower gain settings. After feature extraction, arrays were analyzed with GeneSpring GX10 software (Agilent). Fold change analysis was used to identify differentially expressed mRNAs between miRNA expressing cells and control cells. The cut off was arbitrarily set to 1.2 fold.

3.8 High Throughput Sequencing

3.8.1 Small RNA sequencing

For sequencing of small RNA moieties, RNA from MCVSyn transfected cells and MCC cell lines was subjected to library preparation using the TruSeq Small RNA Sample Preparation Kit (Illumina) or the NEBNext® Small RNA Library Prep Set for Illumina. Small RNA libraries were sequenced on the Illumina HiSeq 2500 platform with a depth of 50 mio reads using 50 bp single end flow cells. Using CLC Genomics Workbench v7.5.1 (Quiagen) reads were mapped to the MCVSyn genome and to human mature miRNAs deposited in (miRBase release 21 [178]). To ensure detection of isomiRs, an offset of 5 nucleotides along the annotated pre-miRNAs was allowed.

3.8.2 RNA sequencing

RNA-Seq libraries were produced from total RNA isolated with RNA Bee as described above. Prior to library preparation, the integrity of RNA was examined on a Bioanalyzer with the RNA Nano Kit (Agilent). Only RNA samples with an RNA Integrity Number (RIN) of 9 or higher were used for HTS analysis.

Library preparation for strand specific RNA sequencing was carried out using the ScriptSeq™ v2 RNA-Seq Library Preparation Kit (Epicentre) or NEXTflex™ Directional RNA-Seq Kit (Bioo Scientific). Libraries were sequenced on the Illumina HiSeq 2500 platform.

Splice junctions were mapped to MCVSyn and MCVSyn-hpko using TopHat2 v 2.0.13 [163]. Since it is not possible to map reads to circular genomes, two concatenated copies of the MCVSyn or MCVSyn-hpko were used as references for mapping to allow detection of splice junctions which extend over the origin of replication. The number of unspliced reads that extended over splice junctions was subsequently counted to determine splice site efficacy and frequency of individual junctions. For each combination of splice junctions that mapped within either the early or late gene cassettes the relative strand-specific frequency value was calculated by multiplying observed frequency values for individual donor sites. The ratio of late to early transcripts was subsequently estimated by calculation of normalized RPKM (reads per kilobase per million mapped reads) for each of the transcripts.

For the analysis of cellular gene expression levels upon expression of mcv-miR-M1, RNA-Seq data was analyzed for differential expression with the CLC Genomics Workbench v7.5.1 (Quiagen). Alternatively, reads were mapped to the human transcriptome with the algorithm TopHat2 v 2.0.13 [163] and differential gene expression was evaluated with R [129].

3.8.3 Sequencing of 5'RACE products

5' RACE products were analyzed on the Illumina HiSeq 2500 or MiSeq platform after library preparation with the NEBnext® Ultra™ DNA Library prep Kit from Illumina. Reads were mapped on the MCVSyn genome with TopHat2 v 2.0.13 [163] and results were visualized using the Integrative Genomics Viewer (v2.2.13) [159]. The number of reads starting at each genomic position was counted and the frequency of putative TSS was calculated for the early and late region, respectively.

3.8.4 ChIP-Seq

HTS libraries from ChIP samples (chapter 3.6) were prepared using the NEXTflex™ ChIP-Seq Kit (Bioo Scientific). Libraries were sequenced on the Illumina HiSeq platform with a depth of 20-40 mio reads. Mapping of reads to MCVSyn and to the human genome hg19 was performed by Bowtie (v 0.12.9) [153] and results were visualized with the Integrative Genomics Viewer (v2.2.13) [159]. Peaks were called with the algorithm MACS 1.4.2 [160] and annotated with bedtools. Using the program CEAS 1.0.2 [154,155] statistics about the distribution of peaks over the human genome was created. Motif discovery and analysis for specific motifs in peak regions was carried out by meme or tomtom, respectively using the package FIMO 4.1 [158].

3.9 Statistical analyses

Statistical analyses were performed using Prism 5 (GraphPad Software). Student's unpaired t-test was used to evaluate if differences in two datasets are statistically significant.

4 Results

Eukaryotic cells produce miRNAs to post transcriptionally modulate their gene expression. This mechanism plays a role in a wide range of processes e.g. cell fate, apoptosis, proliferation and oncogenesis [114]. Viruses exploit the cellular miRNA biogenesis machinery to produce their own miRNAs either to influence their own or the host's gene expression. Since their first discovery in 2004 [128], viral miRNAs are the subject of extensive research, although their functions are still not comprehensively understood. miRNAs are predominantly produced by viruses that have a persistent or latent life cycle which might be attributed to the fact that they are not immunogenic and cause generally very mild changes in gene expression and therefore do not disrupt cellular homeostasis. Among them, a group of human and animal polyomaviruses produce miRNAs [50-53,147,179]. These viruses each encode for a single pre-miRNA, which shows high sequence diversity among different polyomavirus clades although the relative genomic location of polyomavirus miRNAs, antisense to the early coding region, and thus their assumed function to downregulate early viral gene expression is conserved among all polyomaviruses, suggesting a significant role in viral replication, persistence or infection. However, the biological significance of polyomavirus miRNAs is not yet understood. One of the reasons for that is the lack of appropriate model systems, especially for human polyomaviruses. The narrow host range and cell specificity of polyomaviruses hamper the development of cell culture and animal models. Recent advances in establishing Merkel cell polyomavirus (MCPyV) model systems have created the prerequisite to investigate the MCPyV miRNA mcv-miR-M1 during viral replication. Since the natural reservoir of MCPyV is not known, the so far developed model systems employ a synthetic MCPyV genome which has to be excised from a bacterial plasmid and re-ligated prior to transfection into immortalized human cell lines [72,73]. Different human cell lines support viral replication, gene expression and a limited amount of particle production [72-74]. A drawback of this approach is that the replication system might not reflect every aspect of the natural biology of the virus and that the influence of mcv-miR-M1 on progeny production, an important output for virus success, cannot be analyzed.

In this study, the semi-permissive MCPyV replication system developed by Neumann, Borchert et al. [72] was employed to investigate the function of mcv-miR-M1 during viral replication. The synthetic MCPyV genome (MCVSyn) used in this system was designed as a consensus sequence from different MCC field isolates and shows viral replication, gene expression and a limited amount of particle production [72].

4.1 Expression of mcv-miR-M1 from replicating MCPyV genomes

MCPyV encodes for a single pre-miRNA hairpin which was discovered by *in silico* prediction in antisense orientation to the early gene region at the genomic coordinates 1168 to 1251. Ectopic overexpression of the pre-miRNA hairpin resulted in processing of the pre-miRNA into a 5p and a less abundant 3p mature miRNA at positions 16-37 or 50-71 of the pre-miRNA hairpin, respectively [51]. The following part investigates whether mcv-miR-M1 is expressed during viral replication, if mcv-miR-M1 maturation is altered in MCC cell lines and how the expression of mcv-miR-M1 is regulated.

4.1.1 Small RNA sequencing reveals high level expression of mcv-miR-M1 during MCVSyn replication and a dimorphism of the mature miRNA

The pre-miRNA sequence of mcv-miR-M1 is highly conserved among MCPyV isolates and is also contained in the synthetic MCVSyn genome. However, it was not yet analyzed if and how mcv-miR-M1 is expressed in the context of viral replication.

The expression of mcv-miR-M1 5p in PFSK-1 cells transfected with MCVSyn was measured by quantitative stem-loop RT-PCR (SL RT-qPCR). Expression of mcv-miR-M1 was detectable as early at 2 d post transfection (d.p.t.), peaked at 4-8 d.p.t. and then dropped again (Figure 6).

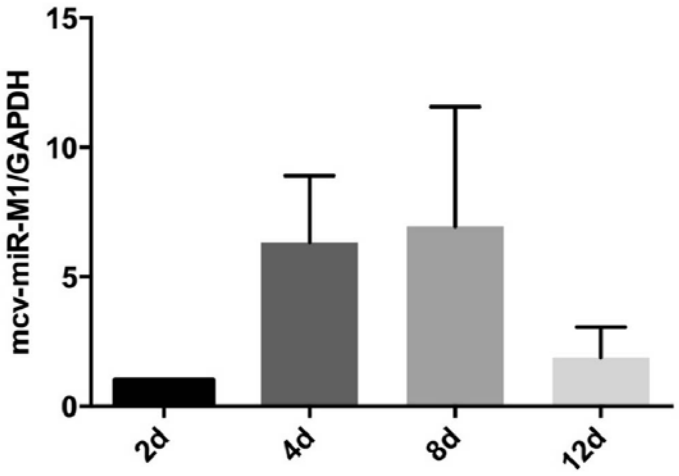

Figure 6: Expression of mcv-miR-M1 5p in PFSK-1 cells after transfection with re-ligated MCVSyn

Expression of mcv-miR-M1 relative to GAPDH as measured by SL RT-qPCR at 2 d, 4 d, 8 d and 12 d.p.t. of PFSK-1 cells with MCVSyn. Expression levels are shown relative to day 2 (n=4).

In 2011 Lee and colleagues reported the identification of a second mcv-miR-M1 5p species in MCC tumor material after HTS analysis of small RNA moieties (small RNA-Seq) [148]. This novel miRNA was proposed to be processed from the mcv-miR-M1 pre-miRNA into a mature 5p

miRNA that differs from the originally described 5p miRNA by a shift of 2 nt and comprises the nucleotides 14-35 of the pre-miRNA hairpin [148]. Although a 2 nt shift might not appear remarkable at first, it leads to a significant alteration in the target defining miRNA seed-region which comprises the nucleotides 2-8 of a mature miRNA (Figure 7). In the following, the two miRNA variants are referred to as $5p_{17\text{-}23}$ (first decribed in [51]) and $5p_{15\text{-}21}$ (first described in [148]) in respect to their seed positions.

Figure 7 depicts how the 2 nt shift between $5p_{17\text{-}23}$ and $5p_{15\text{-}21}$ significantly alters the target defining miRNA seed region and thereby the predicted targetome of mcv-miR-M1 5p among human 3'UTRs. Since the $5p_{15\text{-}21}$ miRNA was found in primary tumor material, it was hypothesized that mcv-miR-M1 undergoes differential processing in MCC cells to regulate a different set of targets in transformed cells.

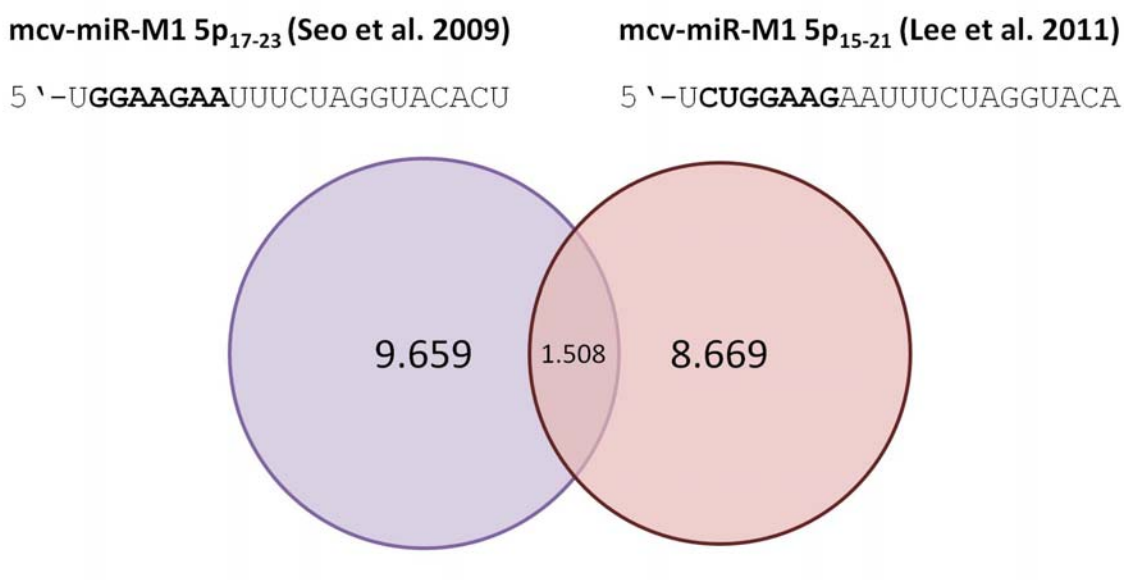

Figure 7: IsomiRs of mcv-miR-M1 5p and resulting differences in predicted targetomes
Sequences of the mature 5p mcv-miR-M1 as first identified by Seo et al. [51] and later found in MCC material by Lee et al. [148]. Seed sequences (nucleotides 2-8 of the mature miRNA) are highlighted in bold letters. To visualize the implication of the shifted seed sequence, the predicted targetomes of both miRNA variants were determined using TargetScanHuman (Release 6.1) [124]. The number of individual and shared putative target transcripts (human 3'UTRs) is depicted for both miRNA variants.

Although the primers for the detection of mcv-miR-M1 by SL RT-qPCR were designed to detect the $5p_{17\text{-}23}$ miRNA, it cannot be excluded that they can also amplify the $5p_{15\text{-}21}$ miRNA variant, given that both sequences can serve as template for the forward PCR primer.

To determine which seed variant of mcv-miR-M1 is predominantly expressed during MCVSyn replication and to analyze if this is indeed the only miRNA expressed by MCVSyn, sequencing of

small RNA moieties from cells harboring replicating MCVSyn at 4 d.p.t. was performed using the NEBnext small RNA Sequencing kit. Sequencing of small RNAs was generally performed on the Illumina HiSeq 2500 platform using 50 bp single end flow cells with a depth of 50 mio reads. The obtained reads were mapped on all human miRNAs as annotated in miRbase (Release 20) and on the MCVSyn genome using CLC Genomics Workbench v7.5.1.

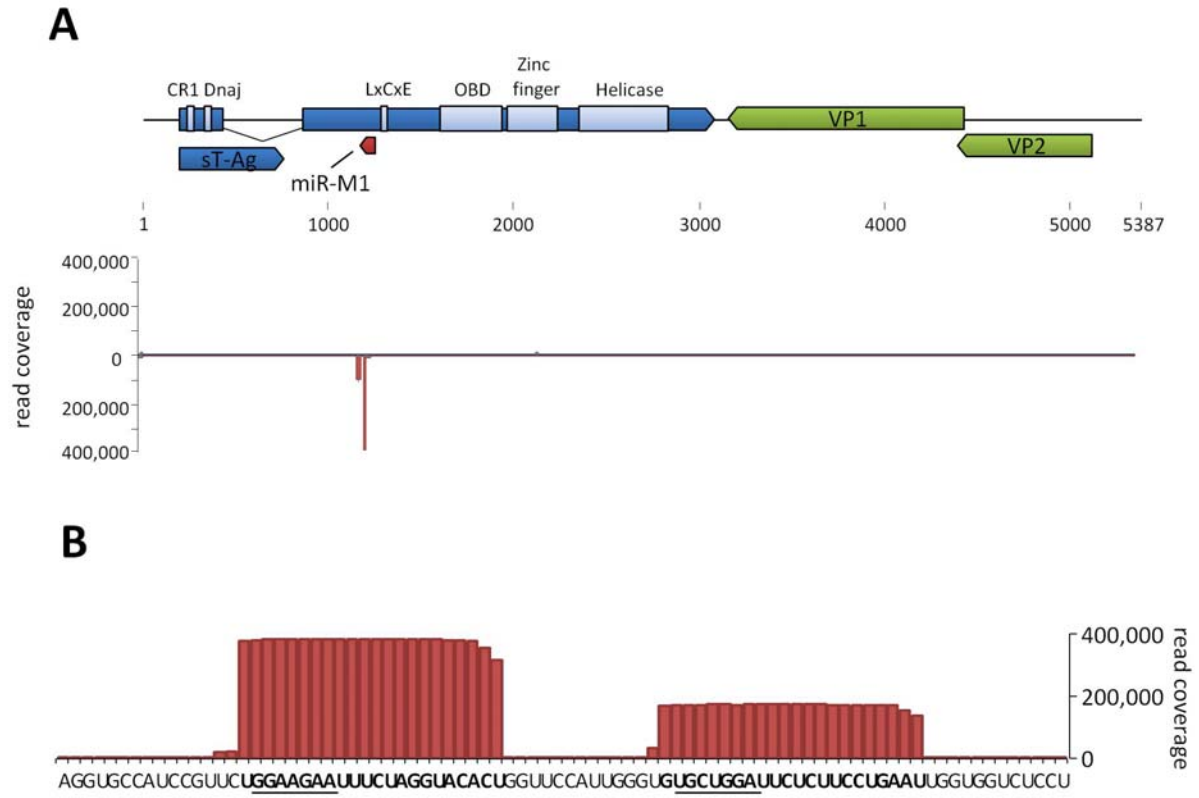

Figure 8: Expression of mcv-miR-M1 from replicating MCVSyn
(A) Top panel: Schematic depiction of the MCVSyn genome with open reading frames for early (blue) and late (green) genes. The mcv-miR-M1 hairpin is shown as red arrow. Lower panel: coverage track of small RNAs mapping to MCVSyn in early (positive axis) or late (negative axis) orientation. Sequencing reads cover the predicted 5p and 3p mature mcv-miR-M1 whereas the 5p mature miRNA is detected at a higher frequency compared to the 3p miRNA. **(B)** Detailed read coverage at the mcv-miR-M1 hairpin (late orientation). The mature miRNA sequence is highlighted in bold letters and the seed region is underlined.

The read coverage on the MCVSyn sequence in Figure 8 shows that only two mature miRNAs are expressed from the viral genome which correspond to the published sequences of the 5p and 3p mature mcv-miR-M1 described by Seo et al. [51]. The 5p mature miRNA ranked among the top 10 of all sequenced miRNAs while the 3p miRNA was approximately two fold less abundant but still ranked at position 14 of all detected miRNAs (Table 47). Low amounts of additional reads mapping to MCVSyn, not visible at the scale in Figure 8A, were found randomly distributed over the viral genome and therefore can be considered as representing break-down RNA fragments.

Figure 8B shows a detailed view of the read distribution on the pre-miRNA sequence and reveals that the majority of miRNA sequences correspond to the $5p_{17\text{-}23}$ miRNA variant, although low coverage of nts 15 and 16 of the pre-miRNA hairpin is detectable as well.

Small RNA-Seq libraries were further prepared from PFSK-1 cells transfected with a mcv-miR-M1 expression plasmid (pCDNA3.1 mcv-miR-M1) as well as from the MCC cell lines MKL-1 and WaGa. All libraries were prepared from total RNA using the NEBnext small RNA Sequencing kit and were sequenced on the Illumina HiSeq platform.

Remarkably, in comparison to the other samples, MCVSyn was found to express the highest levels of mcv-miR-M1 even despite the low amount of cells harboring MCVSyn (3-5%) compared to transfection with pCDNA3.1 mcv-miR-M1 (40-50%). Expression of mcv-miR-M1 in the MCC cell lines MKL-1 and WaGa was generally very low. Overall, the 5p mature miRNA was expressed at significantly higher levels than the 3p miRNA (Figure 9).

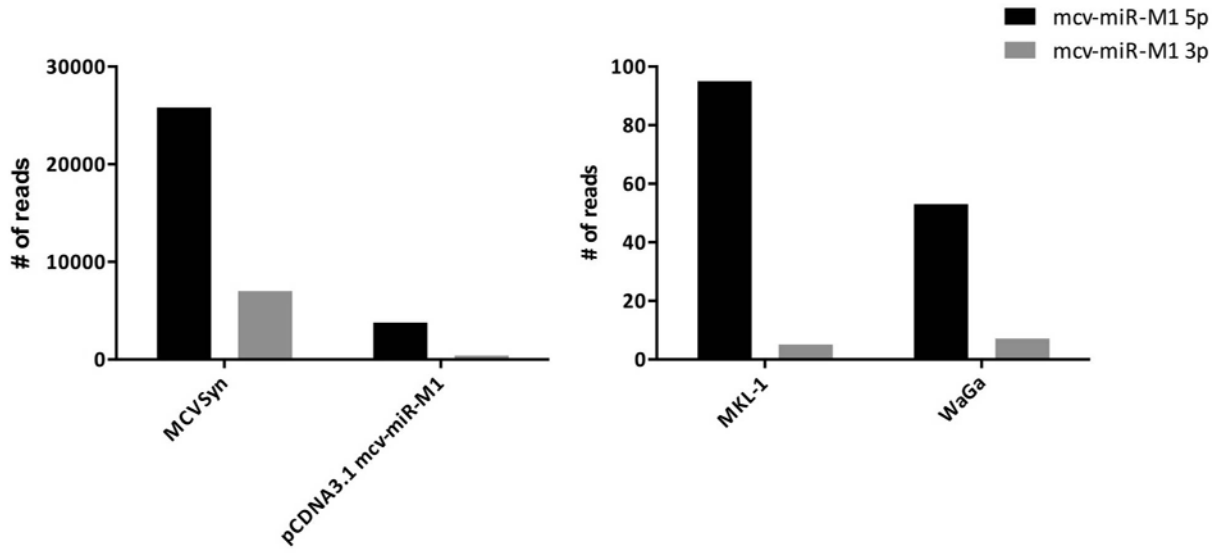

Figure 9: Total number of small RNA-Seq reads for mcv-miR-M1 5p and 3p
The total read count for sequences mapping to the 5p or 3p region of pre-mcv-miR-M1 is shown for small RNA-Sequencing using the NEBnext small RNA library preparation kit. In all samples, the 5p mature mcv-miR-M1 is much higher expressed than the 3p miRNA. Highest expression of mcv-miR-M1 can be observed in PFSK-1 cells transfected with re-ligated MCVSyn. In MCC cell lines, expression of mcv-miR-M1 was generally very low.

Closer analysis of the sequence of mcv-miR-M1 reads revealed that several variants of the 5p and 3p mature miRNA, so-called isomiRs, were captured which are shifted by a few nucleotides with respect to the first described sequences. Variations at the 3' ends of miRNAs are known to result from variable Dicer processing or degradation. Since they do not generally influence the miRNA function, 3' end variations of mv-miR-M1 were not closer examined. However, isomiRs which differ in their 5p end, have different target defining seed-regions and can thus target different mRNAs. To differentiate between mcv-miR-M1 seed variants, miRNA reads were

grouped in respect to their seed sequences. In all samples, the predominant mcv-miR-M1 species were the $5p_{17-23}$ and $3p_{52-58}$ miRNAs, although the $5p_{15-21}$ miRNA could be identified as secondly most abundant species. A 3p mcv-miR-M1 isomiR shifted by 1 nt ($3p_{51-57}$) accounted for approximately 10% of all 3p miRNA reads (Figure 10A). These results strongly suggest that the $5p_{17-23}$ miRNA variant is the predominant mature mcv-miR-M1 species in MCC cells as well as during viral replication.

After reports about biases in small RNA-Seq experiments caused by different small RNA-Seq library kits [180-185], a second miRNA library kit, the Illumina TruSeq small RNA library kit, was used to produce libraries from the same material as previously sequenced. Sequencing of these libraries resulted in a reversed pattern of seed distributions of 5p mcv-miR-M1 isomiRs. In cells transfected with pCDNA3.1 or re-ligated MCVSyn as well as in the MCC cell lines the $5p_{15-21}$ miRNA was the predominant form with a relative abundance of more than 90% (Figure 10B, left panel). The 3p mcv-miR-M1 isomiR $3p_{51-57}$ was detected at higher frequency using the Illumina library preparation protocol (Figure 10B, right panel).

Notably, although the distribution of mcv-miR-M1 isomiRs significantly differed between datasets obtained from the TruSeq and NEBnext library preparation methods, in both cases, the distribution of mcv-miR-M1 5p isomiRs was identical between replicating and integrated MCPyV (Figure 10).

The explanation for the ambiguous results using different library preparation methods can be ascribed to differences in the small RNA library preparation protocols, which introduce a strong kit-dependent bias and enrich for certain miRNA species dependent on their 5p end base composition [180,182,184-186]. Generally, results obtained from the NEBnext small RNA library preparation kit are described to be more accurate and usually match the results from miRNA identification by cloning [184].

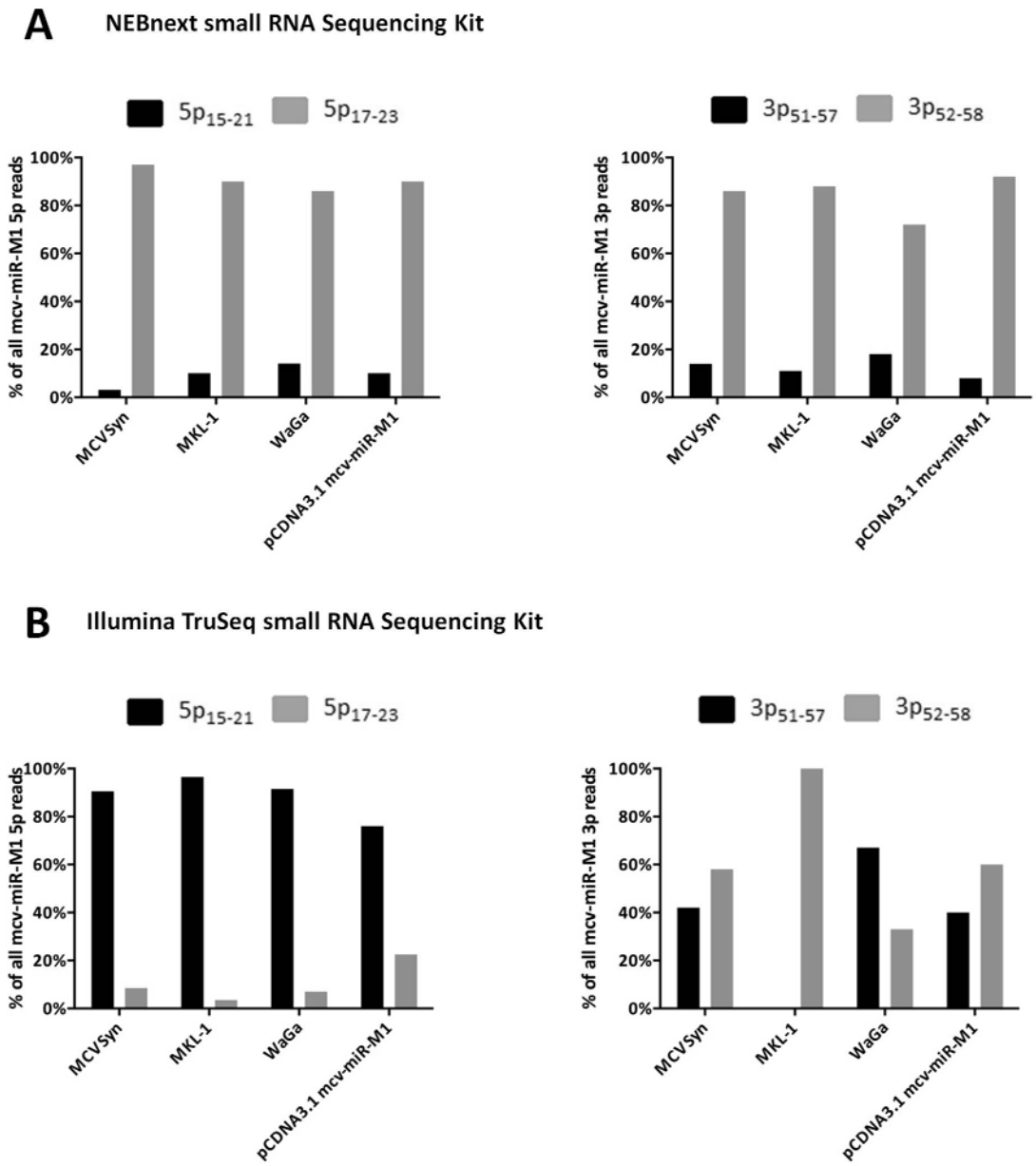

Figure 10: Distribution of mcv-miR-M1 isomiRs using different small RNA-Seq library preparation mehods
Small RNA-Seq results obtained by **(A)** NEBnext small RNA Sequencing library preparation or **(B)** Illumina TruSeq library preparation from PFSK-1 cells transfected with re-ligated MCVSyn or a mcv-miR-M1 expression plasmid (pCDNA3.1 mcv-miR-M1) as well as from the MCC cell lines MKL-1 and WaGa. The percentage of mcv-miR-M1 5p or 3p seed variants among all reads mapping to the 5p or 3p region of the pre-miRNA hairpin is shown.

For library preparation with either the NEBnext or the TruSeq kit the exact same RNA samples were used and consequently both HTS datasets should display the same miRNA profile. Notably, the expression ranks of mcv-miR-M1 between the TruSeq and the NEBnext kit differ significantly (Table 47) and indicate that the TruSeq kit could not retrieve mcv-miR-M1 as efficiently as the NEBnext kit.

Table 47: Expression ranks of mcv-miR-M1-5p and mcv-miR-M1-3p among all miRNAs in small RNA-Seq NEBnext and TruSeq datasets

	MCVSyn		WaGa		MKL-1	
miRNA species	NEBnext	TruSeq	NEBnext	TruSeq	NEBnext	TruSeq
mcv-miR-M1-5p	7	101	467	529	442	399
mcv-miR-M1-3p	14	159	414	839	416	695

Taken together, the pre-miRNA mcv-miR-M1 can give rise to different isomiRs. Importantly, there is no difference in the expression of mcv-miR-M1 seed variants between replicating MCVSyn and MCC cell lines. This observation argues against the hypothesis of Lee et al. regarding the existence of a tumor specific mcv-miR-M1 expression pattern.

Another interesting observation from the small RNA-Seq results is, that mcv-miR-M1 is highly expressed during replication of MCVSyn (Figure 9 and Table 47) whereas expression in MCC cell lines is very low (Figure 9 and Table 47). These differences are even more significant when taking into account that only 3-5% of all cells in a replication assay harbor replicating MCVSyn whereas MCPyV is integrated into every single cell of MCC cell lines. This observation suggests that the viral miRNA does not play a significant role in MCC tumor cells, but also raises the question how miRNA expression is controlled during viral replication and why it is almost completely silenced in MCC cells harboring integrated MCPyV. The regulation of mcv-miR-M1 expression is addressed in the following chapter.

4.1.2 Mapping of 3'UTRs of MCVSyn transcripts reveals inefficient transcriptional termination of late transcription as a possible mechanism for expression of mcv-miR-M1

Even before the first viral miRNAs were identified, murine Polyomavirus (MuPyV) was shown to produce read-through transcripts from the late gene region which caused downregulation of early gene expression [155,158]. Ever since the identification of polyomavirus encoded miRNAs, it was assumed that they were generally produced from these late read-through transcripts [52]. Although being a plausible hypothesis given the location of polyomavirus miRNAs in the orientation of late transcription and the observation that, at least during a productive infection, SV40 miRNAs are expressed at late time-points in infection, this hypothesis has thus far not been experimentally investigated.

To find out if this mechanism also plays a role in the expression of mcv-miR-M1 from replicating MCVSyn, the presence of inefficiently terminated late transcripts which proceed over the miRNA locus was examined by 3'RACE analysis.

3'RACE was performed using an anchored polyT primer and gene specific nested primer sets for early and late transcripts (Figure 12). PCR products (Figure 11) were cloned into pCR2.1 vector and individual clones were subjected to Sanger sequencing.

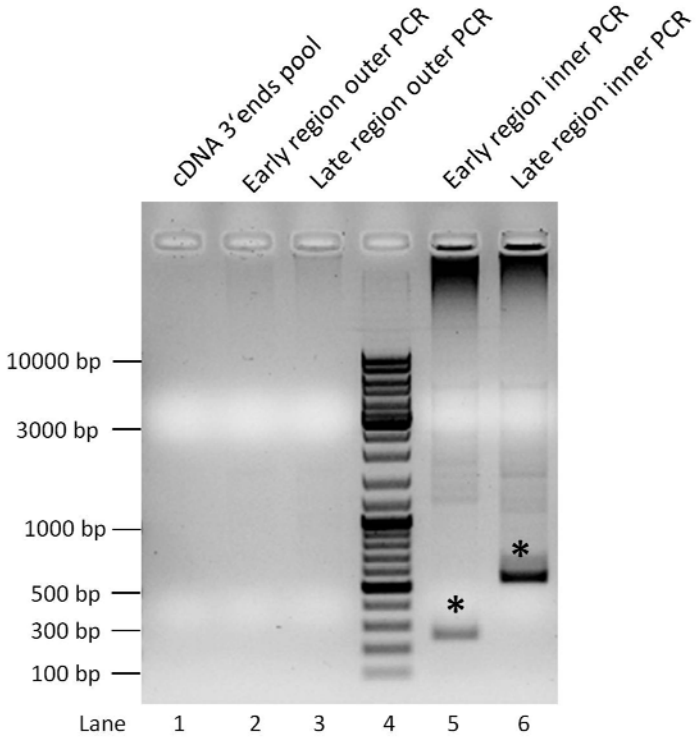

Figure 11: Agarose gel analysis of 3'RACE products
3'RACE PCR products were separated on a 1.2% agarose gel. No distinct band is visible after cDNA-synthesis (lane 1) and outer PCR (lanes 2 and 3). After inner (nested) PCR, a major amplicon of approximately 250 nt was identified for the early viral region and a PCR product of approximately 500 nt was detectable for the late gene region (lanes 5 and 6, marked with an asterisk).

The eukaryotic core polyadenylation signal (pA-signal) comprises a highly conserved AAUAAA motif 10–30 nucleotides upstream of the cleavage site (usually a CpA dinucleotide) and a more diffuse GU-rich downstream sequence element [131,132,187]. In the viral early orientation, one single pA signal overlaps with the early region stop codon and further downstream, a GU-rich region can be identified (Figure 12A and B). All analyzed 3'RACE clones for the early region contained a 3'UTR of only 13 nt between the early region stop codon and the start of the pA-tail (pA site E, Figure 12B).

Similar to MuPyV and SV40 [154,155,158,187,188], two pA-signals and slightly GU enriched regions can be found downstream of the stop codon of the late coding region (Figure 12A and C). However, these signals seem to be inefficient, as is evidenced by the observation of read-through transcripts. About half of all 3'RACE clones from the late orientation terminate at the pA site L1, downstream of the first pA-signal (Figure 12C, upper panel). 8 of 26 3'RACE clones contained a

3'UTR which proceeds behind the first pA-signal and terminates downstram of the second pA-signal at the pA site L2 (Figure 12C, lower panel). The remaining clones of late 3'RACE products were comprised of A-rich sequences further downstream of the second pA site L2, indicating a mispriming of the polyT primer on read-through transcripts that extend beyond both late pA-signals. These long late transcripts could serve as pri-miRNA substrates for Drosha cleavage and could thus be processed into mature mcv-miR-M1.

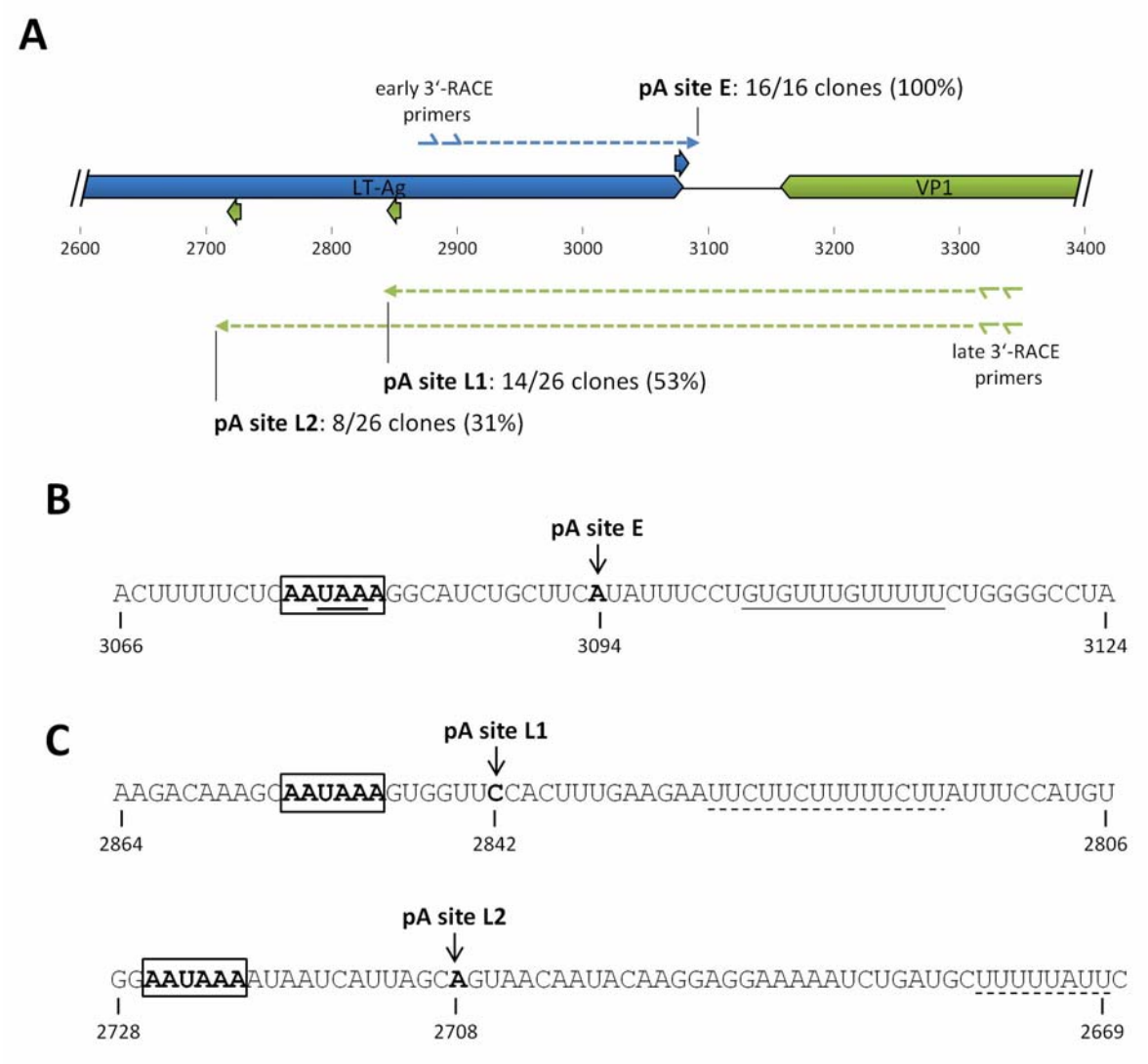

Figure 12: Mapping of 3'UTRs in MCPyV early and late transcripts using 3'RACE analysis
(A) Schematic overview of 3'RACE results. pA signals are depicted as bold arrows in the early (blue) and late (green) orientation. Positions of nested primers for 3'RACE are indicated as double arrows. Dashed arrows show positions of 3'RACE PCR products and resulting polyA sites (pA site E and pA sites L1 and L2) as identified by Sanger sequencing. The abundance of 3'RACE sequences is given in total number of clones and percentage of all sequences in early or late orientation, respectively. **(B and C)** Detailed positions of pA sites in early (B) and late (C) orientation. pA sites are shown in bold letters and are marked by an arrow. pA signals are highlighted by boxes and GU-rich regions are underlined. In (B), the early region stop-codon which overlaps with the boxed pA signal is underlined.

4.1.3 mcv-miR-M1 can be expressed by NCCR-independent transcription

To investigate if all viral transcripts originate in the NCCR, transcriptional start sites (TSS) of viral mRNAs from PFSK-1 cells 4 d after transfection with re-ligated MCVSyn were mapped using a cap-dependent 5'RACE approach. Due to the low abundance of viral transcripts, gene specific primers for the early and late transcripts were used (Figure 14A).

PCR products from RACE analysis, as shown in Figure 13, were first cloned and analyzed by Sanger sequencing. However, especially in the late orientation, PCR products were found to be markedly divergent in size. Due to the relatively low number of 5'RACE PCR products that can be analyzed by this method, is was not possible to precisely define major as well as minor transcriptional start sites. To analyze a higher number of 5'RACE PCR products, HTS libraries were prepared from early and late region 5'RACE PCR products using the NEBnext Ultra DNA library kit for Illumina sequencing, omitting the fragmentation step. Reads were mapped to MCVSyn and the reads starting at each individual nucleotide position were counted.

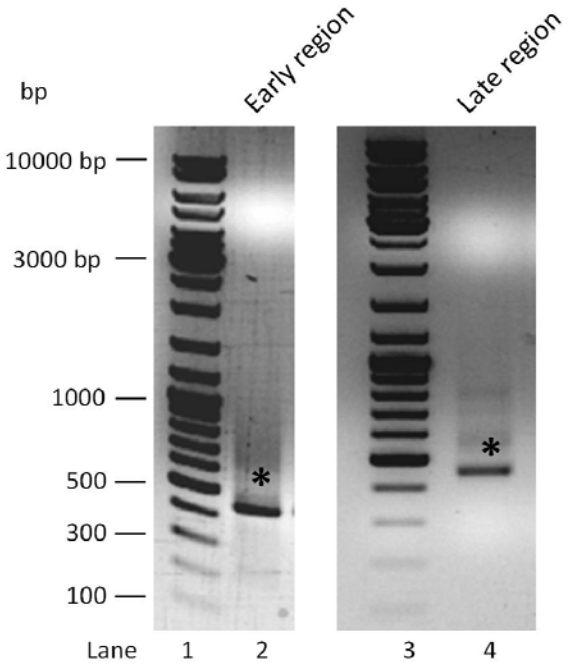

Figure 13: Agarose gel analysis of 5'RACE products
5'RACE derived amplicons were separated on a 1.2% agarose gel, showing each one major amplicon of approximately 450 bp for early (lane 2, marked with an asterisk) as well as for late transcripts (lane 4, marked with an asterisk).

In the early orientation, the majority of reads started 47 nt upstream of the T-Ag start codon, 24 nts downstream of a canonical TATA-box with most of the remaining reads starting only 1 nt up- or downstream of this position (Figure 14B and C). A second distinct TSS was identified upstream of the TATA-Box. All transcripts starting at this initiation site were found to be spliced

to a splice acceptor downstream of the early region start codon. These transcripts could lead to the expression of ALTO as described in chapter 4.2.2.

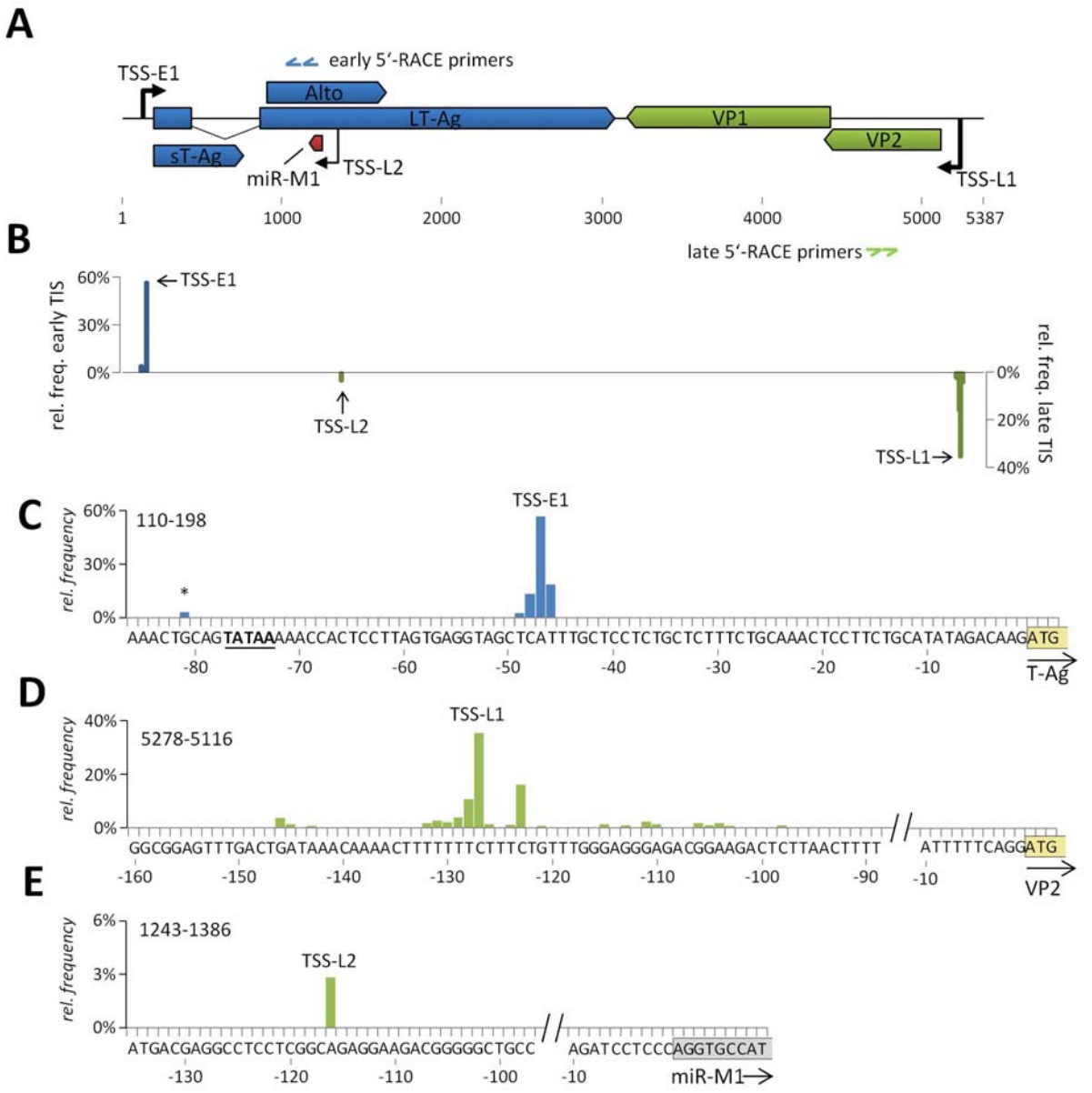

Figure 14: Identification of transcriptional start sites (TSS) for early and late gene expression by HTS analysis
(A) Positions of major TSS (TSS-E1, TSS-L1 and TSS-L2) on the MCVSyn genome as identified by 5'RACE analysis. Positions of nested gene specific primers are indicated by double arrows. **(B)** Positions and frequencies of TSS in early (blue) and late (green) orientation as obtained by HTS. **(C-E)** Details of read coverage at the three major TSS regions ((C) TSS-E1, (D) TSS-L1 and (E) TSS-L2). The frequency of reads is shown relative to all reads in early or late orientation. Positions of reads are shown relative to the nearest start codon (C and D) or the pre-miRNA (E). The TATA-Box in the early region is marked by bold letters and a second TSS in the early region is marked with an asterisk. For all graphs, the absolute position on the MCVSyn genome is given in the upper left corner.

In the late orientation the exact site of transcriptional initiation is less clearly defined with the majority of reads starting in a region approximately 100-150 nt upstream of the late region AUG (Figure 14D). In this case, the RNA polymerase seems to be less accurately positioned, resulting in dispersed TSS as it is typically observed for TATA-less promoters [77]. Interestingly, a single additional distinct transcriptional initiation site was found outside of the late NCCR associated TSS cluster, 116 nt upstream of the pre-miRNA hairpin (Figure 14E), suggesting that mcv-miR-M1 might be expressed from a NCCR independent transcript.

To further investigate this hypothesis, the early coding region of MCVSyn was cloned into a mammalian expression plasmid with mcv-miR-M1 lying in sense (pCMV:ER-AS) or antisense (pCMV:ER-S) orientation relative to the CMV-promoter (Figure 15A). As expected for CMV-promoter driven transcription, mcv-miR-M1 is strongly expressed from pCMV:ER-AS as observed in Northern blot analysis and SL RT-qPCR two days after transfection. Unexpectedly, comparable amounts of mcv-miR-M1 were also produced from pCMV:ER-S, in which the major transcriptional direction should be antisense to the miRNA locus (Figure 15B and C). Thus, antisense transcription might play a role in expression of mcv-miR-M1 e.g. by enhancing the accessibility of the miRNA promoter. Additionally, transfection of the early coding region alone, either as PCR product or in context of a promoter-less plasmid (pER-AS) also resulted in measurable expression of mcv-miR-M1, although to a significantly lower amount (Figure 15B and C).

These results suggest that mcv-miR-M1 can be expressed independently from promoter regions at the viral NCCR, presumably from an intrinsic miRNA-promoter which is embedded in the early region. Moreover, transcription of the early region might further stimulate mcv-miR-M1 expression.

To analyze if exogenous promoter independent expression of mcv-miR-M1 is dependent on RNA polymerase II, PFSK-1 cells, 24 h post transfection with the pER-AS plasmid, were treated with α-amanitin, a potent inhibitor of RNA-pol II activity [189] for 24 h. Transcript levels of mcv-miR-M1, GAPDH and tRNA methionine were measured by RT-qPCR. In comparison to cells that were not treated with α-amanitin, the transcript levels for GAPDH and mcv-miR-M1 but not for the RNA-pol III dependent tRNA methionine, were significantly decreased, indicating a RNA-pol II dependent transcription mechanism for mcv-miR-M1 in absence of the NCCR (Figure 15D).

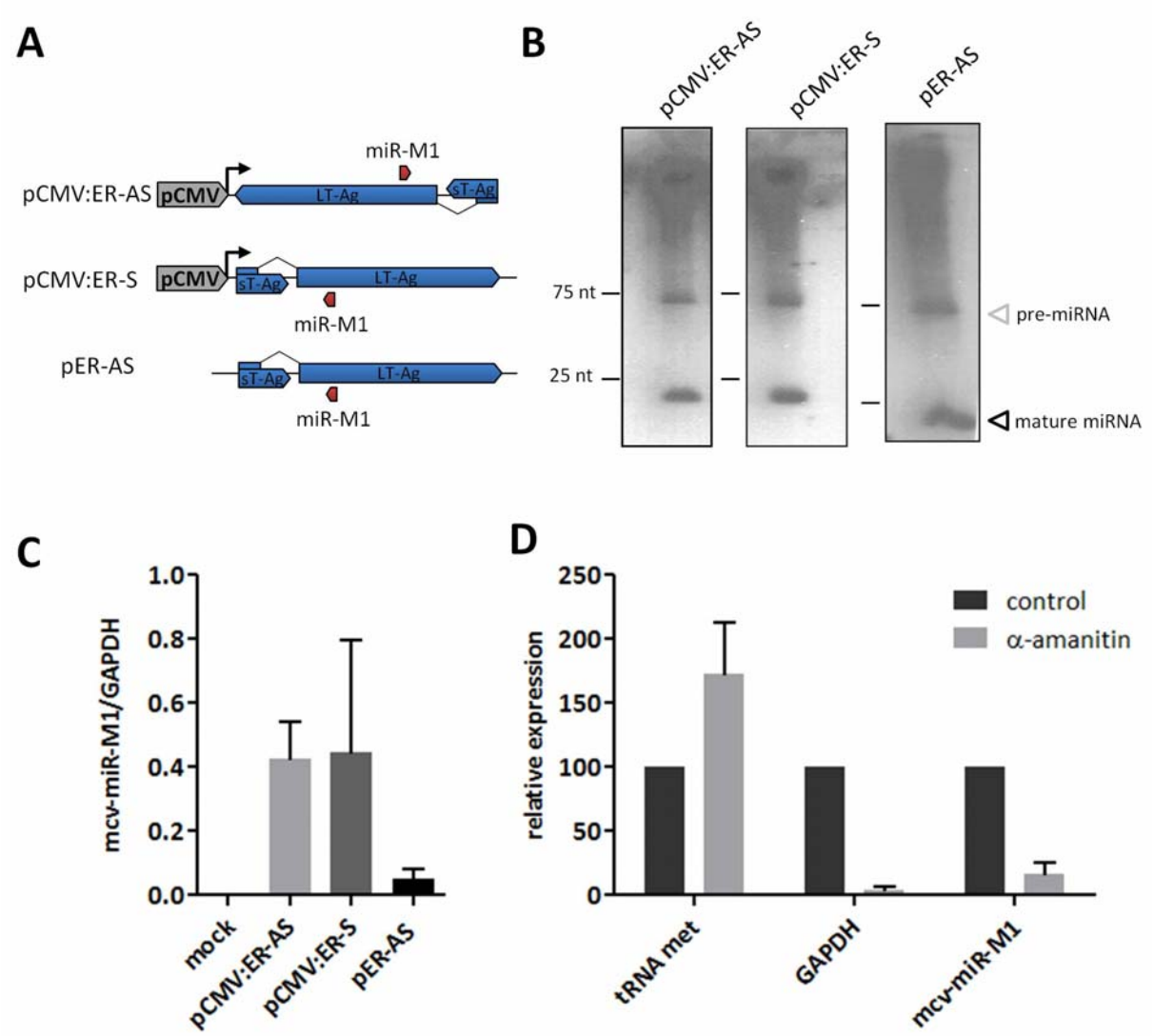

Figure 15: Late transcription independent expression of mcv-miR-M1
(A) Constructs for ectopic expression of mcv-miR-M1. The early coding region of MCVSyn was cloned into a pCMV2b plasmid either in sense (pCMV:ER-S) or anti-sense (pCMV:ER-AS) to the CMV promoter. pER-AS contains the same viral region in a pCR2.1 plasmid without promoter region. **(B)** Small RNA Northern blot analysis of mcv-miR-M1 expression in the context of the early viral region in the presence or absence of a promoter sequence 2d.p.t. in HEK293 cells. Positions of the pre- and mature miRNAs are indicated by arrows. **(C)** Stem-loop RT-qPCR detecting miRNA expression 2 d.p.t. in HEK293 cells transfected with different miRNA expression constructs. **(D)** Quantification of tRNA-methionine, GAPDH transcripts and mcv-miR-M1 by SL RT-qPCR 2 d.p.t. of PFSK-1 cells with pER-AS. 24 h after transfection, cells were treated with α-amanitin for 24 h. Transcript levels are shown relative to cells that were not treated with α-amanitin (control) (n=3).

4.1.4 ChIP-Seq analysis of replicating MCVSyn detects activating histone modifications at the NCCR and upstream of mcv-miR-M1

The identification of a mcv-miR-M1 specific transcriptional start site as well as the observation that mcv-miR-M1 can be expressed in the absence of the viral promoters at the NCCR strongly suggest that mcv-miR-M1 can be transcribed by a so far unannotated promoter, which is

embedded in the early viral region. To further test this hypothesis, replicating MCVSyn was investigated for the presence of the activating histone modification tri-methylation of Lysine 4 at Histone 3 (H3K4-me3), which is often associated with open chromatin structures and transcriptional activity. Using chromatin-immunoprecipitation (ChIP) in combination with HTS analysis (ChIP-Seq), replicating genomes of MCVSyn in PFSK-1 cells were examined for the presence of H3K4-me3 modification. Mapping of HTS reads on the MCVSyn genome revealed two distinct peaks of H3K4-me3 (Figure 16B). As expected, the promoter regions for early and late gene expression at the NCCR displayed H3K4-me3 marks, indicating transcriptional activity in this region. But, interestingly, a second peak of H3K4-me3 was also clearly discernible upstream of the pre-miRNA sequence of mcv-miR-M1, an observation which strongly supports the hypothesis of a miRNA promoter activity in this region.

To investigate, if the mcv-miR-M1 promoter function can be disrupted by sequence alterations, a MCVSyn promoter mutant (MCVSyn-pmt) was designed which is mutated in every third nucleotide of the LT-Ag ORF in a region of 226 nt upstream of mcv-miR-M1 to preserve the aa sequence of LT-Ag. In comparison to wildtype MCVSyn, the MCVSyn-pmt mutant showed decreased levels of H3K4-me3 in the miRNA promoter region (Figure 16C). This result suggests that the mutated region upstream of mcv-miR-M1 contains important factors for promoter activity such as transcription factor binding sites. Scanning this region for human transcription factor binding sites using JASPAR software [134] revealed that 191 putative binding sites with a score threshold of > 80% for 62 different transcription factors are present in the 226 nt long region upstream of mcv-miR-M1. After introduction of pmt-mutations, 22 of the 62 transcription factors would be expected to no longer bind to the putative promoter region. However, many of the detected transcription factor binding motifs are rather short and degenerated and it remains to be investigated if they play a role in miRNA transcription.

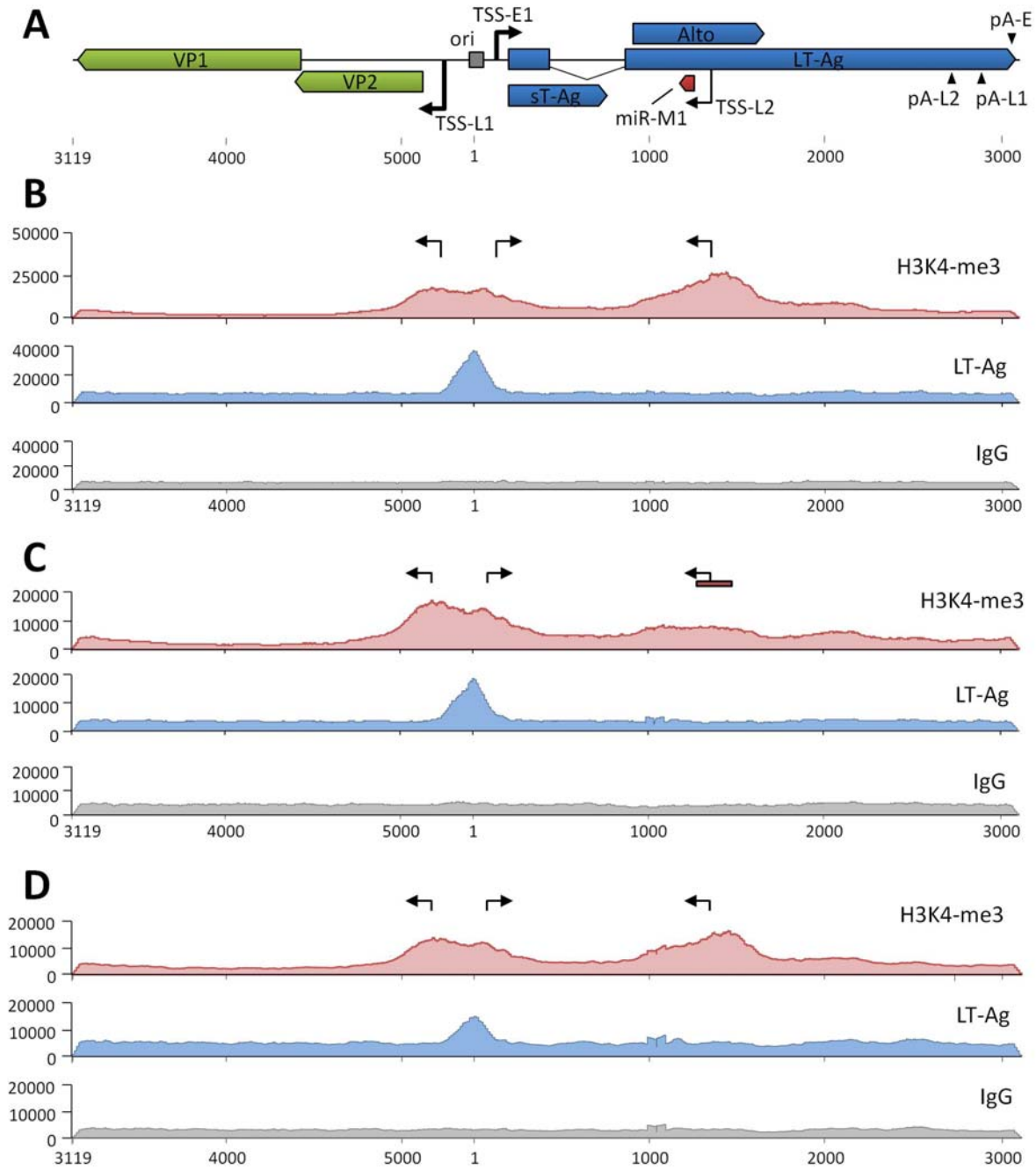

Figure 16: ChIP-Seq analysis for H3K4-me3 and LT-Ag on replicating MCVSyn, MCVSyn-pmt and MCVSyn-ltb
(A) Schematic illustration of the MCVSyn genome including positions of transcriptional start sites (TSS-E1, TSS-L1 and TSS-L2) and of pA sites (pA-E, pA-L1 and pA-L1) **(B-D)** Mapped reads after ChIP-Seq analysis of replicating MCVSyn (B), MCVSyn-pmt (C) and MCVSyn-ltb (D) for H3K4-me3 (red), Large T-Ag (blue) and IgG (grey). Transcriptional start sites are marked by arrows. The mutated region in MCVSyn-pmt is marked by a red bar.

Yet, while H3K4-me3 at the miRNA promoter of MCVSyn-pmt seems to be nearly abolished, expression of mcv-miR-M1 from replicating MCVSyn was observed to be reduced only by half (Figure 17), suggesting that mcv-miR-M1 expression is not solely mediated by the identified

miRNA promoter but that a second expression mechanism is simultaneously active. 3'RACE analysis revealed that late transcripts are not efficiently terminated and that they can proceed over the miRNA locus (Figure 12). These long late transcripts might serve as pri-miRNAs and be processed into mature miRNAs, as suggested for other polyomavirus miRNAs. Thus, expression of mcv-miR-M1 in replicating MCPyV seems to be driven by synergistic activity of the late promoter and a newly discovered autonomous miRNA promoter (Figure 34A).

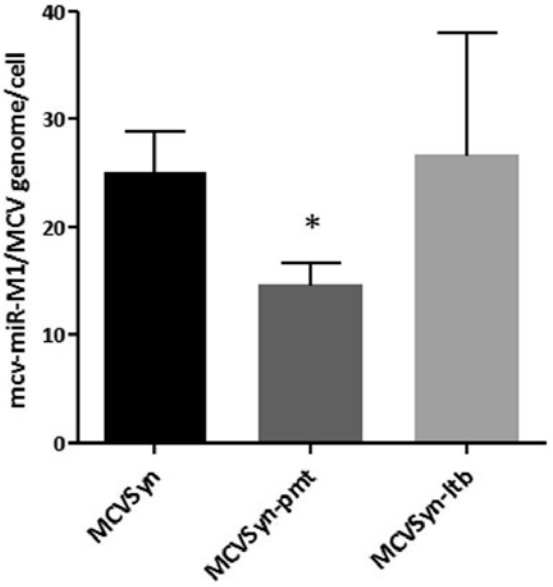

Figure 17: Expression of mcv-miR-M1 in MCVSyn-pmt and MCVSyn-ltb
SL RT-qPCR for detection of mcv-miR-M1 expression in PFSK-1 cells 4 d.p.t. with MCVSyn, MCVSyn-pmt or MCVSyn-ltb. miRNA expression is shown relative to viral copy number per cell (n=3).

Interestingly, besides several transcription factor binding sites, the putative miRNA promoter region contains 6 GRGGC motifs (Figure 18B), which are known binding sites for LT-Ag. A cluster of GRGGC motifs is located at the viral origin of replication and binding of LT-Ag at these sites is required for viral replication [64] (Figure 18A). Notably, the accumulation of GRGGC motifs at the miRNA promoter region is the only significant cluster of GRGGC motifs besides the origin of replication, suggesting that LT-Ag might influence miRNA-expression by binding in this region. The GRGGRC sites upstream of mcv-miR-M1 are also conserved in the primate polyomaviruses GggPyV1 and PtvPyV2a, suggesting an evolutionary conserved and hence important function. Notably, these viruses are close relatives to MCPyV (Figure 1) and their miRNAs are orthologues of mcv-miR-M1 [50,190].

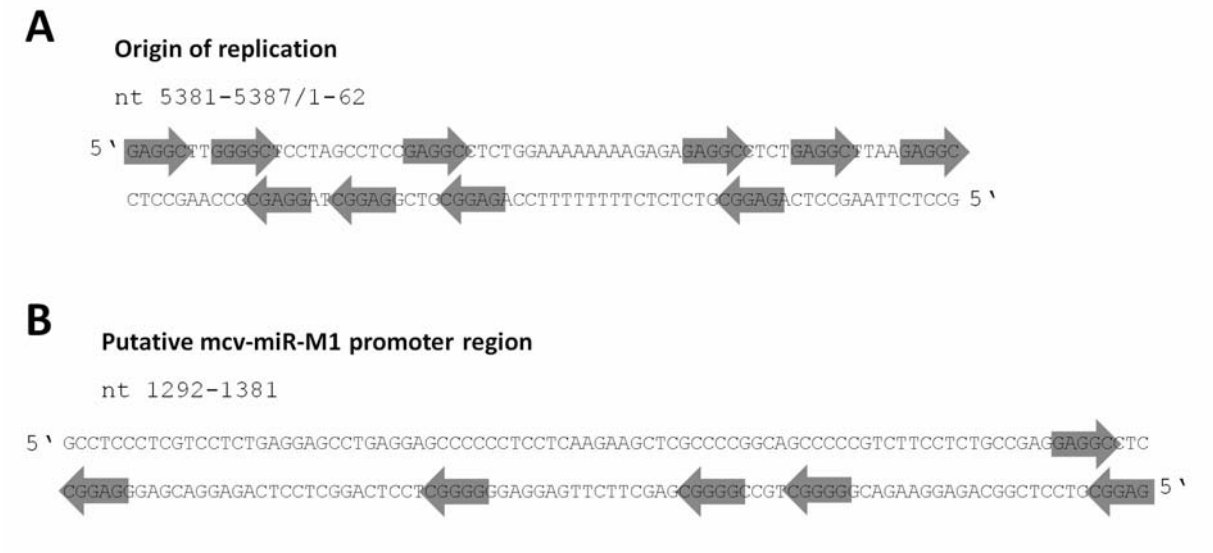

Figure 18: Clustering of LT-Ag binding sites at the origin of replication and upstream of mcv-miR-M1
(A) Positions of GRGGC motifs at the origin of replication and **(B)** upstream of the mcv-miR-M1 pre-miRNA. Sequences are shown as double strands with LT binding sites highlighted by grey arrows.

To investigate if LT-Ag influences the putative miRNA promoter region, a MCVSyn mutant lacking the LT-Ag binding sites upstream of the miRNA hairpin was created (MCVSyn-ltb), without altering the LT-Ag amino acid (aa) sequence. ChIP-Seq analysis of this mutant showed that H3K4-me3 was not impaired in this mutant in comparison to the wildtype (Figure 16C).
To directly investigate if endogenous LT-Ag binds to replicating viral genomes, ChIP-Seq with the LT antibody CM2B4 was carried out for MCVSyn and the MCVSyn mutants. This experiment revealed that LT-Ag binds exclusively at the viral origin of replication in MCVSyn as well as in the MCVSyn mutants (Figure 16B-D). No binding of LT-Ag could be observed at the mcv-miR-M1 promoter region. However, the affinity of the α-LT-Ag antibody CM2B4 is rather low [191] and it remains possible that binding of LT-Ag at the miRNA promoter region is weaker compared to the origin of replication or that LT-Ag binding plays a role during a different phase of the viral life cycle.
The expression of mcv-miR-M1 was not reduced upon introduction of pmt mutations in comparison to the wildtype (Figure 17), suggesting that, at least in the employed replication system and cell type, the GRGGC motifs in the miRNA promoter region are not required for mcv-miR-M1 expression

4.2 Elucidating the role of mcv-miR-M1 in viral replication

The previous chapter showed that mcv-miR-M1 is strongly expressed from replicating but not from integrated MCPyV and that the miRNA can be expressed independently from the NCCR-driven late gene expression, indicating that it plays a role in the absence of particle production, which might be the case during early phases of the viral life cycle or during non-productive persistence. The development of MCPyV replication systems allows investigation of the role of mcv-miR-M1 during early phases of the viral life cycle and unexpectedly also during long-term maintenance of viral episomes. The following chapters investigate the biological function of mcv-miR-M1 by characterization of a MCVSyn miRNA knockout virus mutant.

4.2.1 A mcv-miR-M1 knockout mutant exhibits stronger LT-Ag expression and increased viral DNA replication

All hitherto known polyomavirus encoded miRNAs are located in antisense orientation to the early gene region [50-53,147,179]. This leads to a perfect base pairing between polyomavirus miRNAs and early transcripts, suggesting that PyV miRNAs downregulate early gene expression, which has been experimentally confirmed for SV40, MuPyV, BKPyV and JCPyV miRNAs [51-53,147]. It was also shown that ectopically expressed mcv-miR-M1 has the potential to mediate downregulation of a complementary region in a luciferase assay [51]. However, due to the lack of a model system for MCPyV, the influence of mcv-miR-M1 on viral early gene expression and its implications for the viral life cycle remained elusive.

In order to investigate the functions of mcv-miR-M1, mutations in the region of the miRNA stem-loop were introduced in the context of MCVSyn without altering the LT-Ag aa sequence. This hairpin mutant of MCVSyn (MCVSyn-hpko) is impaired in folding into the correct miRNA hairpin structure (Figure 19A). Northern blot analysis of small RNAs shows efficient expression of mcv-miR-M1 from MCVSyn whereas no miRNA expression is detectable in case of replicating MCVSyn-hpko (Figure 19B).

Western blot analysis of MCVSyn and MCVSyn-hpko at four days after transfection of PFSK-1 cells shows that the expression of LT-Ag is strongly increased in the hairpin mutant in comparison to the wildtype (Figure 19C). Thus, as expected from its perfect complementarity to early viral transcripts, mcv-miR-M1 indeed downregulates the expression of LT-Ag during authentic viral replication.

Interestingly, the miRNA-mediated reduction of LT-Ag expression has marked influence on viral replication as is evident by the observation that MCVSyn-hpko displays a more robust replication of viral DNA in comparison to the wildtype (Figure 19D). Hence, mcv-miR-M1

indirectly also regulates viral DNA replication as has been previously shown for the BKPyV miRNA [151].

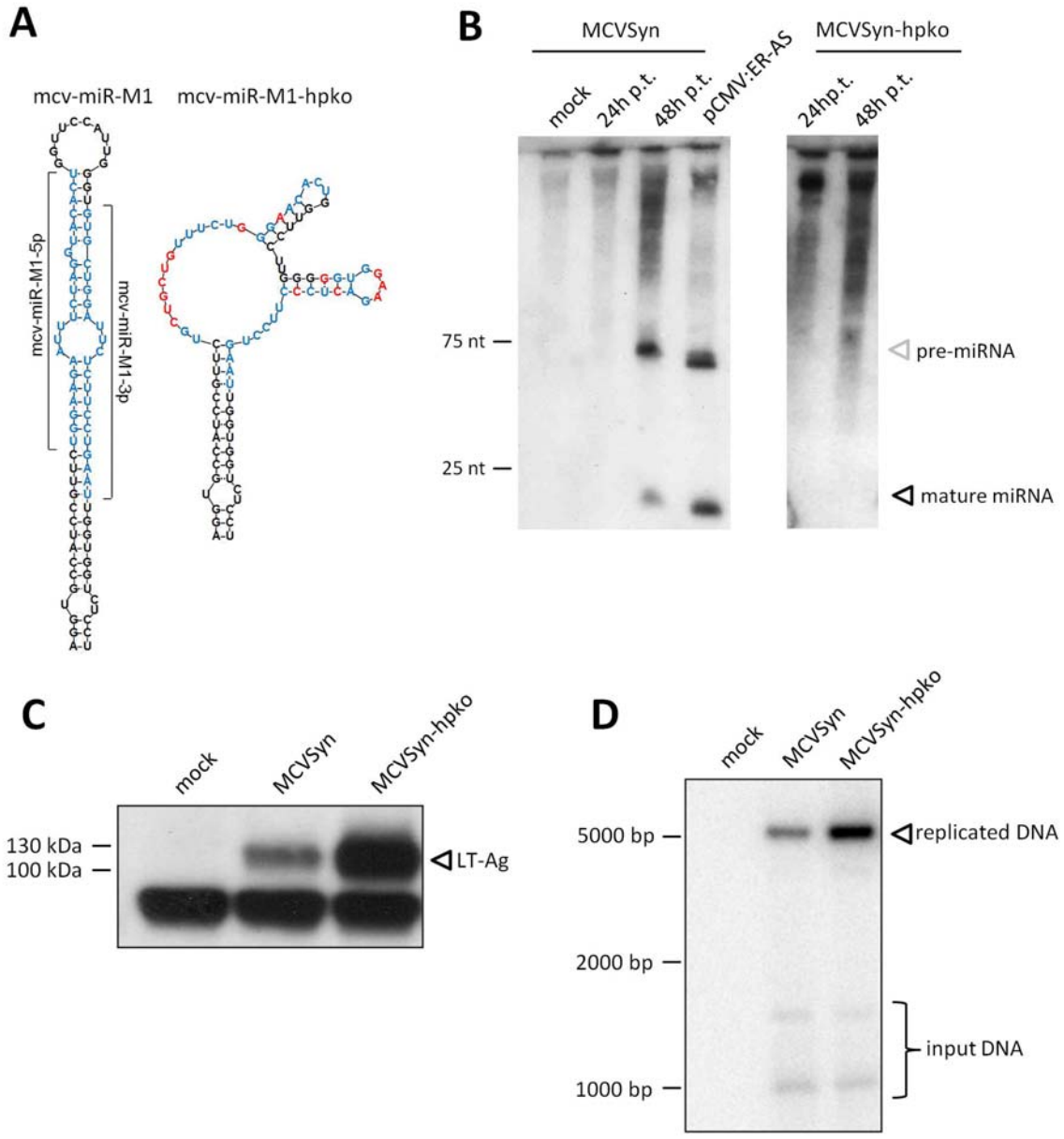

Figure 19: LT-Ag expression and viral DNA replication of MCVSyn and the miRNA mutant MCVSyn-hpko

(A) Predicted pre-miRNA secondary structures (nucleotides 1168-1251) of mcv-miR-M1 and mcv-miR-M1 hpko. **(B)** Small RNA Northern blot analysis 48 h after transfection with MCVSyn or MCVSyn-hpko, using nucleotides 1215-1249 as a probe. mcv-miR-M1 expression is detectable in HEK293 cells transfected with a plasmid containing the miRNA hairpin downstream of a CMV-promoter (positive control) as well as in MCVSyn. HEK293 cells transfected with MCVSyn-hpko do not show mcv-miR-M1 expression 24 or 48 h post transfection. **(C)** Western blot analysis of LT-Ag expression in PFSK-1 cells transfected with MCVSyn or MCVSyn-hpko 4 d.p.t. The band corresponding to LT-Ag is marked by an arrow. **(D)** Analysis of *de novo* replicated viral DNA by Southern blot analysis of HIRT extracts prepared from PFSK-1 cells 4 d.p.t. with MCVSyn or MCVSyn-hpko. Linearized replicated viral DNA is marked with an arrow. Bacterial derived input DNA is sensitive to *DpnI* digestion and is detectable at the bottom of the blot (*input DNA*).

4.2.2 RNA-Seq analysis of replicating MCVSyn and MCVSyn-hpko detects novel splice variants and a putative ORF for a MCPyV agnoprotein

To ensure that the observed phenotype of replicating MCVSyn-hpko was not caused by alteration in the viral transcriptome induced by the introduced hpko-mutations, RNA-Seq was used to analyze and compare the transcriptional profile of MCVSyn and MCVSyn-hpko.

So far, viral transcripts have been characterized by *in silico* methods [12], analysis of MCC samples and ectopic overexpression of early gene cassettes from different MCC tumors ([66], Figure 20D and E). However, it remained unclear if this faithfully reflects the transcriptional profile of non-integrated MCPyV during authentic viral life cycle.

Strand specific RNA-Seq libraries were prepared from total RNA of MCVSyn or MCVSyn-hpko transfected PFSK-1 cells at 4 d.p.t. Sequencing reads were mapped on the MCVSyn sequence and the distribution of reads on the viral genome was plotted as a coverage track showing the relative coverage of each nucleotide for MCVSyn or MCVSyn-hpko, respectively (Figure 20B and C). Both profiles confirm the results of 3' RACE analysis by revealing that transcription in the early orientation efficiently terminates behind the pA-signal, whereas in the late orientation, inefficiently terminated transcripts are found to extend beyond both late pA-signals. Although both profiles look similar at first sight, two major differences can be noted. First, the ratio between early and late transcripts is altered in MCVSyn-hpko, which is either caused by accumulation of early transcripts or reduced late transcription. Second, in MCVSyn-hpko early transcript reads accumulate immediately downstream of the miRNA cleavage site. Both effects can be attributed to the absence of functional mcv-miR-M1, which in MCVSyn reduces early transcript levels and mediates cleavage of early mRNAs as visible by reduced coverage of early transcripts at the miRNA locus.

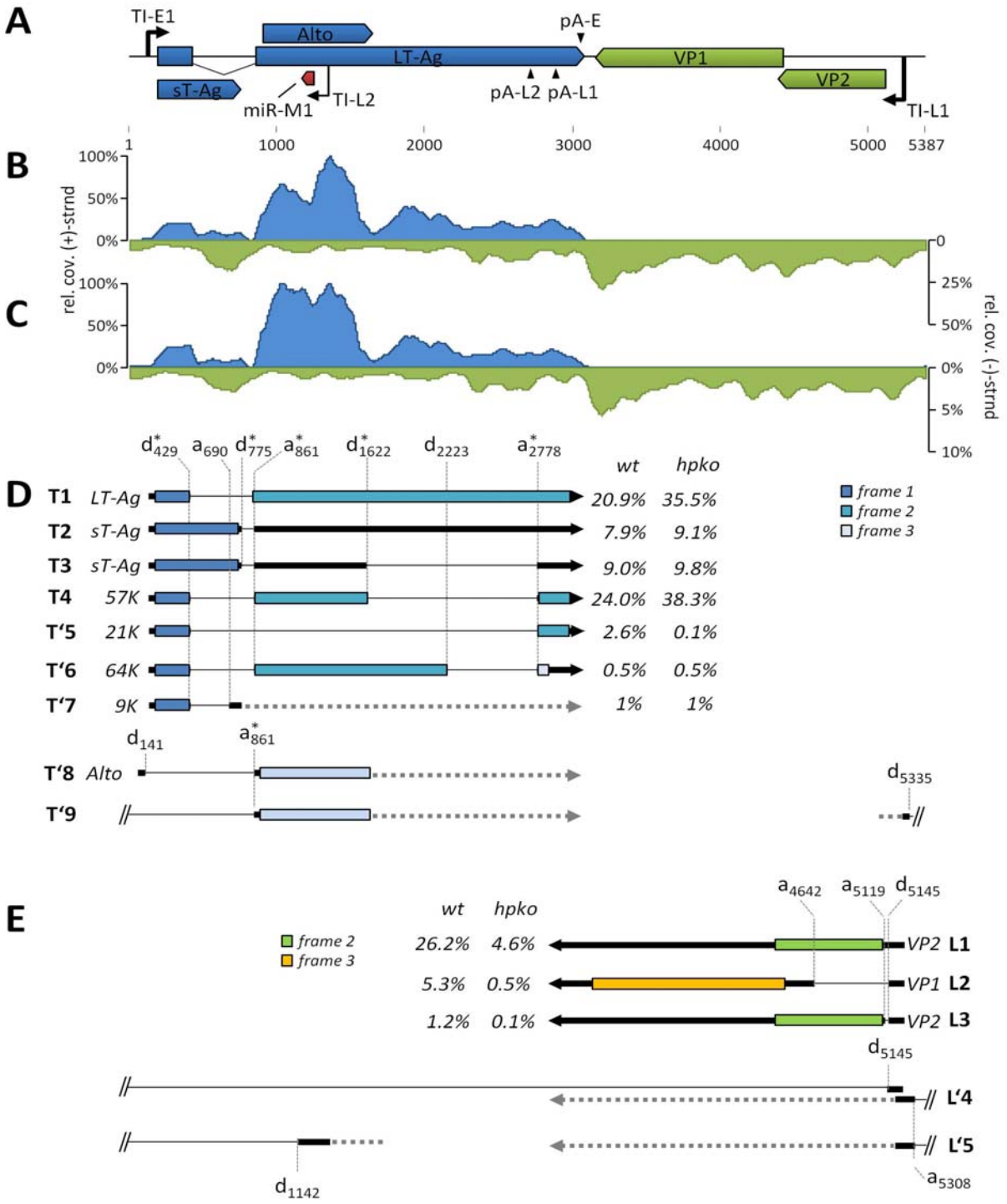

Figure 20: Mapping of MCVSyn transcripts by RNA-Seq
(A) Schematic illustration of the MCPyV genome with the open reading frames encoding the T-Ags (blue), the predicted open reading frames for the structural proteins VP1-3 (green) and the pre-miRNA mcv-miR-M1 (red). pA-sites and TSS are marked by arrows. **(B and C)** Strand specific RNA-Seq coverage of replicating MCVSyn 4 d.p.t. in PFSK-1 cells in early and late orientation for MCVSyn (B) and MCVSyn-hpko (C). **(D and E)** Viral transcripts detected by RNA-Seq in early (D) and late (E) orientation and their relative abundance among all viral transcripts for MCVSyn and MCVSyn-hpko are shown. Splice donors (d) and acceptors (a) were named according to their nucleotide position. Previously identified splice donors and acceptors are marked with an asterisk. ORFs are shown as colored bars and are shaded according to their frame.

Four early viral transcripts (T1-T4 in Figure 20D) were previously identified by RACE and Northern blot analysis of MCC tumors by Shuda et al. [66]. These viral transcripts can be translated into the T-Antigens LT-Ag, sT-Ag and 57kT-Ag. The protein ALTO was described to be expressed by ribosomal leaky scanning of LT-Ag and sT-Ag mRNAs [67].

The described mRNAs T1-T4 were also detected by RNA-Seq of cells transfected with MCVSyn and account for the majority of all viral transcripts in MCVSyn as well as MCVSyn-hpko. In addition, RNA-Seq identified three minor splice sites (T'5-T'7) in the early region with relatively low abundance. The resulting transcripts were named 21K, 64K and 9K according to the calculated size of their expected protein products. All three transcripts use the splice donor d_{429}, which is also used for LT-Ag and 57kT-Ag transcripts, but differ in the usage of 3' splice acceptor sites.

Two additional splice donor sites were identified by sequencing of 5'RACE products. The donor sites at nucleotide positions 141 and 5335 were joined to the acceptor site at position 861, the acceptor site which is also used to produce LT-Ag and 57kT-Ag transcripts. Remarkably, both splice donor sites are located upstream of the major early start codon. The resulting transcripts (T'8 and T'9) therefore lack the major T-Ag translational start site but retain the start codon for the ALTO protein and could serve as ALTO-specific transcripts. The occurrence of these splice events was confirmed by RT-PCR from RNA of replicating MCVSyn (Figure 21).

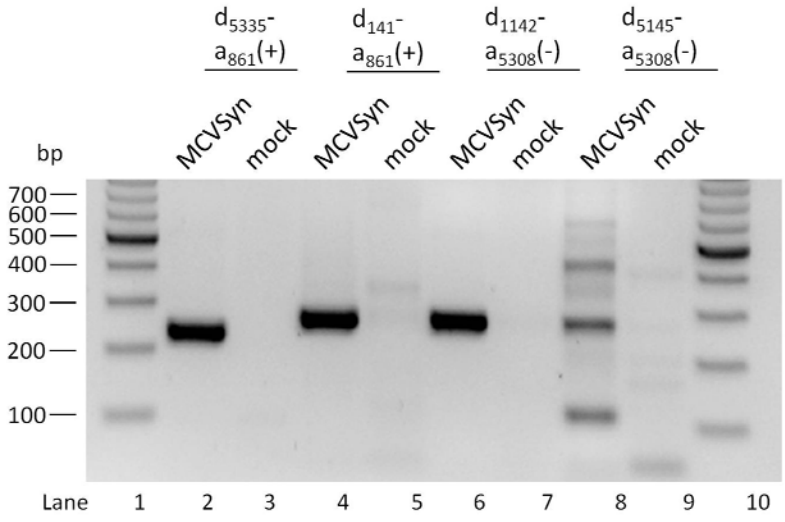

Figure 21: RT-PCR confirms usage of novel splice sites and the occurrence of leader-to-leader splicing

Agarose gel analysis of RT-PCR products from mock or MCVSyn transfected PFSK-1 cells 4 d.p.t. Exon-spanning primers were used to investigate usage of newly discovered splice sites. Splice sites are named by their donor (d) and acceptor (a) positions and their orientation in early (+) or late (-) transcription direction.

So far, transcripts in the late orientation have not been experimentally mapped. It was hypothesized that three capsid proteins VP1-3 are expressed from alternatively spliced

transcripts. RNA-Seq revealed the usage of two splice acceptors and one splice donor in the late orientation (Figure 20E). However, the majority of late transcripts was not spliced and thus is likely translated into the VP2 protein (L1). The acceptor site at position 4642 and the donor site at position 5145 were already predicted by *in silico* analysis and RNA-Seq confirmed that joining of these sites leads to VP1 specific transcripts (L2). Rare usage of a so far unknown acceptor site at position 5119 appears to remove a very small intron of only 26 nt. The resulting transcript L3 seems to be an alternative VP2 transcript. The predicted non-canonical splice for VP3 could not be detected which is in accord with a recent publication by Schowalter et al. [47], showing that MCPyV belongs to a group of polyomaviruses which do not express VP3.

So called leader-to-leader splicing in the late orientation was previously reported for MuPyV [83]. This mechanism requires transcription from the late promoter to proceed over the whole viral genome for multiple rounds. By joining the late leader sequences, almost genome sized introns are removed. Finally, a transcript comprising multiple leader sequences and one copy of the late coding region is generated.

Since late transcription was found to proceed over the whole viral genome, leader-to-leader splicing might also occur in MCVSyn. Mapping of RNA-Seq reads on a single copy of MCVSyn does not allow detection of such transcripts, but mapping on concatenated MCVSyn genomes revealed that leader-to-leader splicing indeed occurs during replication of MCVSyn. The above described donor at position 5145 was found to be joined to an upstream acceptor at position 5308 (L'4). RT-PCR revealed that transcripts with multiple leader sequences are generated during MCVSyn replication (Figure 21). Late transcripts spanning the whole genome might be also spliced from the donor site at 5145 to the acceptor site 5119 thereby removing not only a 26 nt intron but rather a more than genome sized intron. However, since the acceptor site is located downstream of the donor site, this cannot be detected in RNA-Seq data.

Unexpectedly, another donor, immediately downstream of mcv-miR-M1, was observed to be spliced to the acceptor of leader-to-leader splicing at position 5308 (L'5). This miRNA-to-leader splicing was not yet described for any other polyomavirus. Using primers spanning the exon boundaries, this splice event was again confirmed by RT-PCR (Figure 21). Embedded in this transcript, a novel ORF was identified being a promising candidate for a MCPyV Agnoprotein. So far, MCPyV together with MuPyV were thought to lack an Agnogene whereas, among others, SV40, BKPyV and JCPyV express a functional Agnoprotein [48]. Agnoproteins are structurally diverse and also play various roles in PyV infection e.g. in the regulation of gene expression, viral maturation and virus release [48,192,193].

The putative MCPyV Agnogene ORF contains a CUG-start codon and thus translation into Agnoprotein would initiate with Leucine instead of Methionine [110,140] (Figure 22).

```
5303  CTGGCATTGACTCATTTCCTGGAGAGGCGGAGTTTGACTGATAAACAAAACTTTTTTTCT
       L  A  L  T  H  F  L  E  R  R  S  L  T  D  K  Q  N  F  F  S
5243  TTCTGTTTGGGAGGGAGACGGAAGACTCTTAACTTTTTTTCAACAAGGGAGGCCCGGAGG
       F  C  L  G  G  R  R  K  T  L  N  F  F  S  T  R  E  A  R  R
5183  CTTTTTTTTCTCTTACAAAGGGAGGAGGACATTAAAAGAGTAAGTATCCTTATTTATTTT
       L  F  F  L  L  Q  R  E  E  D  I  K  R  V  S  I  L  I  Y  F
5123  TCAGGATGGGGGGCATCATCACACTGCTGGCCAATATTGGTGAAATTGCTACTGAACTAA
       S  G  W  G  A  S  S  H  C  W  P  I  L  V  K  L  L  L  N  *
```

Figure 22: Nucleotide and amino acid (aa) sequence of the predicted MCPyV Agnoprotein
Putative MCPyV Agnoprotein ORF in the late orientation (nucleotides 5303-5064 of MCVSyn) and translation of the coding sequence. The non-canonical start-codon (CUG) is marked in red.

In addition to this discrepancy in regard to other Agnoproteins, MCPyV Agnoprotein contains no significant sequence similarity to SV40, JCPyV or BKPyV (as determined using blastp, NCBI). The putative MCPyV Agnoprotein is slightly larger than Agnoproteins of SV40, BKPyV and JCPyV (Table 48), however even larger Agnoproteins exist e.g. in Mastomys polyomavirus (MaPyV), which encodes an Agnoprotein with a length of 154 aa. Interestingly, most Agnoproteins are characterized by a basic isoelectric point (pI, Table 48, [48]). The predicted pI of MCPyV Agnoprotein is in very good accord to the values predicted for JCPyV, BKPyV and SV40 Agnoproteins (Table 48), indicating that this protein might at least have a certain degree of functional homology to other Agnoproteins.

Table 48: Biochemical properties of selected Agnoproteins
Theoretical pI and weight were calculated using the compute pI/Mw tool (ExPASy [145]).

	SV40	BKPyV	JCPyV	MCPyV
Length (aa)	61	66	71	78
Theoretical weight (kD)	7.3	7.4	8.1	9.4
pI	10.12	10.01	10.10	10.50

In regard to the initial question it can be stated that the mutations introduced in MCVSyn-hpko do not impair the splicing pattern of the hairpin mutant virus. However, the lack of mcv-miR-M1 leads to increased abundance of early mRNAs, particularly LT-Ag and 57kT-Ag specific transcripts, and reduced abundance of late transcripts relative to all viral transcripts.

4.2.3 MCVSyn establishes persistence in PFSK-1 cells

The above described results from replicating MCVSyn were obtained from cells harvested at four days after transfection. This time point was chosen because it guarantees robust genome amplification as well as strong expression of early viral genes. However, it was not investigated

how long viral episomes are detectable after transfection. Generally, plasmids are lost from transfected cells only few days after initial transfection, which is mainly due to loss of extrachromosomal DNA during cell division. Plasmids containing the SV40 origin of replication can replicate in cells expressing SV40 LT-Ag to allow more efficient expression of genes encoded on these plasmids. However, as depicted in Figure 23, these plasmids are also lost from transfected cells after few days. These facts underscore the significance of the discovery made during this study that MCVSyn is able to continuously replicate for several months in transfected cells. In three independent experiments *de novo* replicated viral DNA could be detected for at least 100 d by qPCR and Southern blotting of *DpnI* resistant DNA (Figure 24A, B, D and E).
Hence, the MCVSyn replication system is the first model system, which allows investigation of polyomavirus persistence.

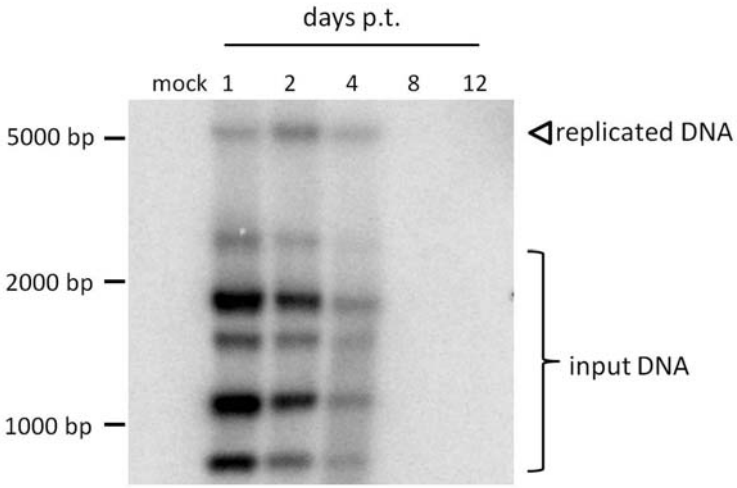

Figure 23: A SV40 LT-Ag expression plasmid with the SV40 origin of replication replicates less efficiently than MCVSyn
Southern blot analysis of *de novo* replicated DNA from PFSK-1 cells transfected with pCDNA3.1 SV40 LT-Ag at day 2-12 after transfection.

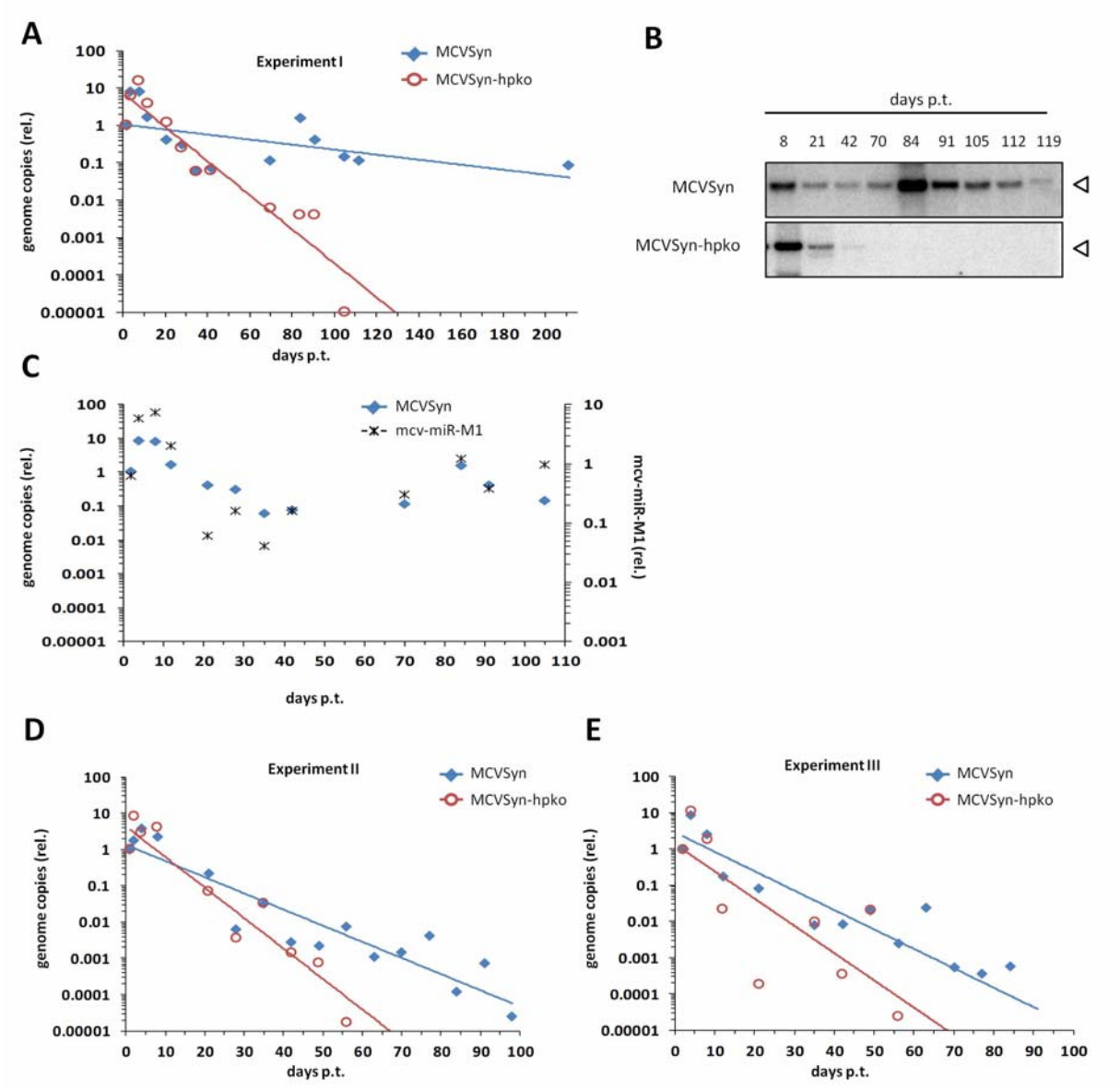

Figure 24: Long-term persistence of MCVSyn and accelerated loss of MCVSyn-hpko in PFSK-1 cells
(A) MCVSyn copy number relative to day 2 of replicating MCVSyn (blue line) and MCVSyn-hpko (red line) as determined by qPCR analysis at the indicated time-points from genomic DNA. **(B)** Southern blot analysis for the detection of newly replicated viral DNA of HIRT extracts from MCVSyn and MCVSyn-hpko at the indicated time-points in lon-term experiment I. **(C)** mcv-miR-M1 expression as measured by stem-loop RT-qPCR and MCVSyn copy numbers per cell in long-term experiment I. **(D)** and **(E)** replicate experiments of long-term experiments.

To assess the number of MCVSyn positive cells and to monitor viral genomes over time, a FISH assay for MCVSyn was established. As a control for the sensitivity and specificity of the MCVSyn FISH assay, the MCPyV positive MCC cell lines, MKL-1 and WaGa cells were analyzed. One or two MCPyV integration sites were discernible in MKL-1 and WaGa cells, respectively (Figure 25). In contrast to the faint but distinct FISH signals in MCC cell lines, larger dots were visible by FISH in MCVSyn transfected cells. Viral genomes were found in enlarged nuclei in multiple dots. At 4

days post transfection approximately 2.5% of all cells were found to harbor detectable amounts of MCVSyn DNA, organized in foci. As indicated in Figure 25, the frequency of MCVSyn positive cells dropped over time and reached a steady level of an estimated 0.01% positive cells. In agreement with the qPCR and Southern blotting results of the first long-term experiment, MCVSyn positive cells were detectable by FISH analysis for more than 200 d after transfection.

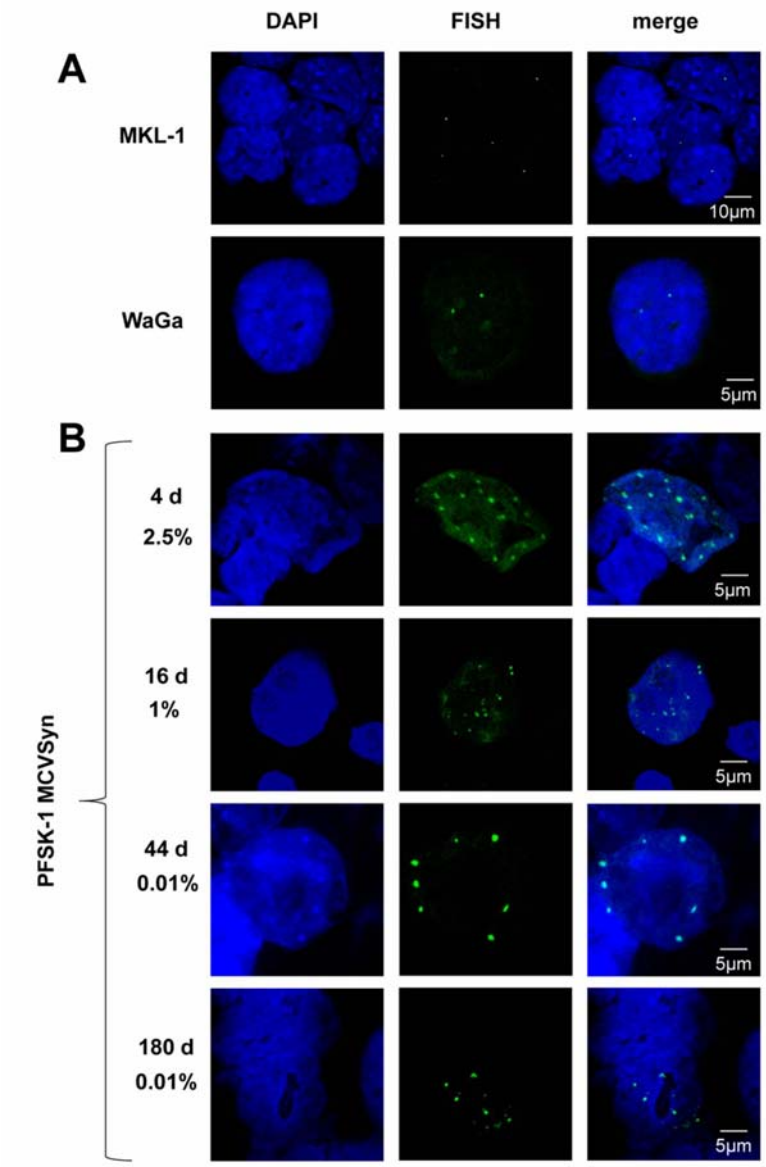

Figure 25: FISH analysis detects MCPyV genomes in MCC cell lines and MCVSyn transfected cells
(A) FISH analysis of the MCPyV positive MCC cell lines MKL-1 and WaGa. **(B)** FISH analysis of MCVSyn transfected PFSK-1 cells at early and late time-points after initial transfection. The estimated frequency of MCVSyn positive cells is indicated for each time point.

Besides detection of DNA, long-term transfected cells were analyzed for LT-Ag and miRNA expression. mcv-miR-M1 was found to be robustly expressed at all time-points (Figure 24C). Comparable to the slow loss of viral DNA, LT-Ag levels were decreasing over time as observed by

Western blot analysis (Figure 26A). However, immunofluorescence analysis for LT-Ag (Figure 26B) shows that not the amount of LT-Ag expressed per cell but rather the number of cells positive for LT-Ag expression dropped over time.

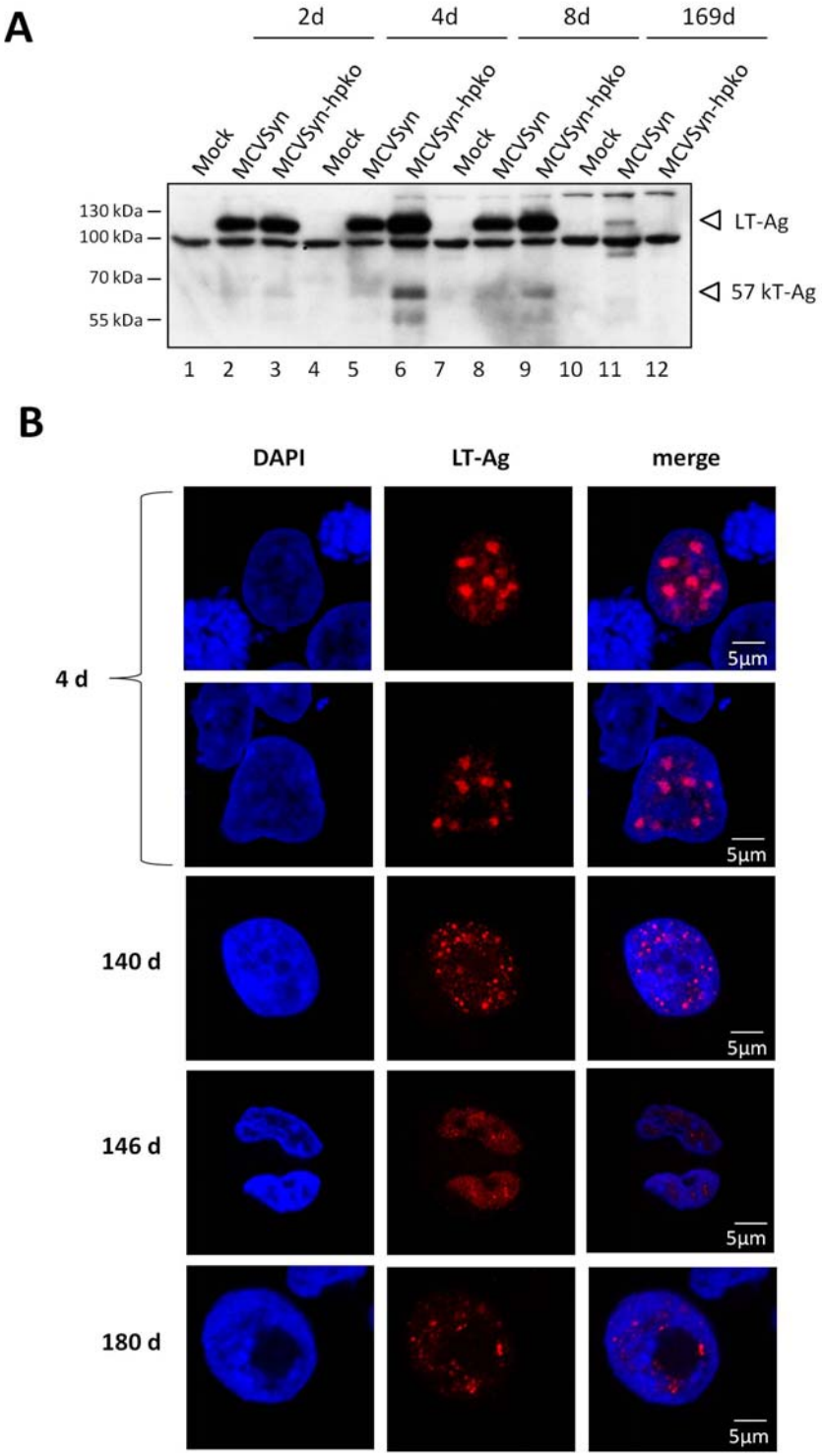

Figure 26: LT-Antigen expression during long-term persistence of MCVSyn
(A) Western blot analysis for LT-Ag expression in PFSK-1 cells transfected with MCVSyn at the indicated time-points. Protein was isolated from long-term experiment I (Figure 24). **(B)** PFSK-1 cells transfected with MCVSyn genome were analyzed for the expression of LT-Ag at the indicated time-points by immunofluorescence staining and laser scanning confocal microscopy.

4.2.4 MCVSyn persists as extrachromosomal episome

Based on strategies of other viruses that persist in their host cells, different mechanism for MCVSyn to maintain persistence come into consideration:

1. The viral DNA integrates into the host cell genome (e.g. retroviruses)
2. The viral DNA persists episomally as a minichromosome and is replicated and portioned upon cell division together with the cellular DNA (e.g. gammaherpesviruses and papillomaviruses)
3. Small amounts of particles are constantly shed to infect neighboring cells

FISH analysis showed that MCVSyn DNA is found in the nucleus in the form of large dots, which presumably represent replication centers (Figure 25). However, this technique cannot determine whether or not the viral genome is present as an episome, or if it has integrated into the cellular DNA. To address this question, a rolling circle amplification (RCA) experiment was carried out. By RCA, circular DNA, in contrast to linear DNA, is strongly amplified by the bacteriophage Phi29 DNA polymerase owing to its strand displacement activity.
DpnI digested genomic DNA (gDNA) from PFSK-1 cells transfected with MCVSyn at 4 d.p.t. and 136 d.p.t. was tested for the presence of circular DNA by RCA. As control for integrated viral DNA, gDNA from MKL-1 and WaGa cells was used as template in RCA. H_2O and gDNA from mock transfected PFSK-1 cells were used as controls for the presence of unspecific amplification or contaminants. After amplification, 1 µl of product was digested with the MCVSyn single cutter *BamHI* and analyzed by agarose gel electrophoresis. Figure 27 shows that MCVSyn transfected cells at 4 d.p.t. as well as at 136 d.p.t. (lanes 5 and 11) produced an amplicon of the same unit length size as MCVSyn excised from the plasmid pMCVSyn by *SacI* digestion (lane 13), revealing that MCVSyn persists in episomal form in PFSK-1 cells.

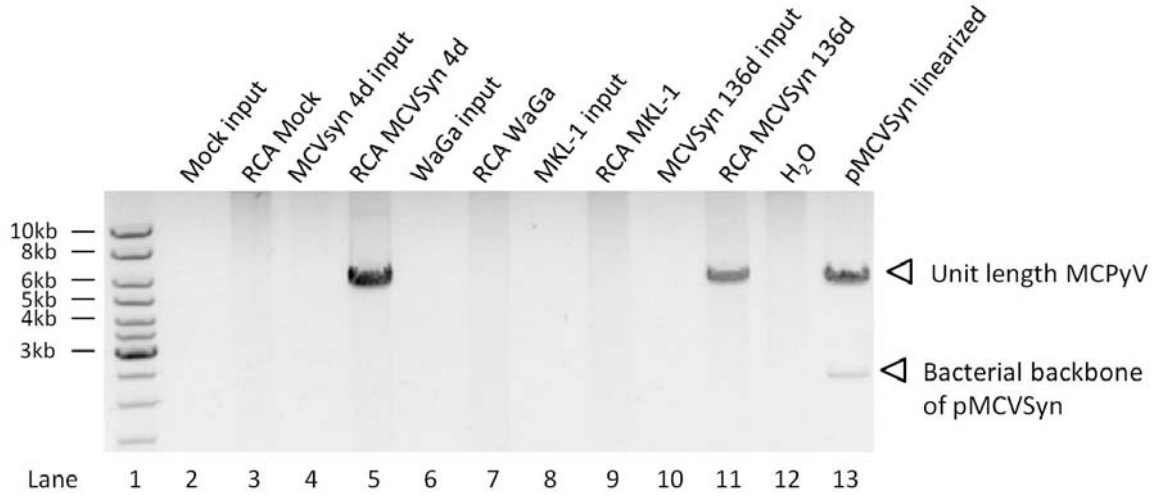

Figure 27: Rolling circle amplification of MCC cell lines and MCVSyn replication assays
Agarose gel analysis of RCA products. For each sample, the amount of DNA used as input as well as 1 µl of circularized RCA product were separated on a 0.8% agarose gel. Amplicons with the size of unit length MCPyV are detectable after RCA with MCVSyn transfected cells at 4 d.p.t. as well as 136 d.p.t.

4.2.5 A mcv-miR-M1 knockout mutant is impaired in maintaining persistence

As described above, mcv-miR-M1 negatively affects the expression of LT-Ag and thereby limits replication of MCVSyn DNA. This could lead to the expectation that MCVSyn-hpko, by exhibiting a more robust DNA replication, might more efficiently establish persistence. However, while MCVSyn-hpko shows a stronger DNA replication at early time-points, it was lost at an accelerated rate in comparison to MCVSyn in three independent experiments (Figure 24A, D and E). As expected, the expression of LT-Ag was also no longer detectable after loss of the viral DNA (Figure 26A). These results strongly suggest that the loss of mcv-miR-M1 expression reduces the ability to maintain persistence of MCVSyn.

4.3 Investigation of the influence of mcv-miR-M1 on MCVSyn persistence

The striking result that mcv-miR-M1 expression is required for long-term replication of MCVSyn immediately raises the question, how mcv-miR-M1 contributes to the maintenance of persistence. Several possible mechanisms could explain this observation. Firstly, in the absence of mcv-miR-M1, early gene expression is not limited and LT-Ag accumulates. This could e.g. influence the switch to late gene expression and thereby impair particle production and re-infection ability. Secondly, the LT-Ag C-terminus was shown to have cell-growth inhibitory functions. An excess of LT-Ag therefore might lead to a growth disadvantage of MCVSyn-hpko transfected cells. Lastly, apart from increased LT-Ag expression, mcv-miR-M1 might target cellular transcripts and thus modulate the cellular environment to allow efficient establishment

of persistence. The following chapter describes how these scenarios were investigated and provides possible explanations for the role of mcv-miR-M1 in persistence.

4.3.1 MCVSyn and MCVSyn-hpko produce low amounts of infectious viral particles

To examine potential differences between the production of viral particles by MCVSyn and MCVSyn-hpko, DNase I resistant DNA was isolated from cell culture supernatants and quantified by qPCR. No significant difference in DNase I resistant MCVSyn specific DNA could be detected between supernatants from MCVSyn and MCVSyn-hpko transfected cells (Figure 28A). However, it must be noted that the effectivity of DNase I digestion could not be controlled and it might be possible that residual non-encapsidated DNA was still present in qPCR.

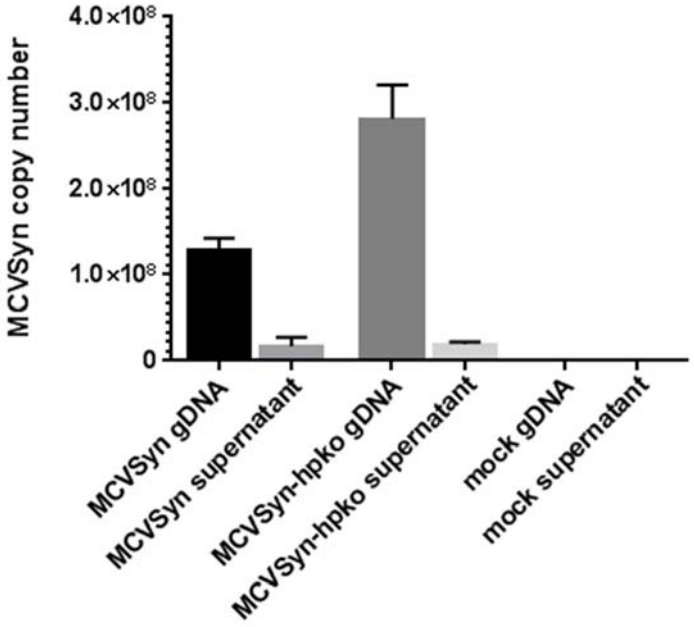

Figure 28: Quantification of MCVSyn genomes in DNase I treated supernatants from MCVSyn and MCVSyn-hpko transfected cells
The amount of DNase I resistant MCVSyn DNA from supernatants of MCVSyn or MCVSyn-hpko transfected PFSK-1 cells at 4 d.p.t. was measured by qPCR and is shown in comparison to MCVSyn copy number in the genomic DNA of transfected cells (n=3).

The presence of DNase I resistant DNA in the cell culture supernatant does not necessarily reflect the production of actually infectious particles. To investigate if MCVSyn and MCVSyn-hpko derived particles are infectious and whether any difference in viral titers can be observed, cells from replication assays were lysed by several freeze-thaw cycles. Lysates were filtered and used to infect fresh cells which were probed for the presence of MCVSyn DNA by qPCR at different time-points after infection.

As depicted in Figure 29, no significant difference in viral DNA amounts between cells infected with lysates from MCVSyn and MCVSyn-hpko could be observed. Generally, viral titers were very low, questioning the efficiency of naturally occurring re-infection.

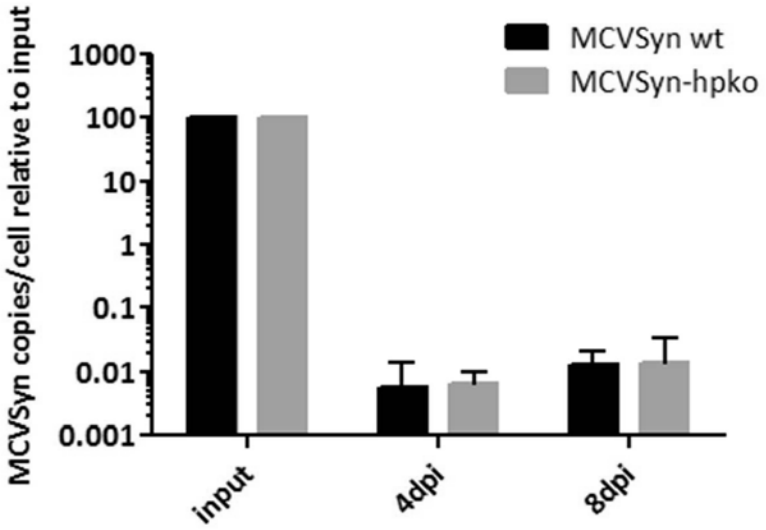

Figure 29: No difference in re-infection ability between MCVSyn and MCVSyn-hpko derived particles
MCVSyn copy number per cell as measured by qPCR from MCVSyn or MCVSyn-hpko transfected cells used for particle isolation (input) and 4 d.p.t. and 8 d.p.t. of PFSK-1 cells (n=3).

As described by Neumann, Borchert et al. [72], the amount of particles produced by MCVSyn is very low and in comparison to particles derived from SV40, they are more irregularly shaped. While infectious particles of SV40 are released upon cell lysis, this was not observed for MCPyV in semi-permissive replication systems. However, it is possible that MCVSyn derived particles can infect neighboring cells by cell-cell-contact and that continuous re-infection is necessary for the maintenance of viral persistence. Hence, it cannot be ruled out that the accelerated loss of MCVSyn-hpko is caused by a reduced production of particles at late time-points of persistence although the general influence of re-infection on the maintenance of persistence requires additional investigation.

4.3.2 LT-Ag has a negative effect on cellular proliferation

The MCPyV LT-Ag has diverse functions during the viral life cycle. Among them, the C-terminal Origin Binding Domain (OBD) of LT-Ag binds to GRGGC pentanucleotides at the viral origin of replication and induces unwinding of the viral DNA by its helicase activity [64,194]. Besides its requirement for viral replication, the LT-Ag C-terminus induces DNA damage response (DDR)

and has growth inhibitory influence on the host cell, which seems to be partly caused by activation of the p53 pathway but remains incompletely understood [165,195].

Since MCVSyn-hpko transfected cells have elevated LT-Ag levels, it can be hypothesized that due to the growth inhibitory functions of the LT-Ag C-terminus, these cells have a proliferation disadvantage. Because of the high background of non-transfected cells in the MCVSyn replication assay, a proliferation defect of MCVSyn-hpko transfected cells cannot be determined directly. To measure the effect of different LT-Ag levels on cellular proliferation, PFSK-1 cells were transfected with increasing amounts of a LT-Ag expression plasmid (pCMV:ER-S) and cellular proliferation was determined by MTT-assays. Figure 30 shows that the cellular growth is reduced in response to LT-Ag levels. These observations suggest that MCVSyn-hpko compared to MCVSyn transfected cells are slightly impaired in their proliferation rate and may be overgrown during long term cultivation at an accelerated rate.

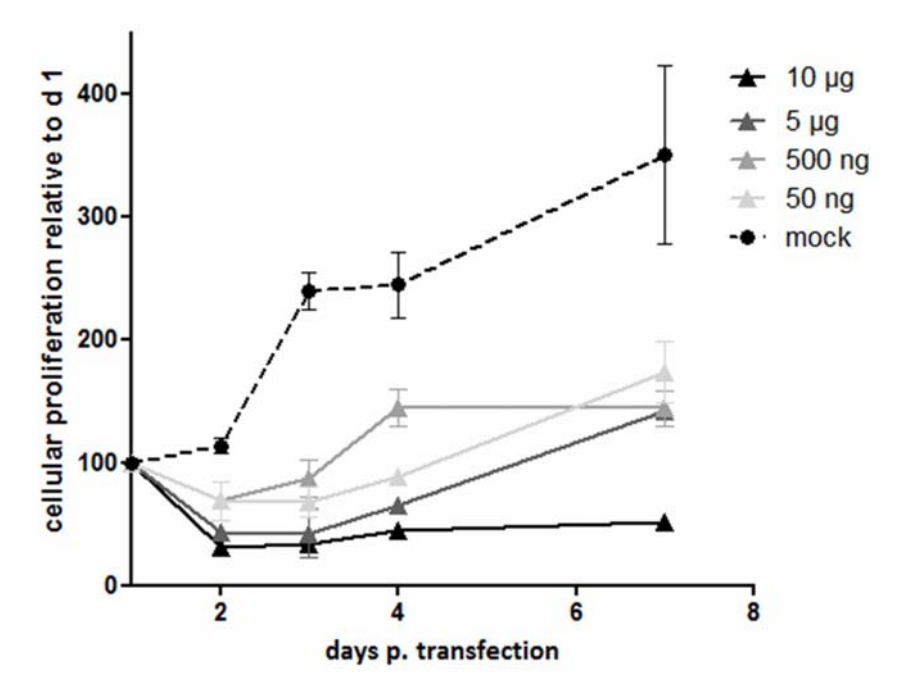

Figure 30: Proliferation of PFSK-1 cells dependent on LT-Ag levels
Proliferation of PFSK-1 cells transfected with increasing amounts of a LT-Ag expression construct (pCMV:ER-S) was measured by MTT-assays at the indicated time-points and is shown relative to the first measurement at 24 h post transfection.

LT-Ag mediated induction of DDR by activation of the p53 pathway was described to play a role in the reduced proliferation rate of LT-Ag expressing cells [165]. The downstream targets of p53 p21, MDM2 and GADD45 were shown to be significantly increased upon expression of LT-Ag [165]. However, using the same primer sets as employed by Li et al. [165], a significant increase of mRNA levels of these target genes upon LT-Ag expression could not be observed in PFSK-1 cells (Figure 31). These diverging observations might be attributed to the usage of different cell

lines and it remains to be investigated if other DDR markers are activated upon expression of LT-Ag in PFSK-1 cells.

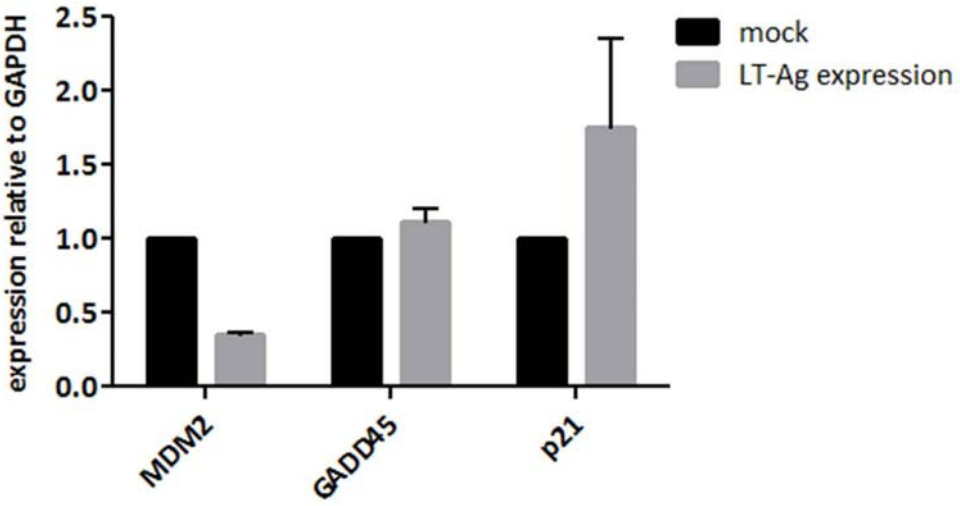

Figure 31: No induction of p53 downstream targets upon expression of LT-Ag
Relative transcript levels of p21, MDM2 and GADD45 in PFSK-1 cells with or without expression of LT-Ag as determined by RT-qPCR (n=3).

Nevertheless, other effects of LT-Ag on the host cell, not mutually exclusive to DDR, might play a role in the reduction of cellular growth. As already hypothesized by Li et al., the LT-Ag helicase domain might be involved in reducing cellular proliferation [165], probably by effects on the host cell genome. Due to its high affinity to viral DNA, it is conceivable that MCPyV LT-Ag also binds to cellular DNA and might cause genomic instability or influence cellular gene expression. Therefore, ChIP-Seq analysis for global identification of LT-Ag binding sites on the cellular genome was carried out.

4.3.3 ChIP-Seq analysis reveals that LT-Ag binds in a distinct pattern to the cellular genome

Transfection of re-ligated MCVSyn usually results in a frequency of only 3-5% of cells harboring replicating viral DNA and thus producing LT-Ag. Due to the fact that the majority of cells do not express LT-Ag, it is difficult to identify LT-binding sites in ChIP-Seq datasets from MCVSyn transfected cells. Therefore, full length (FL) LT-Ag was ectopically expressed in PFSK-1 cells prior to ChIP-Seq experiments using the anti-LT antibody CM2B4. To investigate the importance of the LT-Ag C-terminus in DNA binding, ChIP-Seq analysis was also performed for tumor derived truncated LT-Ag (trunc LT-Ag) in the MCPyV positive MCC cell line WaGa as well as in PFSK-1 cells transfected with a trunc LT-Ag expression plasmid.

After mapping of sequenced reads, peaks were called using macs14 [160]. 1581 peaks were detected for ectopically expressed FL LT-Ag and 125 or 249 for trunc. LT-Ag or WaGa cells, respectively. Visual inspection of these peaks revealed that peaks in the FL LT-Ag dataset were easily distinguished from background signals, while peaks in both trunc LT-Ag datasets appeared less distinct.

The algorithm CEAS (Cis-regulatory Element Annotation System) was used to catalogue peak distribution at genomic features (Figure 32). The analysis roughly discriminates between the coding gene body, downstream regions including 3' UTRs, promoter regions, including 5' UTRs and intergenic regions which are more than 2000 nt up- or downstream of a coding region.

Peaks from ChIP-Seq analysis with ectopically expressed truncated LT-Ag seem to be randomly spread over the genome. In WaGa cells, an enrichment of peaks in distal intergenic regions can be observed. Interestingly, the full length LT-Ag was found to bind preferentially to promoter regions and 5'UTRs, a binding pattern as may be expected for a transcription factor.

LT-Ag is known to bind with high affinity to GRGGC motifs at the viral origin of replication [64,194]. The tool fimo was used to investigate if regions of LT-Ag binding are enriched in GRGGC sequences. The best 100 peaks of all three samples were analyzed for the number of GRGGC sites in a region of 250 nt up- and downstream of the peak summit. The truncated LT-Ag bound to regions with 0.4 or 0.5 (WaGa cells) GRGGC motifs on average whereas the FL LT-Ag was found at sites with 6.9 motifs per 500 nt. Using a dataset with 2000 nt regions upstream of all annotated genes, 1.9 GRGGC motifs per 500 nt were identified. Besides the overrepresentation of GRGGC motifs in FL LT-Ag peaks, the total number of peaks which contain GRGGC sites was significantly higher for FL LT-Ag than for trunc LT-Ag. 95.4% of all FL LT-Ag peaks contained at least one GRGGC sequence, whereas only 69.6% or 60.2% (WaGa cells) of trunc LT-Ag peaks contained at least one GRGGC motif.

The tool meme was employed for motif discovery in ChIP-Seq peaks. For all samples, motifs with significant E-values could be identified but they all displayed a high ambiguity and could not be clearly correlated to known transcription factor binding sites using the resource tomtom. However, it was noticeable that regions of FL LT-Ag peaks at the cellular chromatin were highly GC-rich. Interestingly, most FL LT-Ag peaks at cellular promoter regions map to CpG-islands (Figure 32). Hence, it is possible that LT-Ag does not specifically bind at GRGGC motifs but rather at CpG-islands of open chromatin structures.

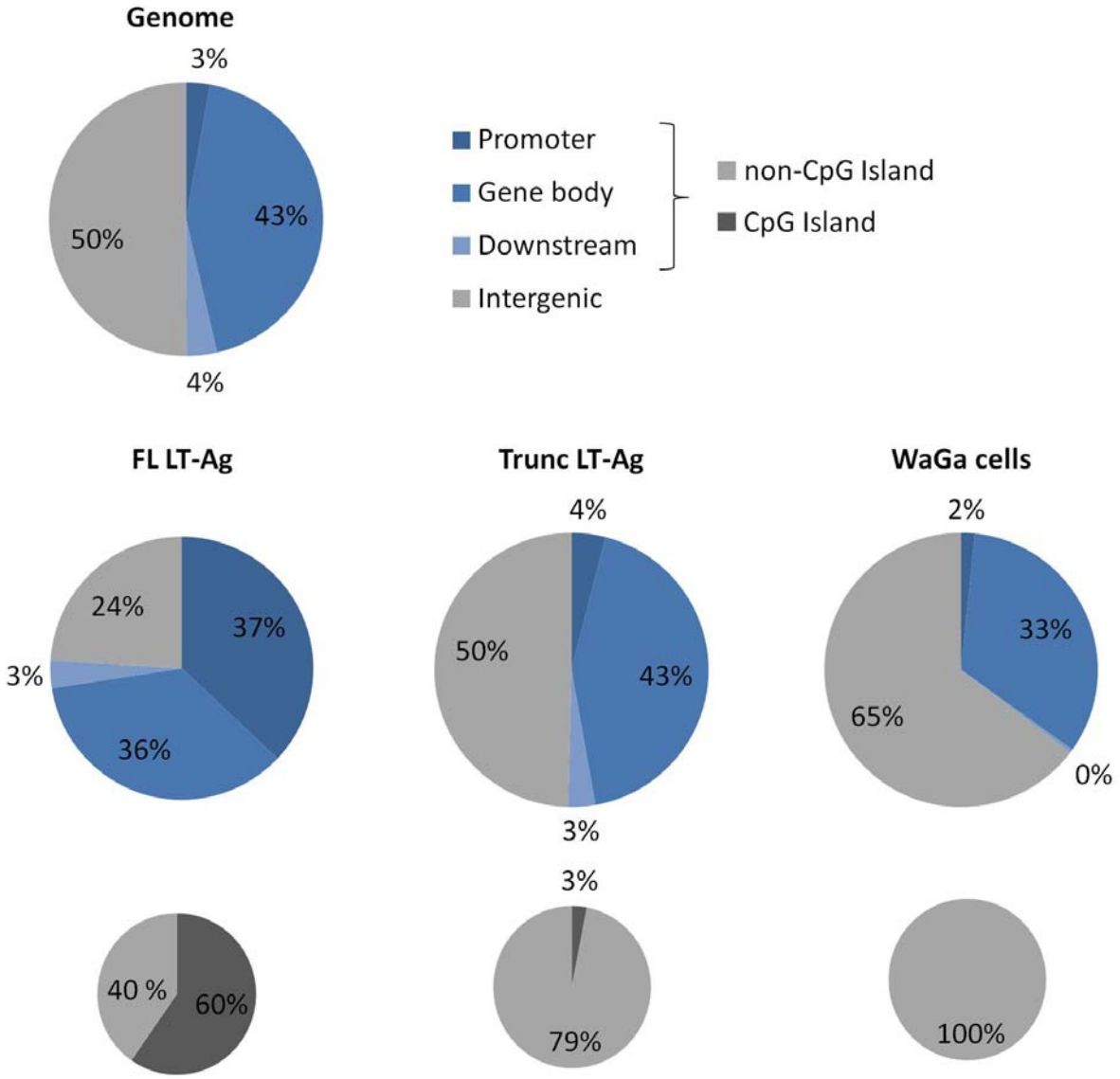

Figure 32: FL LT-Ag but not truncated LT-Ag is enriched at promoter regions of host chromatin and has a high affinity to CpG islands
Clustering of peaks at genomic features after ChIP-Seq analysis of ectopically expressed full length LT-Ag (FL LT-Ag), truncated LT-Ag (trunc LT-Ag) and intrinsic trunc LT-Ag in WaGa cells using CEAS 1.0.2 [154,155]. The diagram "genome" represents the distribution of the given genomic features over the human genome. For all three samples, the percentage of peaks at each genomic feature is shown relative to all peaks. Peaks at the features Promoter, Gene body and Downstream were further invesitaged for the presence of CpG islands. The frequency of CpG islands was determined using MEME 4.1 [161] and is shown as smaller diagrams in the lower panel.

The high enrichment of FL LT-Ag peaks at promoter regions suggests that FL LT-Ag might function as transcription factor and influence host cell gene expression. With bedtools, the nearest TSS for each ChIP peak was identified. Using all peaks from all ChIP samples, 1652 unique TSS were identified in proximity to ChIP-Seq peaks (2000 nt). However, only 5 TSS were common among all three ChIP samples (Figure 33).

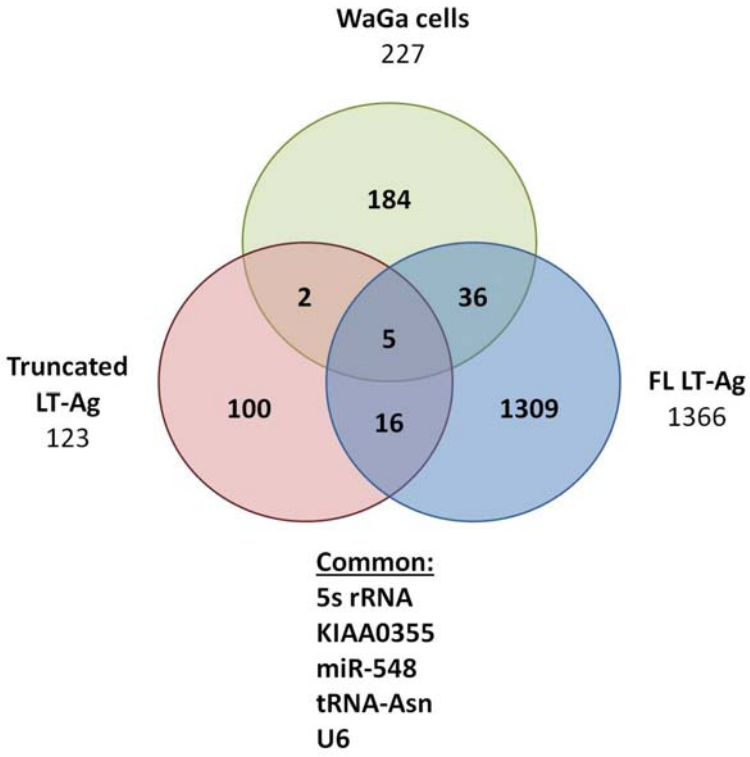

Figure 33: Overlap of genes in proximity to ChIP-Seq peaks between FL LT-Ag and truncated LT-Ag
Venn diagram depicting the number and overlap of genes in vicinity to ChIP-Seq peaks from all three samples. 1652 genes were identified in total but only 5 of them were common among all three samples. These 5 genes do not have a significant functional overlap. The two truncated LT-Ag samples (truncated LT and WaGa), overlap only in 7 genes.

Even among the two truncated LT-Ag samples, the set of overlapping TSS was very small (only 7 of 350 genes were identical). A gene ontology analysis was carried out for FL LT-Ag peaks using DAVID Bioinformatics Resources 6.7 and the 10 highest scoring annotation clusters are found in Table S 1. Interestingly, among them are pathways that are associated with the regulation of apoptosis, chromatin organization, transcription and DNA damage and repair. Thus, binding of FL LT-Ag to cellular chromatin might influence cellular gene expression and modulate infection- or persistence-relevant pathways.

5 Discussion

MCPyV is the first human polyomavirus which is causally linked to cancer [12]. Although, like many other polyomaviruses, MCPyV presumably persists life-long in its host and has a high seroprevalence in the human population [196], MCPyV induced tumorigenesis is very rare. MCPyV gene products involved in tumorigenesis, in particular the early proteins LT-Ag and sT-Ag, are currently under extensive investigation ([96,197,198] and reviewed in [61]). However, due to the difficulties in establishing MCPyV *in vitro* and/or *in vivo* model systems, so far little is known about the natural biology of MCPyV and the mechanisms that lead to the establishment and maintenance of viral persistence. The development of semi-permissive *in vitro* replication systems [72,73] has opened up new possibilities to investigate the biology of MCPyV in context of viral replication and limited particle production and provides the opportunity to analyze the role of the viral miRNA mcv-miR-M1 in authentic viral replication.
The aim of this work was to understand the biological role of mcv-miR-M1 in MCPyV replication by analyzing the miRNA expression and function using the MCVSyn semi-permissive replication system [72].

5.1 Abundance and biogenesis of mcv-miR-M1 during replication of MCVSyn and in MCC cell lines

High throughput sequencing of small RNA moieties revealed that mcv-miR-M1 is highly expressed from replicating MCVSyn and that no additional viral miRNAs are generated. The mcv-miR-M1 pre-miRNA was found to be processed into a 5p and a less abundant 3p mature miRNA. Strikingly, mcv-miR-M1 5p belongs to the most prevalent miRNAs detected in bulk MCVSyn replication assays. The actual abundance of mcv-miR-M1 5p in cells harboring MCVSyn has to be even higher given that only 3-5% of all cells that were analyzed by small RNA-Seq supported viral DNA replication. Thus, mcv-miR-M1 5p seems to dominate the miRNA profile in MCVSyn transfected cells. If mcv-miR-M1 is equally abundant in naturally infected cells remains to be investigated.
Expression of mcv-miR-M1 in the MCC cell lines MKL-1 and WaGa was approximately four orders of magnitude lower, even though, in comparison to MCVSyn transfected bulk cultures, each cell contains at least one copy of integrated MCPyV. mcv-miR-M1 accounts for approximately 0.001% of all miRNAs in the investigated MCC cell lines. This is in accordance with a recent meta-analysis of small RNA-Seq data of MCPyV positive MCC tumors, which further

revealed that more than half of all analyzed MCC samples exhibited no mcv-miR-M1 expression at all [50,148]. Given that shRNA-mediated knockdown of LT-Ag in MCC tumors impairs the proliferation of transformed cells [104], it is to be expected that high-level expression of mcv-miR-M1 from integrated MCPyV would have an equally disadvantageous effect on tumor cell proliferation. Thus, a strong selective pressure to reduce or extinguish mcv-miR-M1 expression in MCPyV transformed cells has to exist. In addition to mutations in the early region leading to expression of a truncated LT-Ag, silencing of mcv-miR-M1 expression therefore seems to be a second prerequisite for MCC formation. Since no mutations of the mcv-miR-M1 pre-miRNA were identified in MCC derived MCPyV sequences, reduction of mcv-miR-M1 expression might be caused by epigenetic silencing of the promoter region(s) responsible for miRNA expression (see Figure 34). Additionally, miRNA expression from the late promoter might be abrogated if, upon integration, the viral DNA linearizes upstream of the mcv-miR-M1 locus. This scenario is supported by the observation that breakpoints of MCPyV are predominantly located downstream of the early region Rb-binding motif [92-94]. Different modes of mcv-miR-M1 expression are discussed in detail in chapter 5.2. While mcv-miR-M1 is unlikely to be biologically active in MCC, the raccoon polyomavirus (RacPyV) expresses a miRNA, which is highly abundant in RacPyV induced tumors [50]. The RacPyV miRNA as well as mcv-miR-M1 both target early viral transcripts but T-Ags are still robustly expressed in RacPyV induced tumors which harbor predominantly episomal and not integrated viral genomes [50]. These observations suggest that MCPyV and RacPyV induced tumors arise via different mechanisms and that some PyV encoded miRNAs might play a role in disease development. While the downregulation of early genes seems to be a conserved functionamong PyV encoded miRNAs [50-53,147], their involvement in the regulation of cellular genes might be highly diverse, an assumption which is underscored by their low sequence conservation.

The discovery of a mcv-miR-M1 5p seed variant in MCC tissue by Lee and colleagues led to the hypothesis that the pre-miRNA hairpin undergoes differential processing in MCC cells, resulting in the expression of isomiRs, i.e. different mature miRNAs produced from the same miRNA precursor [148]. The tumor-specific isomiRs of mcv-miR-M1 were hypothesized to regulate hosts targets that may be involved in MCC tumorigenesis [148]. Although the reported mcv-miR-M1 isomiRs differ by a shift of only 2 nt, this has marked impact on the sequence of the seed region. The seed region of a mature miRNA comprises the nucleotides 2-8 and defines its binding specificity as this part of the miRNA needs to bind with perfect complementarity to the target mRNA to have a regulatory effect [199]. As a consequence, the expected cellular targetomes of mcv-miR-M1 isomiRs would differ dramatically.

However, the hypothesis of integration-dependent differential miRNA processing could be disproven in this study by performing small RNA-Seq from replicating MCVSyn in direct

comparison to MCC cell lines. These experiments revealed that differentially processed mcv-miR-M1 5p and 3p isomiRs are found both in replicating as well as in integrated MCVSyn. Such isomiRs are likely produced by variable Drosha- or Dicer cleavage, a common process in miRNA biogenesis [200,201]. It is now well established that pre-miRNAs are frequently processed into isomiRs with variable lengths and seed-sequences, rather than a single defined miRNA [183,202,203]. Importantly, the side-by-side analysis of replicating and integrated MCPyV revealed that both seed-variants of mcv-miR-M1 5p are found in MCC cell lines as well as during viral replication but that there is no difference in their expression pattern between integrated MCPyV and actively replicating MCVSyn.

The comparison of different small RNA-Seq library preparation methods during this study revealed that the frequencies of mcv-miR-M1 isomiRs in small RNA-Seq datasets differ between small RNA library preparation kits. This can be explained by the fact that several steps during library preparation are highly biased. In particular, the efficiency of 3' adapter ligation is strongly influenced by the small RNA and adapter sequence combinations and the ligase specificity [180,182,184-186]. Thus, the source of the observed differences between different library preparation kits as well as between the reported sequences of mcv-miR-M1 5p by Seo et al. [51] and Lee et al. [148], might be the usage of different 3' adapters, which bind with varying efficiency to one or the other mcv-miR-M1 isomiR.

Since non-biased methods for small RNA sequencing are not yet available, it remains to be investigated which miRNA variant is the predominant form. However, the NEBnext library preparation kit could generally retrieve more mcv-miR-M1 sequences than the TruSeq library preparation kit when using the same RNA sample. Other studies confirmed that the NEBnext kit generally produced higher yields of recovered RNA-Seq libraries with less variation [181]. It thus seems likely that the first described $5p_{17\text{-}23}$ miRNA is the predominant isomiR during MCPyV replication.

Besides differential processing at the 5' end of miRNAs, variations at the 3' ends of miRNAs were observed. In *Drosophila melanogaster*, a 3′- 5′ exoribonuclease called Nibbler which shortens too long mature miRNAs while bound by Ago1, was identified as the cause for miRNA 3' end heterogeneity in flies [204], and it is likely that similar ribonucleases are active in mammalian cells as well. These variations do not influence the miRNA seed sequence and are considered as of minimal importance for miRNA-mRNA binding [205].

5.2 An autonomous promoter allows NCCR-independent expression of mcv-miR-M1 during viral replication

While mechanisms of miRNA-mediated post-transcriptional regulation were extensively studied over the last decade, little is known about how miRNAs themselves are regulated. However, insights into the conditions that induce or silence miRNA expression will most likely significantly contribute to a comprehensive understanding of miRNA functions.

Time-course experiments suggest that the miRNAs of MuPyV, SV40, BKPyV and JCPyV are produced at late stages of infection [52,53,147] and that the BKPyV miRNA expression levels were to correlate with the activity of the late promoter [151]. In MuPyV and SV40 previous findings suggested that the inefficient termination of late transcription results in production of antisense RNAs [158,206] which were later identified as viral precursors for viral miRNAs [52,53]. These observations support the hypothesis that PyV miRNAs are generally produced by inefficiently polyadenylated late transcripts that traverse the entire viral genome [81,82,207]. An exception is the Bandicoot papillomatosis carcinomatosis virus (BPCV) miRNA, which is expressed from its own promoter [208].

3'RACE analysis of replicating MCVSyn indeed revealed that early transcripts are efficiently terminated, whereas a high degree of transcriptional read-through can be observed in the late orientation. Similar to MuPyV, the late pA-signals in MCPyV seem to be rather weak resulting in inefficient termination of late transcription [209]. Thus, late read-through transcripts could likely fold into pre-miRNA hairpin structures and undergo Drosha mediated cleavage.

To investigate if viral transcription indeed exclusively initiates at the NCCR, transcriptional start sites (TSS) were mapped by a cap-dependent 5'RACE analysis. The major early region TSS was found to be located 22-24 nt downstream of a TATA-Box, a distance that is typically observed for eukaryotic promoters, as well as for early promoters of SV40 [210] and MuPyV [211]. The majority of other loci that were detected by early strand 5'RACE were scattered randomly over the viral genome with a very low frequency and can be considered as experimental artifacts with the exception of one low abundant transcriptional start region which was found to be located upstream of the core origin of replication. Transcripts starting upstream of the ori contained a splice donor site immediately downstream of the TSS. RT-PCR with a primer spanning the region between this splice donor and the first early region splice acceptor site confirms the presence of early transcripts starting upstream of the ori during viral replication. Similar observations were made earlier for the polyomaviruses MuPyV [211], SV40 [212] and the human JCPyV [213]. In all cases, immediately after infection, transcription of the early genes starts downstream of the ori whereas after the onset of viral DNA replication, transcriptional start sites are found upstream of the ori. It is hypothesized that different promoter regions and the binding of LT-Ag itself at the

early promoter cause a shift of TSS [212]. Due to the lack of a permissive model system, a similar temporal shift of TSS usage could not be determined for MCPyV.

The major NCCR-associated TSS in the late orientation were found in a broad region with the majority of transcripts starting in a window of 41 nt. This heterogeneity of TSS seems to be a common feature among polyomavirus late transcription as it was already observed in MuPyV [214] and SV40 [215]. Although the consequences of these findings still lack a functional explanation, they underscore the differences of the TATA-box dependent early and TATA-less late promoter. It can be speculated that differences in TATA-dependent and TATA-independent transcription play a role in the regulation of early and late transcription.

HTS analysis of late 5'RACE PCR products revealed that, in addition to TSS at the NCCR, another TSS is located 116 nt upstream of the pre-miRNA hairpin. Transcripts that initiate at this location would most likely lead to expression of mcv-miR-M1, independently of promoter functions in the NCCR. Although this TSS accounts for only 3% of all HTS reads in the late orientation, its abundance is underestimated by the fact that only pri-miRNAs which are not (yet) processed into pre-miRNAs can be captured by this method.

Studies suggest that intronic miRNAs are not necessarily co-transcribed with gene they are located in, suggesting that they might have their own independent promoters [135,136], which was considered to be the case for mcv-miR-M1, as well. Strikingly, mcv-miR-M1 expression could be observed in absence of the late viral promoter and the late gene region, strongly supporting the hypothesis of an independent miRNA transcript which originates from an autonomous promoter.

The expression of mcv-miR-M1 from an early region containing plasmid was significantly reduced by the addition of α-amanitin, a specific inhibitor of RNA-pol II [189], suggesting that the putative miRNA promoter recruits RNA-pol II. The initiation of RNA-pol II transcription in most cases requires the formation of a transcription complex consisting of several transcription factors which are often positioned by binding to a TATA-Box. However, no TATA-Box is found upstream of the miRNA TSS. Several potential transcription factor binding sites could be predicted by *in silico* analysis, however, it remains to be investigated if any of them influence the expression of mcv-miR-M1.

Further, investigation of promoter activity in regions upstream of mcv-miR-M1 using a luciferase reporter assay did not reveal significant promoter activity in this region. However, it is possible that the promoter region overlaps with the pre-miRNA hairpin, which could not be analyzed in luciferase reporter assays.

Taken together, these observations suggest that the miRNA promoter is an atypical promoter, which does not share canonical promoter elements and mechanisms with other eukaryotic promoters. This observation fits into the result that miRNAs can be generally expressed via

various different mechanisms, which in part significantly differ from known transcription mechanisms [135,136,216-218].

Antisense transcription through the miRNA promoter region, as it would usually occur in intact MCPyV, might play a role in promoter activation, as observed in early region expression plasmids. Nevertheless, expression of mcv-miR-M1 was also observed to occur from a promoter-less and consequently antisense-transcription-less plasmid containing the early coding region alone, although the miRNA expression level in this case was very low.

ChIP-Seq experiments support the theory of an autonomous miRNA promoter by revealing that the putative miRNA promoter displays the histone modification H3K4-me3 which is often associated with active transcription. H3K4-me3 is additionally only detectable at the promoter regions for early and late gene expression.

Interestingly, the putative miRNA promoter region displays an accumulation of GRGGC motifs, comparable to the region of GRGGC motifs at the origin of replication. Since LT-Ag was shown to bind specifically at GRGGC motifs at the origin of replication [64,194], it is possible that LT-Ag also has a regulatory role at the miRNA-promoter region. However, binding of LT-Ag at the miRNA promoter region could not be detected by ChIP-Seq, whereas LT-Ag binding at the origin of replication was readily detectable on replicating MCVSyn. A mutant virus, MCVSyn-ltb, in which the GRGGC motifs upstream of the pre-miRNA hairpin were disrupted without altering the LT-Ag aa-sequence, showed ChIP-Seq profiles for H3K4-me3 and LT-Ag, which were identical to MCVSyn. Furthermore, the introduced mutations did not affect the mcv-miR-M1 expression level. Thus, despite of the conspicuous accumulation of GRGGC motifs in the miRNA promoter region, an influence of LT-Ag on mcv-miR-M1 expression could not be observed under the given experimental conditions. It remains possible, however, that the GRGGC pentanucleotide clusters might still play a role during natural viral infection or at another stage of the viral life cycle. It is also possible that smaller amounts of LT-Ag bound to the miRNA promoter region are not detectable with CM2B4 antibody, which was shown to have a rather low affinity [191]. Hence, further experimental investigation, for example by electrophoretic mobility shift assays (EMSA), will be required to investigate whether LT-Ag binds to the promoter region upstream of the viral miRNA.

To further investigate the influence of the miRNA promoter sequence on mcv-miR-M1 expression, a MCVSyn miRNA promoter mutant, MCVSyn-pmk, with point mutations in a region of 226 nt in the miRNA promoter region, which do not alter the LT-Ag aa-sequence, was examined by ChIP-Seq analysis. Interestingly, the MCVSyn-pmk mutant was impaired in acquiring H3K4-me3 marks at the miRNA promoter region. Furthermore, expression of mcv-miR-M1 was significantly reduced in MCVSyn-pmk. These results imply that the mutated region in MCVSyn-pmk contains so far unknown promoter or enhancer elements for the expression of

mcv-miR-M1. It seems possible that residual H3K4-me3 marks at the mutated region and the expression level of mcv-miR-M1 in MCVSyn-pmk might be further reduced if more nucleotides were mutated in the promoter region. However, this would also alter the amino acid sequence of LT-Ag and most likely cause defects in LT-Ag functions. The observation that the expression of mcv-miR-M1 in MCVSyn-pmk is reduced but not fully abolished could be also explained by the observation that late transcriptional read-through transcripts might serve as an additional miRNA expression mechanism.

The model of mcv-miR-M1 expression that can be inferred from these observations is illustrated in Figure 34A: mcv-miR-M1 can be produced from late read-through transcripts but additionally also from autonomous transcripts, which originate from a RNA-pol II promoter upstream of the pre-miRNA hairpin. Both mechanisms seem to be employed during replication of MCVSyn but so far, it can only be speculated why two different expression mechanisms exist side-by-side. Expression levels of the BKPyV miRNA can be regulated by rearrangements of the NCCR which reduces miRNA expression in productively replicating virus [151]. Since no NCCR rearrangements are known to occur in MCPyV, the miRNA promoter might constitute an alternative mechanism to regulate the expression of mcv-miR-M1. Importantly, the miRNA promoter allows miRNA expression independently from late gene expression, implying a functional role of mcv-miR-M1 in non-productive replication e.g. in persistence.

The question remains how mcv-miR-M1 expression from integrated MCPyV genomes is almost completely silenced in MCC. Figure 34B summarizes hypothetical mechanisms causing silencing of mcv-miR-M1 expression in MCC cells. It seems possible that the late promoter, as well as the miRNA promoter, are silenced in integrated MCPyV by epigenetic modifications, which needs to be investigated by further ChIP experiments in MCC cells. Moreover, integration of MCPyV occurs randomly and could lead to linearization downstream of the late promoter preventing late read-through transcripts to reach the mcv-miR-M1 pre-miRNA hairpin. It was shown that MCPyV predominantly integrates at the second exon of LT, mostly downstream of the Rb-binding motif [92-94], which would support this hypothesis. Low levels of mcv-miR-M1 expression might be explained by the fact that MCPyV integrates as concatamers and that low amounts of late read-through transcripts might be produced from a neighboring MCPyV copy, which might be processed into mcv-miR-M1.

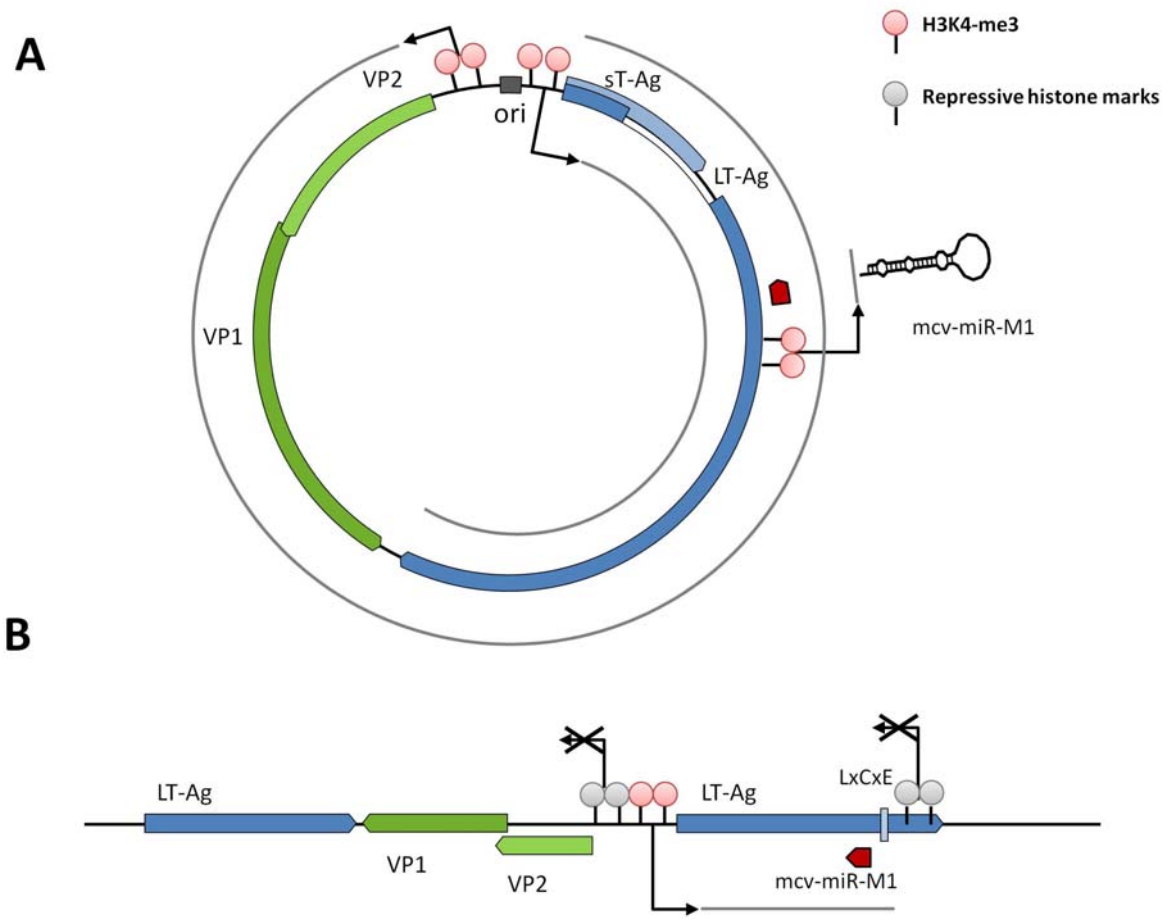

Figure 34: Model of mcv-miR-M1 expression in replicating and integrated MCPyV
(A) Schematic depiction of miRNA expression from circular MCPyV. mcv-miR-M1 is produced from late read-through transcripts or from miRNA specific transcripts originating at the miRNA promoter. Both promoters acquire H3K4-me3 marks during replication of viral episomes. **(B)** Representation of integrated MCPyV in MCC cells. The viral genome is predominantly linearized downstream of the LXCXE motif in the early region. This integration pattern may prevent miRNA expression from the late promoter. The miRNA promoter as well as the late promoter, might be also silenced by repressing histone marks.
Early and late open reading frames are depicted as blue and green bars, respectively. The miRNA is shown as a red arrow. Transcripts are represented by grey lines. Histone modifications are depicted as red (H3K4-me3) and grey (repressive histone modifications) circles.

5.3 The role of mcv-miR-M1 during replication of MCPyV

SV40, MuPyV and the human polyomaviruses JCPyV and BKPyV encode for miRNAs that cause downregulation of early gene expression at late stages of infection [52,53,147]. Luciferase assays suggest a similar function for mcv-miR-M1 [51], although this has so far not been investigated in the context of MCPyV replication.

In this study, a mcv-miR-M1 knockout MCVSyn mutant (MCVSyn-hpko) was used to investigate the role of the miRNA during the viral life cycle. On the microscopic level, no obvious phenotypic

differences between cells transfected with MCVSyn and MCVSyn-hpko were discernible. However, as judged by Western blotting, the expression of LT-Ag from MCVSyn-hpko was markedly enhanced in comparison to the MCVSyn wildtype at four days post transfection. The increased LT-Ag expression in MCVSyn-hpko transfected cells resulted in a more robust viral DNA replication. Thus, similar to other polyomavirus encoded miRNAs [52,53], mcv-miR-M1 downregulates early gene expression during the viral life cycle. As it is the case in BKPyV, this leads to a significant decrease in viral DNA replication [151]. These observations strongly support the notion that, despite their different sequences and genomic locations, all so far identified PyV miRNAs share a conserved functional homology.

5.4 A comprehensive analysis of the MCPyV transcriptome during viral replication

To ensure that the observed effects on LT-Ag levels and viral replication in MCVSyn-hpko did not originate from aberrant gene expression patterns caused by the introduced hpko-mutations, the transcriptomes of MCVSyn and MCVSyn-hpko were analyzed and compared by RNA-Seq. Figure 35 presents a revised model of the transcriptional profile of MCPyV integrating the major findings of this study.

Alternative splicing as well as use of alternative transcriptional start or polyA-sites are long known mechanisms to increase complexity especially for small viruses whose coding capacity is very limited [42]. RNA-Seq analysis provides a high resolution insight into these mechanisms by detecting splice junctions with high sensitivity. So far, no such analysis has been carried out for replicating MCPyV or for any other polyomavirus.

Sequencing of mRNAs from replicating MCVSyn confirmed the existence of alternatively spliced early transcripts that lead to the expression of the four major early mRNAs that were first described by Shuda et al. [66]. Additionally, five minor splicing events were observed in the early orientation which could lead to the expression of so far non-annotated T-Antigens with potentially unique functions. Similar low abundant T'-Antigens were previously detected in JCPyV. At least three T'-mRNAs are produced during productive JCPyV infection by alternative splicing and the resulting T'-Antigens were shown to play a role in viral replication and stabilization of LT-Ag [219]. It remains to be investigated if the T'-mRNAs detected in MCVSyn play comparable roles in MCPyV infection.

Among the newly identified mRNAs, two transcripts use a splice donor site upstream of the early region start region. The loss of the main translational start site could lead to the usage of a downstream start-codon to produce ALTO, a T-Antigen translated from an alternative reading

frame [67]. However, if ALTO is translated from a novel early splice variant by its own mRNA or if occasional read-through of the main start-codon in LT-, sT-, or 57kT-mRNAs is responsible for the expression of ALTO needs to be the further investigated.

In the late orientation non-spliced transcripts encoding for VP2 were most abundant. VP1 specific transcripts which were spliced as predicted could be also detected. Interestingly, another late splice donor was identified which was joined to the VP1-transcript acceptor site seemingly removing an intron of only 26 nt. Although few introns exist which are known to be even less than 20 bp in size [220], most introns have a minimum size of at least 27 bp [138]. Thus, it is likely that the splice donor is not immediately joined to the downstream acceptor site but rather that, enabled by generation of long concatameric transcripts which proceed over the viral genome for several rounds, the splice donor is joined to the acceptor site of a downstream copy of the late region, thereby removing more than genome-sized introns. The resulting mRNA could be translated into VP2 and it remains to be investigated if the splice site has any functional significance e.g. by enhancing transcript stability as shown for MuPyV late transcripts [221].

A third capsid protein, VP3, is expressed by most well studied polyomaviruses. Comparable to these polyomaviruses, a putative VP3 start codon was identified in MCPyV downstream of the VP2 start site. The predicted VP3 coding region lies in frame with the VP2 coding sequence and would thus produce a truncated version of the VP2 protein. *In silico* analysis suggested that a splice might remove the VP2 start codon and function as a VP3 specific transcript. However, RNA-Seq analysis could not confirm the presence of such a transcript during MCVSyn replication. Although VP3 might be also expressed from unspliced transcripts, a recent study revealed that native MCPyV virions lack VP3, thus suggesting that MCPyV belongs to a clade of polyomaviruses which do not express VP3 [47].

Leader-to-leader splicing of late read-through transcripts, as it was described for MuPyV [81,83], was also found in the MCVSyn transcriptome and is illustrated in Figure 35. Late mRNAs of MCVSyn were found to contain multiple copies of the late leader sequence upstream of the coding sequence, which have to be the result of processive RNA polymerase completing several rounds of transcription across the viral genome, combined with inefficient cleavage and polyadenylation of the late-strand transcripts. In MuPyV, leader-to-leader splicing is required for efficient accumulation of late mRNAs and expression of capsid proteins [83]. It is tempting to speculate that increased leader-to-leader splicing in MCVSyn could increase the abundance of late transcripts and lead to an enhanced expression of capsid proteins and particle production. Additionally, introns removed by leader-to-leader splicing might be processed into mature viral miRNAs and thus lead to increased instability of early transcripts. Interestingly, the joining of leader sequences creates a novel ORF which might be translated into an Agnoprotein with comparable biochemical properties to other Agnoproteins.

Future experiments will show if MCVSyn expresses an Agnoprotein in the semi-permissive replication system and if not, whether the putative MCPyV Agnoprotein can facilitate progeny production in semi-permissive cell lines when expressed *in trans*.

In addition to leader-to-leader splicing, a "miRNA-to-leader" splicing was detected by RNA-Seq which was not described for any other PyV so far. In these late transcripts, a splice donor immediately downstream of the miRNA hairpin is joined to the late leader sequence. These mRNAs could be translated into VP proteins and simultaneously serve as miRNA precursors but the significance of miRNA-to-leader splicing remains elusive.

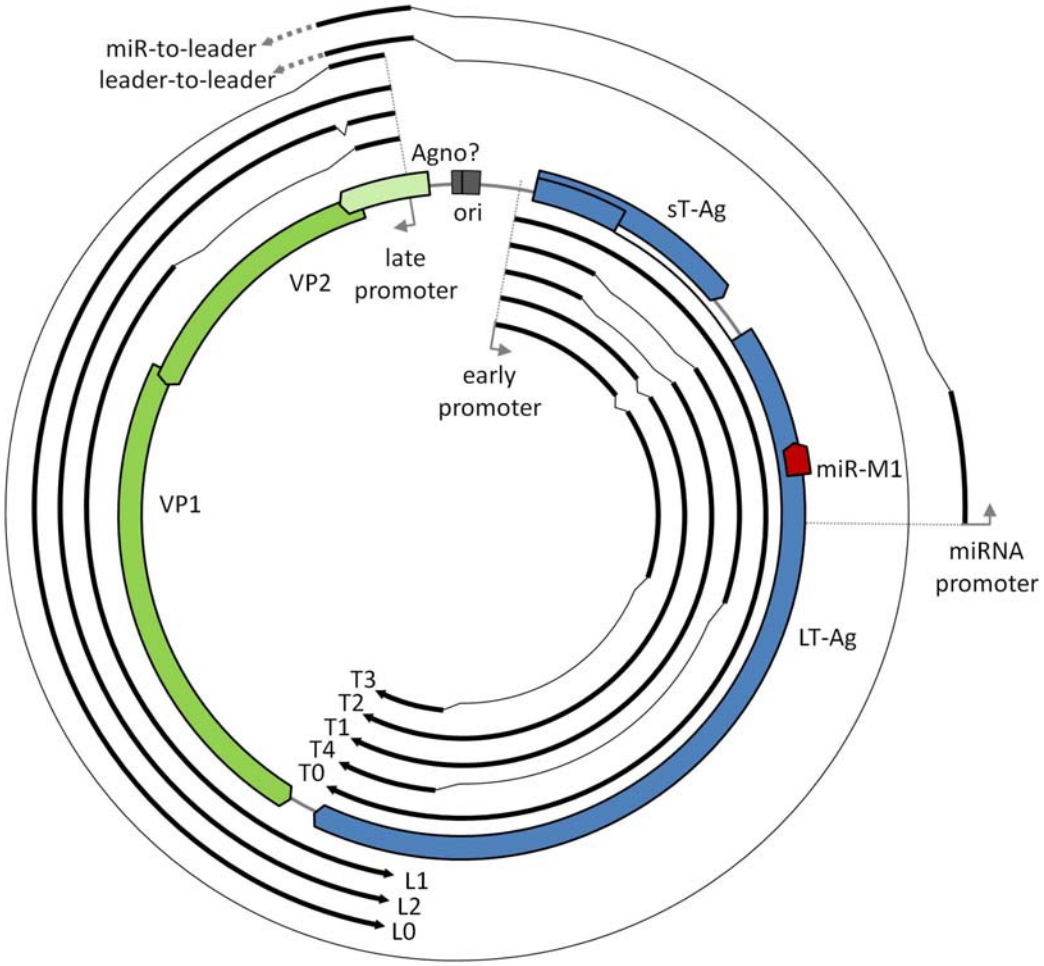

Figure 35: Schematic representation of the transcriptional profile of replicating MCPyV
Revised model of the MCPyV transcriptome including major transcripts and promoter regions. Early and late coding regions are shown as blue and green arrows, respectively. The putative ORF for Agnoprotein is created by leader-to-leader splicing. Transcripts are represented by black arrows with introns as thinner lines.

Although all transcripts described above were detected in wildtype MCVSyn as well as in MCVSyn-hpko, their relative abundances differ significantly. The accumulation of early mRNAs

in MCVSyn-hpko suggests that almost all early transcripts are targeted for degradation by mcv-miR-M1. An exception is the mRNA T'5, which is significantly more abundant in wildtype MCVSyn. This observation is explained by the fact that, due to its splicing pattern, T'5 lacks the complementary region for mcv-miR-M1 and thus does not undergo miRNA-mediated degradation. The accumulation of early transcripts also influences the relative abundance of late mRNAs. While approximately one third of all viral mRNAs in wildtype MCVSyn encode VP1 or VP2, these mRNAs account for only ~5% of all viral mRNAs in MCVSyn-hpko. It must be noted that the abundance of transcripts as determined by RNA-Seq is a relative value. Although, relative to early transcripts, late transcripts in MCVSyn-hpko are less abundant compared to wildtype MCVSyn, it cannot be concluded that the overall amount of late transcripts is reduced in the miRNA mutant.

Comparing the read coverage tracks, which show the overall distribution of transcripts on the viral DNA, no major differences are discernible for MCVSyn and MCVSyn-hpko except for the 5' regions of the second early exon. In MCVSyn, significantly fewer reads were mapped to this region than in MCVSyn-hpko implying that mcv-miR-M1 can cause endonucleolytic cleavage of early mRNAs resulting in turnover of the 5' mRNA fragment and increased stability of the polyadenylated 3' fragment [222]. Thus, mcv-miR-M1 cleaves early transcripts in a siRNA-like fashion.

5.5 mcv-miR-M1 augments episomal persistence of MCPyV

Polyomaviruses are known to establish life-long persistence in their natural host. Although this implies that polyomaviruses must have evolved strategies to evade detection by immune sensors and to maintain asymptomatic viral DNA replication, very little is known about the underlying mechanisms. Therefore, the discovery in this study that MCVSyn episomally persists in transfected cells for several months is a major finding and can significantly contribute to our understanding of MCPyV persistence.

MCVSyn was found to continuously replicate in PFSK-1 cells for several months, which could be monitored by FISH analysis, Southern blotting and qPCR. FISH analysis detecting MCPyV genomes together with immunofluorescence analysis for LT-Ag revealed that 4 d post transfection only approximately 2.5% of all cells harbored MCVSyn genomes. Based on a transfection efficiency of PFSK-1 cells of 30-40%, in less than $1/10^{th}$ of transfected cells correctly re-ligated viral genomes reach the nucleus and initiate viral genome replication. It can be speculated that the cell cycle phase, the acquisition of histone modifications on the viral DNA or other cellular factors play a role in successful establishment of MCVSyn replication. The number of MCVSyn positive cells decreased over time and reached a steady level of estimated 0.01%

positive cells in the population. In contrast to MCC cell lines where integrated MCPyV genomes produce only faint FISH signals, MCVSyn DNA accumulates in large dots in MCVSyn transfected PFSK-1 nuclei. The distribution of LT-Ag in the nuclei was very similar to the overall pattern of MCVSyn DNA positive FISH signals, suggesting that MCVSyn DNA together with LT-Ag forms replication centers as previously shown [69]. The very low frequency of MCVSyn positive cells together with the quantification of viral DNA via qPCR indicates that the replication centers contain high copy numbers of MCVSyn genomes.

By FISH analysis alone, it is not possible to discriminate between extrachromosomal DNA and integrated DNA, as the observed FISH signals could either represent viral minichromosomes clustered in replication foci or concatameric MCVSyn genomes integrated into cellular chromosomes. Rolling circle amplification and HTS analysis revealed that even several weeks after initial transfection, MCVSyn was maintained as circular episome without acquiring any adaptive mutations or undergoing rearrangements of the NCCR.

Considering that the miRNA mutant MCVSyn-hpko exhibits a more robust DNA replication, it seems reasonable to expect that MCVSyn-hpko might be more efficient in the establishment of persistence. Surprisingly, however, MCVSyn-hpko was lost at an accelerated rate compared to wildtype MCVSyn in three replicate experiments. The amount of MCVSyn-hpko DNA dropped below the qPCR detection level at 60-100 days post initial transfection and became undetectable by FISH or Southern blot analysis. LT-Ag expression from MCVSyn-hpko, although initially higher than in MCVSyn transfected cells, showed a diffuse distribution in the nucleus at late time-points and was undetectable after the loss of the viral DNA. Thus, continous expression of mcv-miR-M1 seems to be required for the successful establishment and maintenance of viral persistence.

To further investigate persistence of MCVSyn and the role of mcv-miR-M1 in this process, it is required to isolate or track cells harboring MCVSyn. A direct and simple way to track viruses is the introduction of a reporter gene in the viral genome as it is established for a wide range of viruses [223-226] and which also was attempted during this study. Due to the small size of polyomavirus genomes and the high density of open reading frames, it is difficult to insert a novel gene without affecting viral replication and/or gene expression. To avoid formation of fusion proteins, a FMDV derived 2A sequence was inserted between the early or late ORF and a GFP reporter gene to cause a ribosomal skip [227] (Figure S 1A).

Nevertheless, introduction of the 2A-GFP cassette in the early region resulted in a complete failure of viral DNA replication (Figure S 1B) and gene expression. In contrast to MCVSyn-LT-GFP, MCVSyn-VP1-GFP transfected cells exhibited replication of the episome and expression of early or late viral genes although to a lesser extent than MCVSyn transfected cells (Figure S 1C). Upon sorting of GFP positive cells, a marked enrichment of viral DNA and transcripts was

measurable. Comparable to MCVSyn-hpko, MCVSyn-hpko-VP1-GFP showed a stronger replication and higher LT-Ag mRNA levels at 5 d.p.t. than MCVSyn-VP1-GFP (Figure S 2A).

The MCVSyn-VP1-GFP reporter constructs seemed to be phenotypically similar to MCVSyn in terms of replication, gene expression and miRNA function and even persistence could be observed for several weeks. However, after initial promising experiments, RNA-Seq revealed a marked drawback of the MCVSyn-GFP model by unraveling that the 2A-GFP cassette causes dramatically aberrant splicing of viral transcripts (Figure S 3). Several new donor and acceptor sites were created by insertion of the 2A-GFP cassette and removal of the late region stop codon. More than half of all early splices are produced by joining a novel splice donor in the late region to the acceptor site at position 861. The resulting transcripts lack the major early start codon and thus cannot be translated into any of the predicted T-Antigens except for ALTO. Transcripts in the late orientation of MCVSyn-VP1-GFP accumulate downstream of the 2A-GFP cassette, possibly indicating transcriptional initiation at the GFP locus.

The high amount of read-through transcription in the early orientation and the usage of cryptic splice donor and acceptor sites markedly hamper the usage of MCVSyn-GFP as a reporter virus. Importantly, the effects of the GFP cassette on viral transcription might be extrapolated to other GFP-reporter viruses. Thus, GFP reporter viruses used as model systems should be more closely investigated as it can be hypothesized that they, similar to MCVSyn-GFP, despite replication competence, are influenced by significantly altered transcriptomes.

Although further experiments were carried out in the attempt to isolate MCVSyn harboring cells for a closer investigation of viral persistence and mcv-miR-M1 functions, it was not possible to enrich for MCVSyn positive cells in this study. Future studies will focus on this task e.g. by using fluorescently labeled oligonucleotides.

5.6 The influence of mcv-miR-M1 on viral persistence

A study by Broekema et al. revealed that archetype BKPyV, which is found as persistent form in the healthy population, weakly expresses the early genes, whereas the viral miRNA is robustly expressed and targets early transcripts for degradation, resulting in reduced LT-Ag expression and limited viral replication. In contrast, rearranged BKPyV, which is the disease associated from of BKPyV and produces high levels of progeny, expresses the BKPyV miRNA at only very low levels, leading to strong early gene expression and viral replication [151]. From these observations the authors hypothesize that the BKPyV miRNA might play a role in BKPyV persistence. If that proves correct, the shared function of polyomavirus miRNAs in limiting early gene expression might represent a conserved strategy for establishment and maintainance of lifelong persistence, a feature common among all miRNA expressing polyomaviruses.

However, it remains elusive by which mechanisms polyomavirus miRNAs contribute to viral persistence. Since mcv-miR-M1 reduces early viral gene expression and consequently viral replication, high levels of LT-Ag expression, excessive replication of viral DNA or a combination of both processes might be disadvantageous for persistence of MCPyV. Moreover, cellular targets of mcv-miR-M1 might play a role in the establishment or maintenance of viral persistence.

5.6.1 Progeny production and re-infection

The term persistence describes the continuous presence of viral genomes in the host organism but does not specify how the virus is maintained [228]. Generally, mechanisms of viral persistence include continuous low-level (or chronic) particle production and re-infection or maintenance of the viral genome in the absence of viral particle production.

MCVSyn was shown to be able to produce low amounts of particles in the system used here, however these particles appear to be smaller and irregular in size and shape [229]. PFSK-1 cells could be re-infected with lysates prepared from MCVSyn and MCVSyn-hpko transfected cells after several freeze-thaw-cycles and DNase I resistant MCVSyn DNA was detectable in supernatants of transfected cells. However, no significant difference in particle production or infectivity was detectable between wild type MCVSyn and MCVSyn-hpko.

While a BKPyV archetype miRNA knockout mutant shows increased particle production [151], this could not be reproduced for MCPyV, likely caused by the general block of particle production by MCVSyn in the model system used in this study. A fully permissive MCPyV model system is required to analyze the influence of mcv-miR-M1 on progeny production. So far, the described observations suggest that the accelerated loss of MCVSyn-hpko is not caused by a defect in particle production. However, if MCVSyn persistence is dependent on re-infection needs to be further investigated e.g. by generation of virus mutants which are defective in VP1 expression.

5.6.2 Cellular targets of mcv-miR-M1

Various viral miRNAs were described to regulate host cell gene expression. Among them, some are directly involved in viral persistence and latency or contribute indirectly to maintenance of persistence e.g. by host immune evasion [86,128,130]. Although evidence exists that maintenance of the autoregulatory function of polyomavirus miRNAs is under more stringent evolutionary pressure than the conservation of cellular targets [150], mcv-miR-M1 might still modulate cellular gene expression to create a favorable environment for the maintenance of viral replication, for example by downregulating factors of the innate immune system.

The perfect base pairing between mcv-miR-M1 and early mRNAs and the resulting cleavage of the target transcripts is a rare mechanism in vertebrate cells [230]. Except for plants, miRNAs typically bind with imperfect complementarity at the 3'UTR of the target mRNA and cause instability and translational repression [109,231]. The binding specificity of a miRNA is mainly determined by nucleotides 2-8 of the miRNA, the so-called seed region, which have to bind with perfect base pairing to a target RNA to cause repression of translation [141]. However, searching for seed matches in all 3'UTRs of the human transcriptome usually yields hundreds or thousands of potential target sites. Generally, the rules of miRNA target recognition are still poorly understood and thus most target prediction algorithms are based on finding seed matches in 3'UTRs of mRNAs resulting in exceedingly high rates of false positives [127,232].

In this study, gene expression profiling was carried out by Microarray and RNA-Seq analyses in PFSK-1 cells overexpressing mcv-miR-M1 or harboring either MCVSyn, MCVSyn-hpko, MCVSyn-GFP or MCVSyn-GFP-hpko. These experiments yielded several hundred mildly regulated transcripts. However, no overlap of significantly downregulated genes upon mcv-miR-M1 expression could be observed between individual experiments. Although mRNA levels seem to be largely constant upon mcv-miR-M1 expression, targets repressed at translational level are missed by these methods. Further, PFSK-1 cells are not the natural host cell of MCPyV and thus possible target mRNAs for mcv-miR-M1 might not even be expressed in these cells.

An alternative way of finding miRNA targets is to analyze possible conserved functions of miRNAs. Different herpesvirus and the polyomavirus miRNAs of JCPyV and BKPyV were shown to downregulate the stress induced NKG2D ligands ULBP3 or MICB, respectively to avoid natural killer cells [233,234]. Since these viruses share their persistent life cycle with MCPyV, which likely requires strategies for immune evasion, it is possible that mcv-miR-M1 as well targets NKG2D ligand expression. ULBP3 and MICB expression is not detectable in PFSK-1 cells and thus these genes were not captured with the above described high throughput methods. A prediction of possible mcv-miR-M1 binding sites in NKG2D ligands using RNAhybrid 2.2 [162] revealed high confident seed matches in the 3'UTR of MICB. However, overexpression of mcv-miR-M1 in MICB expressing RKO-cells did not result in reduced MICB protein levels (Figure S 4).

Currently, direct identification of miRNA targets by recently established CLIP-methods (Cross-Linking and ImmunoPrecipitation) in combination with High Throughput Sequencing is carried out to identify mcv-miR-M1 targets with high confidence and to investigate if mcv-miR-M1 shares cellular targets with the cellular miRNA has-miR-7 which is a seed homolog to mcv-miR-M1.

5.6.3 Effects of LT-Ag on cellular proliferation

The SV40 miRNA was shown to reduce the susceptibility to cytotoxic T-lymphocytes by reducing the levels of LT-Ag [52]. However, this mechanism, if applicable to MCPyV, could only play a role *in vivo* but not in cell culture. Besides immunogenic properties of LT-Ag, miRNA mediated reduction of T-Ag levels and viral DNA amplification could be advantageous during MCVSyn persistence since LT-Ag was shown to induce DNA damage response (DDR) leading to inhibition of cellular proliferation [165,235]. Although previously published DNA damage markers [165] were not increased by LT-Ag expression in PFSK-1 cells, LT-Ag expressing cells nevertheless showed a markedly reduced proliferation rate.

ChIP-Seq analysis revealed that in these cells, LT-Ag binds to the host cell DNA, preferably at promoter regions, which are significantly enriched in GRGGC motifs, potential binding sites for LT-Ag. However, it remains to be investigated if binding of LT-Ag at the cellular chromatin is indeed dependent on GRGGC motifs or if they are merely overrepresented in these regions due to the high prevalence of CpG islands in LT-Ag peaks. Approximately 70% of annotated vertebrate promoters contain CpG-islands while most other genomic regions are CpG depleted [107]. Since 60% of LT-Ag peaks in promoter and coding regions contain CpG islands, it is possible that LT-Ag does not bind these regions due to its OBD affinity to GC-rich regions, but rather that binding of LT-Ag is facilitated at open chromatin structures such as actively transcribed regions.

The KSHV protein LANA also binds transcriptionally active regions at the cellular genome, although the cause and consequences of this binding pattern is only incompletely understood [78]. As hypothesized for LANA, LT-Ag might also interact with chromatin remodeling complexes and play a role in formation of the epigenetic profile of MCPyV episomes. Binding of LT-Ag to cellular chromatin might also play a role in tethering of viral episomes during cell division to ensure faithful segregation of the viral DNA into daughter cells. In this case, LT-Ag binding at promoter regions might be a consequence of enhanced accessibility of chromatin at transcriptionally active regions. In papillomaviruses, BRD4 mediates the tethering of viral DNA to host mitotic chromosomes to segregate viral genomes to daughter cells [216]. BRD4 was shown to interact with LT-Ag and is required for viral DNA replication [69]. The association of LT-Ag and BRD4 with viral replication complexes as well as cellular chromatin might play a role in tethering of viral episomes in non-productive infection.

Strikingly, truncated LT-Ag either ectopically overexpressed or produced from integrated MCPyV in WaGa cells did not exhibit an increased affinity to GRGGC motifs on the cellular chromatin most likely caused by the lack of the C-terminus, including origin binding and helicase-ATPase function. Thus, if LT-Ag multimers could assemble at the host cell genome similar to LT-Ag multimers at the viral origin of replication, it may seem possible that the

helicase-ATPase domain catalyzes unwinding of the DNA followed by recruitment of DNA polymerase. The resulting unlicensed replication of DNA fragments might cause genomic instability and induce a DNA damage response. Future experiments will further address if the binding of LT-Ag on the host cell genome has any implications on cellular gene expression, induction of genomic instability, or integration of MCPyV into the host cell genome.

Due to the very low percentage of cells that harbor replicating MCVSyn upon transfection with re-ligated viral genomes, direct measurement of cellular proliferation is not possible in bulk cultures. A method for enrichment or separation of MCVSyn positive cells is required to conduct further studies on the influence of MCPyV on cellular proliferation. Nevertheless, the above described observations indicate that cells harboring MCVSyn-hpko have a growth disadvantage in comparison to MCVSyn transfected cells due to elevated LT-Ag levels and stronger viral replication, probably resulting in activation of pathways which lead to DNA damage response and stalling of cell cycle progression.

Thus, cells harboring MCVSyn and producing LT-Ag have a growth disadvantage in comparison to non-transfected cells and thus are overgrown or diluted out over time. While MCVSyn transfected cells display only moderate LT-Ag levels, LT-Ag expression is not limited by mcv-miR-M1 in MCVSyn-hpko transfected cells. Hence, cells harboring MCVSyn-hpko might be stronger impeded in proliferation and therefore are lost from the bulk culture at an accelerated rate.

5.7 Model of the role of mcv-miR-M1 in viral replication and persistence

The following model integrates the most important findings of this study with regard to mcv-miR-M1 function during MCPyV replication. mcv-miR-M1 is highly expressed during replication of MCVSyn by the combined transcription of the pri-miRNA from the late promoter and an autonomous miRNA promoter, both of which are marked by activating histone modifications (Figure 36). Mature mcv-miR-M1 downregulates early gene expression by perfect base pairing with early transcripts, leading to siRNA-like degradation of bound mRNAs. Thereby, the LT-Ag dependent replication of viral DNA is indirectly limited as well. This observation raises the question how the virus benefits from a reduced LT-Ag expression and genome replication. Strikingly, an effect of these mechanisms was observed during long-term replication of viral episomes in the semi-permissive MCPyV replication system: the loss of mcv-miR-M1 expression reduces the ability of MCVSyn to persist despite the stronger initial viral replication.

The establishment of MCPyV persistence underlies the pressure to maintain viral replication and to prevent damage to the host cell or clearance by the host's immune system. Clearly, LT-Ag is needed for continuous replication of viral episomes, but it also induces DNA damage response

[165] and likely causes T-cell response against infected cells [52]. Figure 36 illustrates the dose-dependent effects of LT-Ag on MCPyV infected cells: persistently infected cells require a minimum of LT-Ag to maintain viral replication, since complete abrogation of LT-Ag expression would lead to loss of viral episomes over time. High levels of LT-Ag, on the other hand, cause massive viral replication and also have detrimental effects on the host cell which might result in clearance of infected cells. LT-Ag markedly reduces the cellular proliferation rate, probably caused by binding of LT-Ag to cellular DNA, induction of DNA damage response or both. Binding of LT-Ag to cellular chromatin might also influence host cell gene expression or facilitate integration of viral DNA, which is massively replicated in presence of high LT-Ag levels, into the host cell genome. Moreover, LT-Ag might induce an immune response leading to virus clearance.
Taken together, the results of this study propose a model in which establishment of MCPyV persistence requires a delicate balance between minimal LT-Ag levels, which are required for episome maintenance, and maximum levels that do not exhibit detrimental effects on cell growth or survival. Since mcv-miR-M1 downregulates LT-Ag expression and a miRNA mutant virus was not able to persist in cell culture as long as the wildtype virus, it can be hypothesized that mcv-miR-M1 acts as a fine tuner of LT-Ag expression to ensure efficient establishment of persistence. A miRNA appears to be the perfect tool to fulfill this task, owing to its non-immunogenic properties and the fact that miRNAs do not cause absolute silencing of target gene expression. Even despite the perfect complementarity of mcv-miR-M1 to LT-Ag transcripts and the presumed siRNA-like cleavage of early transcripts, LT-Ag is still expressed at low levels during replication of MCVSyn. The observation that mcv-miR-M1 can be expressed independently from late gene expression suggests that it is required also in the absence of capsid production, as it might be the case during viral persistence. Additionally, mcv-miR-M1 might influence the expression of host cell genes to create a favorable environment for viral persistence.

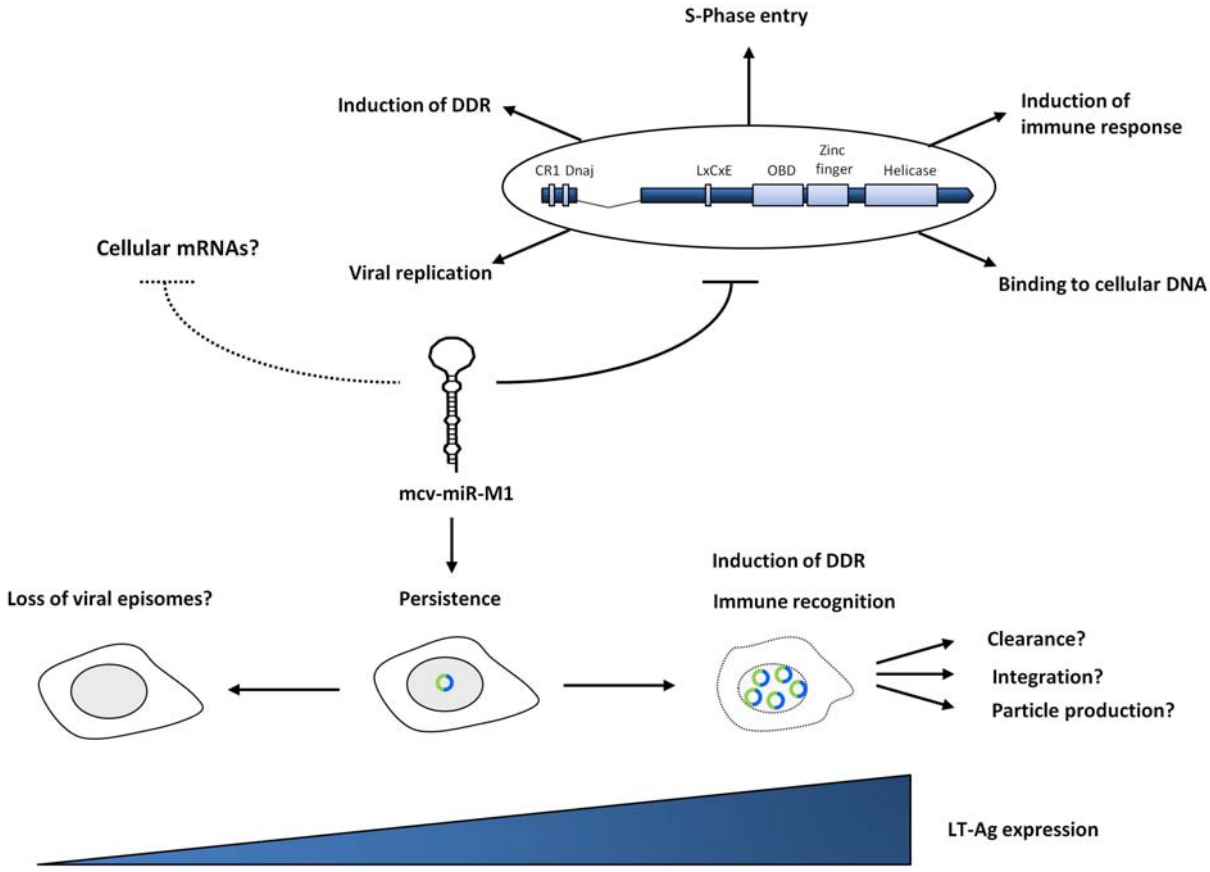

Figure 36: Model of the role of mcv-miR-M1 in the MCPyV life cycle
mcv-miR-M1 supports the establishment of viral persistence by limiting LT-Ag expression and possibly also influencing the hosts cell gene expression. While too high concentrations of LT-Ag not only cause strong viral replication but also have detrimental effects on the host cell, e.g. by induction of DDR or immune responses, abrogation of LT-Ag expression might lead to loss of viral episomes. Thus, a miRNA-mediated regulation of LT-Ag expression represents a plausible mechanism to maintain asymptomatic viral persistence.

5.8 Conclusions and Outlook

Despite decades of polyomavirus research, it is still a major challenge to establish model systems, which support the entire viral life cycle. Consequently, little is known about the natural biology of most of these viruses. Human polyomaviruses establish life-long infections in their host, but the mechanisms that promote establishment and maintenance of persistence are in most parts elusive. However, the rising incidence of severe polyomavirus associated diseases in immunocompromised patients underscores the necessity to gain a more profound understanding not only about the disease-associated processes but also about how polyomaviruses are able to circumvent the host's immune system and establish persistence, which is a prerequisite for the development of polyomavirus associated disorders and, in case of MCPyV, malignancies.

Although MCPyV is so far unique among human polyomaviruses in its ability to cause tumors in its natural host, it also establishes a lifelong asymptomatic maintenance. Thus, MCPyV requires strategies to not disturb cellular homeostasis and evade immune surveillance. But, due to the lack of appropriate model systems, these processes could not yet be investigated.

In this study, for the first time, persistence of MCPyV could be observed and studied using a cell culture based semi-permissive replication system. Strikingly, using this system, the viral miRNA mcv-miR-M1 could be identified as a factor, which contributes to the establishment of long-term viral persistence. It could be demonstrated that mcv-miR-M1 reduces viral replication by limiting expression of LT-Ag during authentic viral replication. Expression of LT-Ag is necessary during viral persistence to maintain viral replication and for driving resting cells into the S-phase. However, LT-Ag also induces DNA damage response [165], hampers cellular proliferation [195] and is immunogenic *in vivo* [52], properties which can be considered as a disadvantage for the maintenance of viral persistence. These observations suggest a model in which mcv-miR-M1 is required to balance LT-Ag expression to allow continuous viral replication but avoid detection by the host cell and thus contributes to the establishment of viral persistence.

A comprehensive epigenetic and transcriptomic analysis of replicating MCPyV, which was carried out in this work, not only allowed for the first time to draw an accurate map of the viral transcriptome, but also revealed that mcv-miR-M1 can be expressed by an autonomous promoter, independently from NCCR initiated transcription. It can be hypothesized that this might play a role when late transcription is reduced, as it might be the case during viral persistence.

Although miRNAs are generally considered as fine tuners of gene expression [233], this work has shown that mcv-miR-M1 has a fundamental influence on the successful establishment of viral persistence and raises the question if this miRNA function is conserved among polyomaviruses, as already suggested for the BKPyV miRNA [151]. The role of mcv-miR-M1 in viral persistence adds further evidence to the hypothesis that many viral miRNAs are important players in viral persistence or latency by modulating the host's or the viral gene expression without being recognized by the infected cell.

Although high throughput sequencing enabled a detailed insight into the MCVSyn transcription program and histone modifications, the potential of this model system is limited by the low amount of cells (3-5%) that harbor replicating viral genomes. The high number of non-transfected cells in replication assay bulk cultures impedes the detailed analysis of replicating and persistent MCVSyn. Thus, at present it is impossible to track individual MCVSyn positive cells to further investigate viral persistence e.g. by analysis of the chromatin status and the influence of cellular proliferation. Furthermore, it remains possible that mcv-miR-M1 may behave differently in a fully permissive system. Therefore it is crucial to find a method to isolate

or to enrich for cells harboring replicating MCVSyn, or to establish a fully permissive model system for MCPyV.

Moreover, besides a higher yield of MCPyV infected cells, a model system that is closer to the natural host cell would provide the possibility to identify authentic cellular targets of mcv-miR-M1. Further, transcriptomic or epigenetic studies could be carried out to analyze the impact of MCPyV infection on the host cell gene expression. The establishment of productive model systems would also provide insight into the impact of mcv-miR-M1 in particle production and infection and to determine if the miRNA is continuously expressed during all phases of the viral replication cycle. However, finding the natural reservoir of MCPyV will be a challenging task. While MCC is thought to arise from Merkel cells [236] or early B-cells [225,227] and MCPyV particles are shed from the skin, these cells do not necessarily have to represent the natural host cell for MCPyV [19,86-89].

Since not all aspects of the MCPyV life cycle can be investigated in cell culture model systems, the establishment of animal models of MCPyV infection is also highly desirable. It will be exciting to investigate if the phenotype observed for mcv-miR-M1 in cell culture is reproducible in an *in vivo* situation and if mcv-miR-M1 has any implications in virus spread.

This study was also the first to show that a polyomavirus miRNA can be expressed independently from the late promoter. Further characterization of the identified miRNA promoter might reveal how this promoter can be activated or silenced in different phases of the viral life cycle. Since MCPyV positive MCC tumor cells are dependent on the presence of viral T-Antigens, reactivation of the mcv-miR-M1 promoter in MCC cells could cause downregulation of T-Antigen expression and thus might represent a future approach for the treatment of MCPyV positive MCC.

This study significantly contributes to our understanding of polyomavirus miRNAs and their possible role in viral persistence. It further underscores the importance of resolving the functions of viral miRNAs to understand viral life cycles and virus-associated diseases.

6 References

1. Toker C (1972) Trabecular carcinoma of the skin. Arch Dermatol 105: 107-110.
2. De Wolff-Peeters C, Marien K, Mebis J, Desmet V (1980) A cutaneous APUDoma or Merkel cell tumor? A morphologically recognizable tumor with a biological and histological malignant aspect in contrast with its clinical behavior. Cancer 46: 1810-1816.
3. Munde PB, Khandekar SP, Dive AM, Sharma A (2013) Pathophysiology of merkel cell. J Oral Maxillofac Pathol 17: 408-412.
4. Tilling T, Moll I (2012) Which are the cells of origin in merkel cell carcinoma? J Skin Cancer 2012: 680410.
5. Zur Hausen A, Rennspiess D, Winnepenninckx V, Speel EJ, Kurz AK (2013) Early B-cell differentiation in Merkel cell carcinomas: clues to cellular ancestry. Cancer Res 73: 4982-4987.
6. Agelli M, Clegg LX (2003) Epidemiology of primary Merkel cell carcinoma in the United States. J Am Acad Dermatol 49: 832-841.
7. Becker JC, Schrama D, Houben R (2009) Merkel cell carcinoma. Cell Mol Life Sci 66: 1-8.
8. Hodgson NC (2005) Merkel cell carcinoma: changing incidence trends. J Surg Oncol 89: 1-4.
9. Heath M, Jaimes N, Lemos B, Mostaghimi A, Wang LC, et al. (2008) Clinical characteristics of Merkel cell carcinoma at diagnosis in 195 patients: the AEIOU features. J Am Acad Dermatol 58: 375-381.
10. Agelli M, Clegg LX, Becker JC, Rollison DE (2010) The etiology and epidemiology of merkel cell carcinoma. Curr Probl Cancer 34: 14-37.
11. Lemos BD, Storer BE, Iyer JG, Phillips JL, Bichakjian CK, et al. (2010) Pathologic nodal evaluation improves prognostic accuracy in Merkel cell carcinoma: analysis of 5823 cases as the basis of the first consensus staging system. J Am Acad Dermatol 63: 751-761.
12. Feng H, Shuda M, Chang Y, Moore PS (2008) Clonal integration of a polyomavirus in human Merkel cell carcinoma. Science 319: 1096-1100.
13. Imperiale MJ, Major EO (2007) Polyomaviruses; Knipe DM, Howley PM, editors. Fields Virology Lippincott Williams and Wilkins, Philadelphia, PA
14. Gross L (1953) A filterable agent, recovered from Ak leukemic extracts, causing salivary gland carcinomas in C3H mice. Proc Soc Exp Biol Med 83: 414-421.
15. Sweet BH, Hilleman MR (1960) The vacuolating virus, S.V. 40. Proc Soc Exp Biol Med 105: 420-427.
16. Girardi AJ, Sweet BH, Slotnick VB, Hilleman MR (1962) Development of tumors in hamsters inoculated in the neonatal period with vacuolating virus, SV-40. Proc Soc Exp Biol Med 109: 649-660.
17. Eddy BE, Borman GS, Grubbs GE, Young RD (1962) Identification of the oncogenic substance in rhesus monkey kidney cell culture as simian virus 40. Virology 17: 65-75.
18. Howley PM, Livingston DM (2009) Small DNA tumor viruses: large contributors to biomedical sciences. Virology 384: 256-259.
19. Pipas JM (2009) SV40: Cell transformation and tumorigenesis. Virology 384: 294-303.
20. Padgett BL, Walker DL, ZuRhein GM, Eckroade RJ, Dessel BH (1971) Cultivation of papova-like virus from human brain with progressive multifocal leucoencephalopathy. Lancet 1: 1257-1260.
21. Gardner SD, Field AM, Coleman DV, Hulme B (1971) New human papovavirus (B.K.) isolated from urine after renal transplantation. Lancet 1: 1253-1257.
22. Schowalter RM, Pastrana DV, Pumphrey KA, Moyer AL, Buck CB (2010) Merkel cell polyomavirus and two previously unknown polyomaviruses are chronically shed from human skin. Cell Host Microbe 7: 509-515.

23. Scuda N, Hofmann J, Calvignac-Spencer S, Ruprecht K, Liman P, et al. (2011) A novel human polyomavirus closely related to the african green monkey-derived lymphotropic polyomavirus. J Virol 85: 4586-4590.
24. Buck CB, Phan GQ, Raiji MT, Murphy PM, McDermott DH, et al. (2012) Complete genome sequence of a tenth human polyomavirus. J Virol 86: 10887.
25. Korup S, Rietscher J, Calvignac-Spencer S, Trusch F, Hofmann J, et al. (2013) Identification of a novel human polyomavirus in organs of the gastrointestinal tract. PLoS One 8: e58021.
26. Allander T, Andreasson K, Gupta S, Bjerkner A, Bogdanovic G, et al. (2007) Identification of a third human polyomavirus. J Virol 81: 4130-4136.
27. Gaynor AM, Nissen MD, Whiley DM, Mackay IM, Lambert SB, et al. (2007) Identification of a novel polyomavirus from patients with acute respiratory tract infections. PLoS Pathog 3: e64.
28. Siebrasse EA, Reyes A, Lim ES, Zhao G, Mkakosya RS, et al. (2012) Identification of MW polyomavirus, a novel polyomavirus in human stool. J Virol 86: 10321-10326.
29. Lim ES, Reyes A, Antonio M, Saha D, Ikumapayi UN, et al. (2013) Discovery of STL polyomavirus, a polyomavirus of ancestral recombinant origin that encodes a unique T antigen by alternative splicing. Virology 436: 295-303.
30. Mishra N, Pereira M, Rhodes RH, An P, Pipas JM, et al. (2014) Identification of a novel polyomavirus in a pancreatic transplant recipient with retinal blindness and vasculitic myopathy. J Infect Dis 210: 1595-1599.
31. van der Meijden E, Janssens RW, Lauber C, Bouwes Bavinck JN, Gorbalenya AE, et al. (2010) Discovery of a new human polyomavirus associated with trichodysplasia spinulosa in an immunocompromized patient. PLoS Pathog 6: e1001024.
32. Kean JM, Rao S, Wang M, Garcea RL (2009) Seroepidemiology of human polyomaviruses. PLoS Pathog 5: e1000363.
33. Pastrana DV, Tolstov YL, Becker JC, Moore PS, Chang Y, et al. (2009) Quantitation of human seroresponsiveness to Merkel cell polyomavirus. PLoS Pathog 5: e1000578.
34. Viscidi RP, Rollison DE, Sondak VK, Silver B, Messina JL, et al. (2011) Age-specific seroprevalence of Merkel cell polyomavirus, BK virus, and JC virus. Clin Vaccine Immunol 18: 1737-1743.
35. Nguyen NL, Le BM, Wang D (2009) Serologic evidence of frequent human infection with WU and KI polyomaviruses. Emerg Infect Dis 15: 1199-1205.
36. Ferenczy MW, Marshall LJ, Nelson CD, Atwood WJ, Nath A, et al. (2012) 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.
37. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32: 1792-1797.
38. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, et al. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol 59: 307-321.
39. Rambaut A (2008) FigTree: Tree figure drawing tool, version 1.4.2.
40. Dela Cruz FN, Jr., Giannitti F, Li L, Woods LW, Del Valle L, et al. (2013) Novel polyomavirus associated with brain tumors in free-ranging raccoons, western United States. Emerg Infect Dis 19: 77-84.
41. DeCaprio JA, Ludlow JW, Figge J, Shew JY, Huang CM, et al. (1988) SV40 large tumor antigen forms a specific complex with the product of the retinoblastoma susceptibility gene. Cell 54: 275-283.
42. Zheng ZM (2010) Viral oncogenes, noncoding RNAs, and RNA splicing in human tumor viruses. Int J Biol Sci 6: 730-755.
43. Lilyestrom W, Klein MG, Zhang R, Joachimiak A, Chen XS (2006) 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.

44. DeCaprio JA, Garcea RL (2013) A cornucopia of human polyomaviruses. Nat Rev Microbiol 11: 264-276.
45. Stubdal H, Zalvide J, Campbell KS, Schweitzer C, Roberts TM, et al. (1997) Inactivation of pRB-related proteins p130 and p107 mediated by the J domain of simian virus 40 large T antigen. Mol Cell Biol 17: 4979-4990.
46. Sowd GA, Fanning E (2012) A wolf in sheep's clothing: SV40 co-opts host genome maintenance proteins to replicate viral DNA. PLoS Pathog 8: e1002994.
47. Schowalter RM, Buck CB (2013) The Merkel cell polyomavirus minor capsid protein. PLoS Pathog 9: e1003558.
48. Gerits N, Moens U (2012) Agnoprotein of mammalian polyomaviruses. Virology 432: 316-326.
49. Bauman Y, Nachmani D, Vitenshtein A, Tsukerman P, Drayman N, et al. (2011) 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.
50. Chen CJ, Cox JE, Azarm KD, Wylie KN, Woolard KD, et al. (2015) Identification of a polyomavirus microRNA highly expressed in tumors. Virology 476: 43-53.
51. Seo GJ, Chen CJ, Sullivan CS (2009) Merkel cell polyomavirus encodes a microRNA with the ability to autoregulate viral gene expression. Virology 383: 183-187.
52. Sullivan CS, Grundhoff AT, Tevethia S, Pipas JM, Ganem D (2005) SV40-encoded microRNAs regulate viral gene expression and reduce susceptibility to cytotoxic T cells. Nature 435: 682-686.
53. Sullivan CS, Sung CK, Pack CD, Grundhoff A, Lukacher AE, et al. (2009) Murine Polyomavirus encodes a microRNA that cleaves early RNA transcripts but is not essential for experimental infection. Virology 387: 157-167.
54. White MK, Gordon J, Khalili K (2013) The rapidly expanding family of human polyomaviruses: recent developments in understanding their life cycle and role in human pathology. PLoS Pathog 9: e1003206.
55. Dalianis T, Hirsch HH (2013) Human polyomaviruses in disease and cancer. Virology 437: 63-72.
56. Carter JJ, Paulson KG, Wipf GC, Miranda D, Madeleine MM, et al. (2009) Association of Merkel cell polyomavirus-specific antibodies with Merkel cell carcinoma. J Natl Cancer Inst 101: 1510-1522.
57. Tolstov YL, Pastrana DV, Feng H, Becker JC, Jenkins FJ, et al. (2009) Human Merkel cell polyomavirus infection II. MCV is a common human infection that can be detected by conformational capsid epitope immunoassays. Int J Cancer 125: 1250-1256.
58. Gustafsson B, Honkaniemi E, Goh S, Giraud G, Forestier E, et al. (2012) KI, WU, and Merkel cell polyomavirus DNA was not detected in guthrie cards of children who later developed acute lymphoblastic leukemia. J Pediatr Hematol Oncol 34: 364-367.
59. Chen T, Hedman L, Mattila PS, Jartti T, Ruuskanen O, et al. (2011) Serological evidence of Merkel cell polyomavirus primary infections in childhood. J Clin Virol 50: 125-129.
60. Tolstov YL, Knauer A, Chen JG, Kensler TW, Kingsley LA, et al. (2011) Asymptomatic primary Merkel cell polyomavirus infection among adults. Emerg Infect Dis 17: 1371-1380.
61. Spurgeon ME, Lambert PF (2013) Merkel cell polyomavirus: a newly discovered human virus with oncogenic potential. Virology 435: 118-130.
62. Bofill-Mas S, Rodriguez-Manzano J, Calgua B, Carratala A, Girones R (2010) Newly described human polyomaviruses Merkel cell, KI and WU are present in urban sewage and may represent potential environmental contaminants. Virol J 7: 141.
63. Foulongne V, Courgnaud V, Champeau W, Segondy M (2011) Detection of Merkel cell polyomavirus on environmental surfaces. J Med Virol 83: 1435-1439.
64. Kwun HJ, Guastafierro A, Shuda M, Meinke G, Bohm A, et al. (2009) The minimum replication origin of merkel cell polyomavirus has a unique large T-antigen loading architecture and requires small T-antigen expression for optimal replication. J Virol 83: 12118-12128.

65. Shuda M, Arora R, Kwun HJ, Feng H, Sarid R, et al. (2009) Human Merkel cell polyomavirus infection I. MCV T antigen expression in Merkel cell carcinoma, lymphoid tissues and lymphoid tumors. Int J Cancer 125: 1243-1249.
66. Shuda M, Feng H, Kwun HJ, Rosen ST, Gjoerup O, et al. (2008) T antigen mutations are a human tumor-specific signature for Merkel cell polyomavirus. Proc Natl Acad Sci U S A 105: 16272-16277.
67. Carter JJ, Daugherty MD, Qi X, Bheda-Malge A, Wipf GC, et al. (2013) 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: 12744-12749.
68. Liu X, Hein J, Richardson SC, Basse PH, Toptan T, et al. (2011) Merkel cell polyomavirus large T antigen disrupts lysosome clustering by translocating human Vam6p from the cytoplasm to the nucleus. J Biol Chem 286: 17079-17090.
69. Wang X, Li J, Schowalter RM, Jiao J, Buck CB, et al. (2012) Bromodomain protein Brd4 plays a key role in Merkel cell polyomavirus DNA replication. PLoS Pathog 8: e1003021.
70. Pallas DC, Shahrik LK, Martin BL, Jaspers S, Miller TB, et al. (1990) Polyoma small and middle T antigens and SV40 small t antigen form stable complexes with protein phosphatase 2A. Cell 60: 167-176.
71. Schowalter RM, Reinhold WC, Buck CB (2012) Entry tropism of BK and Merkel cell polyomaviruses in cell culture. PLoS One 7: e42181.
72. Neumann F, Borchert S, Schmidt C, Reimer R, Hohenberg H, et al. (2011) Replication, gene expression and particle production by a consensus Merkel Cell Polyomavirus (MCPyV) genome. PLoS One 6: e29112.
73. Feng H, Kwun HJ, Liu X, Gjoerup O, Stolz DB, et al. (2011) Cellular and viral factors regulating Merkel cell polyomavirus replication. PLoS One 6: e22468.
74. Schowalter RM, Pastrana DV, Buck CB (2011) Glycosaminoglycans and sialylated glycans sequentially facilitate Merkel cell polyomavirus infectious entry. PLoS Pathog 7: e1002161.
75. Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem 72: 248-254.
76. Neu U, Bauer J, Stehle T (2011) Viruses and sialic acids: rules of engagement. Curr Opin Struct Biol 21: 610-618.
77. Carninci P, Sandelin A, Lenhard B, Katayama S, Shimokawa K, et al. (2006) Genome-wide analysis of mammalian promoter architecture and evolution. Nat Genet 38: 626-635.
78. Hu J, Yang Y, Turner PC, Jain V, McIntyre LM, et al. (2014) LANA binds to multiple active viral and cellular promoters and associates with the H3K4methyltransferase hSET1 complex. PLoS Pathog 10: e1004240.
79. Gai D, Chang YP, Chen XS (2010) Origin DNA melting and unwinding in DNA replication. Curr Opin Struct Biol 20: 756-762.
80. White MK, Safak M, Khalili K (2009) Regulation of gene expression in primate polyomaviruses. J Virol 83: 10846-10856.
81. Hyde-DeRuyscher RP, Carmichael GG (1990) 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: 5823-5832.
82. Hyde-DeRuyscher R, Carmichael GG (1988) 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: 8993-8997.
83. Adami GR, Marlor CW, Barrett NL, Carmichael GG (1989) Leader-to-leader splicing is required for efficient production and accumulation of polyomavirus late mRNAs. J Virol 63: 85-93.
84. Bethge T, Hachemi HA, Manzetti J, Gosert R, Schaffner W, et al. (2015) Sp1 sites in the noncoding control region of BK polyomavirus are key regulators of bidirectional viral early and late gene expression. J Virol 89: 3396-3411.

85. Carswell S, Alwine JC (1986) Simian virus 40 agnoprotein facilitates perinuclear-nuclear localization of VP1, the major capsid protein. J Virol 60: 1055-1061.
86. Moore PS, Chang Y (2010) Why do viruses cause cancer? Highlights of the first century of human tumour virology. Nat Rev Cancer 10: 878-889.
87. zur Hausen H (2008) A specific signature of Merkel cell polyomavirus persistence in human cancer cells. Proc Natl Acad Sci U S A 105: 16063-16064.
88. Oliveira ML, Brochado SM, Sogayar MC (1999) Mechanisms of cell transformation induced by polyomavirus. Braz J Med Biol Res 32: 861-865.
89. Imperiale MJ (2001) Oncogenic transformation by the human polyomaviruses. Oncogene 20: 7917-7923.
90. Kassem A, Schopflin A, Diaz C, Weyers W, Stickeler E, et al. (2008) Frequent detection of Merkel cell polyomavirus in human Merkel cell carcinomas and identification of a unique deletion in the VP1 gene. Cancer Res 68: 5009-5013.
91. Rodig SJ, Cheng J, Wardzala J, DoRosario A, Scanlon JJ, et al. (2012) Improved detection suggests all Merkel cell carcinomas harbor Merkel polyomavirus. J Clin Invest 122: 4645-4653.
92. Laude HC, Jonchere B, Maubec E, Carlotti A, Marinho E, et al. (2010) Distinct merkel cell polyomavirus molecular features in tumour and non tumour specimens from patients with merkel cell carcinoma. PLoS Pathog 6: e1001076.
93. Martel-Jantin C, Filippone C, Cassar O, Peter M, Tomasic G, et al. (2012) Genetic variability and integration of Merkel cell polyomavirus in Merkel cell carcinoma. Virology 426: 134-142.
94. Sastre-Garau X, Peter M, Avril MF, Laude H, Couturier J, et al. (2009) Merkel cell carcinoma of the skin: pathological and molecular evidence for a causative role of MCV in oncogenesis. J Pathol 218: 48-56.
95. Fischer N, Brandner J, Fuchs F, Moll I, Grundhoff A (2010) Detection of Merkel cell polyomavirus (MCPyV) in Merkel cell carcinoma cell lines: cell morphology and growth phenotype do not reflect presence of the virus. Int J Cancer 126: 2133-2142.
96. Borchert S, Czech-Sioli M, Neumann F, Schmidt C, Wimmer P, et al. (2013) 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.
97. Houben R, Shuda M, Weinkam R, Schrama D, Feng H, et al. (2010) Merkel cell polyomavirus-infected Merkel cell carcinoma cells require expression of viral T antigens. J Virol 84: 7064-7072.
98. Shuda M, Kwun HJ, Feng H, Chang Y, Moore PS (2011) Human Merkel cell polyomavirus small T antigen is an oncoprotein targeting the 4E-BP1 translation regulator. J Clin Invest 121: 3623-3634.
99. Rockville Merkel Cell Carcinoma G (2009) Merkel cell carcinoma: recent progress and current priorities on etiology, pathogenesis, and clinical management. J Clin Oncol 27: 4021-4026.
100. Popp S, Waltering S, Herbst C, Moll I, Boukamp P (2002) UV-B-type mutations and chromosomal imbalances indicate common pathways for the development of Merkel and skin squamous cell carcinomas. Int J Cancer 99: 352-360.
101. Guastafierro A, Feng H, Thant M, Kirkwood JM, Chang Y, et al. (2013) Characterization of an early passage Merkel cell polyomavirus-positive Merkel cell carcinoma cell line, MS-1, and its growth in NOD scid gamma mice. J Virol Methods 187: 6-14.
102. Verhaegen ME, Mangelberger D, Harms PW, Vozheiko TD, Weick JW, et al. (2015) Merkel cell polyomavirus small T antigen is oncogenic in transgenic mice. J Invest Dermatol 135: 1415-1424.
103. Spurgeon ME, Cheng J, Bronson RT, Lambert PF, DeCaprio JA (2015) Tumorigenic activity of merkel cell polyomavirus T antigens expressed in the stratified epithelium of mice. Cancer Res 75: 1068-1079.

104. Houben R, Adam C, Baeurle A, Hesbacher S, Grimm J, et al. (2012) 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.
105. Jansen B, Heere-Ress E, Schlagbauer-Wadl H, Halaschek-Wiener J, Waltering S, et al. (1999) Farnesylthiosalicylic acid inhibits the growth of human Merkel cell carcinoma in SCID mice. J Mol Med (Berl) 77: 792-797.
106. Schlagbauer-Wadl H, Klosner G, Heere-Ress E, Waltering S, Moll I, et al. (2000) Bcl-2 antisense oligonucleotides (G3139) inhibit Merkel cell carcinoma growth in SCID mice. J Invest Dermatol 114: 725-730.
107. Saxonov S, Berg P, Brutlag DL (2006) A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. Proc Natl Acad Sci U S A 103: 1412-1417.
108. Mercier A, Arias C, Madrid AS, Holdorf MM, Ganem D (2014) Site-specific association with host and viral chromatin by Kaposi's sarcoma-associated herpesvirus LANA and its reversal during lytic reactivation. J Virol 88: 6762-6777.
109. Lee RC, Feinbaum RL, Ambros V (1993) The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell 75: 843-854.
110. Schwab SR, Shugart JA, Horng T, Malarkannan S, Shastri N (2004) Unanticipated antigens: translation initiation at CUG with leucine. PLoS Biol 2: e366.
111. Kincaid RP, Sullivan CS (2012) Virus-encoded microRNAs: an overview and a look to the future. PLoS Pathog 8: e1003018.
112. Lagos-Quintana M, Rauhut R, Lendeckel W, Tuschl T (2001) Identification of novel genes coding for small expressed RNAs. Science 294: 853-858.
113. Lagos-Quintana M, Rauhut R, Meyer J, Borkhardt A, Tuschl T (2003) New microRNAs from mouse and human. RNA 9: 175-179.
114. Bartel DP (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. Cell 116: 281-297.
115. Lee Y, Jeon K, Lee JT, Kim S, Kim VN (2002) MicroRNA maturation: stepwise processing and subcellular localization. EMBO J 21: 4663-4670.
116. Basyuk E, Suavet F, Doglio A, Bordonne R, Bertrand E (2003) Human let-7 stem-loop precursors harbor features of RNase III cleavage products. Nucleic Acids Res 31: 6593-6597.
117. Lee Y, Ahn C, Han J, Choi H, Kim J, et al. (2003) The nuclear RNase III Drosha initiates microRNA processing. Nature 425: 415-419.
118. Yi R, Qin Y, Macara IG, Cullen BR (2003) Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. Genes Dev 17: 3011-3016.
119. Hutvagner G, McLachlan J, Pasquinelli AE, Balint E, Tuschl T, et al. (2001) A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. Science 293: 834-838.
120. Lau NC, Lim LP, Weinstein EG, Bartel DP (2001) An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. Science 294: 858-862.
121. Grimson A, Farh KK, Johnston WK, Garrett-Engele P, Lim LP, et al. (2007) MicroRNA targeting specificity in mammals: determinants beyond seed pairing. Mol Cell 27: 91-105.
122. Zeng Y, Yi R, Cullen BR (2003) MicroRNAs and small interfering RNAs can inhibit mRNA expression by similar mechanisms. Proc Natl Acad Sci U S A 100: 9779-9784.
123. Bartel DP (2009) MicroRNAs: target recognition and regulatory functions. Cell 136: 215-233.
124. Lewis BP, Burge CB, Bartel DP (2005) Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell 120: 15-20.
125. Djuranovic S, Nahvi A, Green R (2012) miRNA-mediated gene silencing by translational repression followed by mRNA deadenylation and decay. Science 336: 237-240.
126. Park CY, Jeker LT, Carver-Moore K, Oh A, Liu HJ, et al. (2012) A resource for the conditional ablation of microRNAs in the mouse. Cell Rep 1: 385-391.

127. Friedman RC, Farh KK, Burge CB, Bartel DP (2009) Most mammalian mRNAs are conserved targets of microRNAs. Genome Res 19: 92-105.
128. Pfeffer S, Zavolan M, Grasser FA, Chien M, Russo JJ, et al. (2004) Identification of virus-encoded microRNAs. Science 304: 734-736.
129. Team RC (2012) R: A language and environment for statistical computing. . URL http://www.R-project.org/: R Foundation for Statistical Computing, Vienna, Austria
130. Grundhoff A, Sullivan CS (2011) Virus-encoded microRNAs. Virology 411: 325-343.
131. Wahle E, Keller W (1996) The biochemistry of polyadenylation. Trends Biochem Sci 21: 247-250.
132. Colgan DF, Manley JL (1997) Mechanism and regulation of mRNA polyadenylation. Genes Dev 11: 2755-2766.
133. Messeguer X, Escudero R, Farre D, Nunez O, Martinez J, et al. (2002) PROMO: detection of known transcription regulatory elements using species-tailored searches. Bioinformatics 18: 333-334.
134. Sandelin A, Alkema W, Engstrom P, Wasserman WW, Lenhard B (2004) JASPAR: an open-access database for eukaryotic transcription factor binding profiles. Nucleic Acids Res 32: D91-94.
135. Monteys AM, Spengler RM, Wan J, Tecedor L, Lennox KA, et al. (2010) Structure and activity of putative intronic miRNA promoters. RNA 16: 495-505.
136. Ozsolak F, Poling LL, Wang Z, Liu H, Liu XS, et al. (2008) Chromatin structure analyses identify miRNA promoters. Genes Dev 22: 3172-3183.
137. Whisnant AW, Kehl T, Bao Q, Materniak M, Kuzmak J, et al. (2014) Identification of novel, highly expressed retroviral microRNAs in cells infected by bovine foamy virus. J Virol 88: 4679-4686.
138. Kupfer DM, Drabenstot SD, Buchanan KL, Lai H, Zhu H, et al. (2004) Introns and splicing elements of five diverse fungi. Eukaryot Cell 3: 1088-1100.
139. Cai X, Li G, Laimins LA, Cullen BR (2006) Human papillomavirus genotype 31 does not express detectable microRNA levels during latent or productive virus replication. J Virol 80: 10890-10893.
140. Starck SR, Jiang V, Pavon-Eternod M, Prasad S, McCarthy B, et al. (2012) Leucine-tRNA initiates at CUG start codons for protein synthesis and presentation by MHC class I. Science 336: 1719-1723.
141. Brennecke J, Stark A, Russell RB, Cohen SM (2005) Principles of microRNA-target recognition. PLoS Biol 3: e85.
142. Starck SR, Ow Y, Jiang V, Tokuyama M, Rivera M, et al. (2008) A distinct translation initiation mechanism generates cryptic peptides for immune surveillance. PLoS One 3: e3460.
143. Haecker I, Renne R (2014) HITS-CLIP and PAR-CLIP advance viral miRNA targetome analysis. Crit Rev Eukaryot Gene Expr 24: 101-116.
144. Nachmani D, Stern-Ginossar N, Sarid R, Mandelboim O (2009) Diverse herpesvirus microRNAs target the stress-induced immune ligand MICB to escape recognition by natural killer cells. Cell Host Microbe 5: 376-385.
145. Bjellqvist B, Basse B, Olsen E, Celis JE (1994) Reference points for comparisons of two-dimensional maps of proteins from different human cell types defined in a pH scale where isoelectric points correlate with polypeptide compositions. Electrophoresis 15: 529-539.
146. Boussif O, Lezoualc'h F, Zanta MA, Mergny MD, Scherman D, et al. (1995) A versatile vector for gene and oligonucleotide transfer into cells in culture and in vivo: polyethylenimine. Proc Natl Acad Sci U S A 92: 7297-7301.
147. Seo GJ, Fink LH, O'Hara B, Atwood WJ, Sullivan CS (2008) Evolutionarily conserved function of a viral microRNA. J Virol 82: 9823-9828.
148. Lee S, Paulson KG, Murchison EP, Afanasiev OK, Alkan C, et al. (2011) Identification and validation of a novel mature microRNA encoded by the Merkel cell polyomavirus in human Merkel cell carcinomas. J Clin Virol 52: 272-275.

149. Chen CJ, Kincaid RP, Seo GJ, Bennett MD, Sullivan CS (2011) Insights into Polyomaviridae microRNA function derived from study of the bandicoot papillomatosis carcinomatosis viruses. J Virol 85: 4487-4500.
150. Chen CJ, Cox JE, Kincaid RP, Martinez A, Sullivan CS (2013) Divergent MicroRNA targetomes of closely related circulating strains of a polyomavirus. J Virol 87: 11135-11147.
151. Broekema NM, Imperiale MJ (2013) miRNA regulation of BK polyomavirus replication during early infection. Proc Natl Acad Sci U S A 110: 8200-8205.
152. Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26: 841-842.
153. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10: R25.
154. Lai CJ, Dhar R, Khoury G (1978) Mapping the spliced and unspliced late lytic SV40 RNAs. Cell 14: 971-982.
155. Treisman R, Kamen R (1981) Structure of polyoma virus late nuclear RNA. J Mol Biol 148: 273-301.
156. Huang da W, Sherman BT, Lempicki RA (2009) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res 37: 1-13.
157. Huang DW, Sherman BT, Tan Q, Kir J, Liu D, et al. (2007) DAVID Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists. Nucleic Acids Res 35: W169-175.
158. Liu Z, Batt DB, Carmichael GG (1994) Targeted nuclear antisense RNA mimics natural antisense-induced degradation of polyoma virus early RNA. Proc Natl Acad Sci U S A 91: 4258-4262.
159. Thorvaldsdottir H, Robinson JT, Mesirov JP (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform 14: 178-192.
160. Fults D, Pedone CA, Morse HG, Rose JW, McKay RD (1992) Establishment and characterization of a human primitive neuroectodermal tumor cell line from the cerebral hemisphere. J Neuropathol Exp Neurol 51: 272-280.
161. Liu Z, Carmichael GG (1994) Nuclear antisense RNA. An efficient new method to inhibit gene expression. Mol Biotechnol 2: 107-118.
162. Rehmsmeier M, Steffen P, Hochsmann M, Giegerich R (2004) Fast and effective prediction of microRNA/target duplexes. RNA 10: 1507-1517.
163. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, et al. (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol 14: R36.
164. Weber K, Bartsch U, Stocking C, Fehse B (2008) A multicolor panel of novel lentiviral "gene ontology" (LeGO) vectors for functional gene analysis. Mol Ther 16: 698-706.
165. Li J, Wang X, Diaz J, Tsang SH, Buck CB, et al. (2013) Merkel cell polyomavirus large T antigen disrupts host genomic integrity and inhibits cellular proliferation. J Virol 87: 9173-9188.
166. Scotto-Lavino E, Du G, Frohman MA (2006) 3' end cDNA amplification using classic RACE. Nat Protoc 1: 2742-2745.
167. Graham FL, Smiley J, Russell WC, Nairn R (1977) Characteristics of a human cell line transformed by DNA from human adenovirus type 5. J Gen Virol 36: 59-74.
168. DuBridge RB, Tang P, Hsia HC, Leong PM, Miller JH, et al. (1987) Analysis of mutation in human cells by using an Epstein-Barr virus shuttle system. Mol Cell Biol 7: 379-387.
169. Boyd D, Florent G, Kim P, Brattain M (1988) Determination of the levels of urokinase and its receptor in human colon carcinoma cell lines. Cancer Res 48: 3112-3116.
170. Rosen ST, Gould VE, Salwen HR, Herst CV, Le Beau MM, et al. (1987) Establishment and characterization of a neuroendocrine skin carcinoma cell line. Lab Invest 56: 302-312.

171. Mullis K, Faloona F, Scharf S, Saiki R, Horn G, et al. (1986) Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. Cold Spring Harb Symp Quant Biol 51 Pt 1: 263-273.
172. Rector A, Tachezy R, Van Ranst M (2004) A sequence-independent strategy for detection and cloning of circular DNA virus genomes by using multiply primed rolling-circle amplification. J Virol 78: 4993-4998.
173. Hirt B (1967) Selective extraction of polyoma DNA from infected mouse cell cultures. J Mol Biol 26: 365-369.
174. Chomczynski P, Sacchi N (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal Biochem 162: 156-159.
175. Varkonyi-Gasic E, Hellens RP (2011) Quantitative stem-loop RT-PCR for detection of microRNAs. Methods Mol Biol 744: 145-157.
176. Si H, Verma SC, Robertson ES (2006) Proteomic analysis of the Kaposi's sarcoma-associated herpesvirus terminal repeat element binding proteins. J Virol 80: 9017-9030.
177. Gunther T, Grundhoff A (2010) The epigenetic landscape of latent Kaposi sarcoma-associated herpesvirus genomes. PLoS Pathog 6: e1000935.
178. Kozomara A, Griffiths-Jones S (2011) miRBase: integrating microRNA annotation and deep-sequencing data. Nucleic Acids Res 39: D152-157.
179. Cantalupo P, Doering A, Sullivan CS, Pal A, Peden KW, et al. (2005) Complete nucleotide sequence of polyomavirus SA12. J Virol 79: 13094-13104.
180. Hafner M, Renwick N, Brown M, Mihailovic A, Holoch D, et al. (2011) RNA-ligase-dependent biases in miRNA representation in deep-sequenced small RNA cDNA libraries. RNA 17: 1697-1712.
181. Huang X, Yuan T, Tschannen M, Sun Z, Jacob H, et al. (2013) Characterization of human plasma-derived exosomal RNAs by deep sequencing. BMC Genomics 14: 319.
182. Jayaprakash AD, Jabado O, Brown BD, Sachidanandam R (2011) Identification and remediation of biases in the activity of RNA ligases in small-RNA deep sequencing. Nucleic Acids Res 39: e141.
183. Morin RD, O'Connor MD, Griffith M, Kuchenbauer F, Delaney A, et al. (2008) Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. Genome Res 18: 610-621.
184. Raabe CA, Tang TH, Brosius J, Rozhdestvensky TS (2013) Biases in small RNA deep sequencing data. Nucleic Acids Res.
185. Sorefan K, Pais H, Hall AE, Kozomara A, Griffiths-Jones S, et al. (2012) Reducing ligation bias of small RNAs in libraries for next generation sequencing. Silence 3: 4.
186. Raabe CA, Hoe CH, Randau G, Brosius J, Tang TH, et al. (2011) The rocks and shallows of deep RNA sequencing: Examples in the Vibrio cholerae RNome. RNA 17: 1357-1366.
187. Acheson NH (1984) Kinetics and efficiency of polyadenylation of late polyomavirus nuclear RNA: generation of oligomeric polyadenylated RNAs and their processing into mRNA. Mol Cell Biol 4: 722-729.
188. Batt DB, Carmichael GG (1995) Characterization of the polyomavirus late polyadenylation signal. Mol Cell Biol 15: 4783-4790.
189. Lindell TJ, Weinberg F, Morris PW, Roeder RG, Rutter WJ (1970) Specific inhibition of nuclear RNA polymerase II by alpha-amanitin. Science 170: 447-449.
190. Leendertz FH, Scuda N, Cameron KN, Kidega T, Zuberbuhler K, et al. (2011) African great apes are naturally infected with polyomaviruses closely related to Merkel cell polyomavirus. J Virol 85: 916-924.
191. Matsushita M, Nonaka D, Iwasaki T, Kuwamoto S, Murakami I, et al. (2014) A new in situ hybridization and immunohistochemistry with a novel antibody to detect small T-antigen expressions of Merkel cell polyomavirus (MCPyV). Diagn Pathol 9: 65.
192. Resnick J, Shenk T (1986) Simian virus 40 agnoprotein facilitates normal nuclear location of the major capsid polypeptide and cell-to-cell spread of virus. J Virol 60: 1098-1106.
193. Akan I, Sariyer IK, Biffi R, Palermo V, Woolridge S, et al. (2006) Human polyomavirus JCV late leader peptide region contains important regulatory elements. Virology 349: 66-78.

194. Harrison CJ, Meinke G, Kwun HJ, Rogalin H, Phelan PJ, et al. (2011) Asymmetric assembly of Merkel cell polyomavirus large T-antigen origin binding domains at the viral origin. J Mol Biol 409: 529-542.
195. Cheng J, Rozenblatt-Rosen O, Paulson KG, Nghiem P, DeCaprio JA (2013) Merkel cell polyomavirus large T antigen has growth-promoting and inhibitory activities. J Virol 87: 6118-6126.
196. van der Meijden E, Bialasiewicz S, Rockett RJ, Tozer SJ, Sloots TP, et al. (2013) Different serologic behavior of MCPyV, TSPyV, HPyV6, HPyV7 and HPyV9 polyomaviruses found on the skin. PLoS One 8: e81078.
197. Shahzad N, Shuda M, Gheit T, Kwun HJ, Cornet I, et al. (2013) The T antigen locus of Merkel cell polyomavirus downregulates human Toll-like receptor 9 expression. J Virol 87: 13009-13019.
198. Griffiths DA, Abdul-Sada H, Knight LM, Jackson BR, Richards K, et al. (2013) Merkel cell polyomavirus small T antigen targets the NEMO adaptor protein to disrupt inflammatory signaling. J Virol 87: 13853-13867.
199. Lewis BP, Shih IH, Jones-Rhoades MW, Bartel DP, Burge CB (2003) Prediction of mammalian microRNA targets. Cell 115: 787-798.
200. Warf MB, Johnson WE, Bass BL (2011) Improved annotation of C. elegans microRNAs by deep sequencing reveals structures associated with processing by Drosha and Dicer. RNA 17: 563-577.
201. Starega-Roslan J, Krol J, Koscianska E, Kozlowski P, Szlachcic WJ, et al. (2011) Structural basis of microRNA length variety. Nucleic Acids Res 39: 257-268.
202. Wu H, Neilson JR, Kumar P, Manocha M, Shankar P, et al. (2007) miRNA profiling of naive, effector and memory CD8 T cells. PLoS One 2: e1020.
203. Landgraf P, Rusu M, Sheridan R, Sewer A, Iovino N, et al. (2007) A mammalian microRNA expression atlas based on small RNA library sequencing. Cell 129: 1401-1414.
204. Han BW, Hung JH, Weng Z, Zamore PD, Ameres SL (2011) The 3'-to-5' exoribonuclease Nibbler shapes the 3' ends of microRNAs bound to Drosophila Argonaute1. Curr Biol 21: 1878-1887.
205. Doench JG, Sharp PA (2004) Specificity of microRNA target selection in translational repression. Genes Dev 18: 504-511.
206. Ford JP, Hsu MT (1978) Transcription pattern of in vivo-labeled late simian virus 40 RNA: equimolar transcription beyond the mRNA 3' terminus. J Virol 28: 795-801.
207. Acheson NH (1981) Efficiency of processing of viral RNA during the early and late phases of productive infection by polyoma virus. J Virol 37: 628-635.
208. Gottwein E, Corcoran DL, Mukherjee N, Skalsky RL, Hafner M, et al. (2011) Viral microRNA targetome of KSHV-infected primary effusion lymphoma cell lines. Cell Host Microbe 10: 515-526.
209. Lanoix J, Tseng RW, Acheson NH (1986) Duplication of functional polyadenylation signals in polyomavirus DNA does not alter efficiency of polyadenylation or transcription termination. J Virol 58: 733-742.
210. Ghosh PK, Lebowitz P, Frisque RJ, Gluzman Y (1981) Identification of a promoter component involved in positioning the 5' termini of simian virus 40 early mRNAs. Proc Natl Acad Sci U S A 78: 100-104.
211. Kamen R, Jat P, Treisman R, Favaloro J, Folk WR (1982) 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.
212. Buchman AR, Fromm M, Berg P (1984) Complex regulation of simian virus 40 early-region transcription from different overlapping promoters. Mol Cell Biol 4: 1900-1914.
213. Khalili K, Feigenbaum L, Khoury G (1987) Evidence for a shift in 5'-termini of early viral RNA during the lytic cycle of JC virus. Virology 158: 469-472.
214. Cowie A, Tyndall C, Kamen R (1981) Sequences at the capped 5'-ends of polyoma virus late region mRNAs: an example of extreme terminal heterogeneity. Nucleic Acids Res 9: 6305-6322.

215. Gidoni D, Kahana C, Canaani D, Groner Y (1981) Specific in vitro initiation of transcription of simian virus 40 early and late genes occurs at the various cap nucleotides including cytidine. Proc Natl Acad Sci U S A 78: 2174-2178.
216. Song Gao J, Zhang Y, Li M, Tucker LD, Machan JT, et al. (2010) Atypical transcription of microRNA gene fragments. Nucleic Acids Res 38: 2775-2787.
217. Liu X, Chen X, Yu X, Tao Y, Bode AM, et al. (2013) Regulation of microRNAs by epigenetics and their interplay involved in cancer. J Exp Clin Cancer Res 32: 96.
218. Sethupathy P (2013) Illuminating microRNA Transcription from the Epigenome. Curr Genomics 14: 68-77.
219. Trowbridge PW, Frisque RJ (1995) Identification of three new JC virus proteins generated by alternative splicing of the early viral mRNA. J Neurovirol 1: 195-206.
220. Deutsch M, Long M (1999) Intron-exon structures of eukaryotic model organisms. Nucleic Acids Res 27: 3219-3228.
221. Barrett NL, Li X, Carmichael GG (1995) The sequence and context of the 5' splice site govern the nuclear stability of polyoma virus late RNAs. Nucleic Acids Res 23: 4812-4817.
222. Ronemus M, Vaughn MW, Martienssen RA (2006) MicroRNA-targeted and small interfering RNA-mediated mRNA degradation is regulated by argonaute, dicer, and RNA-dependent RNA polymerase in Arabidopsis. Plant Cell 18: 1559-1574.
223. Schneider CL, Hudson AW (2011) The human herpesvirus-7 (HHV-7) U21 immunoevasin subverts NK-mediated cytoxicity through modulation of MICA and MICB. PLoS Pathog 7: e1002362.
224. Skalsky RL, Corcoran DL, Gottwein E, Frank CL, Kang D, et al. (2012) The viral and cellular microRNA targetome in lymphoblastoid cell lines. PLoS Pathog 8: e1002484.
225. Calder KB, Smoller BR (2010) New insights into merkel cell carcinoma. Adv Anat Pathol 17: 155-161.
226. de Vries RD, Lemon K, Ludlow M, McQuaid S, Yuksel S, et al. (2010) In vivo tropism of attenuated and pathogenic measles virus expressing green fluorescent protein in macaques. J Virol 84: 4714-4724.
227. Sharma P, Yan F, Doronina VA, Escuin-Ordinas H, Ryan MD, et al. (2012) 2A peptides provide distinct solutions to driving stop-carry on translational recoding. Nucleic Acids Res 40: 3143-3151.
228. Boldogh I, Albrecht T, Porter DD (1996) Persistent Viral Infections. In: Baron S, editor. Medical Microbiology. 4th ed. Galveston (TX).
229. Bentwich I (2005) Prediction and validation of microRNAs and their targets. FEBS Lett 579: 5904-5910.
230. Desai P, Person S (1998) Incorporation of the green fluorescent protein into the herpes simplex virus type 1 capsid. J Virol 72: 7563-7568.
231. Wightman B, Burglin TR, Gatto J, Arasu P, Ruvkun G (1991) Negative regulatory sequences in the lin-14 3'-untranslated region are necessary to generate a temporal switch during Caenorhabditis elegans development. Genes Dev 5: 1813-1824.
232. Krutzfeldt J, Rajewsky N, Braich R, Rajeev KG, Tuschl T, et al. (2005) Silencing of microRNAs in vivo with 'antagomirs'. Nature 438: 685-689.
233. Baek D, Villen J, Shin C, Camargo FD, Gygi SP, et al. (2008) The impact of microRNAs on protein output. Nature 455: 64-71.
234. Brodersen P, Voinnet O (2009) Revisiting the principles of microRNA target recognition and mode of action. Nat Rev Mol Cell Biol 10: 141-148.
235. Tsang SH, Wang X, Li J, Buck CB, You J (2014) Host DNA damage response factors localize to merkel cell polyomavirus DNA replication sites to support efficient viral DNA replication. J Virol 88: 3285-3297.
236. Tycowski KT, Guo YE, Lee N, Moss WN, Vallery TK, et al. (2015) Viral noncoding RNAs: more surprises. Genes Dev 29: 567-584.

7 Supplementary Material

Table S 1: Gene ontology analysis of LT-Ag ChIP peaks
Top 10 enriched annotation clusters after Gene ontology analysis using nearest TSS to LT-Ag ChIP-Seq peaks

Annotation Cluster 1	Enrichment Score: 7.36	Count	P_Value	Fold Change
GOTERM_CC_FAT	nucleoplasm	65	8.60E-10	2.3
GOTERM_CC_FAT	nuclear lumen	89	3.50E-09	1.9
GOTERM_CC_FAT	organelle lumen	102	2.10E-08	1.7
GOTERM_CC_FAT	intracellular organelle lumen	100	2.80E-08	1.7
GOTERM_CC_FAT	membrane-enclosed lumen	102	5.90E-08	1.7
GOTERM_CC_FAT	nucleoplasm part	37	6.40E-05	2.0
Annotation Cluster 2	**Enrichment Score: 3.1**	**Count**	**P_Value**	**Fold Change**
SP_PIR_KEYWORDS	kinase	44	4.20E-05	1.9
SP_PIR_KEYWORDS	atp-binding	71	5.00E-05	1.6
SP_PIR_KEYWORDS	nucleotide-binding	85	6.70E-05	1.5
GOTERM_BP_FAT	phosphorus metabolic process	58	1.40E-04	1.7
GOTERM_BP_FAT	phosphate metabolic process	58	1.40E-04	1.7
INTERPRO	Protein kinase, core	33	2.40E-04	2.0
GOTERM_MF_FAT	nucleotide binding	111	2.50E-04	1.4
SP_PIR_KEYWORDS	serine/threonine-protein kinase	27	3.70E-04	2.2
GOTERM_MF_FAT	ATP binding	78	4.50E-04	1.5
INTERPRO	Serine/threonine protein kinase, active site	26	5.50E-04	2.1
GOTERM_MF_FAT	adenyl ribonucleotide binding	78	6.60E-04	1.4
GOTERM_MF_FAT	ribonucleotide binding	92	6.80E-04	1.4
GOTERM_MF_FAT	purine ribonucleotide binding	92	6.80E-04	1.4
GOTERM_MF_FAT	adenyl nucleotide binding	81	7.70E-04	1.4
GOTERM_MF_FAT	purine nucleotide binding	95	7.90E-04	1.4
GOTERM_BP_FAT	protein amino acid phosphorylation	41	8.90E-04	1.7
GOTERM_MF_FAT	protein kinase activity	38	1.10E-03	1.7
GOTERM_MF_FAT	purine nucleoside binding	81	1.20E-03	1.4
GOTERM_MF_FAT	nucleoside binding	81	1.50E-03	1.4
INTERPRO	Serine/threonine protein kinase-related	25	1.50E-03	2.0
GOTERM_MF_FAT	protein serine/threonine kinase activity	29	1.80E-03	1.9
INTERPRO	Protein kinase, ATP binding site	29	2.20E-03	1.9
SP_PIR_KEYWORDS	transferase	66	2.50E-03	1.4
GOTERM_BP_FAT	phosphorylation	44	4.70E-03	1.5
INTERPRO	Serine/threonine protein kinase	15	6.40E-02	1.7
SMART	S_TKc	15	1.60E-01	1.4

Annotation Cluster 3	Enrichment Score: 2.75	Count	P_Value	Fold Change
GOTERM_BP_FAT	regulation of apoptosis	52	4.50E-05	1.8
GOTERM_BP_FAT	regulation of programmed cell death	52	5.70E-05	1.8
GOTERM_BP_FAT	regulation of cell death	52	6.20E-05	1.8
GOTERM_BP_FAT	negative regulation of apoptosis	26	9.40E-04	2.1
GOTERM_BP_FAT	negative regulation of programmed cell death	26	1.10E-03	2.0
GOTERM_BP_FAT	negative regulation of cell death	26	1.20E-03	2.0
GOTERM_BP_FAT	positive regulation of apoptosis	29	1.60E-03	1.9
GOTERM_BP_FAT	positive regulation of programmed cell death	29	1.80E-03	1.9
GOTERM_BP_FAT	positive regulation of cell death	29	2.00E-03	1.9
GOTERM_BP_FAT	induction of apoptosis	23	2.60E-03	2.0
GOTERM_BP_FAT	induction of programmed cell death	23	2.70E-03	2.0
SP_PIR_KEYWORDS	Apoptosis	23	7.60E-03	1.8
GOTERM_BP_FAT	cell death	39	9.90E-03	1.5
GOTERM_BP_FAT	death	39	1.10E-02	1.5
GOTERM_BP_FAT	programmed cell death	34	1.20E-02	1.6
GOTERM_BP_FAT	anti-apoptosis	15	1.60E-02	2.0
GOTERM_BP_FAT	apoptosis	33	1.60E-02	1.5
Annotation Cluster 4	**Enrichment Score: 2.72**	**Count**	**P_Value**	**Fold Change**
SP_PIR_KEYWORDS	nucleus	196	1.80E-07	1.4
SP_PIR_KEYWORDS	Transcription	104	9.60E-06	1.5
SP_PIR_KEYWORDS	transcription regulation	102	1.10E-05	1.5
GOTERM_BP_FAT	transcription	104	3.50E-04	1.4
GOTERM_BP_FAT	regulation of transcription	121	1.20E-03	1.3
GOTERM_MF_FAT	transcription regulator activity	69	2.90E-02	1.3
GOTERM_MF_FAT	DNA binding	100	3.70E-02	1.2
SP_PIR_KEYWORDS	dna-binding	75	4.90E-02	1.2
GOTERM_BP_FAT	regulation of RNA metabolic process	78	5.80E-02	1.2
GOTERM_BP_FAT	regulation of transcription, DNA-dependent	74	1.10E-01	1.2
GOTERM_MF_FAT	transcription factor activity	37	4.50E-01	1.1
Annotation Cluster 5	**Enrichment Score: 2.7**	**Count**	**P_Value**	**Fold Change**
GOTERM_BP_FAT	chromatin organization	27	1.10E-03	2.0
GOTERM_BP_FAT	histone modification	13	1.40E-03	3.0
GOTERM_BP_FAT	covalent chromatin modification	13	1.80E-03	2.9
GOTERM_BP_FAT	chromatin modification	21	2.00E-03	2.1
GOTERM_BP_FAT	chromosome organization	31	2.50E-03	1.8
SP_PIR_KEYWORDS	chromatin regulator	16	4.50E-03	2.3

Annotation Cluster 6	Enrichment Score: 2.5	Count	P_Value	Fold Change
SP_PIR_KEYWORDS	dna repair	17	5.70E-04	2.7
GOTERM_BP_FAT	cellular response to stress	36	1.10E-03	1.8
GOTERM_BP_FAT	DNA metabolic process	33	1.20E-03	1.8
SP_PIR_KEYWORDS	DNA damage	16	3.20E-03	2.4
GOTERM_BP_FAT	DNA repair	20	6.60E-03	2.0
GOTERM_BP_FAT	DNA recombination	10	1.30E-02	2.7
GOTERM_BP_FAT	response to DNA damage stimulus	23	1.50E-02	1.7
Annotation Cluster 7	**Enrichment Score: 2.5**	**Count**	**P_Value**	**Fold Change**
SP_PIR_KEYWORDS	ubl conjugation	33	3.60E-03	1.7
SP_PIR_KEYWORDS	isopeptide bond	20	9.20E-03	1.9
Annotation Cluster 8	**Enrichment Score: 2.42**	**Count**	**P_Value**	**Fold Change**
GOTERM_BP_FAT	cell motion	31	1.80E-03	1.8
GOTERM_BP_FAT	cell migration	21	2.20E-03	2.1
GOTERM_BP_FAT	localization of cell	21	7.20E-03	1.9
GOTERM_BP_FAT	cell motility	21	7.20E-03	1.9
Annotation Cluster 9	**Enrichment Score: 2.35**	**Count**	**P_Value**	**Fold Change**
GOTERM_CC_FAT	intracellular non-membrane-bounded organelle	110	1.90E-03	1.3
GOTERM_CC_FAT	non-membrane-bounded organelle	110	1.90E-03	1.3
GOTERM_CC_FAT	cytoskeleton	59	2.50E-02	1.3
Annotation Cluster 10	**Enrichment Score: 2.26**	**Count**	**P_Value**	**Fold Change**
GOTERM_BP_FAT	negative regulation of macromolecule metabolic process	49	3.50E-05	1.9
GOTERM_BP_FAT	negative regulation of nitrogen compound metabolic process	34	9.70E-04	1.8
GOTERM_BP_FAT	negative regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	33	1.50E-03	1.8
GOTERM_BP_FAT	negative regulation of cellular biosynthetic process	34	3.40E-03	1.7
GOTERM_BP_FAT	negative regulation of transcription	29	4.10E-03	1.8
GOTERM_BP_FAT	negative regulation of macromolecule biosynthetic process	33	4.30E-03	1.7
GOTERM_BP_FAT	negative regulation of gene expression	31	4.30E-03	1.7
GOTERM_BP_FAT	negative regulation of biosynthetic process	34	4.60E-03	1.7
GOTERM_MF_FAT	transcription repressor activity	21	1.00E-02	1.8
GOTERM_BP_FAT	negative regulation of RNA metabolic process	22	2.00E-02	1.7

GOTERM_BP_FAT	negative regulation of transcription, DNA-dependent	21	3.10E-02	1.6
GOTERM_BP_FAT	negative regulation of transcription from RNA polymerase II promoter	15	9.50E-02	1.6
GOTERM_BP_FAT	regulation of transcription from RNA polymerase II promoter	33	1.30E-01	1.3

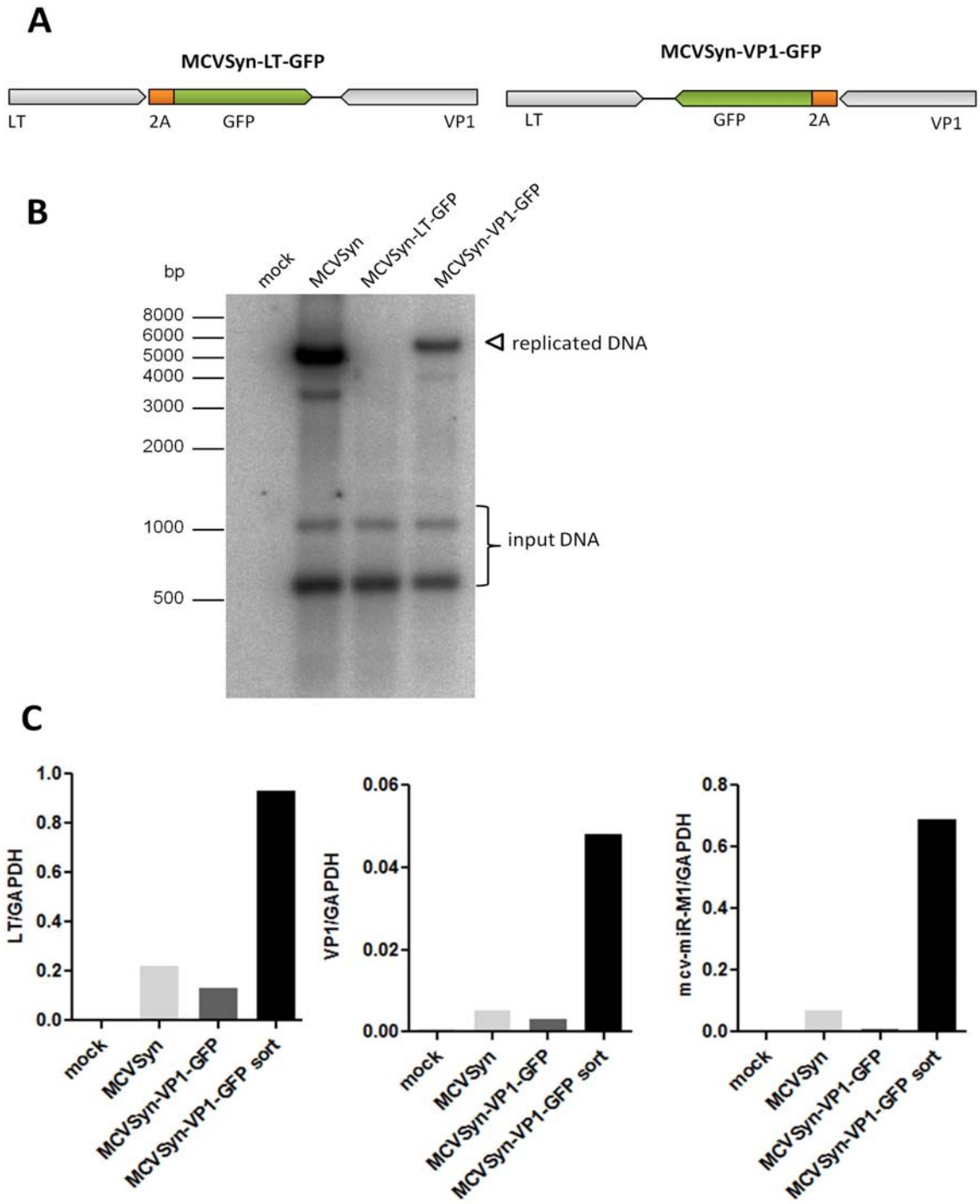

Figure S 1: Replication and gene expression of MCVSyn-GFP episomes
(A) Schematic depiction of MCVSyn-LT-GFP and MCVSyn-VP1-GFP. The 2A-GFP cassette was introduced downstream of the early or late coding region after removal of the respective stop-codon. As a consequence, the early or late regions are transcribed together with GFP in a polycistronic mRNA but during translation, the 2A sequence causes a ribosomal skip resulting in production of two separate proteins. **(B)** Southern blot analysis of PFSK-1 cells transfected with re-ligated MCVSyn or MCVSyn-GFP constructs 4 d.p.t. MCVSyn-VP1-GFP, but not MCVSyn-LT-GFP shows replication of viral DNA. **(C)** LT-mRNA, VP1-mRNA and mcv-miR-M1 levels relative to GAPDH in PFSK-1 cells transfected with re-ligated MCVSyn or MCVSyn-VP1-GFP before and after the sorting of GFP positive cells.

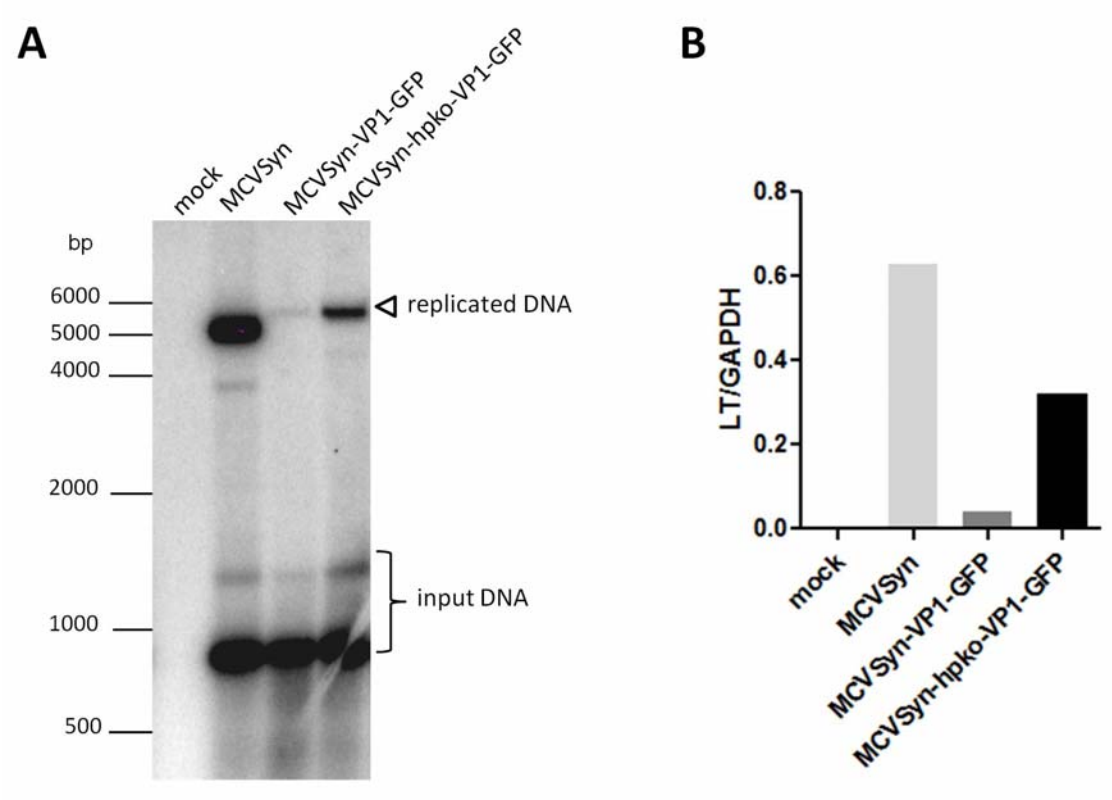

Figure S 2: MCVSyn-hpko-VP1-GFP and MCVSyn-seed-VP1-GFP show increased replication and LT-mRNA levels compared to MCVSyn-VP1-GFP
(A) Southern blot analysis of replication assays with the indicated MCVSyn variants in PFSK-1 cells 5 d.p.t. **(B)** qPCR analysis of LT-mRNA levels relative to GAPDH mRNA levels.

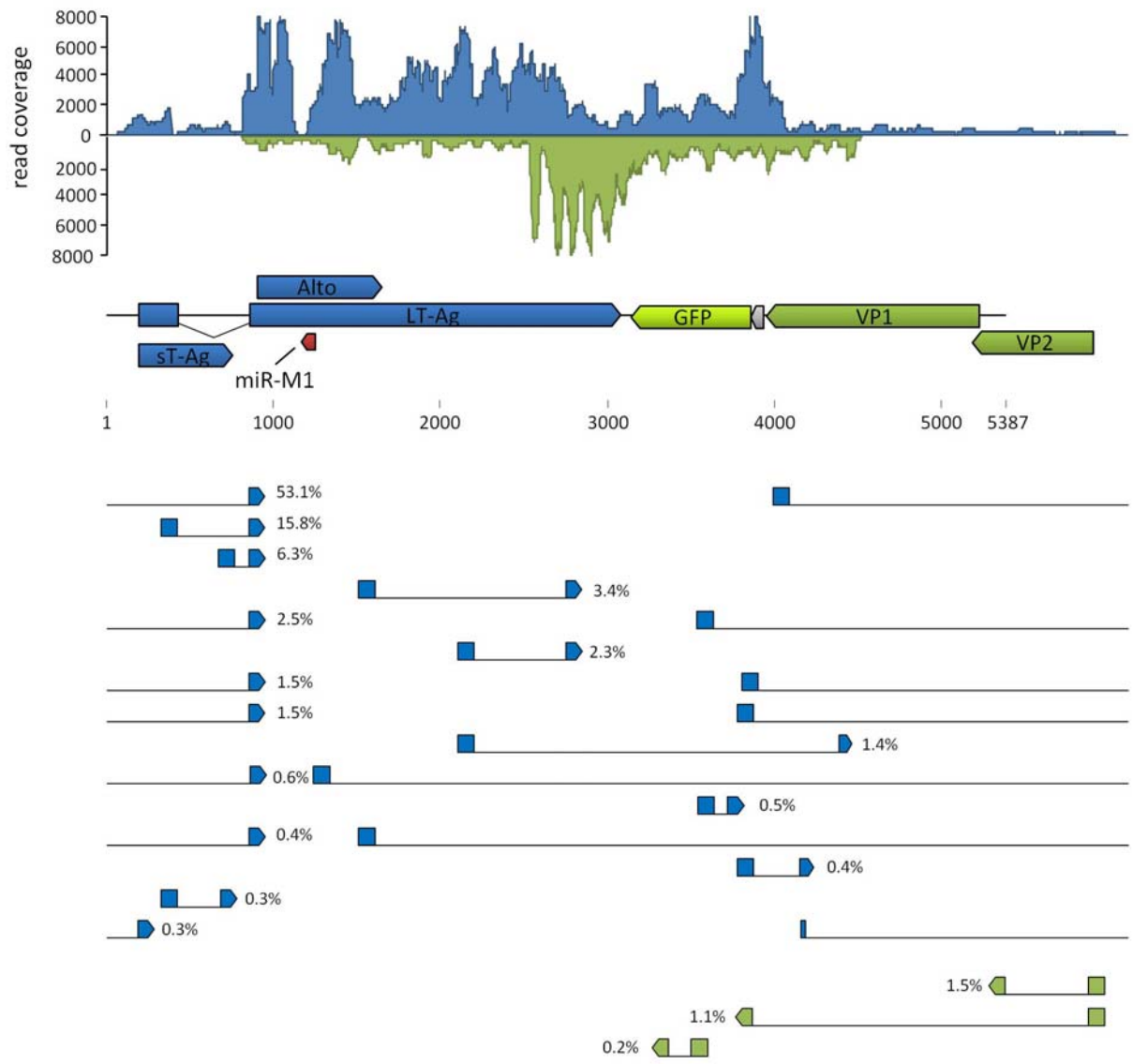

Figure S 3: Introduction of GFP in MCVSyn severely affects the viral splice pattern
RNA-Seq analysis of replicating MCVSyn-VP1-GFP in PFSK-1 cells. Upper panel: Read coverage on the MCVSyn-VP1-GFP genome in early (positive axis, blue) and late (negative axis, green) orientation. Lower panel: Splice sites detected by RNA-Seq. For each splice site the abundance relative to all splices is shown. Only splice sites with a relative abundance of more than 0.1% are depicted.

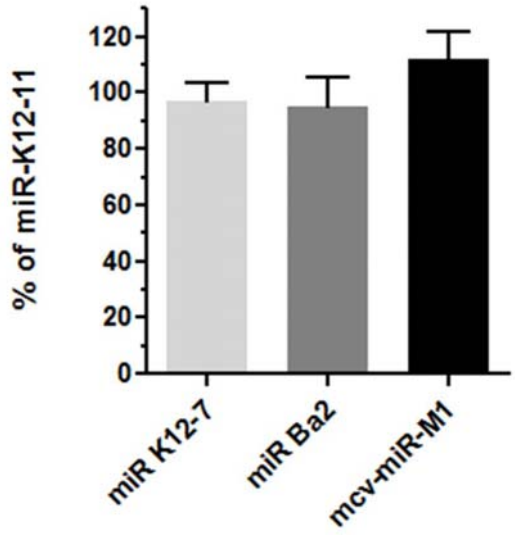

Figure S 4: MICB expression is not reduced upon overexpression of mcv-miR-M1
The ligand MICB is expressed upon cellular stress and presented on the cell surface. Therefore, MICB expression can be measured by a FACS assay using fluorescently labeled antibodies. RKO-cells, a colon carcinoma cell line which constantly expresses MICB, was used to investigate MICB protein levels upon expression of various viral miRNAs after lentiviral transduction. The KSHV encoded miRNA miR-K12-7 and the EBV encoded miRNA miR-Bart2 5p were used as positive control for downregulation of MICB expression. The KSHV miRNA miR-K12-11 was shown not to influence MICB expression levels and was used as background control.
The graph shows MICB protein levels as measured by FACS analysis using PE-labeled anti-MICB antibody. MICB levels are shown relative to MICB levels in RKO cells expressing miR-K12-11.

8 Indices

8.1 Figures

8.2 Tables

8.3 Abbreviations

°C	Degree Celsius
57kT-Ag	57k Tumor-Antigen
A	Adenine
aa	Amino acid
Ago	Argonaute protein
ALTO	Alternate frame of the Large T Open reading frame
approximately	Approximately
ATP	Adenosin tri-phosphate
BFV	Bovine Foamy Virus
BKPyV	BK polyomavirus
BLV	Bovine Leukemia Virus
bp	Base pair
BPCV	Bandicoot papillomatosis carcinomatosis virus
C	Cytosine
C. elegans	*Caenorhabditis elegans*
cDNA	Complemetary DNA
CDS	Coding sequence
ChIP	Chromatin immunoprecipitation
CMV	Cytomegalovirus
CT	Cycle threshold
CTL	Cytotoxic T lymphocytes
Cy3	Cytidine-3
Cy5	Cytidine-5
d	Day(s)
d.p.t.	Day(s) post transfection
DDR	DNA damage response
DEPC	Diethylpyrocarbonate
DMEM	Dulbecco's Modified Eagle Medium
DMSO	Dimethylsulfoxid
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonukleotid-Mix
ds	Double strand(ed)
DTT	Dithiothreitol
E. coli	*Escherichia coli*
EBV	Epstein-Barr Virus
EDTA	Ethylenediaminetetra acetic acid
et al.	Et alteri
FACS	Fluorescence activated cell sorting
FCS	Fetal calf serum
FISH	Fluorescence *in situ* hybridization
FL	Full length
FMDV	Foot-and-mouth disease virus
g	Gram
g	Gravity

G	Guanine
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
GFP	Green fluorescent protein
GggPyV	Gorilla polyomavirus
h	Hour(s)
H3	Histone 3
HCMV	Human Cytomegalovirus
HEK	Human embryonic kidney
HEPES	2-(4-(2-hydroxyethyl)-1-piperazinyl)-ethanesulfonic acid
HIV	Human Immunodeficiency Virus
hpko	Hairpin knockout
HPyV	Human polyomavirus
hsa	Homo Sapiens
HSV-1	Herpes Simplex Virus
HTS	High Throughput Sequencing
Ig	Immunoglobulin
IPTG	Isopropyl-ß-D-thio-galactopyranoside
JCPyV	JC polyomavirus
K4	Lysine 4
kb	Kilobases
kDa	Kilo Dalton
KIPyV	Karolinska Institute polyomavirus
KSHV	Kaposi Sarcoma associated Herpesvirus
L	Liter
LB	Lysogeny Broth
LPyV	African green monkey lymphotropic polyomavirus
LT-Ag	Large Tumor-Antigen
ltb	LT-Ag binding sites
M	Molar
MaPyV	Mastomys polyomavirus
MCC	Merkel cell carcinoma
MCPyV	Merkel cell polyomavirus
me3	Tri-methylation
MGB	Minor groove binder
MICB	MHC class I polypeptide-related sequence B
min	Minute(s)
miRNA	MicroRNA
MOI	Multiplicity of infection
MOPS	4-Morpholinepropanesulfonic acid
MuPyV	Murine polyomavirus
mRNA	Messenger RNA
MTT	3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide
MUR	Merkel cell polyomavirus unique region
MWPyV	Malawi polyomavirus
NCCR	Non-coding control region
NJPyV	New Jersey polyomavirus

NK-cell	Natural killer cell
NOD	Non-Obese Diabetic
nt	Nucleotide
OBD	Origin binding domain
OD	Optical density
oncomiR	Oncogenic miRNA
ORF	Open reading frame
p	Prime
pA	Polyadenylation
PARCLIP	Photoactivatable Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PE	Phycoerythrin
PEI	Polyethylenimine
Pen/Strep	Penicillin-Streptomycin
PFA	Paraformaldehyde
pI	Isoelectric point
PML	Progressive multifocal leukoencephalopathy
pol	Polymerase
polyA	Poly adenylation
pre-miRNA	Precursor miRNA
pri-miRNA	Primary miRNA
pro	Promoter mutant
PtvPyV	Chimpanzee polyomavirus
PVAN	Polyomavirus-associated nephropathy
PyV	Polyomavirus
q	Quantitative
RACE	Rapid amplification of cDNA ends
RacPyV	Raccoon polyomavirus
Rb	Retinoblastoma protein
RbCL2	Rubidium chloride
RCA	Rolling circle amplification
RISC	RNA induced silencing complex
RNA	Ribonucleic acid
RNA-pol	RNA-polymerase
RPKM	Reads per kilobase per million mapped reads
rpm	Rounds per minute
RPMI	Roswell Park Memorial Institute
RT	Room temperature/reverse transcription
s	Second(s)
SA12	Simian Agent 12
SAS-RNA	SV40-associated small RNA
SCID	Severe Combined Immunodeficiency
SDS	Sodium dodecyl sulfate
Seq	Sequencing

shRNA	Short hairpin RNA
siRNA	Small interfering RNA
SL RT-qPCR	Quantitative stem-loop RT-PCR
sT-Ag	Small Tumor-Antigen
STLPyV	St Louis polyomavirus
SV40	Simian Vacuolating Virus 40
T	Thymine
T-Ag	Tumor-Antigen
TM	Taqman
tRNA	Transfer RNA
trunc	Truncated
TSPyV	Trichodysplasia Spinulosa associated polyomavirus
TSS	Transcritional start site
U	Unit
UTR	Untranslated region
UV	Ultraviolet
V	Volt
v	Volume
VZV	Varizella Zoster Virus
wt	Wildtype
WUPyV	Washington University polyomavirus
XGal	ß-D-galactopyranosid

9 Publications and Oral Presentations

Publications:

Theiss JM, Günther T, Alawi M, Neumann F, Tessmer U, Fischer N, Grundhoff A (2015) 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. DOI: 10.1371/journal.ppat.1004974.

Günther T, Theiss JM, Fischer N, Grundhoff A (2015) Investigation of viral and host chromatin by ChIP-PCR or ChIP-Seq analysis. Current Protocols [accepted]

Oral Presentations:

Theiss JM, Günther T, Alawi M, Neumann F, Tessmer U, Fischer N, Grundhoff A
The Merkel Cell Polyomavirus encoded miRNA mcv-miR-M1 autoregulates LT-Ag expression and augments long-term persistence of viral episomes
DNA Tumor Virus Meeting 2015, Trieste (Italy)

Theiss JM, Günther T, Alawi M, Neumann F, Tessmer U, Fischer N, Grundhoff A
The Merkel Cell Polyomavirus encoded miRNA mcv-miR-M1 plays an essential role in the establishment and maintenance of viral persistence
Jahrestagung der Gesellschaft für Virologie e. V. (GfV) 2015, Bochum

Theiss JM, Indenbirken D, Neumann F, Alawi M, Tessmer U, Fischer N, Grundhoff A
The Merkel Cell Polyomavirus encoded miRNA mcv-miR-M1 contributes to long term persistence of viral episomes
DNA Tumor Virus Meeting 2014, Madison (USA)

Theiss JM, Indenbirken D, Neumann F, Alawi M, Tessmer U, Fischer N, Grundhoff A
A genetic system to study replication, transcription and persistence of Merkel Cell Polyomavirus
Jahrestagung der Gesellschaft für Virologie e. V. (GfV) 2013, Kiel

10 Acknowledgements

At the end of my thesis I wish to acknowledge all those people who made this thesis possible and an unforgettable experience for me.

First and foremost I would like to express my deepest gratitude to my supervisor Nicole Fischer for giving me the opportunity to conduct this doctoral thesis. Thank you for your confidence, support and guidance throughout this time, and for the great effort you put into training me in the scientific field.

I also wish to express my sincere thanks to Adam Grundhoff for supervising my doctoral thesis and for giving me the possibility to perform this work in his lab. Thank you for encouraging me to think outside the box and for inspiring ideas and discussions, which greatly advanced the development of this thesis.

I also would like to thank Thomas Dobner for the supervision of this thesis.

The current and former lab members of the Fischer and Grundhoff lab have significantly contributed to the success of this work by fruitful discussions, practical advice, mutual encouragement and by sharing all the ups and downs of everyday lab life. Thomas Günther, Uwe Tessmer, Daniela Indenbirken, Daniel Pohlmann, Kerstin Reumann, Marion Ziegler, Svenja Siebels, Malik Alawi, Michael Spohn, Friederike Neumann, Manja Czech-Sioli, Claudia Schmidt, Jan Knop, Emma Kraus, Christine Dahlke, Nicole Walz, Sophie Borchert – it was a great pleasure to work with you!

I would like to gratefully acknowledge Daniela Indenbirken, Kerstin Reumann and Malik Alawi for their relentless and patient support in producing and analyzing vast amounts of High Throughput Sequencing datasets; Friederike Neumann for spending long hours at the microscope searching for the needle in the haystack; Arne Düsedau for his patient help with sometimes unsatisfactory FACS experiments; Thomas Günther for sharing his knowledge and his contagious enthusiasm with me; Christine Dahlke and Sophie Borchert for their encouragement and support in difficult times; Uwe Tessmer, who did not live to see the end of this work he contributed so much to - you will not be forgotten.

I would like to express my deepest gratitude to Chris Sullivan for giving me the opportunity to join his lab for an unforgettable research stay. Thank you for your confidence, your enthusiasm and encouraging guidance. I also thank the Sullivan lab members for their support, for sharing their expertise with me and for all the fun in the lab.

I would like to acknowlegde Christine Dahlke, Daniela Indenbirken and Svenja Siebels for reading and correcting parts of my thesis.

I am thankful to Carol Stocking for grammatical editing of this dissertation.

This work was supported by a PhD scholarship of the Studienstiftung des deutschen Volkes, providing me the opportunity to conduct my doctoral thesis in a stimulating and interdisciplinary environment, which I am most grateful for.

I am very much indebted to my parents who unconditionally supported me through all phases my life. Thank you for your trust and encouragement and for letting me always pursuit my ambitions.

I am lucky to have a fantastic family and a bunch of wonderful friends who supported me in all possible ways during the last years and who kept me in touch with the world outside the lab.

Thank you, Sönke, for going all these winding roads with me – and beyond. You are wonderfully crazy!

11 Bestätigung der Korrektheit der englischen Sprache

HPI Heinrich-Pette-Institut
Leibniz-Institut für Experimentelle Virologie

HPI · Martinistraße 52 · 20251 Hamburg

Carol Stocking, Ph.D.
Retroviral Pathogenesis, Head
Phone: +49-40-48051 273
Fax: +49-40-48051 187
Email: stocking@hpi.uni-hamburg.de

Studienbüro Biologie
z.H. Frau Sült-Wüpping
MIN Fakultät
Universität Hamburg
Biozentrum Klein Flottbek
Ohnhorststraße 18
22609 Hamburg

29. Juli 2015

Sehr geehrte Damen und Herren,

hiermit bestätige ich, dass die von Frau Juliane Theiss mit dem Titel "Characterization of the Merkel cell polyomavirus (MCPyV) encoded miRNA: Functions in viral life cycle and identification of cellular target mRNAs" vorgelegte Doktorarbeit in korrektem Englisch geschrieben ist.

Mit freundlichen Grüßen,

Carol Stocking

Dr. Carol Stocking
Leiterin der FG Retrovirale Pathogenese
Heinrich-Pette-Institut
(Amerikanerin)

Heinrich-Pette-Institut
Leibniz-Institut für
Experimentelle Virologie

Martinistrasse 52 · 20251 Hamburg
Telefon +49 (0) 40 480 51-0
Telefax +49 (0) 40 48051-103
hpi@hpi.uni-hamburg.de

Bankverbindung
Haspa (200 505 50)
Konto 1001 315 959
www.hpi-hamburg.de

Mitglied der
Leibniz-Gemeinschaft

Heinrich-Pette-Institut
Leibniz-Institut für
Experimentelle Virologie

Martinistraße 52 · 20251 Hamburg
Telefon +49 (0) 40 480 51-0
Telefax +49 (0) 40 480 51-103
hpi@hpi.uni-hamburg.de

Bankverbindung
Haspa (200 505 50)
Konto 1001 315 959
www.hpi-hamburg.de

Mitglied der
Leibniz Gemeinschaft

12 Eidesstattliche Versicherung

Hiermit erkläre ich an Eides statt, dass ich die vorliegende Dissertationsschrift selbst verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe.

Hamburg, den 11.08.2015

Juliane Theiß