Analysis of transcriptional and epigenetic changes induced by Merkel cell polyomavirus tumor antigens in human dermal fibroblasts

Dissertation

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

vorgelegt von
Denise Ohnezeit

Hamburg, März 2022

Gutachter:

1. Prof. Dr. Nicole Fischer
2. Prof. Dr. Stefan Hoth

Datum der Disputation: 12.08.2022

Summary

Merkel cell polyomavirus (MCPyV) is the only human polyomavirus associated with tumorigenesis in its primary host. Initial infection with MCPyV is followed by a mostly lifelong asymptomatic persistent infection that usually does not lead to disease. However, in rare cases such as immunosuppression, the MCPyV genome can be integrated into the host genome, which leads to the formation of a Merkel cell carcinoma (MCC). MCC is a rare but aggressive skin cancer that is characterized by a high metastatic potential and low survival rates. Unfortunately, the use of recently introduced immune checkpoint inhibitors has improved the outcome for only half of all patients. In addition, recurrence is a frequent complication and underscores the importance to develop new treatment options.

Although MCPyV has been shown to be the underlying cause of over 80-90% of all MCCs, there are a few cases of MCCs that do not carry the virus. These tumors are specifically characterized by a high mutational burden that is 100-fold higher compared to MCPyV-positive MCCs. This strongly suggests that in MCPyV-positive MCCs, cellular signaling pathways are modulated independently of genetic aberrations. In fact, the two viral oncoproteins which are expressed in the initial phase of infection and persistence establishment, large Tumor (LT) and small Tumor (sT) antigen have both been shown to interfere with tumor suppressor signaling pathways. While sT expression is maintained in MCCs, the LT antigen is mutated prior to viral genome integration, which results in the expression of a truncated LT (LT*tr*).

The goal of my PhD project was to characterize the influence of MCPyV T antigens on the host genome by applying state-of-the-art transcriptomics and epigenomics to expand our current understanding of MCPyV persistence and pathogenesis. Because there is no *in vitro* infection or tumorigenesis model system available, in this work a lentiviral transduction system was used to overexpress MCPyV T antigens in primary human dermal fibroblasts. Transcriptional and epigenetic profiling gave new insights into the interaction of the T antigens with host cellular gene regulatory processes. First, it was identified that both sT and LT increased the transcription of inflammatory cytokines while LT additionally led to the short-term induction of interferon-stimulated genes (ISGs). ST expression led to a stable repression of ISG transcription which contributed to a diminished induction of ISGs when co-overexpressed with LT. This balancing effect of sT might play an important role during the initial phase of infection by impairing immune recognition and thereby maintaining a persistent infection. When sT was co-

overexpressed with LT*tr* to mimic a tumorigenic environment, the transcription of inflammatory cytokines and chemokines was increased, thereby reflecting the immunogenic feature of MCCs to attract innate and adaptive immune cells. Interestingly, sT expression was observed to result in a downregulation of several *human leukocyte antigen (HLA)* genes which are required for antigen presentation to CD8+ T cells. In addition, while LT*tr* also upregulated type I interferon (IFN) response genes, repression of ISGs by sT may play a role in MCC immune evasion.

This newly identified function was further investigated by analyzing the mechanisms used by sT to suppress the transcription of ISGs. The ISGF3 complex, which is a central transcription factor complex required for IFN signaling and ISG activation, was identified as a target of sT. One component of this complex is the interferon regulatory factor IRF9, which is also an ISG and was downregulated by sT. In addition to stable transcriptional repression, nuclear localization as well as protein levels of IRF9 were strongly reduced by sT. Thus, interference with ISGF3 was revealed in this work as one mechanism of sT to suppress the transcription of ISGs.

The analysis of histone modification changes revealed that none of the T antigens had a significant effect on repressive marks such as H3K27me3. Furthermore, H3K9me3 as a mark of constitutive heterochromatin was not affected by sT expression either. In contrast, transcriptional changes induced by the T antigens correlated with changes in H3K4me3 signal in promoter regions of differentially expressed genes. This applied also to H3K27ac, which was assessed in sT-expressing nHDFS as an additional mark of active promoters but also of enhancers. The reduction of H3K4me3 and H3K27ac specifically in the promoters of ISGs is possibly an effect of the transcriptional repression by sT and underscores the importance of sT in stably suppressing type I IFN response genes.

In summary, this work has revealed a function of sT in subverting the type I IFN response which has not been described so far. By repression of ISG transcription, sT was identified to play an important role in immune evasion as a counterbalance to LT or LT*tr* expression. Conclusively, this function possibly contributes to the establishment of a persistent infection but might also play an important role for tumor progression and survival. Both consequences significantly improve our understanding of MCPyV pathogenesis and might contribute to the development of new therapeutic strategies.

Zusammenfassung

Das Merkelzell-Polyomavirus (MCPyV) ist das einzige humane Polyomavirus, das mit der Tumorentstehung in seinem eigenen Wirt assoziiert ist. Auf die Erstinfektion mit MCPyV folgt eine lange Phase der asymptomatischen Persistenz. In seltenen Fällen, z.B. in immunsupprimierten Patienten, kann das Virus reaktiviert werden und versehentlich in das Wirtsgenom integrieren, was zur Entstehung eines Merkelzellkarzinoms führt (englisch: Merkel cell carcinoma, MCC). Das MCC ist eine aggressive Hauttumorerkrankung, die mit einem hohen Metastasierungspotenzial und niedrigen Überlebenschancen einhergeht. Da der Einsatz von kürzlich eingeführten Immun-Checkpoint-Inhibitoren die Prognose nur für die Hälfte aller Patienten verbessert hat, ist es wichtig, neue Behandlungsmöglichkeiten zu entwickeln.

Obwohl MCPyV nachweislich die Ursache für über 80-90% aller MCCs ist, gibt es einige wenige Fälle ohne bestätigte Virus-Assoziation. Diese Tumore weisen eine besonders hohe Mutationslast auf, die im Vergleich zu MCPyV-positiven MCCs 100-mal höher ist. Dies deutet stark darauf hin, dass bei MCPyV-positiven MCCs zelluläre Signalwege unabhängig von genetischen Aberrationen moduliert werden. Tatsächlich wurde herausgefunden, dass die beiden viralen Onkoproteine, das große Tumor (LT)-Antigen und das kleine Tumor (sT)-Antigen, mit Tumorsuppressor-Signalwegen interferieren. Während die sT-Expression in MCCs aufrechterhalten wird, mutiert das LT-Antigen vor der viralen Genomintegration, was zur Expression eines verkürzten (englisch: truncated) LT (LT*tr*) führt.

Das Ziel meines Promotionsprojekts war es, den Einfluss der T-Antigene auf das Wirtsgenom zu charakterisieren, wofür genomweite Transkriptom- und Epigenom-Analysen durchgeführt wurden. Da es kein geeignetes *in vitro* Infektions- oder Tumorgenese-Modellsystem gibt, wurde in dieser Arbeit ein lentivirales Transduktionssystem verwendet, um die T-Antigene in neonatalen humanen dermalen Fibroblasten (nHDFs) zu überexprimieren. Die Analysen führten zu neuen Erkenntnissen über die Interaktion der T-Antigene mit zellulären Genregulationsprozessen. Es wurde festgestellt, dass sowohl sT als auch LT die Transkription von inflammatorischen Zytokinen erhöhen, während LT zusätzlich zur kurzfristigen Induktion von Interferon-stimulierten Genen (ISGs) führte. Dagegen führte sT-Expression zu einer stabilen Repression der ISG-Transkription und verminderte diese, wenn es gemeinsam mit LT überexprimiert wurde. Dieser ausgleichende Effekt von sT könnte in der Anfangsphase der Infektion eine wichtige Rolle spielen, indem die Immunerkennung beeinträchtigt wird und dadurch die Aufrechterhaltung

einer persistenten Infektion gewährleistet werden kann. Wurde sT zusammen mit LT*tr* überexprimiert, führte dies zu einer erhöhten Transkription von inflammatorischen Zytokinen und Chemokinen, was die immunogene Eigenschaft von MCCs widerspiegelt und zur Aktivierung von angeborenen und adaptiven Immunzellen führt. Dagegen führte die Überexpression von sT zu einer Herunterregulierung von *HLA*-Genen, die für die Antigenpräsentation gegenüber CD8+ T-Zellen erforderlich sind. Durch die beobachtete Hochregulierung von Typ I Interferon (IFN)-gesteuerten Genen durch LT*tr* könnte der reprimierende Effekt von sT auf ISGs eine Rolle in der Immunevasion von MCCs spielen.

Zusätzlich wurden die Mechanismen untersucht, die zur sT-vermittelten Repression der Transkription von ISGs führen könnten. Der ISGF3 Komplex, welcher ein zentraler Transkriptionsfaktorkomplex im IFN-Signalweg ist, wurde als ein Ziel von sT identifiziert. Eine Komponente dieses Komplexes ist der IRF9 (englisch: Interferon regulatory factor), der selbst ein ISG ist, und durch sT auf transkriptioneller Ebene herunterreguliert wurde. Zusätzlich konnte gezeigt werden, dass durch sT die Zellkernlokalisierung reduziert wurde und daher auch verringerte Proteinlevel detektiert werden konnten.

Die Analyse der Histonmodifikationen ergab, dass es keine signifikante Auswirkung auf H3K27me3 gab, welches für fakultative Repression steht. Ebenfalls wurde H3K9me3 als Determinante von konstitutivem Heterochromatin nicht durch die sT-Expression beeinflusst. Im Gegensatz dazu konnten transkriptionelle Veränderungen mit Änderungen des H3K4me3-Signals in Promoterregionen assoziiert werden. Dies galt auch für H3K27ac, das in sT-exprimierenden Zellen als zusätzliche Markierung von aktiven Promotoren, aber auch von sogenannten Enhancern, untersucht wurde. Die Reduktion von H3K4me3 und H3K27ac speziell in den Promotorregionen von ISGs ist möglicherweise ein Effekt der transkriptionellen Repression durch sT.

Zusammenfassend konnte mit dieser Arbeit eine neue Funktion von sT bei der Unterdrückung der Typ-I-IFN-Antwort aufgedeckt werden, die bisher noch nicht beschrieben wurde. Durch die Herunterregulierung von ISGs spielt sT nachweislich eine wichtige Rolle bei der Immunevasion als Gegengewicht zur Expression von LT oder LT*tr*, und könnte somit erheblich zur Etablierung einer persistenten Infektion beitragen. Gleichzeitig könnte es aber auch eine wichtige Rolle im Tumor spielen. Neue Therapiemöglichkeiten sollten diese Erkenntnise hinsichtlich der angeborenen Immunantwort einbeziehen.

Contents

1 Introduction

1.1 Polyomaviruses

Polyomaviruses (PyV) are small (40-45 nm), non-enveloped DNA viruses that enclose their double-stranded (ds)DNA genome with a size of around 5 kbp in an icosahedral capsid. The family name *Polyomaviridae* is derived from the greek terms "poly" (= many) and "oma" (= tumor) which reflects on their large host variability, and the potential to induce tumors (Greenlee & Hirsch, 2016). This tumorigenic characteristic was first described by Gross (1953) who discovered the Murine polyomavirus (MPyV) and revealed an association with tumor formation in mice. The second *Polyomaviridae* member was discovered in 1960 by Sweet and Hilleman (1960) who identified the Simian virus 40 (SV40) as a contaminating agent in Poliovirus vaccine stocks, originating from rhesus monkey kidney cells. Importantly, research on both SV40 and MPyV has contributed substantially to the understanding of basic molecular processes such as transcription, DNA replication, and transformation of eukaryotic cells (Howley & Livingston, 2009).

1.1.1 Human polyomaviruses and associated diseases

The first PyVs that were shown to be infectious to humans were the JC polyomavirus (JCPyV) and the BK polyomavirus (BKPyV). These viruses were recovered from different patients concurrently, and were both named after the patients from which they were isolated (J.C. and B.K.). JCPyV was isolated from a patient suffering from progressive multifocal leukoencephalopathy (PML) by Padgett et al. (1971) and BKPyV was first recovered from a renal transplant patient by Gardner et al. (1971). Since then, BKPyV has been detected in several patients suffering from hemorrhagic cystitis or nephropathy (Arthur et al., 1986). The detection of further PyV species went slow at first, but rapidly accelerated with the advent of molecular and sequencing-based methods. According to the International Committee on Taxonomy of Viruses (ICTV), there are 112 PyV species known to date, that can be assigned to six different genera (Moens et al., 2017, last updated in February 2020), and infect a variety of different species, from birds to fish, to rodents, and other mammals (Greenlee & Hirsch, 2016). Out of the 14 human PyVs that have been identified until today, 13 of them are depicted in the phylogenetic tree in figure 1, sorted according to their large tumor antigen amino acid sequence similarities. In addition, their closest relative animal species are shown in black.

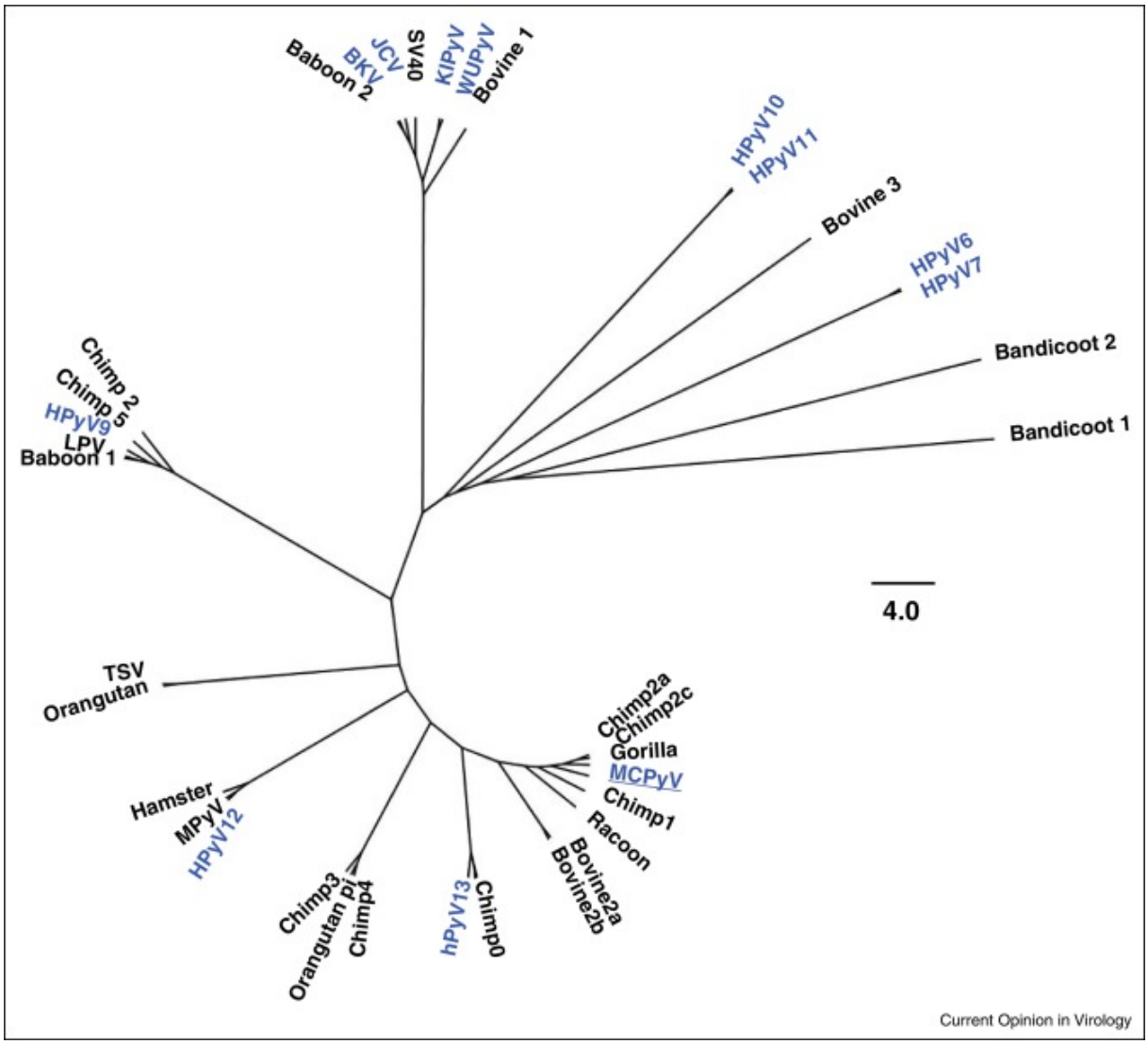


Figure 1 Phylogenetic relationship of human polyomaviruses and their closest animal relatives.
Phylogenetic tree showing 13 human PyVs known until 2015 (blue) and their closest relatives (black). This tree is based on the amino acid sequences of the large tumor antigens (Grundhoff & Fischer, 2015).

Most of the human PyVs are not unequivocally associated with disease but some of them can cause severe disease in the immunosuppressed. In addition to BKPyV and JCPyV, the Trichodysplasia spinulosa polyomavirus (TSPyV) and the Merkel cell polyomavirus (MCPyV) have been shown to be clinically relevant in immunocompromised patients. The TSPyV was recovered from a patient suffering from trichodysplasia spinulosa, a rare skin disease (van der Meijden et al., 2010). MCPyV was discovered in 2008 by digital transcriptome subtraction in samples from patients with Merkel cell carcinoma (MCC), a rare but aggressive tumor of the dermis (Feng et al., 2008). Interestingly, MCPyV is the only human PyV unequivocally shown to cause tumors in its primary host. Among the animal PyVs, this characteristic can only be

attributed to the Racoon polyomaviurs (RacPyV), which causes extremely rare brain tumors in racoons (Dela Cruz et al., 2013). Importantly, although RacPyV shares a high genomic similiarity with MCPyV (figure 1), a major difference relies on the episomal presence of the viral genomes in contrast to MCPyV, which is clonally integrated in MCCs (Feng et al., 2008). Recently, the Human polyomavirus 7 (HPyV7) has been linked to pruritic hyperproliferative keratinopathy in immunocompromised patients, including transplant recipients (Ho et al., 2015; Smith et al., 2018), underlining that ongoing research might reveal further associations of human PyVs with clinical manifestations.

A common feature of human pathogenic PyVs is that they persist benignly in the healthy population but can cause severe disease in immunocompromised patients. The seroprevalence of human PyVs (excluding HPyV 10, 11, 12 and 13) in adults ranges from 25-92%, based on the detection of antibodies against viral capsid proteins. Interestingly, serotyping of BKPyV and TSPyV in children and adults suggests that infection occurs during early childhood and persists for a life-time. Due to the constantly increasing number of patients with autoimmune diseases such as multiple sclerosis or rheumatoid arthritis, as well as transplant recipients, there is also an increase in PyV-associated complications (DeCaprio & Garcea, 2013).

1.1.2 Genome structure of polyomaviruses

Polyomaviruses have a circular dsDNA genome with sizes around 5 kbp which can be divided into three main regions: a non-coding control region (NCCR), an early gene region and a late gene region (figure 2). The NCCR contains viral promoters and the origin of replication which is necessary for the initiation of viral genome replication. In JCPyV and BKPyV, this NCCR can undergo rearrangements which is linked to a gain in pathogenesis and viral replication (DeCaprio & Garcea, 2013; Gosert et al., 2010; Gosert et al., 2008), an effect that has not been identified so far for MCPyV or TSPyV (Moens et al., 2020). The early gene region region of PyVs encodes for the tumor (T) antigens. All PyVs express a small (sT) and a large T (LT) antigen which play important roles in genome replication, deregulation of cellular signaling pathways and pathogenesis. A detailed description of the functions of MCPyV sT and LT can be found in chapter 1.2.4. Some PyVs express additional proteins, for example, the MPyV middle T antigen plays an important role in cellular transformation, described by Courtneidge and Smith (1983). A distantly related homologue is expressed by MCPyV and was named the Alternate frame of

the large T open reading frame (ALTO) (Carter et al., 2013). Several other splicing products have been identified for SV40, BKPyV, JCPyV, MCPyV and TSPyV, however their functions are not completely understood (van der Meijden Els et al., 2015). Several PyVs also express a micro RNA (miRNA) which is located in anti-sense direction of the early region. It is expressed for example by MPyV, RacPyV, SV40, MCPyV, BKPyV and JCPyV and was shown to decrease levels of viral T antigens which is considered an autoregulatory mechanism for a balanced early gene expression (Sullivan et al., 2005; Zou & Imperiale, 2021).

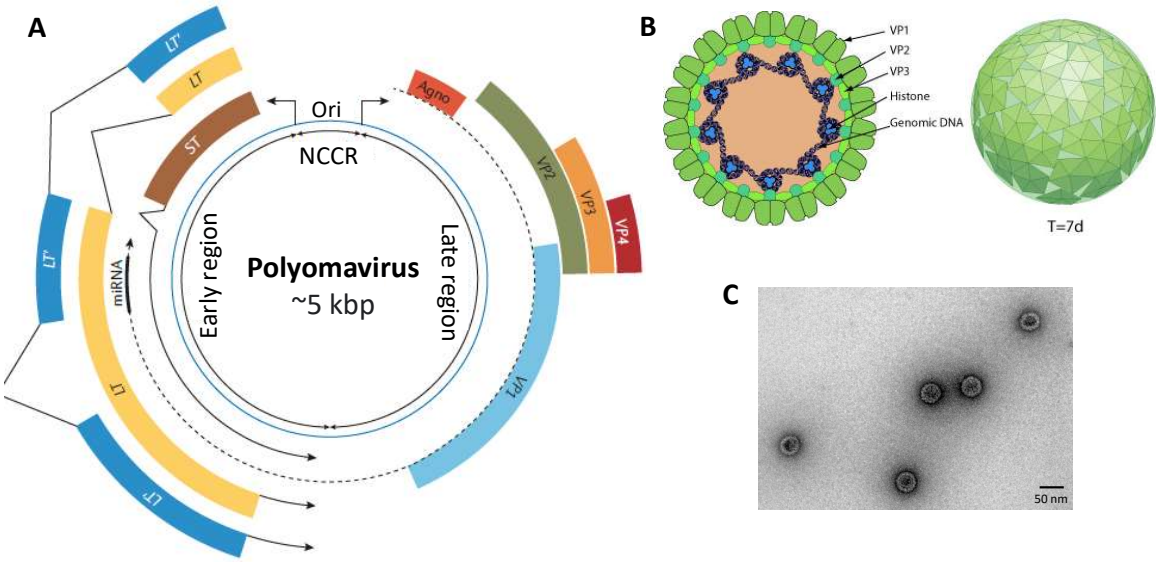


Figure 2 Genomic structure of polyomaviruses and viral particles.
(A) Circular map showing the general composition of PyV genomes (modified from DeCaprio & Garcea, 2013). ori: Origin of replication, NCCR: Non-coding control region, sT: Small T antigen, LT: Large T antigen, LT': Alternatively spliced LT antigens, VP1-4: Viral proteins 1-4, miRNA: MicroRNA, Agno: Agnoprotein. (B) Schematic structure of a human PyV virion. T: Triangulation number (from ViralZone, Swiss Institute of Bioinformatics). (C) Transmission electron microscopic image of recombinant virus like particles (VLPs) from MCPyV, kindly provided by C. Schneider (HPI, Hamburg).

The late region harbors genes important for the assembly of the viral capsids: the viral proteins (VP) 1-4. Figure 2B shows the pentameric formation of the major capsid protein VP1 in the outer layer, resulting in an icosahedral capsid with a T=7 symmetry. Interestingly, it was shown by Salunke et al. (1986) that the VP1 protein is sufficient for the assembly of the capsid. However, PyVs express several minor capsid proteins which are located on the inner side and interact with the viral genome which is wrapped around cellular histones (Wu et al., 2021). Most PyVs express VP2 and VP3, however MCPyV only produces the VP2 protein (Schowalter & Buck, 2013). An additional VP4 is solely expressed by SV40 and was shown to be important for viral egress via membrane pore formation (Daniels et al., 2007; Raghava et al., 2011). In addition to the structural

proteins, some PyVs express a further gene product of the late region - the agnoprotein. It was identified and described first for SV40 (Jay et al., 1981) as a protein expressed late in the lytic cycle and showing high affinity to DNA. While it is present in several other animal PyV species, its expression by human PyVs has only been reported for BKPyV and JCPyV. Its functions are not completely understood but it is required for viral propagation and was shown to interact with other viral and cellular proteins (Gerits & Moens, 2012; Saribas et al., 2016).

1.1.3 Life cycle of polyomaviruses

The life cycle of PyVs is best studied for SV40 and the knowledge can be generally conferred to other PyVs whose life cycles are still incompletely understood due to a lack of permissive cell culture models, e.g. for MCPyV. Nevertheless, many studies have revealed specific receptor types used by different PyVs through tracking viral entry and trafficking by electron microscopic analysis.

The first step for all PyVs is the attachment and entry into the host cell which is strongly dependent on the receptor types the host cell expresses on its surface. It has been observed that PyVs can enter via four different ways: (1) JCPyV by α2,6-glycan lactoseries tetrasaccharide c (LSTc) and the serotonin 5-hydroxytryptamine (5-HT) receptors of subfamily 2 for clathrin-mediated endocytosis (Assetta et al., 2013; Elphick et al., 2004; Liu et al., 1998), (2) caveolin-mediated endocytosis for SV40, BKPyV, MPyV and MCPyV, (3) lipid raft-mediated endocytosis observed for SV40, BKPyV, MPyV, and also for MCPyV (Becker et al., 2019), and (4) by extracellular vesicles (EVs). Both 2 and 3 routes require the attachment to glycosaminoglycans (GAGs) or gangliosides such as GD1b or GT1b. For MCPyV, it was shown that GAGs are used as initial attachment receptors but for efficient entry, sialylated glycans are required (Schowalter et al., 2011). Route 4 was a recently suggested alternative entry mechanism observed for JCPyV and BKPyV, but it is possible that other PyVs might traffic via EVs (Mayberry & Maginnis, 2020). Figure 3 depicts a schematic model of the life cycle of BKPyV and JCPyV with the newly suggested implication of EVs, as reviewed by Helle et al. (2020).

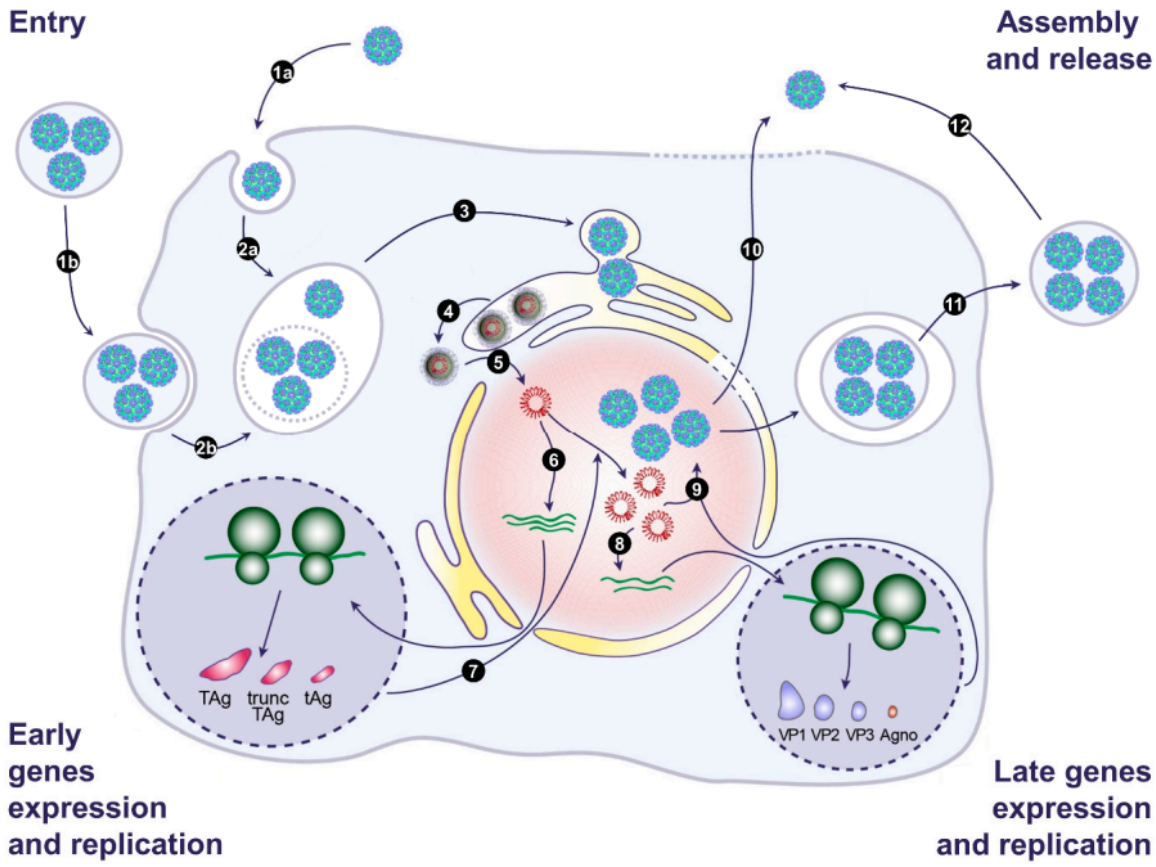


Figure 3 Schematic model of the life cycle of BKPyV and JCPyV.
Schematic model of the life cycle of BKPyV and JCPyV (from Helle et al., 2020). The following steps are denoted: (1a) Entry of the virions occurs via endocytosis (clathrin-mediated for JCPyV or lipid-raft-mediated for BKPyV). Alternatively, virions are enclosed in EVs (1b). (2a+b) Both are transported via endosomes to the endoplasmic reticulum (ER) (3). Viral capsids are partly uncoated from the ER into the cytosol (4) and subsequently, the viral genome enters the nucleus (5) where early genes are transcribed (6). Following translation, the early proteins are translocated into the nucleus (7) and initiate DNA replication and late gene expression (8). After the late proteins are translocated into the nucleus (9), new virions assemble and are released either via cell lysis (10) or via extracellular vesicles (EVs) (11) from which particles can spontaneously be released (12).

After entry, PyVs are taken up into endosomes and are transported to the endoplasmic reticulum (ER) from which they are partly uncoated into the cytosol. Several studies have suggested that PyVs exploit the ER degradation machinery to reach the cytosol and nucleus. However, the exact mechanisms of viral uncoating and nuclear entry are not completely understood (Dupzyk & Tsai, 2016). For SV40 as well as for BKPyV it was shown that the nuclear localization signal which is present in VP2 and VP3, plays an important role for nuclear entry (Bennett et al., 2015; Nakanishi et al., 2007).

Once reaching the nucleus, cellular polymerase II recognizes the viral promoters on the NCCR and transcribes the early genes. Splicing of the different T antigens is also conducted by the

cellular splicing machinery. After translation in the cytosol into the early proteins, they re-locate into the nucleus (Greenlee & Hirsch, 2016). The expression of the early proteins, including LT and sT, has substantial effects on host cell cycle and signaling processes which is discussed for MCPyV in more detail in section 1.2.3 and 1.2.4. Importantly, the LT antigen binds the viral origin of replication (ori), which initiates replication. This has been extensively studied for SV40 and has contributed to the general understanding of DNA replication (Sowd & Fanning, 2012). Replication of the viral genome results in the additional expression of the late proteins. Replicated viral genomes are equipped with cellular histones and the capsid proteins are assembled to them, resulting in new viral particles (Greenlee & Hirsch, 2016).

In order to complete the life cycle and infect new cells, the viral particles must be released from the host cells. One possibility is to lyse the cells to release the progeny virus which has been observed for SV40, BKPyV, and JCPyV. Interestingly, for SV40 it was reported that the structural protein VP4 is not incorporated into the viral capsid but is involved in the induction of cell lysis (Daniels et al., 2007) by forming pores for membrane disruption (Raghava et al., 2011). However, PyVs are also able to leave the host cell via non-lytic processes which are only starting to be unraveled. The first evidence of a non-lytic egress process was described for SV40 (Clayson et al., 1989). Similarly, it was demonstrated by Evans et al. (2015) that BKPyV can be released without cell lysis involving anion homeostasis. Recent studies provide evidence that both JCPyV and BKPyV use EVs to leave infected cells (Handala et al., 2020; Morris-Love et al., 2019). Whether other PyVs also use this mechanism for viral egress remains to be determined.

1.1.4 The type I interferon response and its role in human polyomavirus infection

The innate immune response represents the first line of defense against incoming pathogens. The interferon (IFN) molecule was first identified by Isaacs and Lindemann (1957) as a substance that interferes with the Influenza virus. It turned out that this cytokine is part of a complex web of signaling cascades which is initiated in response to the detection of pathogen-associated molecular patterns (PAMPs) by pattern-recognition receptors (PRRs). Toll-like receptors (TLRs) are one type of PRRs which are located either on the cell surface or in endosomes. There is a large variety of TLRs recognizing different viral PAMPs, including TLR9 which detects CpG-unmethylated dsDNA or TLR3 which senses dsRNA. Both of them are essential for the sensing of herpes viral endosomal DNA or replication-derived dsRNA, respectively (Paludan

et al., 2011). In addition to TLRs, there are several nucleic acid sensors that are located in the cytosol such as retinoic acid inducible gene I (RIG-I)-like receptors (RLRs) detecting foreign RNA, which transmit their signal via the MAVS adaptor protein. The main regulator of cytosolic DNA recognition is the stimulator of interferon genes (STING) which is activated by multiple DNA sensors such as cGAS, AIM2 or IFI16, with the latter being important for nuclear detection of foreign DNA (reviewed in Iwasaki, 2012; Schlee & Hartmann, 2016).

The engagement of PRRs results in the transcriptional activation of cytokines including IFNs, mediated by the nuclear factor kappa-light-chain-enhancer of activated B-cells (NFκB), as well as the interferon regulatory factors (IRFs) IRF3 and IRF7. While the NFκB pathway mostly results in the upregulation of pro-inflammatory cytokines, IRF3/7 activation leads to the production of type I IFNs, including IFNα and IFNβ (Iwasaki, 2012). IFNs can act in an autocrine or paracrine manner by binding to the IFNα receptor (IFNAR1/2) which consists of two subunits (Kim et al., 1997). Subsequently, janus kinase 1 (JAK1) and tyrosine kinase 2 (TYK2) are activated and initiate the phosphorylation of signal transducers and activators of transcription, STAT1 and STAT2. Together with IRF9, they form the interferon-stimulated gene factor 3 (ISGF3) complex which translocates to the nucleus. Via binding to interferon-responsive elements (ISREs) in promoter regions, ISGF3 activates the transcription of a large number of interferon-stimulated genes (ISGs), as reviewed in Schneider et al. (2014). It is estimated that there are over 450 genes that can be classified as ISGs, concerting a wide range of antiviral activity. PRRs, IRFs or STAT proteins are required for the initial activation of ISGs but are also ISGs themselves and thus reinforce the IFN response. In addition, there is a large variety of ISGs directly acting against multiple steps in the viral life cycle. IFITM1, IFITM2, or IFTIM3 for example, inhibit viral entry; MX1 and MX2 reduce nuclear import or mRNA synthesis; IFIT proteins act against viral protein translation; and OAS genes can directly degrade viral RNAs and thus intefere with genome replication (Schoggins, 2019).

Sensing of PyV PAMPs is largely unknown. However, a common feature of PyVs is their establishment of persistent infections in healthy individuals without causing any symptoms. Only in rare cases can they cause severe disease in immunosuppressed patients, suggesting an important role for the immune system in controlling PyV infection. Interestingly, recent studies about BKPyV and JCPyV revealed that the type I IFN response plays a decisive role for the infection outcome. Assetta et al. (2016) showed that infection of human renal epithelial cells with BKPyV results in cell lysis while JCPyV established a persistent infection in these

cells. Interestingly, both viruses induce the production of IFNβ, but only in JCPyV-infected cells a complete antiviral response, involving the production of ISGs, was observed. Similarly, it was reported by An et al. (2019) that BKPyV infection of human renal epithelial cells leads to complete cell lysis without inducing an IFN response. In contrast, when they infected vascular endothelial cells, BKPyV established a persistent infection and the cells responded with the production of ISGs.

These findings highlight that PyVs may directly target the IFN response for immune evasion. However, to date there are only a few studies that have addressed this. For example, a recent report showed that the JCPyV sT antigen antagonizes RIG-I-mediated signaling by preventing the activity of the E3 ubiquitin ligase TRIM25 (Chiang et al., 2021). They observed a similar role for BKPyV but its exact mechanism remains elusive. Still, it is unclear which PAMPs are presented during PyV infections but a study from Forero et al. (2014) indicates that the induction of the DNA damage response (DDR) by the SV40 LT antigen is linked to the upregulation of IFN-dependent ISG expression.

In conclusion, current research focuses on the question how PyVs are sensed and controlled by the innate immune system and which mechanisms PyVs have evolved to interfere with it. There is recent evidence that the type I IFN response plays an important role in PyV infection but further studies are needed for a better mechanistic understanding. The current knowledge about the immune response in MCPyV infection and tumorigenesis is discussed in chapter 1.2.5.

1.2 Merkel cell polyomavirus and Merkel cell carcinoma

MCPyV is unequivocally linked to MCC which is a skin tumor of neuroendocrine origin that was initially described by Toker (1972). The name originates from structural and immunohistochemical similarities of MCCs with Merkel cells which are located in the epidermis. However, due to several discrepancies it is largely accepted that Merkel cells do not represent the cell of origin of MCC. One of the most striking arguments is that Merkel cells are terminally differentiated while MCCs are described as heterogeneous and highly proliferative, which hints at a cell of origin that is not post-mitotic (Tilling & Moll, 2012).

MCC presents as red nodules (see figure 4A) which are fast-growing and have a high metastatic potential. Diagnosis of MCC relies on skin biopsies and immunostaining of cytokeratin 20 (CK20) which is uniquely expressed in >90% of MCCs compared to other neuroendocrine tumors (figure 4B+C). There is a large number of markers such as chromogranin-A or synaptophysin that can be additionally used for the differential diagnosis of MCC, as reviewed in Becker et al. (2017).

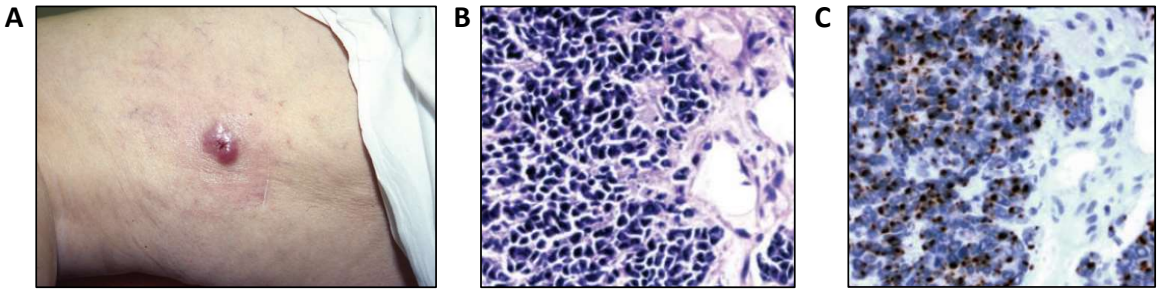


Figure 4 Merkel cell carcinoma.
(A) Clinical appearance of a Merkel cell carcinoma. (B) Hematoxylin and eosin staining from the dermis showing MCC clusters (20x magnification). (C) Immunostaining of an adjacent section with anti-CK20 antibody (20x magnification). (From Chang & Moore, 2012)

The discovery of the MCPyV in 2008 led to the observation that in 80-90% of MCCs, the viral genome is integrated into the host genome (Feng et al., 2008; Fischer et al., 2010; Shuda et al., 2009). The remaining 10-20% of MCCs however don't express any viral sequence and are referred to as virus-negative MCCs. While MPCyV-positive tumors are phenotypically identical to the virus-negative MCCs, they differ substantially on a genetic level. Intriguingly, virus-negative MCCs with 40 mutations/Mb were shown to have a 100-fold higher number of mutations compared to virus-positive MCCs with 0.4-0.75 mutations/Mb. In addition, they display a

strong UV mutational signature affecting typical tumor-associated genes such as *TP53*, *Rb1*, *Notch1* or *HRas* (Goh et al., 2016; Starrett et al., 2017). Similar tumor-associated pathways are also affected in the virus-positive tumors which is largely driven by the expression of the T antigens and discussed in more detail in chapter 1.2.4. While virus-negative tumors often appear in the epidermis which is exposed to UV light, MCPyV-positive MCCs are located in the dermal layer and are better protected from UV light, thus explaining the differences in the mutational burden. It was therefore suggested that virus-positive and virus-negative tumors might arise from different cells of origin (Sunshine et al., 2018). By screening several cell types that were challenged with MCPyV, Liu et al. (2016) showed that human dermal fibroblasts (HDFs) support infection. However, due to a lack of independent confirmations and *in vivo* model systems to study tumorigenesis, the cell(s) of origin of MCCs are still under debate.

1.2.1 Prevalence, epidemiology and clinical aspects

With an incidence of 0.7/100,000 person-years in the USA in 2013, MCC is a rare type of skin cancer which makes up less than 1% of all non-melanoma skin cancers. However, from 2000 to 2013 the incidence has increased over 95% and is estimated to further rise (Paulson et al., 2018; Samimi, 2019). MCC cases are especially high among the 85+ age group with an incidence of 9.8/100,000 person-years in 2013 (Paulson et al., 2018). The global distribution of MCC cases highlights that the majority of cases occurs in Northern America, Europe and Australia. The map in figure 5 depicts the global incidence rates including data from 2003-2007, representatively shown for the male population. In Australia, the incidence rates are among the highest worldwide, with 1.6/100,000 person-years (Youlden et al., 2014). Interestingly, it was observed that in Australia over 70% of MCCs are virus-negative which is most probably linked to the strong UV light exposure (Garneski et al., 2009). In contrast, in Europe and Northern America the majority of MCCs are causally linked to MPCyV.

Despite the low frequency of MCC cases, a high seroprevalence of MCPyV is common within the general population, ranging from 60-90% (Carter et al., 2009; Martel-Jantin et al., 2013; Tolstov et al., 2009). It is considered part of the normal skin microbial flora as complete MCPyV genomes can be recovered from the skin in 40% of healthy adults (Schowalter et al., 2010). In a serological study based on the VP1 protein from Martel-Jantin et al. (2013), it was shown that over 60-80% of children at the age of four to five were positive and by the age of 20, the percentage increased

to 94%. Interestingly, already from birth until the age of four months, seroprevalence was about 60-70% but then decreased to 0% at the age of 15-16 months. This suggests that children acquire an infection with MCPyV after the maternal antibodies have disappeared and it is believed that asymptomatic primary infection occurs via fecal-oral transmission (Loyo et al., 2010).

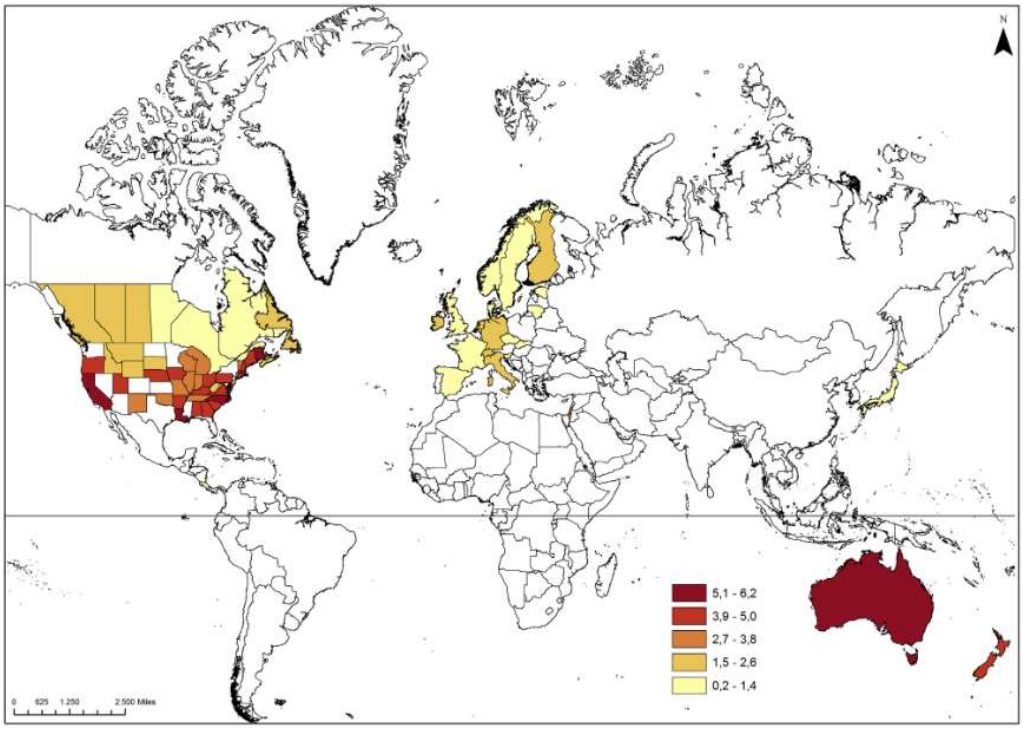


Figure 5 Worldwide distribution of Merkel cell carcinoma.
Worldwide distribution of Merkel cell carcinoma cases, representatively shown for the male population. The numbers refer to age-standardized incidence rates (cases per million person-years) between 2003 and 2007. (From Stang et al., 2018)

Although MCC is a rare disease, occurring mostly in the elder population or in immunosuppressed patients, it is a highly aggressive skin cancer with an overall mortality rate of 30%, the highest among all types of skin cancer (Becker et al., 2018; Eisemann et al., 2016; Samimi, 2019). One reason is its fast progression and metastatic potential with half of all cases having already metastasized at the time of the initial diagnosis (Grandhaye et al., 2015). While for local MCC, the five-year survival rate is around 51%, this number decreases to 35% when the tumor has metastasized and already reached the lymph nodes, dropping to 14% upon further dissemination (Harms et al., 2016). Thus, an early diagnosis is decisive for the survival chance of the patients but since MCCs can be misdiagnosed as benign lesions or other malignancies, the correct diagnosis is often delayed. Hence, the acronym "AEIOU" was introduced to facilitate

diagnosis by taking into consideration the following terms: Asymptomatic, expanding rapidly, immunosuppressed, >50 years of age and UV-exposed (Heath et al., 2008).

There are several treatment options available with a frequent combinatorial application. However, inefficient treatment and recurrence are challenging factors in the treatment of MCCs. The classical and standard treatment option is the direct excision of the primary tumor. It is often combined with radiotherapy and chemotherapy and in approximately 60% of all cases this combined treatment is efficient, although only 10% of the patients can be completely cured (Cowey et al., 2017; Nghiem et al., 2017; Tai et al., 2000). Furthermore, 30-40% of patients that received standard treatment face recurrence within two years (Allen et al., 2005; van Veenendaal et al., 2018). Similarly, patients with advanced disease that are treated with chemotherapy often become chemoresistant already after three months (Samimi, 2019).

With the advent of immune checkpoint inhibitors (CPIs), immunotherapy was introduced as a new approach and revolutionized the treatment of MCC. The cancer is highly immunogenic, displayed for example by the fact that CD8+ T cell infiltration of MCCs improves the outcome (Paulson et al., 2011) and that MCC immune evasion is frequently observed for example by downregulation of major histocompatibility complex-class I (MHC-I) receptors to prevent elimination by CD8+ T cells (Paulson et al., 2014). Furthermore, it was shown that in the context of MCC, T cells show an increased expression of inhibitory receptors such as programmed cell death 1 (PD-1) which normally functions as a checkpoint to inhibit T cell activation (Afanasiev et al., 2013). Moreover, its ligand PD-L1 was reported to be upregulated in MCCs, representing a mechanism to inhibit T cell activation (Lipson et al., 2013). Therefore, PD-L1 inhibitors were especially promising for the treatment of MCC. In fact, Avelumab which targets PD-L1, was approved in 2017 as the first immune CPI to treat advanced MCC in the USA and Europe and is currently considered the standard treatment of advanced MCC. It was shown for Avelumab and other CPIs that 22% of all patients can be completely cured and 29% show at least a partial response (Walker et al., 2018). Although CPIs improved the overall treatment, especially of advanced MCC, there are many patients that don't or only temporally respond to CPIs which highlights the need to develop better treatment options. So far, there are no prognostic markers available to identify if a patient is responsive to this treatment (Samimi, 2019).

MCPyV is the underlying cause for the majority of MCCs and as previously described, virus-positive MCCs differ substantially in their mutational landscape. However, their clinical presentation is nearly identical and immunotherapeutic intervention results in a similar responsiveness.

Interestingly, while virus-negative tumors are characterized by mutations in central tumor signaling genes such as *TP53*, *Rb1*, *Notch1* or *HRas*, MCPyV-positive tumors have nearly any mutations in these genes but similar pathways are targeted by the T antigens (Samimi, 2019). This highlights that tumorigenesis of virus-positive MCCs relies on the potential of T antigens to perturb cellular signaling pathways. With the aim to enable the development of new therapies, identifying the cellular targets of the T antigens and understanding their role in pathogenesis might be relevant for the treatment of both types of MCC.

1.2.2 Merkel cell polyomavirus genome structure

Belonging to the family of *Polyomaviridae*, MCPyV has a small dsDNA genome that can be separated into the NCCR, an early region and a late region. The NCCR contains promoters as well as the viral ori required for the initiation of viral replication. The early region encodes for four non-structural proteins: LT, sT, 57kT and ALTO. While the functions of 57kT and ALTO are not completely understood to date, the tumor antigens sT and LT are considered the main contributors to pathogenesis. The genome structure and ORFs of sT and LT are depicted in figure 6. Both sT and LT share an N-terminal DNA J domain which is followed in the case of LT by a MCPyV unique region (MUR) which flanks an LxCxE motif (Leu-X-Cys-X-Glu) which is required for binding of the retinoblastoma protein (pRb). Furthermore, the C-terminal region harbors the origin-binding domain (OBD), a zink finger and leucine zipper as well as a helicase domain which are required for viral genome replication. In MCC, the C-terminal region of LT is mutated, resulting in the expression of a truncated LT antigen (LT*tr*) which is replication-incompetent due to the loss of the OBD and helicase domain. While mutations can occur in different sites of the C-terminal region, the pRb-binding region is retained. The structure of sT is characterized by multiple domains, including an EP400-binding site, a LT-stabilizing domain (LSD), a PP4 binding site, as well as multiple PP2A binding sites. Their functions are further discussed in chapter 1.2.4

The late region harbors the structural capsid proteins VP1 and VP2 which are expressed after initiation of viral DNA replication (Grundhoff & Fischer, 2015; Harms et al., 2018; Shuda et al., 2008). Furthermore, MCPyV also encodes for miR-M1, a miRNA which is located in anti-sense direction to the early region and was shown to autoregulate early gene expression and limit viral replication (Seo et al., 2009; Theiss et al., 2015).

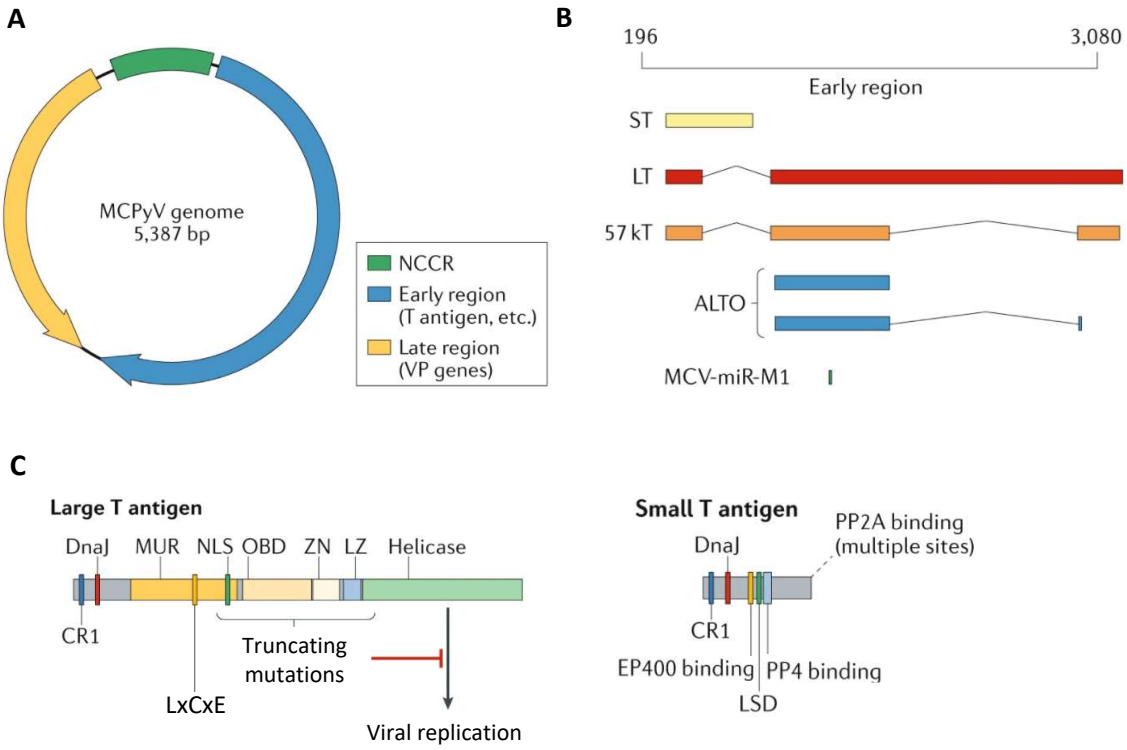


Figure 6 Genome structure of Merkel cell polyomavirus.
(A) Circular map of the MCPyV genome (modified from Harms et al., 2018). The non-coding control region (NCCR) is depicted in green, the early gene region in blue and the late gene region in yellow. VP: viral protein. (B) ORFs of multiple transcripts generated from the early gene region including small T (sT), large T (LT), 57kT and the alternative frame of the large T open reading frame (ALTO). Additionally, the microRNA MCV-miR-M1 is shown which is located in anti-sense direction to the early gene region. (C) Functional domains within LT are shown on the left side and within sT on the right side. LT and sT share the N-terminal DNA J domain with a conserved region (CR). In the case of LT, this is followed by a MCPyV unique regions (MUR) which flanks the LxCxE motif, followed by a nuclear localization signal (NLS). The origin-binding domain (OBD) with an adjacent zinc finger (ZN) and leucine zipper (LZ), as well as the helicase domain are required for viral replication. In MCC, a stop mutation leads to the elimination of the C-terminal domain, resulting in a truncated version of LT (LT*tr*). The mutation sites differ in each MCC, but the LxCxE motif and in some cases the NLS are retained. The DNA J domain in the sT antigen is followed by an EP400-binding site, the LT-stabilizing domain (LSD) and a PP4 binding site, as well as multiple PP2A binding sites.

1.2.3 Merkel cell polyomavirus life cycle and pathogenesis

The host cell type allowing permissive infection is still under debate, as well as the cell of origin of MCC. Due to the fact that MCPyV virions can be shed from the skin of healthy adults and virus-positive MCCs are located in the dermal skin layer (Schowalter et al., 2010), it is likely that the viral reservoir might be also located in the skin. Liu et al. (2016) identified that HDFs support MCPyV entry and productive replication, two factors which characterize a permissive cell. However, so far no other group has reported the usage of HDFs to study MCPyV infection

and it remains unclear if HDFs represent the reservoir for MCPyV due to the lack of confirmation experiments *in vivo*.

Prior to the infection model reported by Liu et al. (2016), a semi-permissive system had been established by Neumann et al. (2011) which was used to analyze MCPyV replication in different cell lines. It is based on transfection using a synthetic MCPyV genome and was shown to achieve replication, early and late transcription and particle formation in HEK293, H1299 and PFSK-1 cells out of 17 tested cell lines. However, re-transmission of viral particles was not observed, as confirmed also by Feng et al. (2011).

The life cycle of MCPyV can generally be divided into an early phase and a late phase. Entry of MCPyV into a cell requires the attachment to GAGs and to sialylated glycans as co-receptors (Schowalter et al., 2011). Most likely, MCPyV particles reach the nucleus via the endosomal pathway, similar to other PyVs. A recent study supports this hypothesis and showed that MCPyV particles were taken up into endosomes via caveolar/lipid raft-mediated endocytosis and transported to the ER (Becker et al., 2019). It remains to be revealed how and where uncoating takes place and how the genome gets into the nucleus.

Once MCPyV has reached the nucleus the early genes are transcribed from the viral DNA. This DNA is decorated with histones, resembling the eukaryotic genome structure enabling cellular polymerase II recognition (Greenlee & Hirsch, 2016). The LT protein is essential for the subsequent replication of viral DNA by binding to the viral ori. Similar to other PyV LT antigens, binding occurs via its OBD to conserved GRGGC motifs (Feng et al., 2011; Harrison et al., 2011; Kwun et al., 2013). Subsequently, the helicase function of LT enables unwinding of the DNA which makes replication factors accessible. Another important function of LT is the regulation of the cell cycle. Via its conserved LxCxE motif , the LT antigen can bind and inactivate pRb (Borchert et al., 2014; Shuda et al., 2008). Being a tumor suppressor protein, pRb functions as a master regulator of cell cycle processes such as G1 checkpoint-mediated blocking of S-phase entry and is inactivated in most cancers (Giacinti & Giordano, 2006). In addition to LT which is indispensable for efficient viral genome replication, auxiliary functions have been ascribed to the sT antigen. It was shown to stabilize the LT protein via its LSD which enables binding to the E3-ubiquitin ligase Fbw7, thereby preventing proteasomal degradation of LT (Kwun et al., 2013). Furthermore, sT was shown to enhance viral DNA replication by providing iron-sulfur (Fe/S) clusters important for DNA processing enzymes (Tsang et al., 2015). Further interactions of sT and LT with cellular factors are discussed in more detail in section 1.2.4.

Besides replication of the viral genome which is initiated following the expression of the early T antigens, the late proteins are transcribed which results in the production of the viral capsid proteins VP1 and VP2. In a permissive cell, the virus completes its life cycle and new virions are produced. For several PyVs it was shown that new virions can be released by cell lysis, as described for SV40, BKPyV or JCPyV. Non-lytic egress represents another form of viral release via the usage of EVs, as was shown in BKPyV and JCPyV, although not yet verified in MCPyV (Helle et al., 2020).

A summarizing model of the current knowledge about the life cycle of MCPyV and possible factors contributing to different scenarios is shown in figure 7. As the virus persists in 80% of the general population, there must be a cell type allowing a state of latency or persistence. While its site of persistence is not known to date, there is some evidence that MCPyV uses specific mechanisms required for viral latency. One example was described by Kwun et al. (2017) who showed that MCPyV uses a unique form of viral latency based on the regulation of LT expression by cellular ubiquitin ligases. Furthermore, Theiss et al. (2015) reported that expression of the MCPyV miR-M1 reduces the replication of viral episomes and contributes to persistence establishment.

It is possible that risk factors such as UV light or immunosuppression lead to a switch from the persistent state to actively replicating virus. Nevertheless, it is completely unknown if MCPyV has different sites of productive infection and persistence and which events lead to the accidental integration. As Kwun et al. (2017) suggested, MCPyV persistence might substantially differ from other chronic DNA viruses. However, integration of the viral genome is observed in nearly all virus-induced MCCs. Interestingly, the site of integration is different in every tumor but is consistent within a patient. It is therefore assumed that integration occurs prior to tumorigenesis (Feng et al., 2008; Rodig et al., 2012; Shuda et al., 2008). Recent studies have revealed mechanistic insights into the integration process. While integration is generally observed randomly in different regions of the human genome, Czech-Sioli et al. (2020a) showed that it preferentially occurs in euchromatic regions, i.e. transcriptionally active regions. Furthermore, two types of integration patterns were presented: (1) Non-homologous end joining (NHEJ) which involves the integration of linear viral genomes into the host genome at sites of DNA breaks, or (2) microhomology-mediated break-induced replication (MMBIR), characterized by an amplification of cellular DNA flanking at the viral integration site (Czech-Sioli et al., 2020a). The hypothesis that integration occurs at sites of dsDNA break resulting in an amplification of

virus-host DNA intermediates was also confirmed by Starrett et al. (2020). Both studies revealed that viral integration occurs after LT truncation because within a tumor the LT mutations are identical (Czech-Sioli et al., 2020a; Starrett et al., 2020).

Once the LT antigen is mutated and the viral genome is integrated into the host genome, the virus is not able to replicate due to the loss of the C-terminal OBD and helicase domain (Feng et al., 2008; Schmitt et al., 2012; Shuda et al., 2008). In addition, it was shown that the loss of the C terminus abrogates its inhibitory growth functions (Cheng et al., 2013; Li et al., 2013). Interestingly, in each tumor the mutations are located in different sites of the LT antigen. Further, the mutations never affect the LxCxE motif (see figure 6), reflecting the necessity for the virus to maintain the ability to inactivate pRb (Borchert et al., 2014; Cheng et al., 2013; Houben et al., 2012; Shuda et al., 2008). Furthermore, MCC cell lines were shown to depend on sT and LT*tr* expression, highlighting that both T antigens probably contribute to tumorigenesis (Houben et al., 2012; Houben et al., 2010; Shuda et al., 2014).

Tumorigenesis is associated with the loss of the ability to replicate and thus represents a "dead-end" for the virus while it is detrimental for the host. It is therefore important to understand the exact mechanisms that lead to tumorigenesis, maintain MCC and drive metastasis. Despite the lack of suitable *in vitro* and *in vivo* models to study the pathogenesis of MCPyV, transgenic mouse models such as xenograft models were utilized in the past. In combination with patient-derived MCC cell lines, several aspects of tumorigenesis and metastasis were addressed. For example, it was shown that the MCC cell line MS-1 induced the formation of tumors when injected into scid mice (Guastafierro et al., 2013). Verhaegen et al. (2015) observed a transforming activity of the sT antigen in several transgenic mouse models and they found that sT co-expression with the atonal bHLH transcription factor 1 (Atoh1) leads to MCC-like phenotypes (Verhaegen et al., 2017). Furthermore, when injecting MCC cell lines into scid mice, the formation of distant metastases was observed which correlated with increased sT expression (Knips et al., 2017). These studies highlight that the T antigens play an important role in MCPyV pathogenesis and are therefore described in more detail in the following chapter.

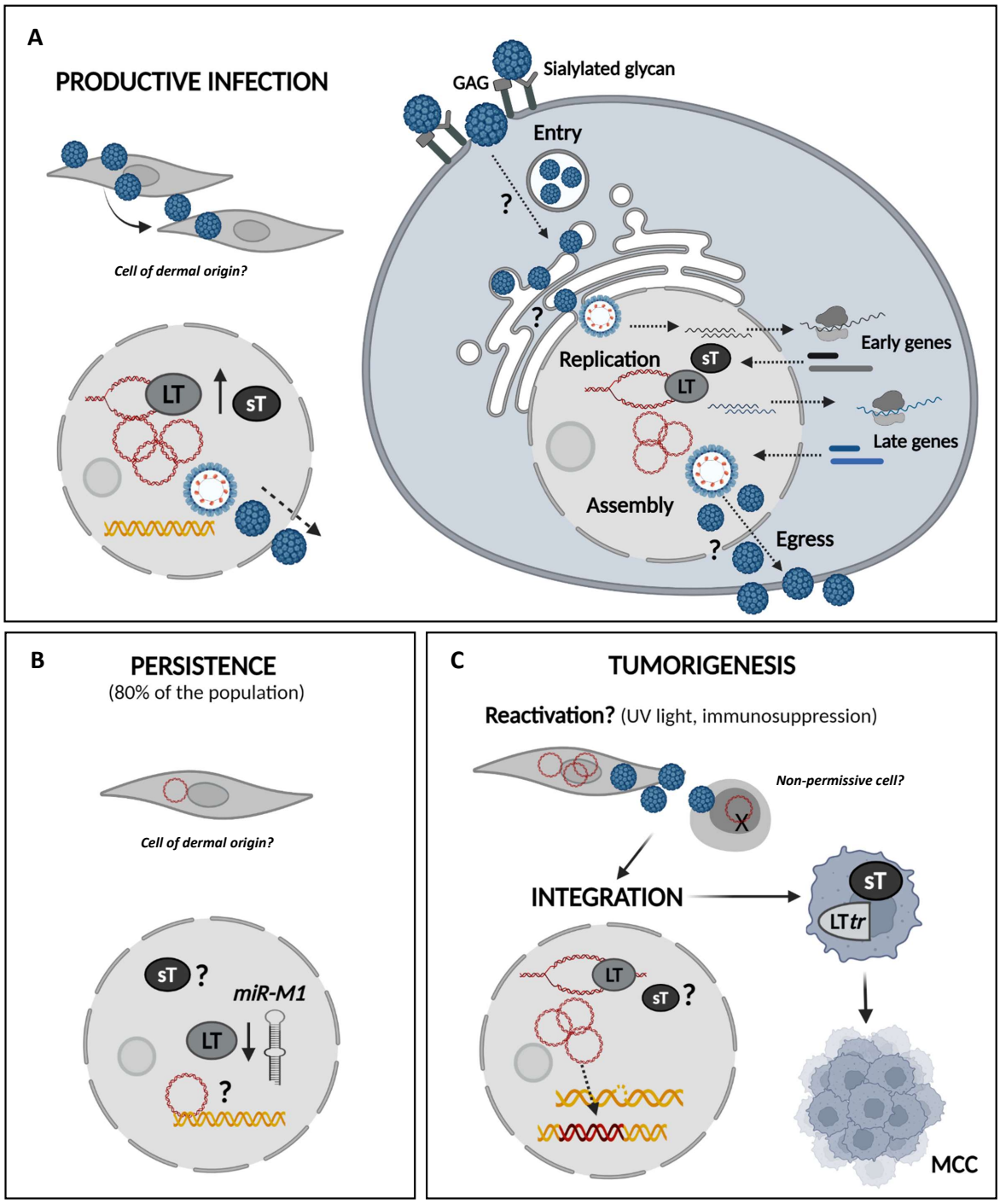


Figure 7 Life cycle of Merkel cell polyomavirus.
(A) MCPyV entry requires the attachment to glycosaminoglycans (GAGs) and sialylated glycans. Entry probably occurs via lipid-raft-mediated endocytosis and the endoplasmic reticulum (ER) is most likely involved in the process of uncoating and nuclear entry. Once reached the nucleus, the early region is transcribed followed by cytoplasmic translation into early proteins, including small tumor (sT) and large tumor (LT) antigens. LT localizes into the nucleus where it binds to the viral origin of replication and initiates full genome replication. In addition, the late viral proteins VP1 and VP2 are produced. It is not completely understood how assembly of the viral particles and egress take place. (B) A state of persistence is observed in 80% of the population but the mechanisms how and in which site of the human body viral persistence is established are unknown. However, persistence might involve the autoregulation of LT by the miR-M1, while the role for sT remains elusive. (C) Tumorigenesis represents a rare process that is not completely understood either. Risk factors such as UV light or immunosuppression might lead to the re-activation of the virus, increasing the risk of accidental integration as a result of a replication block upon infection of a non-permissive cell. LT truncation precedes integration and leads to the deregulation of cellular processes and to the loss of viral replication. In addition to LT*tr*, sT is known to interfere with several cellular factors and is considered the main driver of tumorigenesis. (created with Biorender)

1.2.4 Merkel cell polyoamvirus tumor antigens and their roles in pathogenesis

To date, four early gene products have been identified for MCPyV. While a lot of functions during the MCPyV life cycle have been ascribed to sT and LT; the roles of the 57kT and ALTO proteins are not completely understood. The 57kT antigen shares the N-terminal region with LT and parts of the second exon (see figure 6). While it retains the LxCxE and NLS domains, the OBD and helicase domain are not functional. Although it was not observed to increase viral replication, it is hypothesized to play an indirect role in MCPyV propagation (Kwun et al., 2009).

ALTO was identified by Carter et al. (2013) as an overprinting gene which is defined by the utilization of an alternative ORF and overlaps with the second exon of the LT ORF (see figure 6). Overprinting in general, may be a mechanism to increase the coding capacity of small genomes such as the MCPyV genome. MCPyV ALTO, as a protein of approximately 27 kDa, is expressed during viral genome replication although not required (Carter et al., 2013). MCPyV ALTO has similarities to the MPyV middle T antigen (van der Meijden et al., 2015) which is known to have transforming activity (Cheng et al., 2009), however the exact role of MCPyV ALTO remains elusive. Interestingly, Yang et al. (2021) have identified circular RNAs that encode ALTO variants that were found enriched in exosomes derived from MCPyV-positive MCC cell lines. It is suggested that they may play a role in transcriptional regulation of genes involved in MCPyV pathogenesis. Further studies are needed to elucidate the role of 57kT and ALTO, in order to understand MCPyV-associated pathogenesis.

It is well established that MCPyV sT and LT are the main actors in the viral life cycle and pathogenesis process. Still, there are some open questions that require further examination regarding the involvement of the T antigens. The most important interactions with cellular targets that have been revealed so far, are described in the following separately for LT and sT.

Large tumor antigen

As noted previously, LT plays a substantial role in inactivating the tumor suppressor protein pRb thereby constantly pushing the cells into the S-phase. Interestingly, while the LxCxE motif is expressed by all T antigens except for sT, the N-terminal DNA J domain is present in all T antigens. This domain was shown to be essential for binding to the cellular heat shock protein HSC70 thereby promoting cellular growth (Houben et al., 2015; Kwun et al., 2009). The

interaction of LT with pRb via LxCxE and the DNA J domain activates the E2 factor (E2F) family of transcription factors (Borchert et al., 2014; Hesbacher et al., 2016; Houben et al., 2015). Being key regulators in cell division control, E2F members regulate the transcription of DNA synthesis and cell cycle regulatory genes (Dimova & Dyson, 2005). Targeting the pRb/E2F pathway is a common feature of a variety of oncogenic viruses. Similar to MCPyV LT, SV40 LT, human papillomavirus E7 protein or KSHV LANA protein are known to inhibit pRb (Mesri et al., 2014). This function is always retained by LT*tr* expressed in MCCs (Shuda et al., 2008), which further highlights the indispensability of the tumor cells to inhibit growth suppression.

The LxCxE motif is flanked by the MUR, which is characteristic for MCPyV as it does not show any genomic similarity with other PyVs. The MCPyV MUR has been shown to interact with the lysosomal protein hVAM6p (Liu et al., 2011) which was suggested to contribute to viral genome replication. Concomitantly, Houben et al. (2015) showed that it is largely dispensable for MCC cellular growth although the second part of the MUR may be necessary for LT stability. Further studies are needed to understand its functional relevance. As elaborated in chapter 1.2.3, LT is required for the initiation of viral genome replication by binding to the viral ori. Likewise, there have been identified several cellular factors that contribute to viral replication. For example, bromodomain protein-4 (BRD4) was shown to co-localize with LT in replication centers and enhance viral DNA replication by recruiting the replication factor C (RFC) which is essential for DNA replication (Wang et al., 2012). In contrast to SV40 LT, which is known to inactivate the tumor suppressor protein p53, MCPyV LT does not interact with this protein but rather indirectly increases p53 expression by inducing a DDR via a so far undefined mechanism (Li et al., 2013). Activation of DDR factors was shown to support viral DNA replication as it was observed that Ataxia telangiectasia mutated (ATM) and Rad3-related (ATR) DDR factors co-localized with LT in replication centers (Tsang et al., 2014). Interestingly, ATM-mediated activation of the KRAB-associated protein 1 (Kap1), which was identified by Czech-Sioli et al. (2020b) as a novel interaction partner of LT, results in the induction of senescence during MCPyV genome replication in HDF cells (Siebels et al., 2020). Furthermore, Li et al. (2015) have shown that the ATM kinase phosphorylates LT at serine 816 which results in an apoptotic phenotype that can be rescued when inhibiting serine 816 phosphorylation.

These studies highlight that LT has both growth-promoting but also growth-inhibiting functions (Cheng et al., 2013) which is required for a balance of sustaining genome replication and preventing uncontrolled cell growth resulting in "dead-end" cellular transformation. With the

deletion of the C-terminal region, LT*tr* loses the ability to induce viral genome replication but on the other hand it also abolishes its growth-inhibiting properties. Together with the sT antigen, LT*tr* was shown to transform primary rat epithelial cells *in vitro* while expression of the early region containing sT and full-length LT had a decreased transforming potential (Borchert et al., 2014). This may be explained by the higher stability of LT*tr* compared to full-length LT which can be ascribed to a lower degree of ubiquitination (Czech-Sioli et al., 2020b). The fact that MCC cell lines depend on the expression of both sT and LT*tr* (Houben et al., 2010) highlights that the growth-promoting functions of LT may play an important role in tumorigenesis although the exact mechanisms are not completely understood. Nevertheless, it is well established that sT is the main driver of oncogenesis (Shuda et al., 2011) by interacting with a variety of different cellular targets which will be discussed in more detail in the following.

Small tumor antigen

With a size of around 20 kDa, sT is expressed from the early region as an extended form of the first exon it shares with the other MCPyV T antigens. Thus, it has a DNA J domain which is followed by a unique domain that provides it with additional functions (see figure 6). Although there are still many open questions regarding tumorigenesis, it is well established that sT is the main contributor by exploiting many different cellular signaling pathways. The first target of sT was reported by Shuda et al. (2011) who showed that sT interferes with the mammalian target of rapamycin (mTOR) pathway by increasing hyperphosphorylation of the 4E-binding protein 1 (4E-BP1) which is required for cap-dependent translation. It was suggested that by targeting 4E-BP1, sT might enhance viral replication. As mentioned earlier, sT has evolved several ways to increase replication, for example by stabilizing LT via the LSD (Kwun et al., 2013). Interestingly, the LSD was also suggested to be required for the ability of sT to transform rodent fibroblasts *in vitro* and in an *in vivo* mouse model (Kwun et al., 2015; Shuda et al., 2011; Verhaegen et al., 2015). Via its interaction with Fbw7, sT is able to inhibit not only the proteasomal degradation of LT but also of cellular oncogenes such as the cell cycle regulators c-Myc or cyclin E (Kwun et al., 2013). Another function linked to the LSD was discovered by Zhao et al. (2020) who observed that via its LSD, sT activates non-canonical NFκB signaling which results in the transcriptional activation of senescence-associated phenotype (SASP) genes including cytokines and growth factors such as *Il1*β, *IL6*, *IL8* or *GM-CSF*.

While sT was shown to activate non-canonical NFκB signaling, there are opposing reports on its interaction with canonical NFκB signaling. Berrios et al. (2016) observed that overexpression of

sT in IMR90 human diploid lung fibroblasts substantially perturbs the cellular transcriptome and activates the transcription of genes involved in cancer signaling pathways such as *mTORC1*, E2F targets, glycolysis and also *TNFα* via NFκB. In contrast, Griffiths et al. (2013) reported that sT inhibits NFκB-mediated transcription but it needs to be noted that this effect was observed in response to inflammatory stimuli. It was later identified that binding of sT to the protein phosphatase 4 catalytic subunit (PP4C) was essential for the interaction with the NFκB essential modulator (NEMO) adaptor protein. As a result, sT inhibits the phosphorylation of IκB which is required for nuclear localization of the NFκB heterodimer (Abdul-Sada et al., 2017). In addition to PP4C, sT specifically interacts with the protein phosphatase 2A (PP2A) Aα subunit. In contrast to SV40 sT, which binds to the PP2A Aβ subunit, this effect is not required for transformation but is suggested to play a role in viral replication (Griffiths et al., 2013; Hahn et al., 2002; Shuda et al., 2011; Verhaegen et al., 2017). Furthermore, the interaction of sT with PP2A might be involved in the interaction with PP4C (Griffiths et al., 2013; Kwun et al., 2015).

The interaction with PP4C has also been linked to the induction of increased cell motility and migration by sT. Specifically, the interaction was identified to be involved in an increased formation of filopodia, in the destabilization of the cytoskeleton by desphosphorylation of stathmin, and in Rho GTPase-mediated actin remodeling, thereby increasing cell motility and migration (Knight et al., 2015; Stakaitytė et al., 2018). In line with these findings, sT was shown to upregulate the cellular sheddases A disintegrin and metalloproteinase domain-containing protein 10 (ADAM10) and ADAM17 which disrupt cell integrity and thus increase cell dissociation and invasiveness (Nwogu et al., 2018). As recently shown, these effects might be based on the sT-induced activation of the p38-MAPK-kinase signaling pathway (Dobson et al., 2020) which highlights a possible central role of sT in increasing cell motility and migration thereby contributing to invasion and metastasis formation.

Another target of sT associated with a tumorigenic phenotype was identified by Wu et al. (2016) who observed that sT, via a so far unknown mechanism, is able to induce hyperphosphorylation of the transcription factor c-Jun which regulates several processes such as proliferation, metastasis and angiogenesis. The importance of interacting with transcription factors was reinforced with the recent observation that sT recruits the MYCL-MAX heterodimer to the EP400 transcription factor and chromatin remodeling complex. The authors found that the sT/MYCL/EP400 complex localizes to active promoters enriched with H3K4me3 marks to activate a subset of genes that is specifically involved in cancer-associated processes (Cheng

et al., 2017). Two downstream targets of the complex included the proto-oncogene *MDM2* and the protein kinase *CK1α* which, together with *MDM4* promoted p53 degradation (Park et al., 2019). This finding gave insight into a new possible mechanism of interfering with p53, which is mutated in MCPyV-negative tumors but expressed in its wild-type form in virus-positive tumors (Knepper et al., 2019). As LT is not able to directly bind and inactivate p53 (Cheng et al., 2013; Park et al., 2019), as it is the case for other viral oncogenes such as EBV BZLF1 or HPV E6 (reviewed in Tornesello et al., 2018) sT might indirectly fulfill this function. Further targets of the sT/MYCL/EP400 complex that were investigated by Park et al. (2020) included the *lysine-specific histone deacetylase* (*LSD1*), *REST co-repressor 2* (*RCOR2*) and *INSM Transcriptional Repressor 1* (*INSM1*). By activating this transcriptional repressor complex, sT indirectly downregulates the expression of a subset of targets that confer tumor suppression. Interestingly, the expression of those target genes was antagonized by the non-canonical BRG1/BRM associated factor (BAF) complex which has chromatin-remodeling functions.

These studies underscore the importance of using state-of-the-art genome-wide analysis tools to investigate the interactions of sT with the cellular transcriptional and epigenetic regulatory machinery that might be exploited by MCPyV to establish persistence or contribute to tumorigenesis. Although it is well established that both LT and sT have a large capacity to interact with cellular signaling pathways as shown by the plethora of interaction partners and functional consequences described above, it is less well understood how these functions act in consortium to contribute to either persistence establishment or tumorigenesis.

1.2.5 Role of the immune system in Merkel cell polyomavirus pathogenesis

An important factor for any viral infection is represented by the interaction with the immune system. In the case of persistent oncogenic viruses, this is in particular interesting with regard to the different phases of infection. In the phase of acute and productive infection, the innate immune system is activated by the large amount of PAMPs that are exposed during the viral life cycle. In order to develop persistence, it is essential from the viral perspective to dampen immune recognition in order to maintain its genome. This can be achieved for example by low-level replication or by tethering the viral genome to the host genome which is a known function of the KSHV LANA protein, for example (Ballestas et al., 1999).

Reactivation or accidental infection of non-permissive cells can result in tumorigenesis. As elaborated in the previous chapter, expression of sT and LT*tr* in MCCs drives tumorigenesis and ensures tumor cell survival. In this scenario there is another level of immunoediting required for the tumor cell survival. In fact, MCCs have established several mechanisms to bypass the immune response, especially with regard to the T cell response. Intriguingly, restoration of immune checkpoint functions is part of a new MCC treatment strategy (see chapter 1.2.1).

In addition to inhibiting T cell activation and function which is considered part of the adapative immune system, the innate immune system might also play a role in immune evasion. Interestingly, Liu et al. (2020) observed that the cytosolic DNA sensing molecule STING is silenced in MCCs. Restoration of STING expression resulted in the production of inflammatory cytokines and chemokines, leading to increased T cell migration and infiltration. Furthermore, in a large amount of MCCs it was observed that the MHC- I chain-related proteins MIC-A and MIC-B are downregulated via promoter hypoacetylation, resulting in the inhibition of natural killer (NK) cell-mediated cell lysis (Ritter et al., 2016). Furthermore, in a recent study conducted in our group we showed that MCPyV sT increases the cell surface expression of CD47 on MCPyV+ MCCs, thereby inhibiting macrophage-induced phagocytosis (Schlemeyer et al., in revision).

Although these studies highlight that MCCs, similarly to most other cancers, require immune evasion mechanisms for their survival, little is known about the immunogenicity of the viral infection preceding tumor formation. As discussed in chapter 1.1.2, the role of the type I IFN response might be central for the outcome of PyV infections. In fact, Shahzad et al. (2013) showed that both MCPyV LT and sT inhibited the expression of TLR9, which is an endosomal PRR recognizing dsDNA, by targeting C/EBPβ which is a transactivator of the TLR9 promoter. In a recent study it was reported that infection of HDFs with MCPyV induces an innate immune response with associated expression of ISGs and inflammatory cytokines (Krump et al., 2021). The authors suggest that during the later stage of infection, MCPyV DNA might be sensed by IFI16 or cGAS-STING and via IRF3 and NfκB activate the transcription of ISGs and inflammatory cytokines. Interestingly, in a study from Akhbari et al. (2018), it was shown that the viral miRNA miR-M1 was associated with decreased CXCL8 expression via targeting the host gene SP100. This anti-inflammatory effect resulted in reduced neutrophil chemotaxis and was suggested to play a role during MCPyV infection. However, future studies are needed that further address the role of the innate immune response in MCPyV infection and uncover immune evasion mechanisms used by MCPyV to establish either persistent infection or tumorigenesis.

1.3 Epigenetic modifications

From an etiologic point of view, the term "epigenetics" contains the Greek word *epi* which means *over*, and thus describes the study of mechanisms resulting in heritable genomic functions that underlie changes which are not associated with the DNA sequence itself. Epigenetic changes can be directly linked to DNA such as methylation of the fifth carbon of cytosine (5meC) which is generally associated with stable gene repression (Jones, 2012). It is well established that DNA methylation plays important roles in genomic imprinting, X-chromosome inactivation, silencing of endogenous transposable elements, or during cell development and differentiation (Greenberg & Bourc'his, 2019; Jones, 2012). Aberrant DNA methylation has been linked to several disease phenotypes such as neurological disorders, immunodeficiencies and cancer with the latter being one of the best studied examples (Pfeifer, 2018).

One of the most diverse studied types of epigenetic modification is post-translational modification (PTM) of histones, which is discussed in more detail in the following section.

1.3.1 Histone modifications

The genomic DNA inside a human cell is estimated to be 2 m long and thus it needs to be compacted and well organized in order to be accessible at the right time. Therefore, it is wrapped around a nucleosomal core of histone proteins that are arranged as octamers consisting of two copies of four histone proteins - H2A, H2B, H3 and H4. The linker histone H1 is additionally located outside of the nucleosomal core and interconnects several nucleosomes to stabilize higher-order chromatin structures (Lawrence et al., 2016). The discovery that these histones can undergo PTM (Phillips, 1963) has opened up a new research field of epigenetic gene regulation. There is a large variety of histone modifications which mostly occur on the "tails" of the histones protruding from the nucleosomes and influence the accessibility of DNA for transcription factors. Modifications such as methylation, acetylation, ubiquitination, phosphorylation or sumoylation have been described for different amino acids and are associated with distinct functions.

In general, histone modifications can be classified into activating marks associated with euchromatin such as H3K4me3, H3K9ac or H3K14ac (Karmodiya et al., 2012; Santos-Rosa et al., 2002). On the other hand, repressive marks including H3K27me3 or H3K9me3, can be located in heterochromatic regions (Bernstein et al., 2006; Peters et al., 2001). However, they are not

always mutually exclusive, as both activating and repressive marks can co-occur at the same region, also named bivalent chromatin. This term was first introduced by Bernstein et al. (2006) who observed a large amount of H3K27me3, occuring with smaller amounts of H3K4me3 at gene promoters in embryonic stem cells, allowing a poised state in which genes are generally silenced but predisposed to be activated if necessary. Figure 8 gives a brief overview of the most frequent and best studied histone modifications, including H3K4me3, H3K27me3, H3K9me3 and H3K27ac which were also addressed in this work. Furthermore, as each histone mark is regulated by specific reader, writer and eraser proteins, some of the most important writer proteins are depicted in figure 8.

Transcriptionally permissive chromatin can be marked by mono-, di- or tri-methylation of histone 3 at lysine 4 (H3K4me1, H3K4me2, H3K4me3). All of them can be mediated by the SET domain-containing proteins SET1A and SET1B. Mono- and di-methylation can also be conferred by the histone-lysine N-methyltransferase 2A proteins KMT2A and KMT2B. Mono-methylation is additionally accomplished by KMT2C or KMT2D (Skvortsova et al., 2018). Demethylation can be established by lysine demethylases (KDMs) which are classified into eight subfamilies. The lysine-specific demethylases LSD1 or LSD2, for example, confer demethylation of H3K4me1 and H3K4me2. However, H3K4me3 can be demethylated by KDM2B (i.e. FBXL10) or KDM5A-D (i.e. JARID1A-D) (Arifuzzaman et al., 2020). H3K27ac represents another mark which is found in active promoters, often coinciding with H3K4me3, and is mediated by the transcription coactivators CREB binding protein (CBP) and p300 (Tie et al., 2009). Deacetylation activity has been shown for at least 18 histone deacetylase (HDAC) enzymes (Park & Kim, 2020). Interestingly, H3K27ac can also be found in active enhancer regions where it co-occurs with H3K4me1 (Creyghton et al., 2010).

Much of the genome of higher eukaryotes is stably heterochromatinized and often marked by H3K9me3, also referred to as constitutive heterochromatin. These regions include repetitive sequences such as transposable elements, or centromeric and telomeric repeats that need to be constantly silenced to maintain genomic integrity. In addition, H3K9me3 is also found within gene regulatory regions thereby influencing transcription of genes. H3K9me3 can, among others, be mediated by the writer proteins SU(VAR)3-9 homologs SUV39H1/2 and SETDB1 which are recruited by the reader protein HP1 that binds to H3K9me3, resulting in the condensation of chromatin. Removal of H3K9me3 can be mediated by several KDMs such as KDM3 or KDM4 proteins (Ninova et al., 2019), but also by KDM1A (i.e. LSD-1), as observed by Metzger

et al. (2005). In contrast to the stable silencing factor H3K9me3, H3K27me3 can be found in regions with facultative heterochromatin. While H3K9me3 generally silences genes, it retains the ability to switch to a state of euchromatin in a temporal or spatial manner, a characteristic which is referred to as developmentally regulated chromatin (Trojer & Reinberg, 2007). Another important histone modification is H3K27me3, mediated by the polycomb repressive complex 2 (PRC2) which consists of the enhancer of zeste homolog 2 (EZH2) and the co-factors SUZ12 and EED. Removal of H3K27me3 is normally mediated by KDM6A (i.e. UTX) and KDM6B (i.e. JMJD3), as reviewed by Cloos et al. (2008).

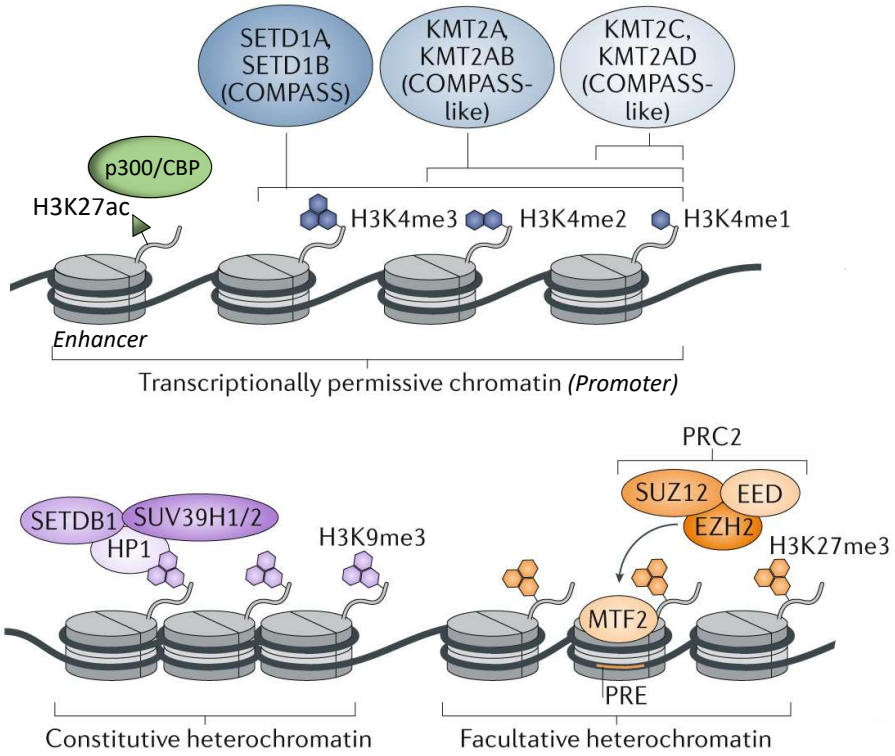


Figure 8 Histone modifications and their writer proteins.
Schematic representation of exemplary histone modifications and their writer proteins mediating chromatin accessibility, either allowing or prohibiting transcription. Transcriptionally permissive chromatin can be marked by H3K4me1, H3K4me2 or H3K4me3 in the promoters of actively transcribed genes which is mediated by SET domain-containing proteins (SETDs) or histone-lysine N-methyltransferases (KMTs). H3K27ac, mediated by p300/CBP, can be present both in active promoter regions as well as active enhancer regions. Repressive chromatin is classified into constitutive and facultative heterochromatin, with the latter referring to a more dynamic state of chromatin conformation allowing to switch from heterochromatin to euchromatin. H3K9me3 is found at constitutive heterochromatin, mediated by SETDB1 and SUV39H1 that is often recruited by HP1. H3K27me3 is mediated by the PRC2 complex which is composed of EZHC, SUZ12 and EDD, for example, and is found in regions of facultative heterochromatin. (modified from Skvortsova et al., 2018)

The histone modifications, their writers, readers and erasers summarized above only represent a few examples of histone modifications that act in combination to influence transcription, by either allowing or inhibiting chromatin accessibility for transcription factors and polymerase II. Intriguingly, aberrant changes in histone modification patterns can have tremendous effects on the cellular transcription machinery and deregulate basic cellular processes, thereby providing an attractive target for several pathogens. Since many viruses also hijack the cellular epigenetic machinery to replicate, establish persistence, or evade from the host immune system, it is likely that these functions are accompanied by deregulation of host cellular processes (Fischer, 2020).

1.3.2 Epigenetic deregulation in Merkel cell carcinoma

Because epigenetic modifications play an important role in the regulation of gene expression, their deregulation is often associated with disease phenotypes including cancer. One of the best studied examples is DNA methylation which is often deregulated in several types of cancers. Similarly, changes in DNA methylation have been reported also in MCC. For example, members of the *RASSF* tumor suppressor protein were observed to be hypermethylated in several MCCs, independent of the patient's age or disease progression and also irrespective of the viral status (Richter et al., 2013). Furthermore, Sahi et al. (2014) reported that the promoter of the tumor suppressor gene *Rb* was hypermethylated in several MCCs, including both MCPyV-positive and -negative MCCs. In a systematic study from Gravemeyer et al. (2021), using an array-based approach it was shown that MCPyV-positive and -negative MCCs had similar DNA methylation patterns. These observations highlight that DNA methylation in MCCs is frequently observed but has not been described to be viral-induced.

Concerning histone modifications it was reported that the expression of the PRC2 member EZH2 was increased in MCCs compared to normal skin tissues which was accompanied by increased disease progression (Harms et al., 2017; Veija et al., 2017). While these studies did not differentiate between MCPyV-positive and -negative tumors, a report from Matsushita et al. (2019) revealed that virus-positive MCCs had a higher degree of H3K27me3 than the virus-negative tumors. However, a study from Busam et al. (2017) stated the exact opposite, showing lower H3K27me3 levels in MCPyV-positive compared to virus-negative tumors. In conclusion, H3K27me3 is probably deregulated in MCCs, however the impacts and the underlying causes remain to be investigated.

In addition to the inactivation of tumor suppressor genes for example by DNA methylation, interfering with the host immune response has been associated with epigenetic deregulation. It could be shown by Ritter et al. (2016) that the expression of MICA and MICB was observed to be completely lacking in several MCC cell lines. Furthermore, the authors revealed that H3K9ac, which is associated with active promoters, was dramatically reduced in the MICA and MICB promoters. Consequently, H3K9ac could be reverted by treating the cells with HDAC inhibitors by restoring immune recognition of MCCs. Furthermore, Ritter et al. (2017) observed that the treatment with HDAC inhibitors also restored HLA class-I surface expression which had been previously reported to be reduced in MCC cell lines and tissues, resulting in impaired CD8+ T cell activation and infiltration (Paulson et al., 2014). *LMP2*, *LMP7*, *TAP1* and *TAP2*, which are key components of the antigen presentation machinery (APM), were shown to be transcriptionally silenced in the majority of MCCs which was suggested as the underlying cause for HLA-I downregulation (Ritter et al., 2017). Although HDAC inhibition resulted in increased mRNA levels of APM genes, the exact underlying mechanism remains to be determined.

These studies underscore that epigenetic deregulation is a common feature in both MCPyV-positive and MCPyV-negative MCCs but its complete impact and functions are not understood to date. Although most of the studies did not investigate the engagement of MCPyV T antigens in epigenetic deregulation, there is some evidence that they potentially interfere with transcriptional processes via epigenetic mechanisms. For example, one of the targets of the sT-MYCL-EP400 complex included *LSD1* that mediates H3K4/K9 methylation. In fact, inhibition of *LSD1* was observed to induce cell cycle arrest and cell death, and abrogate tumor growth of MCPyV-driven MCC (Park et al., 2020). Nevertheless, the majority of the studies reported epigenetic aberrations in both virus-positive and -negative MCCs, which suggests that epigenetic changes might be induced by T antigens but can also be established independent of their expression. Intriguingly, as suggested by several reports, targeting histone modifying enzymes represents a promising therapeutic approach in the context of both viral-induced and virus-negative MCCs (Ritter et al., 2017; Ritter et al., 2016; Song et al., 2021). The present study aimed at identifying whether MCPyV T antigens interfere with cellular transcriptional processes by targeting cellular regulation of histone modifications.

2 Aim of the Study

MCPyV is the only human PyV that is associated with tumorigenesis in its primary host. Although the formation of MCC represents a rare event, it is associated with a rapid growing skin cancer with high metastatic potential. Pathogenesis of MCPyV is mainly mediated by the oncoproteins sT, LT and LT*tr* which have been shown to interfere with a large variety of cellular processes. On the molecular level, LT was shown to bind and inhibit the tumor suppressor protein pRb thereby increasing the transcription of E2F-dependent S-phase genes (Borchert et al., 2014; Shuda et al., 2008). Both sT and LT*tr* were identified to activate the transcription of genes that contribute to cellular proliferation and tumorigenesis (Richards et al., 2015). An important finding was also that sT specifically regulates the transcription of genes involved in tumor growth by recruiting the MYCL-EP400 transcription factor and chromatin remodeling complex (Cheng et al., 2017). Although MCPyV T antigens have been shown to have diverse functions in the life cycle of MCPyV and interact with host cellular processes such as gene regulation, many open questions regarding pathogenesis remain, which is mostly due to the lack of knowledge about the host cell types allowing productive infection or persistence. It is therefore incompletely understood how persistence establishment is supported by the T antigens. Because the cell of origin of MCC is also not known, deciphering the events that lead to transformation is challenging, and so far the roles of sT and LT*tr* are only partially understood.

The aim of this study is to systematically analyze transcriptional and epigenetic changes induced by sT, LT and LT*tr* thereby detecting possible new functions that might contribute to MCPyV pathogenesis. Overexpression of sT together with full-length LT, resembling the initial phase of infection, and overexpression of sT and LT*tr*, reflecting a scenario present in a transformed cell, were conducted due to the lack of *in vitro* infection or tumorigenesis models. Different time points of investigation were chosen to reveal immediate changes induced by T antigen expression, as well as possible stable effects. Transcriptional profiling combined with ChIP-Seq analysis address the question whether transcriptional changes induced by MCPyV T antigens are associated with changes in the composition of histone modifications. The systematic genome-wide analysis conducted in this work contributes to the understanding of the role of the T antigens in MCPyV pathogenesis by revealing an important involvement of the innate immune response.

3 Materials and Methods

3.1 Laboratory reagents, chemicals and commercial systems

In this work, if not stated differently, all standard chemicals were purchased from Roth (Karlsruhe), Sigma-Aldrich (München) or Merck (Darmstadt). An overview of all commercial systems that were used in this work is given in table 1. Further reagents required for the experiments conducted in this work are listed in table 2.

Table 1 Commercial systems used in this work

Name	**Manufacturer**
Biorad Protein Assay	Biorad (München)
DNeasy© Blood and Tissue Kit	Qiagen (Hilden)
DNA Plasmid Purification Kit NucleoBond© PC100	Macherey-Nagel (Düren)
DNA-free^{TM} DNA Removal Kit	Invitrogen (Darmstadt)
ECL Western Blotting Detection Reagents	Santa Cruz Biotechnology (Heidelberg)
Infinium© Methylation EPIC Kit	Illumina (San Diego, CA, USA)
femtoLucent Plus HRP Kit	G-Biosciences (St. Louis, USA)
PCR clean-up, Gel extraction Kit	Macherey-Nagel (Düren)
Qubit© RNA BR /DNA HS Assay Kits	Invitrogen (Darmstadt)
RNeasy© Mini Kit	Qiagen (Hilden)

Table 2 Reagents utilized in this work

Name	**Manufacturer**
Ampicillin Sodium Salt	Sigma (München)
Aqua B.	Braun (Melsungen)
BSA (Bovine Serum Albumin)	GE Healthcare (Freiburg)
Color Protein Marker	NEB (Frankfurt)

cOmplete™ Protease Inhibitor cocktail	Roche Diagnostics (München)
Coverslips	A. Hartenstein (Würzburg)
DNA loading buffer	Thermo Fisher Scientific (Waltham, MA, USA)
Dithiothreitol (DTT)	Sigma (München)
dNTPs (dATP, dCTP, dGTP, dTTP)	Invitrogen (Karlsruhe)
Glass slides	Karl Hecht (Sondheim)
Maxima SYBR Green/ROX©qPCR Master Mix (2x)	Thermo Fisher Scientific (Waltham, MA, USA)
Millex-GP Syringe Filter Units (0.22 µm)	Merck Millipore (Darmstadt)
Paraformaldehyde (16%)	Electron Microscopy Sciences (Hatfield, PA, USA)
Pefablock© SC-Protease-Inhibitor	Roth (Karlsruhe)
PhosSTOP™ Phosphatase Inhibitor-cocktail	Roche Diagnostics (München)
Pierce™ ChIP-grade Protein A/G Magnetic beads	Thermo Fisher Scientific (Waltham, MA, USA)
Polybrene©	Sigma (München)
Polyethylenimine (PEI)	Polyscience Inc. (Washington, USA)
Powdered milk	Roth (Karlsruhe)
QIAshredder	Qiagen (Hilden)
Roti©-GelStain	Roth (Karlsruhe)
Roti©-PVDF-Membrane (0.45 µm)	Roth (Karlsruhe)
Sodium azide (NaN_3)	Sigma (München)
TRIzol™ Reagent	Thermo Fisher Scientific (Waltham, MA, USA)
Ultra Pure™ Agarose	Invitrogen (Darmstadt)
Vectashield mounting medium with DAPI	Vector Laboratories Inc. (Burlingame, CA, USA)

Whatman paper	A. Hartenstein (Würzburg)
X-ray films	Fuji-Film Europe (Düsseldorf)

Several enzymes were used for different experimental applications. An overview of them is given in table 3.

Table 3 Enzymes used in this work

Name	**Application**	**Manufacturer**
EcoRI	plasmid digestion	NEB (Frankfurt a.M.)
NotI	plasmid digestion	NEB (Frankfurt a.M.)
NdeI	plasmid digestion	NEB (Frankfurt a.M.)
MNase	nChIP	NEB (Frankfurt a.M.)
PureLink^{TM} RNase A	nChIP	Thermo Fisher Scientific (Waltham, MA, USA)
DNase I	RNA isolation	Invitrogen (Karlsruhe)
Superscript IV^{TM}	cDNA synthesis	Invitrogen (Karlsruhe)
murine RNase Inhibitor cocktail	cDNA synthesis	Invitrogen (Karlsruhe)

The compositions of all buffers and chemical solutions that were used in general are summarized in table 4.

Table 4 General buffers and chemical solutions used in this work

Name	**Composition**
PBS (Phosphate buffered saline), pH 7.5	140 mM NaCl
	25 mM KCl
	0.5 mM $MgCl_2$
	1 mM $CaCl_2$

	10 mM Na_2 HPO_4
PBS-Tween	1x PBS
	0.05% (v/v) Tween 20
TBS (Tris buffered saline), pH 7.5	50 mM Tris-HCl
	150 mM NaCl
TBS-Tween	1x TBS
	0.05% (v/v) Tween 20
TAE, pH 7.8	2 M Tris-HCl
	0.2 M sodium acetate
	1 mM EDTA
WB Buffer A, pH 10.4	0.3 M Tris
	10% (v/v) methanol
WB Buffer B, pH 10.4	0.025 M Tris
	10% (v/v) methanol
WB Buffer C, pH 9.4	0.025 M Tris
	0.025 M aminohexanoic acid
	10% (v/v) methanol
Permeabilization buffer (IF)	1x PBS
	1% Triton X-100
	0.1% sodium citrate
Blocking buffer (IF)	1x PBS
	1% Triton X-100
	0.5% Tween 20
	3% BSA
RIPA (lysis) buffer	50 mM Tris-HCl pH 7.5
	150 mM NaCl
	1% NP40

	0.5% DOC
	0.1% SDS
	2 mM β-glycerolphosphate
	1 mM Na_3VO_4
	0.4 mM PMSF
	1 mM NaF
	1 mM EGTA
	Protease inhibotor 1 tablet/50 mL
4x Protein loading buffer	12% SDS
	40% (v/v) glycerol (86%)
	0.2 M Tris-HCl, pH 7
	0.004% (x/v) bromphenolblue
	5% (v/v) β-mercaptoehtanol
SDS running buffer	25 mM Tris
	0.192 M glycine
	3.5 mM SDS
10x PBS	140 mM NaCl
	25 mM KCL
	0.5 mM $MgCl_2$
	1 mM $CaCl_2$
	10 mM Na_2HPO_4
Lentivirus concentration buffer	8.5% PEG6000
	0.3 M NaCl
	1x PBS

3.2 Instruments and equipment

For the experiments conducted in this work, if not stated differently, the equipment and devices were used from the following manufacturers: Agfa, Agilent, Biorad, BD Biosciences, Binder, Diagenode, Eppendorf, Heraeus, Illumina, Invitrogen, Leica, Nikon, Sarstedt, PeqLab, Qiagen, Sorvall, Thermo Fisher Scientific, Zeiss, 10x Genomics.

3.3 Antibodies

Antibodies were used for immunofluorescence (IF) analysis, Western Blots (WBs) or ChIP assays and are listed in table 5.

Table 5 Primary and secondary antibodies used in this work

Target	Name	Application	Concentration	Company
H3K4me3	anti-trimethyl-histone H3 (Lys4), rabbit monoclonal (clone MC315)	ChIP	1:200	Merck Millipore
H3K27me3	anti-trimethyl-histone H3 (Lys27), rabbit monoclonal (C36B11)	ChIP	1:300	Cell Signaling
H3K9me3	anti-histone H3K9me3, rabbit polyclonal	ChIP	1:200	Active Motif
H3K27ac	anti-histone H3 (acetyl K27), rabbit polyclonal	ChIP	1:200	abcam
rabbit IgG	normal Rabbit IgG, rabbit polyclonal	ChIP	1:200	Merck Millipore
MCPyV sT/LT*tr*	2T2 antibody	WB	1:1	self-made
MCPyV LT	CM2B4 (sc-136172)	WB	1:1000	Santa Cruz Biotechnology

actin (all isoforms)	anti-Actin, clone 4 (mouse monoclonal)	WB	1:10,000	Merck Millipore
pSTAT1	anti-phospho-STAT1 (rabbit monoclonal)	WB	1:1000	Cell Signaling
STAT1	anti-STAT1 (rabbit monoclonal)	WB	1:1000	Cell Signaling
mouse IgG	Peroxidase-AffiniPure goat anti-mouse IgG	WB	1:10,000	dianova
rabbit IgG	anti-rabbit IgG, HRP-linked antibody	WB	1:3000	Cell Signaling
IRF9	anti-IRF9, rabbit monoclonal (D2T8M)	WB, IF	1:1000	Cell Signaling
rabbit-IgG-633	goat anti-rabbit IgG secondary antibody, Alexa Fluor 633, polyclonal	IF	1:200	Thermo Fisher Scientific

3.4 Plasmids

Plasmids that were used in this work are listed in table 6.

Table 6 Plasmids used in this work

Name	**Vector**	**Insert, Promoter**	**Origin**
Lenti gag-pol	pMD	HIV-gag/pol, CMV	addgene
Lenti rev	pRSV-Rev	HIV-gp6/rev, RSV	addgene
Lenti VSV env	phCMV	VSV-G Env, CMV	addgene
LeGO-iC2	LeGO-iC2	IRES-mCherry, SFFV	Weber et al. (2008)
LeGO-iC2-MCPyV-sT	LeGO-iC2	sT-IRES-mCherry, SFFV	cloned by C. Hartig

LeGO-iG2	LeGO-iG2	IRES-eGFP	Weber et al. (2008)
LeGO-iG-MCPyV-LT	LeGO-iG2	LT-IRES-eGFP, SFFV	cloned by C. Hartig
LeGO-iG-MCPyV-LT*tr*	LeGO-iG2	LT*tr*-IRES-GFP, SFFV	cloned by C. Hartig

3.5 Oligonucleotides

Oligonucleotides were purchased from Sigma or Eurofins. Table 7 summarizes the sequences of all primers used in this work. Molecular probes that were required for the TaqMan™ qPCR assays are listed in table 8.

Table 7 Oligonucleotides used in this work

Name	**Sequence**	**Application**
ADH5 fw	GCATAATTGAGCCTACGCC	ChIP-qPCR
ADH5 rev	GCAGAGGTGTTTGTTACGTG	ChIP-qPCR
C1orf43 fw	AGTGGGTGGAGAATGCAGAC	ChIP-qPCR
C1orf43 rev	GAGATTACCCCACCCCATTC	ChIP-qPCR
HOXC12 fw	AGTAGTTCGCCCCCAGATTT	ChIP-qPCR
HOXC12 rev	GCGGAAGGGAGGTAGAGAA	ChIP-qPCR
STAT1 fw	CATTCACATGGGTGGAGCG	ChIP-qPCR
STAT1 rev	GGGTTCAACCGCATGGAAG	ChIP-qPCR
ZNF268 fw	AATGCATTTCCACACTGCAA	ChIP-qPCR
ZNF268 rev	AAAGAGGTTGCTGCCAAGAC	ChIP-qPCR
GAPDH fw	GACCTCAACTACATGGTTTACATGTTCC	RT-qPCR
GAPDH rev	GCAAATGAGCCCCAGCCTTC	RT-qPCR
HPRT1 fw	TGACCTTGATTTATTTTGCATACC	RT-qPCR
HPRT1 rev	CTCGAGCAAGACGTTCAGTC	RT-qPCR
IRF9 fw	ACCAGGATGCTGCCTTCTTC	RT-qPCR

IRF9 rev	CCTGGTGGCAGCAACTGATA	RT-qPCR
OAS2 fw	GAACACCATCTGTGACGTCCT	RT-qPCR
OAS2 rev	GTACCATCGGAGTTGCCTCT	RT-qPCR
sT fw	GTCTCGCCAGCATTGTAGTC	RT-qPCR
sT rev	CCTCGTCAACACAGAGGAAG	RT-qPCR
LT fw	TGGTTGTTTTTGAGGATGTG	RT-qPCR
LT rev	AAGGTTGTATCAGGCAAG	RT-qPCR
LT*tr* fw (MCC350)	TATGTTTGATGAGGTTGACG	RT-qPCR
LT*tr* rev (MCC350)	AGGTATATCGGGTCCTCTGGACTGGG	RT-qPCR
MCPyV VP1 fw	AAAACACCCAAAAGGCAATG	qPCR
MCPyV VP1 rev	GCAGAGACACTCTTGCCACA	qPCR

Table 8 Molecular probes required for TaqMan™ qPCR assays

Name	**Sequence**	**Application**
GAPDH-TaqMan™	VIC-GTGGCGCTGAGTACGTCGTGGAGTC	RT-qPCR
MCPyV VP1-TaqMan™	6FAM-GATCTGGAGATGATCCCTTTGGCTG	qPCR

3.6 Software and online tools

For the data generation, analysis and visualization, the following software and online tools were used: Microsoft Office (Excel, Powerpoint, Word), CLC DNA Workbench (Version 8.0.1, Qiagen), Photoshop 21.2.1 and Illustrator 24.2.3 (Adobe Inc.), Graph Pad Prism 5, Rotorgene Q Series Software (Qiagen), BD FACSDiva™ Software (BD Biosciences), Gel Doc XR (Bio Rad), Leica Confocal Software, Image J, Volocity Demo (Version 6.1.1, Perkin Elmer), STAR 2.6.0 (Dobin et al., 2013), MACS 2.1.2 (Zhang et al., 2008), bedtools 2.27.1 (Quinlan & Hall, 2010), diffReps 1.55.6 (Shen et al., 2013), IGV 2.10.2 (Thorvaldsdóttir et al., 2013), EaSeq 1.11 (Lerdrup et al.,

2016), Ingenuity Pathway Analysis (IPA©, Qiagen), DAVID Bioinformatics Resources 6.8 (Huang et al., 2009), R Studio (1.2.5042), BioRender.

3.7 Methods of prokaryotic cell culture

Bacteria were used for plasmid amplification and they were generally maintained in Luria-Broth (LB) medium and on LB agar plates. The medium was prepared from powder (Roth, Karlsruhe) using the indicated instructions. Ampicillin was added at a final concentration of 100 µg/mL as all the plasmids used in this work contained an ampicillin resistance gene.

3.7.1 Transformation of chemically competent bacteria

Bacteria from the strain *E.coli Top Ten* that were to be transformed for plasmid amplification were made chemically competent using the following protocol. First, a bacterial pre-culture was generated by inoculating 20 mL LB medium with a colony of *E.coli Top Ten*. After an overnight incubation at 37°C, a new culture was prepared by diluting the pre-culture with fresh LB medium at a ratio of 1:100. 1 M $MgSO_4$ and 1 M KCl were added to the culture which was incubated at 37°C until it reached an OD_{600} between 0.3 and 0.5. Subsequently, the culture was cooled on ice for 15 min and pelleted at 6000 x g and 4°C for 5 min. The pellet was resuspended in 60 mL transformation buffer I and incubated for another hour on ice. Following a centrifugation step at 6000 x g and 4°C for 5 min, the pellet was resuspended in 8 mL transformation buffer II and 100 µL aliquots were immediately flash-frozen using liquid nitrogen and stored at -80°C. The composition of the transformation buffers is shown in table 9.

Table 9 Composition of buffers used for the preparation of chemically competent *E.coli*

Buffer	**Components**
Transformation buffer I, pH 5.8	15% (v/v) glycerin
(adjusted with CH_3COOH)	10 mM $CaCl_2$
	100 mM $RbCl_2$
	50 mM $MnCl_2$

Transformation buffer II, pH 6.8	15% (v/v) glycerin
(adjusted with KOH)	10 mM MOPS
	75 mM $CaCl_2$
	10 mM $RbCl_2$

3.7.2 Plasmid amplification using chemically competent bacteria

In order to amplify plasmids required for the transfection of eukaryotic cells, chemically competent bacteria were transformed using the following procedure. An aliquot of *E.coli Top Ten* was thawed on ice and 100 ng plasmid DNA was added to 50 µL bacterial stock. After an incubation on ice for 30 min, a heat shock was conducted in a heating block at 42°C for 45 s, followed by an immediate cooling step on ice for 5 min. Subsequently, 900 µL fresh LB medium were added to the bacteria and the sample was incubated for 1.5 h at 37°C. The Lentiviral Gene Ontology (LeGO) vectors were incubated at only 30°C in order to prevent recombination events. Subsequently, bacteria were plated on LB agar plates containing 100 µg/mL ampicillin and incubated overnight at 30°C or 37°C.

In the next step, one bacterial colony was transferred into 200 mL fresh LB medium (with ampicillin) and the culture was inoculated overnight at 30°C or 37°C. Total plasmid DNA was isolated using the DNA Plasmid Purification Kit (see table 2) according to the manufacturer's protocol. The final elution step was performed using nuclease-free water and the concentration was determined using a NanoDrop 2000 device (PeqLab, Erlangen).

3.7.3 Restriction enzyme digestion of amplified plasmid DNA

To ensure a pure and correct plasmid amplification, the DNA isolated with the DNA Plasmid Purification Kit was digested with restriction enzymes, followed by agarose gel electrophoretic analysis. All enzymes used in this work are shown in table 3. In general, 1 µg plasmid DNA was mixed with 2 µL 10x Fast Digest buffer and 1 µL of the respective enzyme(s). Water was added to a total volume of 20 µL and the sample was incubated at 37 °C for 10 min. Subsequently, the digestion fragments were separated on an agarose gel as described in the following chapter.

3.7.4 Agarose gel electrophoresis

Agarose gel electrophoresis allows the separation of DNA fragments according to their size in an electric field. Due to its property of being negatively charged, DNA fragments with a high size migrate slower in the gel than smaller fragments. In this work, agarose gel electrophoresis was used to visualize plasmid DNA digested with restriction enzymes (see section 3.7.3) or DNA fragments obtained after MNase digestion for ChIP assays (see chapter 3.12). Briefly, agarose was boiled with 1x TAE buffer at a final concentration of 0.8% or 1%. When completely dissolved, DNA intercalating agents such as 0.005% Roti© GelStain or 0.5 µg/mL ethidium bromide were added and the gel was poured and left at room temperature until it was polymerized. The samples were mixed with 6x DNA Loading Buffer and loaded on the gel which had been transferred into a gel chamber filled with 1x TAE buffer. In addition, a DNA marker with known sizes was loaded in one gel lane. Applying 90-120 V for 45-55 min, the DNA fragments were separated and visualized under UV light using a GelDoc-station.

3.8 Methods of eukaryotic cell culture

3.8.1 Culture and storage of eukaryotic cell lines

Eukaryotic cell lines were cultured in polystyrene dishes or flasks in incubators at 37 °C with 5% CO_2 and a relative humidity of 95%. The different cell lines were cultured in specific media that contained certain supplements. A summary of all media, supplements and reagents that were required for cell culture is given in table 10.

Table 10 Cell culture media, supplements and reagents used in this work

Reagent	Purchased from
Dulbecco's Modified Eagle Medium (DMEM) with Glutamax	Gibco (Bleiswijk, Netherlands)
Dulbecco's Phosphate Buffered Saline (DPBS)	Gibco (Bleiswijk, Netherlands)
Fetal Bovine Serum (FBS)	Gibco (Paisley, UK)
HEPES	ThermoFisher Scientific (Waltham, MA, USA)

Hybridoma-SFM (serum-free medium)	Gibco (Grand Islands, NY, USA)
Medium 199	Gibco (Paisley, UK)
Opti-MEM (Reduced-Serum Minimal Essential Medium)	Gibco (Paisley, UK)
Penicillin/streptomycin (P/S)	Gibco (Grand Islands, NY, USA)
RPMI (Roswell Park Memorial Institute) 1640	Gibco (Paisley, UK)
Sodium pyruvate	ThermoFisher Scientific (Waltham, MA, USA)
Trypsin/EDTA, 0.05%	Gibco (Paisley, UK)
50x HFCS (Hybridoma Fusion and Cloning Supplement, #11363735001)	Roche Diagnostics (München)

All eukaryotic cell lines used in this study are listed in table 11. Furthermore, table 12 depicts all cell lines and their respective media in which they were maintained.

Table 11 Overview of eukaryotic cell lines used in this work

Name	**Characteristics**	**Obtained from**
nHDF Donor I	primary cell line isolated from neonatal foreskin	Lonza (CC-2509)
nHDF Donor II	primary cell line isolated from neonatal foreskin	Lonza (CC-2509)
Lenti-X HEK293T^{TM}	subclone of HEK293 (human embryonic kidney cell line) that constitutively expresses SV40 Large T antigen	Clontech (632180)
Bj5-ta fibroblasts	human foreskin fibroblasts, hTERT-immortalized	ATCC (CRL-4001)

2T2 hybridoma cell line	fusion of murine immunized B-cells and myeloma cells, required for the production of a monoclonal sT antibody	Christopher Buck (NIH, MD, USA)

Table 12 Overview of cell culture media used for the different cell lines

Cell line	**Cell culture medium**	**Supplements**
nHDF Donor I+II	DMEM	10% FBS, 1% P/S
Lenti-X HEK293T^{TM}	DMEM	10% FBS, 1% P/S
Bj5-ta fibroblasts	4/5 DMEM, 1/5 M199	10% FBS, 0.01 mg/mL hygromycin B, 1 mM sodium pyruvate
2T2 hybridoma cell line	RPMI 1640, 1x HFCS	10% FBS, 1% P/S

Primary nHDFs were cultured for a maximum of 10 passages. All cells were passaged two to three times a week at a ratio of 1:5 or 1:10. For splitting of the cells, they were washed once in DPBS and attached from the surface using Trypsin/EDTA. After an incubation time at 37 °C of 5 min, the enzyme was inactivated with DMEM containing supplements and cells were subsequently splitted to the desired concentration. For long-term storage, cells that reached 80% of confluence were trypsinized and pelleted at 200 x g for 5 min. Cells were resuspended in FBS containing 10% of DMSO and were transferred into cryotubes. For cryopreservation of nHDFs, 10% of fresh DMEM were additionally added. To ensure gentle cooling of the cryotubes, they were stored in a Mr. Frosty™- freezing container (Thermo Fisher) at -80°C for at least 24 h. Cells were stored at -80°C for short-term storage or frozen in liquid nitrogen for long-term storage. For the thawing process of cells, the cryotubes were shortly warmed up in a water bath and either directly added to a flask with fresh medium (nHDFs) or washed once in medium and resuspended in fresh medium. 12 h later, the medium was replaced by fresh medium to remove cell debris.

3.8.2 Production of the 2T2 antibody

For the detection of MCPyV sT antigen via WB analysis, the 2T2 antibody was produced from 2T2 hybridoma cell lines, kindly provided by Christopher Buck (NIH, MD, USA). Hybridoma cell lines are a fusion of myeloma cells and B lymphocytes, thus combining their properties of unlimited cell growth and production of antibodies, respectively. The 2T2 hybridoma cell line is derived from murine B cells immunized with MCPyV sT antigen. Cells were maintained in complete RPMI medium supplemented with 1x HFCS which contains IL-6 for example. They were cultured in T-175 flasks at 37°C and constantly expanded. When they reached a cell number of about $1x10^8$, they were centrifuged at 500 x g for 5 min. The cells were suspended in 50 mL 1x Hybridoma-SFM to select for monoclonal antibodies. When the culture medium turned yellow, aliquots were prepared, supplemented with NaN_3 (1:1000) and stored for up to six months in the fridge. Freezing of the 2T2 hybridoma cell lines was performed as described in chapter 3.8.1 with a different freezing medium, composed to 50% of the conditioned supernatant, 40% of FBS and 10% of DMSO.

3.8.3 Lentivirus production

Stable overexpression of MCPyV T antigens was achieved using a third-generation lentiviral transduction system with LeGO vectors as described in Naldini et al. (1996). Each LeGO vector contains a fluorescent marker which is fused to a gene of interest via an internal ribosome entry site (IRES) element, ensuring that both the gene of interest and the fluorescent marker are expressed. A general schematic structure of the LeGO vector containing the genes of interest used in this work is depicted in figure 9A. All plasmids, including those used for lentivirus production, are listed in table 6. Genes encoding for the packaging proteins (gag/pol and rev) and the envelope protein (env) are expressed on individual plasmids and via co-transfection together with the LeGO vector of interest into Lenti-X HEK293T^{TM} producer cells, lentiviral particles can be produced in a safe and efficient manner. Briefly, cells were seeded at a cell number of $1.1x10^7$ per 15 cm dish and one day later transfected using polyethylenimine (PEI). For this, 22.5 µg of each transfer plasmid was mixed with the packaging plasmids (22.5 µg gag/pol, 11.25 µg rev and 4.5 µg env) and diluted in 1 mL of OptiMEM. 338 µL PEI were added for 30 min and after vortexing, the mixture was added to the cells that were previously replenished with 15 mL of fresh medium (DMEM) without supplements. After 6 h, the medium

was changed to fresh DMEM supplemented with 1% FBS, 10% P/S, 20 mM HEPES buffer and 1 mM sodium pyruvate. After 48 and 72 h post transfection, the supernatants were harvested and passed through a filter with the size of 0.22 µm. In order to increase the concentration of lentiviral particles, the volume of the supernatants was reduced by precipitation with a viral concentrator solution consisting of PEG6000 and NaCl (see table 4). For this, the viral supernatants were mixed with the concentrator at a ratio of 4:1 and incubated overnight at 4°C, followed by a centrifugation at 1500 x g and 4°C for 45 min. The supernatant was discarded and the pellets containing the viral particles were resuspended in DMEM using 1/50 of the initial total volume. Aliquots of the virus solutions were stored at -80°C and titers were determined using fluorescence activated cell sorting (FACS) analysis as described in the next chapter.

3.8.4 Determination of viral titers

With the aim to achieve similar transduction rates in nHDFs, the viral particles were measured for their infectious potential via determination of viral titers. Due to the presence of a fluorescent marker on the transfer plasmids, the expression of mCherry or GFP was detected by FACS analysis which allows to determine how many particles are needed to infect a certain amount of cells. For this, $4x10^4$ nHDFs were seeded per well in a 24-well microtiter plate. On the next day, the medium was changed to DMEM containing 10% FBS and 1% P/S, as well as 8 µg/mL polybrene (i.e. hexadimethrine bromide), which is a small cationic agent that binds to cell surfaces and increases the uptake of viral particles. The infected cells were centrifuged at 1000 x g and 37°C for 30 min. After a medium change the day after transduction, cells were cultivated further and analyzed by FACS at 48 h post transduction. For this, the cells were detached from the 24-wells, washed in DPBS (500 x g, 5 min) and suspended in 200 µL of a 1% PFA solution diluted in DPBS. The FACS analysis was performed using an LSR Fortessa™ (BD Biosciences, Heidelberg) device. The percentage of mCherry-or GFP-positive cells was determined and the viral titers were calculated as follows:

$$T = \frac{N \cdot P}{V}$$

with:

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

N = number of cells

P = percentage of transduced cells [%]
V = volume of virus solution added [mL]

For the overexpression experiments conducted in nHDFs, transduction of the cells was performed at a multiplicity of infection (MOI) of one which means that the amount of viral particles equals the amount of cells. This mostly results in a transduction efficiency of $> 50\%$ with the advantage that most of the cells are infected with no more than one viral particle, resulting in similar expression levels in all cells. The MOI can be calculated as follows:

$$V_{MOI} = \frac{N \cdot P}{T}$$

with:
T = viral titer [ffu/µL]
N = number of cells
P = percentage of fluorescent cells [%]
V_{MOI}= Volume that needs to be added for a specific MOI [mL]

3.8.5 Lentiviral transduction and FACS-Sorting

Lentiviral transduction of nHDFs was conducted using the same protocol as for the determination of viral titers. Cells were seeded at appropriate numbers and one day later, viral particles were added at an MOI of one. After 48 h, the cells were detached and suspended in DPBS supplemented with 5% FBS and kept on ice for the subsequent FACS-sorting process. GFP-or mCherry-positive cells were collected with an Aria^{TM} Fusion (BD Biosciences, Heidelberg) device, which was conducted by the FACS Core Unit of the UKE. Subsequently, cells were pelleted at 500 x g for 5 min and further cultivated in complete DMEM until they were harvested for the following experiments.

3.9 SDS-PAGE and Western Blot analysis

Determination and comparison of protein abundance was assessed by WB analysis. For protein lysate preparation, approximately $5x10^5$ - $1x10^6$ cells were detached and pelleted (500 x g, 5 min, 4°C), washed once in cold PBS containing 1x protease inhibitor cocktail and 1x phosphatase inhibitor, and lysed in 200 µL RIPA buffer (see table 4) using a syringe (0.4 µm). Following a 20 min incubation time on ice, lysates were centrifuged at 20,000 x g and 4°C for 30 min and protein concentrations from the supernatants were photometrically quantified after Bradford (1976) using the Bio-Rad Protein Assay kit. Equal amounts ranging from 10 to 100 µg of protein were mixed with 4x SDS loading buffer (see table 4), heated at 95°C for 10 min and subsequently loaded on 10%, 12% or 15% polyacrylamide gels. The preparation of the polyacrylamide gels is depicted in table 13. SDS-PAGE was performed at 120-200 V for 45-80 min, depending on the gel concentration and expected protein sizes. For WB analysis, proteins were transferred from the gel to a PVDF membrane using a semi-dry blotting system. The composition of the buffers required for transfer is shown in table 4. The transfer was conducted at 400 mA for 40 min and the membrane was blocked in 5% BSA-TBST or 5% milk-PBST, depending on the antibody. Following a 20 min incubation time, the primary antibody solution was added to the membrane for an overnight incubation at 4°C. The membrane was washed three times for 10 min with PBST or TBST and a secondary antibody conjugated to a horse radish peroxidase (HRP) was added and incubated at room temperature for 1.5 h. Following another washing step with PBST or TBST (three times for 5 min), the membrane was moistened with an enhanced chemiluminescence substrate which reacts with the HRP and results in light emission. This signal was measured using an X-ray machine.

Table 13 Composition of polyacrylamide gels for SDS-PAGE

SDS Stacking gel	5% (v/v) acryl-bisaccharylamide
	125 mM Tris-HCl, pH 6.8
	0.1% (w/v) SDS
	0.1% (v/v) TEMED
	0.1% ammonium persulfate
SDS Running gel	10-15% (v/v) acryl-bisaccharylamide
	375 mM Tris-HCl, pH 6.8

	0.1% (w/v) SDS
	0.1% (v/v) TEMED
	0.1% ammonium persulfate

3.10 Immunofluorescence analysis

IF analysis allows to localize structures of interest inside a cell via direct staining using fluorescent dyes or indirect labeling of proteins using antibodies coupled to fluorophores. All primary and secondary antibodies used in this work for IF analysis are listed in table 5.

NHDFs that were supposed to be analyzed by IF were seeded on coverslips that were coated with a thin layer of gelatin solution. Approximately $2x10^4$ cells were seeded per coverslip and incubated at 37°C. Two days later the cells were fixed using 4% PFA, which was incubated for 20-30 min at room temperature. The coverslips were washed three times using PBS and incubated for 10 min with permeabilization buffer. Following another washing step, a blocking buffer was added for another 30 min. The composition of the buffers is summarized in table 4. Subsequently, a primary antibody solution diluted in blocking buffer was incubated for 1-2 h in a humid chamber to prevent desiccation. The coverslips were washed three times in PBS and incubated with the secondary antibody (diluted in blocking buffer) for another hour in a humid chamber protected from light. The dilutions used for the respective antibodies are depicted in table 5. After a final washing step with PBS, the coverslips were mounted on glass slides using DAPI-containing Vectashield. They were briefly air-dried and stored at 4°C overnight until microscopic analysis was conducted with a confocal laser scanning microscope (Leica TCS SP5) using a 63x oil immersion objective. For the quantification of the Pearson correlation coefficient, a minimum of 10 cells per replicate was analyzed using the Volocity Demo software.

3.11 Analysis of cell proliferation

To assess the proliferative behavior of nHDFs, the Incucyte© Live-Cell Analysis System was applied. With this system, cells are cultivated under normal cell culture conditions and real-time measurements can be conducted for a desired amount of time. For the proliferation analysis

performed with nHDFs, 1000 cells per condition were seeded in triplicates in 96-well plates and the percentage of confluence was recorded by measuring the occupied area over 120 h, including 61 time points in total.

3.12 Native Chromatin Immunoprecipitation

Native chromatin immunoprecipitation (nChIP) was conducted as described by Brind'Amour et al. (2015) with minor modifications. As standard ChIP protocols use crosslinked chromatin that is fragmented by a mechanical shearing process, high cell numbers are often a pre-requisite. In contrast, the protocol from Brind'Amour et al. (2015) is based on native chromatin and an MNase enzyme which unspecifically cuts the chromatin in histone-free regions and was reported to work also for lower cell numbers in the range of $1x10^4$-$1x10^6$. However, it requires a stable interaction and is thus suited for histone modifications, rather than transcription factors. Briefly, cells were counted and adjusted to a specific cell number. In general, $2x10^5$ cells were required per IP and 1/4, i.e. $5x10^4$ cells were additionally included as an input control. The number of cells required for an experiment (mostly $4.5x10^5$) were pelleted by centrifugation at 300 x g for 5 min and the pellet was washed in ice-cold PBS supplemented with 1x protease inhibitor cocktail and 1 mM Pefabloc© . Pellets were either flash-frozen in liquid nitrogen and stored at -80°C or directly processed as described in the following.

In a first step, cells were resuspended in 50 µL nuclear isolation buffer supplemented with 1x protease inhibitor and 1 mM Pefablock©, pipetted 20-30 times to release all nuclei and placed back on ice. The composition of the nuclear isolation buffer is summarized in table 14.

Table 14 Composition of the nChIP nuclear isolation buffer

Component	Concentration
Hepes-KOH	50 mM
NaCl	140 mM
glycerol	10%
NP40	0.50%
Triton X-100	0.25%

In the next step, an MNase digestion master mix was prepared as shown in table 15. 10 µl of MNase master mix were added to each sample by pipetting up and down for 20-30 times and they were incubated at 37°C for 12 min.

Table 15 Reaction mix for the nChIP MNase digestion

MNase master mix / reaction	**Volume**
10x MNase buffer	6 µL
0.1 M DTT	0.88 µL
H_2O	2.52 µL
MNase 1:10	0.6 µL

The reaction was stopped by adding $1/10^{th}$ of the total volume EDTA (100 mM) to the sample and mixing 20-30 times with a pipette. Subsequently, $1/10^{th}$ of the reaction volume of a solution of 1% Triton-X-100 and 1% deoxycholate were added and the samples were rested for 15 min on ice. After vortexing the samples for 30 s (medium setting) complete immunoprecipitation buffer containing 1x protease inhibitor and 1 mM Pefablock© was added to the digested chromatin at a total volume of 200 µL per IP and 50 µL per input. The chromatin was rotated at 4°C for 1 h and following another vortexing step for 30 s (medium setting), 50 µL were taken out and stored at -20°C as an input control. The rest of the chromatin was processed for immunoprecipitation as described in the following. An overview of the buffers used for the immunoprecipitation and washing steps is given in table 16. For the analysis of H3K27ac, 5 mM sodium butyrate were added to the complete IP buffer as an inhibitor of histone deacetylases.

Table 16 Composition of the buffers used for nChIP

Component	**Concentration**
Complete immunoprecipitation buffer	
Tris-HCl, pH 8	20 mM
EDTA	2 mM
NaCl	150 mM
Triton-X 100	0.10%

Low salt washing buffer	
Tris-HCl, pH 8	20 mM
EDTA	2 mM
NaCl	150 mM
Triton-X 100	1% (v/v)
SDS	0.1% (w/v)
High salt washing buffer	
Tris-HCl, pH 8	20 mM
EDTA	2 mM
NaCl	500 mM
Triton-X 100	1% (v/v)
SDS	0.1% (w/v)

During the incubation of the chromatin at 4°C, the ChIP protein A/G magnetic beads were washed three times in complete immunoprecipitation buffer containing supplements. In general, washing was performed by placing the tubes on a magnetic rack and after approximately 3 min, the supernatant was removed and the beads were suspended in fresh buffer. In a first step, the chromatin was pre-cleared using 10 µL per IP, i.e. per 200 µL. For this, the samples were rotated at 4°C for 5 h. In parallel, the desired antibodies were pre-incubated with 10 µL washed beads and rotated at 4°C for 5 h. The concentration of the different antibodies used is shown in table 5. After incubation, the antibody-bead complexes were placed on a magnetic rack and the supernatant was discarded. The pre-cleared chromatin was transferred to the antibody-bead complexes and the samples were rotated overnight at 4°C.

Following the overnight incubation, several washing steps were performed with the buffers listed in table 16 which were supplemented with 1x protease inhibitor and 1 mM Pefablock©. First, the beads were washed two times in low-salt buffer and two times in high-salt buffer. For the last washing step, the beads were transferred into a new tube and after pulse-spinning the remaining buffer was completely removed. The beads were resuspended in 30 µL freshly prepared elution buffer, consisting of 50 mM Tris-HCl (pH 8), 10 mM EDTA and 1% (w/v) SDS. The DNA was eluted during an incubation at 65°C for 1-1.5 h, vortexing the solution

every 15 min approximately. The samples were pulse-spinned and placed on a magnetic rack. The supernatant containing the eluted DNA was transferred into a 5PRIME phase lock tube (Quantabio), the remaining beads were washed once with 70 µL 10 mM Tris-HCl (pH 8.5) and the supernatant was added to the phase-lock tubes. In the meantime, the input samples were thawed on ice, mixed with 10% SDS and 45 µL 10 mM Tris-HCl (pH 8.5) to adjust the total volume to 100 µL. Together with the immunoprecipitated samples, they were processed for DNA extraction using phenol-chloroform extraction. First, 100 µL Phenol/Chloroform/Isoamyl (PCI) alcohol (25:24:1, Roth) were added and mixed with the samples by vortexing for 30 s and they were centrifuged at 21,000 x g and 4°C for 30 min. This step was repeated once with PCI and then with chloroform. After another centrifugation, the aqueous phase containing DNA was carefully transferred to a new tube and the DNA was precipitated by adding 1055 µL cold ethanol (100%), 6 µL NaCl (5 M) and 3 µL glycogen (5 mg/mL) and the samples were left overnight at -80°C. The next day, the samples were centrifuged at 21,000 x g and 4°C for 30 min and the pellets were washed with 70% freshly prepared ethanol and centrifuged for another 15 min at 21,000 x g and 4°C. The supernatant was discarded, ensuring a complete removal of the ethanol, and the pellet was dried in a thermoblock at 37°C, and was finally resuspended in nuclease-free water. 10 µL of the input samples were analyzed on a 1% agarose gel to control for efficient chromatin fragmentation. An ideal MNase digestion resulted in a size distribution of 100-300 bp, representing the size of single or multiple nucleosomes. As a further quality control, 1.5 µL per sample were analyzed by qPCR to assess the enrichment using control primers, as described in chapter 3.15.1.

3.13 RNA extraction from eukaryotic cells

RNA that was supposed to be used for RNA-Seq analysis was extracted using the RNeasy mini kit (Machery-Nagel, Düren) according to the manufacturer's protocol. For the RNA-Seq experiments, in general around $5x10^5$ cells were harvested, washed once in ice-cold DPBS (500 x g, 4°C, 5 min) and immediately transferred on dry ice and stored at -80°C or directly processed for RNA extraction.

For the quantification of mRNA levels via RT-qPCR, RNA was extracted using TRIzol™ reagent. TRIzol™ is a monophasic solution consisting of guanidine isothiocyanate and phenol. When adding chloroform, different phases separate containing RNA, DNA and protein and the

RNA can be precipitated from the aqueous phase using isopropanol. Briefly, pellets that were harvested from approximately 5x10^5 cells were kept on ice and suspended in 200 µL TRIzol™. Samples were either stored at -80°C or directly processed as described in the following. After adding 100 µL chloroform, the samples were vortexed for 20 s and incubated on ice for 10 min. Subsequently, they were centrifuged at 12,000 x g and 4°C for 15 min. The aqueous solution was carefully transferred into a new tube, 250 µL isopropanol were added and the samples were incubated at room temperature for 10 min. Following a centrifugation step at 21,000 x g and 4°C for 5 min the pellets were washed three times with 75% freshly prepared ethanol. To remove any residual ethanol after the washing procedure, the tubes were inverted and dried at room temperature. The pellets were resuspended in RNase-free water, incubated for 5 min on ice and resuspended by pipetting.

The concentration of the isolated RNA was measured using a Qubit fluorometer (Life Technologies, Carlsbad, CA, USA). The RNA extracted with TRIzol™ was additionally assessed for its purity using a NanoDrop 2000 device (PeqLab, Erlangen).

3.14 cDNA synthesis from isolated RNA

The amount of total mRNA present in a certain sample gives information on the expression levels of a gene of interest which can be measured by reverse transcription-quantitative PCR (RT-qPCR). Reverse transcription allows the conversion of mRNA to cDNA which can be amplified by specific primers directed against the genes of interest. For this, cDNA was transcribed from the extracted RNA by using random hexamer-oligonucleotides (dN6) primers and Superscript™ IV as a reverse transcriptase. Prior to this reaction, the RNA was treated with DNase I in order to remove residual DNA from the samples. For this, 1.6 µg RNA were incubated with 2 µL 10x DNaseI buffer and 3 µL DNase I enzyme at 37°C for 30 min. The reaction was stopped by adding 3 µL DNase inactivation reagent, vortexing and incubating the samples at room temperature for 2 min. After a centrifugation step at 10,000 x g for 90 s, 11.4 µL of the supernatants were transferred into PCR reaction tubes that had been filled with 0.6 µL dN6 primer (50 µM) and 1 µL dNTPs (10 mM). After a pre-incubation step at 65°C for 5 min, the samples were cooled on ice before 6 µL of the reaction master mix were added containing the following components: 4.5 µL 5x SSIV buffer, 1 µL DTT (0.1 M) and 1 µL murine RNase-inhibitor. Finally, 0.5 µL SuperScript© IV was added to the tubes (200 U/µL) except for one sample that

served as a negative control which did not contain the reverse transcriptase (-RT control). The samples were incubated in a thermocycler using the following program: 10 min at 23°C, 30 min at 55°C and 15 min at 80°C. The cDNA was diluted 1:5 with RNase-free water and stored at -80°C.

3.15 Quantitative realtime PCR (qPCR)

In general, qPCRs were conducted using either cDNA or dsDNA as a template which includes the DNA obtained from the ChIP protocol. In the following, the conditions used specifically for the quality control of the ChIP-DNA (ChIP-qPCR) and for quantification of mRNA levels are described separately.

3.15.1 ChIP-qPCR

An overview of the general pipetting scheme for a qPCR reaction is shown in table 17 and the general cycling conditions in table 18.

Table 17 Pipetting scheme for qPCR reactions

Component	**Volume**
RT-qPCR mixture - SYBR Green	
Maxima SYBR Green/Rox qPCR Master Mix 2x	5 µL
Fw primer (10 mM)	0.3 µL
Rev primer (10 mM)	0.3 µL
H_2O	2.9 µL
DNA	1.5 µL
Total volume	10 µL

Table 18 Cycling conditions for qPCR reactions

Cycling conditions	Temperature	Time
SYBR Green		
Hold	95°C	10 min
Cycling 50x	95°C	15 s
	57°C	15 s
	72°C	15 s
Melt	70-95°C	5 s for each temperature

3.15.2 Relative quantification of mRNA levels

RT-qPCR was used to quantify relative mRNA levels using cDNA that was transcribed from cellular total RNA. Table 19 depicts the pipetting schemes for the RT-qPCR reaction mixtures. Most genes were measured using SYBR Green, except for GAPDH which was quantified using a Taqman™ assay. SYBR Green is a fluorescent dye which emits fluorescence upon binding to dsDNA, thus its signal quantity is directly proportional to the DNA amplified in the PCR reaction. The Taqman™ probes rely on a similar principle with the difference that the fluorophore is initially quenched and upon DNA strand synthesis the probe is displaced and the quencher is cleaved so that the fluorophore is released and fluorescence can be measured. In general, primers for RT-qPCR are designed such that they cover exon/exon-splice regions to avoid amplification of genomic DNA. Nevertheless, the -RT control is used as an additional control because it should contain only RNA and should not be amplified. In every run and for every primer pair, a standard curve was included to measure the relative amounts of the gene of interest which was normalized to the expression of two house-keeping genes (*HPRT* and *GAPDH*). The RT-qPCR reactions were conducted using a Rotor-Gene Q Realtime Cycler (Qiagen). The different cycling conditions are shown in table 20.

Table 19 Pipetting scheme for RT-qPCR reactions

Component	**Volume**
RT-qPCR mixture - SYBR Green	
Maxima SYBR Green/Rox qPCR Master Mix 2x	5 µL
Fw primer (10 mM)	0.5 µL (1 µL LT*tr*)
Rev primer (10 mM)	0.5 µL (1 µL LT*tr*)
H_2O	3 µL (2 µL LT*tr*)
cDNA	1 µL
Total volume	10 µL
RT-qPCR mixture - Taqman™ (multiplexing)	
Rotorgene Master Mix 2x	5 µL
Probe-Primer Mix 20x	0.5 µL
Probe-Primer Mix 20x	0.5 µL
H_2O	3 µL
cDNA	1 µL
Total volume	10 µL

Table 20 Cycling conditions for RT-qPCR reactions

Cycling conditions	**Temperature**	**Time**
SYBR Green		
Hold	95°C	10 min
Cycling 50x	95°C	10 s (30 s LT*tr*)
	60°C (53°C LT*tr*)	15 s (30 s LT*tr*)
	72°C	20 s (30 s LT*tr*)
Melt	60-95°C	5 for each temperature
Taqman™	**Temperature**	**Time**
Hold	95°C	10 min
Cycling 50x	95°C	15 s
	60°C	15 s

3.16 High-Throughput Sequencing

Preparation of the libraries for RNA-Seq and ChIP-Seq analysis and the whole genome sequencing were kindly conducted by members of the HPI Sequencing Facility using the Illumina NextSeq© 500 platform.

3.16.1 RNA-Sequencing

RNA that was extracted as described in section 3.13 was examined on a Bioanalyzer with the RNA Nano kit (Agilent). When the samples passed an RNA Integrity Number (RIN) of 0.7 they were enriched using the NEBNext Poly(A) Magnetic Isolation Module (New England Biolabs). This step allows to enrich for poly-adenylated (poly(A)) RNA which is a characteristic of mRNA and thus reduces the amount of different RNA types such as ribosomal or tranfer RNAs (rRNA/tRNA) that were concomitantly isolated by RNA extraction. Subsequently, a library was prepared with the NEXTflex^{TM} Rapid Directional qRNA-Seq Kit (Bioo Scientific) as described in the manufacturer's protocol. The quality of the library was examined on a Bioanalyzer with the High Sensitivity DNA Kit (Agilent) and samples with a read length of 75 bp were sequenced with a depth of 20-30 mio reads.

3.16.2 ChIP-Sequencing

DNA enriched by ChIP as described in chapter 3.12 was sequenced using an intended sequencing depth of 20 mio reads for H3K4me3 and 40 mio reads for H3K27me3, H3K9me3 and H3K27ac. The numbers of unique reads for all experiments analyzed in this work are depicted in table S1. Prior to sequencing, the library was prepared using the NEXTFlex^{TM} ChIP-Seq Library Prep Kit (Bioo Scientific) according to the manufacturer's instructions. This protocol includes a size-selection step which enriches all fragment with a size of 100-500 bp. The quality of the library was assessed on a Bioanalyzer using the High Sensitivity DNA Kit (Agilent).

3.17 Bioinformatic analysis

Processing of the raw data obtained from RNA-Seq and ChIP-Seq analysis was kindly conducted by Dr. Jiabin Huang from the UKE (Hamburg). The programs and working steps are briefly

described in the following sections. Quality control of RNA-Seq and ChIP-Seq reads was assessed with the FastQC tool (Andrews, 2010).

3.17.1 Analysis of RNA-sequencing

RNA-Seq reads were mapped to the human genome (hg19) using the STAR 2.6.0c tool (Dobin et al., 2013). Analysis of differentially expressed genes (DEGs) was conducted using the DESeq2 R package as described by Love et al. (2014). DEGs were considered significantly upregulated when the p adjusted (padj.) value was lower than 0.05 and the log2 fold change (log2FC) > 1, and significantly downregulated with a padj. < 0.05 and a log2FC < -1. Volcano plots and heatmaps were generated using R Studio. The gene ontology (GO) analysis was performed with the DAVID tool (Huang et al., 2009) using significantly up- or downregulated genes as input.

3.17.2 Analysis of ChIP-sequencing

ChIP-Seq reads were mapped to hg19 using the BWA tool (Li & Durbin, 2009). Duplicate removal was performed with picard (BroadInstitute, 2018). Narrow peaks such as H3K4me3 and H3K27ac were called with MACS2 (Zhang et al., 2008), while broader peaks, i.e. H3K9me3 and H3K27me3 were called with SICER (Xu et al., 2014). Identification of regions in the human genome with differential histone modification signals between different conditions was conducted with diffReps (Shen et al., 2013). The window sizes for H3K4me3 and H3K27ac were set to 1000 (100 steps) and for H3K9me3 and H3K27me3 to 10,000 (1000 steps). Subsequent analysis and visualization were performed with R Studio and EaSeq (Lerdrup et al., 2016).

3.18 Statistical analysis

Statistical analyses were conducted using GraphPad Prism and for all experiments described in this work that allowed statistical evaluation, two-way ANOVA tests were conducted.

4 Results

The relatively small genome of MCPyV harbors an early gene region with only few ORFs, classified as T antigens. While the functions of the ALTO and 57kT antigens are not completely understood, the LT and sT antigens have been widely studied in terms of their oncogenic potential and their role in MCPyV pathogenesis. An important feature of sT and LT is their ability to interfere with cellular signaling pathways which is described in more detail in the introduction (chapter 1.2.4). Recent studies have revealed that especially sT has the potential to interfere with co-transcription and chromatin remodeling factors thereby contributing to tumorigenesis (Cheng et al., 2017; Park et al., 2019; Park et al., 2020) . It is therefore possible that sT has additional cellular targets that are so far unknown. Furthermore, it is perceivable that in addition to or together with sT, other T antigens expressed in the life cycle might interfere with gene regulatory processes. Therefore, the aim of this study was to systematically unravel transcriptional and epigenetic changes induced by three important T antigens expressed in the life cycle of MCPyV - sT, LT and LT*tr*.

To date, there is no proven evidence of the cell type(s) allowing productive infection or persistence. The finding that HDFs allow productive infection (Liu et al., 2016) has not been independently confirmed. Consistently, infection of the nHDFs used in this work resulted in approximately 5% of positive cells and there was no indication of re-infection, but rather of persistence establishment (supplementary figure S1). The lack of an efficient infection model impairs genome-wide analysis of bulk cell populations and therefore an alternative experimental setup was chosen. Overexpression of sT and LT together in nHDFs aimed at mimicking an infection scenario as both T antigens are known to be expressed during this phase of the life cycle. In addition, both sT and LT were expressed individually with the aim to dissect their respective contributions during productive infection or persistence establishment.

The second question that was addressed in this work was referred to the role of the T antigens in MCPyV-associated tumorigenesis. Similar to the lack of a suitable infection model, there is no tumorigenesis model available, which is in part due to the lack of knowledge about the cell of origin of viral-induced MCC. Hypotheses about possible cell types are discussed in more detail in the introduction (chapter 1.2). Because MCCs are of dermal or epidermal origin, there is a general consensus in the field that the transformation of a specific cell type in the skin precedes the formation of a MCC. Interestingly, rodent fibroblasts were shown to be transformed *in vitro*

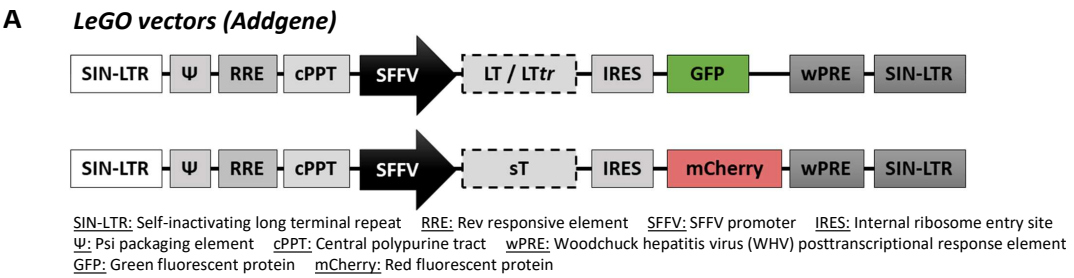
by overexpressing sT (Shuda et al., 2011). Further *in vivo* mouse models showed that sT is the main driver of tumorigenesis (Verhaegen et al., 2017; Verhaegen et al., 2015). These observations led to the hypothesis that also HDFs might be transformed by sT expression. However, soft agar assays performed with the nHDFs used in this work did not result in anchorage- or serum-independent growth when overexpressing sT alone or together with LT*tr* (unpublished observations). Despite the lack of a transformation model that is accompanied by the missing knowledge about the cell of origin for transformation, MCCs are characterized by the expression of sT and LT*tr* which were both shown to be essential for their survival (Houben et al., 2010; Shuda et al., 2014). Therefore, overexpression of sT together with LT*tr* in nHDFs reflected the T antigen expression scenario that characterizes a MCPyV-positive MCC. Expressing sT or LT*tr* individually aimed at dissecting their possible contributions to tumorigenesis.

4.1 Experimental setup to study the effects of MCPyV T antigen expression

Overexpression of proteins in primary cells such as nHDFs is often challenging due to low transfection efficiencies. Furthermore, transfection is only suitable for performing short-term analysis because the proteins of interest are expressed for no longer than three to four days after transfection. Therefore in this work, a third-generation lentiviral transduction system was applied which has several advantages over transfection methods. First, transduction of nHDFs achieves a much higher efficiency. By using a lentiviral system, the integration of the genes of interest into the host genome allows to generate stable cell lines constantly expressing the proteins of interest. Furthermore, via adjusting the MOI it can be controlled how many particles will approximately infect each individual cell, allowing a better comparison between the different conditions.

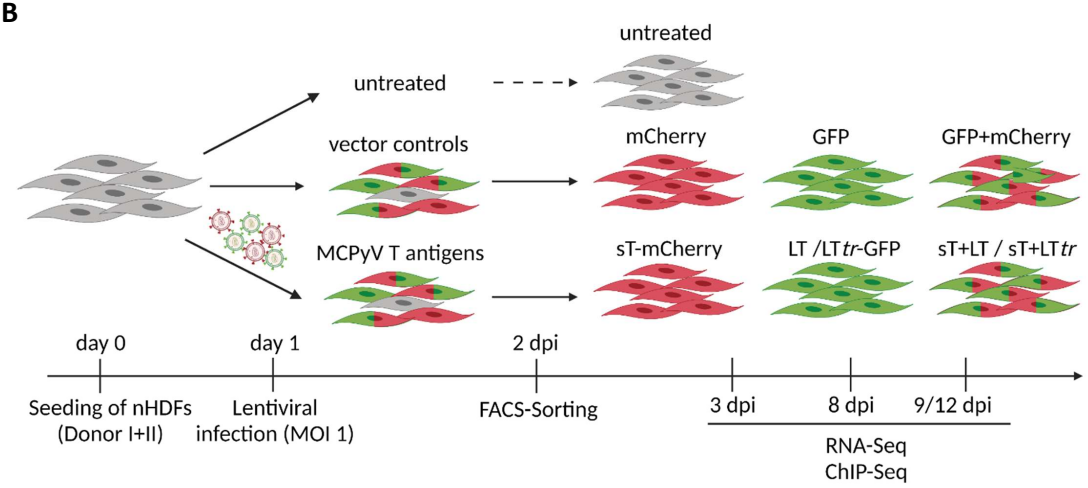
MCPyV T antigens were overexpressed in nHDFs from two different donors using lentiviral plasmids containing the MCPyV T antigens and either GFP (LT/LT*tr*) or mCherry (sT), as shown in figure 9A. Empty vectors containing only the fluorescent markers served as controls, referred to as vector controls (ctrls). Untreated cells that were passaged and harvested in parallel, were included to control for off-target effects induced by transduction or FACS-sorting. Figure 9B gives an overview of the experimental setup and workflow.

A



SIN-LTR: Self-inactivating long terminal repeat RRE: Rev responsive element SFFV: SFFV promoter IRES: Internal ribosome entry site
Ψ: Psi packaging element cPPT: Central polypurine tract wPRE: Woodchuck hepatitis virus (WHV) posttranscriptional response element
GFP: Green fluorescent protein mCherry: Red fluorescent protein

B



C

Condition	RNA-Seq	ChIP-Seq			
		H3K4me3	H3K27me3	H3K9me3	H3K27ac
untreated	+	+	+		
vector control 3 dpi	+	+	+		
vector control 8 dpi	+			+	+
vector control 9/12 dpi	+	+	+		
sT 3 dpi	+	+	+		
LT 3 dpi	+				
sT+LT 3 dpi	+	+			
LT*tr* 3 dpi	+	+	+		
sT 8 dpi	+	+	+	+	+
LT 8 dpi	+	+	+		
LT*tr* 8 dpi	+	+	+		
sT+LT*tr* 9/12 dpi	+	+	+		

Figure 9 Experimental setup and workflow.
(A) Structure of LeGO-vectors obtained from addgene and inserted with MCPyV T antigens (B). NHDFs from two different donors were seeded and infected with lentiviral particles at an MOI of one. In addition to individual transductions using LeGO-sT-mCherry, LeGO-LT-GFP or LeGO-LT*tr*-GFP, LEGO-sT-mCherry was co-transduced with LeGO-LT-GFP or LeGO-LT*tr*-GFP. Transduction with LeGO-mCherry, LeGO-GFP or both served as vector controls. After 2 dpi, cells were FACS-sorted and further cultivated. Cells were harvested at 3, 8, 9 or 12 dpi for RNA-Seq or ChIP-Seq analysis. (C) Summary of all conditions and time points that were analyzed by RNA-Seq or ChIP-Seq.

Two days following lentiviral transduction, the cells were FACS-sorted to increase the percentage of transduced cells to nearly 100%. The cells were harvested at the indicated time points and the following experiments were performed: (1) RNA-Seq from total isolated RNA and (2) ChIP-Seq from DNA enriched with antibodies against different histone modifications. For most conditions, H3K4me3 was chosen as a general activating mark and H3K27me3 as a general polycomb repressive mark. For nHDFs overexpressing sT, two further histone modifications were included: H3K9me3 as a mark of constitutive repressive chromatin and H3K27ac, an activating mark that is often found in correlation with H3K4me3, but can also mark enhancer regions. Due to the fact that NGS-based approaches require high cell numbers and are associated with high costs, ChIP-Seq was not performed for all conditions, as summarized in figure 9C.

The time points of genome-wide analysis were carefully chosen with reference to the two scenarios in the life cycle of MCPyV that were simulated by overexpression of the T antigens - infection *vs* tumorigenesis. Because initial infection is associated with immediate transcription and expression of the early proteins sT and LT, it was important to analyze the transcriptional profile early after transduction, i.e. after 3 dpi. In contrast, in the tumor setting which was represented by overexpression of sT with LT*tr*, the focus was laid on the later time point because the time span between LT truncation, transformation and tumorigenesis is probably larger and transcriptional changes might not be detectable immediately after transduction. Importantly, nHDFs overexpressing sT, LT or LT*tr* were analyzed by RNA-Seq after both 3 and 8 dpi in order to distinguish between transcriptional changes induced as a direct response to T antigen expression, and changes that were only visible after a short adaptation time. Furthermore, this differentiation allowed to detect short-lived *vs* stable transcriptional changes.

To monitor T antigen expression in transduced nHDFs, RT-qPCR as well as WB analysis were performed. Figure 10 summarizes the mRNA and protein levels determined from cDNA or cell lysates that were extracted from nHDFs at the indicated time points. It needs to be noted that for the condition sT+LT*tr* there is only RT-qPCR data for Donor I and for sT+LT there is no WB due to a lack of sufficient cell material. The RT-qPCRs in figure 10A show that the relative levels of LT, LT*tr* or sT were similar in all the experiments with some variance observable between the two donors and the two time points. While co-overexpression of sT+LT resulted in similar levels of sT and LT, there was a strong reduction of both signals in Donor II compared to Donor I.

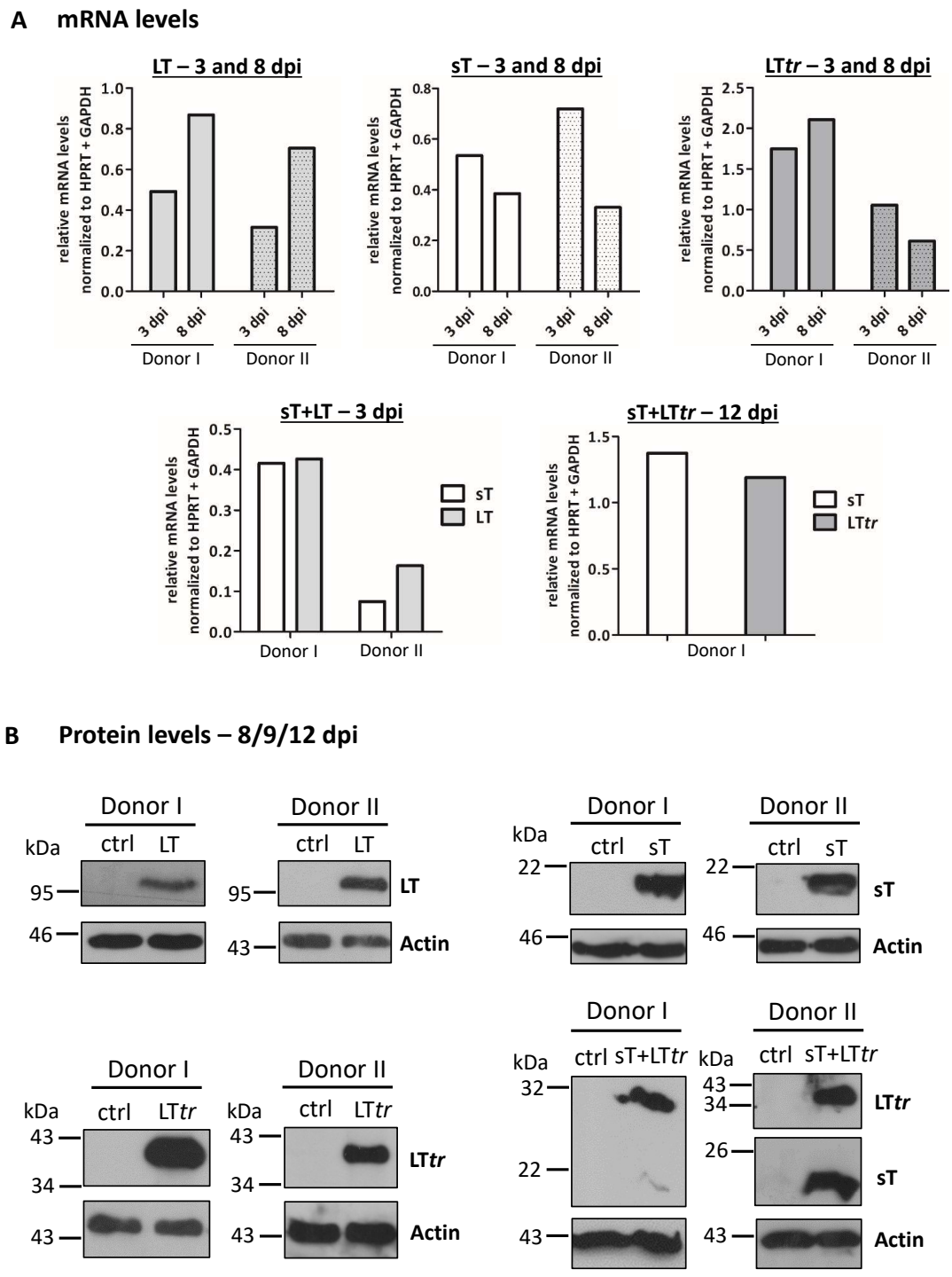


Figure 10 MCPyV T antigen mRNA and protein levels in transduced nHDFs.
(A) mRNA levels of either sT, LT or LT*tr* were measured in nHDFs transduced with lentivirus using RT-qPCR. Experiments were performed with material from the samples that were subsequently used for sequencing approaches and thus one experiment is shown from Donor I and one from Donor II. mRNA levels were determined with primers against *sT*, *LT* or *LTtr* and were normalized to *HPRT* and *GAPDH*. (B) Protein levels were determined by WB analysis using antibodies against sT and LT*tr* (2T2) or LT (CM2B4). Actin was used as a loading control.

The experiment with Donor I in which sT was co-expressed with LT*tr* also showed similar mRNA levels of sT and LT*tr*. Regarding the protein levels, similar quantities were observed for most conditions in both donors. Interestingly, while for Donor II the levels of LT*tr* and sT were quite similar when co-overexpressed, in Donor I the sT signal was very weak in comparison to the LT*tr* signal. It needs to be noted that the antibody that was used for detection of sT and LT*tr* (2T2) is very variable in its power of detection. In summary, in each experiment that was used for sequencing analysis, sufficient levels of either mRNA, protein or both were detected and mostly, the levels were observed to be similar as transduction with lentiviral particles was always performed at an MOI of one.

Before sequencing of immunoprecipitated DNA, qPCRs were performed as a quality control with primers located in the promoter regions of genes that are generally known to be marked by the respective histone modification. A summary of the qPCR results can be found in the supplements (figure S3), as well as information on the sequencing depths of all RNA-Seq and ChIP-Seq samples (table S1).

4.2 Profiling of nHDFs overexpressing MCPyV LT antigen

MCPyV LT is a gene product expressed early in the viral life cycle and is required for viral genome replication. As described in more detail in chapter 1.2.4, LT was shown to interfere with several cellular signaling pathways. Interestingly, viral genome replication initiated by LT was shown to induce senescence in nHDFs (Siebels et al., 2020) and an induction of genome instability by LT can also be ascribed to its involvement in the DDR which has not been completely unraveled to date (Li et al., 2013; Tsang et al., 2014). With the aim to identify the impacts of LT expression on host cellular gene regulation, independent of viral genome replication, RNA-Seq as well as ChIP-Seq experiments were performed in LT-overexpressing nHDFs.

4.2.1 Transcriptional changes induced by LT

NHDFs overexpressing MCPyV LT were assessed for their transcriptional profiles early after transduction, i.e. at 3 dpi and after one further passage at 8 dpi. Figure 11A gives an overview of all DEGs detected in nHDFs overexpressing LT compared to the vector control, using a log2FC threshold of 1 or -1 and a padj. significance level of 0.05. In each volcano plot, genes with the 10 highest log2FCs and lowest significance values were labeled for both up-and downregulated genes. In order to further classify DEGs in terms of biological function, a GO analysis was performed with the Database for Annotation, Visualization and Integrated Discovery (DAVID) tool which groups genes into different GO terms and computes their relative enrichment. For the GO analysis, all significantly up-or downregulated genes were used as an input, and the terms with the 10 highest gene counts (FDR < 0.05) are depicted in Figure 11B.

At 3 dpi, 985 genes were significantly upregulated and 338 significantly downregulated. A large subset of upregulated genes was found in the GO terms *DNA replication*, *Cell division* and *DNA repair* as well as *Regulation of transcription*. Overlapping genes within these GO terms included cell division cycle genes, DNA polymerase subunits, replication factor subunits and replication proteins such as *RPA2* and *RPA3*. In addition, minichromosome maintenance genes (*MCM2-8*) were found in many of the GO terms, as well as DNA repair genes including *BRCA1*. A substantial subset of upregulated genes was present in the GO terms *Type I IFN signaling pathway*, *Defense response to virus* and *(Innate) immune response*. These terms comprised a large amount of ISGs and IRFs such as *IRF1*, *IRF2*, *IRF7* and *IRF9*, as well as *IFNβ*. Finally, the GO term *Inflammatory response* contained 51 input genes including for example *CXCL10* or *CXCL11*.

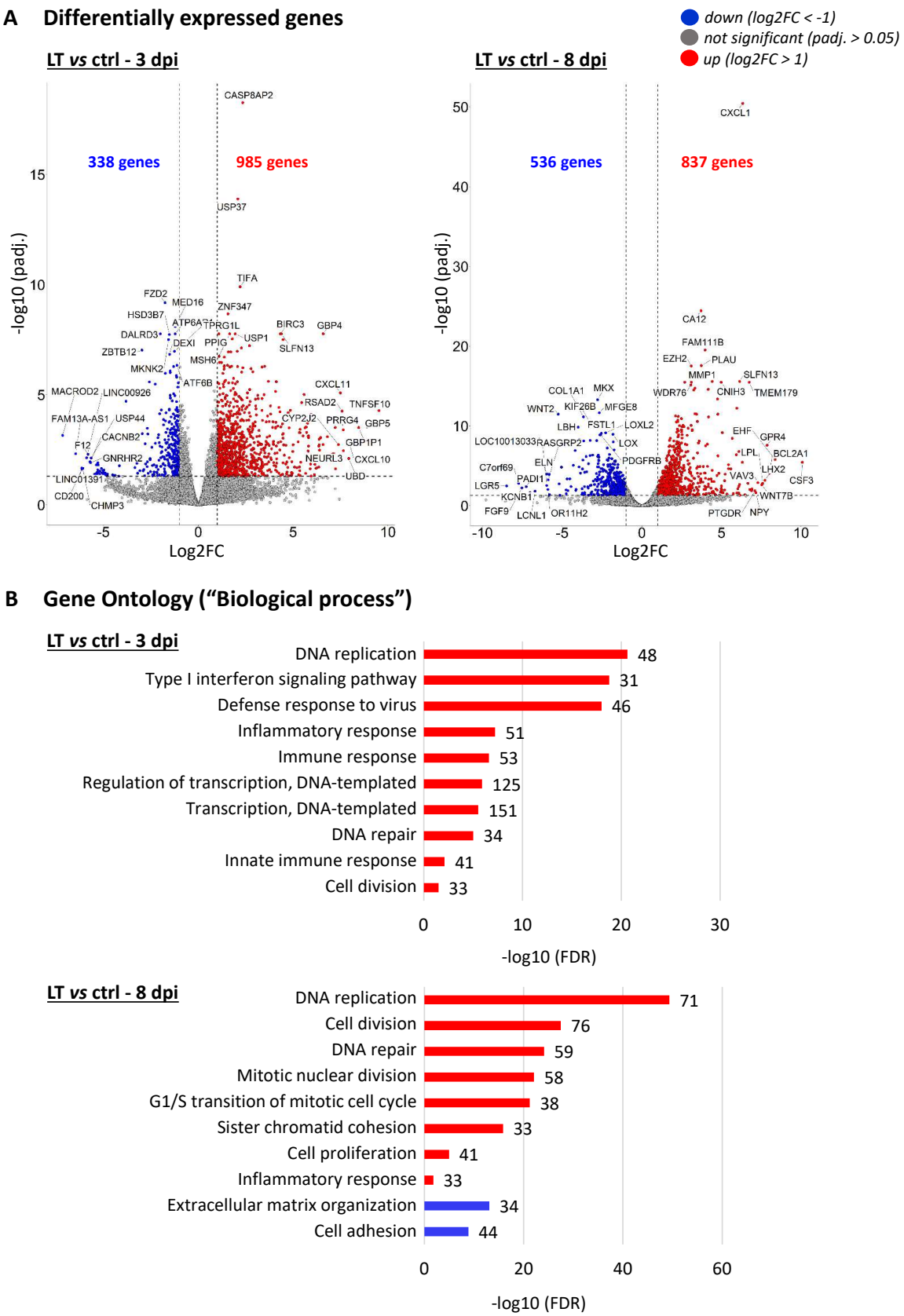


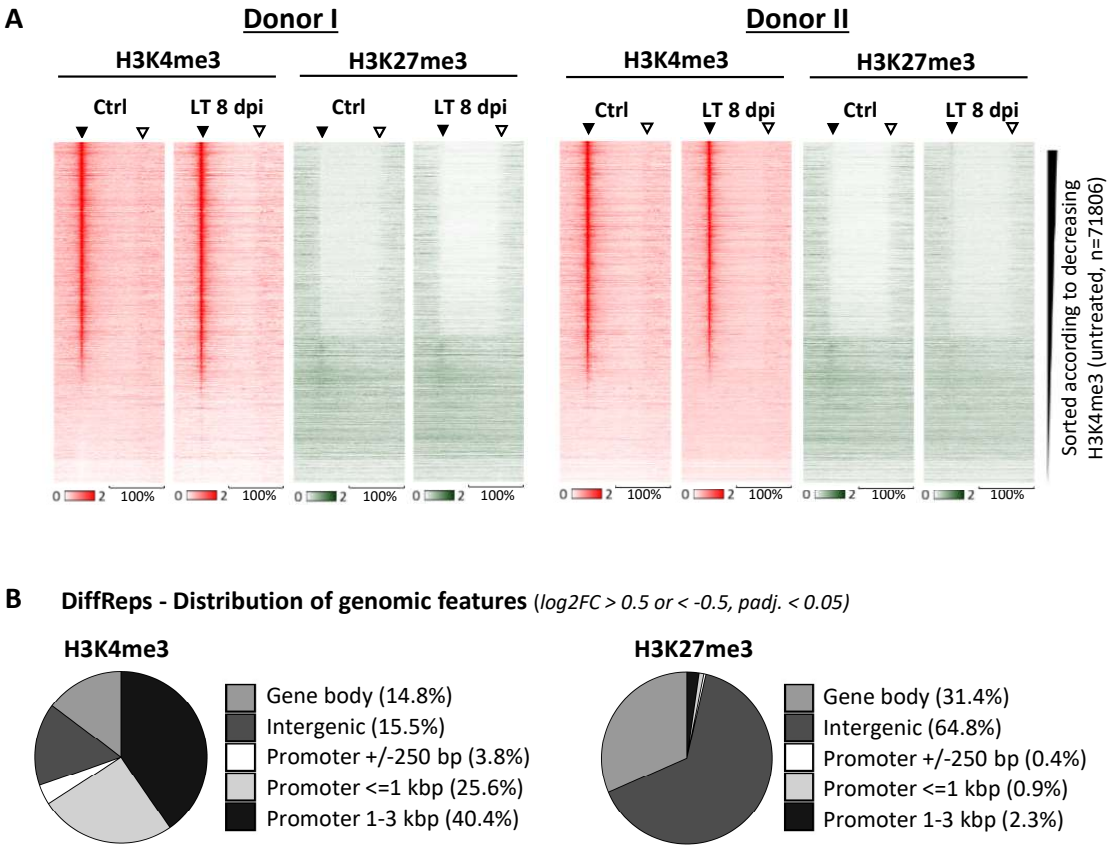
Figure 11 Transcriptional changes induced by LT overexpression in nHDFs.
(A) Volcano plots depicting all DEGs from the RNA-Seq analysis in LT-overexpressing nHDFs compared to the vector control at 3 and 8 dpi. Genes with a log2FC $>$ 1 or $<$ -1 and a padj. value $<$ 0.05 are denoted as significantly upregulated or downregulated, respectively. (B) A GO analysis ("Biological process") was performed with significantly upregulated and downregulated genes using the DAVID tool. The 10 GO terms (FDR $<$ 0.05, gene counts $>$ 10) containing the highest numbers of DEGs are depicted as red or blue bars, reflecting up- or downregulated genes as input, respectively. The number of DEGs found within each GO term is denoted next to the bars, FDR values are shown on the x-axis.

Most of the GO terms described above were also found at 8 dpi, considering the 837 genes that were upregulated. While *DNA replication*, *Cell division* and *Inflammatory response* were among the highest enriched GO terms, *Type I IFN response* genes were not significantly deregulated anymore. Furthermore, the 536 genes that were significantly downregulated at 8 dpi, were found in the GO terms *ECM organization* and *Cell adhesion*, including e.g. *collagen type I alpha 1 chain* (*COL1A1*) or *elastin* (*ELN*) which were also among the top 10 downregulated genes at 8 dpi. A summary of the most significant and prominent deregulated genes labeled in the volcano plots, including their full gene names and their affiliation to the GO terms from Figure 11B, can be found in the supplements (table S4 and S5).

Collectively, overexpression of LT in nHDFs led to a deregulation of a large amount of genes in both directions. Changes in basic regulatory processes such as transcription or replication were observed consistently at both time points. Interestingly, LT increased the transcription of genes involved in the innate and inflammatory response dominantly at 3 dpi and to a lesser extent at 8 dpi. Conversely, genes involved in ECM organization and cell adhesion were found downregulated only after 8 dpi.

4.2.2 Histone modification changes induced by LT

To investigate if LT-induced transcriptional changes were linked to changes in the composition of histone modifications, ChIP-Seq experiments were conducted. IPs were performed in nHDFs overexpressing LT at 8 dpi, using antibodies against H3K4me3 and H3K27me3. The heatmaps in figure 12A summarize the ChIP-Seq read count signals for H3K4me3 and H3K27me3 in the two different donors. For both ctrl and LT datasets, H3K4me3 and H3K27me3 showed a similar anti-correlative signal distribution with H3K4me3 mostly enriched around transcriptional start sites (TSSs) and H3K27me3 in gene bodies and flanking regions. It needs to be noted that in Donor II, there was some decrease in both H3K4me3 and H3K27me3 signal observable in LT-expressing cells compared to the vector control which needs to be considered when interpreting differential histone modification signals.



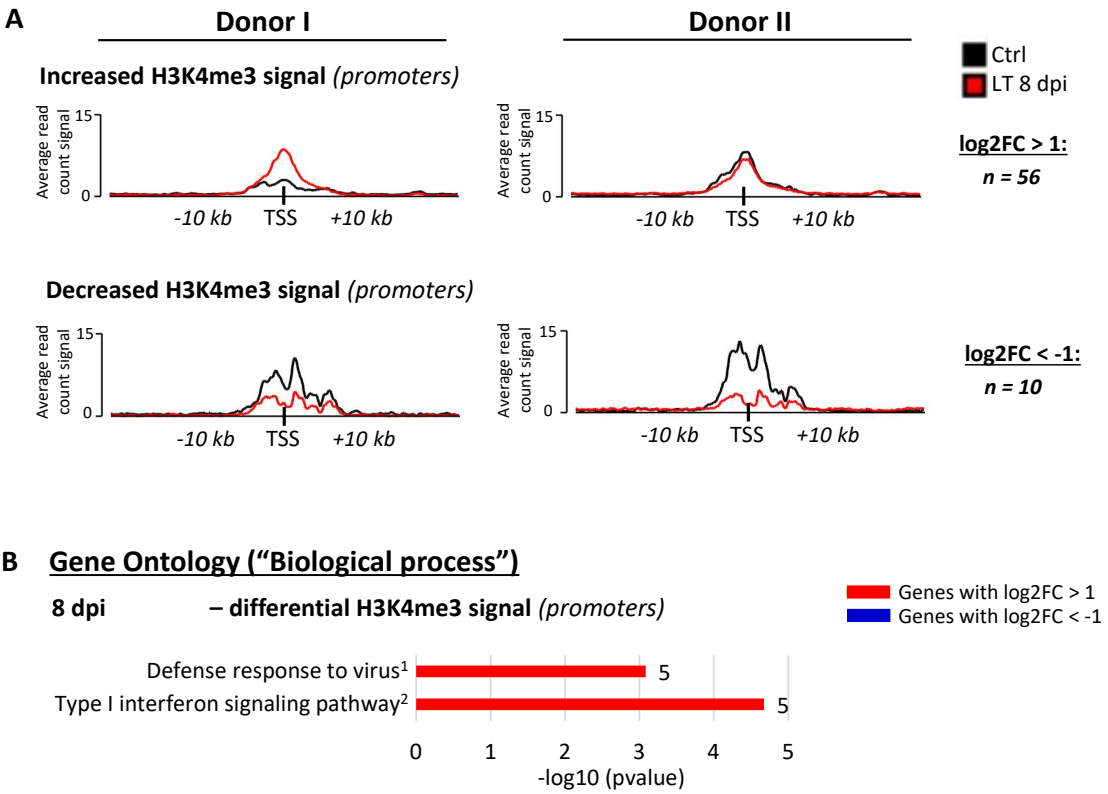
C **Number of diffReps hits**

diffReps	H3K4me3 total	H3K4me3 for annotated genes	H3K27me3 total	H3K27me3 for annotated genes
log2FC > 0.5	2197	1825	1227	**313**
log2FC < -0.5	694	495	143	**49**
log2FC > 1	100	**66**	7	0
log 2FC < -1	40	**26**	0	0

Figure 12 ChIP-Seq analysis in LT-overexpressing nHDFs.
(A) Heatmaps depicting ChIP-Seq read count signals quantified for all genes of hg19 for H3K4me3 and H3K27me3. Relative lengths of each genomic locus are shown for 200%, including TSSs (black triangles) and TTSs (non-filled triangles). Heatmaps are sorted according to decreasing H3K4me3 signal in untreated cells. Density represents DNA fragments per 1 mio reads per 1 kbp. (B) Differential H3K4me3 and H3K27me3 signals were analyzed using the diffReps tool and pie charts depict the proportion of genomic features of all detected sites with differential H3K4me3 or H3K27me3 signal, using padj. < 0.05 and log2FC > 1 or < -1.

To detect changes in histone modifications for specific genes, a diffReps analysis was conducted and the genomic distribution of diffReps hits is summarized in figure 12B. Nearly two third of all differential H3K4me3 reads were located in promoter regions, while the majority of differential H3K27me3 signal was present in intergenic regions. In addition, one third was observed within gene bodies and less than 5% in promoter regions. The total numbers of diffReps hits and the proportion of hits that were assigned to specific genes is depicted in figure 12C. An increase in H3K4me3 signal of at least two-fold between LT *vs* ctrl was observed for 66 genes and a decrease of at least two-fold only for 26 genes. In contrast, there were no genes with a change of at least two-fold in H3K27me3 signal. Only when lowering the log2FC threshold to 0.5 or -0.5, 313 and 49 annotatable diffReps hits were detected, respectively. Due to the fact that these changes were very marginal, they were not considered for the interpretation of the biological relevance. However, a summary of the H3K27me3 analysis can be found in the supplements (figure S8).

As there were several genes with a differential signal in H3K4me3 of at least two-fold, these were further analyzed and interpreted for their biological relevance. The average plots in figure 13A depict the differences in H3K4me3 signal comparing LT *vs* ctrl, plotting the average read count signal for +/- 10 kb around the TSSs for all genes with a significant change in H3K4me3 signal in promoter regions with a log2FC > 1 or < -1. These genes were more specifically analyzed by performing a GO analysis using the DAVID tool (figure 13B). To ensure comparability among all different datasets, a GO analysis was performed only when at least 50 genes showed a significant (padj. < 0.05) differential histone modification signal. All GO terms were filtered for a pvalue of at least 0.05 and a minimal gene count of 5. Applying these criteria, only two GO terms appeared: *Defense response to virus* and *Type I IFN signaling pathway*. In total, six overlapping genes were found enriched in these GO terms, e.g. *OAS* genes and further antiviral effector genes. A complete list of them and information from the diffReps analysis is also shown in figure 13B. It needs to be noted that, in opposition to the GO analysis from the RNA-Seq data, the GO analysis with the ChIP-Seq data did not yield any terms with significant FDR values. Therefore, the less stringent pvalue was chosen for the ChIP-Seq analysis, which applies to all the following data sets.



Symbol	Gene name	log2FC	padj.	Feature
GBP2[2]	Guanylate Binding Protein 2	1.16	0	Promoter1k
OAS1[1,2]	2'-5'-oligoadenylate synthetase 1	1.1	0	Promoter1k
OASL[1,2]	2'-5'-oligoadenylate synthetase like	1.28	0	ProximalPromoter
BST2[1,2]	bone marrow stromal cell antigen 2	2.67	0	Promoter1k
RSAD2[1,2]	radical S-adenosyl methionine domain containing 2	1.28	0	Promoter3k
HERC5[1]	HECT and RLD domain containing E3 ubiquitin protein ligase 5	1.26	0	Promoter1k

Figure 13 Analysis of changes in H3K4me3 induced by LT overexpression in nHDFs.
(A) Average read count signals of H3K4me3 for all genes with differential signal comparing LT *vs* ctrl at 8 dpi. DiffReps analysis was performed as described in figure 12 and the numbers of genes with differential H3K4me3 signal around promoters (+/-3 kbp) of annotated genes are shown next to the average plots. (B) GO analysis ("Biological process") was performed with DAVID using genes from the diffReps analysis described in (A). Only gene sets with n$>$ 50 were used as an input for the GO analysis. From all GO terms with at least 5 genes and a pvalue of $<$ 0.05, the terms with the 10 highest gene counts are shown. The table below lists all genes contained in the GO terms.

Intriguingly, the GO analysis resulted in GO terms that were also found in the RNA-Seq data. To bring these similarities together and detect further overlaps, the RNA-Seq data was correlated with the ChIP-Seq data which is shown in figure 14. First, all genes were quantified for their H3K4me3 and H3K27me3 signals and were plotted for LT on the y-axis *vs* the ctrl on the x-axis. The scatterplots on the right-hand side incorporate the information on the log2FCs from the RNA-Seq analysis, represented by the color code.

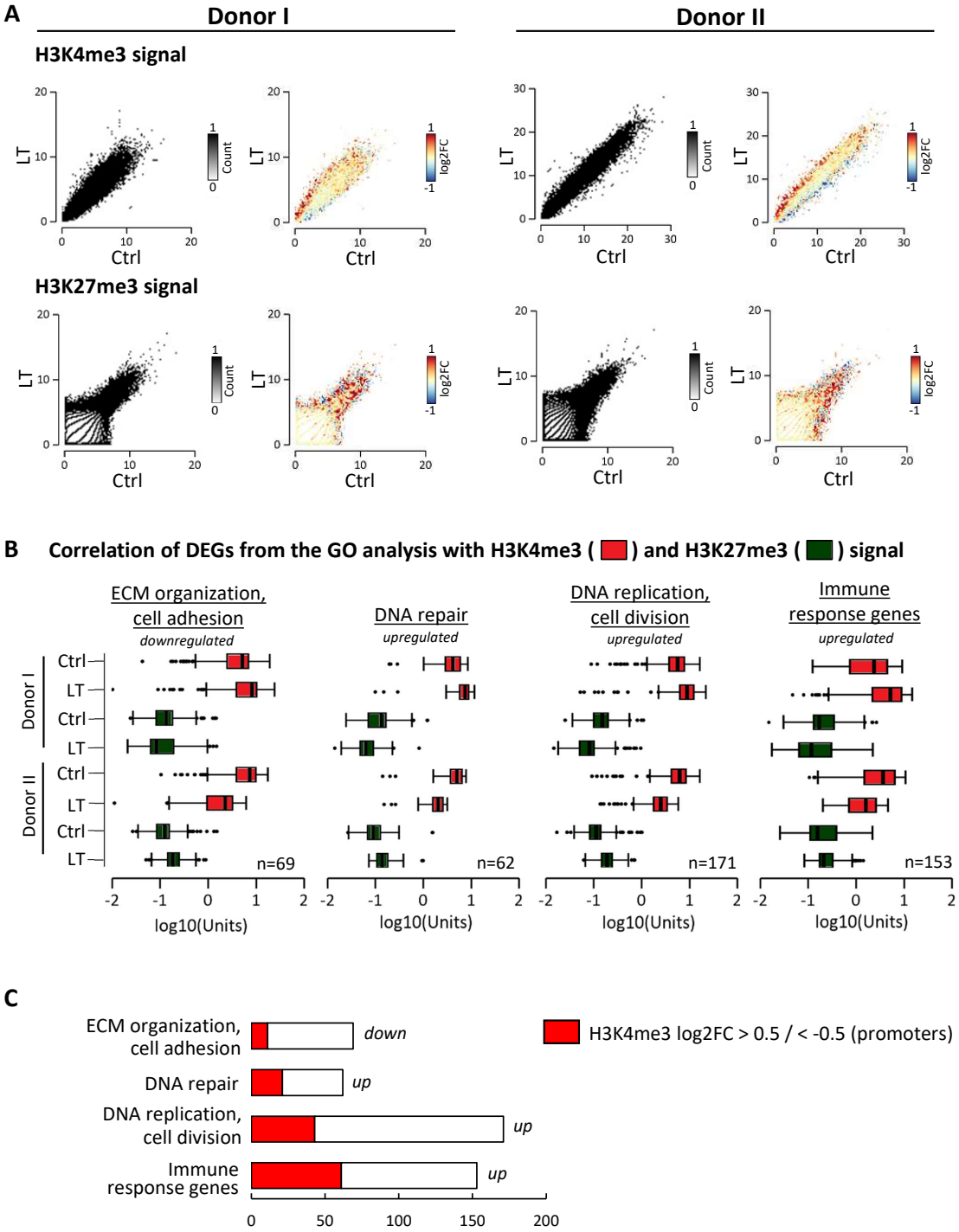


Figure 14 Correlation of ChIP-Seq with RNA-Seq in LT-overexpressing nHDFs.
(A) Quantification of H3K4me3 and H3K27me3 read signals for all genes from hg19 in nHDFs overexpressing ctrl or LT at 8 dpi. Histone modification signals were correlated with changes in gene expression in LT *vs* ctrl. The color code refers to the log2FC of each gene that is plotted by its level of histone modification signal (H3K4me3 and H3K27me3). The x and y-axes were segmented into 100 bins and regions within these bins are depicted by the counts. (B) Boxplots representing the distribution of H3K4me3 and H3K27me3 signals for selected subsets of genes from the RNA-Seq GO analysis (figure 11B). (C) Total numbers of genes from the subsets described above are shown in the bar plots, representing DEGs from the RNA-Seq analysis (white) and their proportion of genes with differential H3K4me3 promoter signal (red).

In both donors, H3K4me3 signal considerably correlated with transcriptional changes. For example, genes with an increase of H3K4me3 in LT-expressing nHDFs also had positive log2FC values, i.e. were upregulated on the transcriptional level. For H3K27me3, the opposite correlation was observable. However, in Donor I there were also many upregulated genes distributed all over the H3K27me3 scatterplot which indicates that a correlation of H3K27me3 with the RNA-Seq data was less consistent. In figure 14B, four gene subsets were selected from the RNA-Seq GO analysis (see figure 11B) and quantified for histone modification signals, summarized by the boxplots for H3K4me3 (red) and H3K27me3 (green). Although the patterns in the scatterplots look very similar in both donors, quantification of histone modification signals of specific gene sets revealed opposite patterns of histone modification signals in the two experiments. More precisely, while H3K4me3 was increased in LT *vs* ctrl for all four gene subsets in Donor I, it was consistently decreased in Donor II. At the same time, while H3K27me3 was decreased in Donor I, it was increased in Donor II. These differences do not allow a precise conclusion whether the transcriptional changes observed for the genes within the GO terms were generally associated with histone modifications or not.

Comparison of ChIP-Seq signals for single genes using the "integrative genomics viewer" (igv) genome browser revealed that for specific genes there was a substantial similarity between Donor I and II (figure 15), reflecting the diffReps result. As an example, *OAS2* and *complement component C3* (*C3*), both genes involved in the innate immune response, showed an increased H3K4me3 promoter signal in LT-expressing nHDFs. The igv tracks reflect this result for both donors around the *C3* promoter. However, the increase in H3K4me3 at the *OAS2* promoter was only visible in Donor I. In the experiment from Donor II there was also a much higher H3K4me3 signal in the *OAS2* promoter region in the ctrl condition as compared to Donor II. This also applied for the other examplary genes with an increased H3K4me3 signal shown in figure 15. In contrast, the two examples from the GO terms *ECM organization* and *Cell adhesion* - NUAK Family Kinase 1 (*NUAK1*) and *ELN* - had reduced track sizes in LT compared to ctrl in both donors.

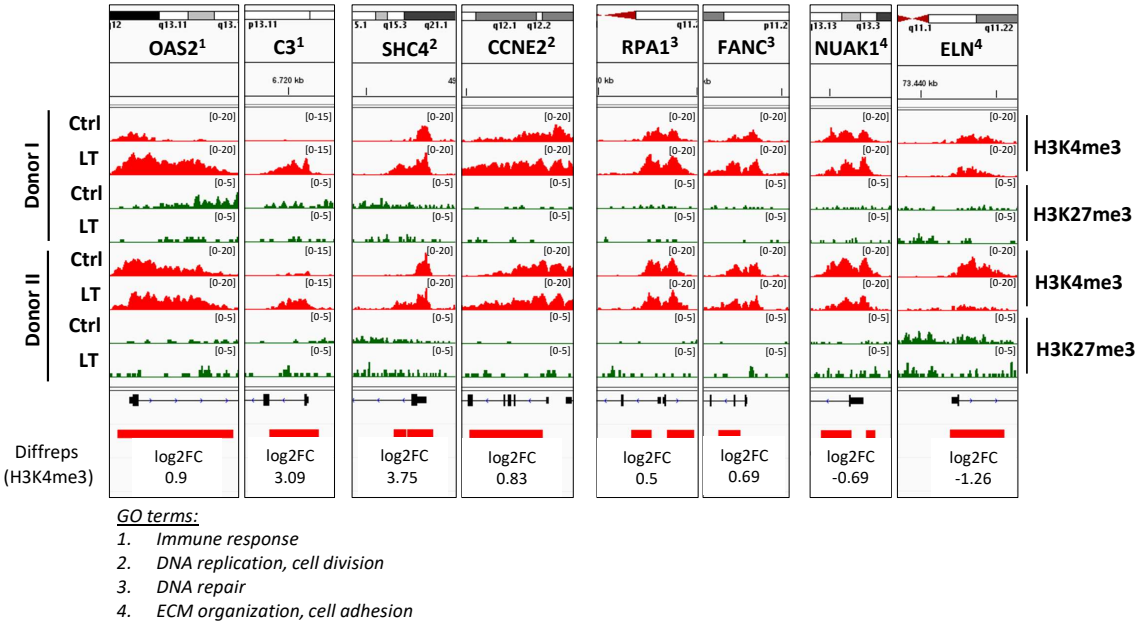


Figure 15 Genome browser tracks from the ChIP-Seq analysis of LT-expressing nHDFs.
Genome browser tracks from two different nHDF donors expressing ctrl or LT at 8 dpi, depicting H3K4me3 and H3K27me3 signals around TSSs of indicated genes. Genes represent two examples of each subset of genes from the GO analysis as described in figure 14 and are denoted below. Log2FCs from the differential H3K4me3 analysis performed with diffReps are indicated below the tracks.

In summary, LT overexpression led to substantial changes in the composition of the activating mark H3K4me3 and less in the repressive H3K27me3 mark. The transcriptional changes observed by RNA-Seq could partly be correlated with changes in H3K4me3, as it was the case for certain innate immune genes that were upregulated at 3 dpi. These showed an increase in H3K4me3 signal at 8 dpi. Nevertheless, the differences observed between the two experiments need to be considered when interpreting the results.

4.3 Profiling of nHDFs overexpressing MCPyV LT*tr* antigen

Essential to MCPyV tumorigenesis is the integration of the viral genome into the host genome. At the same time, there is a stop mutation in the C-terminal region of the full length LT antigen, leading to a truncated version of LT (LT*tr*). It is currently proposed that this truncation occurs prior to the integration event (Czech-Sioli et al., 2020a; Starrett et al., 2020). In this work, the LT*tr* protein was derived from the sequence of a specific MCPyV strain called MCC 350 which is the main genotype found in the USA (Goh et al., 2009). The full length LT antigen is produced exclusively from the non-integrated viral genome and expressed early in the viral life cycle in order to allow replication. In contrast, LT*tr*, which is expressed only in the tumors, has lost the C-terminal domain and viral genome replication is not possible anymore. However, the mechanisms how exactly tumorigenesis is established and driven by LT or LT*tr* remain elusive. With the aim to compare the functions of LT and LT*tr* and dissect common and distinct features, LT*tr* was overexpressed in nHDFs and transcriptional and histone modification changes were analyzed by RNA-Seq and ChIP-Seq.

4.3.1 Transcriptional changes induced by LT*tr*

To address transcriptional changes induced by LT*tr*, nHDFs were transduced with MCPyV LT*tr* and harvested at 3 and 8 dpi, i.e. early after transduction and one passage later. Subseqently, RNA-Seq was performed to detect transcriptional changes induced by LT*tr*. A summary of all DEGs comparing LT*tr* *vs* ctrl at 3 and at 8 dpi is shown in figure 16A. All genes with a log2FC > 1 or < -1 and a padj. value < 0.05 are depicted as red or blue dots, representing up- or downregulated genes, respectively. In total, 74 genes were downregulated and 659 genes were upregulated at 3 dpi. After 8 dpi, 522 genes were downregulated and 944 were upregulated. Genes that are labeled in the volcano plots represent the 10 most up- and downregulated genes referring to the log2FCs. In addition, the 10 genes with the lowest padj. values are marked for up- and downregulated genes. In order to get an overview of the biological function of significantly up- and downregulated genes, a GO term analysis was performed with the DAVID tool and is summarized in figure 16B.

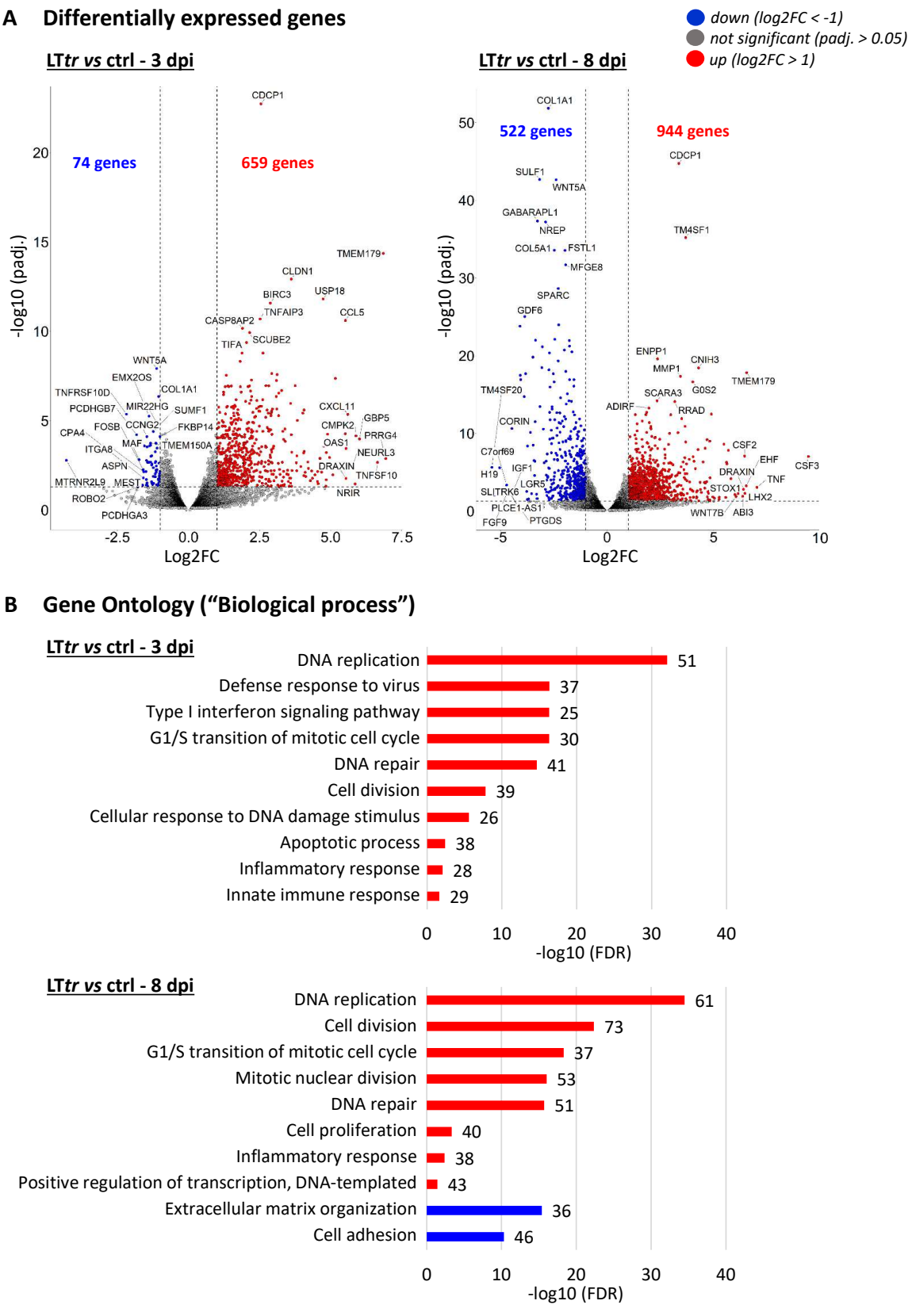


Figure 16 Transcriptional changes induced by LT*tr* overexpression in nHDFs.
(A) Volcano plots depicting all DEGs from the RNA-Seq analysis in LT*tr*-overexpressing nHDFs compared to ctrl at 3 and 8 dpi. Genes with a log2FC > 1 or < -1 and padj. < 0.05 are denoted as significantly upregulated or downregulated, respectively and are marked by red or blue dots. (B) A GO analysis ("Biological process") was performed with significantly upregulated and downregulated genes using DAVID. The 10 GO terms (FDR < 0.05, gene counts > 10) containing the highest numbers of DEGs are depicted as red or blue bars, reflecting up- or downregulated genes as input, respectively. The number of DEGs found within each GO term is noted next to the bars, FDR values are shown on the x-axis.

Remarkably, a large amount of genes marked in the volcano plots also appeared in the GO term enrichment. For example, *TNFAIP3*, *TNFSF10*, *CXCL11*, *OAS1* and *GBP1*, which are genes involved in inflammatory and innate immune signaling, were enriched at day 3 in several GO terms and were also among the top 10 highest upregulated genes. Interestingly, innate immune genes were not significantly regulated at 8 dpi, though inflammatory response genes such as *TNF* were found among the top 10 of upregulated genes and were also enriched in the GO analysis. After both 3 and 8 dpi, several GO terms containing upregulated genes were commonly enriched. These included genes involved in cell cycle, proliferation and cell division, as well as DNA replication and repair. Only a few genes belonging to these GO terms were also labeled in the volcano plots, e.g. *EHF* and *STOX1*. Nevertheless, the majority of upregulated genes at 3 and 8 dpi could be classified into the GO terms mentioned above. Interestingly, the GO terms *DNA damage* and *Apoptosis* were enriched after 3 dpi but not after 8 dpi. In addition to the terms *Inflammatory response*, *Cell cycle* and *DNA replication*, genes from the Fanconi anaemia complementation group (*FANC*) and other DNA damage and repair genes were found enriched in both time points.

Due to the fact that there was a high enrichment of upregulated genes in many different GO terms observable, only two GO terms with downregulated genes were among the 10 GO terms with FDR < 0.05 and gene counts > 10. These two terms - *Extracellular matrix organization* and *Cell adhesion* - are plotted in figure 16B. Both GO terms contained overlapping genes with some of them marked in the volcano plots. These included e.g. collagen genes such as *COL1A1* or *COL5A1*.

In summary, overexpression of LT*tr* had a substantial effect on the host's transcriptome in nHDFs in terms of different biological functions, most of all by upregulating genes involved in the cell cycle, DNA damage, repair and replication, as well as the innate immune response and inflammatory signaling. Furthermore, genes involved in ECM organization and cell adhesion were found among the downregulated genes. A complete list and further information about the genes labeled in the volcano plots, as well as their affiliation to several terms from the GO analysis performed with the DAVID tool, can be found in the supplements (table S6 and S7).

4.3.2 Histone modification changes induced by LT*tr*

To analyze possible changes in histone modification patterns induced by LT*tr*, ChIP-Seq was performed at 3 and 8 dpi, using antibodies against H3K4me3 and H3K27me3. As control datasets, the ChIP-Seq data from the vector control at 3 and 9/12 dpi was used. The heatmaps in figure 17 give an overview of the ChIP-Seq read count signals over the whole genome. Interestingly, there was some global signal increase in H3K4me3 for LT*tr* at 3 and 8 dpi, however this was only the case for Donor I. In Donor II, there was a higher global signal of H3K4me3 in the ctrl dataset. In addition, there was some increase in H3K27me3 for LT*tr* at 3 dpi in Donor I, while a weak signal reduction was observed for in Donor II. Despite these minor differences, the heatmaps showed as a quality control similar anti-correlative signal distributions of H3K4me3 and H3K27me3.

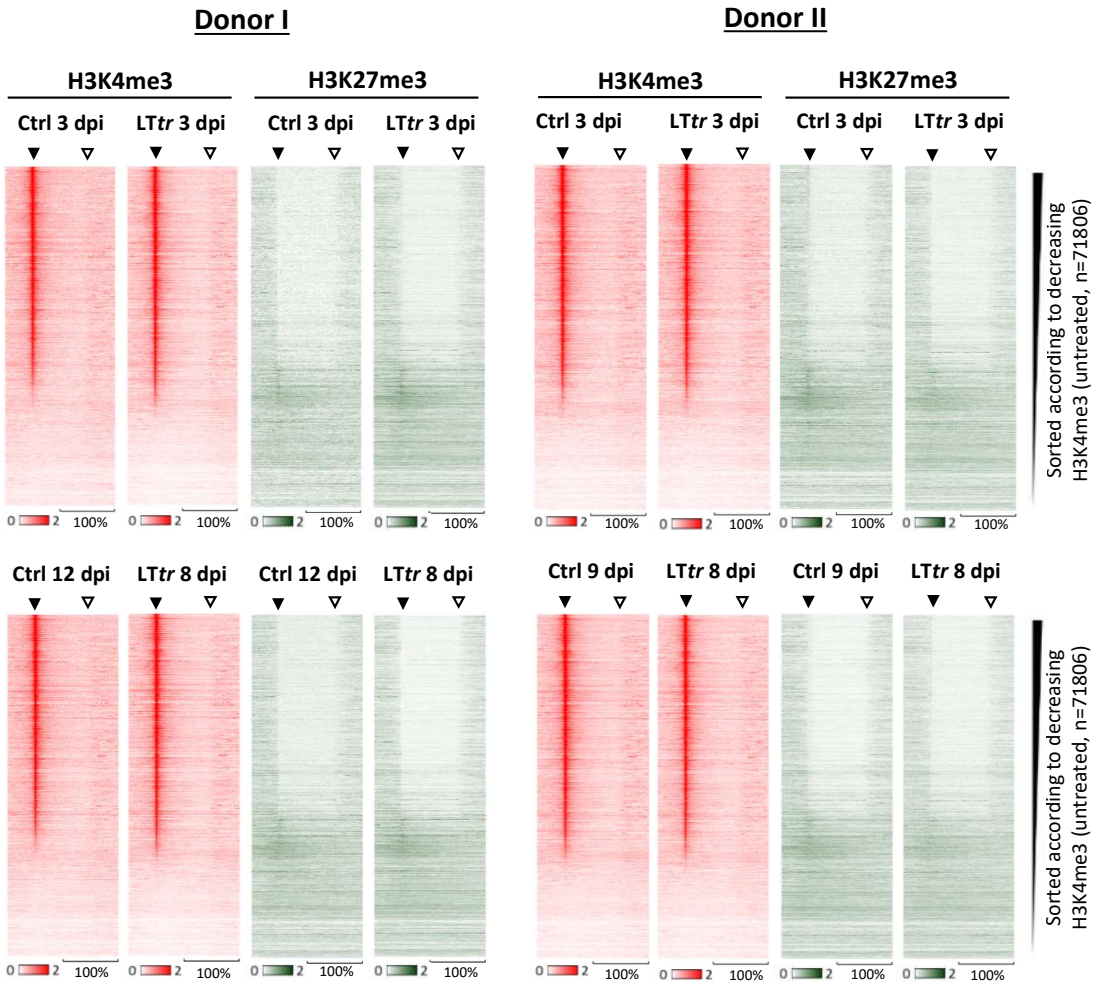


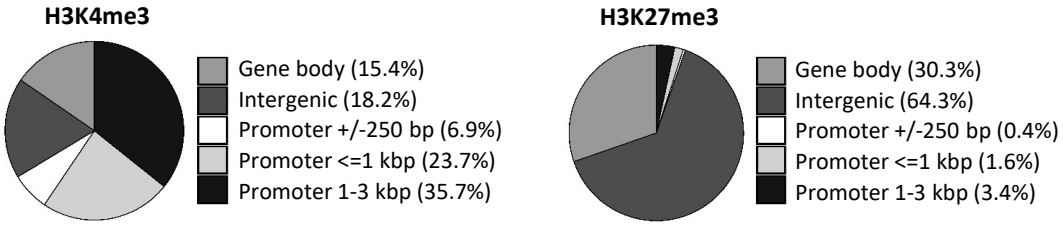
Figure 17 ChIP-Seq analysis in LT*tr*-overexpressing nHDFs.
Heatmaps depicting ChIP-Seq read count signals quantified for all genes of hg19 for H3K4me3 and H3K27me3 at 3 and 8 dpi from nHDFs overexpressing the vector control or LT*tr*. 200% of the genomic loci are shown as relative gene lengths, including TSSs (black triangles) and TTSs (non-filled triangles). Heatmaps are sorted according to decreasing H3K4me3 signal in untreated cells. Density represents DNA fragments per 1 mio reads per 1 kbp.

In the next step, a diffReps analysis was performed with the aim to detect differential histone modification signals comparing LT*tr* to the vector controls. An overview of the diffReps results is given in figure 18. The distribution of genomic features that characterizes each diffReps hit was similar for LT*tr vs* ctrl after 3 and 8 dpi. As expected, two third of all H3K4me3 diffReps hits were located in promoter regions whereas for H3K27me3, most diffReps hits were intergenic. Only around 30% could be assigned to gene bodies and less than 5% were located in promoter regions. For H3K27me3, there were only a few (3 dpi) or no genes at all (8 dpi) that showed a significant increase or decrease of at least two-fold in the diffReps analysis. Lowering the threshold to 0.5 or -0.5 increased the number of hits. These minor changes are probably not of biological relevance and were not further characterized at this point. However, a GO analysis performed with the DAVID tool can be found in the supplements (figure S9).

For H3K4me, 870 genes had a log2FC > 1, i.e. showed at least a two-fold increase in H3K4me at 3 dpi. At 8 dpi, this number decreased to 356. In contrast, there were only two genes at 3 dpi and 17 at 8 dpi which showed a significant decrease in H3K4me3 signal of at least two-fold. These genes, also marked in bold in the tables in figure 18, were further characterized by performing a GO analysis using the DAVID tool. In figure 19A, the average read count signals for H3K4me3 were plotted for all genes with a significant increase or decrease of at least two-fold around promoter regions, comparing LT*tr* to ctrl. It needs to be noted that the increase in Donor I was more pronounced than in Donor II, reflecting the general global increase in H3K4me3 noted in the heatmaps (figure 17). Genes that passed a gene count threshold of at least 50 were taken into consideration for the GO analysis using the same conditions as described previously. These criteria excluded the few genes that had a decreased H3K4me3 signal.

In figure 19B, the top 10 GO terms ("Biological process") are listed for increasing H3K4me3 at 3 and 8 dpi. In general, many genes showed an enrichment in signaling pathways such as *Small GTPase mediated signal transduction* or *Activation of MAPK activity*. In addition, some GO terms specifically referred to neurological processes such as *Nervous system development* or *Chemical synaptic transmission*. Interestingly, *Positive/negative regulation of transcription* were also GO terms that contained many genes with increased H3K4me3 signal, among them transcription factors such as *IRF8*, *TGIF1*, *TP73*, *HDAC9* or *PRDM1*, the two latter being chromatin remodelers. Similar to the observations made for LT, some genes involved in innate immune signaling had an increase in H3K4me3 signal at 8 dpi but were transcriptionally regulated only after 3 dpi.

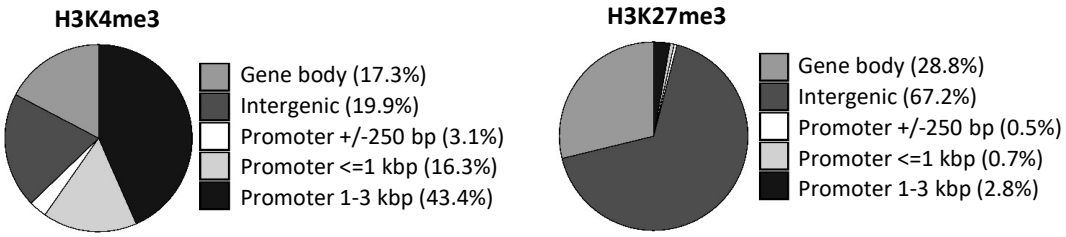
A **3 dpi - Distribution of genomic features from the diffReps analysis** *(log2FC > 0.5 or < -0.5, padj. < 0.05)*



3 dpi - DiffReps hits

diffReps	H3K4me3 total	H3K4me3 for annotated genes	H3K27me3 total	H3K27me3 for annotated genes
log2FC > 0.5	6776	4908	2560	**620**
log2FC < -0.5	268	181	927	**296**
log2FC > 1	1266	**870**	20	8
log 2FC < -1	15	**2**	11	5

B **8 dpi - Distribution of genomic features from the diffReps analysis** *(log2FC > 0.5 or < -0.5, padj. < 0.05)*



8 dpi - DiffReps hits

diffReps	H3K4me3 total	H3K4me3 for annotated genes	H3K27me3 total	H3K27me3 for annotated genes
log2FC > 0.5	5774	4083	516	**159**
log2FC < -0.5	324	166	422	**100**
log2FC > 1	533	**356**	0	0
log 2FC < -1	102	**17**	0	0

Figure 18 Differential histone modification analysis in nHDFs overexpressing LT*tr*.
(A) Distribution of genomic features contained in the diffReps analysis which was performed for H3K4me3 and H3K27me3 signals in nHDFs overexpressing LT*tr vs* vector control at 3 dpi (A) and at 8 dpi (B). All significant hits with a log2FC $> 0.5/< -0.5$ were classified according to their annotations. The tables below summarize the total numbers of diffReps hits with different log2FC cutoffs and show the amount of hits with annotated genes (at promoters or gene bodies). Numbers in bold represent genes that were used for downstream analysis.

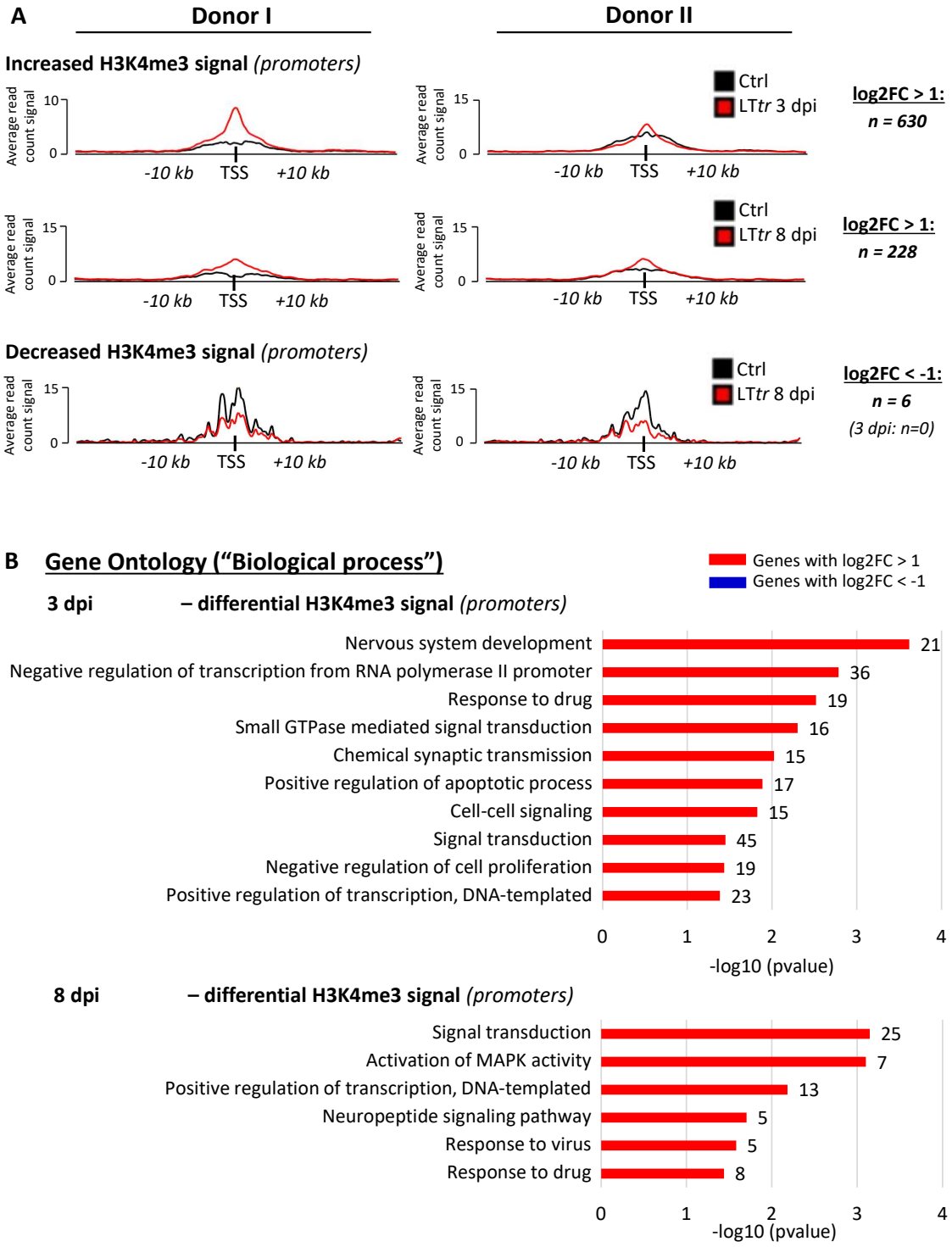


Figure 19 Analysis of changes in H3K4me3 induced by LT*tr* overexpression in nHDFs.
(A) Average read count signals of H3K4me3 for all genes with differential signal comparing LT*tr* *vs* ctrl at 3 and 8 dpi. Total numbers of genes with differential H3K4me3 signal around promoters (+/-3 kbp) of annotated genes are shown next to the average plots. (B) A GO analysis ("Biological process") was performed using DAVID with genes from the diffReps analysis described in (A). Only gene sets with n$>$ 50 were considered for the GO analysis. From all GO terms with at least 5 genes and a pvalue of $<$ 0.05, the 10 GO terms with the highest gene counts are shown.

To specifically resolve if transcriptional changes correlated with histone modification changes, the RNA-Seq and ChIP-Seq results were summarized in the scatterplots shown in figure 20. For this, each gene of the human genome was measured for H3K4me3 and H3K27me3 signal and plotted, with the x-axis representing the ctrl and the y-axis the LT*tr* data sets. On the right side of each plot the signal from the RNA-Seq data was incorporated and is reflected by the color code. The H3K4me3 signal correlated well with the RNA-Seq data in all samples. However, on day 3 this effect was more pronounced than after 8 dpi where many genes with H3K4me3 signal were not as strongly regulated on the transcriptional level. For H3K27me3, as observed for LT, the correlation was unclear. Although at 8 dpi in Donor II there was a shift of upregulated genes towards the right side of the plot, thereby representing genes with less H3K27me3 in LT*tr* compared to ctrl, for the other samples this effect was not reproduced. Rather, there was a general upregulation of genes distributed all over the scatterplots, showing that transcriptional regulation was not directly linked to H3K27me3 signal.

In accordance with the observations made from the diffReps analysis, changes in the repressive H3K27me3 mark were less pronounced and mostly not correlative to transcriptional changes. However, the H3K4me3 signal correlated well with the RNA-Seq data. In order to find genes with common transcriptional regulation and histone modification levels and classify them in terms of biological relevance, gene subsets from the RNA-Seq GO analysis were measured for their signal in H3K4me3 and H3K27me3 (figure 21A). Interestingly, the boxplots show that in Donor I there was a general increase in H3K4me3 for the specific gene subsets in LT*tr vs* ctrl, but for Donor II this effect was slightly reversed, except for the GO term *Inflammatory response*. This reflects the observations from the heatmaps (figure 17) which showed that the global baseline levels of H3K4me3 were higher in the ctrl from Donor I compared to Donor II. Though the H3K27me3 levels showed a decrease for the upregulated genes for all subsets depicted in figure 21, their signals were close to zero and therefore negligible for further interpretation. Regarding the proportion of upregulated genes within the subsets from the RNA-Seq GO analysis, roughly half of them also showed a significant increase of at least one-fold in H3K4me3 (figure 21B). Interestingly, downregulated genes within the GO terms *ECM organization* and *Cell adhesion* did not show a decrease in H3K4me3.

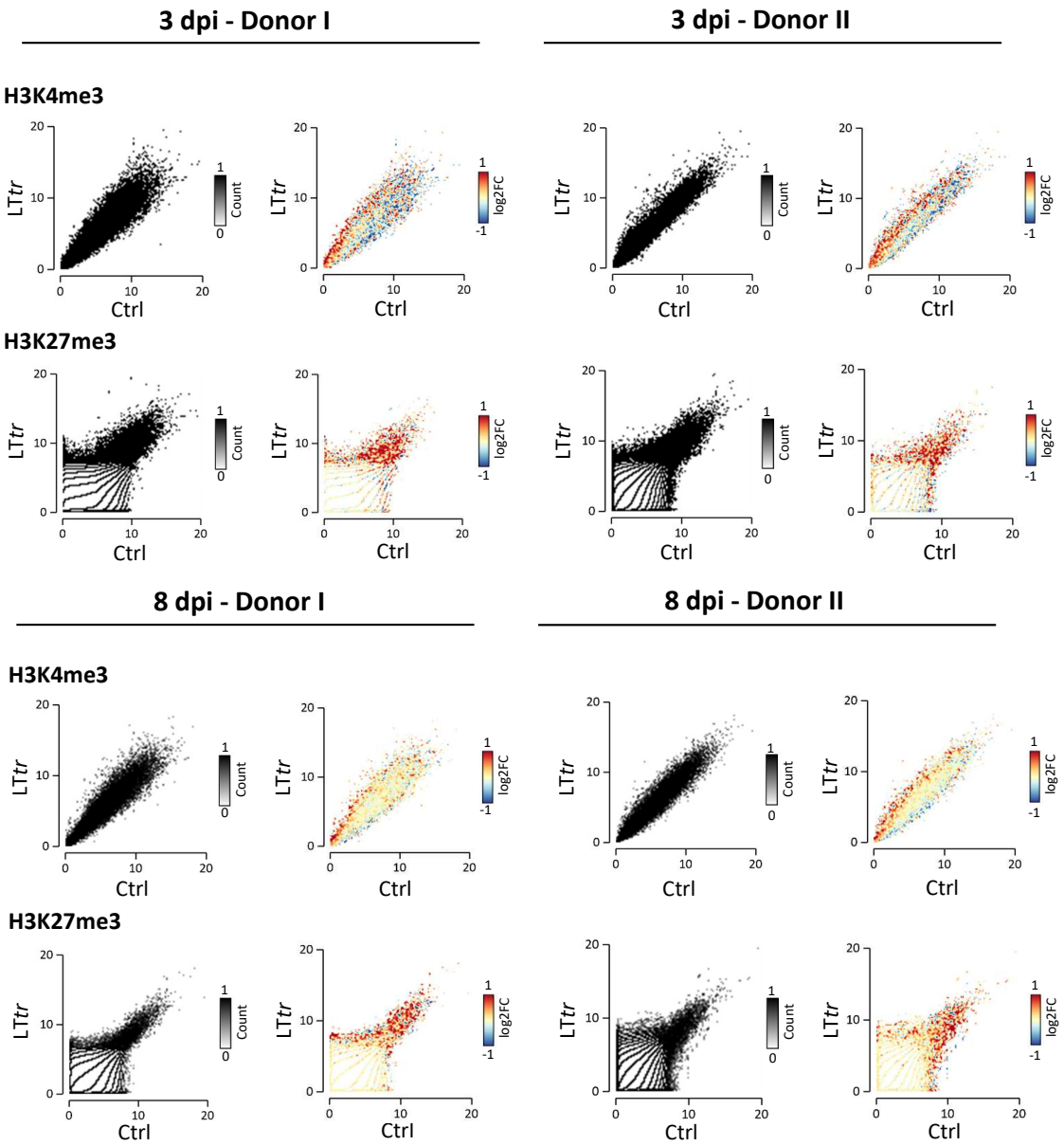


Figure 20 Correlation of ChIP-Seq with RNA-Seq analysis in LT*tr*-overexpressing nHDFs.
Quantification of H3K4me3 and H3K27me3 read signals for all genes from hg19 in nHDFs overexpressing ctrl or LT*tr* at 3 dpi and 8 dpi. Histone modification signals were plotted for the ctrl on the x-axis and for LT*tr* on the y-axis. Right scatterplots incorporate the information on the log2FCs from the RNA-Seq analysis, represented by the color code. The x and y-axes were segmented into 100 bins and regions within these bins are depicted by the counts.

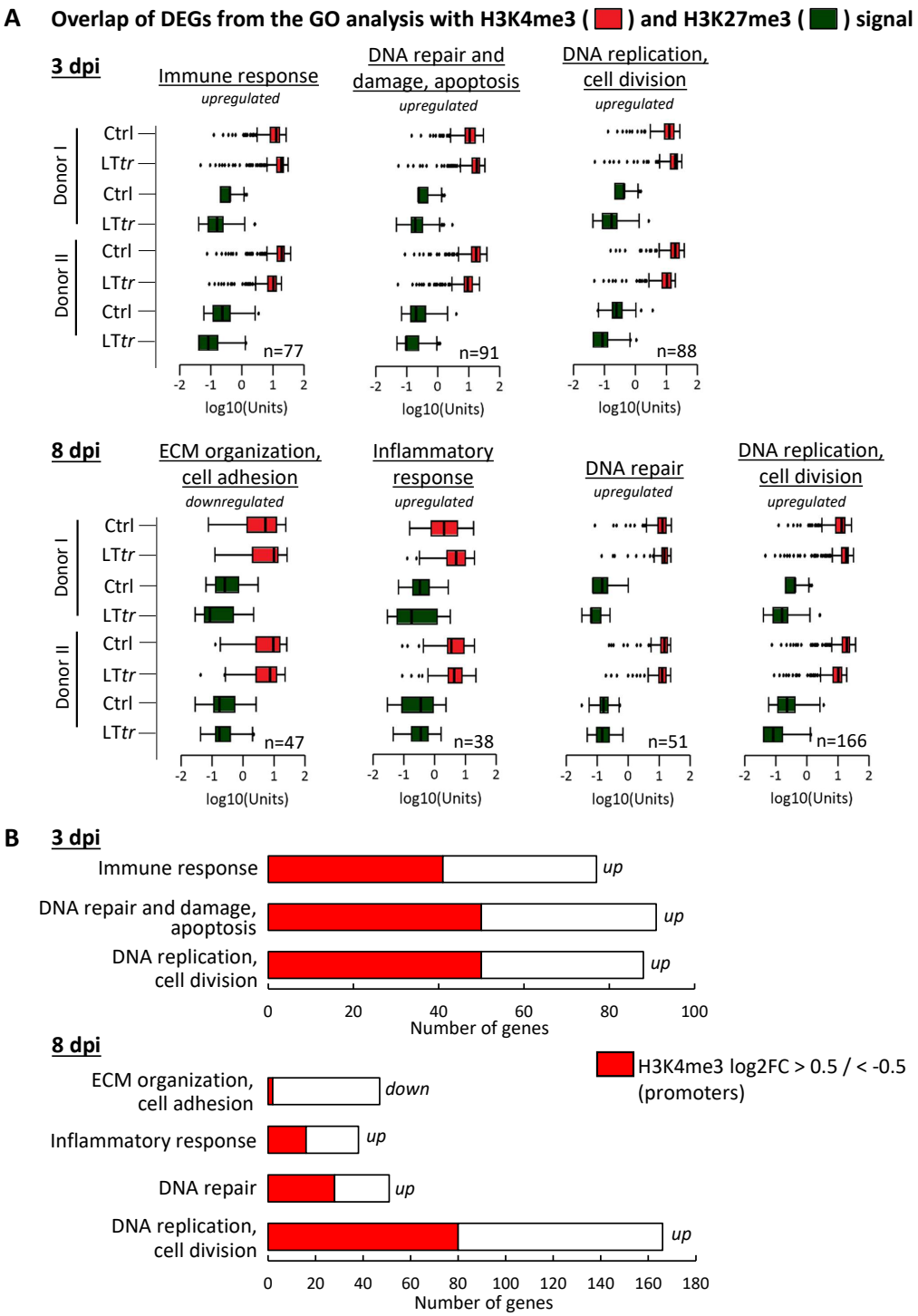


Figure 21 ChIP-Seq signals for a subset of genes from the RNA-Seq GO analysis in LT*tr*-overexpressing nHDFs. (A) Subsets of genes from the RNA-Seq GO analysis (see Figure 16B) quantified for their signal in H3K4me3 and H3K27me3. Boxplots representing the distribution of histone modification signals around the TSSs (<=3 kbp) are shown for H3K4me3 and H3K27me3. (B) Total numbers of genes from the subsets described above are shown in the bar plots, representing DEGs from the RNA-Seq analysis and their proportion of differential H3K4me3 signal (red).

The observations from the correlation of the RNA-Seq and ChIP-Seq analysis revealed changes induced by LT*tr* for specific genes. However, variance between the two experiments performed

in different nHDF donors were detectable as well. Therefore, in figure 22, some genes from the different subsets were selected and visualized with the igv genome browser to detect changes in histone modification signals and variations in the two experiments from different nHDF donors. In fact, for all the eight selected genes, the genome browser tracks from Donor II at 9 dpi showed substantially higher levels in H3K4me3 compared to all other ctrl tracks. However, for all other conditions, LT*tr* showed increased levels of H3K4me3 signal compared to the vector controls. This was particularly pronounced for *OAS2*, *IRF7*, *IL1B* and *CXCL2*, representing genes from the GO terms *Innate immune response* and *Inflammatory response*. Interestingly, the examplary gene from the GO terms *ECM organization* and *Cell adhesion*, *CYP1B1*, had only a reduced H3K4me3 signal at the late time point which was consistent in both donors. As visible also in the previous figures, the tracks for H3K27me3 highlight that there were only minor changes and very low levels of this repressive mark detectable.

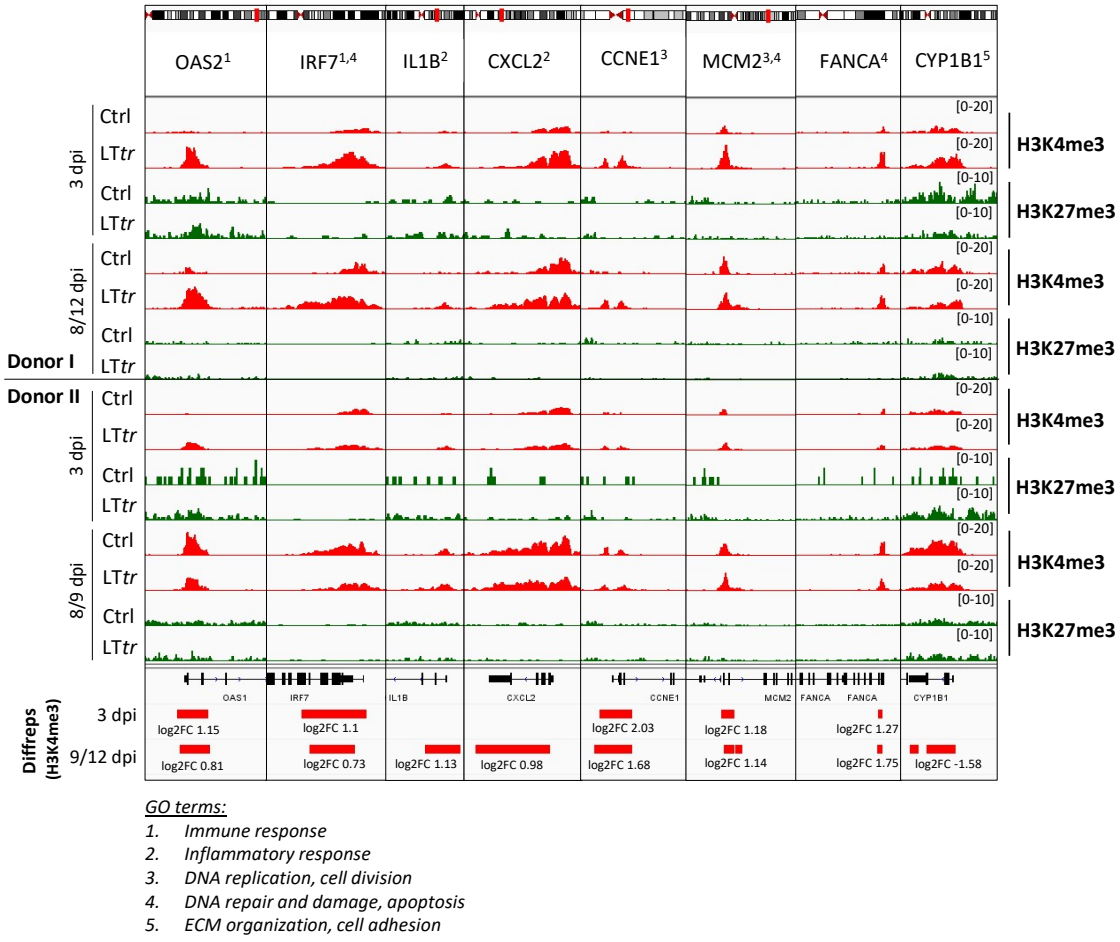


Figure 22 ChIP-Seq genome browser tracks for a subset of DEGs in LT*tr*-overexpressing nHDFs.
Genome browser tracks from two different nHDF donors expressing vector control or LT*tr* at 3 and 8 dpi, showing H3K4me3 and H3K27me3 signals around TSSs of indicated genes. Genes represent two examples of each subset of genes from the GO analysis described in figure 16B. Log2FCs from the differential H3K4me3 analysis are indicated below.

In summary, analysis of histone modification changes in nHDFs overexpressing LT*tr* revealed many similarities to changes observed for LT. Most striking was the fact that innate immune signaling genes were transcriptionally downregulated to a significant degree only at the early time point while H3K4me3 was increased as early as 3 dpi, and lasted also until the later time points. In addition, many genes from the RNA-Seq data that were enriched by the GO analysis also correlated with H3K4me3 signals and half of them showed significantly reduced H3K4me3 levels but no or only minor changes in H3K27me3. Interestingly, at 8 dpi, there were much more genes in LT*tr*-overexpressing nHDFs with an increase in H3K4me3 promoter signal (228 genes, log2FC > 1, padj. < 0.05) compared to LT (56 genes, log2FC > 1, padj. < 0.05). Consequently, the GO analysis for LT*tr* revealed an accumulation of genes involved in signaling processes which were not detected for LT.

4.4 Profiling of nHDFs overexpressing MCPyV sT antigen

MCPy sT is expressed both during the early phase of infection but also in the context of tumorigenesis and metastasis. Importantly, sT has shown in several reports to interfere with different cellular signaling pathways. With the aim to reveal new functions or mechanisms of sT in perturbing the host cellular transcriptome, nHDFs overexpressing sT were assessed for their transcriptional profiles. In addition, ChIP-Seq was performed to determine if sT is able to induce changes in histone modifications.

4.4.1 Transcriptional changes induced by sT

To analyze transcriptional changes induced by sT in nHDFs, sT was overexpressed in two different nHDF donors and harvested at 3 and 8 dpi for RNA isolation and subsequent sequencing. As a control, the data from the nHDFs overexpressing the vector ctrls at 3 or 9/12 dpi were used. The volcano plots in figure 23A give an overview of all DEGs that were detected at the two different time points in Donor I and II. Upon sT overexpression, in total 362 genes were significantly upregulated at 3 dpi and 533 at 8 dpi, representing genes with an increase of at least two-fold. Concerning sT-downregulated genes, 449 genes were detected at 3 dpi and 795 genes at 8 dpi with a decrease of at least two-fold compared to the vector control.

In order to get an overview of the biological functions, a GO analysis was performed applying the DAVID tool. Similar to the analyses conducted with the other T antigens, GO terms with an FDR value < 0.05 and a minimum of 10 gene counts were filtered and the top 10 GO terms are shown in figure 23B. A complete list of the genes labeled in the volcano plots as well as their presence in some of the GO terms can be found in table S8 and table S9.

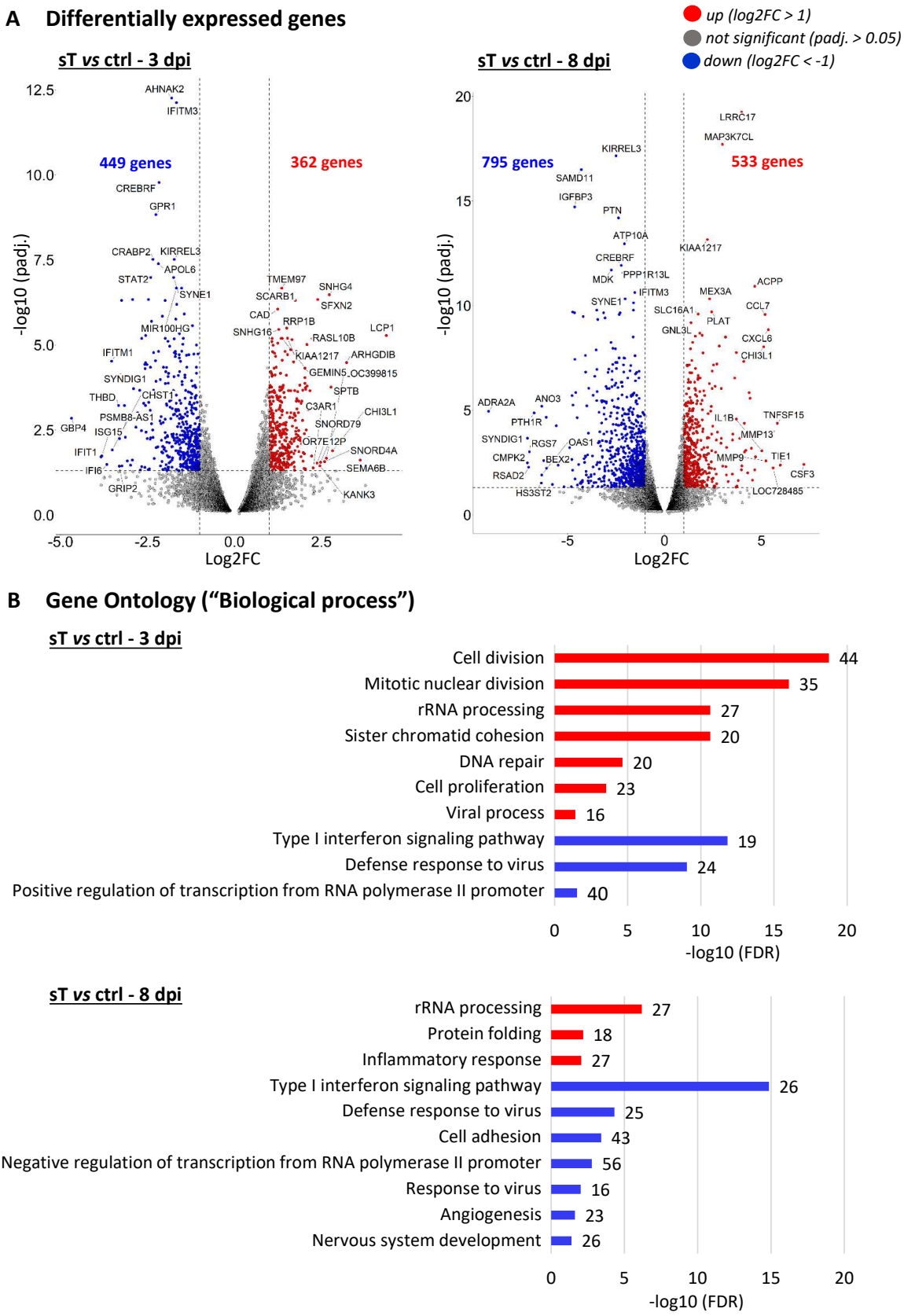


Figure 23 Transcriptional changes induced by sT overexpression in nHDFs.
(A) Volcano plots depicting all DEGs from the RNA-Seq analysis in sT-overexpressing nHDFs compared to the vector control at 3 and 8 dpi. Genes with a log2FC > 1 or < -1 and a padj. value < 0.05 are denoted as significantly upregulated or downregulated, respectively. (B) A GO analysis ("Biological process") was performed using the DAVID tool with significantly upregulated and downregulated genes from the RNA-Seq analysis in sT-overexpressing nHDFs at 3 and 8 dpi. The 10 GO terms (FDR < 0.05, gene counts > 10) containing the highest numbers of DEGs are depicted as red or blue bars, reflecting up- or downregulated genes as input, respectively. The number of DEGs found within each GO term is denoted next to the bars, FDR values are shown on the x-axis.

After 3 dpi, a large proportion of upregulated genes was assigned to the GO terms *Cell division*, including e.g. *cell division cycle* (*CDC*) genes, mitotic checkpoint genes such as *BUB1* or *regulator of chromosome condensation RCC1* and *RCC2*. The GO term *rRNA processing* was enriched both at 3 and 8 dpi and contained ribosomal genes such as the *ribosomal RNA processing 1B* (*RRP1B*) gene which was also labeled in the volcano plot at 3 dpi. The GO term *Viral process* which was enriched at 3 dpi contained genes with different functions, e.g. genes involved in cell cycle and division, as well as chromatin remodelers like *SUV39H1*, a histone lysine methyltransferase. After 8 dpi, there was a strong enrichment of cytokines and inflammatory genes that also appeared among the top regulated genes. *CCL7*, *CXCL6* or *IL1B* are some examples that were enriched in the GO term *Inflammatory genes* and are also labeled in the volcano plots.

A large subset of downregulated genes both at 3 and 8 dpi was assigned to the GO terms *Defense response to virus* and *Type I IFN interferon signaling pathway*, but also to several other terms with overlapping functions. These included a broad spectrum of ISGs such as *IFIT1*, *IFI6*, *ISG15*, *RSAD2* or *OAS1*, but also specifically *IRF7* and *IRF9* as interferon regulatory factors. All of these genes were significantly downregulated by sT at both time points. Interestingly, genes involved in transcriptional regulation, both negative and positive regulators, were also enriched at both time points, including e.g. *IRF7* and *STAT1*, but also many other transcription factor subunits from the *Fos proto-oncogene* (*FOS*), the *JunB proto-oncogene* (*JUNB*) or *CCAAT Enhancer Binding Proteins CEBPDB* and *CEBPDD*, the latter being involved in innate immune signaling.

Finally, there was an enrichment of genes involved in cell adhesion and angiogenesis that were downregulated at 8 dpi. Some examples include the *vascular endothelial growth factors VEGFA* and *VEGFB*, the metallopeptidase *ADAM23* or the *neuronal cell adhesion molecule NRCAM*.

In summary, overexpression of sT in nHDFs led to substantial changes in the host transcriptome and revealed its high potential to interfere with important cellular signaling pathways and transcription factors. A so far unidentified target that was strongly deregulated by sT, as revealed by the transcriptomic data presented here, is the type I IFN response. Therefore, in chapter 4.4.3, the role of sT in perturbing innate immune signaling will be analyzed in more detail.

4.4.2 Histone modification changes induced by sT

sT expression led to the deregulation of a large number of genes. Changes were observed directly after 3 dpi, as well as after 8 dpi. With the aim to assess if sT might be able to induce changes in the composition of histone modifications thereby altering the host transcriptional response, ChIP-Seq was performed for H3K4me3 and H3K27me3. In addition, H3K27ac was assessed as another activating mark that can be present both at promoters and enhancers, and H3K9me3 as a mark of constitutive heterochromatin.

As a quality control and to compare global histone modification signals in the different samples, the heatmaps in figure 24 summarize the results for the four different histone modifications in all the different conditions. For the early time point, i.e. at 3 dpi, only H3K4me3 and H3K27me3 were analyzed. In both donors, they showed an anti-correlative signal distribution and were generally similar in all conditions. However, there was a small increase in H3K4me3 signal observable when comparing sT *vs* ctrl and a small enhancement of H3K27me3 signal in Donor II compared to Donor I.

At 8 dpi, there was still a similar distribution observable, however the H3K4me3 signals were rather decreased in sT *vs* ctrl. Further, the difference in H3K27me3 between the two donors was not as pronounced. H3K27ac signal showed a general overlap with H3K4me3 signal which was expected as they are mostly co-occurring at active promoters. In comparison to the lack of signal around the TTSs in the H3K4me3 heatmaps, there was a consistent signal in H3K27ac observable over all genomic loci with an accumulation around TSSs. Concerning H3K9me3, the heatmaps looked similar in all conditions and there was a lack of signal around the TSSs. The heatmaps suggest that there was only a very weak H3K9me3 signal present in the different samples and no specific enrichment at certain genomic loci.

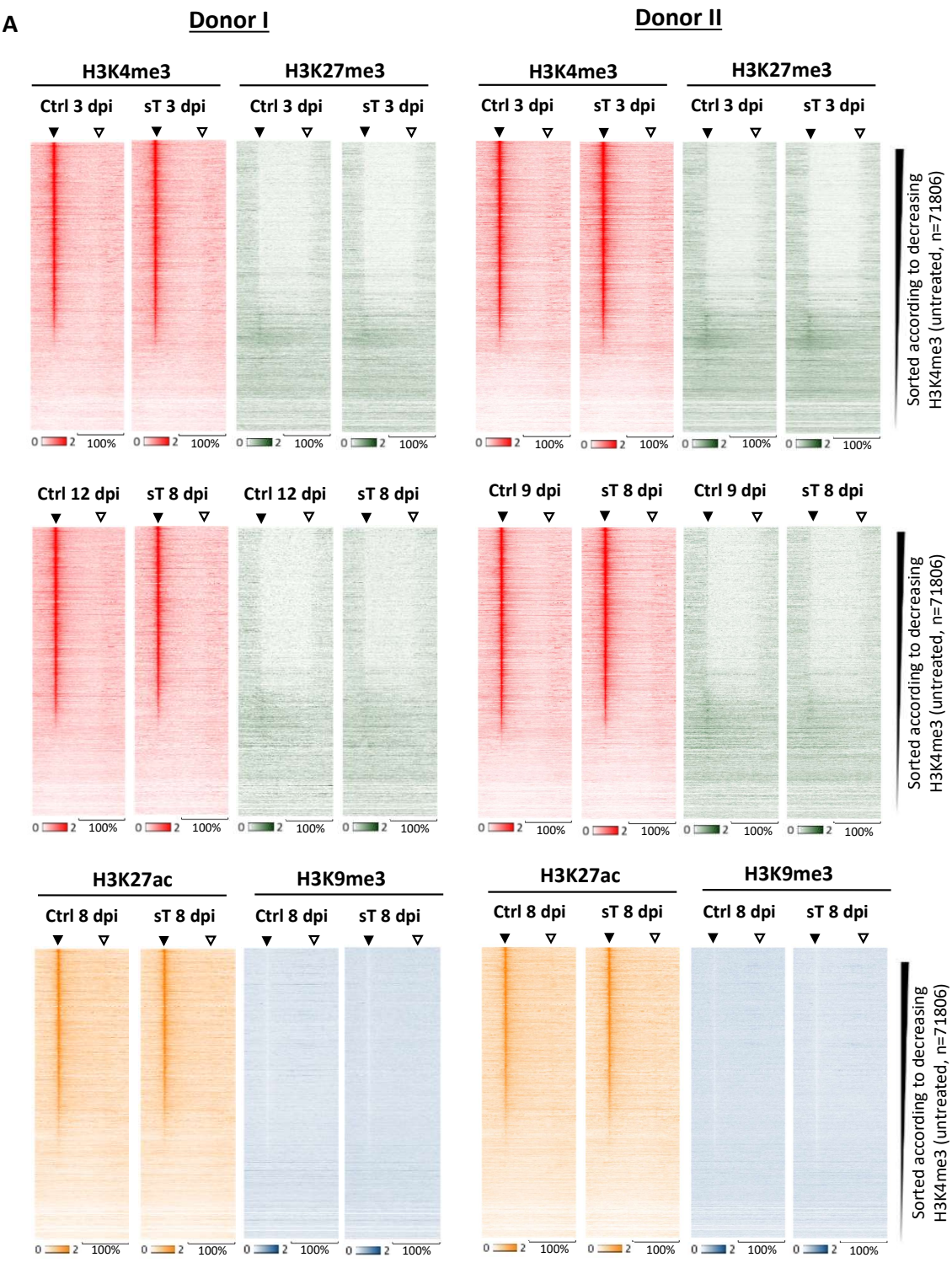


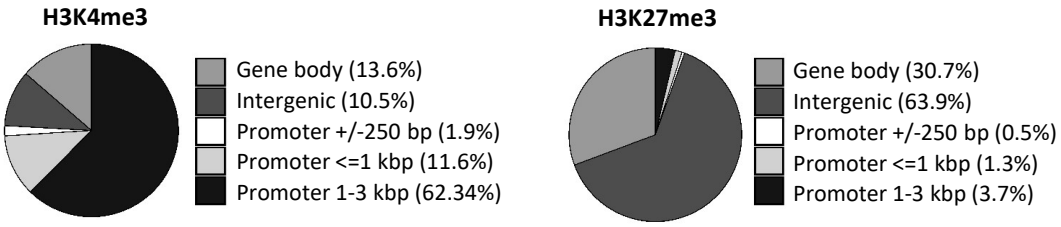
Figure 24 ChIP-Seq analysis in sT-overexpressing nHDFs.
(A) Heatmaps depicting ChIP-Seq read count signals quantified for all genes of hg19 for H3K4me3 and H3K27me3 at 3 and 8 dpi from nHDFs overexpressing the vector control or sT. In addition, signals for H3K27ac and H3K9me3 are shown for 8 dpi. Relative lengths of each genomic locus are depicted for a window size of 200%, including TSSs (black triangles) and TTSs (non-filled triangles). Heatmaps are sorted according to decreasing H3K4me3 signal in untreated cells. Density represents DNA fragments per 1 mio reads per 1 kbp.

In summary, the heatmaps reflect a ChIP-Seq analysis with expected signal distributions in all samples and suggest that there were no global changes induced by sT overexpression, except for some variations in signal intensity. In order to detect specific gene loci with differential histone modification signals upon sT overexpression in nHDFs, a diffReps analysis was performed as described earlier. A summary is given in figure 25. For H3K4me3 at 3 dpi, two third of all diffReps hits were located in promoter regions, with the majority annotated at a distance of 1-3 kbp around TSSs. 357 annotated genes were significantly increased in H3K4me3 with a log2FC of > 1, while only 21 had a significantly decreased H3K4me3 signal of at least two-fold. Interestingly, after 8 dpi, the H3K4me3 distributions shifted towards promoters that were more closely located around the TSSs. However, there was a small decrease in the overall promoter annotation. In contrast to 3 dpi, with 678 genes, most annotated hits had decreased log2FC values, while only 19 genes had a significantly increased H3K4me3 signal of at least two-fold.

For H3K27me3, there was a similar distribution at both 3 and 8 dpi. Nearly two third of all diffReps hits were located in intergenic regions, the remaining proportion was mainly located within gene bodies and around 5% in promoter regions. Only a few annotated genes were more than two-fold increased or decreased. When lowering the log2FC threshold to 0.5 or -0.5, 465 and 583 genes had increased H3K4me3 signal after 3 and 8 dpi, respectively, and 285 and 616 were decreased at 3 and 8 dpi.

H3K27ac diffReps hits were located to a half within gene bodies and roughly 10% were located in promoter regions (figure 25B). Taking a closer look at the total numbers of diffReps hits and genes that were annotated to them, these comprised 1707 genes with significantly increased H3K27ac signal of at least two-fold. Interestingly, 6853 genes had at least two-fold reduced levels of H3K27ac in sT *vs* ctrl. Finally, the H3K9me3 diffReps analysis revealed only a few genes with differential signal. As most of the changes were located in intergenic regions, they were not associated with any specific genes. Nearly no hits were located in promoter regions, but around 40% were found within gene bodies. Even with a log2FC cutoff value of 0.5 or -0.5 there were only 13 genes with increased and 102 with decreased H3K9me3 signal.

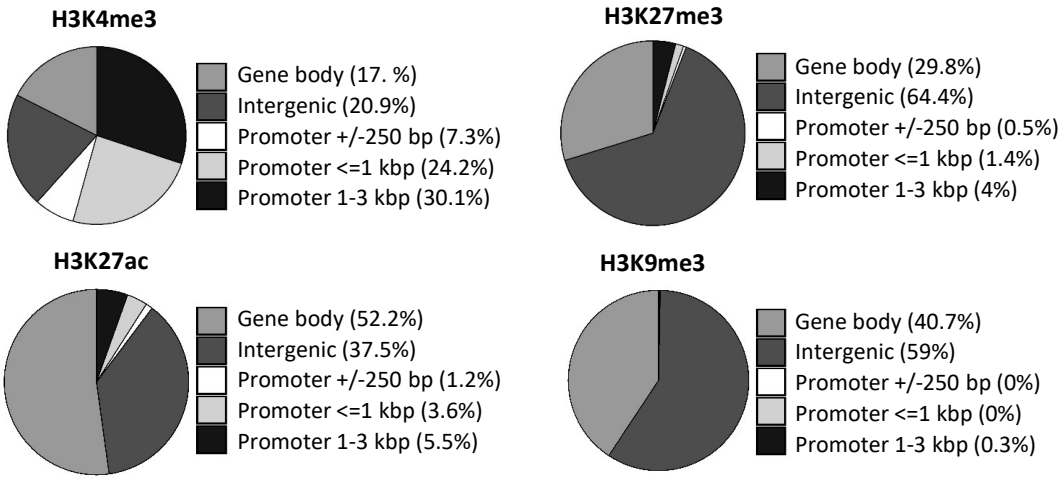
A **3 dpi - Distribution of genomic features from the diffReps analysis** *(log2FC > 0.5 or < -0.5, padj. < 0.05)*



3 dpi - DiffReps hits

diffReps	H3K4me3 total	H3K4me3 for annotated genes	H3K27me3 total	H3K27me3 for annotated genes
log2FC > 0.5	2857	2436	1708	**465**
log2FC < -0.5	320	21	905	**285**
log2FC > 1	483	**357**	28	8
log 2FC < -1	28	**21**	5	7

B **8 dpi - Distribution of genomic features from the diffReps analysis** *(log2FC > 0.5 or < -0.5, padj. < 0.05)*



8 dpi - DiffReps hits

	H3K4me3		H3K27me3		H3K27ac		H3K9me3	
diffReps	**total**	**annotated genes**	**total**	**annotated genes**	**total**	**annotated genes**	**total**	**annotated genes**
log2FC > 0.5	954	857	2067	**583**	11238	6456	36	**13**
log2FC < -0.5	4075	2717	2261	**616**	14335	9532	259	**102**
log2FC > 1	31	**19**	93	25	3458	**1707**	0	0
log 2FC < -1	1035	**678**	20	6	10298	**6853**	2	2

Figure 25 Differential histone modification analysis in nHDFs overexpressing sT.
(A) Distribution of genomic features contained in the diffReps analysis which was performed for H3K4me3 and H3K27me3 signals in nHDFs overexpressing sT *vs* vector control at 3 dpi (A) and for H3K4me3, H3K27me3, H3K27ac and H3K9me3 signals in nHDFs overexpressing sT *vs* vector control at 8 dpi (B). All significant hits with a log2FC $> 0.5/< -0.5$ were classified according to their annotations. The tables below summarize the total numbers of diffReps hits with different log2FC cutoffs and show the amount of hits with annotated genes (at promoters or gene bodies). Numbers in bold represent genes that were used for downstream analysis.

To identify specific genes with changes in histone modification signal and classify them according to their biological relevance, a GO analysis was performed using the conditions described in the previous chapters. Due to the fact that sT overexpression led to only marginal differences in H3K27me3 and H3K9me3, considering that most of the genes were only significantly changed less than two-fold, these were not further analyzed at this point. Nevertheless, a GO analysis can be found in the supplements (figure S10 and S11).

The GO analysis performed with the genes with differential H3K4me3 signal is summarized in figure 26. The average read count signals for the annotated genes in promoter regions with differential H3K4me3 signal are depicted for sT and ctrl at 3 and 8 dpi and the total numbers of genes are noted next to the plots. At 3 dpi, 245 genes with increased H3K4me3 were detected and a GO analysis was conducted using the DAVID tool. Many GO terms contained genes involved in transcriptional or translational regulation, including ribosomal proteins or transcription factors. In addition, genes involved in inflammatory signaling were found within GO terms such as *Viral entry into host cell* or *Response to lipopolysaccharide*, including *CXCL3*, *CXCL5*, *TLR4* or *TNFRSF9*, for example.

Conversely, a GO analysis using the genes with decreased H3K4me3 signal was conducted with the data from 8 dpi only, because at 3 dpi there were only 14 genes with decreased H3K4me3 promoter signal, while at 8 dpi there were 408 genes. A large amount of these genes comprised transcription regulatory genes, including transcription factors such as *IRF5* or *IRF9*, *PR/SET domains PRDM13* or *PRDM16*, *fibroblast growth factor 7* (*FGF7*) or *cyclin dependent kinase inhibitor 1C* (*CDKN1C*). Intriguingly, the GO term *Defense response to virus* contained many genes involved in innate immune signaling, e.g. *IFIT1* or *IFIT3*, *OAS1* and *OAS2*, as well as *IRF5* and *IRF9*, overlapping with the findings from the RNA-Seq analysis that sT downregulated the transcription of a large subset of ISGs.

In addition to H3K4me3, H3K27ac was also assessed as another mark for active promoters and enhancers. The average read count signals from genes identified with the diffReps analysis are shown in figure 27. Genes with a significant increase or decrease of H3K27ac in annotated promoter regions were plotted for sT and ctrl at 8 dpi. A GO analysis was performed using the same conditions as for H3K4me3. The top 10 GO terms were filtered and are depicted in figure 27B.

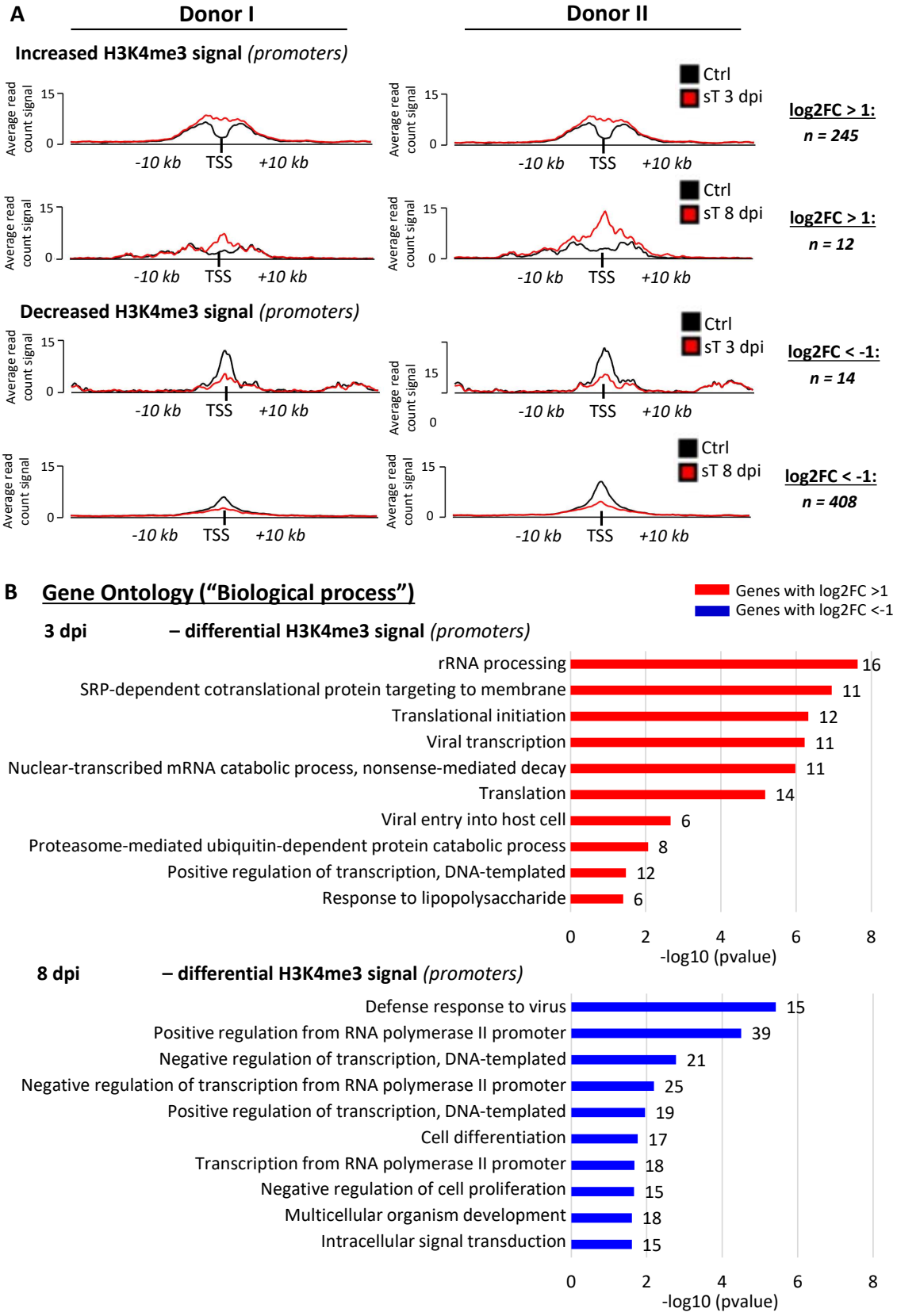


Figure 26 Analysis of changes in promoter H3K4me3 signal induced by sT in nHDFs.
(A) Average read count signals of H3K4me3 for all genes with differential signal comparing sT *vs* ctrl at 3 and 8 dpi. Total numbers of genes with differential H3K4me3 signal around promoters (+/-3 kbp) of annotated genes are shown next to the average plots. (B) A GO analysis ("Biological process") was performed with DAVID using the genes from the diffReps analysis described in (A). Only gene sets with n$>$ 50 were considered for the GO analysis. From all GO terms with at least 5 genes and a pvalue of $<$ 0.05, a maximum of 10 GO terms with the highest gene counts are shown.

Three GO terms contained genes with increased H3K27ac promoter signal, including genes involved in *Translation*, e.g. ribosomal proteins as well as transcriptional regulatory genes such as *cyclin C* (*CCNC*) or the *BAF chromatin remodeling complex subunit BCL11A*. Furthermore, several genes with increased H3K27ac signal were found in the GO term *Protein phosphorylation*, but also genes with decreased H3K27ac signal could be assigned to this GO term. Some of the genes were overlapping, but there were several *mitogen-activated protein kinase* (*MAPK*) genes that specifically showed a reduction in H3K27ac signal. Additional GO terms comprising genes with decreased H3K27ac signal included *Signal transduction*, *Cell adhesion* or *Positive regulation of GTPase activity*. Intriguingly, several GO terms contained genes with decreased H3K27ac signal involved in innate immune signaling, including *IFI6*, *IFI27*, *IRF9*, *MX1* or *MX2*, for example.

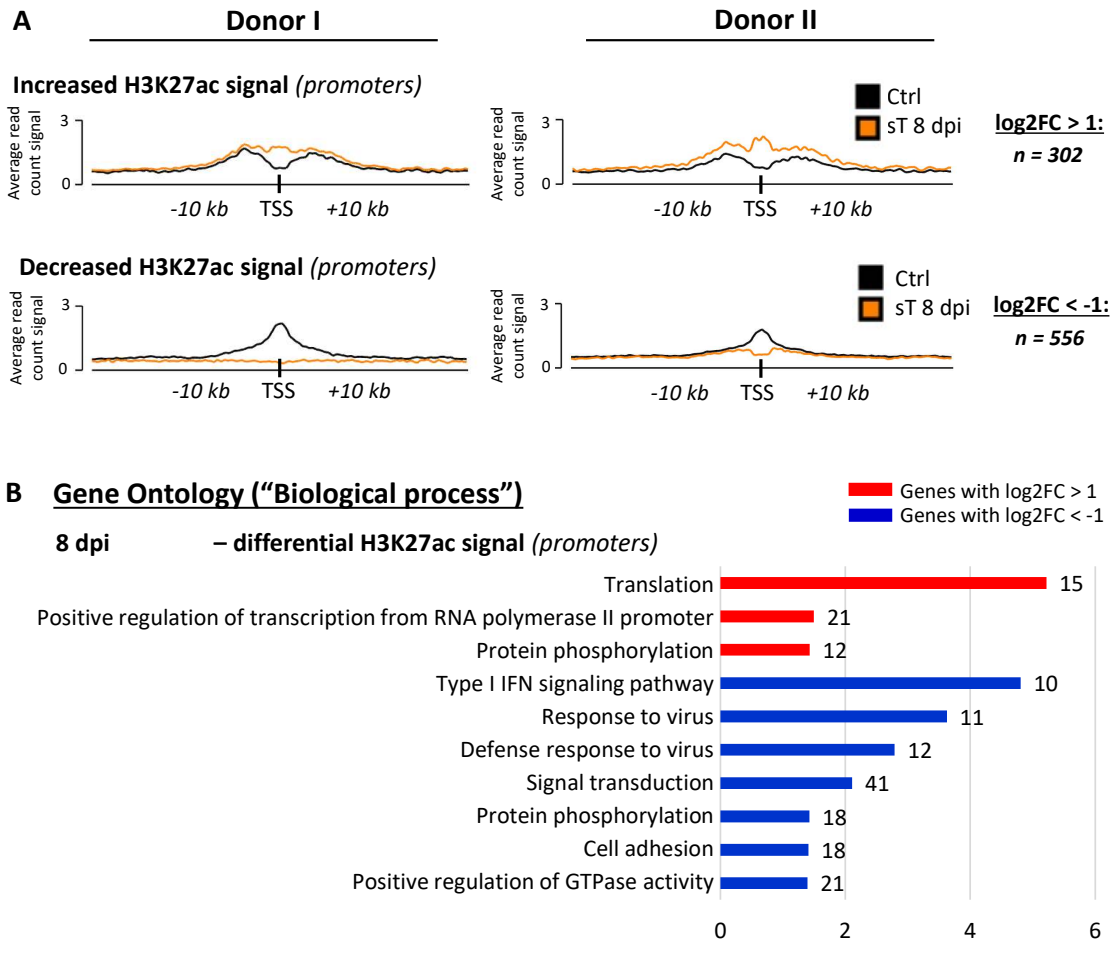


Figure 27 Analysis of changes in H3K27ac promoter signal induced by sT in nHDFs.
(A) Average read count signals of H3K27ac for all genes with differential signal comparing sT *vs* ctrl at 8 dpi. Total numbers of genes with differential H3K27ac signal around promoters (+/-3 kbp) of annotated genes are shown next to the average plots. (B) A GO analysis ("Biological process") was performed using DAVID with genes from the diffReps analysis described in (A). Only gene sets with n$>$ 50 were considered for the GO analysis. From all GO terms with at least 5 genes and a pvalue of $<$ 0.05, the 10 GO terms with the highest gene counts are shown.

In summary, the diffReps analysis from sT-expressing nHDFs revealed substantial changes in the activating marks H3K4me3 and H3K27ac for genes that could be assigned to similar biological processes. Interestingly, a reduction in the signals of both histone modifications was observed for genes involved in innate immune signaling. As it was observed in the RNA-Seq data that sT reduced the transcription of type I IFN response genes, the next step was to analyze whether there was a correlation of these genes and possibly of other gene sets. Therefore, as in the previous chapters, the RNA-Seq data was correlated with the ChIP-Seq data for the different histone modifications (figure 28).

At 3 dpi, the RNA-Seq signal correlated with H3K4me3 signal to a similar extent in both donors, however the effect was more pronounced in Donor II. Concerning the H3K27me3 signal at 3 dpi, there were some genes with matching RNA-Seq and ChIP-Seq signal, however many DEGs were distributed all over the scatterplot and were not specifically modified with H3K27me3.

At 8 dpi, there was a general increase in the total numbers of genes with correlating RNA-Seq and ChIP-Seq signal in all the different conditions. For example, there was a very clear distribution of upregulated genes that had increased H3K4me3 signals in sT *vs* ctrl. Strikingly, H3K27ac signal, which was shown in the heatmaps to co-occur at promoters with H3K4me3 (figure 24), showed the same distribution and a correlation with the RNA-Seq signal in both donors. For H3K27me3 at 8 dpi, only in Donor I there was some correlation of downregulated genes with H3K27me3 signal, which was increased in sT *vs* ctrl. However, in Donor II, downregulated genes were distributed over the whole scatterplot. The same was observed for H3K9me3 in both donors, i.e. there was no pronounced co-occurence of histone modification changes and transcriptional changes. Nevertheless, for both H3K9me3 and H3K27me3, there were some upregulated genes in the scatterplots that displayed less signal comparing sT *vs* ctrl.

In summary, the scatterplots revealed a strong correlation of the RNA-Seq data with changes in the activating marks H3K4me3 and H3K27ac. For a small proportion of genes there was also an overlap with the repressive marks H3K27me3 and H3K9me3 observable, but to a much lower extent. In order to reveal if transcriptional and histone modification changes co-occurred for specific gene subset, the GO analysis from the RNA-Seq data was compared to changes in histone modification signals (figure 29).

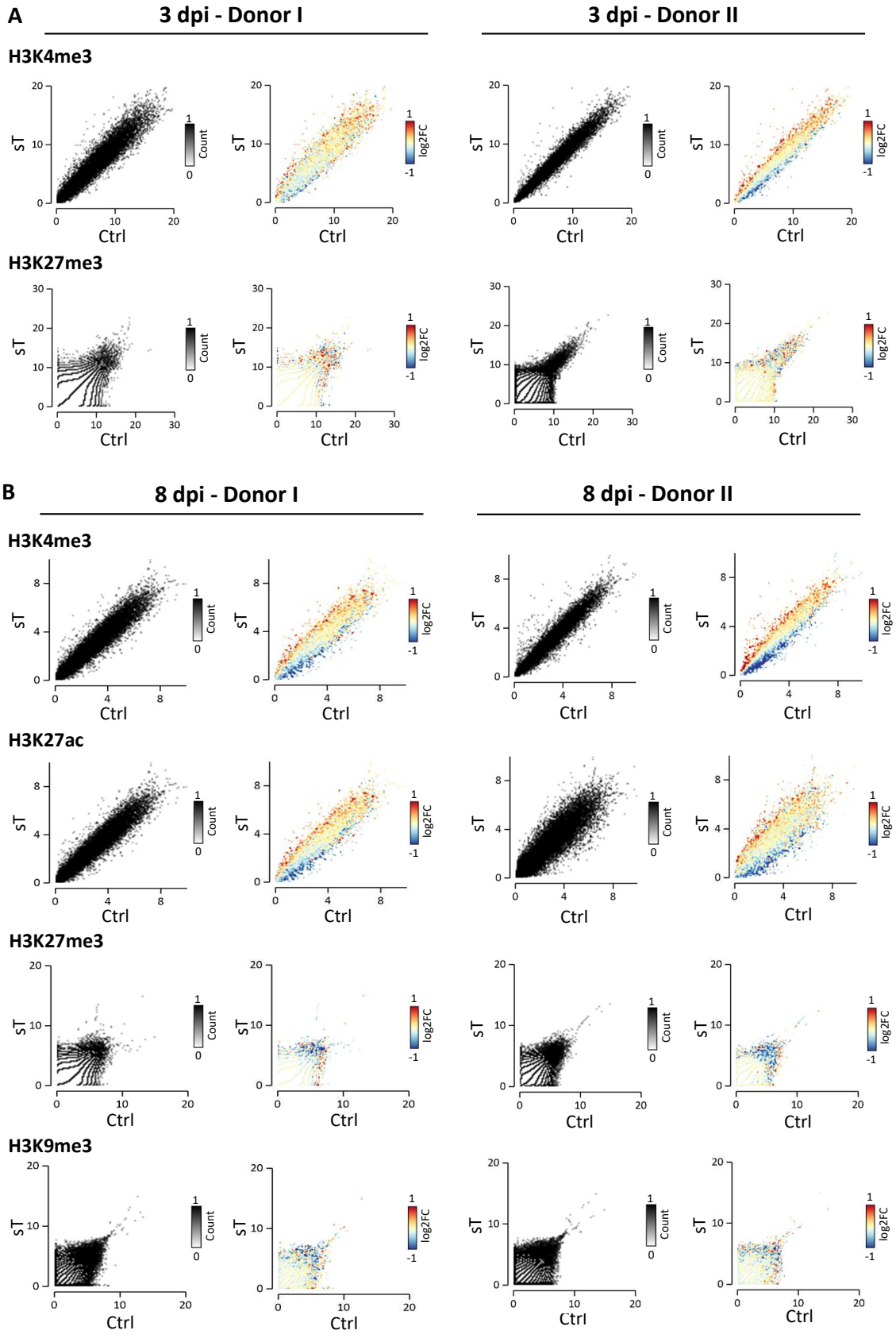


Figure 28 Correlation of ChIP-Seq with RNA-Seq analysis in sT-overexpressing nHDFs.
Quantification of H3K4me3 and H3K27me3 read signals for all genes from hg19 in nHDFs overexpressing ctrl or sT at 3 dpi (A) and of H3K4me3, H3K27me3, H3K27ac and H3K9me3 at 8 dpi (B). Histone modification signals were correlated with changes in gene expression, comparing sT *vs* ctrl. The color code refers to the log2FC of each gene that is plotted by its level of histone modification signal. The x and y-axes were segmented into 100 bins and regions within these bins are depicted by the counts.

Figure 29 summarizes the quantification of histone modification signals for a subset of genes that were contained in the most prominent GO terms from the RNA-Seq analysis described in figure 23B. At 3 dpi, H3K4me3 and H3K27me3 signals were measured in nHDFs expressing sT or the vector control in the two different donors, which is reflected by the boxplots. The total numbers of genes within the respective subsets is noted in the lower part of the boxplots.

There was a general decrease of H3K27me3 in all the subsets that did not match with the RNA-Seq signal, an effect that was observed already in the scatterplots. For H3K4me3, the changes between sT and ctrl were also very marginal and they varied between the two donors, thus at 3 dpi there was no strong overlap at the indicated gene promoters detectable. The barplots below confirmed these observations, as they indicate that the majority of genes that was up-or downregulated on the transcriptional level did not show an increase or decrease in H3K4me3 at 3 dpi. However, the subset *DNA replication and cell division* contained around one third of upregulated genes that also had increased H3K4me3 signals of at least one-fold.

At 8 dpi, there was a general overlap of transcriptional regulation and differential histone modifications and the differences between sT and ctrl were much stronger, at least regarding the activating marks H3K4me3 and H3K27ac. Interestingly, the genes that were downregulated and classified into three different subsets showed a very strong reduction in H3K4me3 signal comparing sT *vs* ctrl. These subsets included GO terms such as *Type I IFN response*, *Negative regulation of transcription* and *Cell adhesion and angiogenesis*. The barplots underneath confirm these observations by showing that over 50% of those genes that were downregulated also had a reduced signal in H3K4me3 of at least one-fold. These findings were also applicable to changes in H3K27ac, however to a lower degree. As shown in the bar plots, the H3K27ac signal was generally weaker than the H3K4me3 signal, and the difference in sT-expressing nHDFs *vs* ctrl was also less pronounced.

Concerning the two repressive marks, the signals for H3K27me3 were very low in both donors at 3 and 8 dpi and as already highlighted in the scatterplots, there was no correlation of H3K27me3 signal and transcriptional changes detectable. Similarly, for H3K9me3 there was only very little signal in general detectable around the promoters of the genes within the depicted subsets and little correlation with transcriptional regulation, reflecting the observations made in the scatterplots.

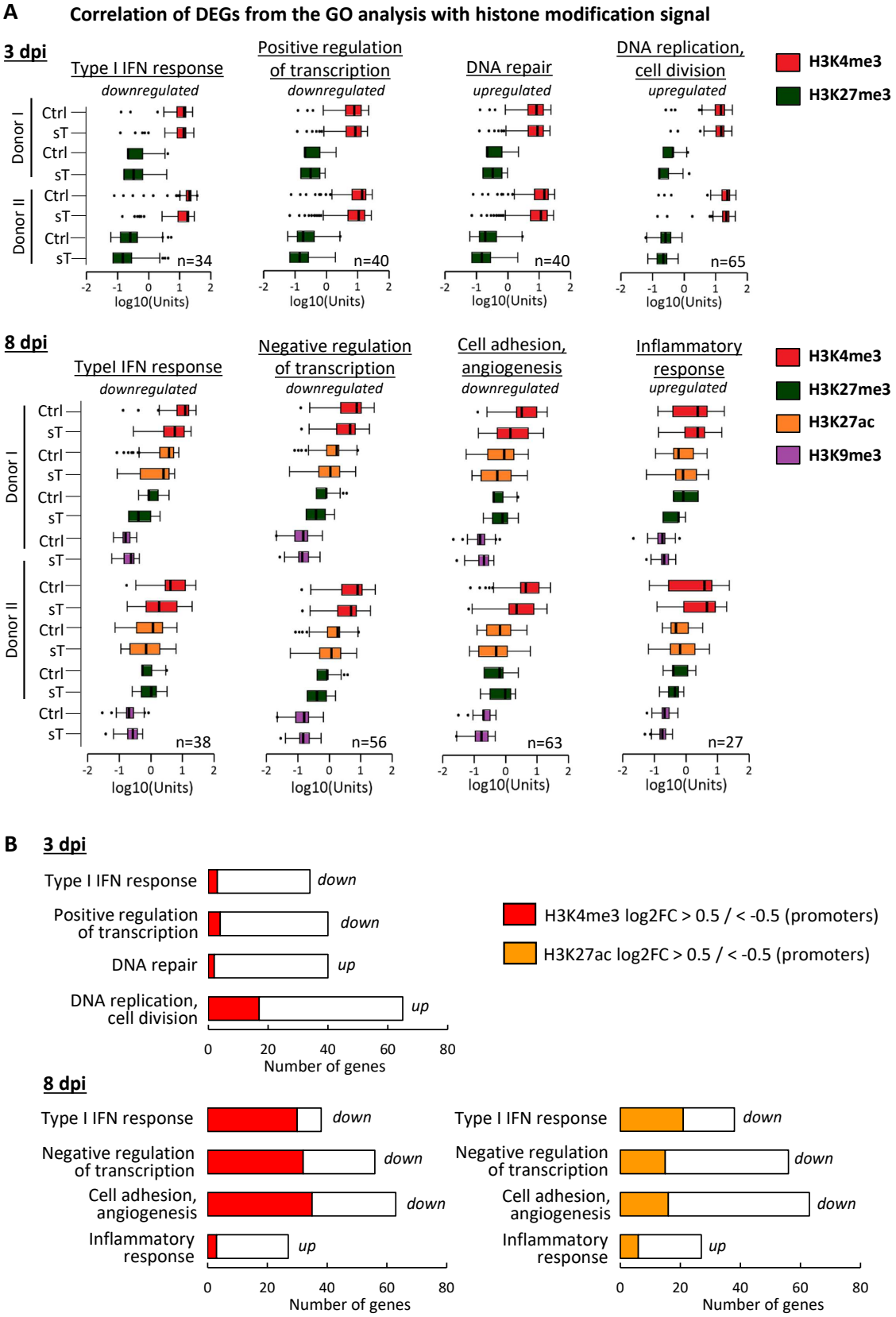


Figure 29 ChIP-Seq signal for a subset of genes from the RNA-Seq GO analysis in sT-expressing nHDFs.
(A) Boxplots representing the distribution of histone modification signals (H3K4me3 and H3K27me3) around the TSSs (<=3 kbp) of DEGs within the GO terms from the RNA-Seq analysis at 3 and 8 dpi. At 8 dpi, histone modification signals were also quantified for H3K27ac and H3K9me3. Data are shown from two different donors for ctrl and sT. (B) Bar plots representing the total numbers of DEGs found within the annotated GO terms and proportion of genes with differential promoter signal in H3K4me3 or H3K27ac (log2FC $>$ 0.5 or $<$ $-$0.5) at 3 and 8 dpi.

In summary, the H3K4me3 and H3K27ac patterns correlated strongly with transcriptional changes induced by sT and were also specific for certain gene subsets. As highlighted before, the type I IFN response represented an important biological function that was donwregulated by sT on the transcriptional level and correlated with decreased H3K4me3 and H3K27ac promoter signals.

4.4.3 MCPyV sT subverts the type I IFN response by transcriptional repression of IFN responsive genes

The experimental setting made use of a lentiviral transduction system which means that the nHDFs were exposed to a virus infection. Furthermore, FACS-sorting represents a process that might affect the cells in their normal behavior. To control for these possible effects, untreated cells were included in the RNA-Seq and ChIP-Seq analysis. A summary of the RNA-Seq and ChIP-Seq analysis can be found in the supplements (figure S4 - S7, table S2 and table S3). These results showed that there was a general deregulation of cellular processes such as cell proliferation, DNA replication or cell adhesion and ECM organization observable when comparing the vector controls to untreated cells. However, most transcriptional changes were not associated with significant changes in H3K4me3 or H3K27me3. An important observation was that among the genes that were upregulated by lentiviral transduction, there was a considerable enrichment of genes involved in the inflammatory and innate immune response, indicating that the cells responded to the lentivirus infection which was reflected on the transcriptional level. To compare which innate immune genes were induced by lentiviral transduction and which ones were specifically downregulated by MCPyV sT, the heatmaps in figure 30 summarize all innate immune genes that were significantly deregulated at least in one of the four conditions.

In total, 44 genes were classified into four different categories regarding their function. For example, some genes were classified into sensing of PAMPs. The *DDX60* gene, which codes for a *DEXD/H box RNA helicase*, was the only one that was significantly downregulated by sT at both 3 and 8 dpi, representing a gene that promotes RIG-I-like receptor-mediated signaling. Downstream of PRR signaling and IFN induction, *IRF7*, *IRF9*, *STAT1* and *STAT2*, all involved in *IFN signaling and regulation*, were significantly downregulated by sT. Except for *STAT2*, their transcription was increased upon lentiviral transduction compared to untreated cells, however this was only significant at 3 dpi. The same distribution was observed for the large subset of

Interferon stimulated genes, which were strongly induced by lentiviral transduction, however to a higher degree at 3 dpi. In contrast, upon sT expression they were mostly downregulated, an effect that was even stronger at 8 dpi compared to 3 dpi, confirming a specific regulation by sT.

Interestingly, an additional subset of genes was assigned to *Antigen presentation and processing*, including *HLA* genes, *HLA complex P5* (*HCP5*), or *transporter associated with antigen processing 1* (*TAP1*). While all of them were significantly downregulated by sT, lentiviral transduction *per se* did not induce their transcription, except for *HCP5* and *TAP1*. A summarizing list of all genes involved in the type I IFN response and in inflammatory signaling, that were observed in this work to be transcriptionally deregulated, is given in table S12.

The genome browser tracks in figure 30B depict seven representative genes that were downregulated by sT on the transcriptional level, and that also had decreased H3K4me3 and H3K27ac signals in their promoters. From the genes involved in PAMP recognition, *STING1* and *TLR3* were chosen as examples for recognition of cytosolic DNA or endosomal DNA, respectively. The genome browser tracks show that H3K4me3 and H3K27ac were not significantly reduced at their promoters, but *TLR3* showed a reduction at the later time points. *IRF7* and *IRF9*, both downregulated by sT and important regulators in IFN signaling, had significantly reduced H3K4me3 and H3K27ac promoter signals at the later time points. For *IRF9*, there was already a small reduction in H3K4me3 visible after 3 dpi. *OAS1* and *MX2* represent two examplary ISGs that were significantly downregulated by sT and that were also strongly reduced in H3K4me3 and H3K27ac promoter signal. Finally, as an example of a gene involved in antigen presentation, *HLA-B* showed a reduction in H3K4me3 and H3K27ac in its promoter and was also downregulated by sT on the transcriptional level.

These observations suggest that sT substantially interferes with type I IFN signaling, comprising a large variety of target genes, especially genes that are classified as ISGs. ISGs are produced in response to activation of the IFN signaling cascade which is initiated by a large variety of PAMPs that are presented during a viral infection, for example. Type I IFNs, including IFNα and IFNβ are recognized by the interferon alpha receptor (IFNAR1/2) which activates a specific signaling cascade (figure 31A). This finally leads to the activation of ISGs, from which a large number was strongly downregulated by sT. Thus, it is possible that sT interferes with either one or several steps in the type I IFN signaling cascade.

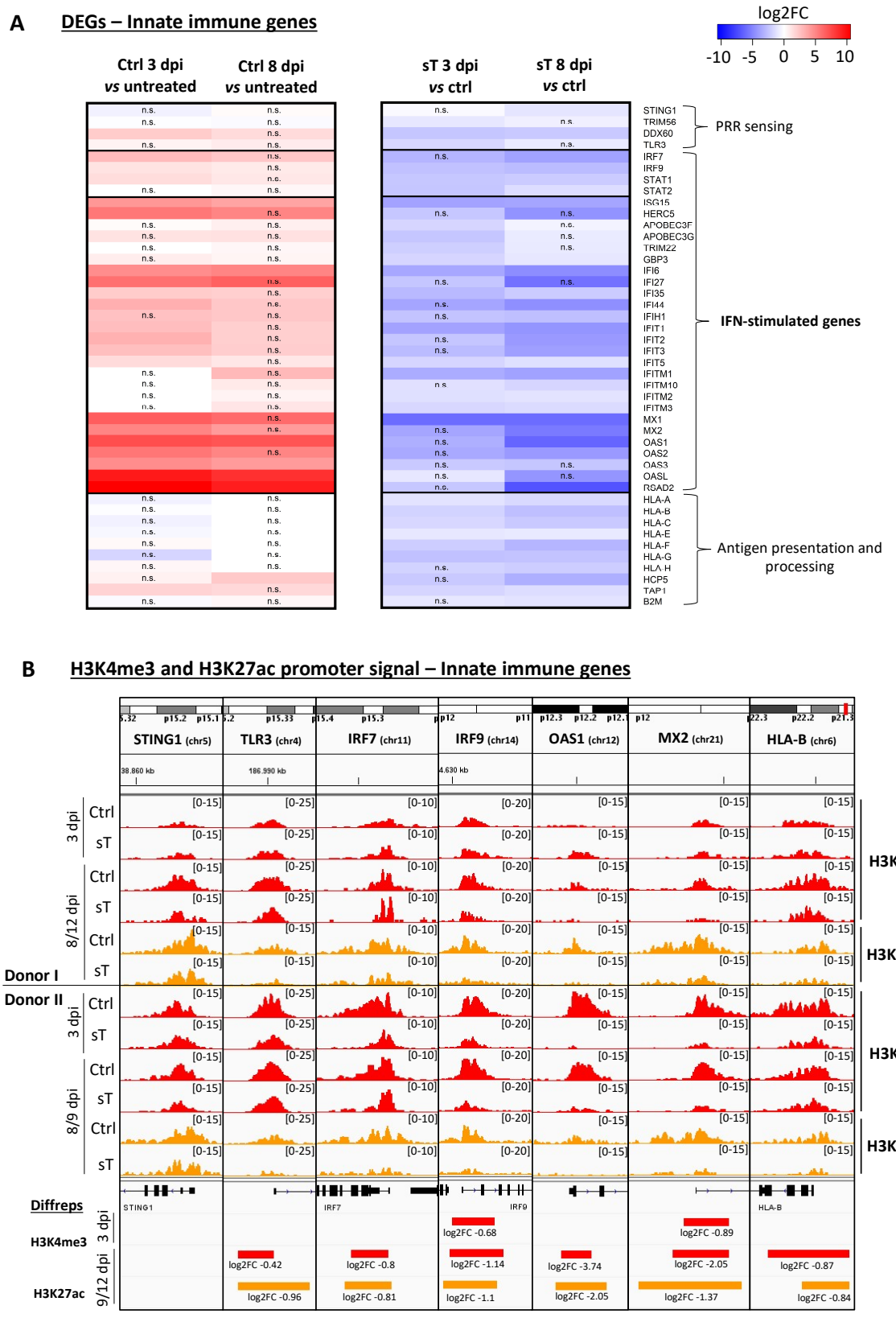


Figure 30 Downregulation of innate immune genes by sT correlates with H3K4me3 and H3K27ac signal.
(A) Heatmaps depicting changes in gene expression for a subset of genes involved in innate immunity, as indicated on the right side. Left heatmap shows the log2FCs from nHDFs expressing ctrl at 3 or 8 dpi compared to untreated cells. Right heatmap represents log2FCs from nHDFs expressing sT compared to ctrl at 3 and 8 dpi. (B) Genome browser tracks representing the read count signals for H3K4me3 and H3K27ac for representative innate immune genes selected from (A). Single tracks show ctrl- and sT-expressing cells at 3 and 8 dpi from the individual experiments performed in two different nHDF donors.

The possible points of interference by sT with the type I IFN signaling cascade are highlighted in figure 31A, and figures 31B-D summarize the results from the experiments performed to address the respective steps. First, it was confirmed by RT-qPCR that sT downregulated *OAS2*, *IRF9* and *STAT1* as exemplary ISGs. The two latter are also members of the ISGF3 complex, which is required for downstream promoter activation of ISGs. These three genes were shown by the RNA-seq data as strongly downregulated and also the mRNA levels were confirmed in four independent experiments to be significantly reduced in nHDFs overexpressing sT compared to the vector control at 2 and 8 dpi (figure 31B). In addition to transcriptional downregulation, the ChIP-Seq analysis revealed that H3K4me3 and H3K27ac signals were strongly reduced in the promoter regions of many of these genes (see figure 29 and 30).

Subsequently, it was analyzed if sT interfered with the protein composition of the ISGF3 complex or with phosphorylation of STAT1 (pSTAT1), which is required for the activation of this complex. The results from the WB analysis of IRF9, STAT1 and pSTAT1 are highlighted in figure 31C. A quantification of the protein levels was performed in three independent experiments using nHDFs expressing sT or the vector control and representative blots are shown from one experiment. Interestingly, sT only induced a small reduction in the protein levels of pSTAT1, but not of total STAT1 levels at 2 dpi, however this was not significant. Importantly, IRF9 protein levels were consistently decreased by sT expression with a strong reduction from 2 to 8 dpi.

The next step in the signaling cascade is the translocation of the ISGF3 complex to the nucleus where it binds to ISRE motifs which are located in the promoters of ISGs. It is thus a crucial step for their transcriptional activation. As the protein levels of IRF9 were only reduced to a low extent at 2 dpi, IFN stimulation possibly leads to a normal formation and activation of the ISGF3 complex. Therefore, it was analyzed if initial nuclear localization was impaired by sT (figure 31D). The IF analysis showed that while in untreated cells there was a weak IRF9 signal in the nucleus, nHDFs that were transduced with the vector control showed an increased nuclear signal. However, in sT-expressing nHDFs, the IRF9 signal was very weak and comparable to the signal in untreated cells. The overlap of IRF9 signal and nuclear DAPI staining was quantified in three independent experiments for sT *vs* ctrl at 3 dpi. The Pearson correlation coefficient, which measures the signal correlation of DAPI and IRF9-staining, was significantly reduced in sT-expressing nHDFs compared to the vector control, highlighting that sT reduced the availability of IRF9 in the nucleus.

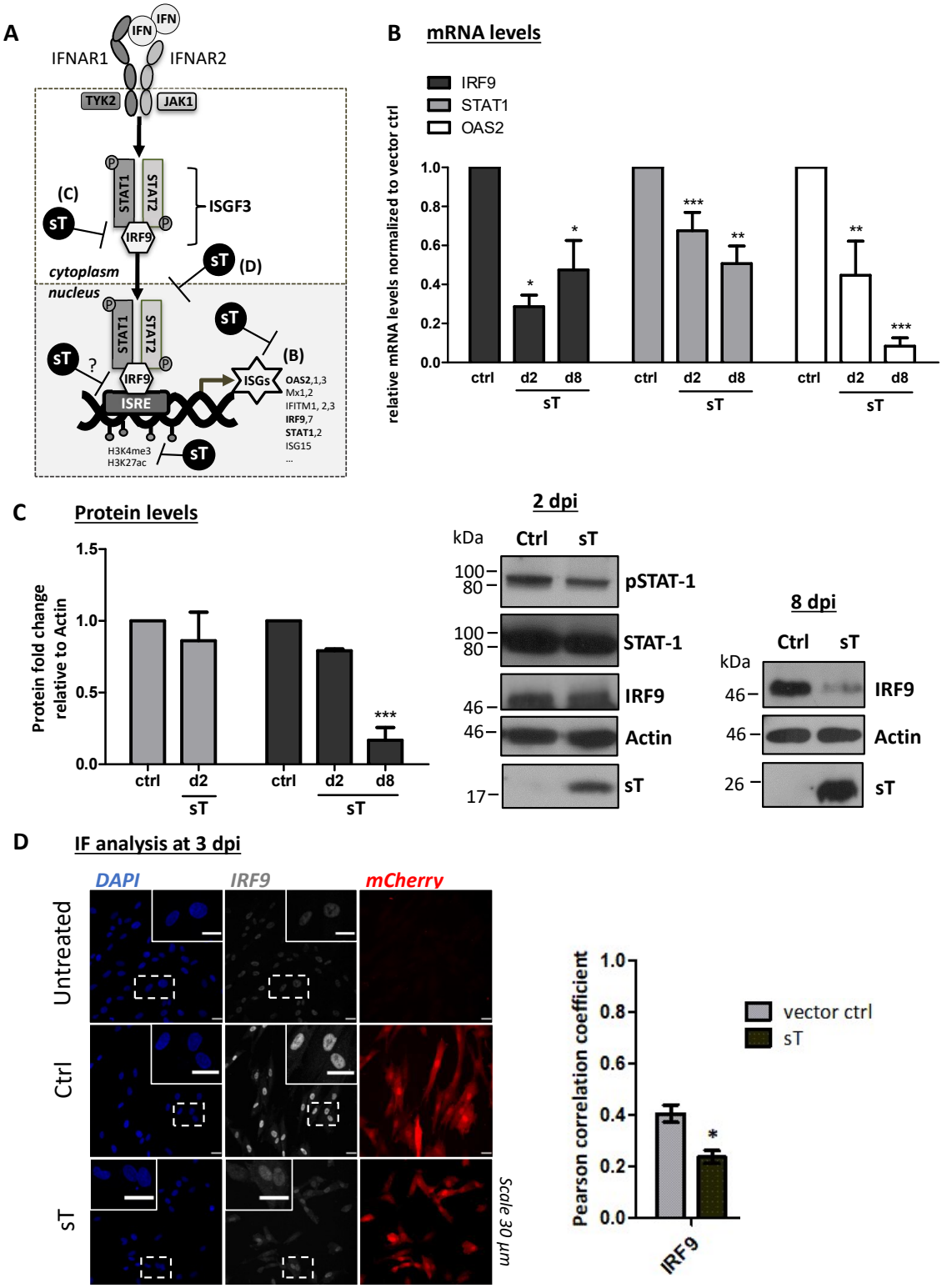


Figure 31 sT interferes with several steps in the type I IFN signaling cascade.
(A) Overview of IFN signaling and possible steps of interference by sT. Upon stimulation of the IFNAR1/2 receptor with type I IFNs, via JAK1 and TYK2 activation, the ISGF3 complex (pSTAT1, pSTAT2, IRF9) translocates to the nucleus and binds to ISREs in the promoters of ISGs, thereby activating their transcription. (B) Confirmation of sT-induced downregulation of transcription of selected IFN regulatory and stimulated genes: *IRF9*, *STAT1* and *OAS2*. mRNA levels were measured by RT-qPCR in sT-or ctrl-expressing nHDFs at 2 (pre-sort) and 8 dpi (sorted; n=4, except for *OAS2* at 8 dpi, n=3). (C) Western Blot analysis of (p)STAT1 and IRF9 at 2 (pre-sort) and 8 dpi. Representative blots are shown on the right side from sT- and ctrl-expressing cells, using antibodies against pSTAT1, STAT1, IRF9 and sT (2T2). Quantification was performed using actin as a loading control (n=3). (D) Immunofluorescent staining of IRF9 in nHDFs expressing sT or ctrl at 3 dpi. Exemplary images are shown from untreated nHDFs expressing ctrl or sT at 3 dpi (Donor II), showing DAPI as a nuclear stain, IRF9-staining and mCherry which marks sorted cells. Pearson correlation coefficient was determined for ctrl- or sT-expressing cells (n=3). Statistical significance was assessed by two-way ANOVA (*$p < 0.05$, **$p < 0.001$, ***$p < 0.001$).

In summary, the observation that sT represses the transcription of a large variety of type I IFN response genes highlights the possible interference of sT with one or several IFN regulatory genes. Importantly, sT had a substantial effect on the IFN signaling cascade and reduced the transcription, protein production and nuclear availability of IRF9 to a large extent. Whether sT interferes with the binding of ISGF3 to the ISRE motif in the IRF9 promoter remains to be determined.

4.5 Profiling of nHDFs expressing MCPyV sT together with LT or LT*tr*

In the previous chapters, MCPyV T antigens were individually overexpressed in order to determine their effects on the host transcriptome and epigenome. Due to a lack of a suitable infection model to date, it remains unclear how exactly MCPyV infection can perturb host cellular processes to enable infection and establish persistence, but also in the context of tumorigenesis and metastasis, many open questions remain. Therefore, in order to mimic the two different scenarios - persistent infection *vs* tumorigenesis - the subsequent experiments were performed in nHDFs overexpressing sT together with the full length LT, reflecting the first scenario, or sT plus the truncated LT, thereby reflecting a tumor setting.

4.5.1 Transcriptional changes induced by sT+LT and sT+LT*tr*

NHDFs were co-transduced with a combination of sT and either LT or LT*tr*. sT+LT overexpression was used as a model to study how both T antigens contribute to infection and persistence establishment by analyzing their effects on the transcriptome and epigenome. RNA-Seq was performed early after 3 dpi in order to detect changes that might immediately occur in response to sT and LT overexpression. In addition, sT+LT*tr*, which were co-expressed to mimic a tumor scenario, were assessed for their effects on the host's transcriptome one passage later, referring to 9 dpi in Donor II and 12 dpi in Donor I. These later time points were chosen because nHDFs co-overexpressing both proteins were shown to regulate cancer signaling processes rather at the later time points. The volcano plots in figure 32A depict all DEGs comparing sT+LT or sT+LT*tr* to the respective vector controls.

At 3 dpi, 236 genes were significantly downregulated, while 302 genes were upregulated in nHDFs overexpressing sT+LT. Similarly, 302 genes were significantly downregulated and 286 were upregulated in nHDFs overexpressing sT+LT*tr*. Genes that are labeled in the volcano plots represent the genes with the highest log2FCs and lowest padj. values. Strikingly, many of these genes had considerably high log2FC values close to or even higher than 10. Further information on the genes labeled in the volcano plots is given in the supplements (table S10 and S11). In addition, a GO analysis using the DAVID tool was performed as previously described with all significantly up-and downregulated genes, which is summarized in figure 32B.

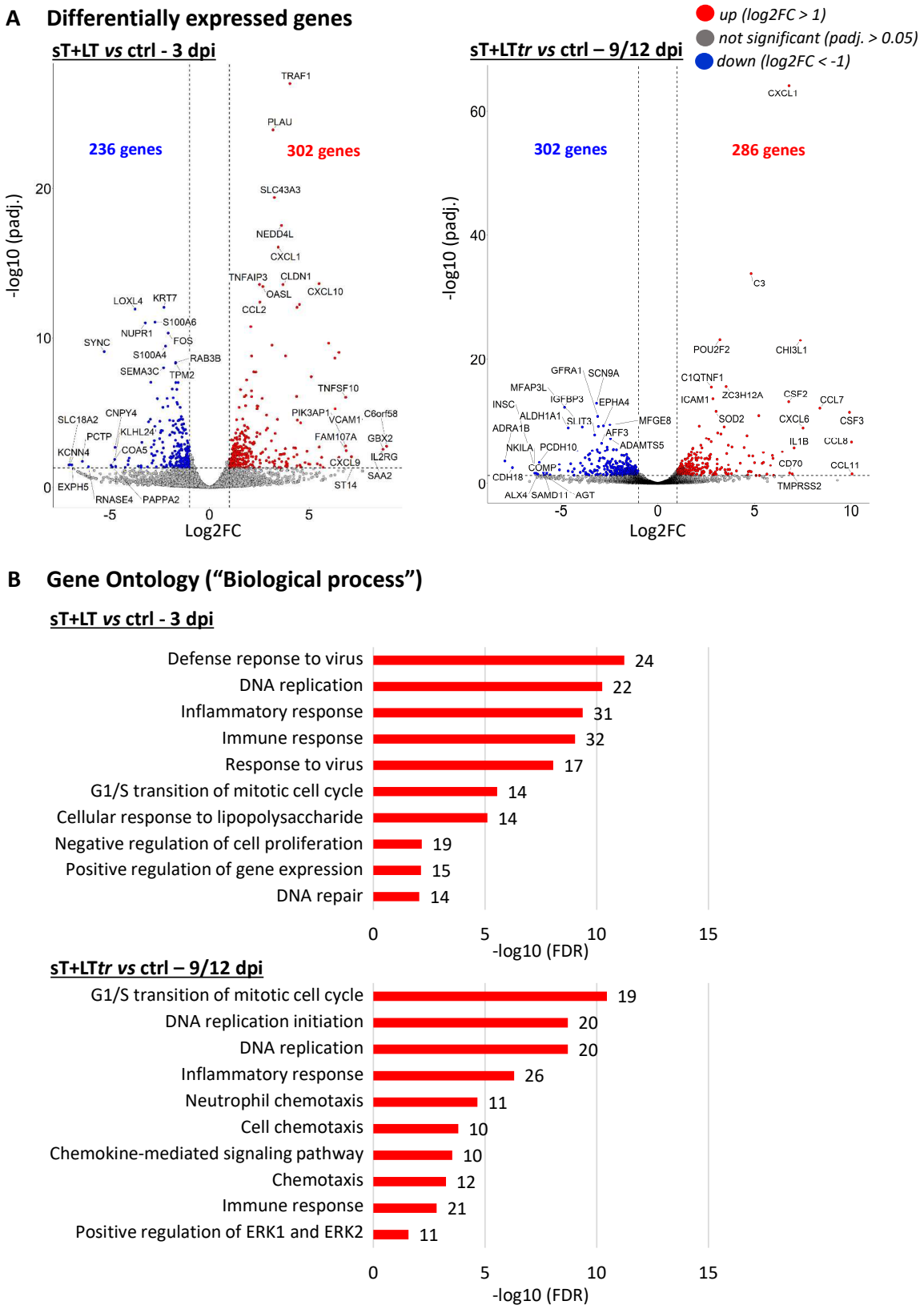


Figure 32 Transcriptional changes induced by co-overexpression of sT with LT or LT*tr*.
(A) Volcano plots depicting all DEGs from the RNA-Seq analysis in nHDFs co-expressing sT plus LT or LT*tr* compared to the vector control at 3 or 9/12 dpi. Genes with a log2FC > 1 or < -1 and a padj. value < 0.05 are denoted as significantly upregulated or downregulated, respectively. (B) A GO analysis ("Biological process") was performed using the DAVID tool with significantly upregulated and downregulated genes from the RNA-Seq analysis shown in (A). The 10 GO terms (FDR < 0.05, gene counts >= 10) containing the highest numbers of DEGs are depicted as red or blue bars, reflecting up- or downregulated genes as input, respectively. The number of DEGs found within each GO term is denoted next to the bars, FDR values are shown on the x-axis.

At 3 dpi, sT+LT expression in nHDFs increased the transcription of genes belonging to the term *Immune response*, including several ISGs such as *OASL*, *IFIT1* or *IFIT3*. Furthermore, *Inflammatory response* genes such as *CXCL1*, *CXCL9*, *CXCL10* or the *interleukin 2 receptor subunit gamma* (*IL2RG*) were included in these GO terms. Many of those genes were also highly upregulated and are labeled in the volcano plot. In addition, GO terms such as *Cell proliferation*, *DNA replication* or *DNA repair* were enriched, as observed also previously upon expression of either sT or LT. Concerning downregulated genes, there were some interesting genes that are labeled in the volcano plot that are involved in tumor-associated processes such as the *FOS* proto-oncogene, *RAB3B* gene, a member of the RAS oncogene family, or genes involved in transcriptional regulation such as the *nuclear protein 1* (*NUPR1*).

Overexpression of sT+LT*tr* in nHDFs also led to a large enrichment of upregulated genes that were assigned to similar biological processes when performing a GO analysis with DAVID. These terms included *DNA replication* and *G1/S transition of mitotic cell cycle*, including *CDC* genes, *minichromosome maintenance complex* (*MCM*) components or DNA polymerase subunits. In addition, the GO term *Inflammatory response* was highly enriched including a large amount of cytokines and chemokines from which many genes are also labeled in the volcano plots because of their high log2FCs. Some examples include *CXCL1*, *CXCL2*, *CXCL5*, *CXCL6* and *CXCL8*, as well as *IL1B*, *IL6*, *CCL2*, *CCL7*, *CCL8* and *CCL20*. Many of these genes were also enriched in five other GO terms which specifically suggests a role of sT+LT*tr* in chemotactic processes.

Concerning downregulated genes, although there was a large amount detectable in the volcano plots, they did not cluster together in significantly enriched GO terms. Nevertheless, some interesting and highly downregulated genes included *cadherin 5* (*CADH5*) or *ADMAMTS5*, thus representing genes involved in cell adhesion and ECM organization.

In summary, co-overexpression of sT with LT or LT*tr* similarly induced the transcription of cell cycle and DNA replication genes. For sT+LT, a substantial amount of genes involved in the innate immune response and inflammatory response appeared highly upregulated at 3 dpi. Overexpression of sT+LT*tr* at 9/12 dpi specifically induced the transcription of genes involved in chemotaxis and cellular signaling pathways, but also inflammatory signaling.

4.5.2 Histone modification changes induced by sT+LT and sT+LT*tr*

MCPyV T antigens were all shown to induce substantial transcriptional changes that partly correlated with changes in histone modification signal. Therefore, nHDFs overexpressing a combination of sT and LT or sT and LT*tr* were also assessed for changes in histone modification patterns. For sT+LT, H3K4me3 was analyzed at 3 dpi and for sT+LT*tr*, both H3K4me3 and H3K27me3 ChIPs were performed at the later timer point, i.e. at 9/12 dpi. The heatmaps in figure 33 reflect a dense signal of H3K4me3 around TSSs which was anti-correlative to the H3K27me3 signal. The signal distribution appeared similar in all the different conditions with the exception of sT+LT at 3 dpi, for which the H3K4me3 signal was increased for several genes around the TSSs.

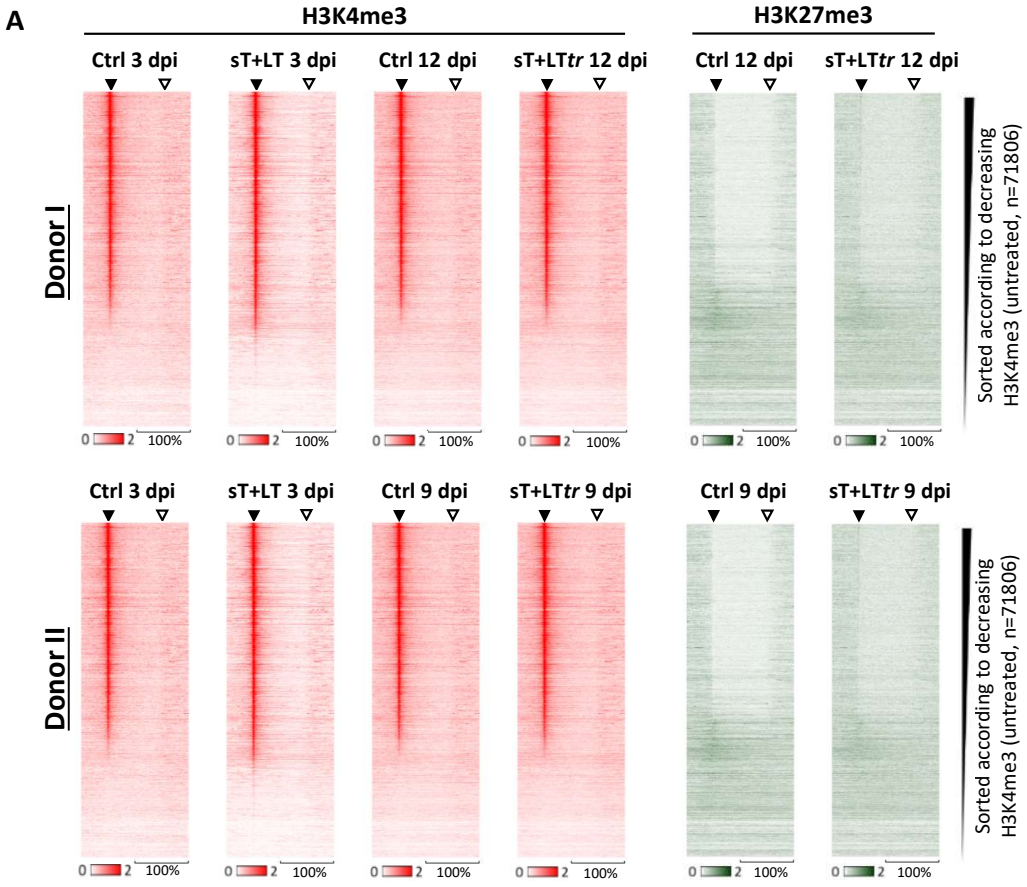
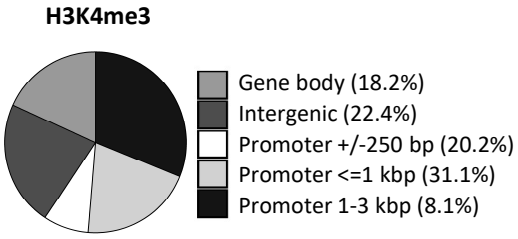


Figure 33 ChIP-Seq analysis in nHDFs overexpressing sT together with LT or LT*tr*.
(A) Heatmaps depicting ChIP-Seq read count signals quantified for all genes of hg19 for H3K4me3 and H3K27me3 at 3 and 8 dpi from nHDFs overexpressing sT+LT or ctrl at 3 dpi and sT+LT*tr* or ctrl at 9/12 dpi. 200% of the genomic loci are shown as relative gene lengths, including TSSs (black triangles) and TTSs (non-filled triangles). Heatmaps are sorted according to decreasing H3K4me3 signal in untreated cells. Density represents DNA fragments per 1 mio reads per 1 kbp.

In order to detect histone modification changes for specific genes, a diffReps analysis was performed as described in the previous chapters. A summary of the distribution of genomic features for the diffReps hits and the total numbers of hits and annotated genes with different log2FC cutoffs is shown in figure 34. For sT+LT at 3 dpi, nearly one third of all hits were assigned to promoter regions and with 3366 hits with a log2FC > 1, a large amount of genes showed an increased signal in H3K4me3 in promoter regions or gene bodies. In contrast, only 25 genes were associated with a decrease in H3K4me3 in promoter regions or gene bodies.

A **sT+LT at 3 dpi - Distribution of genomic features from the diffReps analysis**

(log2FC > 0.5 or < -0.5, padj. < 0.05)

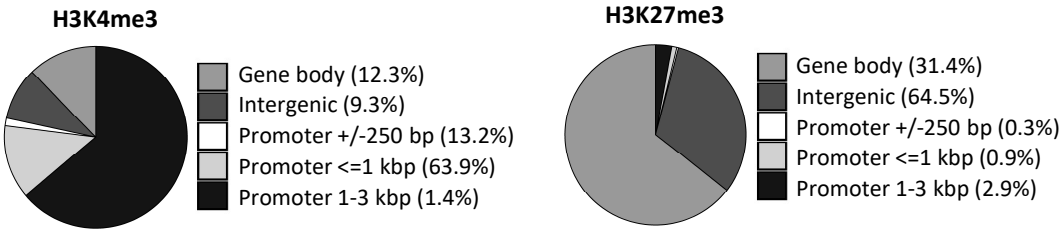


DiffReps hits

diffReps	H3K4me3 total	H3K4me3 for annotated genes
log2FC > 0.5	14918	9365
log2FC < -0.5	1084	747
log2FC > 1	5845	**3366**
log 2FC < -1	109	**25**

B **sT+LT*tr* at 9/12 dpi - Distribution of genomic features from the diffReps analysis**

(log2FC > 0.5 or < -0.5, padj. < 0.05)



DiffReps hits

	H3K4me3		H3K27me3	
diffReps	**total**	**annotated genes**	**total**	**annotated genes**
log2FC > 0.5	3219	2874	1123	**335**
log2FC < -0.5	756	569	650	**180**
log2FC > 1	104	**82**	17	3
log 2FC < -1	90	**68**	6	2

Figure 34 Differential histone modification analysis in nHDFs overexpressing sT+LT or sT+LT*tr*.
(A) Results from the diffReps analysis performed in nHDFs overexpressing sT+LT (A) or sT+LT*tr* for H3K4me3 and H3K27me3. Distribution of diffReps hits (log2FC $> 0.5/< -0.5$) according to genomic features is depicted in the pie charts. The tables summarize the total numbers of diffReps hits with different log2FC cutoffs and depict the amount of hits with annotated genes (at promoters or gene bodies). Numbers in bold represent genes that were used for downstream analysis.

Concerning sT+LT*tr* at 9/12 dpi, the majority of H3K4me3 diffReps hits was located in promoter regions, however most of them (63.9%) were located within a distance of 1-3 kbp around the TSS. In total, 82 genes had a log2FC > 1 and 68 genes < -1. For H3K27me3, most of the diffReps hits were located within gene bodies or were intergenic. In addition, most of the annotated hits were only one-fold increased or decreased, representing 335 and 180 genes in total, respectively. Due to the fact that these differences were very small and probably not relevant with regard to biological processes, a GO analysis can be found in the supplements (figure S12).

In order to determine the biological relevance for changes observed in H3K4me3 signal, a GO analysis was performed with the genes that were detected to have increased H3K4me3 in their promoters of at least two-fold (figure 35). The average plots show the increased or decreased H3K4me3 promoter signals and the total numbers of genes that were considered are noted next to the plots. As only gene sets containing a minimum of 50 genes were considered, the GO analysis was only performed with genes that displayed increased H3K4me3 promoter signal in both conditions. The bar plots in figure 35B reveal that genes with increased H3K4me3 in nHDFs expressing sT+LT at 3 dpi were strongly enriched in GO terms reflecting transcriptional regulation and signal transduction. Interestingly, the GO analysis from nHDFs expressing sT+LT*tr* showed a considerable overlap with the GO terms from the RNA-Seq analysis. Again, many genes involved in *Chemotaxis* as well as the *Inflammatory response* and *Immune response* were enriched.

With the aim to specifically compare transcriptional changes with changes in H3K4me3 in both conditions, the RNA-Seq and ChIP-Seq data from sT+LT and sT+LT*tr* were correlated with each other, as summarized in the scatterplots in figure 36. For sT+LT at 3 dpi, the scatterplots show that there was some correlation of H3K4me3 and transcriptional changes observable. However, a large number of genes was distributed over the scatterplot without showing a strong correlation. In contrast, the correlation was more pronounced for sT+LT*tr*. Concerning H3K27me3, which was only assessed in nHDFs expressing sT+LT*tr*, there was a specific set of downregulated genes with increased H3K27me3 in Donor II, while in Donor I there was no distinct correlation observable.

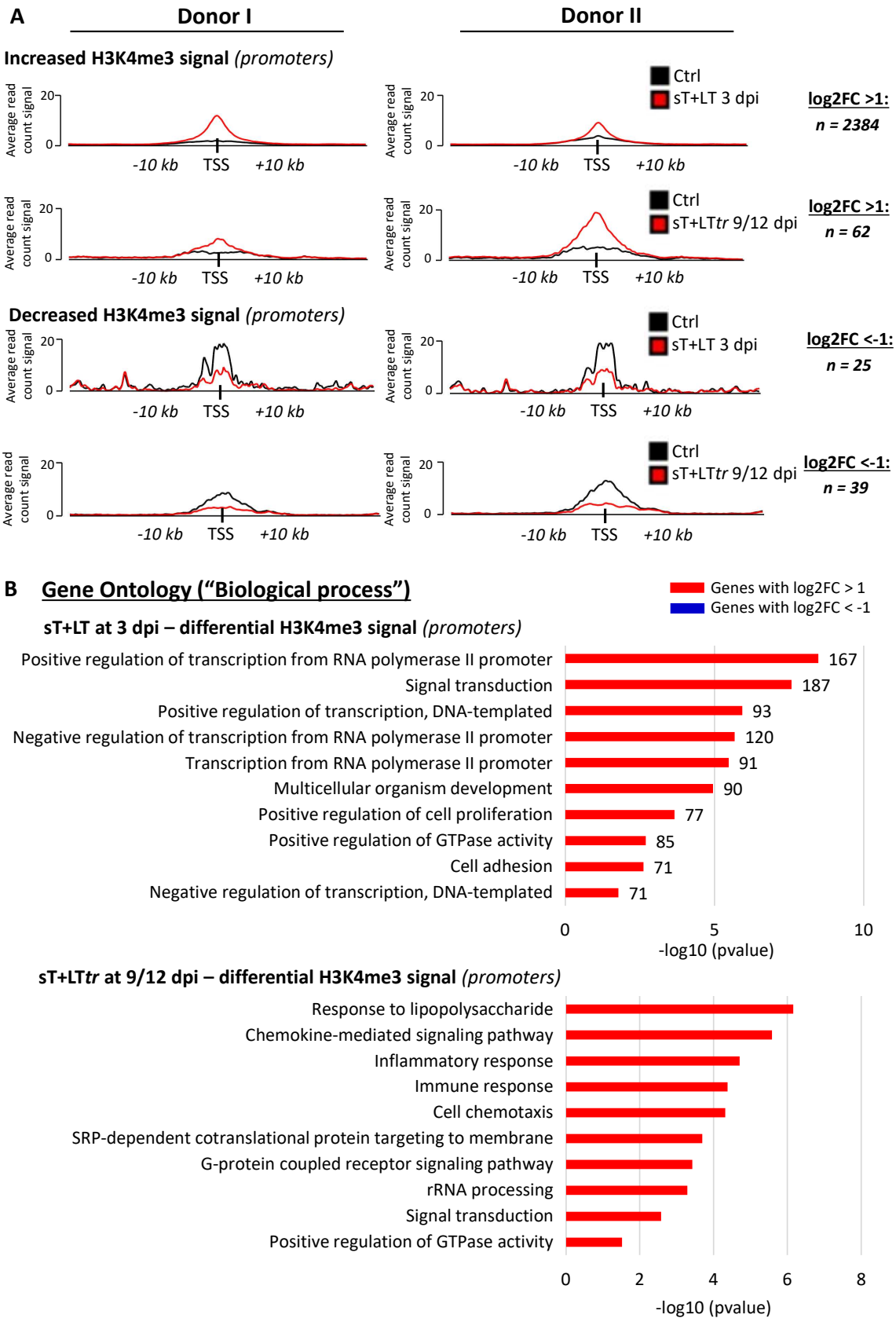


Figure 35 Analysis of changes in H3K4me3 induced by overexpression of sT+LT or sT+LT*tr*.
(A) Average read count signals of H3K4me3 for all genes with differential promoter signal comparing sT+LT *vs* ctrl at 3 dpi or sT+LT*tr* *vs* ctrl at 9/12 dpi. Total numbers of genes with differential H3K4me3 signal around promoters (+/- 3kbp) of annotated genes are shown next to the average plots. (B) A GO analysis ("Biological process") was performed using DAVID with genes from the diffReps analysis described in (A). Only gene sets with n$>$ 50 were considered for the GO analysis. From all GO terms with at least 5 genes and a pvalue of $<$ 0.05, the 10 GO terms with the highest gene counts are shown.

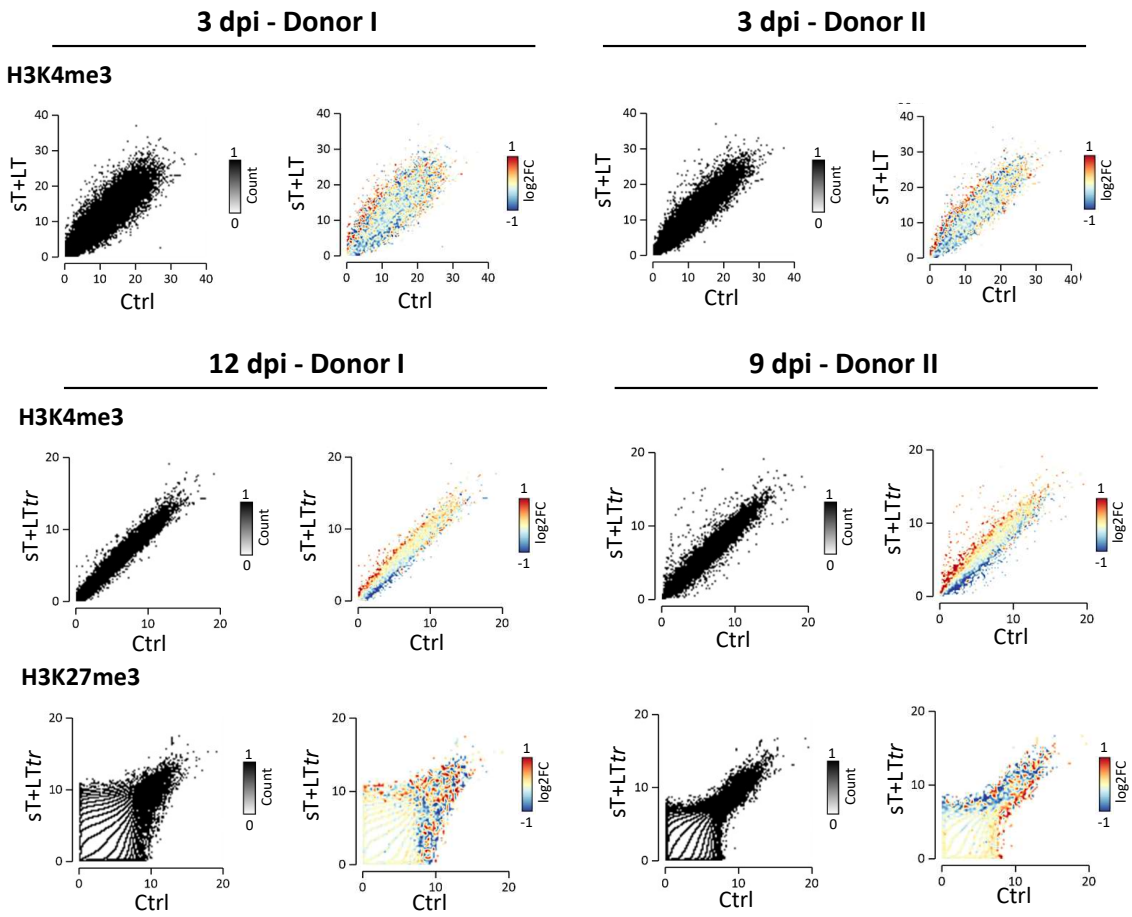


Figure 36 Correlation of ChIP-Seq with RNA-Seq analysis in nHDFs co-expressing sT and LT or LT*tr*. Quantification of H3K4me3 and H3K27me3 read signals for all genes from hg19 in nHDFs overexpressing ctrl, sT+LT or sT+LT*tr*. Histone modification signals were correlated with changes in gene expression from the RNA-Seq analysis. The color code refers to the log2FC of each gene that is plotted by its level of histone modification signal. The x and y-axes were segmented into 100 bins and regions within these bins are depicted by the counts.

The scatterplots highlight that RNA-Seq and ChIP-data generally showed a correlation and thus it was analyzed whether genes identified in the RNA-Seq GO term analysis as highly enriched, also differed in histone modification patterns. In figure 37A, this was assessed for nHDFs expressing sT+LT at 3 dpi. Although the average read count signals represented by the box plots did not differ much comparing sT+LT *vs* ctrl, the majority of genes that were upregulated and classified in the four depicted gene subsets also had increased H3K4me3 signals in their promoters. Concerning sT+LT*tr*, the box plots highlight that the H3K4me3 promoter signal correlated well with the transcriptional regulation of the genes contained in the depicted subsets. All of them had an increased H3K4me3 average signal and as depicted in the bar plots, half of all upregulated genes had at least a one-fold increase of H3K4me3 promoter signal. In contrast, H3K27me3 signal was generally very low and did not correlate with transcriptional changes, as it was observed also in the scatterplots.

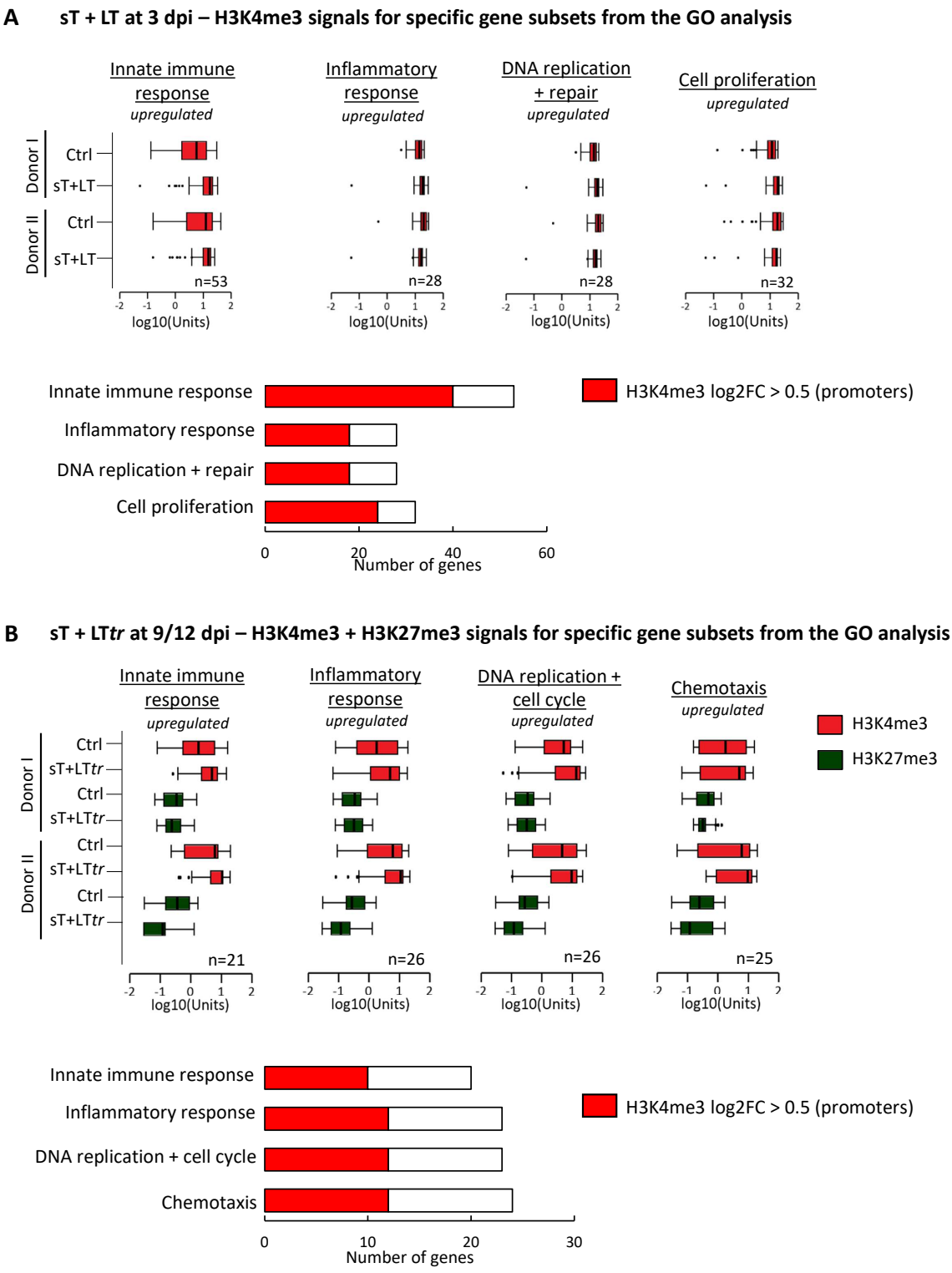


Figure 37 ChIP-Seq signals for a subset of genes from the RNA-Seq GO analysis in nHDFs overexpressing sT+LT or sT+LT*tr*

(A) Boxplots representing the distribution of histone modification signals (H3K4me3) around the TSSs (<= 3kbp) of DEGs contained in the GO terms from the RNA-Seq analysis in nHDFs overexpressing sT+LT at 3 dpi. Bar plots underneath depict the proportion of upregulated genes within denoted gene subsets with increased H3K4me3 signal (log2FC > 0.5). (B) At 9/12 dpi, histone modification signals for H3K4me3 and H3K27me3 were measured for gene subsets that were enriched in the RNA-Seq GO analysis in nHDFs expressing sT+LT*tr*. Total numbers of DEGs found within the annotated GO terms and the proportion of genes with significant differential signal in H3K4me (log2FC > 0.5) are shown in the bar plot below.

4.5.3 MCPyV T antigens have opposing effects on the type I IFN response

Co-overexpression of sT with LT or LT*tr* had considerable effects on the host transcriptome and H3K4me3 patterns. In particular for sT+LT*tr* the upregulation of genes involved in cell chemotaxis and cancer-related signaling pathways was a very striking observation. An interesting observation for sT+LT was the high quantity of genes with a differential H3K4me3 signal and the high enrichment of genes involved in transcriptional regulation.

As a common feature, in both conditions a lot of changes were detected regarding innate immune regulation and signaling. Due to the fact that the innate immune response appeared as a common target of all MCPyV T antigens, the heatmap in figure 38 summarizes the transcriptional changes induced by all T antigens regarding a subset of genes involved in innate immune regulation and signaling. The heatmap highlights that while both LT and LT*tr* strongly enhanced the transcription of type I IFN response genes at 3 dpi, sT downregulated the majority of them. At the late time point, the effect was even enhanced in sT-expressing nHDFs, meaning that a stable transcriptional repression was observed. In contrast, the initial enhancement of the type I IFN response by both LT antigens was not significant anymore at the later time point. Further, when overexpressing LT*tr* together with sT, these genes were not significantly deregulated at 9/12 dpi. In contrast, when overexpressing sT+LT and analyzing the transcriptional response at 3 dpi, the induction of the type I IFN response genes was less pronounced and mostly not even significant compared to LT expression alone.

In addition to the fact that MCPyV T antigens had opposing effects on the transcription of type I IFN response genes, for most of them there was also a correlation with H3K4me3 signal observable, as shown in figure 39. The genome browser tracks depict the H3K4me3 signal from nHDFs overexpressing the T antigens. Exemplary genes from the heatmaps from figure 38 are shown for all the different conditions in both donors. Figure 39A depicts the tracks from several important type I IFN regulatory factors. Interestingly, for *IFNβ1*, which was upregulated at 3 dpi by LT and LT*tr*, no H3K4me3 signal was detected at any site of the gene in none of the conditions. For all other depicted genes, sT+LT-expressing nHDFs showed very high H3K4me3 signals, as reflected by the increased signal range of 0-50. In contrast, except for *IRF3* and *STAT2*, sT induced the reduction of H3K4me3 levels compared to the vector control.

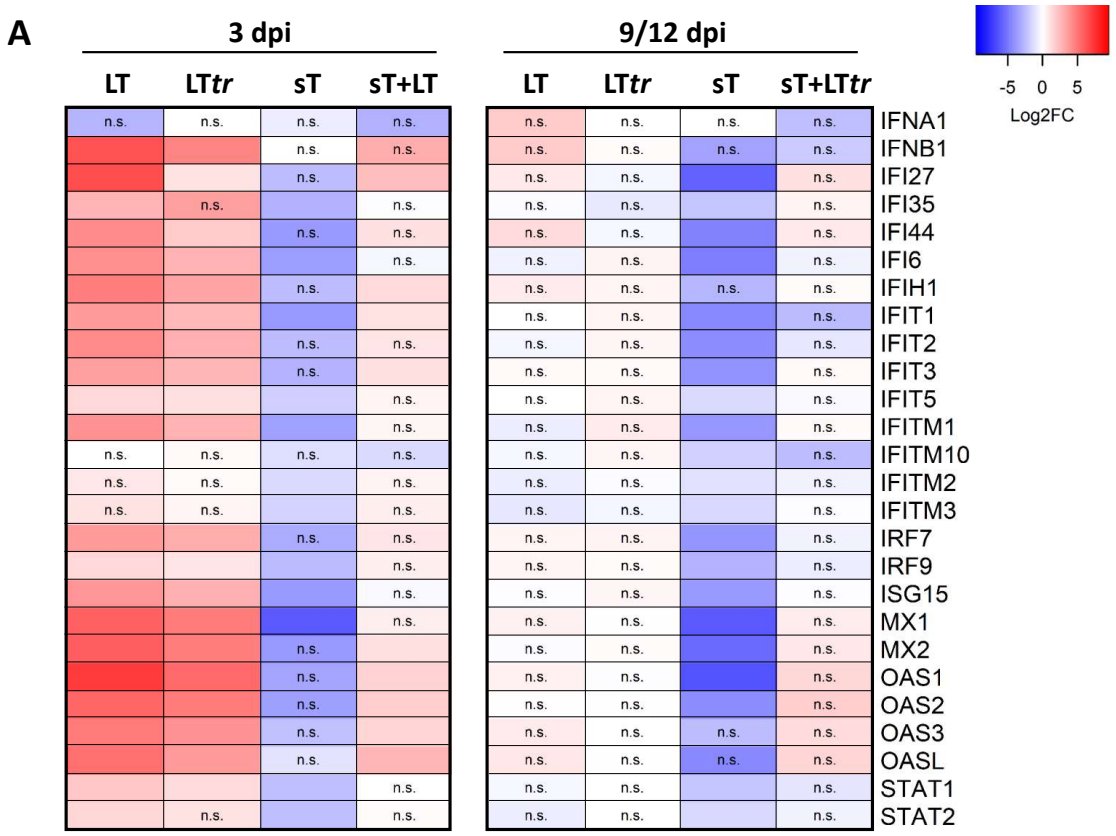


Figure 38 MCPyV T antigens have opposing effects on the transcription of type I IFN response genes
Heatmaps summarizing the results from the RNA-Seq analysis from nHDFs overexpressing the MCPyV T antigens, either individually or in combination, compared to the respective lentiviral vector controls. The color code refers to the log2FCs and n.s. is denoted for genes with a padj. value > 0.05 (not significant).

At the later time point, a similar picture was observed, as sT expression in nHDFs reduced the H3K4me3 levels for all genes except for *IRF3* and *STAT2*, while LT expression, especially in Donor I, led to increased H3K4me3 levels for all genes except for *IRF3* and *STAT2*. Similarly, for LT*tr*, increased H3K4me3 levels were observed around the promoters of *IRF7*, *IRF9* and *STAT1* compared to the vector control, which was also more pronounced in Donor I. Interestingly, the H3K4me3 signal in nHDFs expressing sT+LT*tr* was slightly decreased in comparison to LT*tr* alone, but still considerably higher compared to sT alone.

In summary, MCPyV T antigens specifically induced changes in the transcription and H3K4me3 patterns of genes involved in the type I IFN response with opposing effects comparing sT to LT or LT*tr*. Interestingly, the induction observed by both LT antigens was dampened when sT was co-overexpressed. However, only when expressed alone, sT was able to significantly repress a large number of ISGs and specifically *IRF9* as a master regulator of type I IFN signaling.

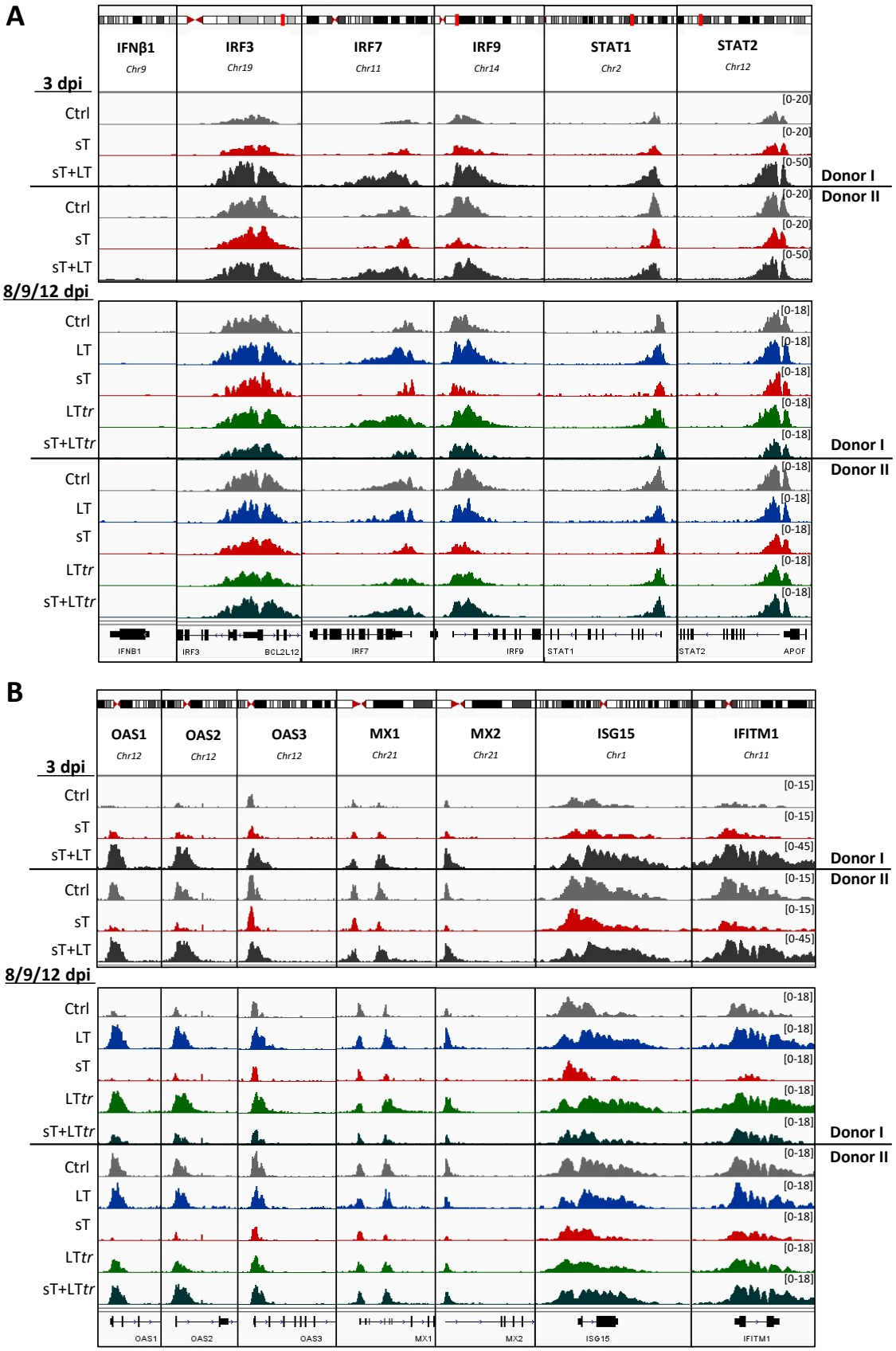


Figure 39 Genome browser tracks depicting H3K4me3 signal for selected type I IFN response genes in T antigen-expressing nHDFs

Genome browser tracks for a subset of type I IFN regulatory genes (A) and for a subset of ISGs (B). The tracks represent H3K4me3 signal from nHDFs expressing ctrl, sT or sT+LT at 3 dpi and ctrl, LT, sT, LT*tr* and sT+LT*tr* at 9/12 dpi from two different donors.

5 Discussion

MCPyV establishes a life-long persistence in the majority of the general population without causing any symptoms, and only rarely under immunosuppression leads to the formation of an aggressive skin tumor. In fact, from the viral point of view, the stage of tumorigenesis represents a "dead-end" because the viral genome is mutated, resulting in the abrogation of its replication. However, the T antigens and especially sT drive tumorigenic and metastatic processes (Dobson et al., 2020; Knips et al., 2017; Skvortsova et al., 2018; Stakaitytė et al., 2018, Schlemeyer et al., in revision). Therapeutic approaches are limited with more recently focusing on immune checkpoint control such as PD1/PD-L1 and combinatorial therapies. However, immune therapy shows its limitation, such as only 40% of MCCs respond to immune therapy and even more, hyperproliferation can occur in some cases (Walker et al., 2018). Therefore, an alternative approach to MCC treatment could be to develop vaccines for particularly vulnerable groups such as immunosuppressed patients. The development of such vaccines however requires a thorough understanding of the life cycle of MCPyV and of the events that lead to tumorigenesis. Since essential factors such as the site of persistence, the natural viral reservoir, and the cell of origin of transformation are still unknown, the life cycle has only been partially elucidated.

With its relatively small genome, MCPyV belongs to a family of viruses that does not have a large coding capacity for proteins that are necessary in the viral life cycle. Because MCPyV LT is the only viral encoded protein with enzymatic function, MCPyV strongly depends on the exploitation of host cellular processes such as the transcriptional or splicing machinery to replicate and produce new viral particles. While many interactions of the T antigens have been described to play important roles in the viral life cycle, their effects on host cell gene regulatory processes are just being explored with the advent of genome-wide analysis tools.

The aim of this study was to further dissect the influence of T antigen expression on gene regulatory processes in nHDFs. The results that were obtained in this work revealed that all T antigens potentially perturb cellular transcription and change the abundance of H3K4me3 at specific gene promoters. Interestingly, while there was a general overlap of distinct cellular signaling pathways modified on the transcriptional level by all T antigens, opposing effects were mainly observed between sT and the LT antigens, specifically regarding the innate immune response, which was identified in this work as a central target of all T antigens. These commonalities and differences, as well as their relevance for MCPyV pathogenesis, are discussed in the

following sections. Furthermore, the experimental setup is carefully evaluated in the context of MCPyV-associated pathogenesis.

5.1 Evaluation of the experimental setup

In this work, MCPyV T antigens were overexpressed in nHDFs to analyze transcriptional and histone modification changes in two different scenarios. On the one hand, expression of sT and LT reflected a combination of T antigen expression that is present in the early phase of infection. The second scenario of MCPyV-associated tumorigenesis was mimicked by the overexpression of sT and LT*tr*. Due to the lack of *in vitro* infection or tumorigenesis models, overexpression of MCPyV T antigens using a lentiviral transduction system represented a trade-off between a non-physiological model and the possibility of controlled T antigen expression.

The scenario of initial infection was assessable in this work by overexpressing sT and LT. Although the viral life cycle is incompletely understood, it is known that both T antigens are expressed early after infection, and especially LT is required for genome replication. Interestingly, LT expression was shown to be reduced by upregulation of miR-M1 (Theiss et al., 2015) which results in a reduction of genome replication. Another mechanism of reducing LT levels was identified by Kwun et al. (2017) who observed that phosphorylated LT is recognized by cellular ubiquitin ligases, leading to LT degradation. This autoregulation of a viral encoded protein that is usually required for genome replication was suggested a novel form of viral latency. Kwun et al. (2017) highlight that by regulating protein turnover of LT, the abrogation of genome replication mediates a persistent state which is in opposition to concepts of other persistent viruses. For example, during herpes virus latency, viral gene transcription is largely repressed and the viral genome is maintained via specific interactions with the host genome, which can be mediated for example by viral encoded proteins such as LANA for KSHV (Ballestas et al., 1999). The mechanisms used by MCPyV to establish persistence might be completely different from the classical forms of viral persistence, and need to be further investigated in the future.

Concomitantly with the incomplete understanding of its life cycle, MCPyV is a unique member of PyVs such that it is the only human PyV unequivocally linked to tumorigenesis in its primary host. Furthermore, it has not been associated with lytic replication, which is in strong opposition to other persistent PyVs such as BKPyV and JCPyV. Both of them causally contribute to disease in immunosuppressed patients, as a consequence of lytic reactivation (Imperiale & Jiang, 2016).

In contrast, reactivation of MCPyV leads to the integration of the viral genome into the host genome, which is the result of a replication block (Feng et al., 2008; Rodig et al., 2012; Shuda et al., 2008). It remains therefore unknown if MCPyV *per se* induces lytic infection in any cell type, or alternatively maintains a low level of virus production in a specific cell type to establish persistent infection.

Human pathogenic PyVs usually infect distinct sites of the human body and establish persistence in a large variety of organs that are different from the sites of reactivation and disease outbreak (Imperiale & Jiang, 2016). It is therefore unlikely that for MCPyV, fibroblasts represent a cell type for both scenarios of initial infection and tumorigenesis. Nevertheless, the lack of knowledge about these sites impairs the establishment of *in vitro* models to answer these questions, representing a major research issue in the field. Although the nHDFs used in this work allowed low-level infection (figure S1), it remains elusive whether dermal fibroblasts represent a permissive cell type or a site of persistence. Furthermore, as transformation of these cells by sT and LT*tr* was not observed (unpublished observations), they are probably not the cell of origin of MCC either. However, they represented the most suitable cell type for the experiments performed in this work, and their transcriptional profiles corresponded well to either scenario of infection or tumorigenesis, matching with decisive observations described in the literature.

5.2 Changes in host cell gene regulation in the infection scenario

Primary infection with MCPyV is suggested to occur normally during early childhood, possibly via fecal-oral transmission (Loyo et al., 2010). This primary infection is asymptomatic and interestingly, MCPyV is shed from the skin in 40% of healthy adults (Schowalter et al., 2010). The majority of the population is thus persistently infected with MCPyV for a life-time which suggests that the immune system is not able to clear initial viral infection. However, the mechanisms of persistence establishment and maintenance are poorly understood. Due to the lack of knowledge about the site of persistence there is no model system available to study questions related to MCPyV persistence. Therefore, in this work, overexpression of sT with LT in nHDFs aimed to mimic the early phase of MCPyV infection. Furthermore, by expressing sT or LT alone, their individual effects on gene regulation were analyzed in order to put them into the greater context of how they contribute to persistence establishment.

5.2.1 Transcriptional changes induced by sT and LT

Both expression of sT and LT induced the deregulation of many different genes that were enriched in several GO terms, highlighting that both T antigens have the potential to interfere with gene regulatory processes. While sT has been shown to specifically regulate the transcription of genes via the MYCL-EP400 transcription factor and chromatin remodeling complex (Cheng et al., 2017), interference of LT with gene regulatory processes is less well understood. It is known that LT indirectly activates E2F transcription factors by inhibiting pRb that normally functions to inhibit E2F which is an important regulator of S-phase related genes (Borchert et al., 2014; Hesbacher et al., 2016; Houben et al., 2015). Besides activation of cell cycle-related genes, which was also reflected in the transcriptome data in nHDFs, LT is known to have growth-inhibitory functions (Cheng et al., 2013). However, the transcriptome data only revealed an upregulation of cell cycle and proliferation genes. This might be explained by the fact that LT-expressing nHDFs were seeded at a higher density after FACS-sorting because of a reduction in cell proliferation that was observed in general by LT expression in nHDFs (see figure S2). In order to harvest sufficient cells for genome-wide analysis, this effect was counteracted by increasing the cell density which probably resulted in enhanced positive proliferative signals secreted by nHDFs.

Similar to LT, expression of sT or sT together with LT resulted in an enrichment of upregulated genes involved in cell cycle regulation and proliferation. This is most likely the result of the general stimulatory effect on cell proliferation that is described for sT and which is associated with its oncogenic potential. While sT was shown in MCC cell lines to be required for their growth and survival, sT expression also increased cell proliferation in human immortalized BJ-TERT fibroblasts (Shuda et al., 2011). Intriguingly, transformation of these cells could not be observed which is consistent with the observation that the nHDFs used in this work were also not transformed by overexpression of sT (data not shown).

Further gene sets that were positively regulated on the transcriptional level in all three conditions included *DNA replication* and *DNA repair*. These observations were anticipated for LT as it is known to interfere with DNA replication. However, as *DNA replication* was also upregulated in the other two conditions, this might also be an indirect effect of increased cell proliferation. Concerning the term *DNA repair*, it is well known that LT induces a DDR which normally functions to augment viral genome replication (Li et al., 2013; Tsang et al., 2014). The role for sT is less well studied but the observations from the transcriptional data presented here overlap with findings from Wu et al. (2019) who identified that sT interferes with the DDR pathway by activation of ATM and subsequent hyperphosphorylation of downstream targets. Collectively, the transcriptional upregulation of DDR-related genes by both sT and LT reinforces the cooperative function of sT to enhance viral genome replication which is mainly mediated by LT.

Another interesting gene set commonly upregulated by sT, LT and sT plus LT included *Inflammatory response* genes which showed the largest gene enrichment in LT-expressing cells (see figure 11). The fact that sT activates inflammatory cytokine expression had been observed already by Richards et al. (2015) and confirms the strong upregulation of several cytokines and chemokines by sT. However, as Richards et al. (2015) only overexpressed sT together with LT*tr* in human fibroblasts, a thorough comparison with the transcriptomic data from nHDFs overexpressing sT and LT*tr* presented here will be given in chapter 5.3.

For full-length LT, activation of inflammatory cytokines has not been described so far. This observation however indirectly overlaps with recent findings from Krump et al. (2021) who studied the innate immune response to MCPyV infection in nHDFs and found that inflammatory cytokines as well as ISGs were upregulated upon infection. They suggest that MCPyV DNA is sensed during replication by cGAS, resulting in the activation of inflammatory cytokines and

ISGs via the STING-TBK1-IRF3 axis as well as via the NFκB pathway. Whether LT expression *per se* also has an influence on the induction of inflammatory cytokines in a natural infection remains to be determined.

The observation that LT induced the upregulation of ISGs has so far not been described for MCPyV. Interestingly, Rathi et al. (2010) showed that SV40 LT activates the expression of a large amount of ISGs in mouse embryonic fibroblasts (MEFs). The authors identified that the pRb-binding motif was necessary for the upregulation of ISGs. Further, when overexpressed in human fibroblasts, SV40 LT was specifically shown to induce an ATR-dependent DDR that resulted in the production of IFNβ via IRF-1 which finally led to the expression of several ISGs (Forero et al., 2014). This study suggested an important link of the DDR to the induction of ISGs, which might also be relevant in the transcriptomic data presented here, as both expression of LT and LT together with sT increased the transcription of DDR genes, as well as ISGs and inflammatory cytokines. Further evidence that PyV T antigens activate ISGs was given by Giacobbi et al. (2015) who overexpressed the early regions from BKPyV, JCPyV and SV40 in MEFs and observed an antiviral state associated with upregulation of ISGs, which depended on STAT1. However, they could not rule out that the sT antigens were co-expressed as a splicing result of the early gene region. Therefore, future studies will be required to assess the impacts of sT and LT on the type I IFN response, especially during the course of acute and persistent infection.

Collectively, co-overexpression of sT and LT resulted in a transcriptional response that generally resembled an infection scenario. Intriguingly, when sT was overexpressed in nHDFs without LT, a specific and strong suppression of ISGs and upstream IRFs was observed. When co-overexpressing sT with LT, this effect was reverted. However, in comparison to expression of LT alone, sT dampened the transcriptional activation of many ISGs (see figure 38). This interesting function of sT, which has not been described so far, is discussed in more detail in chapter 5.4.

5.2.2 Histone modification changes induced by sT and LT

In addition to changes in transcription, the second major question assessed in this work was whether MCPyV T antigens induce histone modification changes as a possible way of interfering with gene regulation. Collectively, overexpression of sT, LT or sT plus LT did not induce global aberrant H3K4me3 or H3K27me3 patterns. H3K27me3 was significantly altered only for a small

amount of genes without an overall correlation with the RNA-Seq data. However, significant changes of H3K4me3 promoter signal were observed for specific gene sets and correlated well with transcriptional changes. These results highlight that the T antigens probably do not regulate host gene transcription by interfering with the polycomb repressive system which would be visible by changes in H3K27me3.

In contrast, H3K4me3 signal was specifically altered in the promoters of several gene sets that were also deregulated on the transcriptional level. While it is well known that H3K4me3 deposition is associated with active gene transcription by polymerase II, the order of events is controversially discussed in the field. Howe et al. (2017) suggest that instead of inducing transcription, H3K4me3 may also be deposited in response to transcription. Determination of the spatiotemporal distribution of H3K4me3 and transcription were not within the scope of the experimental setup, but due to the analysis of both RNA-Seq and ChIP-Seq analysis at different time points, some conclusions are drawable. The data from nHDFs expressing sT was assessed both at 3 and 8 dpi, while for LT only the later time point was considered. In contrast, nHDFs expressing sT plus LT were analyzed only after 3 dpi to detect changes induced as a direct response to T antigen expression. Interestingly, the transcriptional changes of *Type I IFN response* or *Innate immune genes* were detectable already after 3 dpi in all three conditions. Although after 8 dpi the induction of type I IFN response genes was not significant in contrast to 3 dpi in LT-expressing nHDFs, changes in H3K4me3 promoter signal of IFN-responsive genes were still detectable after 8 dpi (see figure 13). Furthermore, H3K4me3 in sT-expressing nHDFs, which showed a repression of ISGs on the transcriptional level both after 3 and 8 dpi, was more prominently reduced at the later time point. These observations hint at an initial transcriptional response that is followed by H3K4me changes.

In fact, maintaining open chromatin states at promoters of IFN responsive genes has been considered a mechanism of pre-activation, ensuring a faster response to external stimuli (Au-Yeung & Horvath, 2018). For example, the BAF chromatin remodeling complex was reported to accelerate the activation of IFN response genes after IFNα stimulation (Cui et al., 2004). Interestingly, the genome browser tracks from nHDFs transduced with the vector controls indicate a pre-existing H3K4me3 signal at the promoters of several ISGs (figure 39). However, this effect was more pronounced in Donor II compared to Donor I, which reflects an expected phenotype, owing to the fact that nHDFs are primary cells.

Regarding T antigen-expressing nHDFs, LT expression led to an increase of H3K4me3 signal only in a small amount of promoters of genes that were mostly involved in the *Type I IFN response*. In contrast, sT had a larger impact on H3K4me3 changes, as summarized in figure 25. These differences might reflect the ability of sT to interfere with gene regulation via the MYCL-EP400 complex which was shown to be recruited by sT to promoters of active genes (Cheng et al., 2017). While this mechanism was initially identified to activate genes involved in tumorigenesis, this complex also activated the LSD1/CoREST transcriptional repressor complex which in turn inhibited the transcription of differentiation genes negatively influencing tumor cell growth (Park et al., 2020). These findings underline the hypothesis that sT might both activate and repress genes via recruitment of co-transcriptional factors. In fact, while in sT-expressing nHDFs the majority of differential H3K4me3 signal was increased after 3 dpi, this effect was reverted after 8 dpi (see figure 26). Interestingly, genes with decreased H3K4me3 promoter signal were enriched for genes involved in *Defense response to virus* and *Transcriptional regulation*, which were new observations. Furthermore, genes involved in *Cell differentiation* were also detected by sT expression thereby confirming observations from Park et al. (2020).

In accordance with the observation that sT possibly interferes with gene regulation via chromatin remodeling, it was a striking observation that overexpression of sT with LT resulted in the highest amount of changes in H3K4me3 signal. As shown in figure 35, 2384 genes were detected with significantly increased H3K4me3 promoter signal at 3 dpi. Although there was a general correlation of transcriptionally activated genes with H3K4me3 signal, as only 302 genes in total were observed by RNA-Seq to be upregulated, most genes with increased H3K4me3 thus were not directly transcriptionally regulated.

The fact that only co-expression of sT and LT led to such a high amount of genes with differential H3K4me3 signal hints at a cooperative effect of sT and LT. Unpublished observations from our group have identified LT to bind to host chromatin via GRGGC motifs that are also located in the viral origin of replication and were shown to be specifically bound by the OBD of LT to initiate genome replication (Feng et al., 2011; Harrison et al., 2011; Kwun et al., 2013). The functions and consequences of LT binding which was specifically observed to occur in promoter regions marked with H3K4me3, are currently under investigation but it might not be the direct cause for transcriptional regulation of host genes as only a small proportion of LT-bound genes was differentially expressed (unpublished observations). Nevertheless, it might be possible that LT acts in cooperation with sT in chromatin remodeling which is underlined by the here

presented observations. However, to investigate the hypothesis of a cooperative effect of sT and LT in chromatin remodeling, further experiments are required to screen for possible associations with co-transcription factors or chromatin remodeling complexes.

Because most LT-bound regions were not directly associated with transcriptional changes, it is perceivable that transcriptional changes might be regulated via 3D genome interactions. These include interactions of distant promoters, but also with enhancer regions. Enhancers enable physical interactions of gene regulatory elements to regulate the transcription of genes via long-range interactions, even across chromosomes (Plank & Dean, 2014). Active enhancer regions are often marked with H3K27ac combined with H3K4me1 (Creyghton et al., 2010; Wang et al., 2008). Interestingly, H3K27ac, which was assessed only for sT-expressing nHDFs at 8 dpi, correlated mostly with changes in H3K4me3, especially at promoters of transcriptionally deregulated genes such as innate immune genes. The H3K27ac diffReps analysis revealed 302 and 556 genes with increased or decreased promoter signal, respectively. However, there was a large amount of differential signal located in gene bodies, including 1405 genes with increased and 6297 with decreased signal. While the presented data is restricted to the analysis of H3K27ac, future experiments should address whether regions with differential H3K27ac are also marked by H3K4me1, as a co-occurrence of both signals marks the presence of active enhancers (Creyghton et al., 2010; Wang et al., 2008). Incorporation of the knowledge about the newly identified chromatin binding function of LT together with the fact that sT potentially interferes with chromatin remodeling highlights that investigating 3D genomic interactions represents a promising and important future research question.

In summary, the systematic analysis of transcriptional and epigenetic changes in a setting that resembled MCPyV early infection events and persistence establishment, confirmed known functions of sT and LT. In addition, it revealed a so far unknown function of LT in activating inflammatory cytokines and type I IFN response genes and of sT in stably repressing ISG transcription. The presented results add on the importance of MCPyV as a persistent virus to have developed mechanisms to evade immune recognition. Possible mechanistic insights are discussed in more detail in chapter 5.4.

5.3 Changes in host cell gene regulation in the tumor scenario

Transformation of MCPyV-infected cells and development of MCC is a rare event that is incompletely understood. As summarized in the model shown in the introduction (figure 7), it is not known which cell type is the cell of origin of transformation and which events exactly lead to tumorigenesis. It is well established though, that LT truncation and integration of the viral genome into the host genome precede transformation (Czech-Sioli et al., 2020a; Feng et al., 2008; Rodig et al., 2012; Shuda et al., 2008; Starrett et al., 2020). Due to the lack of an *in vitro* transformation model, overexpression of sT and LT*tr* in nHDFs was aimed at reflecting a scenario which is present in a transformed cell. Their individual contributions as well as their impact on host gene regulation with regard to tumor-associated processes are therefore discussed in the following.

5.3.1 Transcriptional changes induced by sT and LT*tr*

Both sT and LT*tr* expression, as well as their co-overexpression resulted in a large amount of genes enriched in the GO terms *DNA replication* and *Cell division*. This confirmed the positive effect of sT and LT*tr* on cell proliferation that is described in the literature to be associated with tumor growth and survival (Cheng et al., 2017; Houben et al., 2010; Park et al., 2020; Shuda et al., 2014; Shuda et al., 2011). Interestingly, Richards et al. (2015) analyzed the transcriptional response to overexpression of LT*tr* alone or together with sT in BJ-hTERT fibroblasts and they found many upregulated genes enriched in GO terms related to *DNA replication*, *Cell division* and *Proliferation* which largely reflects the here presented findings in nHDFs.

The transcriptomic analysis from Richards et al. (2015) revealed additionally that sT and LT*tr* activate inflammatory cytokines and chemokines. This strongly overlaps with the data from nHDFs overexpressing sT and LT*tr* individually or in combination. In contrast to the experimental setup used in this work, Richards et al. (2015) did not include cells only expressing sT. Nevertheless, they found that transcription of several cytokines and chemokines was enhanced when sT was co-expressed with LT*tr* compared to LT*tr* alone. This observation was also made in the RNA-Seq data presented here, as co-overexpression of sT and LT*tr* led to an enrichment of genes involved in inflammation, chemotaxis and cytokine signaling (figure 32). Many of these genes were also detected by Richards et al. (2015), including *IL1β*, *IL6*, *CXCL1*, *CXCL6*, *CSF2*,

MMP1 and *CCL7*. A list of all cytokines and chemokines that were enriched in the top 10 GO terms is given in table S12.

Another major observation was that LT*tr* expression in nHDFs resulted in the upregulation of ISGs, in contrast to sT which stably repressed ISG transcription. The comparative heatmap shown in figure 38 highlights that ISG induction by LT*tr* was only significant after 3 dpi and probably represents a short-term response to the induction of the stimulation that was observed by lentiviral transduction (see figure 30). Thus, at the later time point, there were no significant changes detectable in LT*tr*-expressing nHDFs, similar to LT-expressing cells. Because nHDFs that overexpressed sT+LT*tr* were analyzed only at the late time point with the aim to detect possible long-term changes induced by constant T antigen expression, there were no significant changes detectable for the subset of *type I IFN response* genes. Nevertheless, the fact that sT constantly repressed ISG transcription suggests that sT might be able to respond to external stimuli and thereby counterbalance the general stimulatory effects of LT*tr*.

The observations from the RNA-Seq analysis of nHDFs representing a tumor scenario suggest a possible role of sT in immune evasion which is a hallmark for tumor cell survival due to their constant exposure to innate and adaptive immune cells (Hanahan & Weinberg, 2000). This newly identified function of sT fits well into the observation that MCCs have developed several mechanisms to evade the immune system. In fact, the RNA-Seq analysis from sT-expressing nHDFs revealed an additional set of genes that was strongly repressed both after 3 and 8 dpi. These included several *HLA* members, as well as *TAP1* and *B2M* which are involved in antigen presentation and processing (see figure 30). These findings suggest a possible role of sT in the inhibition of MHC-I-mediated antigen presentation (Paulson et al., 2014). Furthermore, a study that was recently conducted in our group has also identified sT to be the driver of CD47 upregulation which results in the inhibition of macrophage-induced phagocytosis thereby contributing to immune evasion (Schlemeyer et al., in revision).

Collectively, the observation that sT is able to repress the transcription of type I IFN response genes as well as MHC-I-related genes highlights its potential to contribute to MCC survival by evading immune recognition. Although LT*tr* might not constantly induce the upregulation of ISGs, it was shown to enhance a pre-existing stimulus. In a tumor scenario, it is well known that type I IFNs play an important role in acting against tumor cells (Musella et al., 2017). Thus, sT might counterbalance the cellular response to type I IFN stimulation that is enhanced by LT*tr* by repressing ISG transcription to evade immune recognition and tumor elimination.

Interestingly, LT*tr* expression also resulted in the upregulation of a variety of genes that are involved in apoptotic processes which is detrimental for the survival of the tumor cell. Although on the transcriptional level it was not observed that sT directly counterbalances this effect, recent evidence from our group revealed that sT plays a role in inhibiting apoptosis via a so far unknown mechanism (Schlemeyer et al., in revision). Furthermore, as certain ISGs have been described to be involved in apoptosis (Musella et al., 2017), by repressing ISG transcription, sT might indirectly also counterbalance the pro-apoptotic feature of LT*tr*.

Collectively, the RNA-Seq analysis from both sT and LT*tr*-expressing cells revealed that both T antigens interfere with the regulation of genes that are known to play a role for tumorigenesis. Furthermore, the activation of type I IFN response genes by LT*tr* and their suppression by sT suggests an important mechanism for tumor cell survival. Although nHDFs were not transformed by sT and LT*tr* (data not shown), they were characterized by a tumor-associated profile on the transcriptional level. Intriguingly, chemokines that were highly enriched in the GO term analysis, are known to play a decisive role for tumor-associated processes such as growth, metastasis and angiogenesis (Singh et al., 2007). While it was not within the scope of this work to decipher the roles of sT and LT*tr* in these processes, their transcriptional regulation reinforced the experimental setup that was aimed at reflecting a tumor scenario.

5.3.2 Histone modification changes induced by sT and LT*tr*

Similar to what was observed in the infection scenario, there was a general overlap of the transcriptional response with the activating H3K4me3 mark, while H3K27me3 was not specifically targeted by T antigen expression. Interestingly, when sT was co-expressed with LT*tr*, there were only mild changes in H3K4me3 promoter signal observable, comprising 62 genes with increased and 39 genes with decreased signal. This was in opposition to the observation that full-length LT together with sT led to 2384 genes with increased H3K4me3 signal. As discussed in the previous section, one explanation could be that sT and LT cooperate in chromatin remodeling via the newly identified feature of LT to bind to cellular chromatin. However, this function is strongly reduced for LT*tr* due to the lack of the OBD (unpublished observations). These findings might explain the moderate impact of sT+LT*tr* on H3K4me3 promoter signal.

The fact that there was a strong overlap of transcriptionally regulated genes with H3K4me3 suggests that all T antigens potentially interfere with gene regulation via H3K4me3. Whether

the changes in H3K4me3 are a result from the transcriptional regulation or vice versa is not precisely extractable from the here presented results. However, as discussed in the previous section, regulation of type I IFN response genes is often associated with promoter priming by maintaining an open chromatin state to enable a faster response to stimuli (Au-Yeung & Horvath, 2018; Cui et al., 2004). The stimulatory effects that were observed for LT and LT*tr* might thus contribute to priming which would be supported by the observation that for LT*tr* a significant upregulation of ISGs was only observed after 3 and not after 8 dpi, while H3K4me3 was steadily increased even at the later time point. However, future experiments are required to confirm whether transcriptional priming might be induced by the T antigens.

In summary, transcriptional and epigenetic profiling confirmed on the one hand that sT and LT*tr* expression in nHDFs is associated with tumor-related processes. Furthermore, dissection of the individual contributions of sT and LT*tr* revealed a potent stimulation of type I IFN response genes by LT*tr* and its repression by sT. Both effects were associated with increased or decreased H3K4me3 promoter signal, highlighting that LT*tr* might contribute to ISG priming. In addition, it is perceivable that sT uses epigenetic mechanisms to repress ISG transcription thereby contributing to immune evasion which is required for tumor cell survival.

5.4 MCPyV sT acts as a counterbalance to LT and LT*tr*

The large-scale, genome-wide analysis of transcriptional and epigenetic changes performed in this work revealed that all MCPyV T antigens interfere with gene regulatory processes that might have important functions during the viral life cycle. Furthermore, they have an influence on the abundance of activating histone marks such as H3K4me3 and H3K27ac, but they do not interfere with repressive marks such as H3K27me3 or H3K9me3. One of the most striking observations was that sT stably represses the transcription of genes involved in the type I IFN response and thereby possibly counterbalances immunogenic effects that were shown to be induced by both LT and LT*tr*. This so-far unknown function of sT underlines the importance of determining the relationship of MCPyV and MCC with the innate immune system and will be further discussed in the following.

5.4.1 Suppression of ISGs by sT as an immune evasion strategy

The experimental setup used in this work was based on a lentiviral transduction system. Therefore, it was important to assess the possible effects of this type of viral infection on the nHDF transcriptome. In fact, when comparing nHDFs transduced with the vector controls to untreated cells by RNA-Seq analysis, many upregulated genes were enriched in GO terms such as *Defense response to virus* or *Inflammatory response* (see figure S4, table S2 and table S3). This initial stimulus mediated by the lentivirus infection *per se* probably reinforced the detection of the specific responses elicited by MCPyV T antigens on the innate immune response. It is therefore indispensable to further dissect which viral PAMPs were presented by the lentiviral infection and point out possible analogies to PAMPs that are presented during MCPyV infection or tumorigenesis.

Lentiviruses are a genus of *Retroviridae* which are enveloped reverse transcribing viruses with an RNA genome (Ryu & Ryu, 2017). In a natural infection, following viral entry the RNA genome is reverse-transcribed into dsDNA which translocates into the nucleus where it integrates into the host genome. Normally, from the integrated proviral sequence, viral mRNAs and genomic RNAs are synthesized in order to assemble new virions. For lentiviral transduction systems, the life cycle ends with integration of the desired sequences into the host genome. However, before viral integration, there are several steps during which multiple PAMPs are exposed such as viral RNA, DNA and hybrids of both that are sensed by endosomal TLR3, TLR7 or TLR9, or

by cytosolic PRRs such as RIG-I-like receptors or cGAS. The sensing of these PAMPs results in the activation of a signaling cascade that leads to the activation of cytokines and type I IFN.

The extent to which the viral PAMPs presented during lentiviral infection play a role for MCPyV remains elusive because of the incomplete knowledge about its life cycle (see figure 7). However, as it was recently suggested that MCPyV replication activates ISGs and inflammatory cytokines (Krump et al., 2021), nucleic acid sensors probably play an important role in sensing of MCPyV replication products during initial infection. This cannot be conferred on the tumor scenario as LT truncation leads to the loss of genome replication. However, there is increasing evidence that tumor cells are often associated with increased DNA damage. Thus, the DNA is sensed by the cGAS-STING axis resulting in the induction of inflammatory cytokines and IFNs (Barber, 2015). Interestingly, Liu et al. (2020) observed that STING is silenced in MCCs, thereby implying that MCCs also potentially carry damaged DNA which would normally be sensed by STING. They propose the inhibition of STING as a way to evade immune recognition via a so-far unknown mechanism. Interestingly, the transcriptomic data from nHDFs overexpressing sT showed that STING was significantly downregulated after 8 dpi (see figure 30).

Due to the fact that the type I IFN response is possibly a decisive factor for the establishment of persistence for BKPyV and JCPyV (An et al., 2019; Assetta et al., 2016), it is perceivable that also MCPyV elicits an IFN response prior to persistence establishment. Intriguingly, by inducing an antiviral response, it is probably necessary from the viral point-of-view to evade immune recognition. In fact, JCPyV sT was recently identified to have evolved a mechanism to counteract RIG-I-mediated immune recognition. A similar role was also shown for BKPyV sT but the exact mechanism remains to be investigated (Chiang et al., 2021). Similarly, the here observed suppression of ISG transcription by MCPyV sT fits well into these newly identified functions of related PyV sT antigens. Interestingly, it had been observed previously that TLR9 is transcriptionally repressed in epithelial and MCC-derived cells when MCPyV LT or sT were expressed (Shahzad et al., 2013). Although TLR9 was not identified to be significantly deregulated by MCPyV T antigens in nHDFs, the here presented subversion of type I IFN signaling adds on the limited evidence in the field that MCPyV has evolved possibly several mechanisms to evade innate immune recognition.

5.4.2 sT interferes with ISGF3 to suppress ISG activation

The transcriptional data from nHDFs overexpressing sT revealed a downregulation of several innate immune genes classified into three main categories (see figure 30). The majority of them included ISGs which are activated downstream of IFN signaling, depending usually on the activation of ISGF3 which is composed of pSTAT1, pSTAT2 and IRF9. It was shown in this work that sT inhibits IRF9 transcription and reduces its nuclear localization which results in decreased protein levels, indicating a stable repression of IRF9. It remains to be determined if sT physically impairs ISGF3 and as a result inhibits its binding to ISRE in the promoters of ISGs and thereby blocks the transcription of a large subset of ISGs (see figure 40).

Upstream of IFN production, upon sensing of PAMPs, several events take place that are regulated by IRFs, PRRs and adapter molecules. Although sT was shown to specifically target IRF9 and repress ISG transcription, it is possible that sT additionally interferes with steps that are upstream of this. The model in figure 40 highlights that cytosolic and endosomal sensing of nucleic acids result in the activation of IRF3, IRF7 and NfκB. The transcriptional data revealed a downregulation of STING and IRF7, but not of NfκB, IRF3 or type I IFNs *per se*. While NfκB is mainly required for the activation of inflammatory cytokines, IRF3 and IRF7 are mostly required for type I IFN expression, although their cooperation has also been reported (Freaney et al., 2013; Iwanaszko & Kimmel, 2015; Platanitis & Decker, 2018). It is possible that sT might interfere with these or other factors on a non-transcriptional level, for example by PTMs, physical impairment or proteasomal degradation. In fact, sT was shown to interfere with NfκB, although opposing reports on its consequences have been made. While Berrios et al. (2016) observed an increased transcription of cancer-associated genes dependent on NfκB, Griffiths et al. (2013) and Abdul-Sada et al. (2017) observed an inhibition of NfκB-mediated transcription. Interestingly, sT was also shown to activate the transcription of SASP genes via non-canonical NfκB signaling (Zhao et al., 2020). Some SASP genes that were upregulated by sT included *Il1β* and *IL6* which were also upregulated by sT in the here presented RNA-Seq analysis. These findings suggest that sT potentially interferes with NfκB-mediated transcription of a subset of genes involved in the innate immune response. However, it remains to be determined if this has a positive or negative effect on type I IFN transcription and if there are further interactions of sT with IRFs, in addition to IRF9.

Due to the fact that sT is known to interfere with gene regulation for example via MYCL-EP400 (Cheng et al., 2017), it is perceivable that sT represses the transcription of IRFs via co-transcription factor and chromatin remodeling complexes. This hypothesis also builds on the observed reduction of H3K4me3 and H3K27ac in promoter regions of ISGs and IRFs by sT. Although little is known about the ISGF3-independent regulation of IRF9 and IRF7 as ISGs, it was shown for example that FOXO3, a member of the forkhead family of transcription factors, is a negative regulator of IRF7 (Litvak et al., 2012). Further co-transcription factors might associate with sT to interfere with the transcription of IRFs which needs to be addressed in future experiments for example by performing ChIPs using antibodies against sT and IRFs, coupled to mass spectrometric analysis. The schematic model in figure 40 summarizes possible mechanisms used by sT to subvert PRR and type I IFN signaling, including the inhibition of ISGF3 and subsequent repression of ISG transcription which was identified in this work as a novel mechanism contributing to immune evasion.

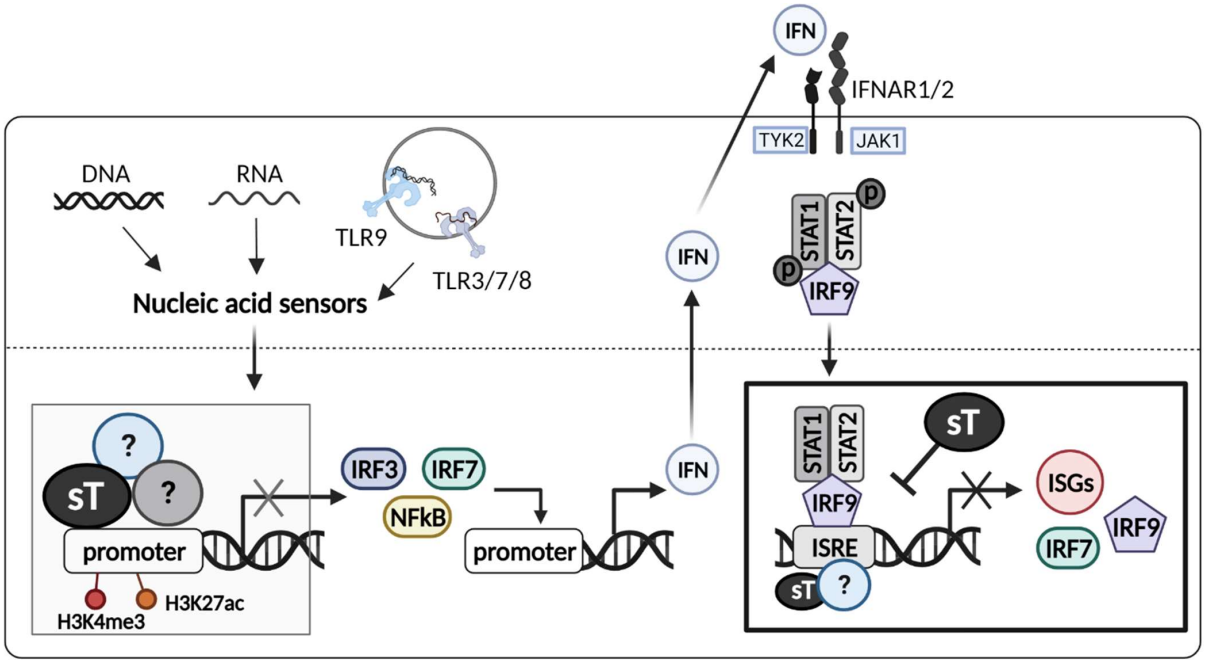


Figure 40 Mechanisms used by sT to repress ISG transcription
Nucleic acid sensing and type I IFN signaling pathway. It was shown in this work that sT efficiently represses transcription, nuclear localization and protein levels of IRF9 which is part of the ISGF3 transcription factor complex. Inhibition of ISGF3 is highlighted in the black box, indicating also a possible involvement of co-transcription factors for IRF9 repression. Gray box highlights the possibility that sT additionally interferes with transcriptional regulation of IRF3, IRF7 or NfκB by recruiting co-transcription and chromatin remodeling complexes. (created with Biorender)

6 Conclusion and Outlook

Although MCPyV was discovered over 14 years ago as the causative agent of the majority of MCCs, there are many open questions regarding its pathogenesis. Despite its abundance on the skin of the majority of the human population, it is unknown which cell types exactly allow productive or persistent infection. Furthermore, the cell of origin of transformation remains elusive but the fact that MCC is a rare disease that only occurs in immunosuppressed patients suggests the host immune response as a decisive factor for the infection outcome. In fact, the here presented data point out that both the inflammatory response as well as the type I IFN response are common targets of MCPyV sT and LT. Strikingly, both T antigens increase the upregulation of inflammatory cytokines but only LT additionally induces the upregulation of ISGs. This is counterbalanced by sT which stably represses ISG transcription, highlighting the relevance of these newly identified functions of sT and LT in the context of persistent infection (see figure 41A).

It has only recently been described that the type I IFN response is a prerequisite for the persistence establishment of the related PyVs BKPyV and JCPyV (An et al., 2019; Assetta et al., 2016). However, the role of the T antigens in persistence establishment was not considered in these studies and represents an interesting question to address, given the association of MCPyV T antigens with the type I IFN response that was observed in this work. It remains to be determined how LT expression alone leads to the temporal upregulation of ISGs. One possible hypothesis would be its association with genotoxic stress induced by DNA damage that was shown for SV40 LT to causally contribute to ISG upregulation (Forero et al., 2014).

Intriguingly, the counterbalancing effect of sT observed in this work adds on the limited knowledge about immune evasion mechanisms used by MCPyV to establish persistence. Due to the use of a non-physiological lentiviral transduction system, future experiments will be required to confirm the relevance of the here presented findings in a physiological MCPyV infection. However, the lack of an *in vitro* model system requires alternative approaches including the following two possibilities: (1) Although the infection of the nHDFs used in this work results in only 5% of infected cells (see figure S1), single cell methods might be useful to detect differences in ISG expression between infected and non-infected cells. In addition to IF analysis, the enrichment of infected cells by a reporter system that is currently established in our group, would allow to conduct single cell RNA-Seq analysis to compare the transcriptional profiles of

infected *vs* non-infected cells. (2) The second approach would be to overexpress MCPyV sT in a persistent infection system of a related PyV such as BKPyV which would serve as a surrogate infection model. Furthermore, characterization of the interaction of BKPyV T antigens with the type I IFN response represents another interesting research question which would add on the understanding of both MCPyV and BKPyV persistence establishment and point out possible differences.

Although it is currently unknown which events lead to MCPyV reactivation and MCC formation, tumorigenesis is considered a "dead-end" for the virus due to the loss of the ability to replicate its genome. Although many open questions regarding tumorigenesis remain, MCC progression is mainly shaped by the expression of sT and LT*tr*. Due to the lack of *in vivo* model systems to study tumorigenesis, in this work both T antigens were overexpressed in nHDFs to mimic a tumor-like scenario. The transcriptional profiles confirmed the cancer-associated phenotype induced by both T antigens that was observed already by Richards et al. (2015), including the upregulation of inflammatory cytokines and chemokines. The observation that sT additionally downregulated several *HLA* genes represents a new finding that adds on the reduced MHC-I expression that is often observed in MCCs (Paulson et al., 2014). Further studies should address the underlying mechanisms and investigate if interference with sT expression reverts the expression of MHC-I molecules on MCC cell lines and restores CD8+ T cell activation. Likewise, in a cell surface marker screen which was conducted in our group with MCC cell lines and a sT knockdown system, it was identified that sT changes the expression of many cell surface receptors, including CD47 as an important signal for macrophage-induced phagocytosis (Schlemeyer et al., in revision). Interestingly, HLA-A2 was also increased upon sT knockdown which is in concordance with the here presented finding that sT overexpression led to downregulation of *HLA-A* on the transcriptional level (see figure 30).

Interestingly, similar to full-length LT, LT*tr* expression resulted in the temporal upregulation of ISGs. Their repression by sT might be important for counterbalancing the effects of possible exposures of MCCs to IFNs secreted in the tumor microenvironment that might be enhanced by LT*tr* expression (see figure 41B). Future experiments should address the effects of stimulating MCC cell lines with type I IFNs in order to characterize their response to it. In addition, the effect of IFNγ might also be of special interest because of its engagement in anti-tumor activity, its interplay with type I IFNs and its involvement in the regulation of antigen presentation (Lee & Ashkar, 2018).

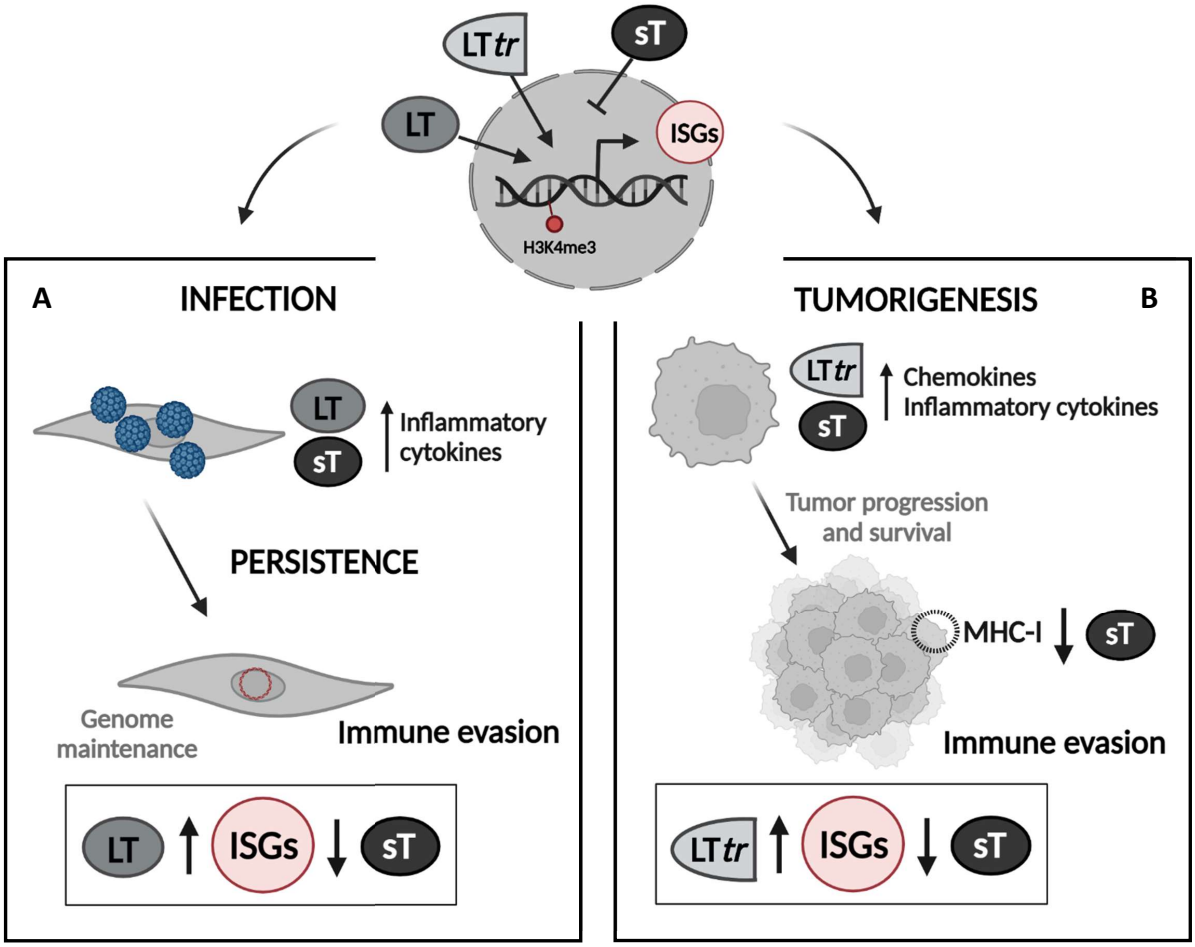


Figure 41 MCPyV sT represses ISG transcription in counterbalance to LT and LT*tr*.
Schematic model summarizing the counterbalancing role of sT against LT during persistent infection (A) and LT*tr* in the tumor scenario (B). While LT and LT*tr* upregulate the transcription of ISGs, sT leads to their suppression. The induction of inflammatory cytokines was observed by sT, LT and LT*tr*, and contributes to the immunogenicity elicited by the LT antigens. In a natural infection, sT might contribute to persistence establishment by counterbalancing the ISG response elicited by LT (A). Likewise, this counterbalancing effect might also play a role in the tumor scenario because ISGs are also upregulated by LT*tr* (B). In addition, transcriptional profiling revealed that sT downregulates the expression of *HLA* genes thereby possibly influencing MHC-I expression and contributing to immune evasion. (created with Biorender)

Finally, the transcriptional profiling has revealed that sT stably suppresses ISGs, and it was shown explicitly that interference with ISGF3 represents at least one underlying cause. Further studies will be required to investigate if sT directly impairs ISGF3 binding to ISREs thereby contributing to ISG transcriptional repression. Moreover, it is perceivable that sT, in addition to ISGF3, has different targets in the upstream signaling cascade, as summarized in figure 40. While investigating these mechanisms will give important new insights into the mode of action of sT and its consequence for MCPyV pathogenesis, they may also enable the development of new targeted therapies.

This study has revealed new insights into the relevance of MCPyV T antigens in shaping the host cellular transcriptional response. In addition, it has precisely elucidated the type I IFN response as a direct target of sT, LT and LT*tr* that has not been described so far. Therefore, future studies based on the observations presented in this work may contribute to understanding the interactions of the T antigens with the innate immune response that possibly play an important role in MCPyV pathogenesis but are incompletely understood to date. Furthermore, future studies will investigate the mechanisms used by sT that result in the stable suppression of ISG transcription, which is probably not restricted to the here presented inhibition of ISGF3 by sT. A thorough understanding will contribute to the development of new therapeutic strategies, for example by restoring a functional type I IFN response.

List of Figures

List of Tables

Abbreviations

ALTO Alternative frame of the large T open reading frame

ANOVA analysis of variance

APM antigen presentation machinery

ATM Ataxia telangiectasia mutated

ATR ATM- and Rad3-Related

BAF BRG1- or BRM-associated factors

BKPyV BK polyomavirus

bp base pair

BSA bovine serum albumin

c centi

cDNA complementary DNA

°C degrees celsius

cGAS cyclic GMP-AMP synthase

ChIP chromatin immunoprecipitation

CK20 cytokeratin 20

CPI checkpoint inhibitor

ctrl control

DAPI 4',6-diamidino-2-phenylindole

DAVID Database for Annotation, Visualization and Integrated Discovery

DDR DNA damage response

DEG differentially expressed gene

DMEM Dulbecco's Modified Eagle Medium

DMSO dimethyl sulfoxide

DNA deoxyribonucleic acid

dpi days post infection

ds double-stranded

ECM extracellular matrix

EDTA ethylenediaminetetraacetic acid

ER endoplasmic reticulum

EV extracellular vesicle

FACS fluorescence activated cell sorting

FBS fetal bovine serum

FDR false discovery rate

fw forward

g gram

GAG glycosaminoglycan

GO gene ontology

h hours

HDAC histone deacetylase

HEPES 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

HFCS Hybridoma Fusion and Cloning Supplement

hg19 human genome (version 19)

HLA human leukocyte antigen

HRP horse radish peroxidase

hTERT human telomerase reverse transcriptase

IF immunofluorescence

IFN interferon

igv integrative genomics viewer

IL-6 interleukin 6

IRES internal ribosome entry site

IRF interferon regulatory factor

ISG interferon-stimulated gene

ISGF3 interferon-stimulated gene factor 3

ISRE interferon-responsive element

JCPyV JC polyomavirus

kbp kilo base pairs

kDa kilodalton

KDM lysine demethylase

KSHV Kaposi's sarcoma-associated herpesvirus

L liter

LANA latency-associated nuclear antigen

LB Luria-Broth

log2FC log 2 fold change

LSD LT-stabilizing domain

LSD1 lysine-specific demethylase 1

LT large Tumor antigen

LT*tr* truncated large Tumor antigen

m milli

µ micro

M molar

mA milliampere

Mb mega base pairs

MCPyV Merkel cell polyomavirus

MEF mouse embryonic fibroblast

MHC-I major histocompatibility complex-class I

MICA/B MHC class I polypeptide–related sequence A/B

min minutes

mio million

miRNA microRNA

MOI multiplicity of infection

MPyV Murine polyomavirus

mRNA messenger RNA

mTOR mammalian target of rapamycin

MUR MCPyV unique region

n.s. not significant

n nano

NaCl sodium chloride

NCCR non-coding control region

NfκB nuclear factor kappa-light-chain-enhancer of activated B-cells

nHDF neonatal human dermal fibroblast

nm nanometer

NP40 nonidet P-40

OBD origin-binding domain

OD optical density

ORF open reading frame

P/S penicillin/streptomycin

p phosphorylated

padj. p adjusted

PAMP pathogen-associated molecular pattern

PBS Phosphate buffered saline

PD-1 programmed cell death 1

PEG polyethylene glycol

PEI polyethylenimine

PFA paraformaldehyde

PML progressive multifocal leukoencephalopathy

PP2A protein phosphatase 2A

PP4C protein phosphatase 4 catalytic subunit

pRb retinoblastoma protein 1

PRC2 polycomb-repressive complex 2

PRR pattern-recognition receptor

PTM post-translational modification

PVDF polyvinylidene difluoride

PyV Polyomavirus

qPCR quantitative polymerase chain reaction

RacPyV Racoon polyomavirus

rev reverse

RLR retinoic acid inducible gene-I-like receptor

RNA ribonucleic acid

RPMI Roswell Park Memorial Institute

RT reverse transcription

s seconds

SASP senescence-associated phenotype

SDS sodium dodecylsulfate

SDS-PAGE sodium dodecylsulfate polyacrylamide gel electrophoresis

Seq sequencing

SFM serum-free medium

S-phase synthesis phase

sT small Tumor antigen

STAT signal transducer and activator of transcription

STING stimulator of interferon genes

SV40 Simian virus 40

TEMED tetramethylethylenediamine

TLR Toll-like receptor

Tris 2-amino-2-hydroxymethyl-propane-1,3-diol

TSPyV Trichodysplasia spinulosa polyomavirus

TSS transcriptional start site

TTS transcriptional termination site

U units

UV ultraviolet

v/v volume per volume

V volt

VP viral protein

w/v weight per volume

WB Western Blot

x g times gravity

Supplementary material

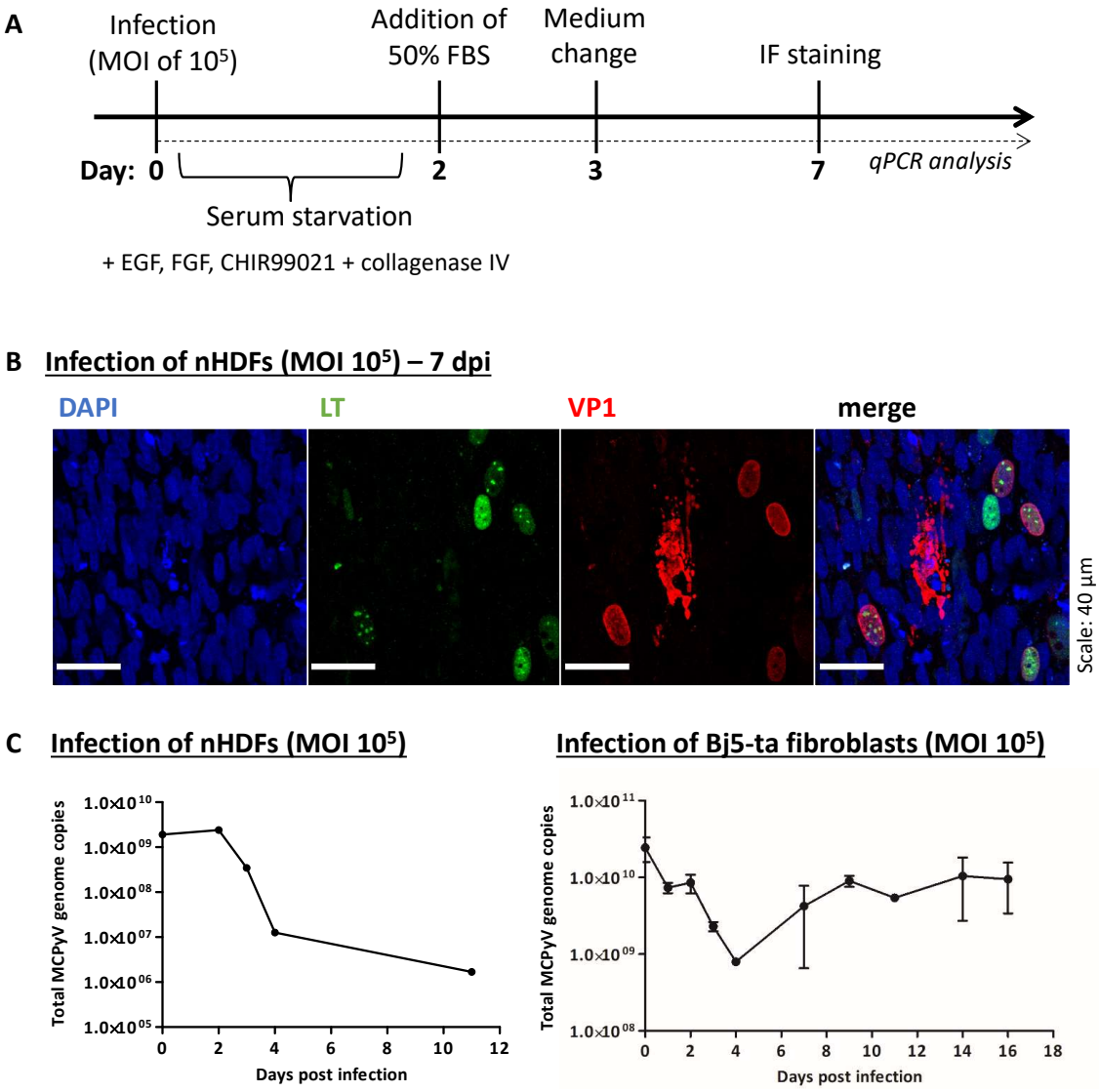


Figure S1 Infection of fibroblasts with MCPyV.
(A) Time-line depicting the protocol used to infect nHDFs or Bj5-ta immortalized fibroblasts with MCPyV as described in Liu et al. (2016). An MOI of 100,000 was used for infection. For two days, cells were starved of FBS which was added to a total of 50% at 2 dpi. During the whole infection time, several supplements were contained in the cell culture medium. EGF: Epidermal growth factor; FGF: Fibroblast growth factor; CHIR99021: Glycogen synthase kinase 3 (GSK3) inhibitor (B) Representative IF images from nHDFs (Donor I) infected for 7 days with MCPyV. DAPI, LT and VP1 were stained as indicated. Scale bar represents 40 µm. (C) Total MCPyV genome copies were determined from the supernatant of either nHDFs (Donor I, n=1) or immortalized Bj5-ta hTert fibroblasts (n=2) that were infected with MCPyV. The x-axis represents the days post infection and the y-axis the total genome copies which were determined using primers against the VP1 gene.

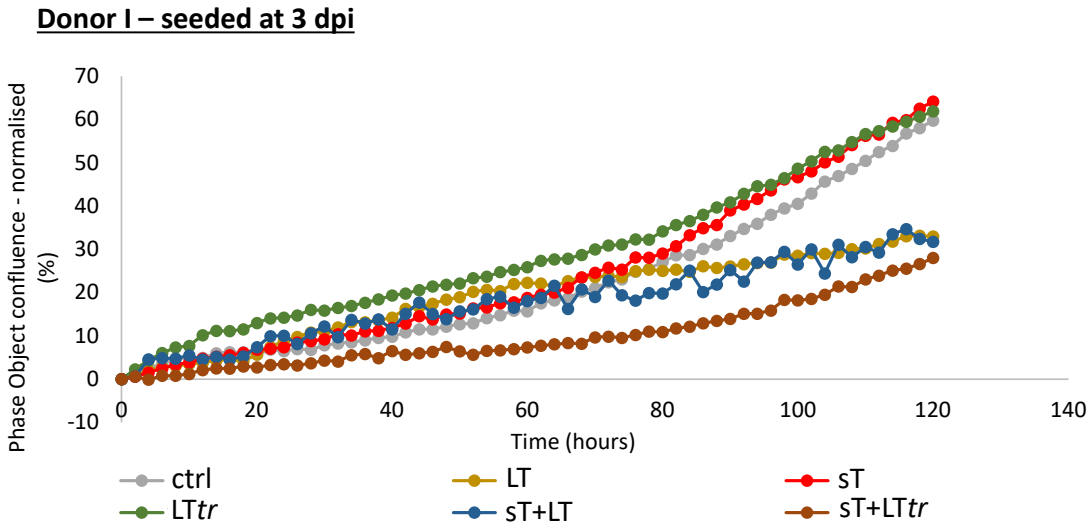


Figure S2 Comparative proliferation rates of transduced nHDFs
(A) NHDFs from Donor I were seeded in similar numbers at 3 dpi and their proliferation rate was assessed using the Incucyte® Live-Cell Analysis System. The x-axis represents the time in hours while the y-axis depicts the percentage of confluence. Every condition was measured in technical triplicates and the mean values of one biological replicate are depicted.

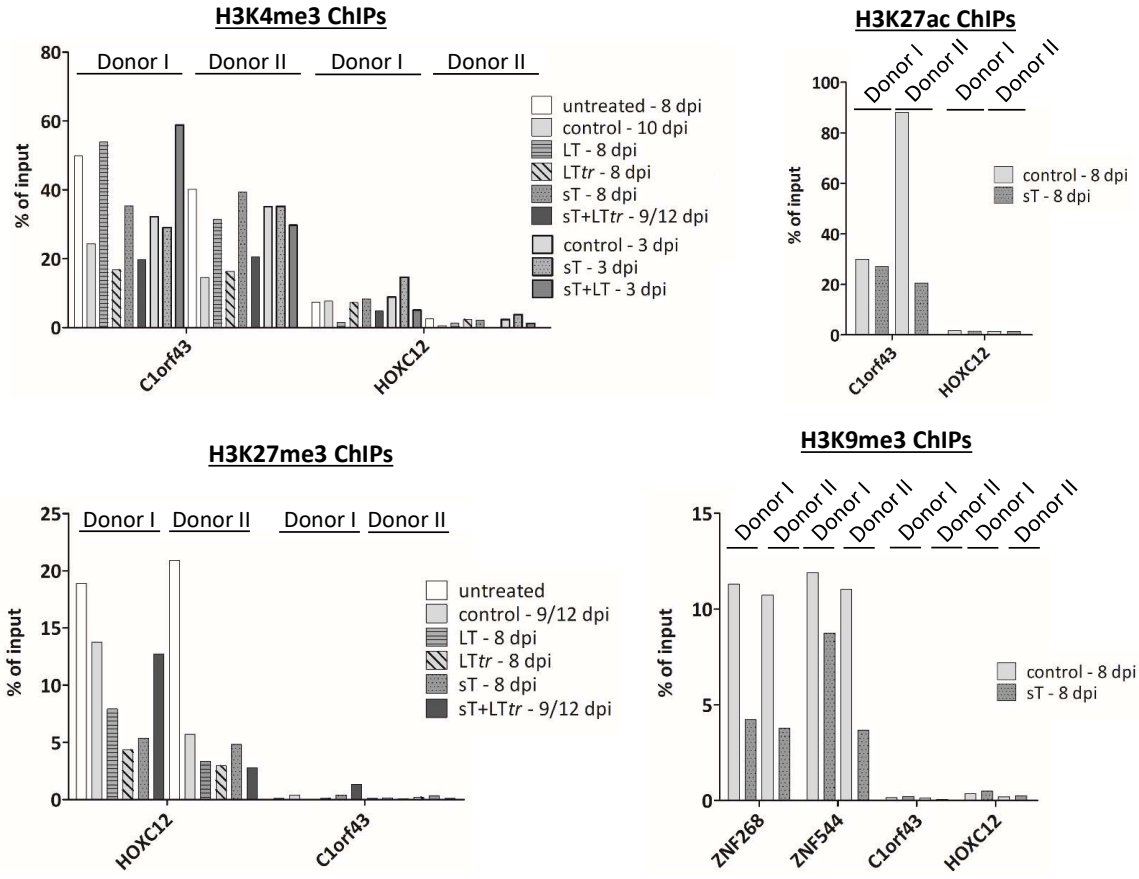


Figure S3 Quality control of sequenced ChIP samples
DNA obtained from ChIP experiments was used for qPCR analysis to test the efficiency of the conducted ChIPs. Enrichment is represented by the % input shown on the y-axis. *C1orf43* was used as a positive control for H3K4me3 and H3K27ac, while it served as a negative control for H3K27me3 and H3K9me3. *HOXC12* was used as a positive control for H3K27me3 and as a negative control for H3K4me3, H3K27ac and H3K9me3. For H3K9me3, *ZNF268* and *ZNF544* served as positive control primers.

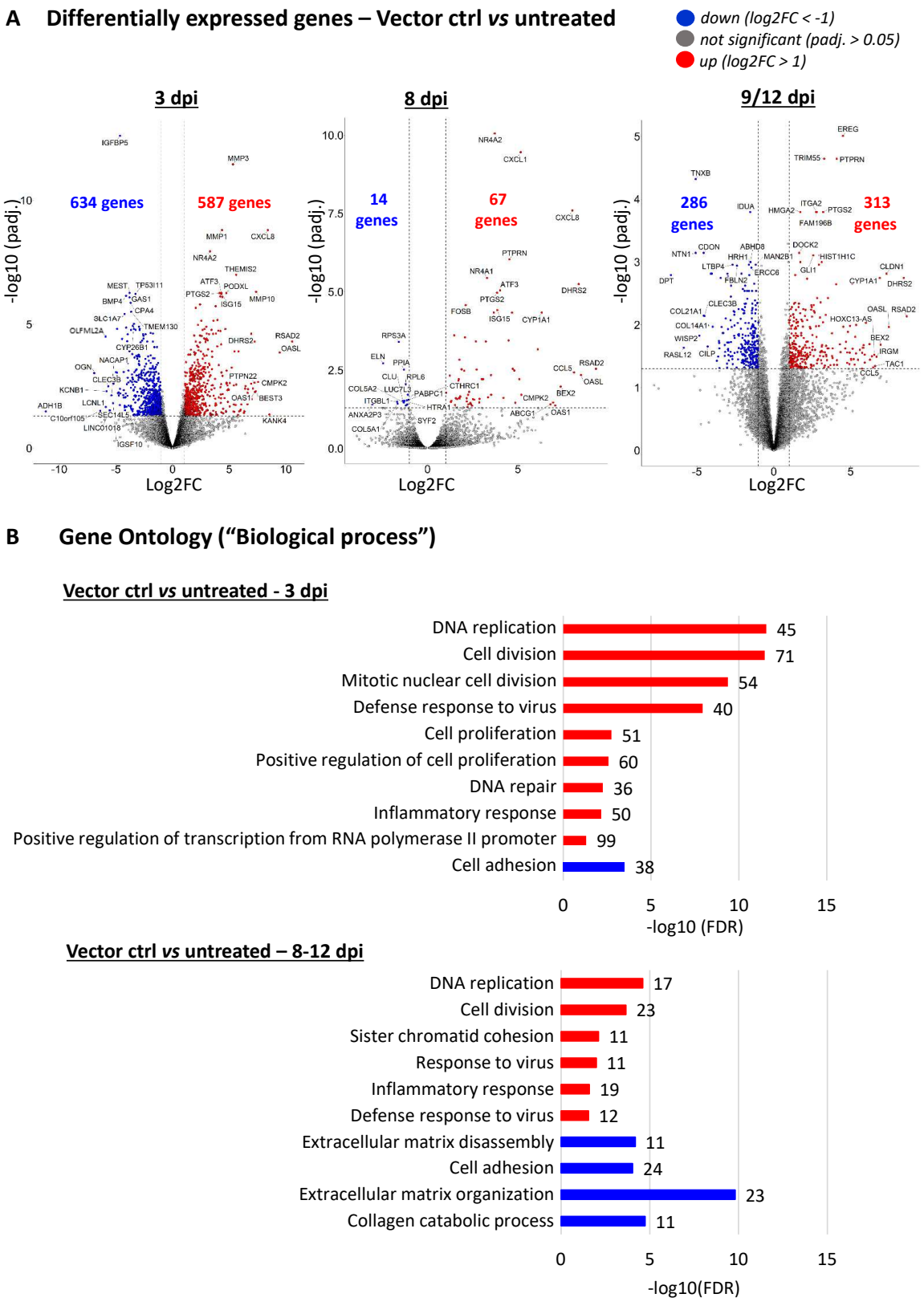


Figure S4 Transcriptional changes induced by lentiviral transduction and FACS-sorting.
(A) Volcano plots depicting all DEGs from the RNA-Seq analysis performed in two different nHDF donors comparing cells transduced with the vector ctrl *vs* untreated cells. Genes with a log2FC > 1 or < -1 and padj. < 0.05 are marked as red or blue dots, respectively. Genes with the 10 lowest padj. and the 10 lowest and highest log2FC values are labeled. (B) A GO analysis ("Biological process") was performed using the DAVID tool with significantly upregulated and downregulated genes at 3 and 8-12 dpi, summarizing the DEGs from the time points 8, 9 and 12 dpi. The 10 GO terms (FDR < 0.05, gene counts > 10) containing the highest numbers of DEGs are depicted as red or blue bars, reflecting up- or downregulated genes respectively. The numbers of DEGs found within each GO term are noted next to the bars, FDR values are shown on the x-axis.

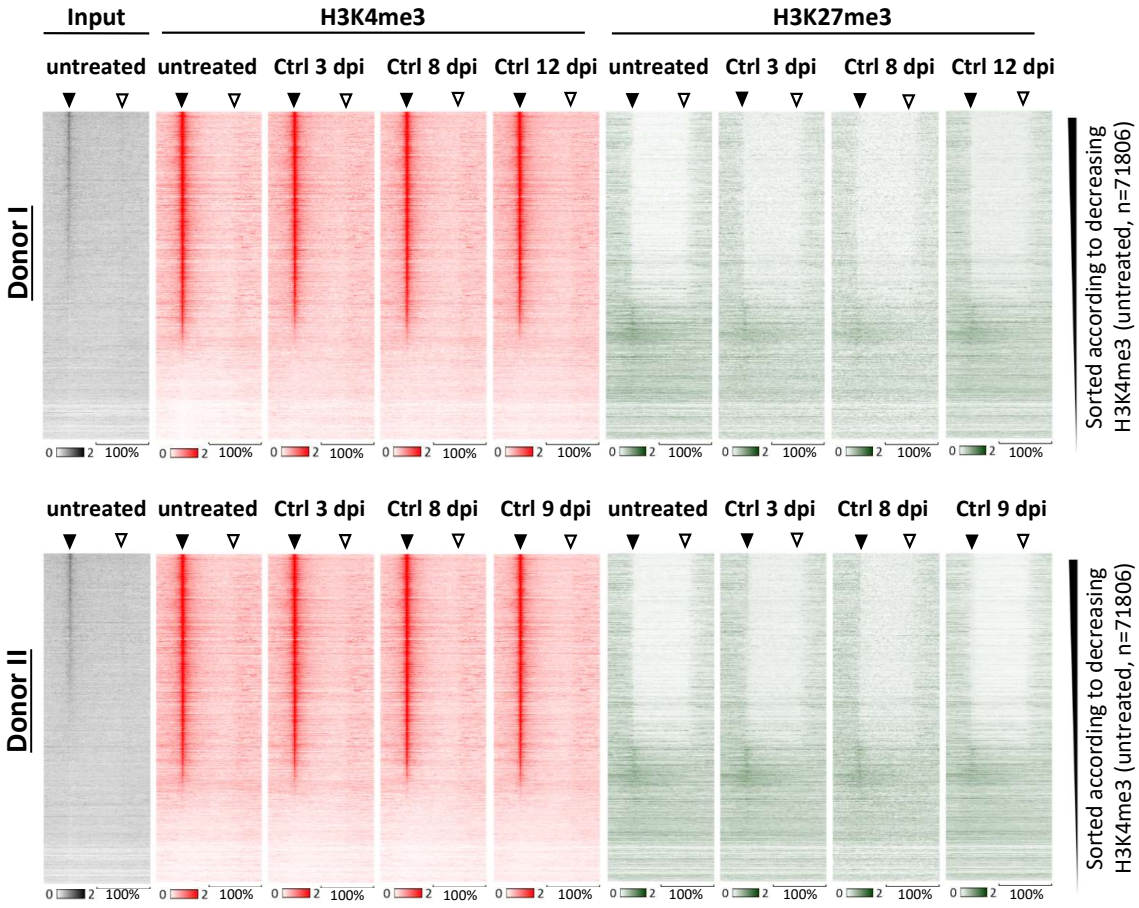
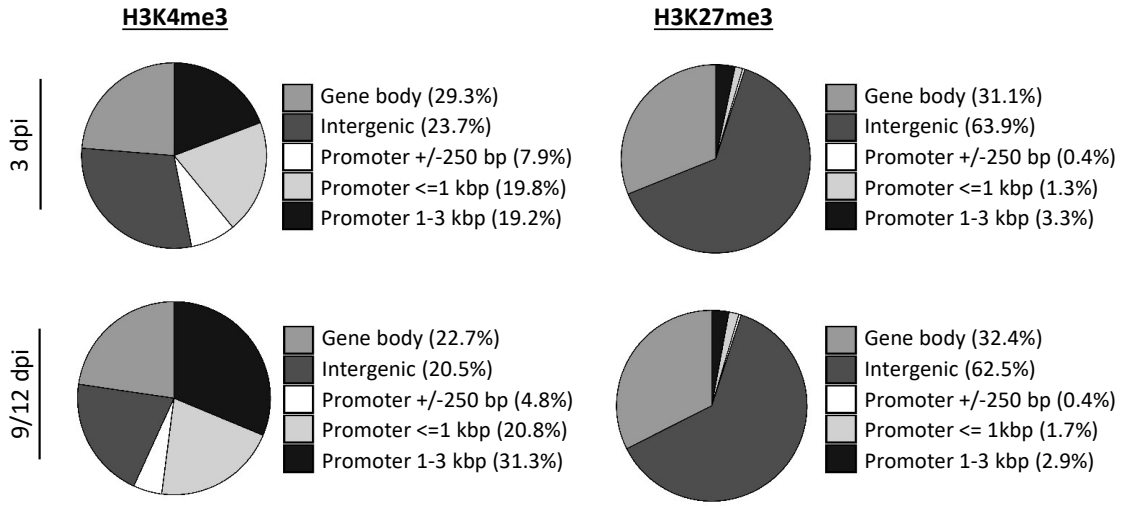


Figure S5 ChIP-Seq analysis for H3K4me3 and H3K27me3 in vector control-expressing nHDFs.
Heatmaps depicting ChIP-Seq read count signals quantified for all genes of the human genome (hg19) for input, H3K4me3 and H3K27me3 from two different donors of untreated nHDFs or cells overexpressing the vector controls (at 3, 8, 9 or 12 dpi). Relative lengths of each genomic locus are shown for 200%, including TSSs (black triangles) and TTSs (non-filled triangles). Heatmaps are sorted according to decreasing H3K4me3 signal in untreated cells. Density represents DNA fragments per 1 mio reads per 1 kbp.

A **DiffReps - Distribution of genomic features** (*log2FC > 0.5 or < -0.5, padj. < 0.05)*



B **Number of diffReps hits**

	diffReps	H3K4me3 total	H3K4me3 for annotated genes	H3K27me3 total	H3K27me3 for annotated genes
3 dpi	**log2FC > 0.5**	351	262	989	261
	log2FC < -0.5	1081	758	1328	427
	log2FC > 1	29	9	44	8
	log 2FC < -1	363	252	9	2
9/12 dpi	**log2FC > 0.5**	2140	1718	1435	341
	log2FC < -0.5	2118	1389	409	184
	log2FC > 1	176	65	21	3
	log 2FC < -1	482	321	4	3

Figure S6 Differential histone modification analysis in response to lentiviral transduction and FACS-sorting. (A) Distribution of genomic features contained in the diffReps analysis, which was performed for H3K4me3 and H3K27me3 ChIPs. All significant hits with a log2FC $> 0.5/< -0.5$ were classified according to their annotations. (B) Summary of total numbers of diffReps hits from the comparison of vector ctrl *vs* untreated after 3 or 9/12 dpi for H3K4me3 and H3K27me3.

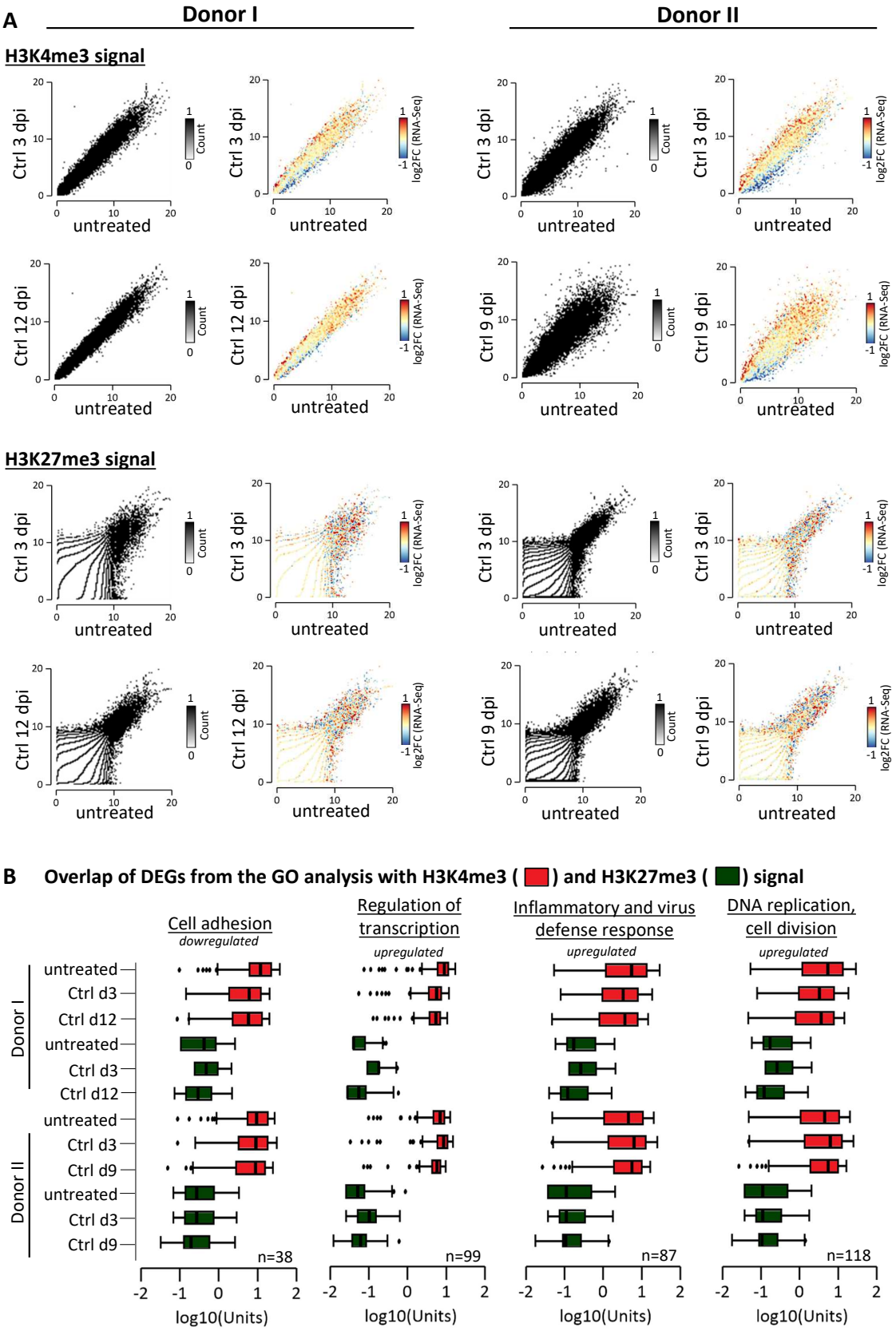


Figure S7 Correlation of ChIP-Seq with RNA-Seq analysis in transduced and untreated nHDFs.
(A) Quantification of H3K4me3 and H3K27me3 read signals for all genes from hg19 in nHDFs overexpressing the vector controls or untreated cells. The colored scatterplots incorporate the information on the log2FCs from the RNA-Seq analysis, represented by the color code. The x and y-axes were segmented into 100 bins and regions within these bins are depicted by the counts. (B) Subset of genes from the RNA-Seq GO analysis (see figure S4B) quantified for their signal in H3K4me3 and H3K27me3. Boxplots represent the distribution of histone modification signals around the TSSs (<= 3 kbp) for H3K4me3 (red) and H3K27me3 (green).

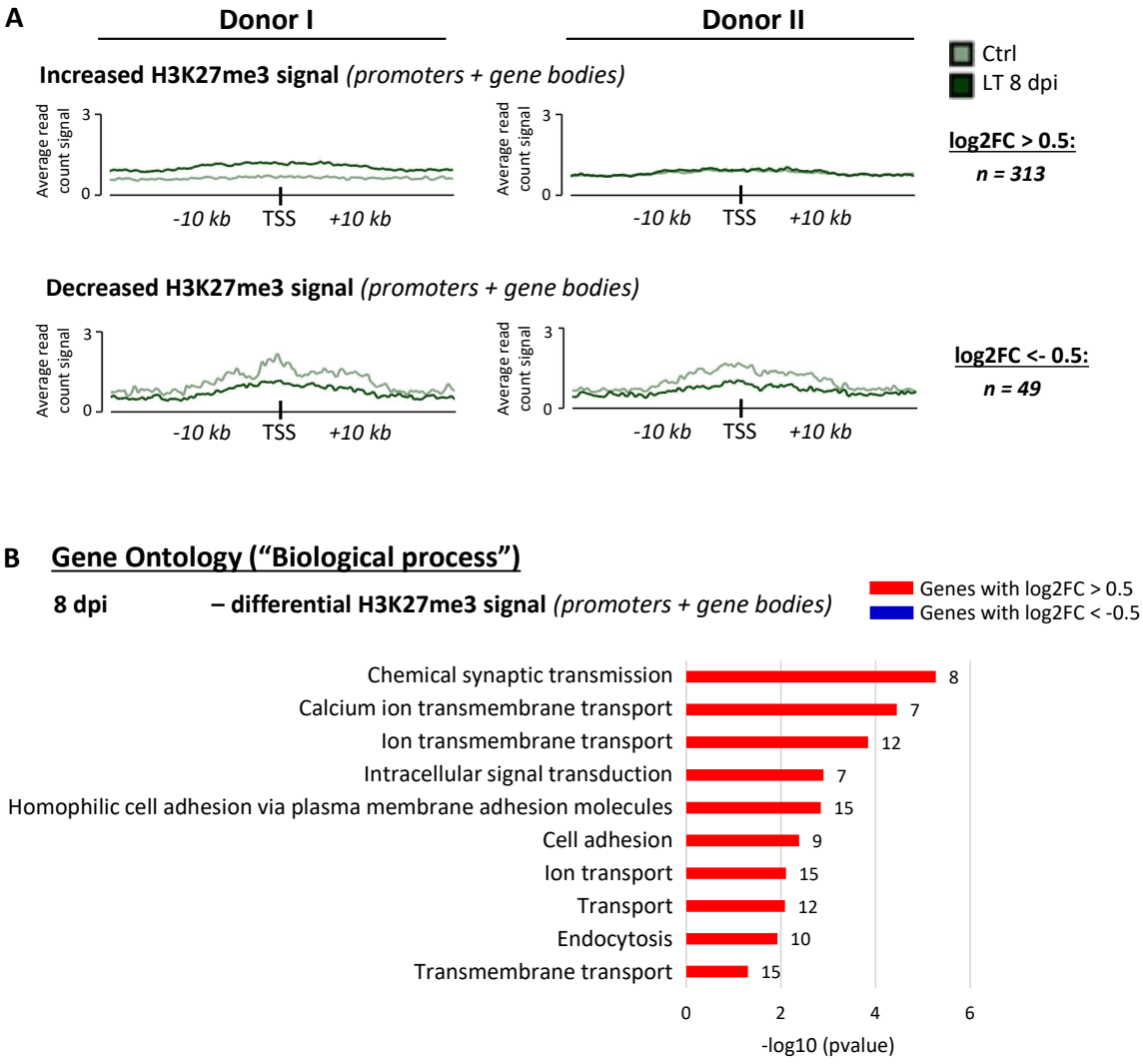


Figure S8 H3K27me3 diffReps analysis from LT *vs* ctrl at 8 dpi.
(A) Average read count signals of H3K27me3 for all genes with differential signal comparing LT *vs* ctrl at 8 dpi. The total numbers of genes with differential H3K27me3 signal around promoters (+/-3 kbp) or within gene bodies of annotated genes are shown next to the average plots. (B) A GO analysis ("Biological process") was performed with DAVID using genes from the diffReps analysis described in (A). Only gene sets with n$>$ 50 were used as an input for the GO analysis. From all GO terms with at least 5 genes and a pvalue $<$ 0.05, a maximum of 10 GO terms with the highest gene counts are shown.

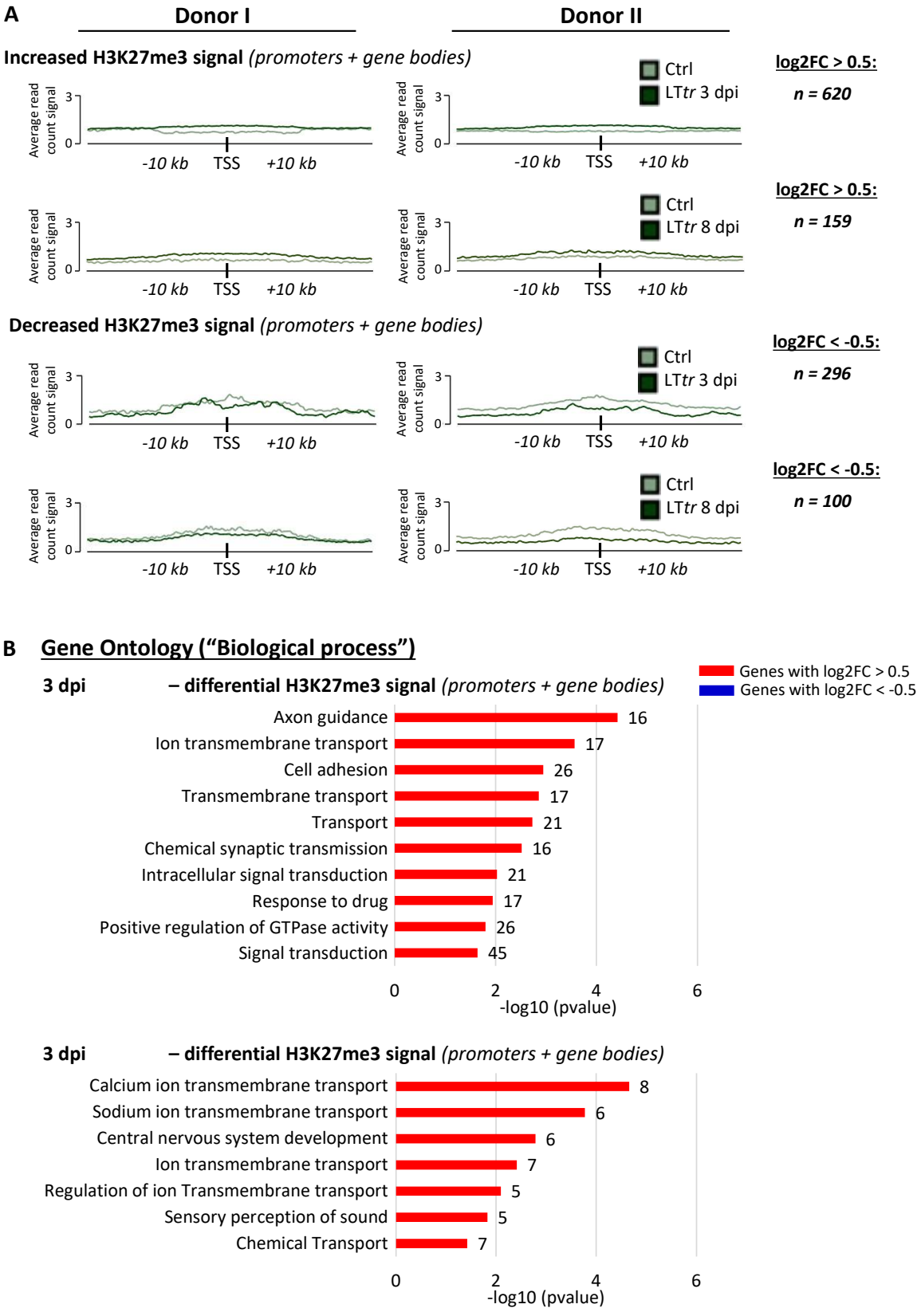


Figure S9 H3K27me3 diffReps analysis from LT*tr* *vs* ctrl at 3 and 8 dpi.
(A) Average read count signals of H3K27me3 for all genes with differential signal comparing LT*tr* vs ctrl at 3 and 8 dpi. DiffReps analysis was performed using two different donors and the total numbers of genes with differential H3K27me3 signal around promoters (+/- 3 kbp) or within gene bodies of annotated genes are shown next to the average plots. (B) A GO analysis ("Biological process") was performed with DAVID using genes from the diffReps analysis described in (A). Only gene sets with n> 50 were used as an input for the GO analysis. From all GO terms with at least 5 genes and a pvalue < 0.05, a maximum of 10 GO terms with the highest gene counts are shown.

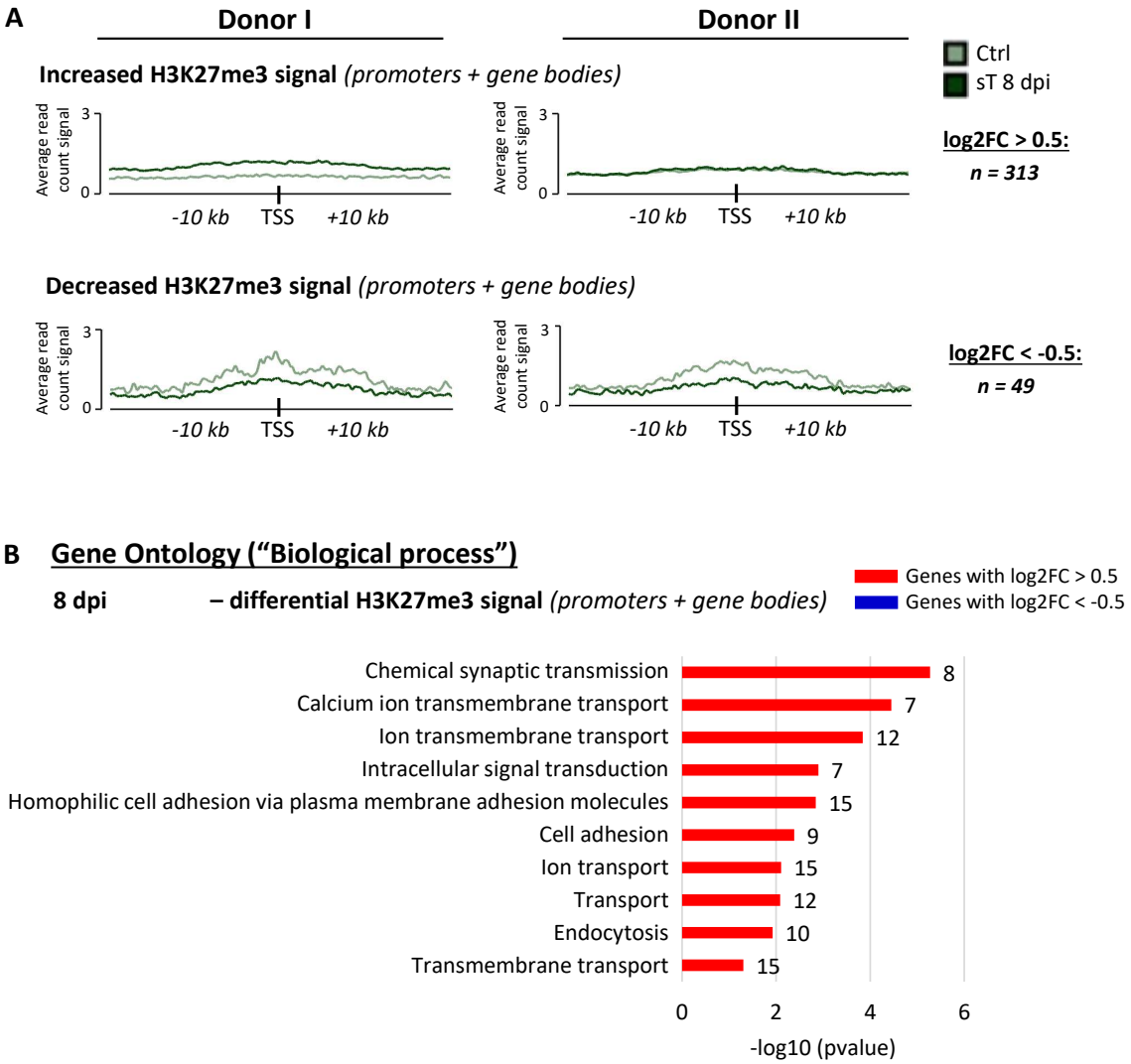


Figure S10 H3K27me3 diffReps analysis from sT *vs* ctrl at 8 dpi. (A) Average read count signals of H3K27me3 for all genes with differential signal comparing sT *vs* ctrl at 8 dpi. The total numbers of genes with differential H3K27me3 signal around promoters (+/-3 kbp) or within gene bodies of annotated genes are shown next to the average plots. (B) A GO analysis ("Biological process") was performed with DAVID using genes from the diffReps analysis described in (A). Only gene sets with n> 50 were used as an input for the GO analysis. From all GO terms with at least 5 genes and a pvalue < 0.05, a maximum of 10 GO terms with the highest gene counts are shown.

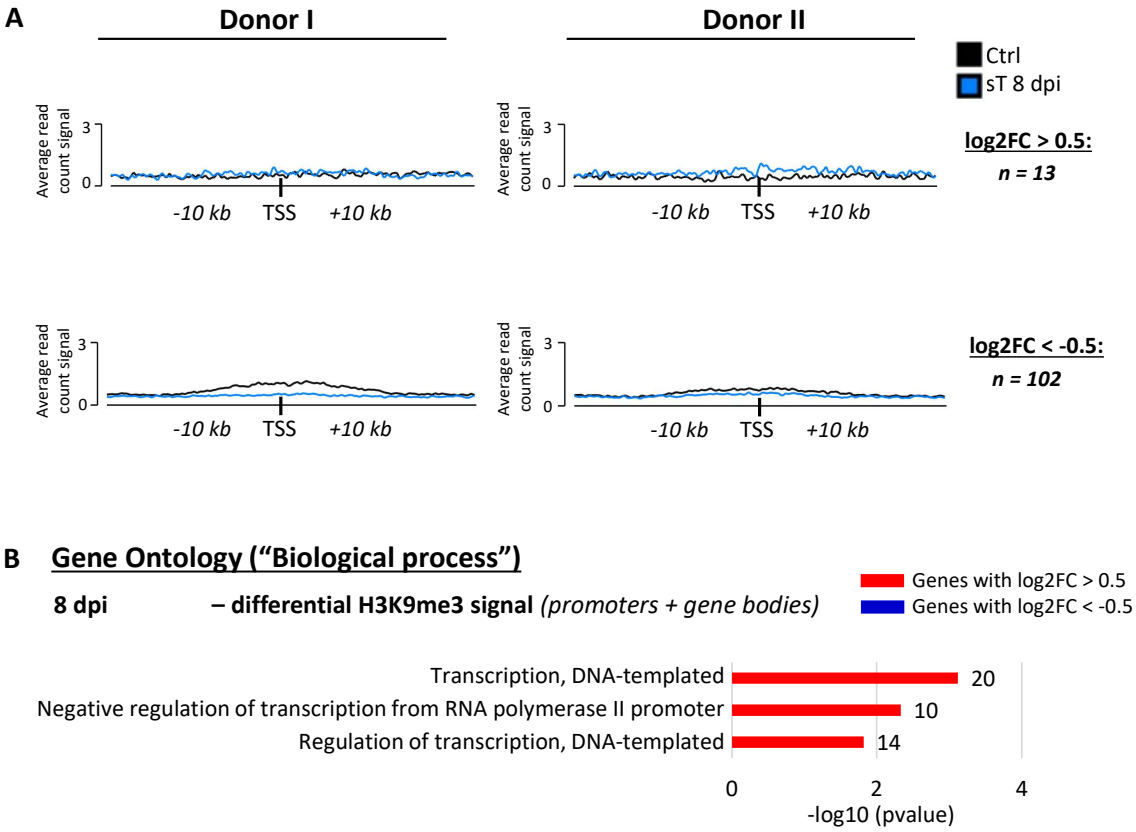


Figure S11 H3K9me3 diffReps analysis from sT *vs* ctrl at 8 dpi.
(A) Average read count signals of H3K9me3 for all genes with differential signal comparing sT *vs* ctrl at 8 dpi. The total numbers of genes with differential H3K9me3 signal around promoters (+/-3 kbp) or within gene bodies of annotated genes are shown next to the average plots. (B) A GO analysis ("Biological process") was performed with DAVID using genes from the diffReps analysis described in (A). Only gene sets with n> 50 were used as an input for the GO analysis. From all GO terms with at least 5 genes and a pvalue < 0.05, a maximum of 10 GO terms with the highest gene counts are shown.

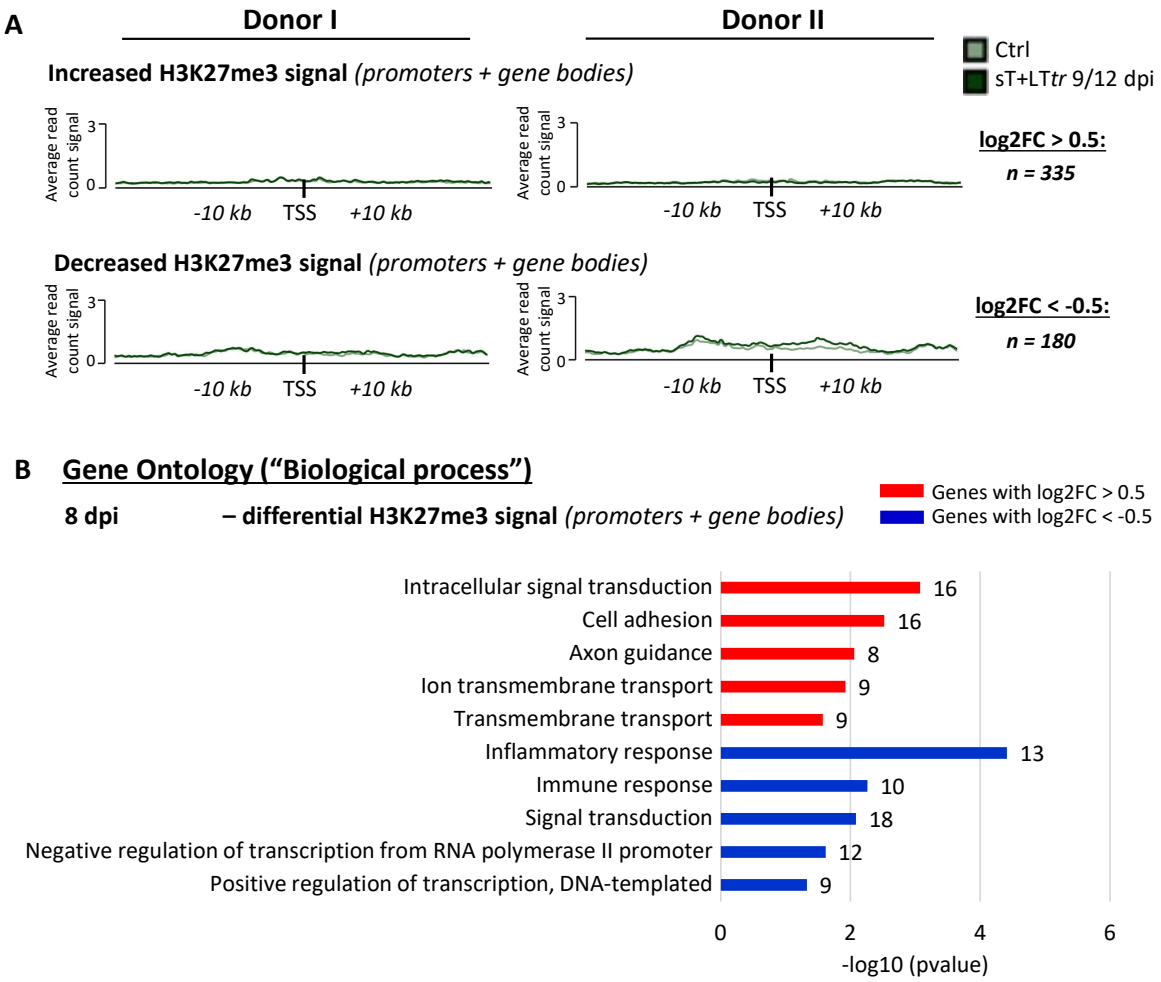


Figure S12 H3K27me3 diffReps analysis from sT+LT*tr vs* ctrl at 9/12 dpi.
(A) Average read count signals of H3K27me3 for all genes with differential signal comparing sT+LT*tr vs* ctrl at 9/12 dpi. The total numbers of genes with differential H3K27me3 signal around promoters (+/-3 kbp) or within gene bodies of annotated genes are shown next to the average plots. (B) A GO analysis ("Biological process") was performed with DAVID using genes from the diffReps analysis described in (A). Only gene sets with n$>$ 50 were used as an input for the GO analysis. From all GO terms with at least 5 genes and a pvalue $<$ 0.05, a maximum of 10 GO terms with the highest gene counts are shown.

Table S1 Sequencing depths of the RNA-Seq and ChIP-Seq analysis.

(a) Total numbers of unique RNA-Seq reads. M=million.

RNA-Seq				
Condition	**Donor I**		**Donor II**	
untreated	14M		20M	
	3 dpi	**8 dpi**	**3 dpi**	**8 dpi**
GFP control	28M	32M	20M	16M
LT	17M	19M	20M	25M
LT*tr*	26M	32M	18M	17M
mCherry control	26M	22M	21M	21M
sT	27M	21M	21M	22M
sT+LT	9M	-	7M	-
	12 dpi		**9 dpi**	
GFP+mCherry control	18M		18M	
sT+LT*tr*	18M		21M	

(b) Total numbers of ChIP-Seq unique reads. M=million, D=Donor.

ChIP-Seq								
Condition	**H3K4me3**		**H3K27me3**		**H3K9me3**		**H3K27ac**	
	DI	**DII**	**DI**	**DII**	**DI**	**DII**	**DI**	**DII**
untreated	10M	11M	10M	15M	-	-	-	-
vector control 3 dpi	7M	11M	5M	17M	-	-	-	-
vector control 8 dpi	-	-	-	-	37M	26M	30M	29M
vector control 9/12 dpi	11M	20M	15M	32M	-	-	-	-
sT 3 dpi	7M	12M	7M	15M	-	-	-	-
LT 3 dpi	-	-	-	-	-	-	-	-
sT+LT 3 dpi	14M	20M	-	-	-	-	-	-
LT*tr* 3 dpi	18M	19M	22M	23M	-	-	-	-
sT 8 dpi	6M	13M	6M	7M	29M	31M	35M	12M
LT 8 dpi	16M	28M	24M	24M	-	-	-	-
LT*tr* 8 dpi	21M	20M	32M	21M	-	-	-	-
sT+LT*tr* 9/12 dpi	14M	22M	13M	32M	-	-	-	-

Table S2 Top DEGs from ctrl at 3 dpi *vs* untreated nHDFs

(a) Top 10 upregulated/downregulated and most significant genes

Symbol	Gene name	log2FC	padj.
RSAD2[8]	radical S-adenosyl methionine domain containing 2	10.56	5.00E-05
OASL[8]	2′-5′-oligoadenylate synthetase like	9.47	1.40E-04
KANK4	KN motif and ankyrin repeat domains 4	8.55	4.30E-02
CXCL8[6]	C-X-C motif chemokine ligand 8	8.42	1.60E-09
CMPK2	cytidine/uridine monophosphate kinase 2	7.55	2.20E-03
MMP10	matrix metallopeptidase 10	7.38	5.00E-07
BEST3	bestrophin 3	7.34	5.10E-03
OAS1[8]	2′-5′-oligoadenylate synthetase 1	7.27	5.50E-03
DHRS2	dehydrogenase/reductase 2	7.24	5.00E-05
PTPN22	protein tyrosine phosphatase non-receptor type 22	7.23	2.30E-03
THEMIS2	thymocyte selection associated family member 2	5.62	1.00E-07
MMP3	matrix metallopeptidase 3	5.34	3.60E-12
PODXL	podocalyxin like	4.75	5.50E-07
MMP1	matrix metallopeptidase 1	4.37	1.60E-09
ISG15[8]	ISG15 ubiquitin like modifier	4.32	6.00E-07
ATF3[1,6]	activating transcription factor 3	4.3	5.50E-07
PTGS2[6]	prostaglandin-endoperoxide synthase 2	4.14	5.50E-07
NR4A2[1]	nuclear receptor subfamily 4 group A member 2	3.31	1.20E-08
ADH1B	alcohol dehydrogenase 1B (class I), beta polypeptide	-11.19	3.30E-02
OGN	osteoglycin	-6.91	9.40E-04
C10orf105	chromosome 10 open reading frame 105	-6.39	3.10E-02
TM4SF20	transmembrane 4 L six family member 20	-5.92	3.10E-05
LCNL1	lipocalin like 1	-5.89	2.20E-02
KCNB1	potassium voltage-gated channel subfamily B member 1	-5.82	5.10E-03
CLEC3B	C-type lectin domain family 3 member B	-5.68	3.20E-03
SEC14L5	SEC14 like lipid binding 5	-5.56	3.00E-02
LINC01018	long intergenic non-protein coding RNA 1018	-5.53	3.00E-02
IGSF10	immunoglobulin superfamily member 10	-5.25	1.50E-02
IGFBP5	insulin like growth factor binding protein 5	-4.63	2.50E-13
SLC1A7	solute carrier family 1 member 7	-4.24	3.90E-06
BMP4	bone morphogenetic protein 4	-4.1	7.20E-07
GAS1	growth arrest specific 1	-3.8	8.30E-07
MEST	mesoderm specific transcript	-3.79	5.40E-07
CPA4	carboxypeptidase A4	-3.69	1.40E-06
TMEM130	transmembrane protein 130	-3.65	3.20E-06
OLFML2A	olfactomedin like 2A	-3.52	1.20E-05
CYP26B1	cytochrome P450 family 26 subfamily B member 1	-3.4	9.70E-06
TP53I11	tumor protein p53 inducible protein 11	-3.27	6.00E-07

(b) Top 10 GO terms

	Gene name	Gene count	-log10 (FDR)
1	Positive regulation of transcription from RNA polymerase II promoter	99	1.3
2	Cell division	71	11.5
3	Positive regulation of cell proliferation	60	2.6
4	Mitotic nuclear division	54	9.4
5	Cell proliferation	51	2.7
6	Inflammatory response	50	2.2
7	DNA replication	45	11.6
8	Defense response to virus	40	7.9
9	Cell adhesion	38	3.5
10	DNA repair	36	2.3

Table S3 Top DEGs from ctrl at 8-12 dpi *vs* untreated nHDFs.

(a) Top 10 upregulated/downregulated and most significant genes

Symbol	Gene name	log2FC	padj.
RSAD2[5,7]	radical S-adenosyl methionine domain containing 2	9.32	2.90E-03
OASL[5,7]	2′-5′-oligoadenylate synthetase like	8.5	4.50E-03
DHRS2	dehydrogenase/reductase (SDR family) member 2	8.46	1.80E-03
CCL5[6,7]	C-C motif chemokine ligand 5	8.11	3.80E-03
CXCL8[6]	C-X-C motif chemokine ligand 8	8.02	2.50E-08
BEX2	brain expressed X-linked 2	7.37	1.10E-02
CLDN1	claudin 1	7.33	1.50E-03
OAS1[5,7]	2′-5′-oligoadenylate synthetase 1	7.07	4.30E-02
CMPK2	cytidine/uridine monophosphate kinase 2	6.95	3.50E-02
CYP1A1[7]	cytochrome P450, family 1, subfamily A, polypeptide 1	6.9	1.80E-03
ABCG1	ATP binding cassette subfamily G member 1	6.8	3.90E-02
IRGM[6]	immunity-related GTPase family, M	6.69	3.20E-02
CXCL1[6]	C-X-C motif chemokine ligand 1	5.16	3.50E-10
PTPRN	protein tyrosine phosphatase receptor type N	4.54	9.20E-07
EREG	epiregulin	4.5	9.90E-06
ATF3	activating transcription factor 3	4.02	9.00E-06
PTGS2[5]	prostaglandin-endoperoxide synthase 2	3.85	1.10E-05
NR4A2	nuclear receptor subfamily 4 group A member 2	3.72	8.80E-11
NR4A1	nuclear receptor subfamily 4 group A member 1	3.3	3.60E-06
TRIM55	tripartite motif containing 55	3.28	2.30E-05
DPT[3]	dermatopontin	-6.7	1.60E-03
RASL12	RAS-like, family 12	-5.85	2.30E-02
NTN1	netrin 1	-5.08	7.20E-04
TNXB	tenascin XB	-5.08	4.80E-05
WISP2[3]	WNT1 inducible signaling pathway protein 2	-4.84	1.50E-02
CDON[3]	cell adhesion associated, oncogene regulated	-4.56	7.20E-04
COL21A1	collagen, type XXI, alpha 1	-4.55	7.20E-03
CLEC3B	C-type lectin domain family 3, member B	-4.51	7.30E-03
COL14A1[2]	collagen, type XIV, alpha 1	-4.34	1.10E-02
CILP	cartilage intermediate layer protein, nucleotide pyrophosphohydrolase	-4.3	2.20E-02
ABCA9	ATP-binding cassette, sub-family A (ABC1), member 9	-4.27	1.00E-02
PRELP	proline/arginine-rich end leucine-rich repeat protein	-4.1	1.50E-03
CFD	complement factor D (adipsin)	-4.02	1.50E-03
LTBP4	latent transforming growth factor beta binding protein 4	-2.68	1.10E-03
FBLN2	fibulin 2	-2.39	1.10E-03
HRH1	histamine receptor H1	-1.62	1.10E-03
RPS3A	ribosomal protein S3A	-1.59	3.90E-04
IDUA	iduronidase, alpha-L-	-1.53	1.60E-04
ABHD8	abhydrolase domain containing 8	-1.51	1.00E-03
MAN2B1	mannosidase, alpha, class 2B, member 1	-1.17	1.10E-03

(b) Top 10 GO terms

	Gene name	Gene count	-log10 (FDR)
1	Collagen catabolic process	11	4.7
2	Extracellular matrix organization	23	9.8
3	Cell adhesion	24	4
4	Extracellular matrix disassembly	11	4.2
5	Defense response to virus	12	1.5
6	Inflammatory response	19	1.6
7	Response to virus	11	2
8	Sister chromatid cohesion	11	2.1
9	Cell division	23	3.7
10	DNA replication	17	4.6

Table S4 Top DEGs from LT *vs* ctrl at 3 dpi.

(a) Top 10 upregulated/downregulated and most significant genes

Symbol	Gene name	log2FC	padj.
TNFSF10[5]	TNF superfamily member 10	9.53	5.20E-05
GBP5[4]	guanylate binding protein 5	8.89	2.80E-04
GBP1P1	guanylate binding protein 1 pseudogene 1	8.46	3.10E-04
UBD	ubiquitin D	7.94	7.80E-03
CXCL10[3,4,5]	C-X-C motif chemokine ligand 10	7.91	1.00E-03
NEURL3	neuralized E3 ubiquitin protein ligase 3	7.65	3.50E-03
PRRG4	proline rich and Gla domain 4	7.65	3.90E-04
RSAD2[2,3]	radical S-adenosyl methionine domain containing 2	7.58	5.50E-05
CXCL11[4,5]	C-X-C motif chemokine ligand 11	7.49	8.20E-06
CYP2J2	cytochrome P450 family 2 subfamily J member 2	7.39	1.80E-03
CASP8AP2[6,7]	caspase 8 associated protein 2	2.35	5.50E-19
USP37[10]	ubiquitin specific peptidase 37	2.09	1.30E-14
TIFA	TRAF interacting protein with forkhead associated domain	2.21	1.20E-10
ZNF347[6,7]	zinc finger protein 347	1.57	2.10E-09
GBP4	guanylate binding protein 4	6.59	1.70E-08
SLFN13	schlafen family member 13	4.38	1.70E-08
BIRC3	baculoviral IAP repeat containing 3	4.33	1.70E-08
USP1	ubiquitin specific peptidase 1	1.95	1.70E-08
MSH6[8]	mutS homolog 6	1.66	1.70E-08
PPIG	peptidylprolyl isomerase G	1.09	1.70E-08
MACROD2	mono-ADP ribosylhydrolase 2	-7.15	7.00E-04
FAM13A-AS1	FAM13A antisense RNA 1	-6.46	4.80E-03
LINC01391	long intergenic non-protein coding RNA 1391	-6.14	2.10E-02
CD200	CD200 molecule	-6.13	2.40E-02
CHMP3	charged multivesicular body protein 3	-6.08	2.20E-02
LINC00926	long intergenic non-protein coding RNA 926	-5.88	7.20E-03
F12	coagulation factor XII	-5.81	5.00E-03
CACNB2	calcium voltage-gated channel auxiliary subunit beta 2	-5.71	1.10E-02
GNRHR2	gonadotropin releasing hormone receptor 2 (pseudogene)	-5.68	1.20E-02
USP44	ubiquitin specific peptidase 44	-5.65	7.40E-03
FZD2	frizzled class receptor 2	-1.75	6.60E-10
MED16	mediator complex subunit 16	-1.21	7.90E-09
DALRD3	DALR anticodon binding domain containing 3	-2	1.70E-08
ATP6AP1	ATPase H+ transporting accessory protein 1	-1.22	1.70E-08
HSD3B7	hydroxy-delta-5-steroid dehydrogenase, 3 beta- and steroid delta-isomerase 7	-1.53	1.80E-08
MKNK2	MAPK interacting serine/threonine kinase 2	-1.57	3.10E-08
ZBTB12	zinc finger and BTB domain containing 12	-2.96	9.30E-08
TPRG1L	tumor protein p63 regulated 1 like	-1.25	1.10E-07
DEXI	Dexi homolog	-1.51	1.50E-07
ATF6B	activating transcription factor 6 beta	-1.12	4.60E-07

(b) Top 10 GO terms

	Gene name	Gene count	-log10 (FDR)
1	DNA replication	48	20.6
2	Type I interferon signaling pathway	31	18.8
3	Defense response to virus	46	18
4	Inflammatory response	51	7.2
5	Immune response	53	6.6
6	Regulation of transcription, DNA-templated	125	5.9
7	Transcription, DNA-templated	151	5.5
8	DNA repair	34	5
9	Innate immune response	41	2.1
10	Cell division	33	1.5

Table S5 Top DEGs from LT *vs* ctrl at 8 dpi.

(a) Top 10 upregulated/downregulated and most significant genes

Symbol	Gene name	log2FC	padj.
CSF3	colony stimulating factor 3	10.07	3.69E-06
EHF[9]	ETS homologous factor	8.37	1.64E-06
BCL2A1	BCL2 related protein A1	8.02	8.36E-05
GPR4	G protein-coupled receptor 4	7.86	2.63E-08
LHX2	LIM homeobox 2	7.73	6.43E-04
LPL	lipoprotein lipase	7.59	1.58E-03
WNT7B	Wnt family member 7B	7.53	2.44E-03
VAV3	vav guanine nucleotide exchange factor 3	7.28	1.30E-03
NPY[9]	neuropeptide Y	7.16	2.38E-02
PTGDR	prostaglandin D2 receptor	7.1	1.33E-02
CXCL1[9,10]	C-X-C motif chemokine ligand 1	6.33	3.69E-51
CA12	carbonic anhydrase 12	3.72	3.57E-25
FAM111B	family with sequence similarity 111 member B	3.97	3.03E-20
PLAU	plasminogen activator, urokinase	3.73	2.69E-18
EZH2	enhancer of zeste 2 polycomb repressive complex 2 subunit	3.1	2.97E-18
MMP1	matrix metallopeptidase 1	4.42	2.60E-16
SLFN13	schlafen family member 13	6.14	2.63E-16
TMEM179	transmembrane protein 179	6.74	3.29E-16
CNIH3	cornichon family AMPA receptor auxiliary protein 3	4.98	3.29E-16
WDR76	WD repeat domain 76	3.11	3.29E-16
MKX	mohawk homeobox	-2.78	5.11E-14
MFGE8[8]	milk fat globule-EGF factor 8 protein	-2.67	2.23E-12
WNT2	Wnt family member 2	-5.25	3.38E-12
COL1A1[7,8]	collagen type I alpha 1 chain	-3.67	7.21E-12
KIF26B	kinesin family member 26B	-3.31	2.77E-11
LBH	LBH regulator of WNT signaling pathway	-4	1.43E-10
LOXL2[8]	lysyl oxidase like 2	-2.28	7.30E-10
FSTL1	follistatin like 1	-2.54	8.31E-10
LOX[7]	lysyl oxidase	-1.82	1.11E-09
PDGFRB	platelet derived growth factor receptor beta	-2.59	1.28E-09
LGR5	leucine rich repeat containing G protein-coupled receptor 5	-8.49	3.34E-03
C7orf69	chromosome 7 open reading frame 69	-7.74	1.60E-03
FGF9	fibroblast growth factor 9	-7.54	5.91E-03
PADI1	peptidyl arginine deiminase 1	-7.26	4.10E-03
KCNB1	potassium voltage-gated channel subfamily B member 1	-7.02	2.39E-02
LCNL1	lipocalin like 1	-6.71	1.53E-02
LOC100130331	POTE ankyrin domain family, member F pseudogene	-5.97	2.94E-03
ELN[7]	elastin	-5.97	1.11E-04
OR11H2	olfactory receptor family 11 subfamily H member 2	-5.9	9.33E-03
RASGRP2	RAS guanyl releasing protein 2	-5.89	5.67E-04

(b) Top 10 GO terms

	Gene name	Gene count	-log10 (FDR)
1	DNA replication	71	49.3
2	Cell division	76	27.5
3	DNA repair	59	24.1
4	Mitotic nuclear division	58	22.1
5	G1/S transition of mitotic cell cycle	38	21.2
6	Sister chromatid cohesion	33	15.9
7	Extracellular matrix organization	34	13.1
8	Cell adhesion	44	8.9
9	Cell proliferation	41	5
10	Inflammatory response	33	1.8

Table S6 Top DEGs from LT*tr vs* ctrl at 3 dpi.

(a) Top 10 upregulated/downregulated and most significant genes

Symbol	Gene name	log2FC	padj.
PRRG4	proline rich and Gla domain 4	6.95	1.30E-03
TMEM179	transmembrane protein 179	6.86	4.40E-15
TNFSF10[8,9]	TNF superfamily member 10	6.65	2.20E-03
GBP5[9]	guanylate binding protein 5	6.02	1.00E-04
CMPK2	cytidine/uridine monophosphate kinase 2	5.89	7.10E-05
NRIR	negative regulator of interferon response	5.86	3.40E-02
NEURL3	neuralized E3 ubiquitin protein ligase 3	5.78	3.60E-03
CXCL11[9]	C-X-C motif chemokine ligand 11	5.6	4.40E-06
OAS1[2,3]	2′-5′-oligoadenylate synthetase 1	5.54	3.60E-04
DRAXIN	dorsal inhibitory axon guidance protein	5.54	1.70E-02
CCL5[9]	C-C motif chemokine ligand 5	5.52	2.50E-11
CDCP1	CUB domain containing protein 1	2.55	1.90E-23
CLDN1	claudin 1	3.62	1.20E-13
USP18	ubiquitin specific peptidase 18	4.74	1.50E-12
BIRC3[8]	baculoviral IAP repeat containing 3	2.88	2.60E-12
TNFAIP3[8,9]	TNF alpha induced protein 3	2.52	2.00E-11
CASP8AP2	caspase 8 associated protein 2	1.9	6.80E-11
SCUBE2	signal peptide, CUB domain and EGF like domain containing 2	2.15	1.20E-10
TIFA	TRAF interacting protein with forkhead associated domain	2.04	4.20E-10
MTRNR2L8	MT-RNR2 like 8	-4.3	1.60E-03
ASPN	asporin	-2.47	4.90E-03
TNFRSF10D	TNF receptor superfamily member 10d	-2.19	4.30E-06
PCDHGB7	protocadherin gamma subfamily B, 7	-1.83	6.20E-05
PCDHGA3	protocadherin gamma subfamily A, 3	-1.78	5.00E-02
MAF	MAF bZIP transcription factor	-1.74	1.50E-03
MEST	mesoderm specific transcript	-1.68	3.90E-02
ROBO2	roundabout guidance receptor 2	-1.62	4.10E-02
FOSB	FosB proto-oncogene, AP-1 transcription factor subunit	-1.6	5.50E-03
CPA4	carboxypeptidase A4	-1.59	1.60E-02
ITGA8	integrin subunit alpha 8	-1.58	7.20E-03
WNT5A	Wnt family member 5A	-1.13	1.20E-08
COL1A1	collagen type I alpha 1 chain	-1.05	4.40E-07
EMX2OS	EMX2 opposite strand/antisense RNA	-1.39	5.60E-06
MIR22HG	MIR22 host gene	-1.12	1.40E-05
SUMF1	sulfatase modifying factor 1	-1.24	4.00E-05
CCNG2	cyclin G2	-1.47	7.80E-05
FKBP14	FKBP prolyl isomerase 14	-1.01	7.80E-05
TMEM150A	transmembrane protein 150A	-1.34	1.80E-04

(b) Top 10 GO terms

	Gene name	Gene count	-log10 (FDR)
1	DNA replication	51	32.1
2	Defense response to virus	37	16.4
3	Type I interferon signaling pathway	25	16.4
4	G1/S transition of mitotic cell cycle	30	16.3
5	DNA repair	42	14.7
6	Cell division	39	7.8
7	Cellular response to DNA damage stimulus	26	5.6
8	Apoptotic process	38	2.4
9	Inflammatory response	28	2.1
10	Innate immune response	29	1.7

Table S7 Top DEGs from LT*tr vs* ctrl at 8 dpi.

(a) Top 10 upregulated/downregulated and most significant genes

Symbol	Gene name	log2FC	padj.
CSF3	colony stimulating factor 3	9.45	9.40E-08
TNF[7]	tumor necrosis factor	7.03	8.70E-04
TMEM179	transmembrane protein 179	6.55	1.50E-18
EHF[1,2]	ETS homologous factor	6.54	5.60E-04
LHX2	LIM homeobox 2	6.47	5.20E-03
CSF2	colony stimulating factor 2	6.46	8.30E-08
DRAXIN	dorsal inhibitory axon guidance protein	6.41	1.60E-03
ABI3	ABI family member 3	6.37	1.30E-02
STOX1[2,4]	storkhead box 1	6.15	5.70E-03
WNT7B	Wnt family member 7B	6.06	1.30E-02
CDCP1	CUB domain containing protein 1	3.37	1.80E-45
TM4SF1	transmembrane 4 L six family member 1	3.69	6.20E-36
ENPP1	ectonucleotide pyrophosphatase/phosphodiesterase 1	2.37	2.60E-20
CNIH3	cornichon family AMPA receptor auxiliary protein 3	4.3	3.80E-19
MMP1	matrix metallopeptidase 1	3.45	4.50E-18
G0S2	G0/G1 switch 2	4.02	2.40E-17
SCARA3	scavenger receptor class A member 3	2.35	6.40E-15
RRAD	RRAD, Ras related glycolysis inhibitor and calcium channel regulator	3.18	8.10E-15
ADIRF	adipogenesis regulatory factor	1.96	5.50E-14
TM4SF20	transmembrane 4 L six family member 20	-5.41	2.30E-06
SLITRK6	SLIT and NTRK like family member 6	-5.29	8.90E-03
H19	H19 imprinted maternally expressed transcript	-5.04	2.70E-06
PTGDS	prostaglandin D2 synthase	-4.75	1.30E-03
C7orf69	chromosome 7 open reading frame 69	-4.72	4.40E-04
FGF9	fibroblast growth factor 9	-4.69	3.90E-03
PLCE1-AS1	PLCE1 antisense RNA 1	-4.5	5.30E-03
CORIN	corin, serine peptidase	-4.47	2.40E-11
IGF1	insulin like growth factor 1	-4.44	4.30E-03
LGR5	leucine rich repeat containing G protein-coupled receptor 5	-4.15	3.60E-03
COL1A1[9,10]	collagen type I alpha 1 chain	-2.75	1.40E-52
SULF1	sulfatase 1	-3.17	2.10E-43
WNT5A	Wnt family member 5A	-2.39	2.30E-43
GABARAPL1	GABA type A receptor associated protein like 1	-3.27	4.60E-38
NREP	neuronal regeneration related protein	-2.88	6.10E-38
COL5A1[9,10]	collagen type V alpha 1 chain	-2.48	2.60E-34
FSTL1	follistatin like 1	-1.98	2.80E-34
MFGE8[9,10]	milk fat globule-EGF factor 8 protein	-1.94	2.00E-32
SPARC	secreted protein acidic and cysteine rich	-2.29	2.30E-29
GDF6	growth differentiation factor 6	-3.86	9.10E-26

(b) Top 10 GO terms

	Gene name	Gene count	-log10 (FDR)
1	DNA replication	61	34.5
2	Cell division	73	22.3
3	G1/S transition of mitotic cell cycle	37	18.3
4	Mitotic nuclear division	53	16
5	DNA repair	51	15.7
6	Cell proliferation	40	3.3
7	Inflammatory response	38	2.4
8	Positive regulation of transcription, DNA-templated	43	1.4
9	Extracellular matrix organization	36	15.4
10	Cell adhesion	46	10.3

Table S8 Top DEGs from sT *vs* ctrl at 3 dpi.

(a) Top 10 upregulated/downregulated and most significant genes

Symbol	Gene name	log2FC	padj.
LCP1	lymphocyte cytosolic protein 1	4.37	5.30E-06
SEMA6B	semaphorin 6B	3.62	2.40E-02
ARHGDIB	Rho GDP dissociation inhibitor beta	3.22	3.30E-05
CHI3L1	chitinase 3 like 1	2.82	1.30E-02
SPTB	spectrin beta, erythrocytic	2.78	1.70E-04
SNORD4A	small nucleolar RNA, C/D box 4A	2.59	2.70E-02
KANK3	KN motif and ankyrin repeat domains 3	2.46	3.60E-02
SFXN2	sideroflexin 2	2.4	4.60E-07
OR7E12P	olfactory receptor family 7 subfamily E member 12 pseudogene	2.38	3.00E-02
C3AR1	complement C3a receptor 1	2.26	4.10E-02
TMEM97	transmembrane protein 97	1.36	2.10E-07
SFXN2	sideroflexin 2	2.4	4.60E-07
SCARB1	scavenger receptor class B member 1	1.75	4.90E-07
CAD	carbamoyl-phosphate synthetase 2, aspartate transcarbamylase, and dihydroorotase	1.25	8.90E-07
RRP1B[3]	ribosomal RNA processing 1B	1.5	3.20E-06
SNHG16	small nucleolar RNA host gene 16	1.27	3.50E-06
LCP1	lymphocyte cytosolic protein 1	4.37	5.30E-06
KIAA1217	KIAA1217	1.53	6.40E-06
RASL10B	RAS like family 10 member B	2.24	6.40E-06
GBP4	guanylate binding protein 4	-4.69	1.40E-03
IFIT1[9,10]	interferon induced protein with tetratricopeptide repeats 1	-3.84	2.00E-02
ISG15[9,10]	ISG15 ubiquitin like modifier	-3.81	1.80E-02
IFI6[9]	interferon alpha inducible protein 6	-3.67	3.30E-02
IFITM1[9,10]	interferon induced transmembrane protein 1	-3.53	3.00E-05
CHST1	carbohydrate sulfotransferase 1	-3.51	1.20E-02
THBD	thrombomodulin	-3.33	6.00E-04
GRIP2	glutamate receptor interacting protein 2	-3.31	3.70E-02
PSMB8-AS1	PSMB8 antisense RNA 1 (head to head)	-3.31	5.70E-03
SYNDIG1	synapse differentiation inducing 1	-3.27	7.80E-05
AHNAK2	AHNAK nucleoprotein 2	-1.81	5.50E-13
IFITM3[9,10]	interferon induced transmembrane protein 3	-1.67	7.50E-13
CREBRF	CREB3 regulatory factor	-2.17	1.70E-10
GPR1	G protein-coupled receptor 1	-2.26	1.50E-09
CRABP2	cellular retinoic acid binding protein 2	-2.34	3.10E-08
KIRREL3	kirre like nephrin family adhesion molecule 3	-1.73	3.10E-08
APOL6	apolipoprotein L6	-2.19	4.10E-08
STAT2[9,10]	signal transducer and activator of transcription 2	-2.41	1.00E-07
SYNE1	spectrin repeat containing nuclear envelope protein 1	-1.75	1.00E-07
MIR100HG	mir-100-let-7a-2-mir-125b-1 cluster host gene	-1.67	2.10E-07

(b) Top 10 GO terms

	Gene name	Gene count	-log10 (FDR)
1	Cell division	44	18.8
2	Mitotic nuclear division	35	16
3	rRNA processing	27	10.7
4	Sister chromatid cohesion	20	10.7
5	DNA repair	20	4.6
6	Cell proliferation	23	3.5
7	Viral process	16	1.4
8	Positive regulation of transcription from RNA polymerase II promoter	40	18.8
9	Type I interferon signaling pathway	19	11.8
10	Defense response to virus	24	9

Table S9 Top DEGs from sT *vs* ctrl at 8 dpi.

(a) Top 10 upregulated/downregulated and most significant genes

Symbol	Gene name	log2FC	padj.
CSF3	colony stimulating factor 3	7.19	3.90E-03
TIE1	tyrosine kinase with immunoglobulin like and EGF like domains 1	5.98	4.20E-03
TNFSF15	TNF superfamily member 15	5.83	4.20E-05
MMP13	matrix metallopeptidase 13	5.61	1.00E-04
LOC728485	uncharacterized LOC728485	5.61	5.70E-03
CXCL6[3]	C-X-C motif chemokine ligand 6	5.36	1.40E-09
MMP9	matrix metallopeptidase 9	5.24	2.60E-03
CCL7[3]	C-C motif chemokine ligand 7	5.19	2.70E-10
CHI3L1[3]	chitinase 3 like 1	5.12	9.30E-09
IL1B[3]	interleukin 1 beta	5.02	8.70E-04
LRRC17	leucine rich repeat containing 17	3.98	5.70E-20
MAP3K7CL	MAP3K7 C-terminal like	2.99	2.00E-18
KIAA1217	KIAA1217	2.21	7.00E-14
ACPP	acid phosphatase, prostate	4.66	1.20E-11
MEX3A	mex-3 RNA binding family member A	2.33	4.80E-11
PLAT	plasminogen activator, tissue type	2.43	2.00E-10
SLC16A1	solute carrier family 16 member 1	1.74	2.50E-10
CCL7[3]	C-C motif chemokine ligand 7	5.19	2.70E-10
GNL3L	G protein nucleolar 3 like	1.37	6.60E-10
ADRA2A	adrenoceptor alpha 2A	-9.08	1.10E-05
CMPK2	cytidine/uridine monophosphate kinase 2	-7.08	3.30E-03
SYNDIG1	synapse differentiation inducing 1	-7.08	2.20E-04
RSAD2[4,5,8]	radical S-adenosyl methionine domain containing 2	-7.02	5.30E-03
RGS7	regulator of G protein signaling 7	-6.97	9.70E-04
PTH1R	parathyroid hormone 1 receptor	-6.73	1.30E-05
HS3ST2	heparan sulfate-glucosamine 3-sulfotransferase 2	-6.36	3.00E-02
OAS1[4,5,8]	2′-5′-oligoadenylate synthetase 1	-6.34	1.20E-02
ANO3	anoctamin 3	-6.34	6.40E-06
BEX2	brain expressed X-linked 2	-6.23	2.60E-03
KIRREL3	kirre like nephrin family adhesion molecule 3	-2.5	7.00E-18
SAMD11	sterile alpha motif domain containing 11	-4.29	3.20E-17
IGFBP3	insulin like growth factor binding protein 3	-4.64	1.90E-15
PTN[10]	pleiotrophin	-2.38	6.50E-15
ATP10A	ATPase phospholipid transporting 10A (putative)	-2.07	1.10E-13
CREBRF[7]	CREB3 regulatory factor	-2.23	1.20E-12
MDK[10]	midkine	-2.75	2.00E-12
PPP1R13L[7]	protein phosphatase 1 regulatory subunit 13 like	-2.14	2.70E-12
IFITM3[4,5,8]	interferon induced transmembrane protein 3	-1.53	2.40E-11
SYNE1	spectrin repeat containing nuclear envelope protein 1	-2.03	4.80E-11

(b) Top 10 GO terms

	Gene name	Gene count	-log10 (FDR)
1	rRNA processing	27	6.2
2	Protein folding	18	2.2
3	Inflammatory response	27	2
4	Type I interferon signaling pathway	26	14.9
5	Defense response to virus	25	4.3
6	Cell adhesion	43	3.4
7	Negative regulation of transcription from RNA polymerase II promoter response	56	2.8
8	Response to virus	16	2
9	Angiogenesis	23	1.6
10	Nervous system development	26	1.4

Table S10 Top DEGs from sT+LT *vs* ctrl at 3 dpi.

(a) Top 10 upregulated/downregulated and most significant genes

Symbol	Gene name	log2FC	padj.
C6orf58	chromosome 6 open reading frame 58	8.9	1.80E-03
GBX2	gastrulation brain homeobox 2	8.72	2.80E-03
IL2RG[4]	interleukin 2 receptor subunit gamma	8.64	5.20E-03
SAA2	serum amyloid A2	8.23	1.10E-02
VCAM1	vascular cell adhesion molecule 1	7.62	2.80E-05
ST14	suppression of tumorigenicity 14	7.29	2.10E-02
CXCL9[1,3,4]	C-X-C motif chemokine ligand 9	7.13	8.80E-03
FAM107A	family with sequence similarity 107 member A	6.87	3.50E-03
TNFSF10[4]	TNF superfamily member 10	6.84	9.30E-07
PIK3AP1	phosphoinositide-3-kinase adaptor protein 1	6.84	1.70E-03
TRAF1	TNF receptor associated factor 1	4.04	9.40E-28
PLAU	plasminogen activator, urokinase	3.19	1.20E-24
SLC43A3[1,5,8,9]	solute carrier family 43 member 3	3.26	4.00E-20
NEDD4L	NEDD4 like E3 ubiquitin protein ligase	3.62	2.90E-18
CXCL1[3,4,8]	C-X-C motif chemokine ligand 1	3.46	8.40E-17
CXCL10[1,3,4,7]	C-X-C motif chemokine ligand 10	5.51	2.30E-14
CLDN1	claudin 1	3.69	2.60E-14
TNFAIP3[3,7]	TNF alpha induced protein 3	2.51	2.60E-14
OASL[1,5]	2′-5′-oligoadenylate synthetase like	2.68	3.60E-14
CCL2[3,4,7]	C-C motif chemokine ligand 2	2.53	4.00E-13
SLC18A2	solute carrier family 18 member A2	-7.06	3.10E-02
KCNN4	potassium calcium-activated channel subfamily N member 4	-7.01	3.00E-02
EXPH5	exophilin 5	-6.91	3.20E-02
PCTP	phosphatidylcholine transfer protein	-6.38	1.80E-02
RNASE4	ribonuclease A family member 4	-6.08	4.10E-02
SYNC	syncoilin, intermediate filament protein	-5.28	8.20E-10
COA5	cytochrome c oxidase assembly factor 5	-4.94	3.30E-02
PAPPA2	pappalysin 2	-4.88	4.40E-02
KLHL24	kelch like family member 24	-4.76	1.30E-02
CNPY4	canopy FGF signaling regulator 4	-4.74	2.10E-03
KRT7	keratin 7	-2.28	9.00E-13
LOXL4	lysyl oxidase like 4	-3.74	1.20E-12
S100A6	S100 calcium binding protein A6	-2.73	8.70E-12
NUPR1	nuclear protein 1, transcriptional regulator	-3.22	9.80E-12
FOS	Fos proto-oncogene, AP-1 transcription factor subunit	-2.07	4.70E-11
S100A4	S100 calcium binding protein A4	-2.2	3.50E-10
RAB3B	RAB3B, member RAS oncogene family	-1.7	8.20E-10
TPM2	tropomyosin 2	-1.7	4.40E-09
SEMA3C	semaphorin 3C	-2.3	4.80E-09

(b) Top 10 GO terms

	Gene name	Gene count	-log10 (FDR)
1	Defense response to virus	24	11.2
2	DNA replication	22	10.2
3	Inflammatory response	31	9.4
4	Immune response	32	9
5	Response to virus	17	8.1
6	G1/S transition of mitotic cell cycle	14	5.5
7	Cellular response to lipopolysaccharide	14	5.1
8	Negative regulation of cell proliferation	19	2.2
9	Positive regulation of gene expression	15	2.1
10	DNA repair	14	2.1

Table S11 Top DEGs from sT+LT*tr vs* ctrl at 9/12 dpi.

(a) Top 10 upregulated/downregulated and most significant genes

Symbol	Gene name	log2FC	padj.
CCL11[1,4,5,6,7,9,10]	C-C motif chemokine ligand 11	10.05	2.90E-02
CCL8[1,4,5,6,7,9,10]	C-C motif chemokine ligand 8	10.02	2.20E-07
CSF3[9]	colony stimulating factor 3	9.91	3.90E-12
CCL7[4,5,7,8,10]	C-C motif chemokine ligand 7	8.39	6.90E-13
CXCL6[4,9,6,7,8]	C-X-C motif chemokine ligand 6	7.51	1.10E-09
CHI3L1[4,10]	chitinase 3 like 1	7.38	8.00E-24
IL1B[4,5,9]	interleukin 1 beta	7.05	1.80E-06
TMPRSS2	transmembrane serine protease 2	6.96	2.90E-02
CD70[9]	CD70 molecule	6.82	2.10E-02
CXCL1[4,9,6,7,8]	C-X-C motif chemokine ligand 1	6.79	6.90E-65
C3[4,9]	complement C3	4.83	1.70E-34
POU2F2	POU class 2 homeobox 2	3.21	5.60E-24
ZC3H12A	zinc finger CCCH-type containing 12A	3.54	2.30E-16
C1QTNF1	C1q and TNF related 1	2.77	2.80E-16
ICAM1[10]	intercellular adhesion molecule 1	2.85	2.40E-14
CSF2[10]	colony stimulating factor 2	6.78	6.90E-14
SOD2	superoxide dismutase 2	3.02	2.80E-12
ADRA1B	adrenoceptor alpha 1B	-7.91	2.50E-04
CDH18	cadherin 18	-7.52	3.50E-03
NKILA	NF-kappaB interacting lncRNA	-6.37	2.20E-02
INSC	INSC spindle orientation adaptor protein	-6.29	2.30E-02
ALX4	ALX homeobox 4	-6.21	3.20E-02
PCDH10	protocadherin 10	-6.13	4.20E-04
COMP	cartilage oligomeric matrix protein	-6.09	5.60E-03
SAMD11	sterile alpha motif domain containing 11	-5.91	1.90E-02
AGT	angiotensinogen	-5.88	4.70E-02
PPIL6	peptidylprolyl isomerase like 6	-5.75	3.10E-02
SCN9A	sodium voltage-gated channel alpha subunit 9	-3.17	1.20E-13
IGFBP3	insulin like growth factor binding protein 3	-4.82	5.20E-13
GFRA1	GDNF family receptor alpha 1	-3.11	1.50E-11
MFGE8	milk fat globule-EGF factor 8 protein	-2.48	3.10E-10
EPHA4	EPH receptor A4	-2.81	5.10E-10
MFAP3L	microfibril associated protein 3 like	-3.9	7.80E-10
ALDH1A1	aldehyde dehydrogenase 1 family member A1	-4.63	1.10E-09
SLIT3	slit guidance ligand 3	-3.27	1.40E-08
ADAMTS5	ADAM metallopeptidase with thrombospondin type 1 motif 5	-2.45	6.80E-08
AFF3	AF4/FMR2 family member 3	-2.9	3.50E-07

(b) Top 10 GO terms

	Gene name	Gene count	-log10 (FDR)
1	G1/S transition of mitotic cell cycle	19	10.5
2	DNA replication initiation	12	9.1
3	DNA replication	20	8.7
4	Inflammatory response	26	6.3
5	Neutrophil chemotaxis	11	4.7
6	Cell chemotaxis	10	3.8
7	Chemokine-mediated signaling pathway	10	3.5
8	Chemotaxis	12	3.3
9	Immune response	21	2.8
10	Positive regulation of ERK1 and ERK2	11	1.6

Table S12 Description of genes involved in inflammatory and type I IFN signaling

Symbol	Gene name
APOBEC3F	apolipoprotein B mRNA editing enzyme catalytic subunit 3F
APOBEC3G	apolipoprotein B mRNA editing enzyme catalytic subunit 3G
B2M	beta-2-microglobulin
C3	complement C3
CCL11	C-C motif chemokine ligand 11
CCL2	C-C motif chemokine ligand 2
CCL5	C-C motif chemokine ligand 5
CCL7	C-C motif chemokine ligand 7
CCL8	C-C motif chemokine ligand 8
CSF2	colony stimulating factor 2
CSF3	colony stimulating factor 3
CXCL1	C-X-C motif chemokine ligand 1
CXCL10	C-X-C motif chemokine ligand 10
CXCL11	C-X-C motif chemokine ligand 11
CXCL6	C-X-C motif chemokine ligand 6
CXCL9	C-X-C motif chemokine ligand 9
DDX60	DExD/H-box helicase 60
GBP3	guanylate binding protein 3
GBP5	guanylate binding protein 1 pseudogene 1
HCP5	HLA complex P5
HERC5	HECT and RLD domain containing E3 ubiquitin protein ligase 5
HLA-A	major histocompatibility complex, class I, A
HLA-B	major histocompatibility complex, class I, B
HLA-C	major histocompatibility complex, class I, C
HLA-E	major histocompatibility complex, class I, E
HLA-F	major histocompatibility complex, class I, F
HLA-G	major histocompatibility complex, class I, G
HLA-H	major histocompatibility complex, class I, H
IFI27	interferon alpha inducible protein 27
IFI35	interferon induced protein 35
IFI44	interferon induced protein 44
IFI6	interferon alpha inducible protein 6
IFIH1	interferon induced with helicase C domain 1
IFIT1	interferon induced protein with tetratricopeptide repeats 1
IFIT2	interferon induced protein with tetratricopeptide repeats 2
IFIT3	interferon induced protein with tetratricopeptide repeats 3
IFIT5	interferon induced protein with tetratricopeptide repeats 5
IFITM1	interferon induced transmembrane protein 1
IFITM10	interferon induced transmembrane protein 10
IFITM2	interferon induced transmembrane protein 2
IFITM3	interferon induced transmembrane protein 3
IFNA1	interferon alpha 1
IFNB1	interferon beta 1
IL1B	interleukin 1 beta

IL2RG	interleukin 2 receptor subunit gamma
IRF7	interferon regulatory factor 7
IRF9	interferon regulatory factor 9
ISG15	ISG15 ubiquitin like modifier
MMP1	matrix metallopeptidase 1
MX1	MX dynamin like GTPase 1
MX2	MX dynamin like GTPase 2
OAS1	2′-5′-oligoadenylate synthetase 1
OAS2	2′-5′-oligoadenylate synthetase 2
OAS3	2′-5′-oligoadenylate synthetase 3
OASL	2′-5′-oligoadenylate synthetase like
RSAD2	radical S-adenosyl methionine domain containing 2
STAT1	signal transducer and activator of transcription 1
STAT2	signal transducer and activator of transcription 2
TAP1	transporter 1, ATP binding cassette subfamily B member
TLR3	toll like receptor 3
TMEM173 / STING1	transmembrane protein 173 / stimulator of interferon response cGAMP interactor 1
TNF	tumor necrosis factor
TNFAIP3	TNF alpha induced protein 3
TNFSF10	TNF superfamily member 10
TRIM22	tripartite motif containing 22
TRIM56	tripartite motif containing 56

References

Abdul-Sada, H., Müller, M., Mehta, R., Toth, R., Arthur, J. S. C., Whitehouse, A., & Macdonald, A. (2017). The pp4r1 sub-unit of protein phosphatase pp4 is essential for inhibition of nf-κb by merkel polyomavirus small tumour antigen. *Oncotarget*, *8*, 25418–25432.

Afanasiev, O. K., Yelistratova, L., Miller, N., Nagase, K., Paulson, K., Iyer, J. G., Ibrani, D., Koelle, D. M., & Nghiem, P. (2013). Merkel polyomavirus-specific t cells fluctuate with merkel cell carcinoma burden and express therapeutically targetable pd-1 and tim-3 exhaustion markers. *Clinical cancer research : an official journal of the American Association for Cancer Research*, *19*, 5351–60.

Akhbari, P., Tobin, D., Poterlowicz, K., Roberts, W., & Boyne, J. R. (2018). Mcv-mir-m1 targets the host-cell immune response resulting in the attenuation of neutrophil chemotaxis. *Journal of Investigative Dermatology*, *138*(11), 2343–2354.

Allen, P. J., Bowne, W. B., Jaques, D. P., Brennan, M. F., Busam, K., & Coit, D. G. (2005). Merkel cell carcinoma: Prognosis and treatment of patients from a single institution. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, *23*, 2300–9.

An, P., Sáenz Robles, M. T., Duray, A. M., Cantalupo, P. G., & Pipas, J. M. (2019). Human polyomavirus bkv infection of endothelial cells results in interferon pathway induction and persistence. *PLoS pathogens*, *15*, e1007505.

Andrews, S. (2010). Fastqc. a quality control tool for high throughput sequence data.

Arifuzzaman, S., Khatun, M. R., & Khatun, R. (2020). Emerging of lysine demethylases (kdms): From pathophysiological insights to novel therapeutic opportunities. *Biomedicine pharmacotherapy = Biomedecine pharmacotherapie*, *129*, 110392.

Arthur, R. R., Shah, K. V., Baust, S. J., Santos, G. W., & Saral, R. (1986). Association of bk viruria with hemorrhagic cystitis in recipients of bone marrow transplants. *N Engl J Med*, *315*(4), 230–234.

Assetta, B., De Cecco, M., O'Hara, B., & Atwood, W. J. (2016). Jc polyomavirus infection of primary human renal epithelial cells is controlled by a type i ifn-induced response. *mBio*, *7*.

Assetta, B., Maginnis, M. S., Gracia Ahufinger, I., Haley, S. A., Gee, G. V., Nelson, C. D. S., O'Hara, B. A., Allen Ramdial, S.-a. A., & Atwood, W. J. (2013). 5-ht2 receptors facilitate jc polyomavirus entry. *Journal of virology*, *87*, 13490–8.

Au-Yeung, N., & Horvath, C. M. (2018). Transcriptional and chromatin regulation in interferon and innate antiviral gene expression (2018/10/22). *Cytokine growth factor reviews*, *44*(30509403), 11–17.

Ballestas, M. E., Chatis, P. A., & Kaye, K. M. (1999). Efficient persistence of extrachromosomal kshv dna mediated by latency-associated nuclear antigen. *Science (New York, N.Y.)*, *284*, 641–4.

Barber, G. N. (2015). Sting: Infection, inflammation and cancer. *Nature reviews. Immunology*, *15*, 760–70.

Becker, J. C., Stang, A., DeCaprio, J. A., Cerroni, L., Lebbé, C., Veness, M., & Nghiem, P. (2017). Merkel cell carcinoma. *Nature reviews. Disease primers*, *3*, 17077.

Becker, J. C., Stang, A., Hausen, A. Z., Fischer, N., DeCaprio, J. A., Tothill, R. W., Lyngaa, R., Hansen, U. K., Ritter, C., Nghiem, P., Bichakjian, C. K., Ugurel, S., & Schrama, D. (2018). Epidemiology, biology and therapy of merkel cell carcinoma: Conclusions from the eu project immomec. *Cancer immunology, immunotherapy : CII*, *67*, 341–351.

Becker, M., Dominguez, M., Greune, L., Soria-Martinez, L., Pfleiderer, M. M., Schowalter, R., Buck, C. B., Blaum, B. S., Schmidt, M. A., & Schelhaas, M. (2019). Infectious entry of merkel cell polyomavirus. *Journal of virology*, *93*.

Bennett, S. M., Zhao, L., Bosard, C., & Imperiale, M. J. (2015). Role of a nuclear localization signal on the minor capsid proteins vp2 and vp3 in bkpyv nuclear entry. *Virology*, *474*, 110–116.

Bernstein, B. E., Mikkelsen, T. S., Xie, X., Kamal, M., Huebert, D. J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., Jaenisch, R., Wagschal, A., Feil, R., Schreiber, S. L., & Lander, E. S. (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell*, *125*, 315–26.

Berrios, C., Padi, M., Keibler, M. A., Park, D. E., Molla, V., Cheng, J., Lee, S. M., Stephanopoulos, G., Quackenbush, J., & DeCaprio, J. A. (2016). Merkel cell polyomavirus small t antigen promotes pro-glycolytic metabolic perturbations required for transformation. *PLoS pathogens*, *12*(27880818), e1006020–e1006020.

Borchert, S., Czech-Sioli, M., Neumann, F., Schmidt, C., Wimmer, P., Dobner, T., Grundhoff, A., & Fischer, N. (2014). High-affinity rb binding, p53 inhibition, subcellular localization, and transformation by wild-type or tumor-derived shortened merkel cell polyomavirus large t antigens. *Journal of virology*, *88*, 3144–60.

Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical biochemistry*, *72*, 248–54.

Brind'Amour, J., Liu, S., Hudson, M., Chen, C., Karimi, M. M., & Lorincz, M. C. (2015). An ultra-low-input native chip-seq protocol for genome-wide profiling of rare cell populations. *Nature communications*, *6*, 6033.

BroadInstitute. (2018). Picard. *http://broadinstitute.github.io/picard/*.

Busam, K. J., Pulitzer, M. P., Coit, D. C., Arcila, M., Leng, D., Jungbluth, A. A., & Wiesner, T. (2017). Reduced h3k27me3 expression in merkel cell polyoma virus-positive tumors. *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc*, *30*, 877–883.

Carter, J. J., Daugherty, M. D., Qi, X., Bheda-Malge, A., Wipf, G. C., Robinson, K., Roman, A., Malik, H. S., & Galloway, D. A. (2013). Identification of an overprinting gene in merkel cell polyomavirus provides evolutionary insight into the birth of viral genes. *Proc Natl Acad Sci USA*, *110*(31), 12744.

Carter, J. J., Paulson, K. G., Wipf, G. C., Miranda, D., Madeleine, M. M., Johnson, L. G., Lemos, B. D., Lee, S., Warcola, A. H., Iyer, J. G., Nghiem, P., & Galloway, D. A. (2009). Association of merkel cell polyomavirus-specific antibodies with merkel cell carcinoma. *Journal of the National Cancer Institute*, *101*, 1510–22.

Chang, Y., & Moore, P. S. (2012). Merkel cell carcinoma: A virus-induced human cancer. *Annual review of pathology*, *7*, 123–44.

Cheng, J., DeCaprio, J. A., Fluck, M. M., & Schaffhausen, B. S. (2009). Cellular transformation by simian virus 40 and murine polyoma virus t antigens. *Seminars in cancer biology*, *19*, 218–28.

Cheng, J., Park, D. E., Berrios, C., White, E. A., Arora, R., Yoon, R., Branigan, T., Xiao, T., Westerling, T., Federation, A., Zeid, R., Strober, B., Swanson, S. K., Florens, L., Bradner, J. E., Brown, M., Howley, P. M., Padi, M., Washburn, M. P., & DeCaprio, J. A. (2017). Merkel cell polyomavirus recruits mycl to the ep400 complex to promote oncogenesis. *PLoS pathogens*, *13*, e1006668.

Cheng, J., Rozenblatt-Rosen, O., Paulson, K. G., Nghiem, P., & DeCaprio, J. A. (2013). Merkel cell polyomavirus large t antigen has growth-promoting and inhibitory activities. *Journal of virology*, *87*, 6118–26.

Chiang, C., Dvorkin, S., Chiang, J. J., Potter, R. B., & Gack, M. U. (2021). The small t antigen of jc virus antagonizes rig-i-mediated innate immunity by inhibiting trim25's rna binding ability. *mBio*, *12*.

Clayson, E. T., Brando, L. V., & Compans, R. W. (1989). Release of simian virus 40 virions from epithelial cells is polarized and occurs without cell lysis. *Journal of virology*, *63*, 2278–88.

Cloos, P. A. C., Christensen, J., Agger, K., & Helin, K. (2008). Erasing the methyl mark: Histone demethylases at the center of cellular differentiation and disease. *Genes development*, *22*(18451103), 1115–1140.

Courtneidge, S. A., & Smith, A. E. (1983). Polyoma virus transforming protein associates with the product of the c-src cellular gene. *Nature*, *303*(5916), 435–439.

Cowey, C. L., Mahnke, L., Espirito, J., Helwig, C., Oksen, D., & Bharmal, M. (2017). Real-world treatment outcomes in patients with metastatic merkel cell carcinoma treated with chemotherapy in the usa. *Future oncology (London, England)*, *13*, 1699–1710.

Creyghton, M. P., Cheng, A. W., Welstead, G. G., Kooistra, T., Carey, B. W., Steine, E. J., Hanna, J., Lodato, M. A., Frampton, G. M., Sharp, P. A., Boyer, L. A., Young, R. A., & Jaenisch, R. (2010). Histone h3k27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National Academy of Sciences of the United States of America*, *107*, 21931–6.

Cui, K., Tailor, P., Liu, H., Chen, X., Ozato, K., & Zhao, K. (2004). The chromatin-remodeling baf complex mediates cellular antiviral activities by promoter priming. *Molecular and cellular biology*, *24*, 4476–86.

Czech-Sioli, M., Günther, T., Therre, M., Spohn, M., Indenbirken, D., Theiss, J., Riethdorf, S., Qi, M., Alawi, M., Wülbeck, C., Fernandez-Cuesta, I., Esmek, F., Becker, J. C., Grundhoff, A., & Fischer, N. (2020a). High-resolution analysis of merkel cell polyomavirus in merkel cell carcinoma reveals distinct integration patterns and suggests nhej and mmbir as underlying mechanisms. *PLoS pathogens*, *16*, e1008562.

Czech-Sioli, M., Siebels, S., Radau, S., Zahedi, R. P., Schmidt, C., Dobner, T., Grundhoff, A., & Fischer, N. (2020b). The ubiquitin-specific protease usp7, a novel merkel cell polyomavirus large t-antigen interaction partner, modulates viral dna replication. *Journal of virology*, *94*.

Daniels, R., Sadowicz, D., & Hebert, D. N. (2007). A very late viral protein triggers the lytic release of sv40. *PLoS pathogens*, *3*, e98.

DeCaprio, J. A., & Garcea, R. L. (2013). A cornucopia of human polyomaviruses. *Nature Reviews Microbiology*, *11*(4), 264–276.

Dela Cruz, F. N., Jr, Giannitti, F., Li, L., Woods, L. W., Del Valle, L., Delwart, E., & Pesavento, P. A. (2013). Novel polyomavirus associated with brain tumors in free-ranging raccoons, western united states. *Emerging infectious diseases*, *19*(23260029), 77–84.

Dimova, D. K., & Dyson, N. J. (2005). The e2f transcriptional network: Old acquaintances with new faces. *Oncogene*, *24*, 2810–26.

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). Star: Ultrafast universal rna-seq aligner. *Bioinformatics (Oxford, England)*, *29*, 15–21.

Dobson, S. J., Anene, A., Boyne, J. R., Mankouri, J., Macdonald, A., & Whitehouse, A. (2020). Merkel cell polyomavirus small tumour antigen activates the p38 mapk pathway to enhance cellular motility. *The Biochemical journal*, *477*, 2721–2733.

Dupzyk, A., & Tsai, B. (2016). How polyomaviruses exploit the erad machinery to cause infection. *Viruses*, *8*(27589785), 242.

Eisemann, N., Jansen, L., Castro, F. A., Chen, T., Eberle, A., Nennecke, A., Zeissig, S. R., Brenner, H., & Katalinic, A. (2016). Survival with nonmelanoma skin cancer in germany. *The British journal of dermatology*, *174*, 778–85.

Elphick, G. F., Querbes, W., Jordan, J. A., Gee, G. V., Eash, S., Manley, K., Dugan, A., Stanifer, M., Bhatnagar, A., Kroeze, W. K., Roth, B. L., & Atwood, W. J. (2004). The human polyomavirus, jcv, uses serotonin receptors to infect cells. *Science (New York, N.Y.)*, *306*, 1380–3.

Evans, G. L., Caller, L. G., Foster, V., & Crump, C. M. (2015). Anion homeostasis is important for non-lytic release of bk polyomavirus from infected cells. *Open biology*, *5*.

Feng, H., Kwun, H. J., Liu, X., Gjoerup, O., Stolz, D. B., Chang, Y., & Moore, P. S. (2011). Cellular and viral factors regulating merkel cell polyomavirus replication. *PloS one*, *6*, e22468.

Feng, H., Shuda, M., Chang, Y., & Moore, P. S. (2008). Clonal integration of a polyomavirus in human merkel cell carcinoma. *Science (New York, N.Y.)*, *319*, 1096–100.

Fischer, N. (2020). Infection-induced epigenetic changes and their impact on the pathogenesis of diseases. *Seminars in immunopathology*, *42*(32318841), 127–130.

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. *International journal of cancer*, *126*, 2133–42.

Forero, A., Giacobbi, N. S., McCormick, K. D., Gjoerup, O. V., Bakkenist, C. J., Pipas, J. M., & Sarkar, S. N. (2014). Simian virus 40 large t antigen induces ifn-stimulated genes through atr kinase. *Journal of immunology (Baltimore, Md. : 1950)*, *192*, 5933–42.

Freaney, J. E., Kim, R., Mandhana, R., & Horvath, C. M. (2013). Extensive cooperation of immune master regulators irf3 and nfκb in rna pol ii recruitment and pause release in human innate antiviral transcription. *Cell reports*, *4*, 959–73.

Gardner, S. D., Field, A. M., Coleman, D. V., & Hulme, B. (1971). New human papovavirus (b.k.) isolated from urine after renal transplantation. *Lancet (London, England)*, *1*, 1253–7.

Garneski, K. M., Warcola, A. H., Feng, Q., Kiviat, N. B., Leonard, J. H., & Nghiem, P. (2009). Merkel cell polyomavirus is more frequently present in north american than australian merkel cell carcinoma tumors (2008/07/24). *The Journal of investigative dermatology*, *129*(18650846), 246–248.

Gerits, N., & Moens, U. (2012). Agnoprotein of mammalian polyomaviruses. *Virology*, *432*(2), 316–326.

Giacinti, C., & Giordano, A. (2006). Rb and cell cycle progression. *Oncogene*, *25*, 5220–7.

Giacobbi, N. S., Gupta, T., Coxon, A. T., & Pipas, J. M. (2015). Polyomavirus t antigens activate an antiviral state (2015/01/10). *Virology*, *476*(25589241), 377–385.

Goh, G., Walradt, T., Markarov, V., Blom, A., Riaz, N., Doumani, R., Stafstrom, K., Moshiri, A., Yelistratova, L., Levinsohn, J., Chan, T. A., Nghiem, P., Lifton, R. P., & Choi, J. (2016). Mutational landscape of mcpyv-positive and mcpyv-negative merkel cell carcinomas with implications for immunotherapy. *Oncotarget*, *7*, 3403–15.

Goh, S., Lindau, C., Tiveljung-Lindell, A., & Allander, T. (2009). Merkel cell polyomavirus in respiratory tract secretions. *Emerging infectious diseases*, *15*, 489–91.

Gosert, R., Kardas, P., Major, E. O., & Hirsch, H. H. (2010). Rearranged jc virus noncoding control regions found in progressive multifocal leukoencephalopathy patient samples increase virus early gene expression and replication rate. *Journal of virology*, *84*, 10448–56.

Gosert, R., Rinaldo, C. H., Funk, G. A., Egli, A., Ramos, E., Drachenberg, C. B., & Hirsch, H. H. (2008). Polyomavirus bk with rearranged noncoding control region emerge in vivo in renal transplant patients and increase viral replication and cytopathology. *The Journal of experimental medicine*, *205*, 841–52.

Grandhaye, M., Teixeira, P. G., Henrot, P., Morel, O., Sirveaux, F., Verhaeghe, J.-L., & Blum, A. (2015). Focus on merkel cell carcinoma: Diagnosis and staging. *Skeletal radiology*, *44*, 777–86.

Gravemeyer, J., Spassova, I., Verhaegen, M. E., Dlugosz, A. A., Hoffmann, D., Lange, A., & Becker, J. C. (2021). Dna-methylation patterns imply a common cellular origin of virus- and uv-associated merkel cell carcinoma. *Oncogene*.

Greenberg, M. V. C., & Bourc'his, D. (2019). The diverse roles of dna methylation in mammalian development and disease. *Nature reviews. Molecular cell biology*, *20*, 590–607.

Greenlee, J. E., & Hirsch, H. H. (2016). Polyomaviruses. *Clinical Virology, John Wiley Sons, Ltd, page 599-623*.

Griffiths, D. A., Abdul-Sada, H., Knight, L. M., Jackson, B. R., Richards, K., Prescott, E. L., Peach, A. H. S., Blair, G. E., Macdonald, A., & Whitehouse, A. (2013). Merkel cell polyomavirus small t antigen targets the nemo adaptor protein to disrupt inflammatory signaling. *Journal of virology*, *87*, 13853–67.

Gross, L. (1953). A filterable agent, recovered from ak leukemic extracts, causing salivary gland carcinomas in c3h mice. *Proceedings of the Society for Experimental Biology and Medicine. Society for Experimental Biology and Medicine (New York, N.Y.)*, *83*, 414–21.

Grundhoff, A., & Fischer, N. (2015). Merkel cell polyomavirus, a highly prevalent virus with tumorigenic potential. *Current opinion in virology*, *14*, 129–37.

Guastafierro, A., Feng, H., Thant, M., Kirkwood, J. M., Chang, Y., Moore, P. S., & Shuda, M. (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. *Journal of virological methods*, *187*, 6–14.

Hahn, W. C., Dessain, S. K., Brooks, M. W., King, J. E., Elenbaas, B., Sabatini, D. M., DeCaprio, J. A., & Weinberg, R. A. (2002). Enumeration of the simian virus 40 early region elements necessary for human cell transformation. *Molecular and cellular biology*, *22*, 2111–23.

Hanahan, D., & Weinberg, R. A. (2000). The hallmarks of cancer. *Cell*, *100*, 57–70.

Handala, L., Blanchard, E., Raynal, P.-I., Roingeard, P., Morel, V., Descamps, V., Castelain, S., Francois, C., Duverlie, G., Brochot, E., & Helle, F. (2020). Bk polyomavirus hijacks extracellular vesicles for en bloc transmission. *Journal of virology*, *94*.

Harms, K. L., Chubb, H., Zhao, L., Fullen, D. R., Bichakjian, C. K., Johnson, T. M., Carskadon, S., Palanisamy, N., & Harms, P. W. (2017). Increased expression of ezh2 in merkel cell carcinoma is associated with disease progression and poorer prognosis. *Human pathology*, *67*, 78–84.

Harms, K. L., Healy, M. A., Nghiem, P., Sober, A. J., Johnson, T. M., Bichakjian, C. K., & Wong, S. L. (2016). Analysis of prognostic factors from 9387 merkel cell carcinoma cases forms the basis for the new 8th edition ajcc staging system. *Annals of surgical oncology*, *23*, 3564–3571.

Harms, P. W., Harms, K. L., Moore, P. S., DeCaprio, J. A., Nghiem, P., Wong, M. K. K., & Brownell, I. (2018). The biology and treatment of merkel cell carcinoma: Current understanding and research priorities. *Nature reviews. Clinical oncology*, *15*, 763–776.

Harrison, C. J., Meinke, G., Kwun, H. J., Rogalin, H., Phelan, P. J., Bullock, P. A., Chang, Y., Moore, P. S., & Bohm, A. (2011). Asymmetric assembly of merkel cell polyomavirus large t-antigen origin binding domains at the viral origin (2011/04/09). *Journal of molecular biology*, *409*(21501625), 529–542.

Heath, M., Jaimes, N., Lemos, B., Mostaghimi, A., Wang, L. C., Peñas, P. F., & Nghiem, P. (2008). Clinical characteristics of merkel cell carcinoma at diagnosis in 195 patients: The aeiou features. *Journal of the American Academy of Dermatology*, *58*, 375–81.

Helle, F., Handala, L., Bentz, M., Duverlie, G., & Brochot, E. (2020). Intercellular transmission of naked viruses through extracellular vesicles: Focus on polyomaviruses. *Viruses*, *12*.

Hesbacher, S., Pfitzer, L., Wiedorfer, K., Angermeyer, S., Borst, A., Haferkamp, S., Scholz, C.-J., Wobser, M., Schrama, D., & Houben, R. (2016). Rb1 is the crucial target of the merkel cell polyomavirus large t antigen in merkel cell carcinoma cells. *Oncotarget*, *7*, 32956–68.

Ho, J., Jedrych, J. J., Feng, H., Natalie, A. A., Grandinetti, L., Mirvish, E., Crespo, M. M., Yadav, D., Fasanella, K. E., Proksell, S., Kuan, S.-F., Pastrana, D. V., Buck, C. B., Shuda, Y., Moore, P. S., & Chang, Y. (2015). Human polyomavirus 7-associated pruritic rash and viremia in transplant recipients (2014/09/17). *The Journal of infectious diseases*, *211*(25231015), 1560–1565.

Houben, R., Adam, C., Baeurle, A., Hesbacher, S., Grimm, J., Angermeyer, S., Henzel, K., Hauser, S., Elling, R., Bröcker, E.-B., Gaubatz, S., Becker, J. C., & Schrama, D. (2012). An intact retinoblastoma protein-binding site in merkel cell polyomavirus large t antigen is required for promoting growth of merkel cell carcinoma cells. *International journal of cancer*, *130*, 847–56.

Houben, R., Angermeyer, S., Haferkamp, S., Aue, A., Goebeler, M., Schrama, D., & Hesbacher, S. (2015). Characterization of functional domains in the merkel cell polyoma virus large t antigen. *International journal of cancer*, *136*, E290–300.

Houben, R., Shuda, M., Weinkam, R., Schrama, D., Feng, H., Chang, Y., Moore, P. S., & Becker, J. C. (2010). Merkel cell polyomavirus-infected merkel cell carcinoma cells require expression of viral t antigens. *Journal of virology*, *84*, 7064–72.

Howe, F. S., Fischl, H., Murray, S. C., & Mellor, J. (2017). Is h3k4me3 instructive for transcription activation? *BioEssays : news and reviews in molecular, cellular and developmental biology*, *39*, 1–12.

Howley, P. M., & Livingston, D. M. (2009). Small dna tumor viruses: Large contributors to biomedical sciences (2009/01/10). *Virology*, *384*(19136134), 256–259.

Huang, D. W., Sherman, B. T., & Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using david bioinformatics resources. *Nature protocols*, *4*, 44–57.

Imperiale, M. J., & Jiang, M. (2016). Polyomavirus persistence. *Annual review of virology*, *3*, 517–532.

Isaacs, A., & Lindemann, J. (1957). Virus interference. i. the interferon. *Proceedings of the Royal Society of London. Series B, Biological sciences*, *147*, 258–67.

Iwanaszko, M., & Kimmel, M. (2015). Nf-κb and irf pathways: Cross-regulation on target genes promoter level. *BMC genomics*, *16*(25888367), 307–307.

Iwasaki, A. (2012). A virological view of innate immune recognition. *Annual review of microbiology*, *66*, 177–96.

Jay, G., Nomura, S., Anderson, C. W., & Khoury, G. (1981). Identification of the sv40 agnogene product: A dna binding protein. *Nature*, *291*, 346–9.

Jones, P. A. (2012). Functions of dna methylation: Islands, start sites, gene bodies and beyond. *Nature reviews. Genetics*, *13*, 484–92.

Karmodiya, K., Krebs, A. R., Oulad-Abdelghani, M., Kimura, H., & Tora, L. (2012). H3k9 and h3k14 acetylation co-occur at many gene regulatory elements, while h3k14ac marks a subset of inactive inducible promoters in mouse embryonic stem cells. *BMC genomics*, *13*, 424.

Kim, S. H., Cohen, B., Novick, D., & Rubinstein, M. (1997). Mammalian type i interferon receptors consists of two subunits: Ifnar1 and ifnar2. *Gene*, *196*, 279–86.

Knepper, T. C., Montesion, M., Russell, J. S., Sokol, E. S., Frampton, G. M., Miller, V. A., Albacker, L. A., McLeod, H. L., Eroglu, Z., Khushalani, N. I., Sondak, V. K., Messina, J. L., Schell, M. J., DeCaprio, J. A., Tsai, K. Y., & Brohl, A. S. (2019). The genomic landscape of merkel cell carcinoma and clinicogenomic biomarkers of response to immune checkpoint

inhibitor therapy. *Clinical cancer research : an official journal of the American Association for Cancer Research*, *25*, 5961–5971.

Knight, L. M., Stakaityte, G., Wood, J. J., Abdul-Sada, H., Griffiths, D. A., Howell, G. J., Wheat, R., Blair, G. E., Steven, N. M., Macdonald, A., Blackbourn, D. J., & Whitehouse, A. (2015). Merkel cell polyomavirus small t antigen mediates microtubule destabilization to promote cell motility and migration. *Journal of virology*, *89*, 35–47.

Knips, J., Czech-Sioli, M., Spohn, M., Heiland, M., Moll, I., Grundhoff, A., Schumacher, U., & Fischer, N. (2017). Spontaneous lung metastasis formation of human merkel cell carcinoma cell lines transplanted into scid mice. *International journal of cancer*, *141*, 160–171.

Krump, N. A., Wang, R., Liu, W., Yang, J. F., Ma, T., & You, J. (2021). Merkel cell polyomavirus infection induces an antiviral innate immune response in human dermal fibroblasts. *Journal of virology*, *95*, e0221120.

Kwun, H. J., Chang, Y., & Moore, P. S. (2017). Protein-mediated viral latency is a novel mechanism for merkel cell polyomavirus persistence. *Proceedings of the National Academy of Sciences of the United States of America*, *114*, E4040–E4047.

Kwun, H. J., Guastafierro, A., Shuda, M., Meinke, G., Bohm, A., Moore, P. S., & Chang, Y. (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. *Journal of virology*, *83*, 12118–28.

Kwun, H. J., Shuda, M., Camacho, C. J., Gamper, A. M., Thant, M., Chang, Y., & Moore, P. S. (2015). Restricted protein phosphatase 2a targeting by merkel cell polyomavirus small t antigen. *Journal of virology*, *89*, 4191–200.

Kwun, H. J., Shuda, M., Feng, H., Camacho, C. J., Moore, P. S., & Chang, Y. (2013). Merkel cell polyomavirus small t antigen controls viral replication and oncoprotein expression by targeting the cellular ubiquitin ligase scffbw7. *Cell host microbe*, *14*(23954152), 125–135.

Lawrence, M., Daujat, S., & Schneider, R. (2016). Lateral thinking: How histone modifications regulate gene expression. *Trends in genetics : TIG*, *32*, 42–56.

Lee, A. J., & Ashkar, A. A. (2018). The dual nature of type i and type ii interferons. *Frontiers in immunology*, *9*, 2061.

Lerdrup, M., Johansen, J. V., Agrawal-Singh, S., & Hansen, K. (2016). An interactive environment for agile analysis and visualization of chip-sequencing data. *Nature structural molecular biology*, *23*, 349–57.

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics (Oxford, England)*, *25*, 1754–60.

Li, J., Diaz, J., Wang, X., Tsang, S. H., & You, J. (2015). Phosphorylation of merkel cell polyomavirus large tumor antigen at serine 816 by atm kinase induces apoptosis in host cells (2014/12/05). *The Journal of biological chemistry*, *290*(25480786), 1874–1884.

Li, J., Wang, X., Diaz, J., Tsang, S. H., Buck, C. B., & You, J. (2013). Merkel cell polyomavirus large t antigen disrupts host genomic integrity and inhibits cellular proliferation. *Journal of virology*, *87*, 9173–88.

Lipson, E. J., Vincent, J. G., Loyo, M., Kagohara, L. T., Luber, B. S., Wang, H., Xu, H., Nayar, S. K., Wang, T. S., Sidransky, D., Anders, R. A., Topalian, S. L., & Taube, J. M. (2013). Pd-l1 expression in the merkel cell carcinoma microenvironment: Association with inflammation, merkel cell polyomavirus and overall survival. *Cancer immunology research*, *1*, 54–63.

Litvak, V., Ratushny, A. V., Lampano, A. E., Schmitz, F., Huang, A. C., Raman, A., Rust, A. G., Bergthaler, A., Aitchison, J. D., & Aderem, A. (2012). A foxo3-irf7 gene regulatory circuit limits inflammatory sequelae of antiviral responses. *Nature*, *490*, 421–5.

Liu, C. K., Wei, G., & Atwood, W. J. (1998). Infection of glial cells by the human polyomavirus jc is mediated by an n-linked glycoprotein containing terminal alpha(2-6)-linked sialic acids. *Journal of virology*, *72*, 4643–9.

Liu, W., Kim, G. B., Krump, N. A., Zhou, Y., Riley, J. L., & You, J. (2020). Selective reactivation of sting signaling to target merkel cell carcinoma. *Proceedings of the National Academy of Sciences of the United States of America*, *117*, 13730–13739.

Liu, W., Yang, R., Payne, A. S., Schowalter, R. M., Spurgeon, M. E., Lambert, P. F., Xu, X., Buck, C. B., & You, J. (2016). Identifying the target cells and mechanisms of merkel cell polyomavirus infection. *Cell host and microbe*, *19*, 775–87.

Liu, X., Hein, J., Richardson, S. C. W., Basse, P. H., Toptan, T., Moore, P. S., Gjoerup, O. V., & Chang, Y. (2011). Merkel cell polyomavirus large t antigen disrupts lysosome clustering by translocating human vam6p from the cytoplasm to the nucleus. *The Journal of biological chemistry*, *286*, 17079–90.

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for rna-seq data with deseq2. *Genome biology*, *15*, 550.

Loyo, M., Guerrero-Preston, R., Brait, M., Hoque, M. O., Chuang, A., Kim, M. S., Sharma, R., Liégeois, N. J., Koch, W. M., Califano, J. A., Westra, W. H., & Sidransky, D. (2010). Quantitative detection of merkel cell virus in human tissues and possible mode of transmission. *International journal of cancer*, *126*(19588496), 2991–2996.

Martel-Jantin, C., Pedergnana, V., Nicol, J. T. J., Leblond, V., Trégouët, D.-A., Tortevoye, P., Plancoulaine, S., Coursaget, P., Touzé, A., Abel, L., & Gessain, A. (2013). Merkel cell polyomavirus infection occurs during early childhood and is transmitted between siblings. *Journal of clinical virology : the official publication of the Pan American Society for Clinical Virology*, *58*, 288–91.

Matsushita, M., Iwasaki, T., Wardhani, L. O., Kuwamoto, S., Nonaka, D., Nagata, K., Kato, M., Kitamura, Y., & Hayashi, K. (2019). Decreased h3k27me3 expression is associated with merkel cell polyomavirus-negative merkel cell carcinoma, especially combined with cutaneous squamous cell carcinoma. *Anticancer research*, *39*, 5573–5579.

Mayberry, C. L., & Maginnis, M. S. (2020). Taking the scenic route: Polyomaviruses utilize multiple pathways to reach the same destination. *Viruses*, *12*.

Mesri, E. A., Feitelson, M. A., & Munger, K. (2014). Human viral oncogenesis: A cancer hallmarks analysis. *Cell host microbe*, *15*, 266–82.

Metzger, E., Wissmann, M., Yin, N., Müller, J. M., Schneider, R., Peters, A. H. F. M., Günther, T., Buettner, R., & Schüle, R. (2005). Lsd1 demethylates repressive histone marks to promote androgen-receptor-dependent transcription. *Nature*, *437*(7057), 436–439.

Moens, U., Calvignac-Spencer, S., Lauber, C., Ramqvist, T., Feltkamp, M. C. W., Daugherty, M. D., Verschoor, E. J., Ehlers, B., & Consortium, I. R. (2017). Ictv virus taxonomy profile: Polyomaviridae (2017/06/22). *The Journal of general virology*, *98*(28640744), 1159–1160.

Moens, U., Prezioso, C., & Pietropaolo, V. (2020). Genetic diversity of the noncoding control region of the novel human polyomaviruses. *Viruses*, *12*.

Morris-Love, J., Gee, G. V., O'Hara, B. A., Assetta, B., Atkinson, A. L., Dugan, A. S., Haley, S. A., & Atwood, W. J. (2019). Jc polyomavirus uses extracellular vesicles to infect target cells. *mBio*, *10*.

Musella, M., Manic, G., De Maria, R., Vitale, I., & Sistigu, A. (2017). Type-i-interferons in infection and cancer: Unanticipated dynamics with therapeutic implications. *Oncoimmunology*, *6*(28638743), e1314424–e1314424.

Nakanishi, A., Li, P. P., Qu, Q., Jafri, Q. H., & Kasamatsu, H. (2007). Molecular dissection of nuclear entry-competent sv40 during infection. *Virus Research*, *124*(1), 226–230.

Neumann, F., Borchert, S., Schmidt, C., Reimer, R., Hohenberg, H., Fischer, N., & Grundhoff, A. (2011). Replication, gene expression and particle production by a consensus merkel cell polyomavirus (mcpyv) genome. *PloS one*, *6*, e29112.

Nghiem, P., Kaufman, H. L., Bharmal, M., Mahnke, L., Phatak, H., & Becker, J. C. (2017). Systematic literature review of efficacy, safety and tolerability outcomes of chemotherapy regimens in patients with metastatic merkel cell carcinoma. *Future oncology (London, England)*, *13*, 1263–1279.

Ninova, M., Fejes Tóth, K., & Aravin, A. A. (2019). The control of gene expression and cell identity by h3k9 trimethylation. *Development (Cambridge, England)*, *146*(31540910), dev181180.

Nwogu, N., Boyne, J. R., Dobson, S. J., Poterlowicz, K., Blair, G. E., Macdonald, A., Mankouri, J., & Whitehouse, A. (2018). Cellular sheddases are induced by merkel cell polyomavirus small tumour antigen to mediate cell dissociation and invasiveness. *PLoS pathogens*, *14*, e1007276.

Padgett, B., Zurhein, G., Walker, D., Eckroade, R., & Dessel, B. (1971). Cultivation of papova-like virus from human brain with progressive multifocal leucoencephalopathy. *The Lancet*, *297*(7712), 1257–1260.

Paludan, S. R., Bowie, A. G., Horan, K. A., & Fitzgerald, K. A. (2011). Recognition of herpesviruses by the innate immune system. *Nature reviews. Immunology*, *11*(21267015), 143–154.

Park, D. E., Cheng, J., Berrios, C., Montero, J., Cortés-Cros, M., Ferretti, S., Arora, R., Tillgren, M. L., Gokhale, P. C., & DeCaprio, J. A. (2019). Dual inhibition of mdm2 and mdm4 in virus-positive merkel cell carcinoma enhances the p53 response. *Proceedings of the National Academy of Sciences of the United States of America*, *116*, 1027–1032.

Park, D. E., Cheng, J., McGrath, J. P., Lim, M. Y., Cushman, C., Swanson, S. K., Tillgren, M. L., Paulo, J. A., Gokhale, P. C., Florens, L., Washburn, M. P., Trojer, P., & DeCaprio, J. A. (2020). Merkel cell polyomavirus activates lsd1-mediated blockade of non-canonical

baf to regulate transformation and tumorigenesis (2020/04/13). *Nature cell biology*, *22*(32284543), 603–615.

Park, S.-Y., & Kim, J.-S. (2020). A short guide to histone deacetylases including recent progress on class ii enzymes. *Experimental molecular medicine*, *52*, 204–212.

Paulson, K. G., Iyer, J. G., Tegeder, A. R., Thibodeau, R., Schelter, J., Koba, S., Schrama, D., Simonson, W. T., Lemos, B. D., Byrd, D. R., Koelle, D. M., Galloway, D. A., Leonard, J. H., Madeleine, M. M., Argenyi, Z. B., Disis, M. L., Becker, J. C., Cleary, M. A., & Nghiem, P. (2011). Transcriptome-wide studies of merkel cell carcinoma and validation of intratumoral cd8+ lymphocyte invasion as an independent predictor of survival. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, *29*, 1539–46.

Paulson, K. G., Park, S. Y., Vandeven, N. A., Lachance, K., Thomas, H., Chapuis, A. G., Harms, K. L., Thompson, J. A., Bhatia, S., Stang, A., & Nghiem, P. (2018). Merkel cell carcinoma: Current us incidence and projected increases based on changing demographics (2017/11/02). *Journal of the American Academy of Dermatology*, *78*(29102486), 457–463.e2.

Paulson, K. G., Tegeder, A., Willmes, C., Iyer, J. G., Afanasiev, O. K., Schrama, D., Koba, S., Thibodeau, R., Nagase, K., Simonson, W. T., Seo, A., Koelle, D. M., Madeleine, M., Bhatia, S., Nakajima, H., Sano, S., Hardwick, J. S., Disis, M. L., Cleary, M. A., ... Nghiem, P. (2014). Downregulation of mhc-i expression is prevalent but reversible in merkel cell carcinoma. *Cancer immunology research*, *2*, 1071–9.

Peters, A. H., O'Carroll, D., Scherthan, H., Mechtler, K., Sauer, S., Schöfer, C., Weipoltshammer, K., Pagani, M., Lachner, M., Kohlmaier, A., Opravil, S., Doyle, M., Sibilia, M., & Jenuwein, T. (2001). Loss of the suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability. *Cell*, *107*, 323–37.

Pfeifer, G. P. (2018). Defining driver dna methylation changes in human cancer. *International journal of molecular sciences*, *19*.

Phillips, D. M. (1963). The presence of acetyl groups of histones. *The Biochemical journal*, *87*(13943142), 258–263.

Plank, J. L., & Dean, A. (2014). Enhancer function: Mechanistic and genome-wide insights come together. *Molecular cell*, *55*, 5–14.

Platanitis, E., & Decker, T. (2018). Regulatory networks involving stats, irfs, and nfκb in inflammation. *Frontiers in immunology*, *9*(30483250), 2542–2542.

Quinlan, A. R., & Hall, I. M. (2010). Bedtools: A flexible suite of utilities for comparing genomic features. *Bioinformatics (Oxford, England)*, *26*, 841–2.

Raghava, S., Giorda, K. M., Romano, F. B., Heuck, A. P., & Hebert, D. N. (2011). The sv40 late protein vp4 is a viroporin that forms pores to disrupt membranes for viral release. *PLOS Pathogens*, *7*(6), e1002116.

Rathi, A. V., Cantalupo, P. G., Sarkar, S. N., & Pipas, J. M. (2010). Induction of interferon-stimulated genes by simian virus 40 t antigens (2010/08/07). *Virology*, *406*(20692676), 202–211.

Richards, K. F., Guastafierro, A., Shuda, M., Toptan, T., Moore, P. S., & Chang, Y. (2015). Merkel cell polyomavirus t antigens promote cell proliferation and inflammatory cytokine gene expression. *The Journal of general virology*, *96*, 3532–3544.

Richter, A. M., Haag, T., Walesch, S., Herrmann-Trost, P., Marsch, W. C., Kutzner, H., Helmbold, P., & Dammann, R. H. (2013). Aberrant promoter hypermethylation of rassf family members in merkel cell carcinoma. *Cancers*, *5*(24252868), 1566–1576.

Ritter, C., Fan, K., Paschen, A., Reker Hardrup, S., Ferrone, S., Nghiem, P., Ugurel, S., Schrama, D., & Becker, J. C. (2017). Epigenetic priming restores the hla class-i antigen processing machinery expression in merkel cell carcinoma. *Scientific reports*, *7*(28536458), 2290–2290.

Ritter, C., Fan, K., Paulson, K. G., Nghiem, P., Schrama, D., & Becker, J. C. (2016). Reversal of epigenetic silencing of mhc class i chain-related protein a and b improves immune recognition of merkel cell carcinoma. *Scientific Reports*, *6*(1), 21678.

Rodig, S. J., Cheng, J., Wardzala, J., DoRosario, A., Scanlon, J. J., Laga, A. C., Martinez-Fernandez, A., Barletta, J. A., Bellizzi, A. M., Sadasivam, S., Holloway, D. T., Cooper, D. J., Kupper, T. S., Wang, L. C., & DeCaprio, J. A. (2012). Improved detection suggests all merkel cell carcinomas harbor merkel polyomavirus. *The Journal of clinical investigation*, *122*, 4645–53.

Ryu, W.-S., & Ryu, W.-S. Chapter 17 - retroviruses. In: *Molecular virology of human pathogenic viruses*. Boston: Academic Press, 2017, pp. 227–246.

Sahi, H., Savola, S., Sihto, H., Koljonen, V., Bohling, T., & Knuutila, S. (2014). Rb1 gene in merkel cell carcinoma: Hypermethylation in all tumors and concurrent heterozygous deletions in the polyomavirus-negative subgroup. *APMIS : acta pathologica, microbiologica, et immunologica Scandinavica*, *122*, 1157–66.

Salunke, D. M., Caspar, D. L., & Garcea, R. L. (1986). Self-assembly of purified polyomavirus capsid protein vp1. *Cell*, *46*, 895–904.

Samimi, M. (2019). Immune checkpoint inhibitors and beyond: An overview of immune-based therapies in merkel cell carcinoma. *American journal of clinical dermatology*, *20*, 391–407.

Santos-Rosa, H., Schneider, R., Bannister, A. J., Sherriff, J., Bernstein, B. E., Emre, N. C. T., Schreiber, S. L., Mellor, J., & Kouzarides, T. (2002). Active genes are tri-methylated at k4 of histone h3. *Nature*, *419*, 407–11.

Saribas, A. S., Coric, P., Hamazaspyan, A., Davis, W., Axman, R., White, M. K., Abou-Gharbia, M., Childers, W., Condra, J. H., Bouaziz, S., & Safak, M. (2016). Emerging from the unknown: Structural and functional features of agnoprotein of polyomaviruses (2016/02/24). *Journal of cellular physiology*, *231*(26831433), 2115–2127.

Schlee, M., & Hartmann, G. (2016). Discriminating self from non-self in nucleic acid sensing. *Nature reviews. Immunology*, *16*, 566–80.

Schmitt, M., Wieland, U., Kreuter, A., & Pawlita, M. (2012). C-terminal deletions of merkel cell polyomavirus large t-antigen, a highly specific surrogate marker for virally induced malignancy. *International journal of cancer*, *131*, 2863–8.

Schneider, W. M., Chevillotte, M. D., & Rice, C. M. (2014). Interferon-stimulated genes: A complex web of host defenses. *Annual review of immunology*, *32*, 513–45.

Schoggins, J. W. (2019). Interferon-stimulated genes: What do they all do? *Annual review of virology*, *6*, 567–584.

Schowalter, R. M., & Buck, C. B. (2013). The merkel cell polyomavirus minor capsid protein. *PLoS pathogens*, *9*, e1003558.

Schowalter, R. M., Pastrana, D. V., & Buck, C. B. (2011). Glycosaminoglycans and sialylated glycans sequentially facilitate merkel cell polyomavirus infectious entry. *PLoS pathogens*, *7*, e1002161.

Schowalter, R. M., Pastrana, D. V., Pumphrey, K. A., Moyer, A. L., & Buck, C. B. (2010). Merkel cell polyomavirus and two previously unknown polyomaviruses are chronically shed from human skin. *Cell host microbe*, *7*, 509–15.

Seo, G. J., Chen, C. J., & Sullivan, C. S. (2009). Merkel cell polyomavirus encodes a microrna with the ability to autoregulate viral gene expression. *Virology*, *383*, 183–7.

Shahzad, N., Shuda, M., Gheit, T., Kwun, H. J., Cornet, I., Saidj, D., Zannetti, C., Hasan, U., Chang, Y., Moore, P. S., Accardi, R., & Tommasino, M. (2013). The t antigen locus of merkel cell polyomavirus downregulates human toll-like receptor 9 expression. *Journal of virology*, *87*, 13009–19.

Shen, L., Shao, N.-Y., Liu, X., Maze, I., Feng, J., & Nestler, E. J. (2013). Diffreps: Detecting differential chromatin modification sites from chip-seq data with biological replicates. *PloS one*, *8*(23762400), e65598–e65598.

Shuda, M., Arora, R., Kwun, H. J., Feng, H., Sarid, R., Fernández-Figueras, M.-T., Tolstov, Y., Gjoerup, O., Mansukhani, M. M., Swerdlow, S. H., Chaudhary, P. M., Kirkwood, J. M., Nalesnik, M. A., Kant, J. A., Weiss, L. M., Moore, P. S., & Chang, Y. (2009). Human merkel cell polyomavirus infection i. mcv t antigen expression in merkel cell carcinoma, lymphoid tissues and lymphoid tumors. *International journal of cancer*, *125*, 1243–9.

Shuda, M., Chang, Y., & Moore, P. S. (2014). Merkel cell polyomavirus-positive merkel cell carcinoma requires viral small t-antigen for cell proliferation. *The Journal of investigative dermatology*, *134*(5), 1479–1481.

Shuda, M., Feng, H., Kwun, H. J., Rosen, S. T., Gjoerup, O., Moore, P. S., & Chang, Y. (2008). T antigen mutations are a human tumor-specific signature for merkel cell polyomavirus. *Proceedings of the National Academy of Sciences of the United States of America*, *105*, 16272–7.

Shuda, M., Kwun, H. J., Feng, H., Chang, Y., & Moore, P. S. (2011). Human merkel cell polyomavirus small t antigen is an oncoprotein targeting the 4e-bp1 translation regulator. *The Journal of clinical investigation*, *121*, 3623–34.

Siebels, S., Czech-Sioli, M., Spohn, M., Schmidt, C., Theiss, J., Indenbirken, D., Günther, T., Grundhoff, A., & Fischer, N. (2020). Merkel cell polyomavirus dna replication induces senescence in human dermal fibroblasts in a kap1/trim28-dependent manner. *mBio*, *11*.

Singh, S., Sadanandam, A., & Singh, R. K. (2007). Chemokines in tumor angiogenesis and metastasis. *Cancer metastasis reviews*, *26*(17828470), 453–467.

Skvortsova, K., Iovino, N., & Bogdanović, O. (2018). Functions and mechanisms of epigenetic inheritance in animals. *Nature reviews. Molecular cell biology*, *19*, 774–790.

Smith, S. D. B., Erdag, G., Cuda, J. D., Rangwala, S., Girardi, N., Bibee, K., Orens, J. B., Prono, M. D., Toptan, T., & Loss, M. J. (2018). Treatment of human polyomavirus-7-associated rash and pruritus with topical cidofovir in a lung transplant patient: Case report and literature review. *Transplant infectious disease : an official journal of the Transplantation Society*, *20*.

Sowd, G. A., & Fanning, E. (2012). A wolf in sheep's clothing: Sv40 co-opts host genome maintenance proteins to replicate viral dna. *PLoS pathogens*, *8*, e1002994.

Stakaitytė, G., Nwogu, N., Dobson, S. J., Knight, L. M., Wasson, C. W., Salguero, F. J., Blackbourn, D. J., Blair, G. E., Mankouri, J., Macdonald, A., & Whitehouse, A. (2018). Merkel cell polyomavirus small t antigen drives cell motility via rho-gtpase-induced filopodium formation. *Journal of virology*, *92*.

Stang, A., Becker, J. C., Nghiem, P., & Ferlay, J. (2018). The association between geographic location and incidence of merkel cell carcinoma in comparison to melanoma: An international assessment. *European journal of cancer (Oxford, England : 1990)*, *94*, 47–60.

Starrett, G. J., Marcelus, C., Cantalupo, P. G., Katz, J. P., Cheng, J., Akagi, K., Thakuria, M., Rabinowits, G., Wang, L. C., Symer, D. E., Pipas, J. M., Harris, R. S., & DeCaprio, J. A. (2017). Merkel cell polyomavirus exhibits dominant control of the tumor genome and transcriptome in virus-associated merkel cell carcinoma. *mBio*, *8*.

Starrett, G. J., Thakuria, M., Chen, T., Marcelus, C., Cheng, J., Nomburg, J., Thorner, A. R., Slevin, M. K., Powers, W., Burns, R. T., Perry, C., Piris, A., Kuo, F. C., Rabinowits, G., Giobbie-Hurder, A., MacConaill, L. E., & DeCaprio, J. A. (2020). Clinical and molecular characterization of virus-positive and virus-negative merkel cell carcinoma. *Genome medicine*, *12*, 30.

Sullivan, C. S., Grundhoff, A. T., Tevethia, S., Pipas, J. M., & Ganem, D. (2005). Sv40-encoded micrornas regulate viral gene expression and reduce susceptibility to cytotoxic t cells. *Nature*, *435*, 682–6.

Sunshine, J. C., Jahchan, N. S., Sage, J., & Choi, J. (2018). Are there multiple cells of origin of merkel cell carcinoma? (2018/01/11). *Oncogene*, *37*(29321666), 1409–1416.

Sweet, B. H., & Hilleman, M. R. (1960). The vacuolating virus, s.v. 40. *Proceedings of the Society for Experimental Biology and Medicine. Society for Experimental Biology and Medicine (New York, N.Y.)*, *105*, 420–7.

Tai, P. T., Yu, E., Winquist, E., Hammond, A., Stitt, L., Tonita, J., & Gilchrist, J. (2000). Chemotherapy in neuroendocrine/merkel cell carcinoma of the skin: Case series and review of 204 cases. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, *18*, 2493–9.

Theiss, J. M., 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 pathogens*, *11*, e1004974.

Thorvaldsdóttir, H., Robinson, J. T., & Mesirov, J. P. (2013). Integrative genomics viewer (igv): High-performance genomics data visualization and exploration (2012/04/19). *Briefings in bioinformatics*, *14*(22517427), 178–192.

Tie, F., Banerjee, R., Stratton, C. A., Prasad-Sinha, J., Stepanik, V., Zlobin, A., Diaz, M. O., Scacheri, P. C., & Harte, P. J. (2009). Cbp-mediated acetylation of histone h3 lysine 27 antagonizes drosophila polycomb silencing. *Development (Cambridge, England)*, *136*, 3131–41.

Tilling, T., & Moll, I. (2012). Which are the cells of origin in merkel cell carcinoma? *Journal of skin cancer*, *2012*, 680410.

Toker, C. (1972). Trabecular carcinoma of the skin. *Archives of dermatology*, *105*, 107–10.

Tolstov, Y. L., Pastrana, D. V., Feng, H., Becker, J. C., Jenkins, F. J., Moschos, S., Chang, Y., Buck, C. B., & Moore, P. S. (2009). Human merkel cell polyomavirus infection ii. mcv is a common human infection that can be detected by conformational capsid epitope immunoassays. *International journal of cancer*, *125*, 1250–6.

Tornesello, M. L., Annunziata, C., Tornesello, A. L., Buonaguro, L., & Buonaguro, F. M. (2018). Human oncoviruses and p53 tumor suppressor pathway deregulation at the origin of human cancers. *Cancers*, *10*.

Trojer, P., & Reinberg, D. (2007). Facultative heterochromatin: Is there a distinctive molecular signature? *Molecular cell*, *28*, 1–13.

Tsang, S. H., Wang, R., Nakamaru-Ogiso, E., Knight, S. A. B., Buck, C. B., & You, J. (2015). The oncogenic small tumor antigen of merkel cell polyomavirus is an iron-sulfur cluster protein that enhances viral dna replication. *Journal of virology*, *90*(26608318), 1544–1556.

Tsang, S. H., Wang, X., Li, J., Buck, C. B., & You, J. (2014). Host dna damage response factors localize to merkel cell polyomavirus dna replication sites to support efficient viral dna replication (2014/01/03). *Journal of virology*, *88*(24390338), 3285–3297.

van Veenendaal, L. M., van Akkooi, A. C. J., Verhoef, C., Grünhagen, D. J., Klop, W. M. C., Valk, G. D., & Tesselaar, M. E. T. (2018). Merkel cell carcinoma: Clinical outcome and prognostic factors in 351 patients. *Journal of surgical oncology*, *117*, 1768–1775.

van der Meijden, E., Janssens, R. W. A., Lauber, C., Bouwes Bavinck, J. N., Gorbalenya, A. E., & Feltkamp, M. C. W. (2010). Discovery of a new human polyomavirus associated with trichodysplasia spinulosa in an immunocompromized patient. *PLoS pathogens*, *6*, e1001024.

van der Meijden, E., Kazem, S., Dargel, C. A., van Vuren, N., Hensbergen, P. J., & Feltkamp, M. C. W. (2015). Characterization of t antigens, including middle t and alternative t, expressed by the human polyomavirus associated with trichodysplasia spinulosa. *Journal of virology*, *89*, 9427–39.

van der Meijden Els, Siamaque, K., Dargel Christina, A., van Vuren Nick, Hensbergen Paul, J., Feltkamp Mariet C., W., & Imperiale M., J. (2015). Characterization of t antigens, including middle t and alternative t, expressed by the human polyomavirus associated with trichodysplasia spinulosa. *Journal of Virology*, *89*(18), 9427–9439.

Veija, T., Koljonen, V., Bohling, T., Kero, M., Knuutila, S., & Sarhadi, V. K. (2017). Aberrant expression of alk and ezh2 in merkel cell carcinoma. *BMC cancer*, *17*, 236.

Verhaegen, M. E., Mangelberger, D., Harms, P. W., Eberl, M., Wilbert, D. M., Meireles, J., Bichakjian, C. K., Saunders, T. L., Wong, S. Y., & Dlugosz, A. A. (2017). Merkel cell polyomavirus small t antigen initiates merkel cell carcinoma-like tumor development in mice. *Cancer research*, *77*, 3151–3157.

Verhaegen, M. E., Mangelberger, D., Harms, P. W., Vozheiko, T. D., Weick, J. W., Wilbert, D. M., Saunders, T. L., Ermilov, A. N., Bichakjian, C. K., Johnson, T. M., Imperiale, M. J., & Dlugosz, A. A. (2015). Merkel cell polyomavirus small t antigen is oncogenic in transgenic mice. *The Journal of investigative dermatology*, *135*, 1415–1424.

Walker, J., Kasturi, V., Lebbe, C., Sandhu, S. K., Grignani, G., Hennessy, M. G., Hildemann, S., Reed, J., & Hariharan, S. (2018). Second-line avelumab treatment of patients (pts) with metastatic merkel cell carcinoma (mmcc): Experience from a global expanded access program (eap). *JCO*, *36*(15_suppl), 9537–9537.

Wang, X., Li, J., Schowalter, R. M., Jiao, J., Buck, C. B., & You, J. (2012). Bromodomain protein brd4 plays a key role in merkel cell polyomavirus dna replication (2012/11/08). *PLoS pathogens*, *8*(23144621), e1003021–e1003021.

Wang, Z., Zang, C., Rosenfeld, J. A., Schones, D. E., Barski, A., Cuddapah, S., Cui, K., Roh, T.-Y., Peng, W., Zhang, M. Q., & Zhao, K. (2008). Combinatorial patterns of histone acetylations and methylations in the human genome. *Nature genetics*, *40*, 897–903.

Weber, K., Bartsch, U., Stocking, C., & Fehse, B. (2008). A multicolor panel of novel lentiviral "gene ontology" (lego) vectors for functional gene analysis. *Molecular therapy : the journal of the American Society of Gene Therapy*, *16*, 698–706.

Wu, J. H., Narayanan, D., Limmer, A. L., Simonette, R. A., Rady, P. L., & Tyring, S. K. (2019). Merkel cell polyomavirus small t antigen induces dna damage response.

Wu, J. H., Simonette, R. A., Nguyen, H. P., Rady, P. L., & Tyring, S. K. (2016). Merkel cell polyomavirus in merkel cell carcinogenesis: Small t antigen-mediates c-jun phosphorylation. *Virus genes*, *52*, 397–9.

Wu, Z., Graf, F. E., & Hirsch, H. H. (2021). Antivirals against human polyomaviruses: Leaving no stone unturned. *Reviews in medical virology*, *31*, e2220.

Xu, S., Grullon, S., Ge, K., & Peng, W. (2014). Spatial clustering for identification of chip-enriched regions (sicer) to map regions of histone methylation patterns in embryonic stem cells. *Methods in molecular biology (Clifton, N.J.)*, *1150*, 97–111.

Yang, R., Lee, E. E., Kim, J., Choi, J. H., Kolitz, E., Chen, Y., Crewe, C., Salisbury, N. J. H., Scherer, P. E., Cockerell, C., Smith, T. R., Rosen, L., Verlinden, L., Galloway, D. A., Buck, C. B., Feltkamp, M. C., Sullivan, C. S., & Wang, R. C. (2021). Characterization of alto-encoding circular rnas expressed by merkel cell polyomavirus and trichodysplasia spinulosa polyomavirus. *PLoS pathogens*, *17*, e1009582.

Youlden, D. R., Soyer, H. P., Youl, P. H., Fritschi, L., & Baade, P. D. (2014). Incidence and survival for merkel cell carcinoma in queensland, australia, 1993-2010. *JAMA dermatology*, *150*, 864–72.

Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., Nusbaum, C., Myers, R. M., Brown, M., Li, W., & Liu, X. S. (2008). Model-based analysis of chip-seq (macs). *Genome biology*, *9*, R137.

Zhao, J., Jia, Y., Shen, S., Kim, J., Wang, X., Lee, E., Brownell, I., Cho-Vega, J. H., Lewis, C., Homsi, J., Sharma, R. R., & Wang, R. C. (2020). Merkel cell polyomavirus small t antigen activates noncanonical nf-$\varkappa$b signaling to promote tumorigenesis. *Molecular cancer research : MCR*, *18*, 1623–1637.

Zou, W., & Imperiale, M. J. (2021). Biology of polyomavirus mirna. *Frontiers in Microbiology*, *12*.

Danksagung

Zunächst möchte ich mich bei Prof. Dr. Nicole Fischer dafür bedanken, dass ich meine Promotion in ihrer Arbeitsgruppe am Institut für Medizinische Mikrobiologie, Virologie und Hygiene durchführen konnte. Vielen Dank für die tolle Zusammenarbeit und die Bereitstellung des interessanten Forschungsprojekts.

Ebenso bedanke ich mich bei meinem Ko-Betreuer Prof. Dr. Stefan Hoth aus der Abteilung für Molekulare Pflanzenphysiologie am Biozentrum Klein Flottbek.

Ein weiterer großer Dank geht an meinen zweiten Ko-Betreuer Prof. Dr. Adam Grundhoff aus der Abteilung für Virus Genomik am Leibniz-Institut für Experimentelle Virologie. Außerdem danke ich für die enge Zusammenarbeit mit allen Mitgliedern der AG Grundhoff. Insbesondere danke ich Daniela, Christina und Kerstin für die exzellente Sequenzierungsarbeit, und Thomas und Jacqueline für die Hilfe bei der Auswertung der Sequenzierdaten.

Ein großer Dank geht auch an Dr. Jiabin Huang, der mir bei den bioinformatischen Analysen in dieser Arbeit geholfen hat. Ich danke außerdem allen Mitgliedern von EPILOG für den wissenschaftlichen Austausch und die netten Abende außerhalb des Labors.

Besonders danke ich allen Mitarbeitern des Instituts für Mikrobiologie, Virologie und Hygiene für die super Stimmung im Labor. Durch die vielen spaßigen Stunden auch außerhalb des Labors seid ihr mir sehr ans Herz gewachsen. Ein großer Dank geht hier vor allem auch an Tabea, die parallel mit mir an ihrer Promotion gearbeitet hat und mit der ich alles besprechen konnte.

Außerdem möchte ich mich herzlich bei Tabea, Vero, Ute, Harsha und Lindsay für das Korrekturlesen dieser Arbeit bedanken.

Ein besonderer Dank geht an dieser Stelle auch an alle meine Freunde, die mich während der gesamten Zeit begleitet haben, mich ablenken konnten und mich immer unterstützt haben. Ihr seid die besten und ich bin froh euch meine Freunde nennen zu können.

Zu guter letzt danke ich meinen Eltern, ohne die ich nicht hier angekommen wäre. Meinem Basti danke ich auch dafür, mir stets den Rücken zu stärken und einfach für mich da zu sein. Danke, dass ihr mich immer unterstützt und vor allem, an mich geglaubt habt.

Eidesstattliche Versicherung

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

Hamburg, den 28.03.2022

Denise Ohnezeit