

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/33437> holds various files of this Leiden University dissertation.

**Author:** Kazem, Siamaque

**Title:** The trichodysplasia spinulosa-associated polyomavirus : infection, pathogenesis, evolution and adaptation

**Issue Date:** 2015-06-17

# **The Trichodysplasia Spinulosa- Associated Polyomavirus**

**Infection, Pathogenesis, Evolution and Adaptation**

**Siamaque Kazem**

## Colophon

- Author:** Siamaque Kazem (MSc, Ing)
- Content:** Doctoral dissertation, Leiden University, Leiden, The Netherlands, 2015
- Title:** The Trichodysplasia Spinulosa-Associated Polyomavirus:  
Infection, Pathogenesis, Evolution and Adaptation
- Cover layout:** Light of the stars and galaxies in the vast outer space. In front, illustration of one of the tiniest virus particles (a polyomavirus) known to human-kind, from which its genetic material and its whereabouts (in this dissertation) are coming to light
- Pictures:** NASA, ESA, and the Hubble SM4 ERO Team; Virusworld, Inst. for Molecular Virology, Madison  
Reprinted with permission and animation by S. Kazem
- Funding:** The research described in this dissertation was carried out at the Department of Medical Microbiology, Leiden University Medical Center (LUMC), Leiden, The Netherlands. The research and printing of this dissertation was financially supported by the Department of Medical Microbiology of LUMC
- Printing:** Uitgeverij BOXPress | | Proefschriftmaken.nl
- ISBN:** 978-94-6295-193-8

**ALL RIGHTS RESERVED.** No part of this dissertation may be reproduced, stored or transmitted in any form or by any means - graphic, electronic or mechanical, including photocopying, recording, taping, web distribution, or information storage retrieval systems - without prior written permission of the author. Copyright of the published articles was transferred to the respective journal or publisher

# **The Trichodysplasia Spinulosa- Associated Polyomavirus**

Infection, Pathogenesis, Evolution and Adaptation

## **Proefschrift**

ter verkrijging van de graad van Doctor aan de Universiteit Leiden  
op gezag van Rector Magnificus prof. mr. C.J.J.M. Stolker,  
volgens het besluit van het College voor Promoties  
te verdedigen op woensdag 17 juni 2015  
klokke 16:15 uur

door

**Siamaque Kazem**

geboren op 23 december 1982  
te Kabul, Afghanistan

# Promotiecommissie

*Promotors:*

**Prof. dr. A.C.M. Kroes**

**Prof. dr. A.E. Gorbalenya**

*Co-promotor:*

**Dr. M.C.W. Feltkamp**

*Overige leden:*

**Prof. dr. E.J. Snijder**

**Prof. dr. M.H. Vermeer**

**Prof. dr. B. Akgül**

***(Universiteit van Keulen)***

**Prof. dr. A. zur Hausen**

***(Maastricht Universitair Medisch Centrum)***

*Everything in the universe was born in the light of a wonderful  
star billions of light-years away. All will transform to dust of  
that beauty once more, all have always been!*

*For my beloved brother Babak Kazem (1985-2013)*



# Table of Contents

		Page
<b>Part I: Introduction</b>		
<b>Chapter 1</b>	<b>General introduction</b> From Stockholm to Malawi: recent developments in studying human polyomaviruses. <i>Journal of General Virology</i> (2013)	<b>11</b>
<b>Part II: TSPyV infection and pathogenesis</b>		
<b>Chapter 2</b>	<b>Clinical and viral aspects of trichodysplasia spinulosa</b> The trichodysplasia spinulosa-associated polyomavirus: virological background and clinical implications. <i>Acta pathologica, microbiologica, et immunologica Scandinavica</i> (2013)	<b>43</b>
<b>Chapter 3</b>	<b>Active TSPyV infection in trichodysplasia spinulosa</b> Trichodysplasia spinulosa is characterized by active polyomavirus infection. <i>Journal of Clinical Virology</i> (2012)	<b>63</b>
<b>Chapter 4</b>	<b>Immunohistopathology of trichodysplasia spinulosa</b> Polyomavirus-associated trichodysplasia spinulosa involves hyperproliferation, pRB phosphorylation and upregulation of p16 and p21. <i>PLoS ONE</i> (2014)	<b>75</b>
<b>Part III: TSPyV host adaptation and evolution</b>		
<b>Chapter 5</b>	<b>Polyomavirus host adaptation</b> Interspecific adaptation by binary choice at <i>de novo</i> polyomavirus T antigen site through accelerated codon-constrained Val-Ala toggling within an intrinsically disordered region. <i>Nucleic Acids Research</i> (2015)	<b>93</b>
<b>Chapter 6</b>	<b>Constrained TSPyV evolution in human population</b> Adaptation of trichodysplasia spinulosa-associated polyomavirus to the human population is mediated by middle T antigen and involves COCO-VA toggling. <i>Submitted for publication</i>	<b>137</b>
<b>Part IV: Discussion and Summary</b>		
<b>Chapter 7</b>	<b>Summarizing discussion</b>	<b>171</b>
<b>Chapter 8</b>	<b>Samenvatting (Dutch) / (Persian/Dari) مختصر</b>	<b>195</b>
<b>Part V: Appendix</b>		
<b>Appendix 1</b>	<b>List of Abbreviations</b>	<b>207</b>
<b>Appendix 2</b>	<b>Acknowledgement</b>	<b>213</b>
<b>Appendix 3</b>	<b>Curriculum Vitae</b>	<b>217</b>
<b>Appendix 4</b>	<b>List of Publications</b>	<b>221</b>



VePyV1 CaPyV JCPyV *mPyV* SA12 RacPyV GHPyV BKPyV

*BatsgU* CPyV OrAPyV1 FPyV MptV APPyV1 SqPyV LPyV CSLPyV

EPyV PRPyV1 APP<sub>g</sub>V2 KIPyV **MIRgV** PtvPyV2c OtPyV1 **STLPyV**

MFPyV1 *KSgU9* SV40 TSPyV CoPyV1 PPPyV CPPyV HPyV12

MXPyV PtvPyV1a PDPyV EIPyV1 AtPPyV1 HaPyV (TggPyV1

CdPyV DRPyV MWPyV APyV CaPyV1 HPyV7 ChPyV MasPyV

WUPyV *HSp<sub>g</sub>V6* BPyV MCPyV OrAPyV2 MMPyV SLPyV HPyV10

VePyV1 CaPyV JCPyV *mPyV* SA12

*BatPyV* CPyV OraPyV1 FPyV MptV APPy

EPyV PRPyV1 APPyV2 KIPyV **MiPyV** P

MFPyV1 *HSPyV9* SV40 TSPyV CoPyV1

MPyV PtvPyV1a PDPyV EiPyV1 At

CdPyV DRPyV MWPyV APyV CAPyV1

WUPyV *HSPyV6* BPyV MCPyV OraPyV2

# Part I

## Introduction

PyV BKPyV

PyV CSLPyV

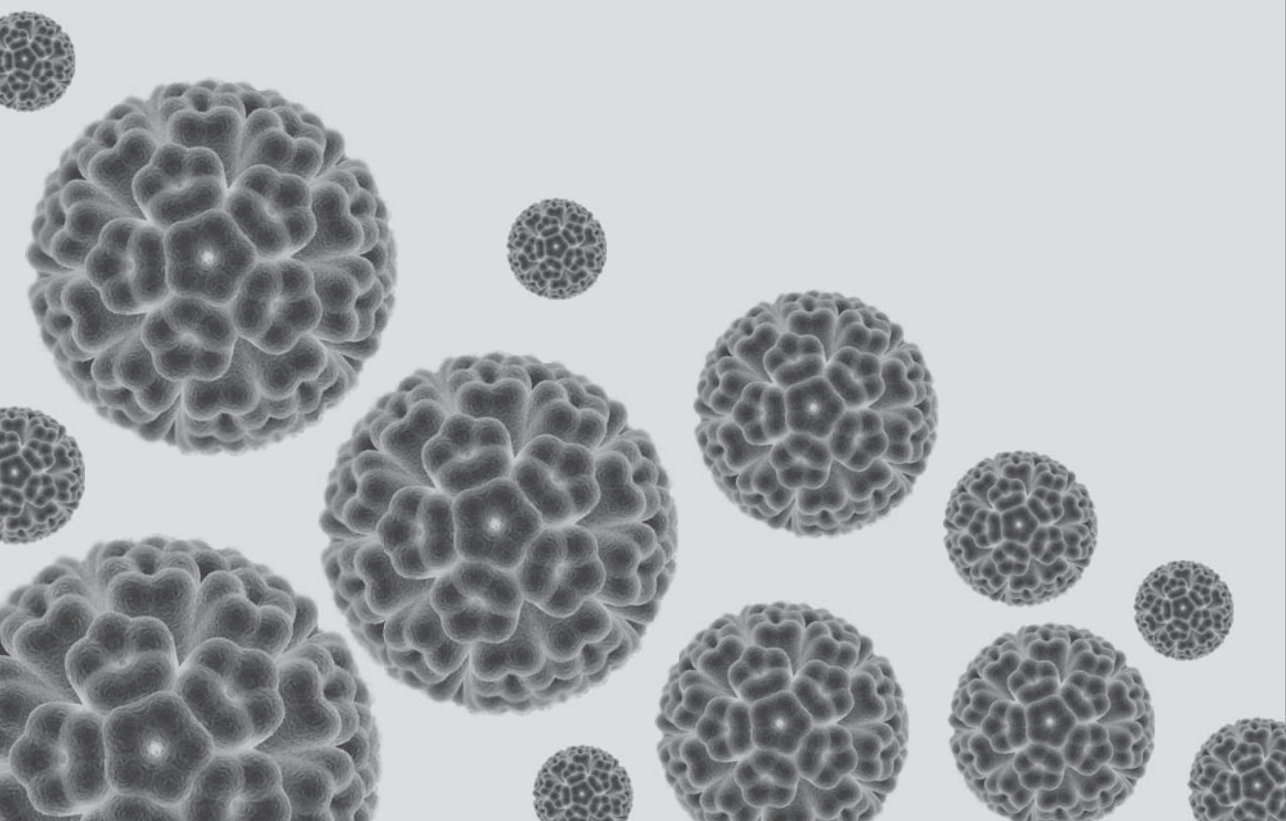
PyV1 STLPyV

PyV HPyV12

PyV GgPyV1

PyV MasPyV

PyV HPyV10



# Chapter 1

## General introduction

*Adapted from\*:*

### **From Stockholm to Malawi: recent developments in studying human polyomaviruses**

*Authors:*

Mariet Feltkamp<sup>1</sup>  
Siamaque Kazem<sup>1</sup>  
Els van der Meijden<sup>1</sup>  
Chris Lauber<sup>1</sup>  
Alexander Gorbalenya<sup>1,2</sup>

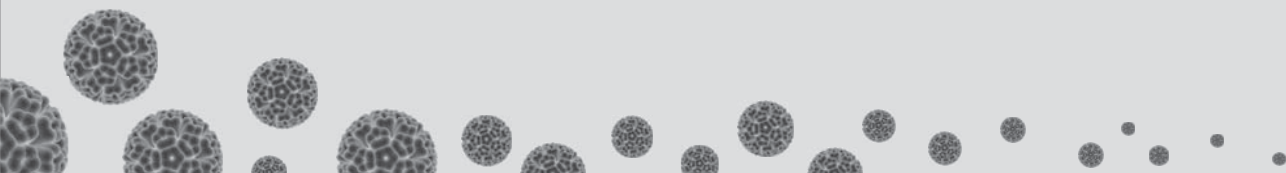
*Affiliations:*

<sup>1</sup> Department of Medical Microbiology, Leiden University Medical Center, Leiden, The Netherlands, <sup>2</sup> Faculty of Bioengineering and Bioinformatics, Lomonosov Moscow State University, 119899 Moscow, Russia

*Published in original form:*

Journal of General Virology, 2013 (94): 482 - 496

*\* Note: Adaptation of this chapter from the original published article concerns textual and figure adjustments. The text has been updated with novel data from the literature until mid-2014. Figures 2, 7 and 8 are updated with novel polyomaviruses. Figures 3, 5 and 6 were not part of the original publication.*



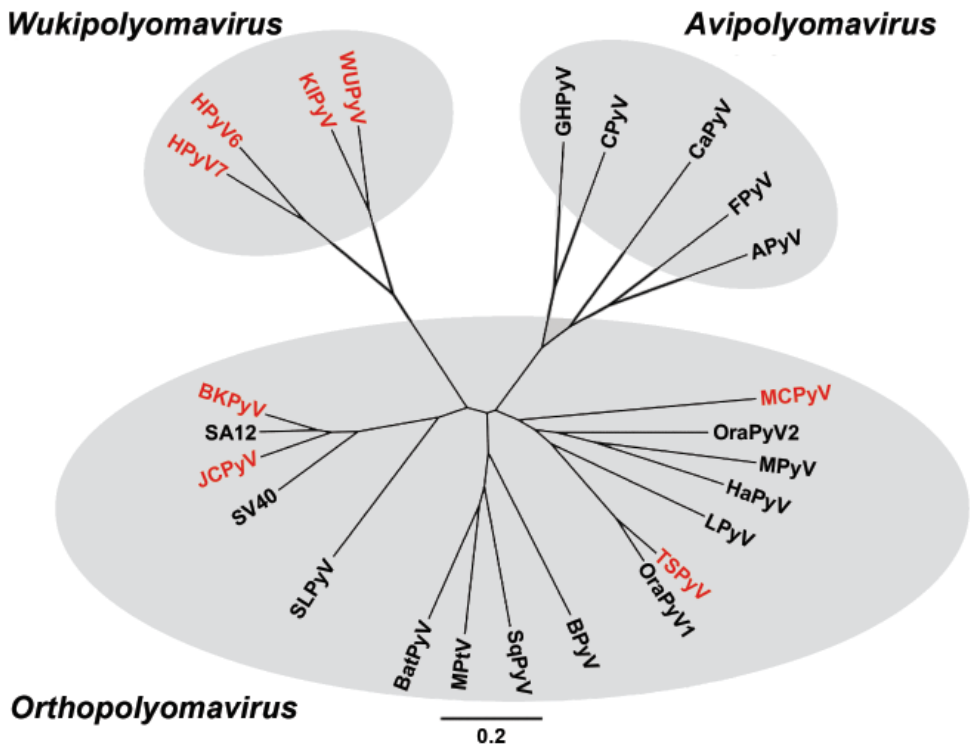
## Abstract

Until a few years ago, the polyomavirus family (*Polyomaviridae*) included only a dozen viruses mostly identified in avian and mammalian hosts. Two of these, the JC and BK-polyomaviruses isolated a long time ago, are known to infect humans and cause severe illness in immunocompromised hosts. Since 2007, tens of new polyomaviruses were identified, including at least eleven that infect humans. Among them is the polyomavirus associated with trichodysplasia spinulosa (TSPyV). In this introductory chapter, the recent developments in studying the novel human polyomaviruses until mid-2014 are summarized, which sets the stage for further investigation into TSPyV infection, pathogenesis, evolution and host adaptation.

Abstract

## Historical background

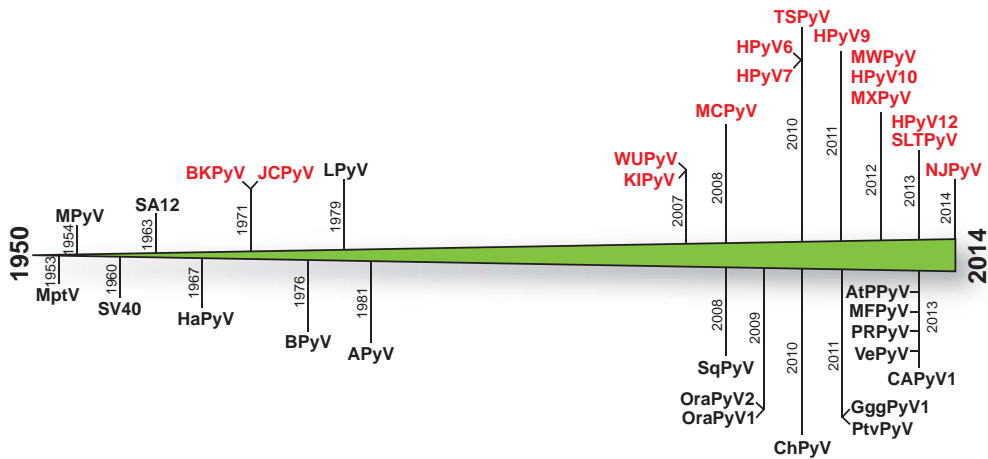
**P**olyomaviruses have been recognized as a separate virus family (*Polyomaviridae*) since 1999. Before that time, they formed the genus *Polyomavirus* in the *Papovaviridae* family that also contained the genus *Papillomavirus* [1, 2]. The first members of the polyomavirus family, mouse polyomavirus (MPyV) and SV40, were identified halfway through the last century as filterable agents that caused tumors in newborn mice and hamsters [3 - 5]. Subsequently, the first two human polyomaviruses were discovered in 1971. The JC-polyomavirus (JCPyV) was identified in a brain tissue extract from a progressive multifocal leukoencephalopathy (PML) patient with the initials J.C. [6]. The first reported BK-polyomavirus (BKPyV) was isolated from the urine of a nephropathic kidney transplant patient with the initials B.K. [7]. Both JCPyV and BKPyV have no oncogenic capability in humans despite being phylogenetically closely related to SV40 that was associated with oncogenic properties in humans, but this issue remained controversial so far (**Figure 1**) [2, 8].



**Figure 1.** Phylogenetic grouping of 25 polyomaviruses that were known up until October 2010, including 8 human viruses (red), as proposed to the International Committee on Taxonomy of Viruses (ICTV). The tree is based on the whole genomic nucleotide sequences. Figure adapted from [2].

Subsequently, more polyomaviruses were identified in rodents, non-human primates, cattle and birds, but not in humans until this century. From 2007 on, at least eleven additional human polyomaviruses (HPyVs) were discovered. For three of those (i.e., MCPyV, TSPyV and NJPyV) involvement in disease development in immunosuppressed and/or elderly patients seems highly likely [9 - 11]. They seem to fit the 'opportunistic' polyomavirus profile already known for JCPyV and BKPyV characterized by symptomatic reactivation after a long period of persistent latent infection. For the other novel HPyVs, disease association has not been demonstrated and their pathogenicity is still unknown.

A timeline of a selection of polyomavirus discoveries is illustrated in **Figure 2**. A list of all polyomaviruses identified up until mid-2014, which prototype the (tentative) name-sake species are shown in **Table 1**, including name abbreviations and GenBank (RefSeq) numbers of genome sequences.



**Figure 2.** A time line is shown that indicates the discovery of mainly human (red) and a number of animal polyomaviruses until June-2014. Additionally discovered mammal and bird polyomaviruses are indicated in **Table 1**. The year of discovery refers to the first description of the virus, not necessarily to the year that the full DNA genome was sequenced. The abbreviated names of the viruses are explained in **Table 1**.

## Identification of new human polyomaviruses

With the development of nucleic acid amplification and detection techniques, the need for cytopathic changes observed in cell culture to detect the presence of a virus was bypassed. Recent versions of these molecular techniques allowed sensitive and high-throughput analyses of large number of clinical samples that led to a revolution of virus discovery. These techniques are frequently combined with strategies to enrich the original sample for viral DNA or RNA by reducing the content of hosts genomic DNA. Here, recent discoveries of human polyomaviruses are summarized.

**Table 1:** List of 58 polyomavirus names identified up to mid-2014

Host type	Polyomaviruses	Year*	RefSeq**
Human	BK polyomavirus (BKPyV)	1971 [6]	NC_001538
	JC polyomavirus (JCPyV)	1971 [7]	NC_001699
	KI polyomavirus (KIPyV)	2007 [12]	NC_009238
	WU Polyomavirus (WUPyV)	2007 [13]	NC_009539
	Merkel cell polyomavirus (MCPyV)	2008 [9]	NC_010277
	Human polyomavirus 6 (HPyV6)	2010 [10]	NC_014406
	Human polyomavirus 7 (HPyV7)	2010 [14]	NC_014407
	Trichodysplasia spinulosa-associated polyomavirus (TSPyV)	2010 [14]	NC_014361
	Human polyomavirus type 9 (HPyV9)	2011 [15]	NC_015150
	Malawi polyomavirus (MWPyV)	2012 [16]	NC_018102
	Human polyomavirus type 10 (HPyV10)	2012 [17]	JX262162
	Mexico polyomavirus (MXPyV)	2012 [18]	JX259273
	St. Louis polyomavirus (STLPyV)	2013 [19]	NC_020106
Human polyomavirus type 12 (HPyV12)	2013 [20]	NC_020890	
New Jersey polyomavirus (NJPyV)	2014 [11]	KF954417	
Non-human primate	Simian virus 40 (SV40)	1962 [3]	NC_001669
	Baboon polyomavirus 1 (SA12)	1963 [21]	NC_007611
	B-lymphotropic polyomavirus (LPyV)	1979 [22]	NC_004763
	Squirrel monkey polyomavirus (SqPyV)	2008 [23]	NC_009951
	Bornean orang-utan polyomavirus (OraPyV1)	2010 [24]	NC_013439
	Sumatran orang-utan polyomavirus (OraPyV2)	2010 [24]	FN356901
	Chimpanzee polyomavirus (ChPyV)	2010 [25]	NC_014743
	Gorilla gorilla gorilla polyomavirus (GggPyV1)	2011 [26]	HQ385752
	Pan troglodytes verus polyomavirus (PtvPyV1a)	2011 [26]	HQ385746
	Pan troglodytes verus polyomavirus (PtvPyV2c)	2011 [26]	HQ385749
	Cebus albifrons polyomavirus (CAPyV1)	2013 [27]	NC_019854
	Ateles paniscus polyomavirus (AtPyV1)	2013 [27]	NC_019853
	Macaca fascicularis polyomavirus (MFPyV1)	2013 [27]	NC_019851
Ptilocolobus rufomitratus polyomavirus (PRPyV1)	2013 [27]	NC_019850	
Vervet monkey polyomavirus (VePyV1)	2013 [27]	NC_019844	
Mammal (other)	Murine pneumotropic virus (Mptv)	1953 [28]	NC_001505
	Murine polyomavirus (MPyV)	1953 [5]	NC_001515
	Hamster polyomavirus (HaPyV)	1967 [29]	NC_001663
	Bovine polyomavirus (BPyV)	1976 [30]	NC_001442
	Bat polyomavirus (BatPyV)	2009 [31]	NC_011310
	California sea lion polyomavirus (CSLPyV)	2010 [32]	NC_013796
	Mastomys polyomavirus (MasPyV)	2011 [33]	AB588640
	Equine polyomavirus (EPyV)	2012 [34]	JQ412134
	Molossus molossus polyomavirus (MMPyV)	2012 [35]	JQ958893
	Desmodus rotundus polyomavirus (DRPyV)	2012 [35]	JQ958892
	Pteronotus parnellii polyomavirus (PPPvV)	2012 [35]	JQ958891
	Pteronotus davi polyomavirus (PDPyV)	2012 [36]	NC_020070
	Artibeus planirostris polyomavirus (APPyV2)	2012 [35]	JQ958890
	Artibeus planirostris polyomavirus (APPyV1)	2012 [35]	JQ958887
	Carollia perspicillata polyomavirus (CPPyV)	2012 [35]	JQ958889
	Sturnira lilium polyomavirus (SLPyV)	2012 [35]	JQ958888
	Chaerephon polyomavirus (CoPyV1)	2012 [36]	NC_020065
	Cardioderma polyomavirus (CdPyV)	2012 [36]	NC_020067
	Eidolon polyomavirus (EiPyV1)	2012 [36]	NC_020068
Miniopterus polyomavirus (MiPyV)	2012 [36]	NC_020069	
Otomops polyomavirus (OtPyV1)	2012 [36]	NC_020071	
Raccoon polyomavirus (RacPyV)	2013 [37]	JQ178241	
African elephant polyomavirus (AelPyV)	2013 [38]	NC_022519	
Bird	Avian polyomavirus (APyV)	1981 [39]	NC_004764
	Goose hemorrhagic polyomavirus (GHPyV)	2000 [40]	NC_004800
	Finch polyomavirus (FPyV)	2006 [41]	NC_007923
	Crow polyomavirus (CPyV)	2006 [41]	NC_007922
	Canary polyomavirus isolate (CaPyV)	2010 [42]	GU345044

\* Year of first description, not necessarily the year of first complete genome sequencing

\*\* Genome sequence according GenBank (RefSeq) number (<http://www.ncbi.nlm.nih.gov/RefSeq/>)



## **KI-polyomavirus (KIPyV) and WU-polyomavirus (WUPyV)**

In 2007, almost coincidentally, two novel HPyVs were reported; the Karolinska Institute polyomavirus (KIPyV) identified in Stockholm (Sweden), and the Washington University polyomavirus (WUPyV) identified in St. Louis (USA). Both were discovered in respiratory tract samples from individuals with (acute) respiratory tract infections, but by different methods.

KIPyV was identified by pooling centrifuged and filtered supernatants from randomly selected DNase-treated nasopharyngeal aspirates, from which DNA and RNA were extracted [12]. The nucleic acids served as a template for random-PCR, and the products were separated by gel electrophoresis, cloned and sequenced. This resulted in several sequence reads that were subsequently trimmed, clustered and sorted using dedicated software. Finally, BLAST searches were performed to look for similarities with known (viral) sequences listed in GenBank.

For the identification of WUPyV, a ‘shotgun sequencing’ strategy was used on a single nasopharyngeal aspirate that proved negative for 17 known viral respiratory pathogens in diagnostic PCR. Total nucleic acid isolated from the sample was randomly amplified and PCR products were cloned and sequenced. The DNA sequence reads were analyzed in a comparable fashion to the above for KIPyV, which revealed several sequences that displayed significant homology with JCPyV and BKPyV [13]. Subsequent analyses of the compiled WUPyV genome showed highest sequence homology to KIPyV.

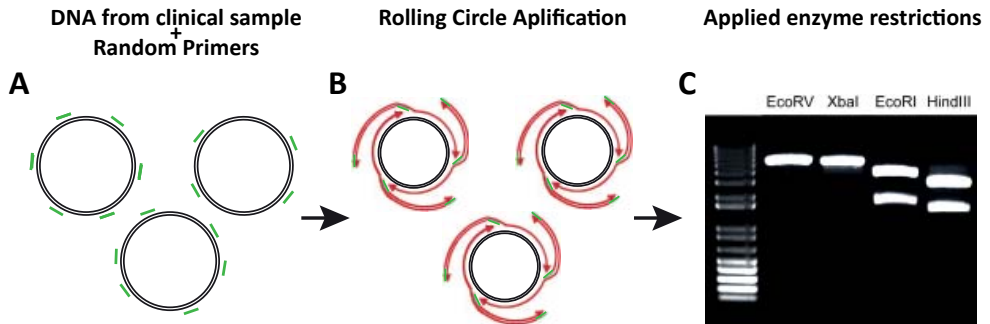
## **Merkel cell polyomavirus (MCPyV)**

The discovery of MCPyV proved the assumptions of polyomavirus involvement in human oncogenesis finally to be true. Merkel cell carcinomas (MCC) are a rare but aggressive type of cutaneous tumors of neuroendocrine origin that typically develop in the elderly and immunocompromized patients (see paragraph “The new human polyomaviruses and their associated diseases”). Using a technique called the digital transcriptome subtraction [43], MCPyV was identified in 8 out of 10 analyzed MCC lesions [9]. This technique compares (subtracts) cDNA libraries generated with random RT-PCR from diseased and healthy tissues and thus can identify messengerRNA transcripts potentially unique for either or both. In this case, with the help of sophisticated software, unique and unknown sequences were followed-up, of which some displayed homology to the monkey B-lymphotrophic polyomavirus (LPyV) and BKPyV. Through primer-genome-walking, subsequently the complete genome was sequenced. New to HPyVs, but known for oncogenic animal polyomaviruses detected in tumors, the MCPyV genome was found to be monoclonally randomly integrated into the genome of the analyzed MCC lesions [9, 44].

## **Human polyomavirus 6 and 7 (HPyV6 and HPyV7)**

In order to detect free, non-integrated, MCPyV genomes on the skin, rolling-circle amplification (RCA) was performed on forehead swab samples of healthy individuals. Performing the RCA technique on samples results into amplification of circular double-stranded viral DNA

by using random primers and strand-displacement DNA polymerase (**Figure 3**) [45]. With some adjustments to boost the RCA sensitivity, this approach revealed the identity of a new polyomavirus, called the type 6 (HPyV6), on the skin of a healthy individual [14]. By using a newly developed degenerate PCR primer-set on the similar pre-treated DNA materials from those healthy skin samples to detect HPyV6- and WUPyV-like sequences, an additional new polyomavirus sequence was identified and completely sequenced, which was subsequently named HPyV7 [14].



**Figure 3.** The rolling-circle amplification (RCA) technique exploits random hexamer primers and a DNA polymerase (Phi-29) with proofreading and strand-displacement capabilities used on clinical DNA samples (A). If a sample contains a relative excess of a particular small circular double-stranded DNA molecule, for instance a plasmid or a (small) viral genome, RCA will produce long, high-molecular DNA molecules with concatenated (linearized) stretches of DNA of potential interest (B). These stretches will appear as sharp bands on gel when the RCA-product is cut with specific restriction enzymes that can subsequently be cloned and sequenced (C). Figure adapted from [10, 45].

### Trichodysplasia spinulosa-associated polyomavirus (TSPyV)

Trichodysplasia spinulosa (TS) is a rare skin disease of immunocompromized patients, which will be described in more detail below and in **Chapter 2**. Already in 1999, Haycox and co-workers proposed that TS has a viral origin, because TS lesions contained clusters of small virus particles [46]. It was only in 2010 that the identity of the virus was revealed [10], with the help of RCA technique (**Figure 3**) [45]. With specific primers used during primer-genome-walking the entire RCA product was sequenced, which revealed the entire TSPyV genome in lesions of a TS patient [10].

### Human polyomavirus 9 (HPyV9)

Previously, several serological studies suggested the presence of ‘cross-reactive’ antibodies in a percentage of human sera that recognized a non-human primate polyomavirus (LPyV isolated from an African green monkey) capsid-protein [47 - 49]. Furthermore, LPyV-derived sequences were obtained from human blood, indicating the occurrence of LPyV(-like) infections in humans [50]. In 2011, a novel polyomavirus that closely resembled LPyV was identified, independently, by two research groups in human serum and skin swabs, respectively,

and tentatively called HPyV9. One group used a degenerate polyomavirus PCR primer-set [26] to screen serum samples, which generated an unknown, putative HPyV sequence, next to some known HPyV sequences. With primer-genome-walking, the rest of the HPyV9 genome was sequenced [15]. The other group isolated DNA from healthy skin swabs and subjected the DNA to high-throughput sequencing, which resulted in millions of reads and several contigs that matched MCPyV, HPyV6 and HPyV7, as well as many human papillomaviruses. Some of the contigs showed the highest homology to LPyV and after filling the contig sequence gaps by designing additional primers, HPyV9 genome was identified [51].

### **Malawi, Mexico and human polyomavirus 10 (MWPyV/MXPyV/HPyV10)**

Another novel human polyomavirus was identified in stool of a child from Malawi [16]. This virus was discovered by suspending a small amount of feces that subsequently was centrifuged and filtered through 0.45 mm and then 0.22 mm pores. After chloroform and DNase treatment, extraction-free-DNA was obtained. The solution that still contained encapsidated DNA was SDS and Proteinase K-treated, and DNA was isolated. Finally, this DNA was subjected to RCA and pyrosequencing. Hundreds of reads were obtained, sequenced and aligned to sequences from the GenBank database. This revealed several polyomavirus-like reads and with the use of primer-genome-walking strategy the complete MWPyV genome sequence was fulfilled [16].

Shortly after, two almost identical viruses were identified, called the HPyV10 and Mexico polyomavirus (MXPyV). HPyV10 genome was isolated from an anal wart of an immunocompromized patient with WHIM syndrome with the help of RCA and primer-genome-walking sequencing [17]. MXPyV was identified in diarrheal stool collected from a child in Mexico [18]. From stool, viral particles were purified in a PBS suspension, containing glass beads and chloroform, by mechanical shaking and centrifugation. Further processing of the filtered supernatant was similar as for MWPyV genome identification described above. Sequencing of the barcoded reads was performed with an Illumina HiSeq 2000. Assembled contigs from 75bp reads were used to design primers directed outward for use in long-range PCR to amplify and sequence the whole genome. Because MWPyV, HPyV10 and MXPyV share at least 95% of their genome sequence, they most probably represent the same polyomavirus species.

### **St. Louis polyomavirus (STLPyV)**

As part of a human gut-microbiome-survey study of healthy and malnourished children from Malawi performed by the same group that identified the WUPyV and MWPyV, another new polyomavirus was discovered and called STLPyV [19]. Prior to DNA extraction, the stool was processed by CsCl ultracentrifugation to generically enrich for viral particles. After DNA was extracted and subsequently amplified by a technique called multiple displacement amplification (MDA), shotgun 454 pyrosequencing was performed. MDA technique is comparable to the RCA method exploiting random hexamer primers and a DNA polymerase (Phi-29) with

proofreading and strand-displacement capabilities (**Figure 3**). After sequencing, two reads were identified in this sample, which shared limited sequence identity to known polyomaviruses. Employing these two initial reads, primers were designed to PCR amplify products that span the whole genome in either direction. STLPyV has a mosaic genome that probably shares an ancestral recombinant origin with the MWPyV because of its 64.2% genome similarity to this polyomavirus.

### Human polyomavirus 12 (HPyV12)

Very recently, a twelfth human polyomavirus was isolated (HPyV12) [20]. HPyV12 genome was tested positive in 11% of liver tissues analyzed in that study. Methods used for genome identification and sequencing were identical/comparable to the identification of HPyV9 [15]. HPyV12 is distantly related to other viruses from the *Orthopolyomavirus-I* genogroup (see paragraph “Phylogeny and evolutionary trends”). Whether this virus is involved in pathogenesis of a liver disease, or other gastrointestinal diseases, remains to be seen.

### New Jersey polyomavirus (NJPyV)

The most recently identified human polyomavirus, called the New Jersey polyomavirus (NJPyV), was isolated from an immunosuppressed patient diagnosed with vasculitic myopathy and retinopathy caused by microthrombosis. In tested biopsies, clusters of viral particles were observed in swollen nuclei of endothelial cells [11]. The authors identified the viral genome after extraction of RNA from sections of formalin-fixed paraffin-embedded muscle tissue, from which they generate cDNA libraries that were subsequently used for sequencing. The obtained sequence reads were all uninformative by nucleotide sequence analysis (BLASTn), but when translated amino acid analysis (BLASTx) was performed, they found approximately 80% sequence homology with chimpanzee polyomaviruses. Future studies should investigate whether this virus has indeed a tropism for vascular endothelial cells and may be involved in vasculitis in general. Interestingly, the patient had to evacuate through floodwaters because of superstorm Sandy weeks prior to development of her symptoms, prompting speculation about NJPyV contagion through sewage-contaminated waters.

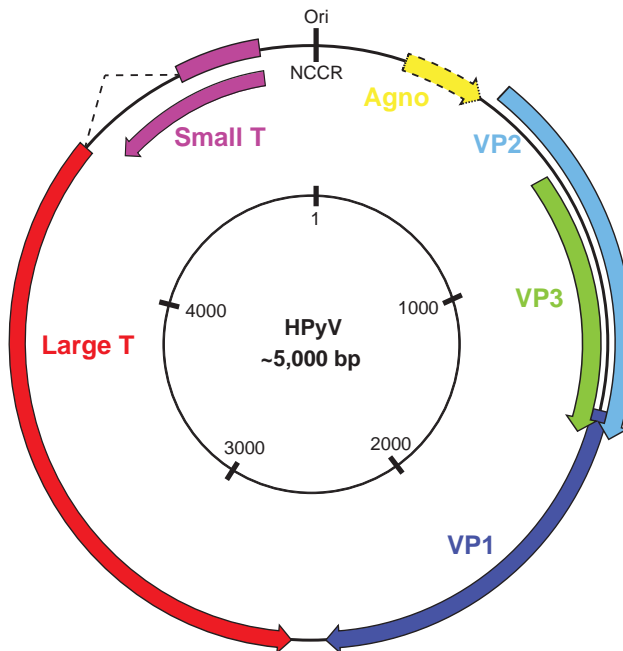
Altogether, since 2007 an explosion in human polyomavirus discoveries took place, among which the discovery of TSPyV is evident. These discoveries are most likely due to improvement in viral DNA purification and amplification methods and particularly the use of high-throughput Next Generation Sequencing to identify viral DNA with low copy-numbers in the background of excess host genomic DNA. Identification of additional viruses of the *Polyomaviridae* family and studies of their genome, epidemiology, pathogenicity and evolution will undoubtedly lead to a better understanding of the role these viruses play in human health and disease.

## The polyomavirus genome and its products

### Genome organization and transcription

Viruses of the *Polyomaviridae* family, including the recently identified human viruses, are among the smallest known to infect humans, both in particle size and in genome length [52]. The non-enveloped virions of about 45-nm harbour a circular double-stranded DNA genome of approximately 5000 basepairs that is divided into three regions; (i) the non-coding control region (NCCR) and in opposing directions, (ii) the early (T-antigen) and (iii) the late (VP) coding regions (**Figure 4**).

The NCCR contains the origin of replication flanked by several large T-antigen-binding sites, transcription promoters and regulatory elements. The novel human polyomaviruses display a similar NCCR organization. In JCPyV- or BKPyV-diseased individuals, the NCCR of virus isolates found in different body compartments can vary from what is called the ‘archetype’ NCCR that is present in transmissible virus found in urine for example. It is believed that the rearranged NCCR increases the T-antigen transcription and replication rate, as was shown for instance for JCPyV [53]. Systematic comparisons of NCCR regions among different isolates found within diseased and healthy individuals have not been reported (yet) for the novel human polyomaviruses.



**Figure 4.** Old schematic representation of human polyomavirus circular double-stranded DNA genome organization as proposed before, showing only the protein-expressing layer. Updated representations, that include middle T-antigen and ALTO protein, are shown in **Chapter 2** and **Chapter 6**. Regions coding for the indicated proteins are depicted in different colors. The agnoprotein (yellow) is encoded only by some HPyVs.

The T-antigen coding region of most HPyVs encodes at least the small and large T-antigens that are translated from an alternatively spliced pre-messengerRNA. In addition, shorter or alternatively spliced products of T-antigens are expressed, such as 17kT, T4 and 57kT [54, 55]. Rodent polyomaviruses are known to express an additional T-antigen, called middle T-antigen [56]. The VP coding region found on the opposite strand codes for at least the VP1 and VP2 structural capsid proteins, and often VP3. The agnoprotein and VP4 expressed by JCPyV, BKPyV and several animal polyomaviruses, including birds, seem not to be encoded by the newly identified HPyVs [57], but for most of these viruses the transcription-patterns have not been solved yet.

Besides the NCCR transcription regulation, for example shown for SV40, T-antigen messengerRNAs are post-transcriptionally regulated by microRNAs [58]. In addition to SV40, also for MCPyV and BKPyV a microRNA was identified that potentially regulates early gene expression [59, 60].

## T-antigens

Polyomavirus T-antigens play essential roles in viral transcription and replication, as well as in host-cell tuning to enable efficient virus replication [52, 61]. As polyomaviruses are completely dependent on the host cell replication machinery, a significant part of this tuning is aimed at inducing S-phase and bypassing cell cycle-control measures. In some cases, these features can derail and result in uncontrolled cell proliferation and eventually tumor formation. So far, MCPyV is the only (newly identified) HPyV for which these phenomena have been noted [62]. Consequently, the T-antigens of MCPyV have been studied already quite extensively (see next, and “MCPyV and MCC” paragraph), whereas for the other new HPyVs this knowledge is almost entirely based on predicted motifs in protein sequences.

The small and large T-antigens share their N-terminus, while having C-terminal regions of different sizes encoded from different reading frames (**Figure 4**). The novel human polyomaviruses seem no exception to this rule. This shared region occupied by the so-called J-domain contains important motifs, such as CR1 and DnaJ, for viral replication and cellular transformation [61]. Downstream, the small T-antigen contains a PP2A subunit-binding motif probably involved in activation of the Akt-mTOR pathway related to (tumor) cell survival [63, 64]. The part of large T-antigen, encoded by the second exon, includes many functionally important domains or motifs such as: (i) a nuclear-localization signal, (ii) an origin-binding domain, (iii) an ATPase-containing helicase domain [61, 65], (iv) a pRB-binding and (v) sometimes a p53-interaction domain that both potentially play a crucial role in cellular transformation. Apart from MCPyV [55], the presence of (most of) these motifs and domains was identified only by bioinformatics approach in TSPyV and HPyV9 [10, 15], but they remain to be characterized functionally.

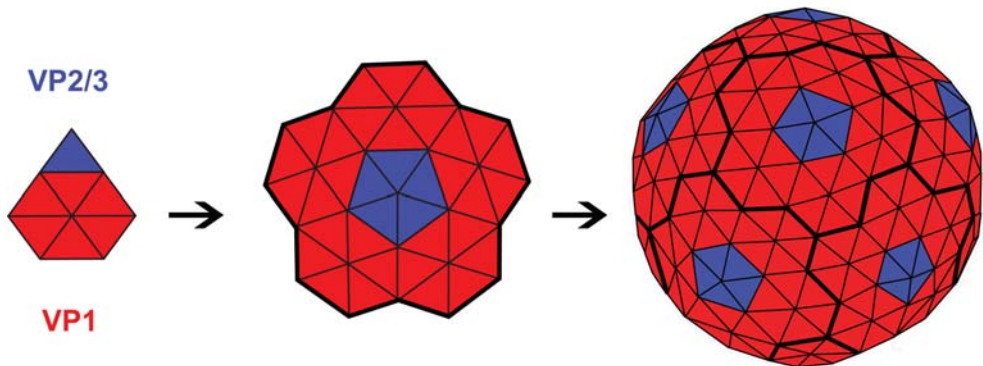
Murine (MPyV) and hamster (HaPyV) polyomaviruses express also a middle T-antigen, which serves as the prime oncoprotein during cellular transformation [56]. Middle T-antigen has a C-terminal transmembrane domain, several Tyr, Ser and Thr phosphoryla-

tion sites, and some other functionally unknown conserved domains. Phosphorylation of the Tyr, Ser and Thr residues of middle T-antigen results into binding and activation of several cellular signal transduction proteins (e.g., Src, PI3K, PLC-gamma1) that promotes cellular transformation and oncogenesis [56, 66 - 68].

### VP capsid proteins

Together, SV40 encoded late VP structural proteins make up the approximately 45-nm polyomavirus capsid [69]. VP1 molecules interact with VP2 or VP3 molecules to auto-assemble into a full capsid-complex (**Figure 5**) [52]. Comparable VP1, VP2 and/or VP3 encoding sequences have been suggested for the novel polyomaviruses. So far, there is no reason to believe that the novel viruses will behave differently with respect to capsid formation, although there is debate whether MCPyV VP3 is expressed or is functional [70, 71]. The VP4 protein, noted for SV40, is expressed very late in infection and serves probably as a porin of cellular membrane that facilitates the viral release [72, 73]. So far, this protein has not been predicted or reported for any of the newly identified HPyVs.

How the polyomaviruses recognize and infect their host target cells is partly known. Most of the known polyomaviruses interact by linkage of the major capsid protein VP1 with terminal sialic acids (alpha2,3 or alpha2,6) of mainly gangliosides on the host cell membrane [74], after which they are internalized using different entry pathways. For MCPyV, it was shown that this virus uses glycosaminoglycans for the initial attachment, while the sialic acids may be important later in the encounter [75]. Future research could reveal whether the other novel human polyomaviruses use similar strategies for host cell recognition.



**Figure 5.** Illustration of polyomavirus icosahedral capsid formation. Six VP1 proteins interact with either a VP2 or VP3 to assemble into a hexameric protomere. Five of these protomeres interact to form a capsomere, which subsequently auto-assemble into a full capsid-complex [76].

## The new human polyomaviruses and their associated diseases

For two of the novel HPyVs a causal relationship with a specific clinical condition is highly probable. In both cases, the disease/cancer has lent its name to the virus, trichodysplasia spinulosa to TSPyV and Merkel cell carcinoma to MCPyV. For MCPyV, an additional association with chronic lymphocytic leukemia that originates from B-cells was suggested, and for TSPyV an association with pilomatricoma that represents a benign hair follicular skin tumor was speculated. However, evidence for these correlations is (yet) weak. For the recently identified NJPyV associated with the vasculatory myositis, retinitis and dermatitis, the evidence is highly suggestive, but limited to one patient observation so far. For the other new human polyomaviruses, the evidence for involvement in disease is too weak at the moment.

### MCPyV and MCC

MCC is a rare and aggressive neuroendocrine tumor of epidermal origin with an increasing incidence of approximately 0.2 - 0.5 per 100,000 depending on age, skin type, (cumulative) sun exposure and immune status (**Figure 6A**) [77, 78]. The association with fair skin and sun exposure suggests a direct effect of UV-radiation in MCC development, but this has not been confirmed experimentally. In 2008, the presence of MCPyV in MCC lesions was shown for the first time [9], a finding that has been confirmed by many research groups worldwide showing that at least 75% of all MCCs have monoclonally integrated MCPyV genome [65].

Several findings indicate direct involvement of MCPyV in Merkel cell carcinogenesis. For instance, in almost all cancer cases MCPyV is found linearized and monoclonally integrated in the host tumor-cell genome, although the site of integration appears random [9, 44, 79]. Furthermore, the MCPyV large T-gene was found to carry 'signature' mutations specific for the carcinomas. Their expression results into truncated large T-antigen that lacks the C-terminal helicase domain, the putative p53-interaction domain and the origin-binding domain [9, 55, 80], consequently rendering the virus inactive for replication. Therefore, the function of large T-antigen in viral progeny production is selectively impaired. Small T-antigen is unaffected by these mutations. Therefore, in an apparent parallel to other virus-associated cancers, e.g. human papillomavirus (HPV)-associated cervical cancer, also in MCPyV-positive MCC there seems to be a strong association with non-productive MCPyV infection. Probably the strongest evidence of pathogenicity (oncogenicity) of MCPyV is provided by studies showing that expression of both small and (truncated) large T-antigens are indispensable in the maintenance of cancer cell phenotype [81, 82].

### TSPyV and trichodysplasia spinulosa

Trichodysplasia spinulosa (TS) is a rare skin disease exclusively found in severely immunocompromized hosts, especially the solid organ transplant recipients (**Figure 6B**) [85, 86]. TS is characterized by gradual development of papules and spicules (spines) on the face, sometimes accompanied by alopecia of the eyebrows and lashes [87]. Worldwide, some 30 cases have been described so far [88, 89]. In 1999, Haycox and co-workers proposed for the





**Figure 6.** (A) Presentation of a 5-cm large Merkel cell carcinoma with concomitant nodal spread and highly vascularized surface on the left cheek of a 72-year-old man [83]. (B) Presentation of trichodysplasia spinulosa showing eyebrow alopecia, spiny spicule eruptions and multiple inflammatory papules on the face of a 5-year-old immunocompromized girl [84].

first time that trichodysplasia spinulosa concerns a viral-associated disease and showed the presence of intranuclear clusters of honeycomb-arranged viral particles in the distended and dysmorphic hair follicles [46]. Subsequent reports confirmed the typical histological findings of hyperplastic hair bulbs and the presence of polyomavirus or papillomavirus-like particles [90 - 94]. Attempts to identify the nature of this virus failed until TSPyV was identified with the help of RCA in 2010 [10]. More detail about TSPyV's virological and pathogenic properties, as well as epidemiologic, diagnostic, and therapeutic aspects will be presented in **Chapter 2**.

## Prevalence, persistence, reactivation and spread of polyomavirus infection

### Seroprevalence

Cross-sectional, serological studies of JCPyV and BKPyV have shown that both infections are highly prevalent in the general population. The seroprevalence of JCPyV is approximately 60% and of BKPyV approximately 90% [47]. Serological studies of the novel human polyomaviruses revealed a considerable variation in the presence of serum antibodies [14, 47, 95 - 101]. The lowest seroprevalence was detected for HPyV9 varying between 15% and 45% [100, 102, 103], the highest for WUPyV, MCPyV, TSPyV and HPyV6 varying between 60% and 95% [14, 47, 97, 98, 104]. In all serological studies, the strongest increase in seroprevalence is observed at a young age, indicating that primary infections generally occur at (early) childhood [47, 96, 98, 99]. So far, the seroprevalences of the novel HPyVs measured with VP1-VLPs and GST-VP1 fusion proteins appear similar, as was shown for instance for TSPyV [97, 98].

## Persistence and reactivation

It is generally believed that primary polyomavirus infection is followed by a persistent, asymptomatic (latent) infection with very low levels of viral replication, which could remain hidden in the body lifelong. For JCPyV and BKPyV reactivation from the latent state and manifestation of symptomatic infection is observed only in the case of reduced immunity. For instance, nephropathy can develop in long-term immunosuppressed solid organ transplant recipients caused by BKPyV, and PML in AIDS patients and Natalizumab-treated multiple sclerosis patients caused by JCPyV [52, 105, 106].

For most polyomaviruses, in healthy individuals the detection rate of viral DNA is much lower than their seroprevalence. In urine of healthy blood donors for example, the detection rate of JCPyV and BKPyV DNA was approximately 3- and 10-fold lower than the seroprevalence measured within the same group, respectively [107]. In immunocompromised patients, however, BKPyV can be detected in urine at very high loads indicative of massive reactivation of infection and shedding of the virus [108, 109].

For the novel HPyVs, the association between diminished immunity and reactivation is less clear. Only for TSPyV, with a seroprevalence of approximately 75% [97, 98, 103], clinical consequences are seen exclusively in severely immunocompromised hosts (see also **Chapter 2**) [85, 86]. However, the TSPyV DNA detection rate, on the skin at least, hardly varies among immunocompetents (2%) and immunocompromised hosts (4%) [10, 85]. Also for MCPyV, the DNA detection rates on the skin do not differ much between immunocompetents and immunocompromised, albeit they are generally much higher compared to TSPyV and resemble the seroprevalence of about 50% [110]. In general, it should be kept in mind that both MCC and TS are very rare conditions implying that (host) factors other than immunity may also play a role in controlling these infections. For KIPyV, WUPyV HPyV6, HPyV7, HPyV9, MWPyV, HPyV10, MXPyV, STLPyV, HPyV12 and NJPyV infections, the relation with immunity is not (yet) solved. This is largely explained by weak disease associations or the lack of an established clinical condition that could be attributed to one of these novel viruses and the unawareness of the organ where these viruses could latently persist.

## Spread of infection

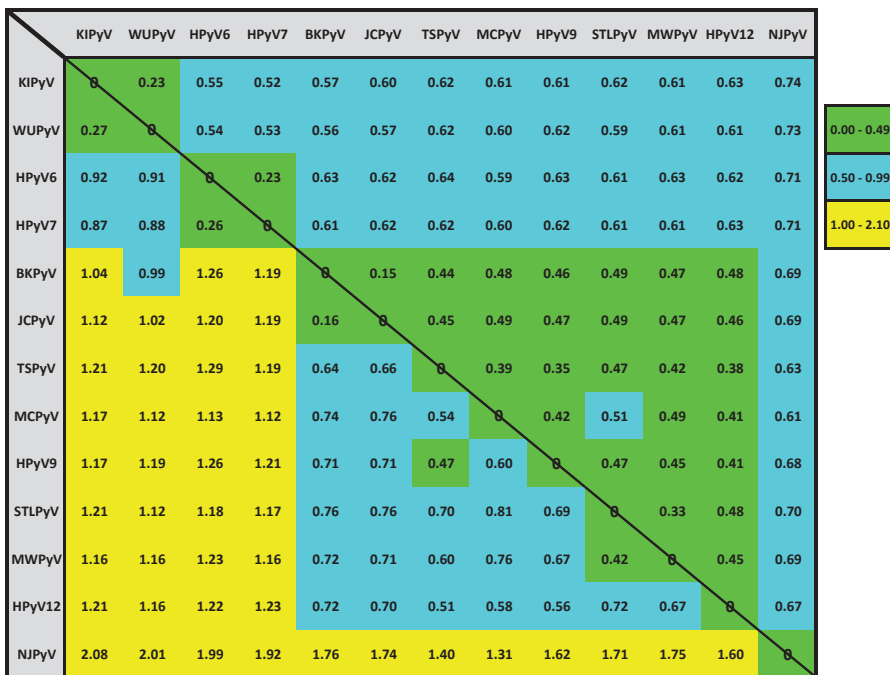
Earlier studies have shown that JCPyV and BKPyV are found in urine, feces and saliva. The excreted virus numbers are so high that these polyomaviruses even serve as markers of sewage pollution or recreational waters [111]. Which route of virus excretion primarily drives human infection is not precisely known, but the efficiency of infection seems clear from the high seroprevalence, for both the known and novel human polyomaviruses.

Regarding possible routes of excretion and transmission of cutaneous polyomaviruses, a recent study showed the presence of MCPyV DNA on 75% of samples from environmental surfaces (door handles, ticket machines etc.). Prior DNase treatment of the samples and subsequent viral load measurement suggested that about 5% of the detected MCPyV DNA was protected, most probably encapsidated, and therefore potentially infectious [112]. If this provides a solid clue regarding the virus transmission, remains to be seen.

## Phylogeny and evolutionary trends

So far, the thirteen known HPyVs comprise the largest host-specific subset of identified polyomaviruses. This probably reflects a bias of the virus discovery efforts that increasingly focus on hunting for human viruses, particularly during the past several years as was detailed above and shown in **Figure 2**. Pairwise divergence between HPyVs varies as much as in the entire polyomavirus family. If measured for the combined conserved regions of VP1, VP2 and LT proteins it ranges from 15% (for BKPyV/JCPyV pair) to 74% (KIPyV/NJPyV), or, if repeated substitutions accounted, from 0.16 (BKPyV/JCPyV) to 2.08 (KIPyV/NJPyV) substitutions per position on average (**Figure 7**). Apart from KIPyV/WUPyV, also the HPyV6/HPyV7 and BKPyV/JCPyV pairs are among the most closely related, which correlates with putative shared tropism for the respiratory tract, skin and the urinary tract infection, respectively.

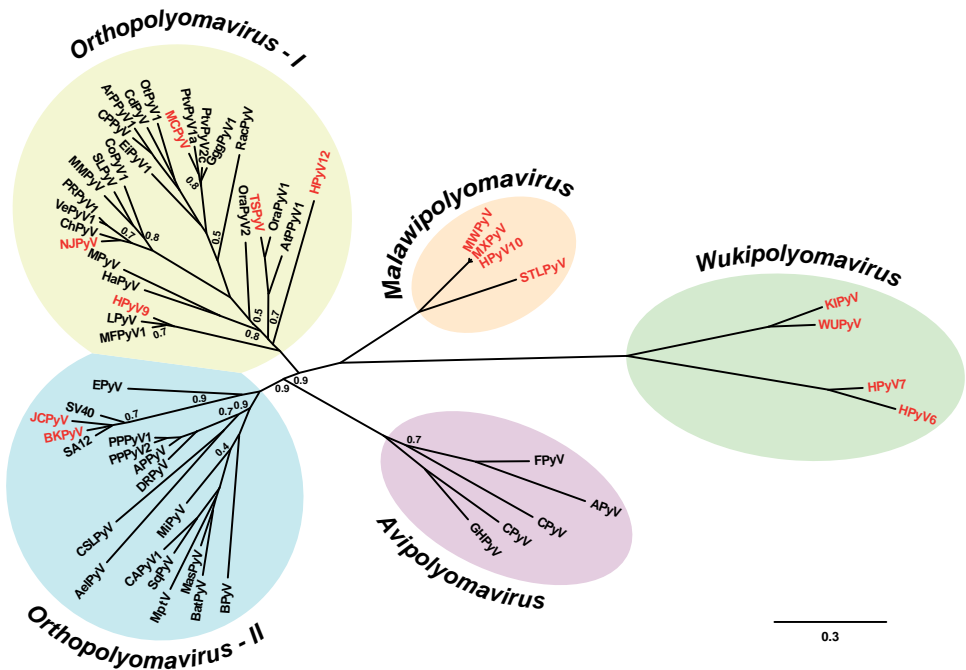
In **Figure 8**, a tentative polyomavirus phylogenetic tree is illustrated. In addition to the older tree shown in **Figure 1** [2], the novel tree includes several newly identified animal and human viruses that represent new putative polyomavirus species. From this tree, it is



**Figure 7.** Genetic comparison between thirteen human polyomaviruses is shown. Two measures of genetic distance are shown for the combined polyomavirus protein-identity of conserved regions of VP1, VP2 and LT proteins. The upper-right triangle shows uncorrected dissimilarity percentages (1 corresponds to 100% dissimilarity; 0 means identical sequences); the lower-left triangle shows evolutionary distances that correct for multiple substitutions at the same site (the scale is mean substitutions per sequence position) under the WAG amino acid substitution model and rate heterogeneity among sites [113] (C. Lauber is acknowledged for updating the figure).

clear that the HPyVs do not form a monophyletic cluster. Rather, they are distributed unevenly among four out of five distant lineages. Two lineages that include human polyomaviruses belong to the genus *Orthopolyomavirus* [2], and are called *Orthopolyomavirus-I* and *-II*, respectively. The third lineage belongs to the genus *Wukupolyomavirus*, and the fourth comprising the genus *Malawipolyomavirus* includes MWPyV (as well as HPyV10 and MXPpyV that are almost identical to MWPyV) and STLPyV all isolated from human stool.

The *Malawipolyomavirus* and *Wukupolyomavirus* lineages are the least populated and include exclusively HPyVs. In contrast, the *Orthopolyomavirus-I* and *-II* include 45 putative species from which five and two, respectively, are composed of HPyVs. Because of the observed phyletic distribution of HPyVs among different lineages, it may be envisaged that a number of human polyomaviruses emerged because of cross-species jumps of zoonotic viruses rather than of virus-host coevolution [114, 115]. Recently, the applicability of this virus-host coevolution model to two species of best-sampled polyomaviruses, BKPyV and JCPyV [116 - 118], as well as the entire family [119] was questioned.



**Figure 8.** Unrooted phylogenetic tree (up until June 2014) of known (putative) polyomavirus species based on the alignment of concatenated VP1, VP2 and LT amino acid sequences. The obtained branching pattern (topology) of basal nodes in the tree matches that proposed by Johne and colleagues [2], shown in **Figure 1**. HPyVs are shown in red. The bar indicates number of substitutions per site. The numbers at branching events represent probability support values ranging from 0 (no support) to 1 (best support). Only support values lower than 1 are shown. The full polyomavirus names can be found in **Table 1** (C. Lauber is acknowledged for updating the figure).

Resolving the uncertainty about the origins of human polyomaviruses will require extending the virus discovery efforts. For example, continuation of hunting for human viruses and characterization of other hosts should lead to considerable improvement in our understanding of polyomaviruses (host) evolution. Moreover, additional polyomavirus studies regarding their gene expression and overprinting mechanisms will facilitate understanding of polyomavirus-specific patterns of speciation and host adaptation.

## **Concluding remarks**

With the availability of sensitive molecular techniques to detect unknown genome sequences in the background of human genomic DNA, the discoveries of polyomaviruses have increased significantly over the past few years. It is expected that these discoveries will continue in the coming years.



## Outline and Scope of this dissertation

In **Chapter 1**, the focus was on the recent developments in studying the newly identified human polyomaviruses, which reviewed several general aspects of virus identification, pathogenicity, epidemiology and complete *Polyomaviridae* family phylogeny. Still, detailed analyses of these aspects for each virus are desperately needed. This dissertation will address some of these issues in particular for TSPyV.

In **Chapter 2**, the virological and clinical aspects of TS disease are discussed in more detail, including epidemiologic, diagnostic, and therapeutic aspects of TSPyV infection.

To study causality between TSPyV infection and TS disease, in **Chapter 3** the prevalence, load and localization of TSPyV infection in a series of TS patients is analyzed in comparison to healthy controls.

In **Chapter 4**, it is investigated which cellular mechanisms are potentially disrupted by TSPyV, thereby contributing to hyperproliferation of infected hair follicle cells and spicule formation. Furthermore, the putative role of TSPyV Large T-antigen in this process is studied.

In **Chapter 5**, we (re)assessed *in-silico* the genome sequences of all available (recently identified) polyomaviruses known so far. A newly identified ORF encoding MT/ALTO-antigen is characterized in more detail, and its role in polyomavirus evolution and adaptation is addressed.

In **Chapter 6**, whole genome sequences of TSPyV are obtained and analyzed from a series of TS cases described in Chapter 3, for the purpose of gaining insight into the evolution of this polyomavirus. The phenomena regarding polyomavirus evolution described in Chapter 5 are employed on TSPyV evolution that acts as a unique *in-silico* model to study the molecular basis of amino-acid exchange and toggling in polyomavirus host adaptation in general.

Finally, in **Chapter 7**, the findings described in this dissertation will be discussed with regard to TSPyV infection, pathogenesis, evolution and host adaptation, and compared to the existing knowledge about polyomaviruses in a broader context. In addition, some thoughts about future research directions will be shared in this chapter.



## References

1. Norkin LC, Allander T, Atwood WJ, Buck CB, Garcea RL, Imperiale MJ, Johne R, Major EO, Pipas JM, Ramqvist T. Family Polyomaviridae. In: Virus Taxonomy, Ninth Report of the International Committee on Taxonomy of Viruses (2012). King AMQ, Adams MJ, Carstens EB, Lefkowitz EJ (editors). London, UK: Academic Press: 279-290.
2. Johne R, Buck CB, Allander T, Atwood WJ, Garcea RL, Imperiale MJ, Major EO, Ramqvist T, Norkin LC. Taxonomical developments in the family Polyomaviridae (2011) *Arch. Virol.* 156: 1627-1634.
3. Eddy BE, Borman GS, Grubbs GE, Young RD. Identification of the oncogenic substance in rhesus monkey kidney cell culture as simian virus 40 (1962) *Virology* 17: 65-75.
4. Girardi AJ, Sweet BH, Slotnick VB, Hilleman MR. Development of tumors in hamsters inoculated in the neonatal period with vacuolating virus, SV-40 (1962) *Proc. Soc. Exp. Biol. Med.* 109: 649-660.
5. Gross L. A filterable agent, recovered from Ak leukemic extracts, causing salivary gland carcinomas in C3H mice (1953) *Proc. Soc. Exp. Biol. Med.* 83: 414-421.
6. Padgett BL, Walker DL, ZuRhein GM, Eckroade RJ, Dessel BH. Cultivation of papova-like virus from human brain with progressive multifocal leucoencephalopathy (1971) *Lancet* 1: 1257-1260.
7. Gardner SD, Field AM, Coleman DV, Hulme B. New human papovavirus (B.K.) isolated from urine after renal transplantation (1971) *Lancet* 1: 1253-1257.
8. Strickler HD, Rosenberg PS, Devesa SS, Hertel J, Fraumeni JF, Jr., Goedert JJ. Contamination of poliovirus vaccines with simian virus 40 (1955-1963) and subsequent cancer rates (1998) *JAMA* 279: 292-295.
9. Feng H, Shuda M, Chang Y, Moore PS. Clonal integration of a polyomavirus in human Merkel cell carcinoma (2008) *Science* 319: 1096-1100.
10. van der Meijden E, Janssens RW, Lauber C, Bouwes Bavinck JN, Gorbalenya AE, Feltkamp MC. Discovery of a new human polyomavirus associated with trichodysplasia spinulosa in an immunocompromized patient (2010) *PLoS Pathog.* 6: e1001024.
11. Mishra N, Pereira M, Rhodes RH, An P, Pipas JM, Jain K, Kapoor A, Briese T, Faust PL, Lipkin WI. Identification of a Novel Polyomavirus in a Pancreatic Transplant Recipient With Retinal Blindness and Vasculitic Myopathy (2014) *J. Infect. Dis.* 15: 1595-1599.
12. Allander T, Andreasson K, Gupta S, Bjerkner A, Bogdanovic G, Persson MA, Dalianis T, Ramqvist T, Andersson B. Identification of a third human polyomavirus (2007) *J. Virol.* 81: 4130-4136.
13. Gaynor AM, Nissen MD, Whiley DM, Mackay IM, Lambert SB, Wu G, Brennan DC, Storch GA, Sloots TP, Wang D. Identification of a novel polyomavirus from patients with acute respiratory tract infections (2007) *PLoS Pathog.* 3: e64.

14. Schowalter RM, Pastrana DV, Pumphrey KA, Moyer AL, Buck CB. Merkel cell polyomavirus and two previously unknown polyomaviruses are chronically shed from human skin (2010) *Cell Host Microbe* 7: 509-515.
15. Scuda N, Hofmann J, Calvignac-Spencer S, Ruprecht K, Liman P, Kuhn J, Hengel H, Ehlers B. A novel human polyomavirus closely related to the african green monkey-derived lymphotropic polyomavirus (2011) *J. Virol.* 85: 4586-4590.
16. Siebrasse EA, Reyes A, Lim ES, Zhao G, Mkakosya RS, Manary MJ, Gordon JI, Wang D. Identification of MW Polyomavirus, a Novel Polyomavirus in Human Stool (2012) *J. Virol.* 86: 10321-10326.
17. Buck CB, Phan GQ, Raiji MT, Murphy PM, McDermott DH, McBride AA. Complete genome sequence of a tenth human polyomavirus (2012) *J. Virol.* 86: 10887.
18. Yu G, Greninger AL, Isa P, Phan TG, Martinez MA, de la Luz SM, Contreras JF, Santos-Preciado JI, Parsonnet J, Miller S, Derisi JL, Delwart E, Arias CF, Chiu CY. Discovery of a novel polyomavirus in acute diarrheal samples from children (2012) *PLoS One* 7: e49449.
19. Lim ES, Reyes A, Antonio M, Saha D, Ikumapayi UN, Adeyemi M, Stine OC, Skelton R, Brennan DC, Mkakosya RS, Manary MJ, Gordon JI, Wang D. Discovery of STL polyomavirus, a polyomavirus of ancestral recombinant origin that encodes a unique T antigen by alternative splicing (2013) *Virology* 436: 295-303.
20. Korup S, Rietscher J, Calvignac-Spencer S, Trusch F, Hofmann J, Moens U, Sauer I, Voigt S, Schmuck R, Ehlers B. Identification of a novel human polyomavirus in organs of the gastrointestinal tract (2013) *PLoS One* 8: e58021.
21. Malherbe H, HARWIN R. The cytopathic effects of vervet monkey viruses (1963) *S. Afr. Med. J.* 37: 407-411.
22. zur Hausen H, Gissmann L. Lymphotropic papovaviruses isolated from African green monkey and human cells (1979) *Med. Microbiol. Immunol.* 167: 137-153.
23. Verschoor EJ, Groenewoud MJ, Fagrouch Z, Kewalapat A, Van Gessel S, Kik MJ, Heeney JL. Molecular characterization of the first polyomavirus from a New World primate: squirrel monkey polyomavirus (2008) *J. Gen. Virol.* 89: 130-137.
24. Groenewoud MJ, Fagrouch Z, Van Gessel S, Niphuis H, Bulavaite A, Warren KS, Heeney JL, Verschoor EJ. Characterization of novel polyomaviruses from Bornean and Sumatran orang-utans (2010) *J. Gen. Virol.* 91: 653-658.
25. Deuzing I, Fagrouch Z, Groenewoud MJ, Niphuis H, Kondova I, Bogers W, Verschoor EJ. Detection and characterization of two chimpanzee polyomavirus genotypes from different subspecies (2010) *Virol. J.* 7: 1-7.
26. Leendertz FH, Scuda N, Cameron KN, Kidega T, Zuberbuhler K, Leendertz SA, Couacy-Hymann E, Boesch C, Calvignac S, Ehlers B. African great apes are naturally infected with polyomaviruses closely related to Merkel cell polyomavirus (2011) *J. Virol.* 85: 916-924.



27. Scuda N, Madinda NF, Akoua-Koffi C, Adjogoua EV, Wevers D, Hofmann J, Cameron KN, Leendertz SA, Couacy-Hymann E, Robbins M, Boesch C, Jarvis MA, Moens U, Mugisha L, Calvignac-Spencer S, Leendertz FH, Ehlers B. Novel polyomaviruses of nonhuman primates: genetic and serological predictors for the existence of multiple unknown polyomaviruses within the human population (2013) *PLoS Pathog.* 9: e1003429.
28. Kilham L, Murphy HW. A pneumotropic virus isolated from C3H mice carrying the Bittner Milk Agent (1953) *Proc. Soc. Exp. Biol. Med.* 82: 133-137.
29. Graffi A, Schramm T, Bender E, Bierwolf D, Graffi I. On a new virus containing skin tumor in golden hamster (1967) *Arch. Geschwulstforsch.* 30: 277-283.
30. Reissig M, Kelly TJ, Jr., Daniel RW, Rangan SR, Shah KV. Identification of the stump-tailed macaque virus as a new papovavirus (1976) *Infect. Immun.* 14: 225-231.
31. Misra V, Dumonceaux T, Dubois J, Willis C, Nadin-Davis S, Severini A, Wandeler A, Lindsay R, Artsob H. Detection of polyoma and corona viruses in bats of Canada (2009) *J. Gen. Virol.* 90: 2015-2022.
32. Colegrove KM, Wellehan JF, Jr., Rivera R, Moore PF, Gulland FM, Lowenstine LJ, Nordhausen RW, Nollens HH. Polyomavirus infection in a free-ranging California sea lion (*Zalophus californianus*) with intestinal T-cell lymphoma (2010) *J. Vet. Diagn. Invest.* 22: 628-632.
33. Orba Y, Kobayashi S, Nakamura I, Ishii A, Hang'ombe BM, Mweene AS, Thomas Y, Kimura T, Sawa H. Detection and characterization of a novel polyomavirus in wild rodents (2011) *J. Gen. Virol.* 92: 789-795.
34. Renshaw RW, Wise AG, Maes RK, Dubovi EJ. Complete genome sequence of a polyomavirus isolated from horses (2012) *J. Virol.* 86: 8903.
35. Fagrouch Z, Sarwari R, Lavergne A, Delaval M, de TB, Lacoste V, Verschoor EJ. Novel polyomaviruses in South American bats and their relationship to other members of the family Polyomaviridae (2012) *J. Gen. Virol.* 93: 2652-2657.
36. Tao Y, Shi M, Conrardy C, Kuzmin IV, Recuenco S, Agwanda B, Alvarez DA, Ellison JA, Gilbert AT, Moran D, Niezgodna M, Lindblade KA, Holmes EC, Breiman RF, Rupprecht CE, Tong S. Discovery of diverse polyomaviruses in bats and the evolutionary history of the Polyomaviridae (2013) *J. Gen. Virol.* 94: 738-748.
37. Dela Cruz FNJ, Giannitti F, Li L, Woods LW, Del VL, Delwart E, Pesavento PA. Novel polyomavirus associated with brain tumors in free-ranging raccoons, western United States (2013) *Emerg. Infect. Dis.* 19: 77-84.
38. Stevens H, Bertelsen MF, Sijmons S, Van RM, Maes P. Characterization of a novel polyomavirus isolated from a fibroma on the trunk of an African elephant (*Loxodonta africana*) (2013) *PLoS One* 8: e77884.
39. Bozeman LH, Davis RB, Gaudry D, Lukert PD, Fletcher OJ, Dykstra MJ. Characterization of a papovavirus isolated from fledgling budgerigars (1981) *Avian Dis.* 25: 972-980.

40. Guerin JL, Gelfi J, Dubois L, Vuillaume A, Boucraut-Baralon C, Pingret JL. A novel polyomavirus (goose hemorrhagic polyomavirus) is the agent of hemorrhagic nephritis enteritis of geese (2000) *J. Virol.* 74: 4523-4529.
41. Johne R, Wittig W, Fernandez-de-Luco D, Hofle U, Muller H. Characterization of Two Novel Polyomaviruses of Birds by Using Multiply Primed Rolling-Circle Amplification of Their Genomes (2006) *J. Virol.* 80: 3523-3531.
42. Halami MY, Dorrestein GM, Couteel P, Heckel G, Muller H, Johne R. Whole-genome characterization of a novel polyomavirus detected in fatally diseased canary birds (2010) *J. Gen. Virol.* 91: 3016-3022.
43. Feng H, Taylor JL, Benos PV, Newton R, Waddell K, Lucas SB, Chang Y, Moore PS. Human transcriptome subtraction by using short sequence tags to search for tumor viruses in conjunctival carcinoma (2007) *J. Virol.* 81: 11332-11340.
44. Martel-Jantin C, Filippone C, Cassar O, Peter M, Tomasic G, Vielh P, Briere J, Petrella T, Aubriot-Lorton MH, Mortier L, Jouvion G, Sastre-Garau X, Robert C, Gessain A. Genetic variability and integration of Merkel cell polyomavirus in Merkel cell carcinoma (2012) *Virology* 426: 134-142.
45. Johne R, Muller H, Rector A, van Ranst M, Stevens H. Rolling-circle amplification of viral DNA genomes using phi29 polymerase (2009) *Trends Microbiol.* 17: 205-211.
46. Haycox CL, Kim S, Fleckman P, Smith LT, Piepkorn M, Sundberg JP, Howell DN, Miller SE. Trichodysplasia spinulosa--a newly described folliculocentric viral infection in an immunocompromised host (1999) *J. Investig. Dermatol. Symp. Proc.* 4: 268-271.
47. Kean JM, Rao S, Wang M, Garcea RL. Seroepidemiology of Human Polyomaviruses (2009) *PLoS Pathog.* 5: e1000363.
48. Viscidi RP, Clayman B. Serological cross reactivity between polyomavirus capsids (2006) *Adv. Exp. Med. Biol.* 577: 73-84.
49. Pawlita M, Clad A, zur HH. Complete DNA sequence of lymphotropic papovavirus: prototype of a new species of the polyomavirus genus (1985) *Virology* 143: 196-211.
50. Delbue S, Tremolada S, Branchetti E, Elia F, Gualco E, Marchioni E, Maserati R, Ferrante P. First identification and molecular characterization of lymphotropic polyomavirus in peripheral blood from patients with leukoencephalopathies (2008) *J. Clin. Microbiol.* 46: 2461-2462.
51. Sauvage V, Foulongne V, Cheval J, Ar Gouilh M, Pariente K, Dereure O, Manuguerra JC, Richardson J, Lecuit M, Burguiere A, Caro V, Eloit M. Human polyomavirus related to african green monkey lymphotropic polyomavirus (2011) *Emerg. Infect. Dis.* 17: 1364-1370.
52. Imperiale MJ, Major EO. Polyomaviruses. In: *Fields VIROLOGY* (2007). Knipe DM, Howley PM (editors). Philadelphia: Wolters Kluwer / Lippincott Williams & Wilkins: 2263-2299.

53. Gosert R, Kardas P, Major EO, Hirsch HH. Rearranged JC virus noncoding control regions found in progressive multifocal leukoencephalopathy patient samples increase virus early gene expression and replication rate (2010) *J. Virol.* 84: 10448-10456.
54. Zerrahn J, Knippschild U, Winkler T, Deppert W. Independent expression of the transforming amino-terminal domain of SV40 large T antigen from an alternatively spliced third SV40 early mRNA (1993) *EMBO J.* 12: 4739-4746.
55. Shuda M, Feng H, Kwun HJ, Rosen ST, Gjoerup O, Moore PS, Chang Y. T antigen mutations are a human tumor-specific signature for Merkel cell polyomavirus (2008) *Proc. Natl. Acad. Sci. U. S. A.* 105: 16272-16277.
56. Fluck MM, Schaffhausen BS. Lessons in signaling and tumorigenesis from polyomavirus middle T antigen (2009) *Microbiol. Mol. Biol. Rev.* 73: 542-563.
57. Johne R, Muller H. Polyomaviruses of birds: etiologic agents of inflammatory diseases in a tumor virus family (2007) *J. Virol.* 81: 11554-11559.
58. Sullivan CS, Grundhoff AT, Tevethia S, Pipas JM, Ganem D. SV40-encoded microRNAs regulate viral gene expression and reduce susceptibility to cytotoxic T cells (2005) *Nature* 435: 682-686.
59. Seo GJ, Chen CJ, Sullivan CS. Merkel cell polyomavirus encodes a microRNA with the ability to autoregulate viral gene expression (2009) *Virology* 383: 183-187.
60. Broekema NM, Imperiale MJ. miRNA regulation of BK polyomavirus replication during early infection (2013) *Proc. Natl. Acad. Sci. U. S. A.* 110: 8200-8205.
61. Sullivan CS, Pipas JM. T Antigens of Simian Virus 40: Molecular Chaperones for Viral Replication and Tumorigenesis (2002) *Microbiol. Mol. Biol. Rev.* 66: 179-202.
62. Chang Y, Moore PS. Merkel cell carcinoma: a virus-induced human cancer (2012) *Annu. Rev. Pathol.* 7: 123-144.
63. Buchkovich NJ, Yu Y, Zampieri CA, Alwine JC. The TORrid affairs of viruses: effects of mammalian DNA viruses on the PI3K-Akt-mTOR signalling pathway (2008) *Nat. Rev. Microbiol.* 6: 266-275.
64. Andrabi S, Hwang JH, Choe JK, Roberts TM, Schaffhausen BS. Comparisons between murine polyomavirus and Simian virus 40 show significant differences in small T antigen function (2011) *J. Virol.* 85: 10649-10658.
65. Arora R, Chang Y, Moore PS. MCV and Merkel cell carcinoma: a molecular success story (2012) *Curr. Opin. Virol.* 2: 489-498.
66. Whitman M, Kaplan DR, Schaffhausen B, Cantley L, Roberts TM. Association of phosphatidylinositol kinase activity with polyoma middle-T competent for transformation (1985) *Nature* 315: 239-242.
67. Cheng SH, Harvey R, Espino PC, Semba K, Yamamoto T, Toyoshima K, Smith AE. Peptide antibodies to the human c-fyn gene product demonstrate pp59c-fyn is capable of complex formation with the middle-T antigen of polyomavirus (1988) *EMBO J.* 7: 3845-3855.

68. Su W, Liu W, Schaffhausen BS, Roberts TM. Association of Polyomavirus middle tumor antigen with phospholipase C-gamma 1 (1995) *J. Biol. Chem.* 270: 12331-12334.
69. Baker TS, Drak J, Bina M. The capsid of small papova viruses contains 72 pentameric capsomeres: direct evidence from cryo-electron-microscopy of simian virus 40 (1989) *Biophys. J.* 55: 243-253.
70. Pastrana DV, Tolstov YL, Becker JC, Moore PS, Chang Y, Buck CB. Quantitation of Human Seroresponsiveness to Merkel Cell Polyomavirus (2009) *PLoS Pathog.* 5: e1000578.
71. Schowalter RM, Buck CB. The Merkel cell polyomavirus minor capsid protein (2013) *PLoS Pathog.* 9: e1003558.
72. Daniels R, Sadowicz D, Hebert DN. A very late viral protein triggers the lytic release of SV40 (2007) *PLoS Pathog.* 3: e98.
73. Raghava S, Giorda KM, Romano FB, Heuck AP, Hebert DN. The SV40 late protein VP4 is a viroporin that forms pores to disrupt membranes for viral release (2011) *PLoS Pathog.* 7: e1002116.
74. Neu U, Stehle T, Atwood WJ. The Polyomaviridae: Contributions of virus structure to our understanding of virus receptors and infectious entry (2009) *Virology* 384: 389-399.
75. Schowalter RM, Pastrana DV, Buck CB. Glycosaminoglycans and sialylated glycans sequentially facilitate Merkel cell polyomavirus infectious entry (2011) *PLoS Pathog.* 7: e1002161.
76. Belnap DM, Olson NH, Cladel NM, Newcomb WW, Brown JC, Kreider JW, Christensen ND, Baker TS. Conserved features in papillomavirus and polyomavirus capsids (1996) *J. Mol. Biol.* 259: 249-263.
77. Schrama D, Ugurel S, Becker JC. Merkel cell carcinoma: recent insights and new treatment options (2012) *Curr. Opin. Oncol.* 24: 141-149.
78. Van KA, Mascre G, Youseff KK, Harel I, Michaux C, De GN, Szpalski C, Achouri Y, Bloch W, Hassan BA, Blanpain C. Epidermal progenitors give rise to Merkel cells during embryonic development and adult homeostasis (2009) *J. Cell Biol.* 187: 91-100.
79. Laude HC, Jonchere B, Maubec E, Carlotti A, Marinho E, Couturaud B, Peter M, Sastre-Garau X, Avril MF, Dupin N, Rozenberg F. Distinct merkel cell polyomavirus molecular features in tumour and non tumour specimens from patients with merkel cell carcinoma (2010) *PLoS Pathog.* 6: e1001076.
80. Sastre-Garau X, Peter M, Avril MF, Laude H, Couturier J, Rozenberg F, Almeida A, Boitier F, Carlotti A, Couturaud B, Dupin N. Merkel cell carcinoma of the skin: pathological and molecular evidence for a causative role of MCV in oncogenesis (2009) *J. Pathol.* 218: 48-56.
81. Houben R, Shuda M, Weinkam R, Schrama D, Feng H, Chang Y, Moore PS, Becker JC. Merkel cell polyomavirus-infected Merkel cell carcinoma cells require expression of viral T antigens (2010) *J. Virol.* 84: 7064-7072.

82. Shuda M, Kwun HJ, Feng H, Chang Y, Moore PS. Human Merkel cell polyomavirus small T antigen is an oncoprotein targeting the 4E-BP1 translation regulator (2011) *J. Clin. Invest.* 121: 3623-3634.
83. Jouary T, Lalanne N, Siberchicot F, Ricard AS, Versapuech J, Lepreux S, Delaunay M, Taieb A. Neoadjuvant polychemotherapy in locally advanced Merkel cell carcinoma (2009) *Nat. Rev. Clin. Oncol.* 6: 544-548.
84. Schwieger-Briel A, Balma-Mena A, Ngan B, Dipchand A, Pope E. Trichodysplasia spinulosa--a rare complication in immunosuppressed patients (2010) *Pediatr. Dermatol.* 27: 509-513.
85. Kazem S, van der Meijden E, Kooijman S, Rosenberg AS, Hughey LC, Browning JC, Sadler G, Busam K, Pope E, Benoit T, Fleckman P, de VE, Eekhof JA, Feltkamp MC. Trichodysplasia spinulosa is characterized by active polyomavirus infection (2012) *J. Clin. Virol.* 53: 225-230.
86. Matthews MR, Wang RC, Reddick RL, Saldivar VA, Browning JC. Viral-associated trichodysplasia spinulosa: a case with electron microscopic and molecular detection of the trichodysplasia spinulosa-associated human polyomavirus (2011) *J. Cutan. Pathol.* 38: 420-431.
87. Wanat KA, Holler PD, Dentchev T, Simbiri K, Robertson E, Seykora JT, Rosenbach M. Viral-associated trichodysplasia: characterization of a novel polyomavirus infection with therapeutic insights (2012) *Arch. Dermatol.* 148: 219-223.
88. Chastain MA, Millikan LE. Pilomatrix dysplasia in an immunosuppressed patient (2000) *J. Am. Acad. Dermatol.* 43: 118-122.
89. Izakovic J, Buchner SA, Duggelin M, Guggenheim R, Itin PH. [Hair-like hyperkeratoses in patients with kidney transplants. A new cyclosporin side-effect] (1995) *Hautarzt* 46: 841-846.
90. Lee JS, Frederiksen P, Kossard S. Progressive trichodysplasia spinulosa in a patient with chronic lymphocytic leukaemia in remission (2008) *Australas. J. Dermatol.* 49: 57-60.
91. Osswald SS, Kulick KB, Tomaszewski MM, Sperling LC. Viral-associated trichodysplasia in a patient with lymphoma: a case report and review (2007) *J. Cutan. Pathol.* 34: 721-725.
92. Sadler GM, Halbert AR, Smith N, Rogers M. Trichodysplasia spinulosa associated with chemotherapy for acute lymphocytic leukaemia (2007) *Australas. J. Dermatol.* 48: 110-114.
93. Sperling LC, Tomaszewski MM, Thomas DA. Viral-associated trichodysplasia in patients who are immunocompromised (2004) *J. Am. Acad. Dermatol.* 50: 318-322.
94. Wyatt AJ, Sachs DL, Shia J, Delgado R, Busam KJ. Virus-associated trichodysplasia spinulosa (2005) *Am. J. Surg. Pathol.* 29: 241-246.

95. Carter JJ, Paulson KG, Wipf GC, Miranda D, Madeleine MM, Johnson LG, Lemos BD, Lee S, Warcola AH, Iyer JG, Nghiem P, Galloway DA. Association of Merkel Cell Polyomavirus-Specific Antibodies With Merkel Cell Carcinoma (2009) *J. Natl. Cancer Inst.* 4: 1510-1522.
96. Chen T, Hedman L, Mattila PS, Jartti T, Ruuskanen O, Soderlund-Venermo M, Hedman K. Serological evidence of Merkel cell polyomavirus primary infections in childhood (2011) *J. Clin. Virol.* 50: 125-129.
97. Chen T, Mattila PS, Jartti T, Ruuskanen O, Soderlund-Venermo M, Hedman K. Seroepidemiology of the Newly Found Trichodysplasia Spinulosa-Associated Polyomavirus (2011) *J. Infect. Dis.* 204: 1523-1526.
98. van der Meijden E, Kazem S, Burgers MM, Janssens R, Bouwes Bavinck JN, de Melker H, Feltkamp MC. Seroprevalence of Trichodysplasia Spinulosa-associated Polyomavirus (2011) *Emerg. Infect. Dis.* 17: 1355-1363.
99. Viscidi RP, Rollison DE, Sondak VK, Silver B, Messina JL, Giuliano AR, Fulp W, Ajidahun A, Rivanera D. Age-specific seroprevalence of Merkel cell polyomavirus, BK virus, and JC virus (2011) *Clin. Vaccine Immunol.* 18: 1737-1743.
100. Trusch F, Klein M, Finsterbusch T, Kuhn J, Hofmann J, Ehlers B. Seroprevalence of human polyomavirus 9 and cross-reactivity to African green monkey-derived lymphotropic polyomavirus (2012) *J. Gen. Virol.* 93: 698-705.
101. Touze A, Gaitan J, Arnold F, Casal R, Fleury MJ, Combelas N, Sizaret PY, Guyétant S, Maruani A, Baay M, Tognon M, Coursaget P. Generation of Merkel cell polyomavirus (MCV)-like particles and their application to detection of MCV antibodies (2010) *J. Clin. Microbiol.* 48: 1767-1770.
102. Nicol JT, Touze A, Robinot R, Arnold F, Mazzoni E, Tognon M, Coursaget P. Seroprevalence and Cross-reactivity of Human Polyomavirus 9 (2012) *Emerg. Infect. Dis.* 18: 1329-1332.
103. van der Meijden E, Bialasiewicz S, Rockett RJ, Tozer SJ, Sloots TP, Feltkamp MC. Different serologic behavior of MCPyV, TSPyV, HPyV6, HPyV7 and HPyV9 polyomaviruses found on the skin (2013) *PLoS One* 8: e81078.
104. Nguyen NL, Le BM, Wang D. Serologic evidence of frequent human infection with WU and KI polyomaviruses (2009) *Emerg. Infect. Dis.* 15: 1199-1205.
105. Bloomgren G, Richman S, Hotermans C, Subramanyam M, Goelz S, Natarajan A, Lee S, Plavina T, Scanlon JV, Sandrock A, Bozic C. Risk of natalizumab-associated progressive multifocal leukoencephalopathy (2012) *N. Engl. J. Med.* 366: 1870-1880.
106. Major EO. Progressive multifocal leukoencephalopathy in patients on immunomodulatory therapies (2010) *Annu. Rev. Med.* 61: 35-47.
107. Egli A, Infanti L, Dumoulin A, Buser A, Samaridis J, Stebler C, Gosert R, Hirsch HH. Prevalence of polyomavirus BK and JC infection and replication in 400 healthy blood donors (2009) *J. Infect. Dis.* 199: 837-846.

108. Bennett SM, Broekema NM, Imperiale MJ. BK polyomavirus: emerging pathogen (2012) *Microbes Infect.* 14: 672-683.
109. Hirsch HH, Randhawa P. BK virus in solid organ transplant recipients (2009) *Am. J. Transplant.* 9 Suppl 4: S136-S146.
110. Wieland U, Silling S, Scola N, Potthoff A, Gambichler T, Brockmeyer NH, Pfister H, Kreuter A. Merkel cell polyomavirus infection in HIV-positive men (2011) *Arch. Dermatol.* 147: 401-406.
111. Korajkic A, Brownell MJ, Harwood VJ. Investigation of human sewage pollution and pathogen analysis at Florida Gulf coast beaches (2011) *J. Appl. Microbiol.* 110: 174-183.
112. Foulongne V, Cournaud V, Champeau W, Segondy M. Detection of Merkel cell polyomavirus on environmental surfaces (2011) *J. Med. Virol.* 83: 1435-1439.
113. Whelan S, Goldman N. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach (2001) *Mol. Biol. Evol.* 18: 691-699.
114. Shadan FF, Villarreal LP. The evolution of small DNA viruses of eukaryotes: past and present considerations (1995) *Virus Genes* 11: 239-257.
115. Perez-Losada M, Christensen RG, McClellan DA, Adams BJ, Viscidi RP, Demma JC, Crandall KA. Comparing Phylogenetic Codivergence between Polyomaviruses and Their Hosts (2006) *J. Virol.* 80: 5663-5669.
116. Shackelton LA, Rambaut A, Pybus OG, Holmes EC. JC virus evolution and its association with human populations (2006) *J. Virol.* 80: 9928-9933.
117. Krumbholz A, Bininda-Emonds OR, Wutzler P, Zell R. Evolution of four BK virus subtypes (2008) *Infect. Genet. Evol.* 8: 632-643.
118. Zheng HY, Nishimoto Y, Chen Q, Hasegawa M, Zhong S, Ikegaya H, Ohno N, Sugimoto C, Takasaka T, Kitamura T, Yogo Y. Relationships between BK virus lineages and human populations (2007) *Microbes Infect.* 9: 204-213.
119. Krumbholz A, Bininda-Emonds OR, Wutzler P, Zell R. Phylogenetics, evolution, and medical importance of polyomaviruses (2009) *Infect. Genet. Evol.* 9: 784-799.





VePyV1 CaPyV JCPyV *mPyV* SA12 RacPyV GHPyV BKPyV

*BatsyV* CPyV OrapyV1 FPyV MptV APPyV1 SqPyV LPyV CSLPyV

EPyV PRPyV1 APP<sub>g</sub>V2 KIPyV **MIRgV** PtvPyV2c OtPyV1 **STLLPyV**

MFPyV1 *KSyV09* SV40 TSPyV CoPyV1 PPPyV CPPyV HPyV12

MXPyV PtvPyV1a PDPyV EIPyV1 AtPPyV1 HaPyV (TggPyV1

CdPyV DRPyV MWPyV APyV CaPyV1 HPyV7 ChPyV MasPyV

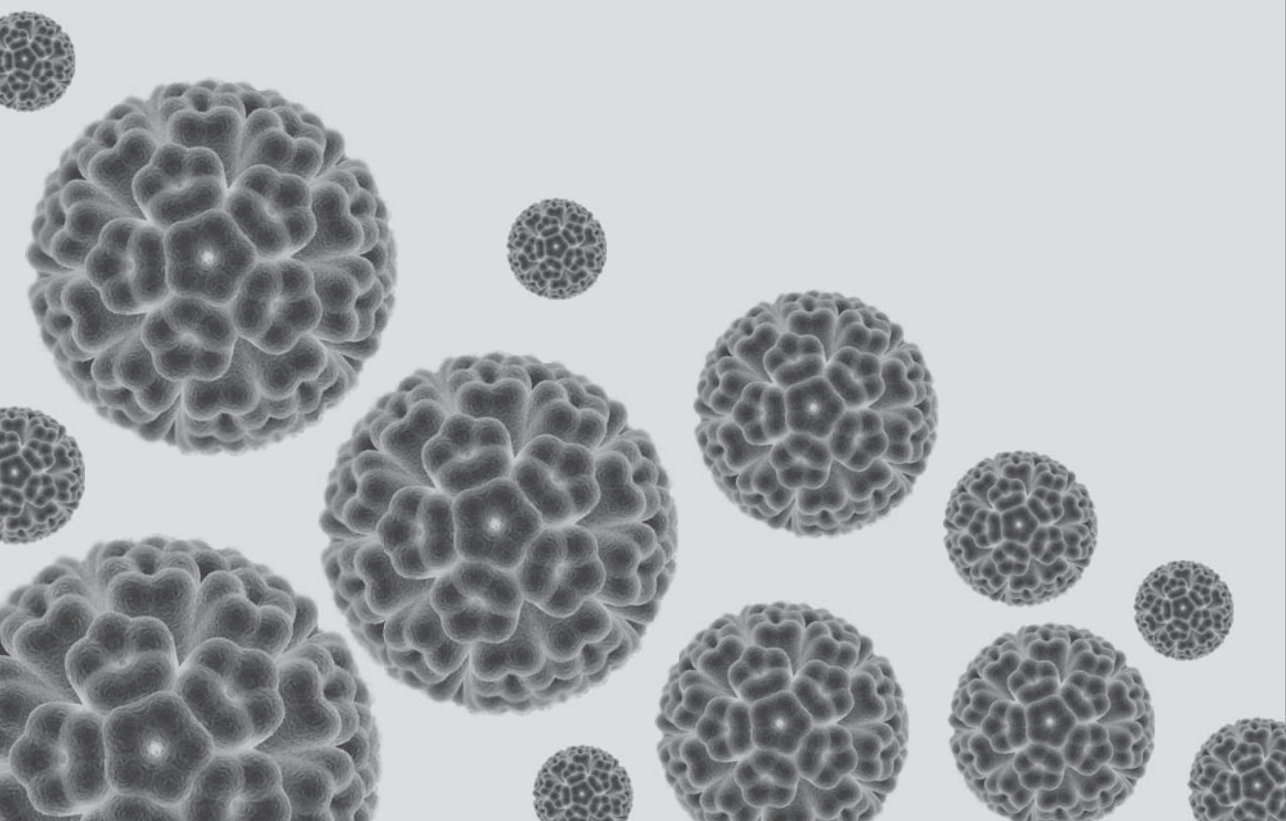
WUPyV *HSPyVc* BPyV MCPyV OrapyV2 MMPyV SLPyV HPyV10

# Part II

## TSPyV Infection and Pathogenesis

**VePyV1** *CaPyV* **JCPyV** *mPyV* **SA12**  
*BatPyV* **CPyV** *OraPyV1* **FPyV** **MptV** *APPy*  
**EPyV** **PRPyV1** *APPyV2* **KIPyV** **MiPyV** **IPyV**  
**MFPyV1** *HPyV9* **SV40** **TSPyV** **CoPyV1**  
**MPyV** *PtvPyV1a* **PDPyV** **EiPyV1** **AtPyV**  
**CdPyV** **DRPyV** *MWPyV* **APyV** **CAPyV1**  
**WUPyV** *HPyV6* **BPyV** **MCPyV** **OraPyV2**

**PyV** **BKPyV**  
**V** **CSLPyV**  
**V1** **STLPyV**  
**V** **HPyV12**  
**V** **GgPyV1**  
**V** **MasPyV**  
**V** **HPyV10**



# Chapter 2

## Clinical and viral aspects of trichodysplasia spinulosa

*Adapted from\*:*

### **The trichodysplasia spinulosa-associated polyomavirus: virological background and clinical implications**

*Authors:*

Siamaque Kazem<sup>1</sup>

Els van der Meijden<sup>1</sup>

Mariet Feltkamp<sup>1</sup>

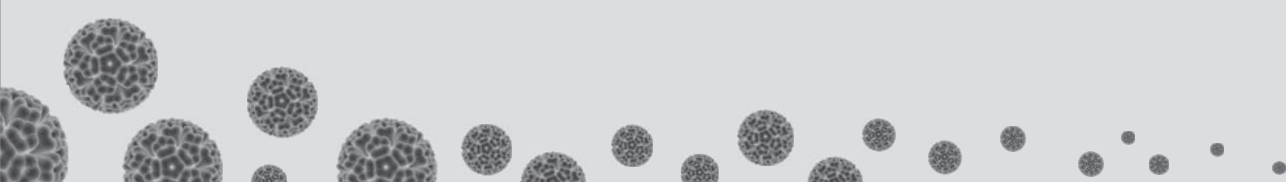
*Affiliations:*

<sup>1</sup> Department of Medical Microbiology, Leiden University Medical Center, Leiden, The Netherlands

*Published in original form:*

Acta pathologica, microbiologica et immunologica Scandinavica, 2013 (121): 770 - 782

*\* Note: Adaptation of this chapter from the original published article concerns textual, tabular and figure adjustments. The text has been updated with novel TSPyV data from the literature until mid-2014. The tables were reviewed and updated until mid-2014. In figure 2, the TSPyV genome map is updated with the newly identified ORF5 that encodes ALTO/MT-antigen.*



## Abstract

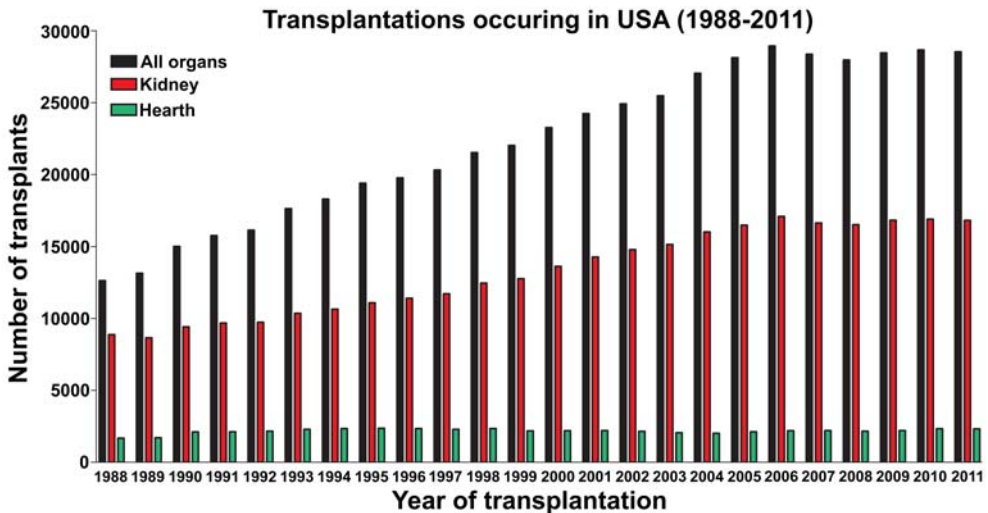
Trichodysplasia spinulosa-associated polyomavirus (TSPyV) is one of the new species of the *Polyomaviridae* family discovered in 2010. TSPyV infects humans and is associated with the development of a rare skin disease called trichodysplasia spinulosa. Trichodysplasia spinulosa is a disease of severely immunocompromized hosts characterized by follicular distention and keratotic spine formation, most notably on the face. Electron microscopy, immunohistochemistry, and viral load measurements suggest an etiological role of active TSPyV infection in the development of this disfiguring disease. This chapter will address some clinical and virological properties of TSPyV, and touches upon epidemiologic, diagnostic, and therapeutic aspects of TSPyV infection.

Abstract

## Pathogenic human polyomaviruses

**H**uman polyomaviruses (HPyVs) are ubiquitous viruses that infect their host without causing apparent disease. After primary infection, they can persist asymptomatically, sometimes producing small quantities of detectable progeny. This state is often referred to as latency. In case of impaired immune function, for example because of AIDS or the use of immunosuppressive drugs, HPyVs can reactivate from latency and cause severe disease. As the number of solid organ transplantations gradually increases, as shown for instance by the United States Organ Procurement and Transplantation network (**Figure 1**) [1], the incidence of HPyV-associated disease in long-term immunosuppressed patients is expected to rise.

So far, at least four out of thirteen HPyVs have been associated with human disease. As pointed out in **Chapter 1**, this includes BKPyV (allograft nephropathy [2, 3]), JCPyV (progressive multifocal leukoencephalopathy [4 - 6]), MCPyV (Merkel cell carcinoma [7 - 10]) and TSPyV. TSPyV, identified in our lab in 2010, is associated with a rare skin disease called trichodysplasia spinulosa (TS) [11]. This disease is observed exclusively in severely immunocompromized hosts [12, 13]. In this chapter, several clinical and virological aspects of TS and TSPyV are presented in more detail.



**Figure 1.** Overview of (solid) organ transplantations in USA between 1988 and 2011. Bars depicted in black represent all solid organ transplantation, i.e., kidney, liver, heart, lung, pancreas, intestine, kidney+pancreas and heart+lung. Red colored bars represent the number of kidney transplantations and green bars the number of heart transplantations (source OPTN, data as of September 2012).

## The trichodysplasia spinulosa-associated polyomavirus

### TSPyV genome and gene products

TSPyV genome consists of a relatively small double-stranded DNA of 5232 basepairs. Its genomic orientation is similar to other known polyomaviruses, with its early T-antigen and late VP products encoded bidirectionally on separate DNA strands. At least five open reading frames can be located, transcription of which is regulated by the non-coding control region (NCCR) and is subject of alternative splicing (**Figure 2**) [11].

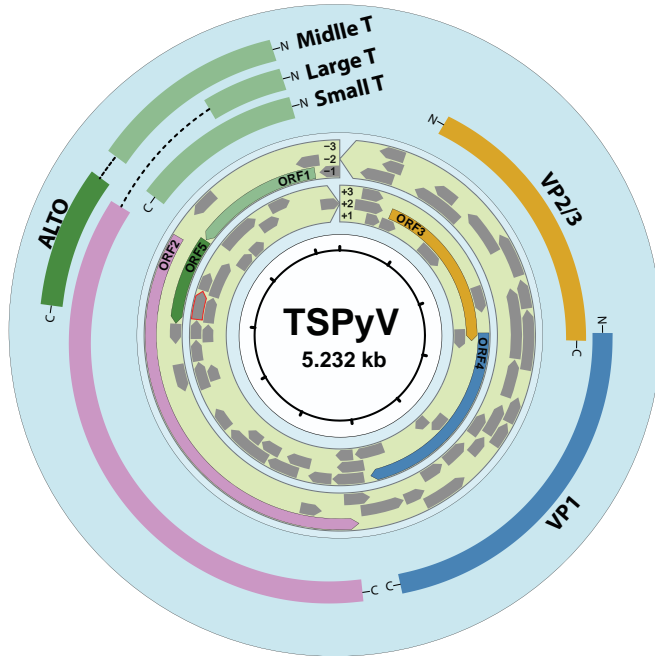
In analogy with other polyomaviruses, upon infection probably first the T genes are expressed that code for small (ST) and large T-antigen (LT), and possibly for middle T-antigen (MT) and ALTO as well (van der Meijden, unpublished results). ST-, MT- and LT-antigen initiate at the same start codon, but are a subject of alternative splicing downstream. As a result, ST-, MT- and LT-antigen share N-terminal region of approximately 80 amino acids which contain some highly conserved regions important for polyomavirus replication and cellular transformation (**Figure 2**) [11]. Furthermore, ST-antigen contains a unique putative PP2A-binding motif, MT-antigen a membrane-spanning domain and several putative Tyr, Ser and Thr phosphorylation sites, and LT-antigen a number of motifs that putatively interact with tumor suppressor and cell cycle regulatory proteins like pRB and p53. Preliminary analyses of messengerRNA and protein expression of the T-region confirmed this pattern for TSPyV that includes ST-, MT-, LT-antigen expression including an additional ALTO product (van der Meijden, unpublished results). The function of TSPyV T-antigens, individually and in concert, related to virus replication and possibly cellular transformation, remains to be studied.

The opposing strand of the TSPyV genome encodes the late structural proteins VP1, VP2, and possibly VP3 that together form the viral capsid. VP1 is the major protein of the pentameric viral capsomere that likely incorporates one copy of either VP2 or VP3 [14]. The icosahedral capsid is constructed of 72 of such capsomers. Antigenetically, VP1 is considered the immunodominant polyomavirus protein and therefore TSPyV VP1 represents an important antigen in measuring host antibody seroreactivity against TSPyV.

From the late pre-messengerRNA, complementary to the early coding region, microRNAs can be encoded, which can regulate gene expression by directing cleavage of the targeted T-antigen messengerRNAs and thereby repressing translation. For BKPyV, this mechanism was recently experimentally identified [15]. For MCPyV this mechanism was postulated as well [16, 17]. Preliminary bioinformatics data suggest that TSPyV may code as well for a microRNA that targets the transcripts of both LT-antigen and MT-antigen/ALTO (**Figure 2**).

### TSPyV prevalence

For BKPyV and JCPyV it is suggested that the primary infection takes place sometime early in life and transmission probably occurs through the feco-oral, urino-oral, or respiratory route



**Figure 2.** TSPyV genome map. The five large open reading frames (ORFs) in different reading frames are depicted by colored inner arrows (ORF1-ORF5). Additional smaller ORFs are depicted by gray inner arrows. Transcription of the early region results into a pre-messengerRNA that contains ORF1, ORF2 and ORF5. Upon alternative splicing events (introns indicated by dotted lines), small T-antigen (light-green), large T-antigen (light-green + pink) and middle T-antigen (light-green + dark-green) translation of the indicated protein will be putatively initiated at the same start codon and all proteins will share the N-terminal region (light-green). Alternatively, internal start codons could be used in RF-1 that might result into translation of ALTO product(s) (dark green). The late coding region contains ORF3 and ORF5 that encode VP1, VP2, and possibly VP3 viral capsid proteins. Putatively, a microRNA that is complementary to the early coding region targeting ORF2 and ORF4, is transcribed from a late pre-messengerRNA (red-lined arrow). The region before the early and late ORFs contains the non-coding control region and the origin of replication.

[18]. However, for none of the HPyVs the exact route of transmission is known, including the TSPyV and MCPyV pathogenic viruses. Although these latter two HPyVs cause skin disease, large differences in skin viral DNA prevalence suggest different routes of infection, transmission, and/or persistence. MCPyV DNA can be detected on healthy and diseased skin in 30–60% of individuals, regardless of the host immune status [19 - 27]. Studies that have looked at TSPyV prevalence in various samples (e.g., skin swabs, skin biopsies, plucked eyebrows, serum/plasma, tonsils, urine etc.) have found less than 5% DNA-positivity in asymptomatic individuals, regardless of their immune status [11, 13, 28, 29] (**Table 1**). Higher TSPyV prevalence and high viral loads were only reported in TS patients [13], which suggest that TSPyV is the causative agent of TS disease (**Chapter 3**). Hypothetically, at some stage in life TSPyV infects the skin, to cause disease there in a small minority of immunocompromised hosts. If, however, the skin represents the major organ of TSPyV persistence, is far from clear, as



the majority of infected (seropositive) individuals are skin TSPyV-negative. Whether TS is a manifestation of TSPyV reactivation from a yet unidentified (transplanted) organ reservoir [29], or an unfortunately timed primary infection in the midst of immunosuppression, is unknown.

**Table 1:** TSPyV prevalence

Samples tested	Number of samples	Number of positives (%)	References
<b>Skin biopsy</b>			
TS, Immunocompromized	1	1 (100)	[11]
TS, Immunocompromized	1	1 (100)	[12]
TS, Immunocompromized	11	11 (100)	[13]
TS, Immunocompromized	1	1 (100)	[49]
TS, Immunocompromized	1	1 (100)	[50]
TS, Immunocompromized	2	2 (100)	[51]
Various other skin disease, Pilomatricoma	193	0 (-)	[52]
	10	0 (-)	[53]
<b>Plucked eyebrow hair</b>			
Immunocompromized (TS)	1	1 (100)	This study
Immunocompromized (RTR)	69	3 (4)	[11]
Immunocompromized (RTR)	81	2 (2)	This study
<b>Tonsillar tissue</b>			
Immunocompetent	229	8 (4)	[29]
<b>Nasopharyngeal swabs</b>			
Immunocompromized (TX)	32	1 (3)	[54]
<b>Feces</b>			
Immunocompromized (TX)	32	1 (3)	[54]
Immunocompetent (GE)	38	0 (-)	[28]
Immunocompetent (GE)	160	2 (1)	This study
<b>Kidney biopsy</b>			
Immunocompromized (TS)	1	1 (100)	[49]
<b>Urine</b>			
Immunocompromized (TS)	1	0 (-)	[49]
Immunocompromized (TX)	179	2 (1)	[28]
Immunocompetent	17	0 (-)	This study
<b>Serum/plasma</b>			
Immunocompromized (TX)	88	0 (-)	[28]
Immunocompetent	19	0 (-)	This study
<b>Cerebrospinal fluid</b>			
Immunocompromized (TX)	74	0 (-)	[28]
Immunocompetent	38	0 (-)	This study

Abbreviations: TS, trichodysplasia spinulosa; RTR, renal transplant recipients; TX, transplant patient (not fully specified in the cited study [28]); GE, gastroenteritis.

## TSPyV seroprevalence

Epidemiological studies looking at polyomavirus seroprevalence in humans indicate that HPyV infections of the general population are highly prevalent and occur at early age, probably without clinical manifestation. In healthy, immunocompetent adults, seroprevalences of 80-100% for BKPyV [30 - 33], 40-70% for JCPyV [30 - 33], 55-90% for KIPyV [31, 32, 34, 35], 70-100% for WUPyV [31, 32, 34, 35], 40-80% for MCPyV [19, 31, 32, 36 - 38], 60% for HPyV6 [19, 38], 35% for HPyV7 [19, 38], 35-50% for HPyV9 [38 - 40] and 17-23% for HPyV12 [41] have been reported so far.

Detection of serum antibodies against TSPyV VP1 have been described in several studies using either a multiplex serology method based on Glutathione-S-transferase (GST)-VP1 fusion proteins [38, 42, 43] or VP1 VLP-based antibody detection by ELISA [44 - 46]. All these studies suggest that TSPyV circulate widely in humans, with seroprevalences of about 75% in adults in all populations tested, including a total of 528 Dutch, 371 and 394 Finnish, 829 Italian, and 799 Australian healthy individuals [38, 42, 44 - 46]. Furthermore, age distribution analysis of TSPyV VP1 seroprevalences revealed an increase from about 41% in children aged 0–9 years to 75% in adults aged 30 years and older, and wane later in life comparable to BKPyV [38, 42]. This distribution pattern is in agreement with other HPyVs, although modest differences in age distribution patterns are observed sometimes [31, 39, 41].

The seroreactivity of TSPyV VP1 antibody responses of seropositive individuals shows a decline in the median seroreactivity from 40 to 49 years of age onwards [38, 42], similar to what was reported before for BKPyV [47]. This waning of antibody levels in the elderly might be due to immunosenescence [45]. Serum cross-reactivity between TSPyV and other (related) polyomaviruses, as shown for instance between HPyV9/LPyV [39, 40] and MCPyV/ChPyV [48], and to a lesser extent between HPyV6/HPyV7 [19, 38] and BKPyV/JCPyV [31], was not observed [48].

Earlier BKPyV serological studies have revealed increased VP1 antibody and seropositivity levels in immunocompromized transplant patients, in accordance with intensity of BKPyV infection or reactivation post-transplantation [55 - 57]. Whether this pattern is also true for TSPyV remains to be seen. So far, only one study has shown a higher (89%) TSPyV VP1 seroprevalence in a renal-transplant patient population compared to healthy individuals (75%) [42]. How to interpret the putative increase in TSPyV seropositivity concomitant with immunosuppressive treatment is not fully understood. Hypothetically, this may be explained by an increase in humoral immunity in response to TSPyV reactivation, possibly viremia, because of iatrogenic suppression of cellular immunity [11].

Altogether, seroepidemiological data indicate that TSPyV infection is common and occurs primarily at young age with a deterioration of antibody levels at the later stages of life. Knowledge about cellular immunity against TSPyV is scarce. A Finnish study suggested that TSPyV Th-cell responses correlate with TSPyV serological responses, but can sometimes also be detected in TSPyV-seronegative individuals [58].

## Trichodysplasia spinulosa

### Clinical description and epidemiology

TS is a rare skin disease first reported by Izakovic and colleagues in 1995 [59]. In 1999, Haycox and colleagues fully described the disease and introduced the term “trichodysplasia spinulosa” [60]. They showed for the first time the presence of virus particles and suggested a viral etiology for TS. Ever since, approximately 30 comparable cases were published (**Table 2**). Many of these reports adopted the term “trichodysplasia spinulosa”, whereas others used different terms for the same condition, such as “trichodysplasia”, “pilomatrix dysplasia of immune suppression”, “cyclosporine-induced folliculodystrophy” or “viral-associated trichodysplasia” [11 - 13, 49 - 51, 61 - 74].

So far, TS has been observed in immunosuppressed organ transplant patients and occasionally in chronic and acute lymphocytic leukemia patients. Despite the disparity in TS terminology, in all cases the disease was clinically characterized by spiny follicular papules distributed largely on the face and ears, and to a lesser extent on extremities, trunk, and scalp. In most patients, non-scarring alopecia of the eyebrows was observed, upon which small hyperkeratotic white-yellowish spicules started to protrude the skin. At the same time, these features manifested also on the nose and ears (**Figure 3A**). As the disease progressed, the skin of eyebrows, ears, and nose thickened to cause disfigurement of the facial appearance, sometimes resulting in a “leonine facies” [70], in combination with increased conspicuous spine expression [11, 71].

### Histology and viral pathology

Histological analysis of TS skin biopsies reveals acanthosis of the epidermis in most cases. In addition, enlargement of the hair follicles with excessive number of proliferative cells is observed [60]. Compared with normal skin, TS follicles are absent of hair shafts, and papilla with abnormal corneocytes filling the infundibula of the follicles (**Figure 3B**). Sometimes a subtle perifollicular lymphocytic infiltrate is seen.

Electron microscopical analyses of TS lesions repeatedly revealed intranuclear, crystalloid-organized, regularly spaced 38 - 45-nm virus particles (**Figure 3E**). The identity of TSPyV remained unknown for many years [12, 60, 63, 64, 66, 67], until in 2010 when the double-stranded DNA viral genome was revealed with the help of rolling-circle amplification, cloning, primer-genome-walking, and sequencing [11]. By now, several research groups have confirmed the presence of TSPyV DNA in lesional TS samples (**Table 2**) [12, 13, 50, 51]. The given name of TSPyV was accepted by the international polyomavirus study group [79], while awaiting official recognition by the ICTV.

Development of TS may be caused by uncontrolled proliferation of TSPyV-infected inner root sheath (IRS) cells. The irregular IRS cells show enlarged, dystrophic and prominent eosinophilic perinuclear globules that probably represent the accumulation of trichohyalin protein (**Figure 3C and D**) [13, 60]. Ki-67 staining of the affected cells suggests that the IRS cells, or a subpopulation thereof, are hyperproliferating [60].

**Table 2A:** Overview of reported TS cases

Publication year	Gender	Age	Medical history	Country	Reference
1995	Male	31	Kidney transplant	GER	[59]
1999 **	Male	44	Kidney-pancreas transplant	USA	[13, 60]
2000	Female	13	Lung transplant	USA	[61]
2004	Female	34	Kidney transplant	USA	[62]
2004	Female	13	Kidney transplant	USA	[63]
2005 (case 1) **	Male	19	ALL (Pre-B cell)	USA	[13, 64]
2005 (case 2)	Male	8	Kidney transplant	USA	[64]
2006	Female	48	Kidney transplant	USA	[65]
2007 (case 1) **	Male	8	ALL (T-cell)	AUS	[13, 67]
2007 (case 2)	Male	6	ALL (T-cell)	AUS	[67]
2007	Male	68	non-Hodgkin's lymphoma	USA	[66]
2008	Male	70	CLL	AUS	[69]
2008 **	Female	37	Heart transplant	USA	[13, 68]
2010 **	Female	5	Heart transplant	CAN	[13, 71]
2010 **	Male	5	Heart transplant	USA	[13, 70]
2010 *	Male	15	Heart transplant	NLD	[11]
2011 **	Female	7	ALL (Pre-B cell)	USA	[12]
2011	Female	27	Kidney transplant	USA	[72]
2011	Female	9	ALL (Pre-B cell)	CAN	[73]
2012 (case 1) **	Male	5	Kidney transplant	USA	[13]
2012 (case 2) **	Female	63	Heart transplant	USA	[13]
2012 (case 3) **	Male	62	Lung transplant	USA	[13, 51]
2012 **	Male	48	Kidney transplant	USA	[49]
2012 **	Female	57	CLL	USA	[50]
2012	Female	46	Kidney transplant	USA	[74]
2013 **	Female	49	Kidney transplant (Lupus)	AUS	[75]
2013 **	Female	20	Lupus glomerulonephritis***	FRA	[76]
2013	Male	5	ALL (Pre-B cell)	LBN	[77]
2013	Female	14	Lung transplant	USA	[78]

**Table 2B:** Summary of reported TS cases shown in **Table 2A**

Total Number of TS cases	Male:Female ratio	Mean age	Transplant type (TX) or underlying disease (n, relative %)	Relative percentage of Countries reporting TS cases (n, relative %)
29	14:15	29	Kidney TX: (10, 34%) Kidney-pancreas TX: (1, 3%) Heart TX: (5, 17%) Lung TX: (3, 10%) ALL (Pre-B cell): (4, 14%) CLL: (2, 7%) ALL (T-cell): (2, 7%) NHL: (1, 3%) LGN: (1, 3%)	USA: (19, 66%) AUS: (4, 14%) CAN: (2, 7%) NLD: (1, 3%) GER: (1, 3%) FRA: (1, 3%) LBN: (1, 3%)

\* Identification of TSPyV [11]

\*\* Presence of TSPyV DNA confirmed

\*\*\* Lupus glomerulonephritis condition was treated with immunosuppression, shortly after pt died

Abbreviations: USA, United states of America; AUS, Australia; CAN, Canada; NLD, The Netherlands; GER, Germany; FRA, France; LBN, Lebanon; ALL, Acute lymphocytic leukemia; CLL, Chronic lymphocytic leukemia; M, Male; F, Female; TX, Transplant type; NHL, non-Hodgkin's lymphoma; LGN, Lupus glomerulonephritis

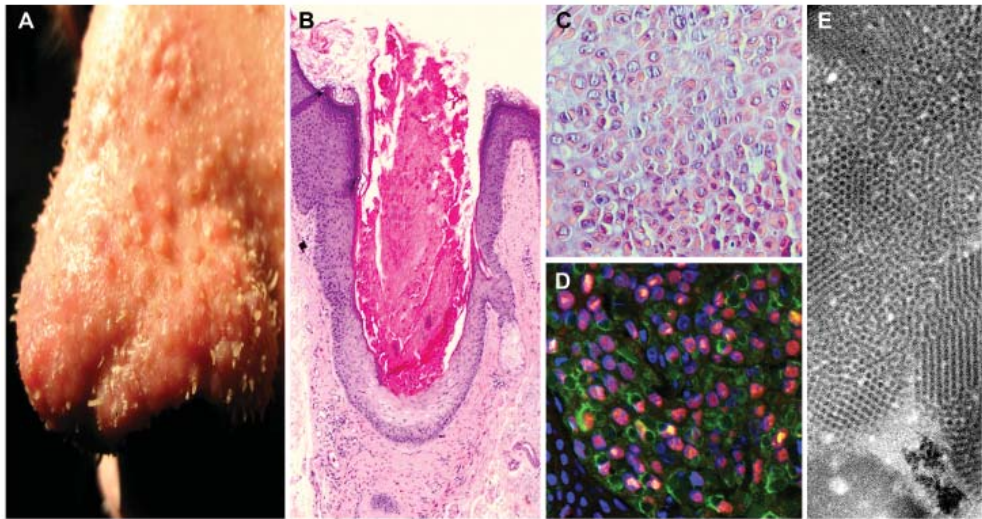


## Patients at risk and diagnosis

TS is exclusively observed in immunocompromized patients, especially kidney(-pancreas) (37%), heart (17%) and lung (10%) transplant recipients (**Table 2B**). It can also develop in otherwise immunocompromized patients, e.g., in pre-B cell (14%) and T-cell (7%) acute lymphocytic leukemia patients, chronic lymphocytic leukemia patients (7%), non-Hodgkin's lymphoma patient (3%), and Lupus glomerulonephritis patient treated with immunosuppression (3%). The solid organ transplant group, especially the kidney graft recipients, seem the most hit by TSPyV. If this reflects the numbers of susceptible immunosuppressed hosts (**Figure 1**), or rather is related pathogenically to the transplanted organ is unknown. For BKPyV it is known that the course and dose of immunosuppression is associated with polyomavirus reactivation and accompanying complications [80]. Whether this is also true for TSPyV remains to be seen.

Diagnosis of TS is primarily established on clinical features, such as visual detection of the spiny follicular papules on the face (**Figure 3A**). The clinical diagnosis can be confirmed by histopathologic analysis of a biopsy of the lesional skin showing enlarged dystrophic hair follicles with eosinophilic perinuclear globules of the IRS cells (**Figure 3C**). TSPyV VP1 staining, as illustrated in **Figure 3D** can be added to the immunohistochemical diagnosis.

Electron microscopy has proven useful for visualization of TSPyV particles in nuclei of IRS cells of affected hair follicles and provided the first clue of TS having a viral etiology.



**Figure 3.** Clinical and (immuno)histopathological characteristics of TS disease. (A) Clinical features of the TS disease seen on the nose with hard small hyperkeratotic white-yellowish spicules protruding the skin. (B) H&E characteristics of lesional skin showing the conspicuous papule and spicule protruding from the epidermis. (C) A magnified picture of an affected hair follicle showing irregular inner root sheath cells with enlarged, dystrophic eosinophilic, and perinuclear globules. (D) Immunohistochemistry staining of an enlarged hair follicle showing colocalization of trichohyalin (green) and TSPyV VP1 (red). (E) Electron microscopy of an inner root sheath cell with intranuclear, crystalloid-organized, regularly spaced 38 - 45-nm virus particles.

Low sensitivity and specificity, however, as well as its cumbersomeness make this method less suitable for TSPyV detection in a clinical setting. Virus culture from fresh materials, in an attempt to identify the nature of the virus particles, was so far unsuccessful [64].

### TSPyV PCR

In addition to the clinical and histopathological approach, TSPyV DNA detection and load measurement by quantitative PCR can be included in the TS diagnosis work-up. All lesional TS samples analyzed so far for TSPyV DNA content, were shown positive with high average viral loads, as described in **Chapter 3** [11, 13]. Normal skin samples of TS patients, as well as skin swabs and plucked eyebrow hairs from unaffected individuals were occasionally shown to be TSPyV DNA-positive, but with low viral loads [11, 13]. Up until now, four primer and probe sets for (quantitative) PCR have been described for TSPyV, one located in the NCCR, one in VP1 and two in the T-antigen coding regions [11, 49].

Preferably, a lesional skin biopsy or plucked spicule is used for TSPyV detection in TS patients [11 - 13]. Proteinase-K treatment and total DNA extraction of the material prior to PCR analysis is indicated to remove excessive protein and protective capsid structures. A non-invasive method to obtain clinical sample would be swabbing of the patient's face, for instance of the forehead that proved to bear high loads of TSPyV DNA in one TS patient sampled in this way [11, 13], suggesting equal sensitivity of skin swab and biopsy to detect viral DNA. However, for viral load calculations normalized for cellular DNA content and the number of cells, only a biopsy or plucked spicule can be regarded a reliable specimen.

PCR analysis of plucked eyebrow hairs has revealed the presence of TSPyV DNA in a small proportion (4%) of asymptomatic renal transplant recipients [11]. How forehead swabs and plucked eyebrow hairs compare with respect to the number of obtained cells, viral DNA detection and load measurement is unknown. Interestingly, in a study a 13-year-old immunocompromized heart transplant patient without TS symptoms sampled 1 month after his immunosuppression, TSPyV DNA was detected in his stool and in a nasopharyngeal swab. Repeated analysis in the following months remained negative [54]. Occasionally also urine and kidney samples of immunocompromized patients were found to be positive for TSPyV DNA [49]. The meaning of these findings, in particular their role in TS diagnosis, remains unclear.

Altogether, skin swabs, plucked hairs, and fresh or fixed lesional biopsies of the affected skin can serve as a proper (diagnostic) sample for viral detection and load measurements by (quantitative) PCR to confirm TSPyV infection and TS diagnosis, and to monitor disease progression and/or treatment efficacy.

### Antiviral treatment

Improving the patient's immunity usually leads to complete resolution of TS symptoms. In case of TS in (organ) transplant recipients, reducing the dose of immunosuppressive drugs

should be considered, obviously without endangering the grafted organ by immune rejection. This strategy has been shown as the best option at the moment to improve the outcome of BKPyV-associated nephropathy [81, 82].

Next to reduction of immunosuppression, antiviral therapy should be considered a treatment option in controlling viral replication and TS progression without jeopardizing the transplanted organ. A number of TS patients treated with topical cidofovir 1 - 3% cream, which serves as a cytosine analogue inhibiting human polymerase activity needed for polyomavirus replication, have demonstrated significant reduction in symptoms [11, 50, 63, 64, 66]. Nephrotoxicity after topical use of this drug was monitored in some patients, but this effect was not reported. Furthermore, oral valganciclovir, a guanosine analogue also inhibiting polymerase activity, has been reported to have modest in some and strong activity in other TS patients [68, 70, 71]. The putative mechanism of action of this drug in the treatment of TS remains unclear at the moment, since valganciclovir requires (viral) thymidine kinase-mediated modifications to exert its polymerase-inhibiting effect, which are not expected to occur upon polyomavirus infection [83].

## Concluding remarks

TSPyV is a ubiquitous virus, similar to most other HPyVs, as concluded from its 75% seroprevalence. When looking at its DNA prevalence in TS-asymptomatic individuals, however, the virus appears difficult to detect with a prevalence not exceeding 5% (**Table 1**). Whether TSPyV persists in human skin and replicates at undetectable levels, or alternatively has its latent reservoir in an extracutaneous site, like for instance the tonsils, remains to be seen. It is also not known how people acquire TSPyV infection, for instance by direct contact or by respiratory transmission, and whether TS symptoms are caused by reactivation as a result of poor immunity or by primary infection later in life in the midst of immunosuppression. In addition, the pathogenic mechanisms used by TSPyV are unknown, but likely include induction of hyperproliferation of IRS cells through its T-antigens.

With the identification of TSPyV as the probable causative infectious agent, improvement of TS diagnosis and clinical care can be achieved. This includes TSPyV-specific DNA detection and quantification, and immunohistochemistry or immunofluorescence in the course of the diagnostic process, and antiviral therapy as a treatment option. Because of the increasing number of immunocompromized patients and the high TSPyV seroprevalence in healthy populations, it is expected that TS will be diagnosed and reported more often in immunocompromized patients than it is at this moment. For this reason, clinicians, in particular nephrologists, dermatologists, and pathologists, should become more aware of this condition, knowing that appropriate viral diagnosis and treatment is at hand.

## References

1. U.S.Department of Health & Human Services. United States Organ Procurement and Transplantation network (OPTN). In: 2012.
2. Nickeleit V, Klimkait T, Binet IF, Dalquen P, Del Zenero V, Thiel G, Mihatsch MJ, Hirsch HH. Testing for polyomavirus type BK DNA in plasma to identify renal-allograft recipients with viral nephropathy (2000) *N. Engl. J. Med.* 342: 1309-1315.
3. Gosert R, Rinaldo CH, Funk GA, Egli A, Ramos E, Drachenberg CB, Hirsch HH. Polyomavirus BK with rearranged noncoding control region emerge in vivo in renal transplant patients and increase viral replication and cytopathology (2008) *J. Exp. Med.* 205: 841-852.
4. Mateen FJ, Muralidharan R, Carone M, van de Beek D, Harrison DM, Aksamit AJ, Gould MS, Clifford DB, Nath A. Progressive multifocal leukoencephalopathy in transplant recipients (2011) *Ann. Neurol.* 70: 305-322.
5. Reid CE, Li H, Sur G, Carmillo P, Bushnell S, Tizard R, McAuliffe M, Tonkin C, Simon K, Goelz S, Cinque P, Gorelik L, Carulli JP. Sequencing and analysis of JC virus DNA from natalizumab-treated PML patients (2011) *J. Infect. Dis.* 204: 237-244.
6. Cinque P, Koranik IJ, Gerevini S, Miro JM, Price RW. Progressive multifocal leukoencephalopathy in HIV-1 infection (2009) *Lancet Infect. Dis.* 9: 625-636.
7. Houben R, Shuda M, Weinkam R, Schrama D, Feng H, Chang Y, Moore PS, Becker JC. Merkel cell polyomavirus-infected Merkel cell carcinoma cells require expression of viral T antigens (2010) *J. Virol.* 84: 7064-7072.
8. Shuda M, Arora R, Kwun HJ, Feng H, Sarid R, Fernandez-Figueras MT, Tolstov Y, Gjoerup O, Mansukhani MM, Swerdlow SH, Chaudhary PM, Kirkwood JM, Nalesnik MA, Kant JA, Weiss LM, Moore PS, Chang Y. Human Merkel cell polyomavirus infection I. MCV T antigen expression in Merkel cell carcinoma, lymphoid tissues and lymphoid tumors (2009) *Int. J. Cancer* 125: 1243-1249.
9. Feng H, Shuda M, Chang Y, Moore PS. Clonal integration of a polyomavirus in human Merkel cell carcinoma (2008) *Science* 319: 1096-1100.
10. Shuda M, Feng H, Kwun HJ, Rosen ST, Gjoerup O, Moore PS, Chang Y. T antigen mutations are a human tumor-specific signature for Merkel cell polyomavirus (2008) *Proc. Natl. Acad. Sci. U. S. A.* 105: 16272-16277.
11. van der Meijden E, Janssens RW, Lauber C, Bouwes Bavinck JN, Gorbalenya AE, Feltkamp MC. Discovery of a new human polyomavirus associated with trichodysplasia spinulosa in an immunocompromized patient (2010) *PLoS Pathog.* 6: e1001024.
12. Matthews MR, Wang RC, Reddick RL, Saldivar VA, Browning JC. Viral-associated trichodysplasia spinulosa: a case with electron microscopic and molecular detection of the trichodysplasia spinulosa-associated human polyomavirus (2011) *J. Cutan. Pathol.* 38: 420-431.



13. Kazem S, van der Meijden E, Kooijman S, Rosenberg AS, Hughey LC, Browning JC, Sadler G, Busam K, Pope E, Benoit T, Fleckman P, de VE, Eekhof JA, Feltkamp MC. Trichodysplasia spinulosa is characterized by active polyomavirus infection (2012) *J. Clin. Virol.* 53: 225-230.
14. Stehle T, Gamblin SJ, Yan Y, Harrison SC. The structure of simian virus 40 refined at 3.1 Å resolution (1996) *Structure* 4: 165-182.
15. Broekema NM, Imperiale MJ. miRNA regulation of BK polyomavirus replication during early infection (2013) *Proc. Natl. Acad. Sci. U. S. A.* 110: 8200-8205.
16. Lee S, Paulson KG, Murchison EP, Afanasiev OK, Alkan C, Leonard JH, Byrd DR, Hannon GJ, Nghiem P. Identification and validation of a novel mature microRNA encoded by the Merkel cell polyomavirus in human Merkel cell carcinomas (2011) *J. Clin. Virol.* 52: 272-275.
17. Seo GJ, Chen CJ, Sullivan CS. Merkel cell polyomavirus encodes a microRNA with the ability to autoregulate viral gene expression (2009) *Virology* 383: 183-187.
18. Bofill-Mas S, Formiga-Cruz M, Clemente-Casares P, Calafell F, Girones R. Potential transmission of human polyomaviruses through the gastrointestinal tract after exposure to virions or viral DNA (2001) *J. Virol.* 75: 10290-10299.
19. Schowalter RM, Pastrana DV, Pumphrey KA, Moyer AL, Buck CB. Merkel cell polyomavirus and two previously unknown polyomaviruses are chronically shed from human skin (2010) *Cell Host Microbe* 7: 509-515.
20. Garneski KM, Warcola AH, Feng Q, Kiviat NB, Leonard JH, Nghiem P. Merkel cell polyomavirus is more frequently present in North American than Australian Merkel cell carcinoma tumors (2009) *J. Invest. Dermatol.* 129: 246-248.
21. Becker JC, Houben R, Ugurel S, Trefzer U, Pfohler C, Schrama D. MC polyomavirus is frequently present in Merkel cell carcinoma of European patients (2009) *J. Invest. Dermatol.* 129: 248-250.
22. Wieland U, Mauch C, Kreuter A, Krieg T, Pfister H. Merkel cell polyomavirus DNA in persons without merkel cell carcinoma (2009) *Emerg. Infect. Dis.* 15: 1496-1498.
23. Foulongne V, Dereure O, Kluger N, Moles JP, Guillot B, Segondy M. Merkel cell polyomavirus DNA detection in lesional and nonlesional skin from patients with Merkel cell carcinoma or other skin diseases (2010) *Br. J. Dermatol.* 162: 59-63.
24. Kassem A, Technau K, Kurz AK, Pantulu D, Loning M, Kayser G, Stickeler E, Weyers W, Diaz C, Werner M, Nashan D, Zur HA. Merkel cell polyomavirus sequences are frequently detected in nonmelanoma skin cancer of immunosuppressed patients (2009) *Int. J. Cancer* 125: 356-361.
25. Dworkin AM, Tseng SY, Allain DC, Iwenofu OH, Peters SB, Toland AE. Merkel Cell Polyomavirus in Cutaneous Squamous Cell Carcinoma of Immunocompetent Individuals (2009) *J. Invest. Dermatol.* 129: 2868-2874.

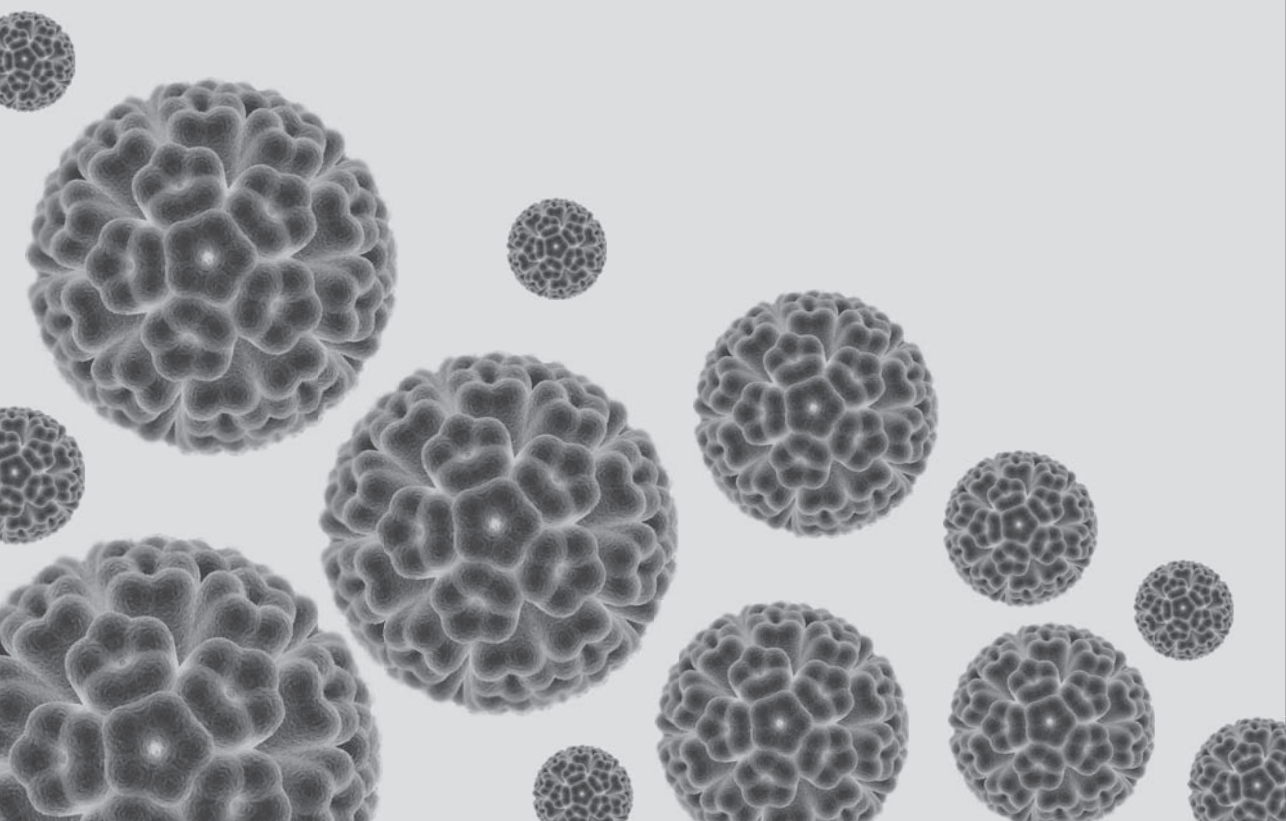
26. Mertz KD, Pfaltz M, Junt T, Schmid M, Fernandez Figueras MT, Pfaltz K, Barghorn A, Kempf W. Merkel cell polyomavirus is present in common warts and carcinoma in situ of the skin (2010) *Hum. Pathol.* 41: 1369-1379.
27. Foulongne V, Kluger N, Dereure O, Mercier G, Moles JP, Guillot B, Segondy M. Merkel cell polyomavirus in cutaneous swabs (2010) *Emerg. Infect. Dis.* 16: 685-687.
28. Scuda N, Hofmann J, Calvignac-Spencer S, Ruprecht K, Liman P, Kuhn J, Hengel H, Ehlers B. A novel human polyomavirus closely related to the african green monkey-derived lymphotropic polyomavirus (2011) *J. Virol.* 85: 4586-4590.
29. Sadeghi M, Aaltonen LM, Hedman L, Chen T, Soderlund-Venermo M, Hedman K. Detection of TS polyomavirus DNA in tonsillar tissues of children and adults: Evidence for site of viral latency (2014) *J. Clin. Virol.* 59: 55-58.
30. Egli A, Infanti L, Dumoulin A, Buser A, Samaridis J, Stebler C, Gosert R, Hirsch HH. Prevalence of polyomavirus BK and JC infection and replication in 400 healthy blood donors (2009) *J. Infect. Dis.* 199: 837-846.
31. Kean JM, Rao S, Wang M, Garcea RL. Seroepidemiology of Human Polyomaviruses (2009) *PLoS Pathog.* 5: e1000363.
32. Carter JJ, Paulson KG, Wipf GC, Miranda D, Madeleine MM, Johnson LG, Lemos BD, Lee S, Warcola AH, Iyer JG, Nghiem P, Galloway DA. Association of Merkel Cell Polyomavirus-Specific Antibodies With Merkel Cell Carcinoma (2009) *J. Natl. Cancer Inst.* 4: 1510-1522.
33. Kjaerheim K, Roe OD, Waterboer T, Sehr P, Rizk R, Dai HY, Sandeck H, Larsson E, Andersen A, Boffetta P, Pawlita M. Absence of SV40 antibodies or DNA fragments in prediagnostic mesothelioma serum samples (2007) *Int. J. Cancer* 120: 2459-2465.
34. Neske F, Prifert C, Scheiner B, Ewald M, Schubert J, Opitz A, Weissbrich B. High prevalence of antibodies against polyomavirus WU, polyomavirus KI, and human bocavirus in German blood donors (2010) *BMC Infect. Dis.* 10: 215.
35. Nguyen NL, Le BM, Wang D. Serologic evidence of frequent human infection with WU and KI polyomaviruses (2009) *Emerg. Infect. Dis.* 15: 1199-1205.
36. Tolstov YL, Pastrana DV, Feng H, Becker JC, Jenkins FJ, Moschos S, Chang Y, Buck CB, Moore PS. Human Merkel cell polyomavirus infection II. MCV is a common human infection that can be detected by conformational capsid epitope immunoassays (2009) *Int. J. Cancer* 125: 1250-1256.
37. Touze A, Gaitan J, Arnold F, Cazal R, Fleury MJ, Combelas N, Sizaret PY, Guyétant S, Maruani A, Baay M, Tognon M, Coursaget P. Generation of Merkel cell polyomavirus (MCV)-like particles and their application to detection of MCV antibodies (2010) *J. Clin. Microbiol.* 48: 1767-1770.
38. van der Meijden E, Bialasiewicz S, Rockett RJ, Tozer SJ, Sloots TP, Feltkamp MC. Different serologic behavior of MCPyV, TSPyV, HPyV6, HPyV7 and HPyV9 polyomaviruses found on the skin (2013) *PLoS One* 8: e81078.

39. Nicol JT, Touze A, Robinot R, Arnold F, Mazzoni E, Tognon M, Coursaget P. Seroprevalence and Cross-reactivity of Human Polyomavirus 9 (2012) *Emerg. Infect. Dis.* 18: 1329-1332.
40. Trusch F, Klein M, Finsterbusch T, Kuhn J, Hofmann J, Ehlers B. Seroprevalence of human polyomavirus 9 and cross-reactivity to African green monkey-derived lymphotropic polyomavirus (2012) *J. Gen. Virol.* 93: 698-705.
41. Korup S, Rietscher J, Calvignac-Spencer S, Trusch F, Hofmann J, Moens U, Sauer I, Voigt S, Schmuck R, Ehlers B. Identification of a novel human polyomavirus in organs of the gastrointestinal tract (2013) *PLoS One* 8: e58021.
42. van der Meijden E, Kazem S, Burgers MM, Janssens R, Bouwes Bavinck JN, de Melker H, Feltkamp MC. Seroprevalence of Trichodysplasia Spinulosa-associated Polyomavirus (2011) *Emerg. Infect. Dis.* 17: 1355-1363.
43. Waterboer T, Sehr P, Michael KM, Franceschi S, Nieland JD, Joos TO, Templin MF, Pawlita M. Multiplex human papillomavirus serology based on in situ-purified glutathione s-transferase fusion proteins (2005) *Clin. Chem.* 51: 1845-1853.
44. Chen T, Mattila PS, Jartti T, Ruuskanen O, Soderlund-Venermo M, Hedman K. Seroepidemiology of the Newly Found Trichodysplasia Spinulosa-Associated Polyomavirus (2011) *J. Infect. Dis.* 204: 1523-1526.
45. Nicol JT, Robinot R, Carpentier A, Carandina G, Mazzoni E, Tognon M, Touze A, Coursaget P. Age-specific seroprevalences of merkel cell polyomavirus, human polyomaviruses 6, 7, and 9, and trichodysplasia spinulosa-associated polyomavirus (2013) *Clin. Vaccine Immunol.* 20: 363-368.
46. Sadeghi M, Aronen M, Chen T, Jartti L, Jartti T, Ruuskanen O, Soderlund-Venermo M, Hedman K. Merkel cell polyomavirus and trichodysplasia spinulosa-associated polyomavirus DNAs and antibodies in blood among the elderly (2012) *BMC Infect. Dis.* 12: 383.
47. Viscidi RP, Rollison DE, Sondak VK, Silver B, Messina JL, Giuliano AR, Fulp W, Ajidahun A, Rivanera D. Age-specific seroprevalence of Merkel cell polyomavirus, BK virus, and JC virus (2011) *Clin. Vaccine Immunol.* 18: 1737-1743.
48. Nicol JT, Liais E, Potier R, Mazzoni E, Tognon M, Coursaget P, Touze A. Serological cross-reactivity between Merkel cell polyomavirus and two closely related chimpanzee polyomaviruses (2014) *PLoS One* 9: e97030.
49. Fischer MK, Kao GF, Nguyen HP, Drachenberg CB, Rady PL, Tyring SK, Gaspari AA. Specific detection of trichodysplasia spinulosa-associated polyomavirus DNA in skin and renal allograft tissues in a patient with trichodysplasia spinulosa (2012) *Arch. Dermatol.* 148: 726-733.
50. Wanat KA, Holler PD, Dentchev T, Simbiri K, Robertson E, Seykora JT, Rosenbach M. Viral-associated trichodysplasia: characterization of a novel polyomavirus infection with therapeutic insights (2012) *Arch. Dermatol.* 148: 219-223.

51. Elaba Z, Hughey L, Isayeva T, Weeks B, Solovan C, Solovastru L, Andea A. Ultrastructural and molecular confirmation of the trichodysplasia spinulosa-associated polyomavirus in biopsies of patients with trichodysplasia spinulosa (2012) *J. Cutan. Pathol.* 39: 1004-1009.
52. Scola N, Wieland U, Silling S, Altmeyer P, Stucker M, Kreuter A. Prevalence of human polyomaviruses in common and rare types of non-Merkel cell carcinoma skin cancer (2012) *Br. J. Dermatol.* 167: 1315-1320.
53. Kanitakis J, Kazem S, van der Meijden E, Feltkamp M. Absence of the trichodysplasia spinulosa-associated polyomavirus in human pilomatricomas (2011) *Eur. J. Dermatol.* 21: 453-454.
54. Siebrasse EA, Bauer I, Holtz LR, Le B, Lassa-Claxton S, Canter C, Hmiel P, Shenoy S, Sweet S, Turmelle Y, Shepherd R, Wang D. Human polyomaviruses in children undergoing transplantation, United States, 2008–2010 (2012) *Emerg. Infect. Dis.* 18: 1676-1679.
55. Bohl DL, Brennan DC, Ryschkewitsch C, Gaudreault-Keener M, Major EO, Storch GA. BK virus antibody titers and intensity of infections after renal transplantation (2008) *J. Clin. Virol.* 43: 184-189.
56. Randhawa P, Bohl D, Brennan D, Ruppert K, Ramaswami B, Storch G, March J, Shapiro R, Viscidi R. Longitudinal analysis of levels of immunoglobulins against BK virus capsid proteins in kidney transplant recipients (2008) *Clin. Vaccine Immunol.* 15: 1564-1571.
57. Bodaghi S, Comoli P, Bosch R, Azzi A, Gosert R, Leuenberger D, Ginevri F, Hirsch HH. Antibody responses to recombinant polyomavirus BK large T and VP1 proteins in young kidney transplant patients (2009) *J. Clin. Microbiol.* 47: 2577-2585.
58. Kumar A, Kantele A, Jarvinen T, Chen T, Kavola H, Sadeghi M, Hedman K, Franssila R. Trichodysplasia spinulosa-associated polyomavirus (TSV) and Merkel cell polyomavirus: correlation between humoral and cellular immunity stronger with TSV (2012) *PLoS One* 7: e45773.
59. Izakovic J, Buchner SA, Duggelin M, Guggenheim R, Itin PH. [Hair-like hyperkeratoses in patients with kidney transplants. A new cyclosporin side-effect] (1995) *Hautarzt* 46: 841-846.
60. Haycox CL, Kim S, Fleckman P, Smith LT, Piepkorn M, Sundberg JP, Howell DN, Miller SE. Trichodysplasia spinulosa--a newly described folliculocentric viral infection in an immunocompromised host (1999) *J. Investig. Dermatol. Symp. Proc.* 4: 268-271.
61. Chastain MA, Millikan LE. Pilomatrix dysplasia in an immunosuppressed patient (2000) *J. Am. Acad. Dermatol.* 43: 118-122.
62. Heaphy MR, Jr., Shamma HN, Hickmann M, White MJ. Cyclosporine-induced folliculodystrophy (2004) *J. Am. Acad. Dermatol.* 50: 310-315.
63. Sperling LC, Tomaszewski MM, Thomas DA. Viral-associated trichodysplasia in patients who are immunocompromised (2004) *J. Am. Acad. Dermatol.* 50: 318-322.

64. Wyatt AJ, Sachs DL, Shia J, Delgado R, Busam KJ. Virus-associated trichodysplasia spinulosa (2005) *Am. J. Surg. Pathol.* 29: 241-246.
65. Campbell RM, Ney A, Gohh R, Robinson-Bostom L. Spiny hyperkeratotic projections on the face and extremities of a kidney transplant recipient (2006) *Arch. Dermatol.* 142: 1643-1648.
66. Osswald SS, Kulick KB, Tomaszewski MM, Sperling LC. Viral-associated trichodysplasia in a patient with lymphoma: a case report and review (2007) *J. Cutan. Pathol.* 34: 721-725.
67. Sadler GM, Halbert AR, Smith N, Rogers M. Trichodysplasia spinulosa associated with chemotherapy for acute lymphocytic leukaemia (2007) *Australas. J. Dermatol.* 48: 110-114.
68. Holzer AM, Hughey LC. Trichodysplasia of immunosuppression treated with oral valganciclovir (2009) *J. Am. Acad. Dermatol.* 60: 169-172.
69. Lee JS, Frederiksen P, Kossard S. Progressive trichodysplasia spinulosa in a patient with chronic lymphocytic leukaemia in remission (2008) *Australas. J. Dermatol.* 49: 57-60.
70. Benoit T, Bacelieri R, Morrell DS, Metcalf J. Viral-associated trichodysplasia of immunosuppression: report of a pediatric patient with response to oral valganciclovir (2010) *Arch. Dermatol.* 146: 871-874.
71. Schwieger-Briel A, Balma-Mena A, Ngan B, Dipchand A, Pope E. Trichodysplasia spinulosa--a rare complication in immunosuppressed patients (2010) *Pediatr. Dermatol.* 27: 509-513.
72. Blake BP, Marathe KS, Mohr MR, Jones N, Novosel TA. Viral-Associated Trichodysplasia of Immunosuppression in a Renal Transplant Patient (2011) *J. Drugs Dermatol.* 10: 422-424.
73. Burns A, Arnason T, Fraser R, Murray S, Walsh N. Keratotic "spiny" papules in an immunosuppressed child. Trichodysplasia spinulosa (TS) (2011) *Arch. Dermatol.* 147: 1215-1220.
74. Brimhall CL, Malone JC. Viral-associated trichodysplasia spinulosa in a renal transplant patient (2012) *Arch. Dermatol.* 148: 863-864.
75. Lee YY, Tucker SC, Prow NA, Setoh YX, Banney LA. Trichodysplasia spinulosa: A benign adnexal proliferation with follicular differentiation associated with polyomavirus (2013) *Australas. J. Dermatol.*
76. Moktefi A, Laude H, Brudy GL, Rozenberg F, Vacher Lavenu MC, Dupin N, Carlotti A. Trichodysplasia Spinulosa Associated With Lupus (2013) *Am. J. Dermatopathol.* 36: e70-4
77. Ghosn S, Abboud M, Kurban M, Abbas O. Spiny follicular hyperkeratotic papules on the face (2013) *JAMA Pediatr.* 167: 867-868.
78. Berk DR, Lu D, Bayliss SJ. Trichodysplasia spinulosa in an adolescent with cystic fibrosis and lung transplantation (2013) *Int. J. Dermatol.* 52: 1586-1588.

79. Johne R, Buck CB, Allander T, Atwood WJ, Garcea RL, Imperiale MJ, Major EO, Ramqvist T, Norkin LC. Taxonomical developments in the family Polyomaviridae (2011) *Arch. Virol.* 156: 1627-1634.
80. Funk GA, Steiger J, Hirsch HH. Rapid dynamics of polyomavirus type BK in renal transplant recipients (2006) *J. Infect. Dis.* 193: 80-87.
81. Trofe J, Hirsch HH, Ramos E. Polyomavirus-associated nephropathy: update of clinical management in kidney transplant patients (2006) *Transpl. Infect. Dis.* 8: 76-85.
82. Cosio FG, Amer H, Grande JP, Larson TS, Stegall MD, Griffin MD. Comparison of low versus high tacrolimus levels in kidney transplantation: assessment of efficacy by protocol biopsies (2007) *Transplantation* 83: 411-416.
83. Piret J, Boivin G. Resistance of herpes simplex viruses to nucleoside analogues: mechanisms, prevalence, and management (2011) *Antimicrob. Agents Chemother.* 55: 459-472.



# Chapter 3

## Active TSPyV infection in trichodysplasia spinulosa

*Adapted from\*:*

### **Trichodysplasia spinulosa is characterized by active polyomavirus infection**

*Authors:*

Siamaque Kazem<sup>1</sup>  
Els van der Meijden<sup>1</sup>  
Sander Kooijman<sup>1</sup>  
Arlene Rosenberg<sup>2</sup>  
Lauren Hughey<sup>3</sup>  
John Browning<sup>4</sup>  
Genevieve Sadler<sup>5</sup>

Klaus Busam<sup>6</sup>  
Elena Pope<sup>7</sup>  
Taylor Benoit<sup>8</sup>  
Philip Fleckman<sup>9</sup>  
Esther de Vries<sup>10</sup>  
Just Eekhof<sup>11</sup>  
Mariet Feltkamp<sup>1</sup>

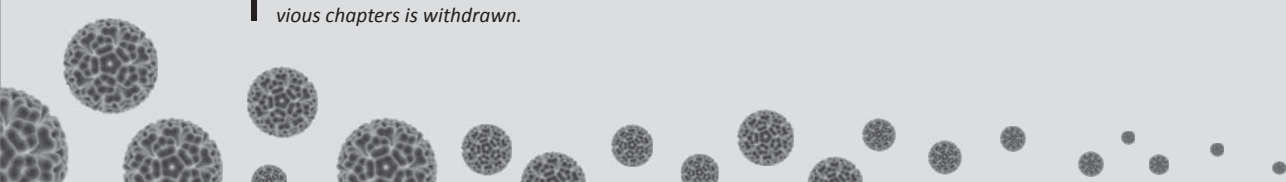
*Affiliations:*

<sup>1</sup> Department of Medical Microbiology, Leiden University Medical Center, Leiden, The Netherlands, <sup>2</sup> Department of Dermatology, MetroHealth Medical Center, Case Western Reserve University, Cleveland, OH, USA, <sup>3</sup> Department of Dermatology, MetroHealth Medical Center, Case Western Reserve University, Cleveland, OH, USA, <sup>4</sup> Department of Dermatology, UT Southwestern Medical Center, Dallas, USA, <sup>5</sup> Department of Dermatology, Sir Charles Gairdner Hospital, Western Australia, Australia, <sup>6</sup> Department of Pathology, Memorial Sloan-Kettering Cancer Center, New York, USA, <sup>7</sup> Department of Pathology, Memorial Sloan-Kettering Cancer Center, New York, USA, <sup>8</sup> Department of Dermatology, Medical University of South Carolina, Charleston, SC, USA, <sup>9</sup> Department of Dermatology, Medical University of South Carolina, Charleston, SC, USA, <sup>10</sup> Department of Pediatrics, Jeroen Bosch Hospital, 's-Hertogenbosch, The Netherlands, <sup>11</sup> Department of Public Health and Primary Care, Leiden University Medical Center, Leiden, The Netherlands

*Published in original form:*

Journal of Clinical Virology, 2012 (53) 225 - 230

*\* Note: Adaptation of this chapter from the original published article concerns only minor textual adjustments. Only in the introduction part, the overlapping information with previous chapters is withdrawn.*





## Abstract

In this chapter, we aimed to corroborate the relationship between TSPyV infection and TS disease by analyzing the presence, load, and precise localization of TSPyV infection in TS patients and in controls. For this purpose, archived TS lesional and non-lesional skin samples were retrieved from TS patients through a PubMed search. Samples were analyzed for the presence and load of TSPyV DNA with quantitative PCR, and for expression and localization of viral protein with immunofluorescence. Findings obtained in TS patients ( $n = 11$ ) were compared to those obtained in healthy controls ( $n = 249$ ). TSPyV DNA detection was significantly associated with disease ( $P < 0.001$ ), with 100% positivity of the lesional and 2% of the control samples. Quantification revealed high TSPyV DNA loads in the lesional samples ( $\sim 10^6$  copies/cell), and low viral loads in the occasionally TSPyV-positive non-lesional and control samples ( $< 10^2$  copies/cell). TSPyV VP1 protein expression was detected only in lesional TS samples, restricted to the nuclei of inner root sheath cells overexpressing trichohyalin. Altogether, the high prevalence and load of TSPyV DNA only in TS lesions, and the abundant expression of TSPyV protein in the affected hair follicle cells demonstrate a tight relationship between TSPyV infection and TS disease, and indicate involvement of active TSPyV infection in TS pathogenesis.

Abstract

## Introduction

As indicated in the previous chapters, TS is a rare follicular skin disease observed only in immunocompromized solid organ transplant and lymphocytic leukemia patients. To our knowledge, approximately 30 cases of TS have been reported in the literature (**Chapter 2**) [1 - 3]. TS is characterized by follicular papules and keratotic protrusions (spicules), most commonly noticed on the face [4 - 6], sometimes accompanied by alopecia of the eyebrows and eyelashes [6]. Histopathology of the affected skin shows enlarged and dysmorphic hair follicles predominantly populated by trichohyalin-overproducing inner root sheath (IRS) cells. A viral cause of TS was suspected since the initial description of TS in 1999 by Haycox *et al.* [5], but attempts to identify the virus from plucked spicules of TS patients failed until the isolation of TSPyV genome by the use of rolling-circle DNA amplification technique [4].

To confirm the association between TSPyV infection and the development of TS, in this chapter we studied the presence and activity of TSPyV infection in a series of TS patients and in healthy controls, comparing presence and load of viral DNA, as well as expression and localization of viral protein.

## Methods and Materials

### Study populations and sample collection

Through a PubMed search on April 2010, we identified twelve articles reporting TS cases (search terms can be requested from the authors). Authors of eligible cases were contacted, asked to collaborate and to provide archival TS samples for additional analyses in our laboratory. This strategy resulted in the inclusion of six TS cases. The remaining six did not reply ( $n = 4$ ), no longer possessed samples ( $n = 1$ ) or declined collaboration ( $n = 1$ ). The sample set was completed with an unpublished case brought forward by one of the contacted authors [7], three new cases sent to us for TSPyV analysis of which one was subsequently published [1], and the TS patient from which TSPyV was originally isolated [4].

From a total of 11 patients, formalin-fixed paraffin-embedded (FFPE) lesional biopsies or sections were obtained, numbered TS1 - TS11 according to the time of arrival at our laboratory (**Table 1**). From TS4 and TS5, we also received a non-lesional skin sample (NLS4 and NLS5). From TS1, also a forehead skin swab was obtained.

Forehead skin swabs were included from a control group of 249 individuals (age 4 - 79 years) participating in a clinical trial that monitored the effects of treatment against common skin warts (Clinical Trial Gov. registration number ISRCTN42730629) [8]. The samples were collected prior to treatment with a cotton-tip brushing uninvolved skin of the central forehead. The cotton-tip was placed in a sterile phosphate-buffered 0.9% saline solution.

**Table 1:** Overview of TS cases analyzed in this study

TS Case <sup>Ref</sup>	Sex	Age	Underlying Disease	TS onset	Description	Country
TS1 <sup>[4]</sup>	Male	15	Heart transplant	2009	TS	NLD
TS2 <sup>[9]</sup>	Male	8	ALL (T cell)	2003	TS	AUS
TS3 <sup>[10]</sup>	Male	19	ALL (pre-B cell)	2004	VATS	USA
TS4 <sup>This study</sup>	Male	5	Renal transplant	2009	TS	USA
TS5 <sup>This study</sup>	Female	63	Heart transplant	2010	TS	USA
TS6 <sup>This study</sup>	Male	62	Lung transplant	2002	TS	USA
TS7 <sup>[7]</sup>	Female	37	Heart transplant	1997	TOI	USA
TS8 <sup>[6]</sup>	Female	5	Heart transplant	2007	TS	CAN
TS9 <sup>[1]</sup>	Female	7	ALL (pre-B cell)	2010	TS	USA
TS10 <sup>[11]</sup>	Male	5	Heart transplant	2008	VATD	USA
TS11 <sup>[5]</sup>	Male	43	Renal transplant	1997	TS	USA

Abbreviations: ALL, acute lymphocytic leukemia; TS, trichodysplasia spinulosa; VATS, virus associated trichodysplasia spinulosa; TOI, trichodysplasia of immunosuppression; VATD, viral-associated trichodysplasia of immunosuppression; NLD, The Netherlands; AUS, Australia; USA, United states of America; CAN, Canada.

## Ethical approval

The Institutional Review Board (IRB) of the Leiden University Medical Center (LUMC) stated that no medical ethical approval was needed to analyze the archival patient samples collected for TS diagnosis for the presence of the TS-associated pathogen.

Medical ethical approval to use skin swabs collected from the healthy controls for TSPyV analysis was obtained from the LUMC IRB. Informed consent (child as well as parental informed consent for participants less than 18 years of age) was obtained from each participant.

## DNA extraction and PCR

FFPE tissue sections were received in sterile Eppendorf tubes (TS2, TS3, TS6, TS7 and TS11) or on glass slides (TS8 and TS10). TS4, TS5, NLS4 and NLS5 were received as paraffin blocks and 5- $\mu$ m sections were cut using a clean blade for each tissue block. One section each was treated with Proteinase-K lysis buffer in a total volume of 100  $\mu$ L. Upon processing, the supernatants were used for PCR analysis. Samples of TS1 and TS9 consisted of previously extracted DNA. The forehead skin swabs were used directly for PCR analysis.

To detect and quantify TSPyV DNA, three previously described real-time quantitative PCRs were performed with primers and probes located in VP1 and LT open reading frames, and in the non-coding control region (NCCR) [4]. In parallel, a pUC19 plasmid-series containing the TSPyV genome titrated in a background of 5 ng/ $\mu$ L human genomic DNA (Promega) was analyzed. To normalize for input DNA and calculate input cell numbers, a  $\beta$ -actin household gene PCR was conducted on the same samples [12]. For the skin swabs no normalization with  $\beta$ -actin household gene was performed, because of lack of sufficient human genomic DNA in the collected skin swabs. In these samples, TSPyV copies are indicated as copies per PCR reaction.

## Histology, immunofluorescence and electron microscopy

Hematoxylin and eosin (H&E) staining for histology and immunofluorescence (IFA) to detect trichohyalin (1:250, AE15 (sc-80607), Santa Cruz) and TSPyV VP1 protein (1:1000), was performed on deparaffinized 4- $\mu$ m sections. To detect VP1, a polyclonal rabbit antiserum was raised against a synthetic TSPyV VP1 peptide (TGNRYRTDYSANDKL). As a TSPyV VP1-specificity control, serum from the same rabbit was used collected prior to VP1 peptide immunization. The sections were incubated in a dark chamber with a secondary Donkey- $\alpha$ -Rabbit-Cy3 antibody (1:1000, Invitrogen), a Goat- $\alpha$ -Mouse-Alexafluor488 antibody (1:300, Jackson) and Hoechst nuclear staining (1:100), and analyzed under a fluorescence microscope.

For electron microscopy (EM), a 50- $\mu$ m FFPE section from lesional sample TS5 was treated as described [13], with slight modifications. In short, the section was deparaffinized, rehydrated and washed with cacodylate buffer. Upon O/N fixation in 2% glutaraldehyde plus 2% paraformaldehyde, it was post-fixed with 1% osmium tetroxide and transferred in an Epon capsule. One-hundred-nm sections were cut and placed on carbon/formvar-coated copper grids, poststained in 7% uranyl acetate/lead citrate, and analyzed with a Tecnai-12 electron microscope (Model Eagle, Fei Company, The Netherlands).

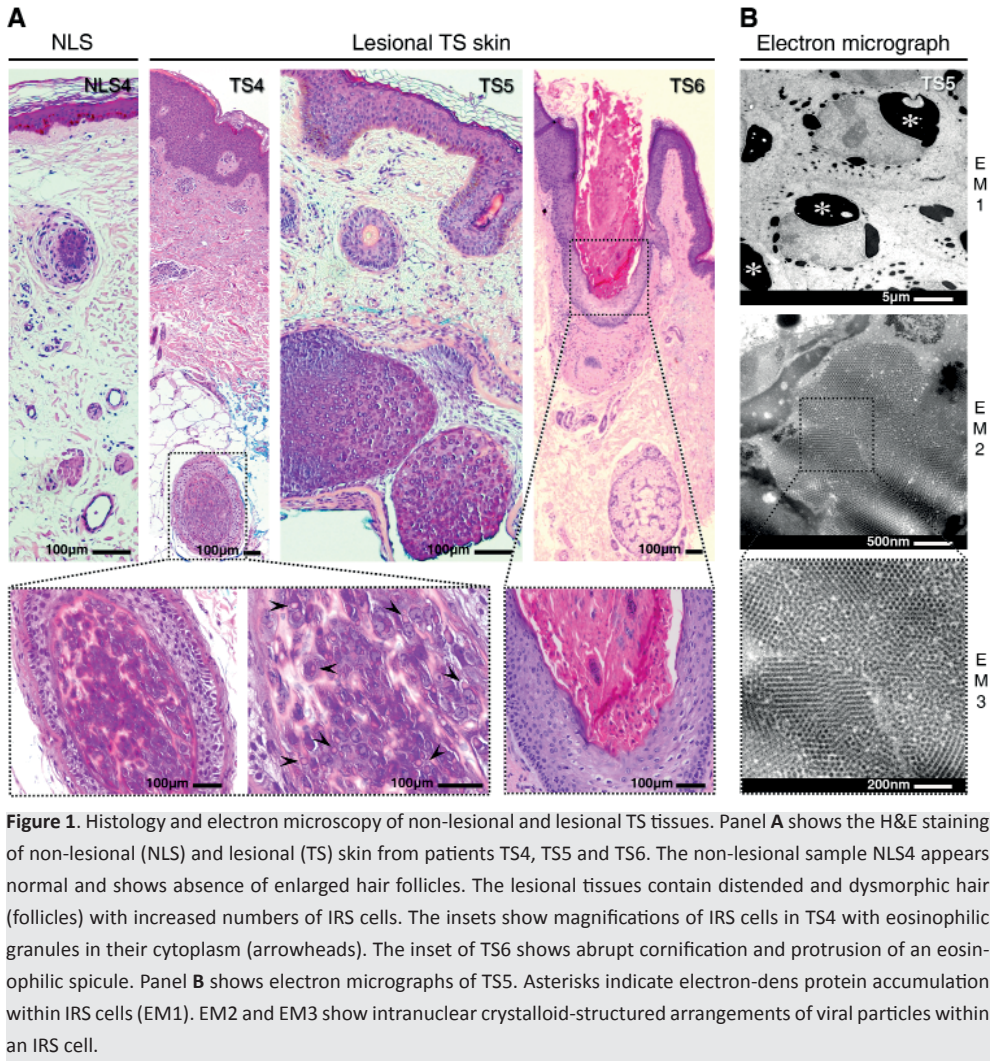
## Results

### Confirmation of trichodysplasia spinulosa in TS4, TS5 and TS6

TS-compatible clinical and histological findings were previously described for patients TS1 - TS3 and TS7 - TS11, including spicules on the face, distention of hair follicles and increase of trichohyalin-overexpressing IRS cells [1, 4 - 7, 9 - 11]. To verify TS in the unpublished cases (**Table 1**), histology was performed on TS4, TS5 and TS6. H&E staining of these samples showed enlarged hair follicles with numerous IRS cells containing large cytoplasmic eosinophilic granules (**Figure 1A**). The non-lesional samples did not display these histological features. In ultra-thin lesional sections of TS5, we confirmed the presence of intranuclear 38 - 42-nm virus particles in nuclei of IRS cells (**Figure 1B**) [10, 14]. These cells also contained electron-dense cytoplasmic protein accumulations, probably trichohyalin [5, 15, 16].

### Presence of TSPyV DNA in TS patients and controls

The presence of TSPyV DNA was assessed by PCR in the lesional and non-lesional TS samples, and in the healthy control samples. TSPyV DNA was detected in all lesional samples (11 of 11, 100%), in one of two non-lesional samples, and in 6 out of 249 healthy controls (2%) (**Table 2**). This revealed a statistically significant association between the presence of TSPyV DNA and symptomatic TS ( $P < 0.001$ , Chi-square test).



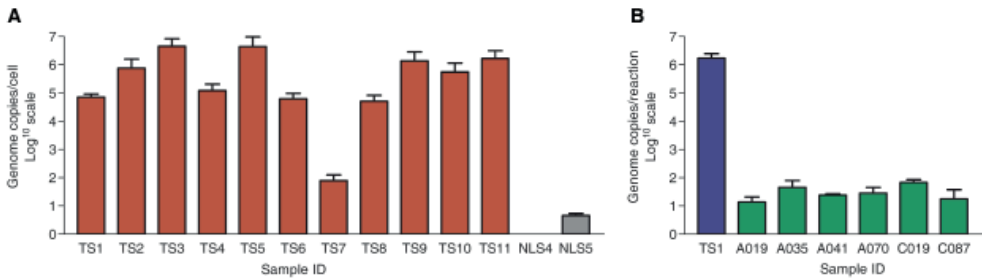
### TSPyV DNA load in TS patients and controls

As a measure of viral replication, we calculated TSPyV DNA load in TSPyV DNA-positive samples. High viral loads above  $10^4$  copies/cell were measured in all, except one, lesional TS samples (91%), with an average of  $\sim 10^6$  (**Figure 2A**, **Table 2**). The non-lesional TSPyV-positive sample NLS5 contained less than 5 copies/cell, a 1,000,000-fold reduction compared to the lesional sample from the same patient (TS5, **Figure 2A**). In the six TSPyV DNA-positive healthy control skin swabs, the average amount of detectable TSPyV DNA copies was 33, being 10,000–100,000-fold lower compared to the forehead skin swab of TS1 (**Figure 2B**).

**Table 2:** Overview of TSPyV DNA-PCR results

Group	TSPyV DNA-positive	TSPyV DNA-load *
	n (%)	average copy number (range)
Lesional TS samples (n=11)	11 (100)	1226845 (78.0 - 4479917.9)
Non-lesional TS samples (n=2)	1	5
Healthy control samples (n=249)	6 (2)	33 (13.6 - 67.3)

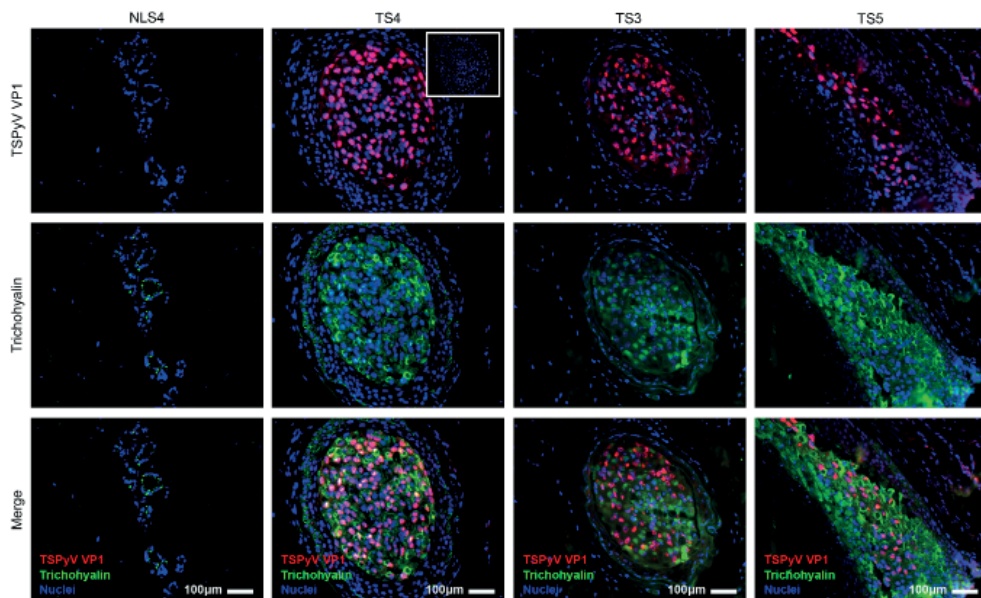
\* Viral copy numbers were calculated per cell for the lesional and non-lesional TS samples, and per PCR reaction for the healthy control samples



**Figure 2.** Quantification of TSPyV DNA in TS patients and healthy controls. Bars in graph **A** illustrate the amount of TSPyV genome copies per cell measured in the lesional (TS1 - TS11) and non-lesional (NLS4 and NLS5) samples. The bars in graph **B** show the amount of TSPyV genome copies per PCR-reaction measured in skin swabs from TS patient 1 (TS1) and from the six TSPyV DNA-positive healthy individuals (A019-C087). The graphs are shown in a Log<sub>10</sub> scale. The error bars represent the standard deviation of the mean viral load calculated with three different quantitative PCRs (NCCR, VP1 and LT).

### Expression and localization of TSPyV protein

To elaborate on the activity of TSPyV infection in TS patients, expression and localization of the TSPyV major capsid protein VP1 was assessed in lesional and non-lesional TS samples. In all of the lesional samples available for this purpose (TS3 - TS5, TS8, TS10 and TS11), TSPyV VP1 was easily detectable within the distended hair follicles with dystrophic morphology (**Figure 3**, and data not shown). In every case, VP1 was observed almost exclusively within the nuclei of IRS cells that overexpressed trichohyalin. In the non-lesional TS samples, no TSPyV VP1 was detected and trichohyalin expression was not increased.



**Figure 3.** TSPyV VP1 protein expression in lesional and non-lesional TS samples. TS non-lesional (NLS4) and lesional (TS3 - TS5) skin sections were co-stained for TSPyV VP1 (red), trichohyalin (green) and chromatin (blue). In the non-lesional sample, no TSPyV VP1 was detected, whereas the lesional samples showed abundant intranuclear VP1 staining of the affected hair follicles (upper row). Staining with a pre-immune rabbit serum collected prior to TSPyV VP1 immunization (inset in TS4), revealed no staining of the lesional samples underscoring the specificity of TSPyV VP1 antiserum. Overexpression of trichohyalin was observed in IRS cells in the lesional samples but not in the non-lesional sample (middle row). Co-localization of both signals was frequently observed in the lesional samples (lower row).

## Discussion

In previously published reports, different names have been used to describe TS, such as trichodysplasia or pilomatrix dysplasia of immunosuppression [7, 15, 17] cyclosporine-induced folliculodystrophy [18] or viral/virus-associated trichodysplasia [1, 5, 9 - 11, 14, 19]. However, despite terminology differences, all cases represented immunocompromized patients that had follicular papules and spiny lesions predominantly on the face histologically characterized by enlarged, dysmorphic hair follicles consisting of IRS cells packed with trichohyalin. By assembling the first representative series of TS samples from almost half of all previously described cases grouped under different names, we provided evidence that these patients indeed suffered from the same disease with TSPyV infection as a shared feature.

When one reasons about the involvement of a certain pathogen in the pathogenesis of a certain disease, the prevalence of the putative pathogen should be considerably higher in the patient group than in healthy controls [20]. As polyomaviruses cause lifelong

latent infections with shedding of small amounts of detectable virus, here the present amount of virus should also be taken into account. TSPyV was detected in all of the lesional TS samples, in one of two non-lesional TS samples and in a minority of controls, resulting in a highly significant statistical association between TSPyV infection and TS. High TSPyV DNA loads were measured exclusively in lesional TS samples indicating that active TSPyV infection with a high rate of replication is only seen in TS-affected skin, suggesting its involvement in the etiology of the disease. Preliminary sequence analysis revealed subtle TSPyV nucleotide substitutions in each of the lesional samples ruling out the possibility of TSPyV plasmid or sample contamination as a source of the positive findings. The observed TSPyV prevalence of 2% in the control swabs is comparable to the 4% prevalence that we measured in plucked eyebrows from a group of asymptomatic immunosuppressed renal transplant patients [4]. This probably reflects latent TSPyV infections, as indicated by serological studies [21, 22]. Unfortunately, we were unable to test the TSPyV DNA status of healthy skin biopsies. However, because comparable load measurements were obtained with the healthy skin swabs and the non-lesional biopsies, and a high TSPyV load was measured in the skin swab of TS1, we feel that skin swabs provide an appropriate sample for TSPyV analysis. Whether TSPyV load measurements in non-invasive skin swabs can be used as a TS diagnostic tool needs further investigation.

In every positive sample, the presence of TSPyV DNA was detected with all three PCRs, each located in a different region of the TSPyV genome (LT, VP1 and NCCR). With all of them comparable viral loads were measured indicating that, if detectable, the whole TSPyV genome was present. Host genomic integration of the viral DNA, as has been shown for MCPyV [23], was not investigated in this study but seems unlikely considering the high TSPyV loads that were measured.

Another indication that active TSPyV infection is involved in the development of TS was found in the presence of VP1 in all lesional TS samples tested, especially because the staining was restricted to the trichohyalin-overexpressing IRS cells that are known to be hyperproliferative in TS and are held responsible for the growth of the characteristic spines [4]. Taken together, with the TSPyV DNA prevalence and load data, despite the limited study size, these findings provide strong evidence that active TSPyV infection is associated with TS and probably involved in the pathogenesis of TS. Further *in vitro* studies are needed to confirm this hypothesis and to reveal the pathogenic mechanisms involved. Furthermore, the use of antiviral treatment, for example with topical cidofovir 1% cream that proved successful in treating some TS patients in the past [4, 10, 18], deserves serious consideration.

## Acknowledgments

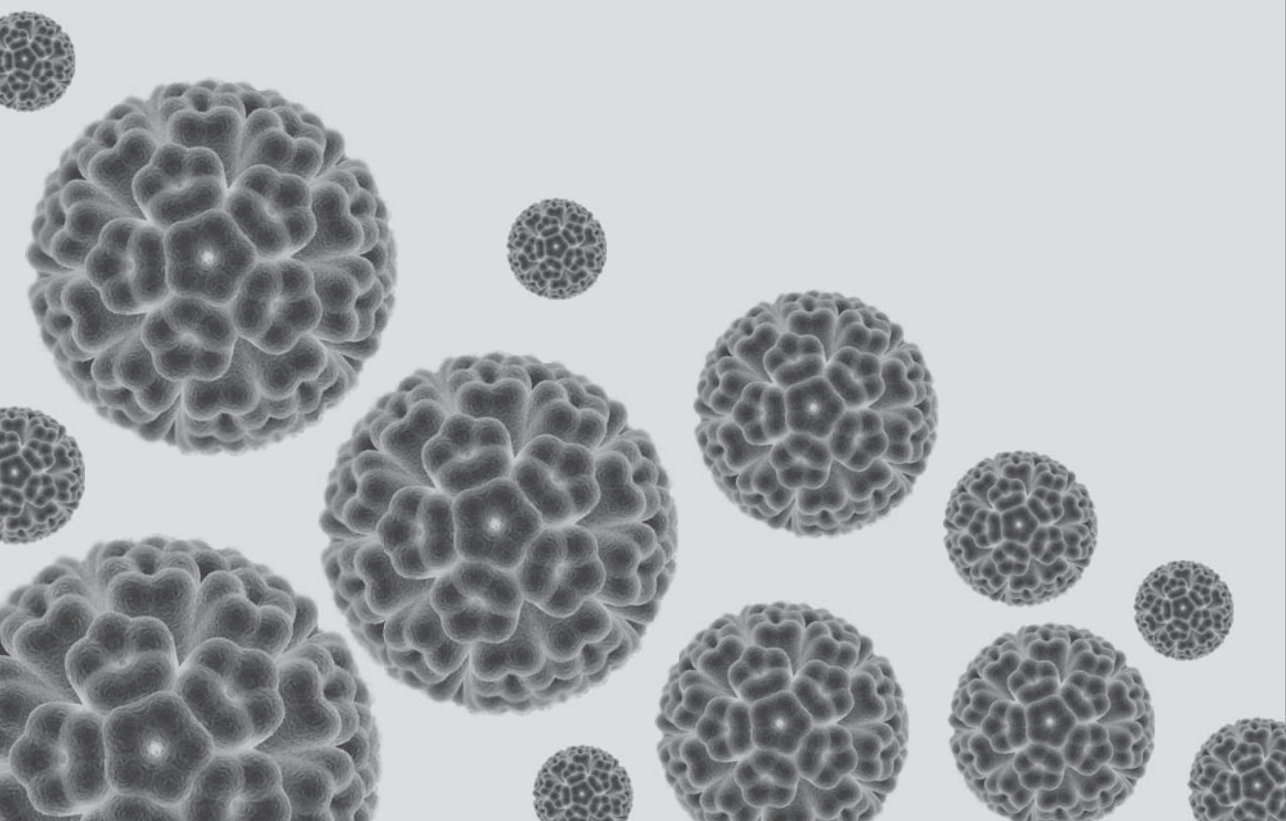
Mieke Mommaas is thanked for her assistance with the EM analysis. Aleodor Andea and Zende Elaba are thanked for the histological pictures of TS6. Marjolein Knoester is thanked for advice regarding the statistics. We also thank the participants and conductors of the WARTS-1 study and DDL Diagnostic Laboratory, Rijswijk, The Netherlands.



## References

1. Matthews MR, Wang RC, Reddick RL, Saldivar VA, Browning JC. Viral-associated trichodysplasia spinulosa: a case with electron microscopic and molecular detection of the trichodysplasia spinulosa-associated human polyomavirus (2011) *J. Cutan. Pathol.* 38: 420-431.
2. Blake BP, Marathe KS, Mohr MR, Jones N, Novosel TA. Viral-Associated Trichodysplasia of Immunosuppression in a Renal Transplant Patient (2011) *J. Drugs Dermatol.* 10: 422-424.
3. Kazem S, van der Meijden E, Feltkamp MC. The trichodysplasia spinulosa-associated polyomavirus: virological background and clinical implications (2013) *APMIS* 121: 770-782.
4. van der Meijden E, Janssens RW, Lauber C, Bouwes Bavinck JN, Gorbalenya AE, Feltkamp MC. Discovery of a new human polyomavirus associated with trichodysplasia spinulosa in an immunocompromized patient (2010) *PLoS Pathog.* 6: e1001024.
5. Haycox CL, Kim S, Fleckman P, Smith LT, Piepkorn M, Sundberg JP, Howell DN, Miller SE. Trichodysplasia spinulosa--a newly described folliculocentric viral infection in an immunocompromised host (1999) *J. Investig. Dermatol. Symp. Proc.* 4: 268-271.
6. Schwieger-Briel A, Balma-Mena A, Ngan B, Dipchand A, Pope E. Trichodysplasia spinulosa--a rare complication in immunosuppressed patients (2010) *Pediatr. Dermatol.* 27: 509-513.
7. Holzer AM, Hughey LC. Trichodysplasia of immunosuppression treated with oral valganciclovir (2009) *J. Am. Acad. Dermatol.* 60: 169-172.
8. Bruggink SC, Gussekloo J, Berger MY, Zaijier K, Assendelft WJ, de Waal MW, Bavinck JN, Koes BW, Eekhof JA. Cryotherapy with liquid nitrogen versus topical salicylic acid application for cutaneous warts in primary care: randomized controlled trial (2010) *CMAJ.* 182: 1624-1630.
9. Sadler GM, Halbert AR, Smith N, Rogers M. Trichodysplasia spinulosa associated with chemotherapy for acute lymphocytic leukaemia (2007) *Australas. J. Dermatol.* 48: 110-114.
10. Wyatt AJ, Sachs DL, Shia J, Delgado R, Busam KJ. Virus-associated trichodysplasia spinulosa (2005) *The American Journal of Surgical Pathology* 29: 241-246.
11. Benoit T, Bacelieri R, Morrell DS, Metcalf J. Viral-associated trichodysplasia of immunosuppression: report of a pediatric patient with response to oral valganciclovir (2010) *Arch. Dermatol.* 146: 871-874.
12. Zhao M, Rosenbaum E, Carvalho AL, Koch W, Jiang W, Sidransky D, Califano J. Feasibility of quantitative PCR-based saliva rinse screening of HPV for head and neck cancer (2005) *Int. J. Cancer* 117: 605-610.

13. Lighezan R, Baderca F, Alexa A, Iacovliev M, Bonte D, Murarescu ED, Nebunu A. The value of the reprocessing method of paraffin-embedded biopsies for transmission electron microscopy (2009) *Rom. J. Morphol. Embryol.* 50: 613-617.
14. Osswald SS, Kulick KB, Tomaszewski MM, Sperling LC. Viral-associated trichodysplasia in a patient with lymphoma: a case report and review (2007) *J. Cutan. Pathol.* 34: 721-725.
15. Chastain MA, Millikan LE. Pilomatrix dysplasia in an immunosuppressed patient (2000) *J. Am. Acad. Dermatol.* 43: 118-122.
16. Lee JS, Frederiksen P, Kossard S. Progressive trichodysplasia spinulosa in a patient with chronic lymphocytic leukaemia in remission (2008) *Australas. J. Dermatol.* 49: 57-60.
17. Campbell RM, Ney A, Gohh R, Robinson-Bostom L. Spiny hyperkeratotic projections on the face and extremities of a kidney transplant recipient (2006) *Arch. Dermatol.* 142: 1643-1648.
18. Heaphy MR, Jr., Shamma HN, Hickmann M, White MJ. Cyclosporine-induced folliculodystrophy (2004) *J. Am. Acad. Dermatol.* 50: 310-315.
19. Sperling LC, Tomaszewski MM, Thomas DA. Viral-associated trichodysplasia in patients who are immunocompromised (2004) *J. Am. Acad. Dermatol.* 50: 318-322.
20. Fredericks DN, Relman DA. Sequence-based identification of microbial pathogens: a reconsideration of Koch's postulates (1996) *Clin. Microbiol. Rev.* 9: 18-33.
21. van der Meijden E, Kazem S, Burgers MM, Janssens R, Bouwes Bavinck JN, de Melker H, Feltkamp MC. Seroprevalence of Trichodysplasia Spinulosa-associated Polyomavirus (2011) *Emerg. Infect. Dis.* 17: 1355-1363.
22. Chen T, Mattila PS, Jartti T, Ruuskanen O, Soderlund-Venermo M, Hedman K. Seroepidemiology of the Newly Found Trichodysplasia Spinulosa-Associated Polyomavirus (2011) *J. Infect. Dis.* 204: 1523-1526.
23. Feng H, Shuda M, Chang Y, Moore PS. Clonal integration of a polyomavirus in human Merkel cell carcinoma (2008) *Science* 319: 1096-1100.



# Chapter 4

## Immunohistopathology of trichodysplasia spinulosa

*Adapted from\*:*

### **Polyomavirus-associated trichodysplasia spinulosa involves hyperproliferation, pRB phosphorylation and upregulation of p16 and p21**

*Authors:*

Siamaque Kazem<sup>1</sup>  
Els van der Meijden<sup>1</sup>  
Richard Wang<sup>2</sup>  
Arlene Rosenberg<sup>3</sup>

Elena Pope<sup>4</sup>  
Taylor Benoit<sup>5</sup>  
Philip Fleckman<sup>6</sup>  
Mariet Feltkamp<sup>1</sup>

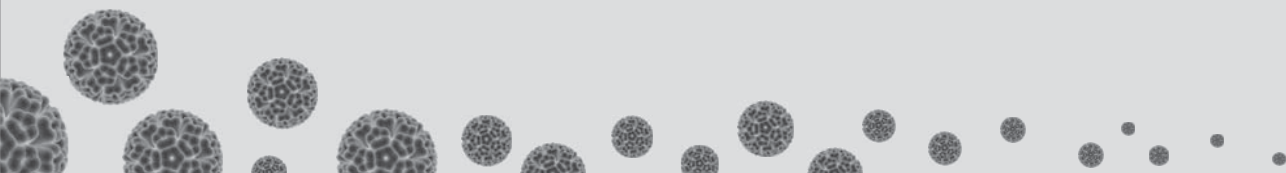
*Affiliations:*

<sup>1</sup> Department of Medical Microbiology, Leiden University Medical Center, Leiden, The Netherlands, <sup>2</sup> Department of Dermatology, UT Southwestern Medical Center, Dallas, USA, <sup>3</sup> Department of Dermatology, MetroHealth Medical Center, Case Western Reserve University, Cleveland, OH, USA, <sup>4</sup> Department of Pathology, Memorial Sloan-Kettering Cancer Center, New York, USA, <sup>5</sup> Department of Dermatology, Medical University of South Carolina, Charleston, SC, USA, <sup>6</sup> Department of Dermatology, Medical University of South Carolina, Charleston, SC, USA

*Published in original form:*

PLoS ONE, 2014 (9) e108947

*\* Note: Adaptation of this chapter from the original published article concerns only minor textual adjustments. Only in the introduction part, the overlapping information with previous chapters is withdrawn.*



## Abstract

Prior data suggested that hair follicle cells in trichodysplasia spinulosa are highly proliferative, but the underlying pathogenic mechanism imposed by TSPyV infection has not been solved yet. By analogy with other polyomaviruses, such as the Merkel cell polyomavirus associated with Merkel cell carcinoma, we hypothesized that TSPyV T-antigen promotes proliferation of infected IRS cells. Therefore, we analyzed TS biopsy sections for markers of cell proliferation (Ki-67) and cell cycle regulation (p16<sup>ink4a</sup>, p21<sup>waf</sup>, pRB, phosphorylated pRB), and the putatively transforming TSPyV early large tumor (LT) antigen. Intense Ki-67 staining was detected especially in the margins of TS hair follicles, which colocalized with TSPyV LT-antigen detection. In this area, staining was also noted for pRB and particularly phosphorylated pRB, as well as p16<sup>ink4a</sup> and p21<sup>waf</sup>. Healthy control hair follicles did not or hardly stained for these markers. Trichohyalin was particularly detected in the center of TS follicles that stained negative for Ki-67 and TSPyV LT-antigen. In summary, we provide evidence for clustering of TSPyV LT-antigen-expressing and proliferating cells in the follicle margins that overproduce negative cell cycle regulatory proteins. These data are compatible with a scenario of TSPyV T-antigen-mediated cell cycle progression, potentially creating a pool of proliferating cells that enable viral DNA replication and drive papule and spicule formation.

16  
Abstract

## Introduction

It is well known that best-studied polyomaviruses, such as SV40 and Merkel cell polyomavirus (MCPyV), encode proteins that counteract host regulatory cellular factors and induce cellular transformation [1, 2]. The polyomavirus large tumor (LT) antigen is generally considered the most potent viral transforming protein that revokes many functions of cellular factors, to the benefit of the virus life cycle [3, 4]. One of its important functions is to induce cell cycle progression by inactivation of the retinoblastoma protein family members (e.g., pRB) [5]. Through its conserved LXCXE motif, LT-antigen interacts with pRB tumor-suppressor protein and deprives it from its cell cycle inhibitory function by inducing pRB hyperphosphorylation [6, 7].

This mechanism of cell cycle deregulation is not exclusive to polyomaviruses, as other DNA viruses like human papillomavirus 16 (HPV16) exploit similar regulatory function through the viral oncogenes E6 and E7 [8, 9]. HPV16-positive cervical cancers are highly proliferative as a result of pRB cell cycle control inhibition that consequently leads to p16<sup>ink4a</sup> and p21<sup>waf</sup> overexpression, and Cyclin-D1 downregulation [10]. p16<sup>ink4a</sup> and p21<sup>waf</sup> are inhibitors of cyclin dependent kinases (CDK), such as CDK4 and CDK6 that promote pRB phosphorylation and G1 to S phase cell cycle transition [11].

Elaborating on a putative interaction between LT-antigen and pRB, in order to explain the proliferative nature of TS, we sought immunohistological evidence of TSPyV LT-antigen-induced hyperproliferation of TS-affected hair follicles. Within a representative set of archived TSPyV DNA-positive TS sections, proliferation, differentiation and cell cycle progression were assessed by analyzing the presence of cell cycle regulation and proliferation markers Ki-67, pRB, p16<sup>ink4a</sup> and p21<sup>waf</sup>. Staining patterns of these markers were correlated with detection of TSPyV LT-antigen and trichohyalin locally, as markers for viral infection and TS disease. The observed staining patterns that indicate disruption of the follicular cell cycle pathway are discussed with regard to the underlying disease mechanism, possibly involving TSPyV LT-antigen, and with regard to histological and clinical symptoms of TS.

## Materials and Methods

### Patients and materials

A set of six formalin-fixed paraffin-embedded (FFPE) TS lesional skin biopsies was retrieved as described previously (**Table 1**) [12]. The FFPE sections of healthy skin biopsies from three healthy donors were used as negative (normal) staining controls. These skin samples were collected after informed written consent and handled according to the declaration of Helsinki principles [13]. As a positive staining control for assessment of cellular proliferation and transformation, sections of previously generated human papillomavirus 16 (HPV16) E6/E7 organotypic raft cultures were used [14]. In brief, these organotypic raft cultures were produced using a dermal-like 3T3-fibroblast-containing collagen-gel matrix that was seeded

with primary human keratinocytes (PHK) stably expressing HPV16 E6/E7 proteins from the plasmid pLZRS [15, 16]. After 10 days in culture, the organotypic raft cultures were fixed with paraformaldehyde, processed for embedding in paraffin and sectioned afterwards.

**Table 1:** List of analyzed TS samples

Case ID <sup>Ref</sup>	Age	Sex	History	Collected	Country	TSPyV load *
TS4 <sup>[12]</sup>	5	Male	Kidney Tx	2009	USA	1.2E+05
TS5 <sup>[12]</sup>	63	Female	Heart Tx	2010	USA	4.4E+06
TS8 <sup>[17]</sup>	5	Female	Heart Tx	2007	Canada	5.1E+04
TS10 <sup>[18]</sup>	5	Male	Heart Tx	2008	USA	5.6E+05
TS11 <sup>[19]</sup>	43	Male	Kidney Tx	1997	USA	1.7E+06
TS13 <sup>**</sup>	43	Female	Kidney Tx	2012	USA	2.1E+04

\*, Viral copies per cell measured as described (all) and reported (TS4-TS11) by Kazem *et al.* [1].

\*\*, TS13 concerned a kidney transplantation patient immunosuppressed with tacrolimus, mycophenolate, and prednisone. TS was diagnosed 10 months after the rash was noted. Symptoms improved after reduction of immunosuppression and remained absent ever since.

## Histology and immunofluorescence analysis

Four  $\mu\text{m}$  paraffin sections were cut for histological and marker-specific immunofluorescence (IFA) analyses. The sections were heated overnight at 60°C on glass slides and the next day deparaffinized in xylene and rehydrated through descending grades of ethanol to distilled water. Slides for histological assessment were directly stained with Hematoxylin and Eosin (H&E). Sections for IFA analysis were subjected to antigen retrieval in citrate-buffer. After blocking, the sections were incubated overnight with the primary antibodies, listed in **Table 2**, in a moist chamber at 4°C. The day after, the slides were incubated with secondary antibodies at dilutions indicated in the table, and supplemented with Hoechst for DNA staining. Slides were kept in dark and analyzed under a fluorescence microscope and representative pictures were taken with Axiovision software (Carl ZEISS Vision, USA).

## Results

### General histological skin features of trichodysplasia spinulosa

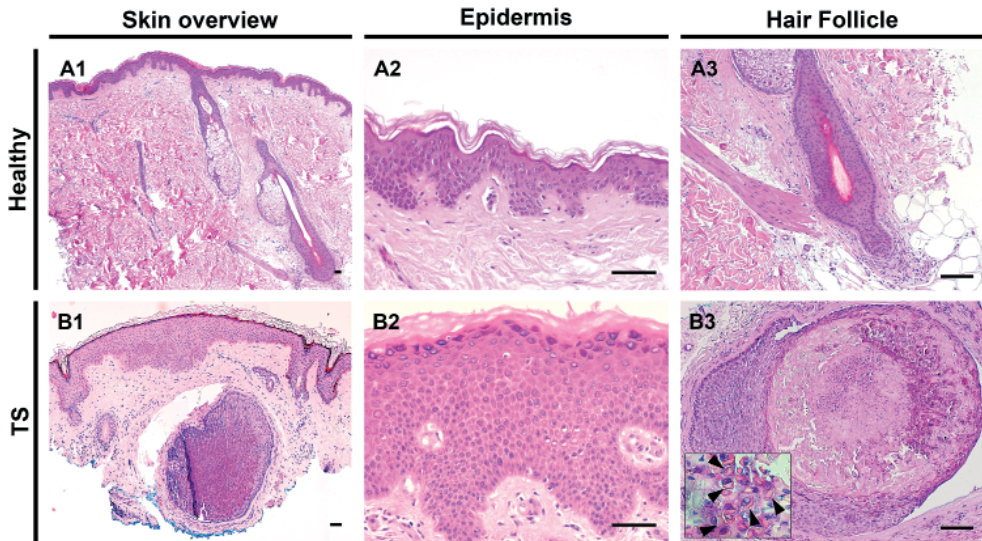
To start, the dermis, epidermis and hair follicles of TS-affected and healthy skin were compared. H&E staining of healthy skin sections demonstrated normal, slim, hair follicles and epidermal stratification (**Figure 1, A1-A3**). In agreement with the literature [12, 17 - 19], in the TS samples enlarged dysmorphic hair follicles were observed (**Figure 1, B1 and B3**). In the center, the TS follicles were inhabited by cells producing eosinophilic protein deposits, possibly trichohyalin protein in the IRS cells. In most TS cases, acanthosis of the epidermis was noted (**Figure 1, B2**).

**Table 2:** Primary and secondary antibodies used for immunofluorescence

Primary antibodies (origin)	Clone	Dilutions	Company
TSPyV LT-antigen (rabbit) *	V5264	1:1000	GenScript, USA
Pre-immune (rabbit)	V5264	1:1000	GenScript, USA
TSPyV VP1-antigen (rabbit) **	V581	1:1000	GenScript, USA
Pre-immune (rabbit)	V581	1:1000	GenScript, USA
Trichohyalin (mouse)	AE15	1:250	Santa Cruz, USA
Trichohyalin (rabbit)	TCHH	1:500	Sigma-Aldrich, USA
Ki-67 (mouse)	MIB-1	1:250	Abcam, USA
p21 <sup>waf</sup> (mouse)	6B6	1:250	BD Biosciences, USA
pRB (mouse)	G3-245	1:250	BD Biosciences, USA
Phospho-pRB (Ser807-811) (rabbit)	D20B12	1:250	Cell Signaling Tech., USA
p16 <sup>ink4a</sup> (mouse)	JC8	1:250	Santa Cruz Biotech, USA
Secondary antibodies (origin)	Clone	Dilutions	Company
Anti-mouse Alexa488-labeled (goat)	A-11001	1:300	Invitrogen, USA
Anti-rabbit Alexa488-labeled (goat)	A-11008	1:300	Invitrogen, USA
Anti-rabbit Cy3-labeled (donkey)	711-165-152	1:1000	Jackson, USA

\*, Rabbit immunized with two TSPyV LT-antigen-derived synthetic peptides FSSQHDVPTQDGRD (AA, 77-90) and NSRRRRRAAPPEDSP (AA, 151-164)

\*\*, Rabbit immunized with TSPyV VP1-antigen-derived synthetic peptide TGNYRTDYSANDKL (AA, 170-183) [12, 20].



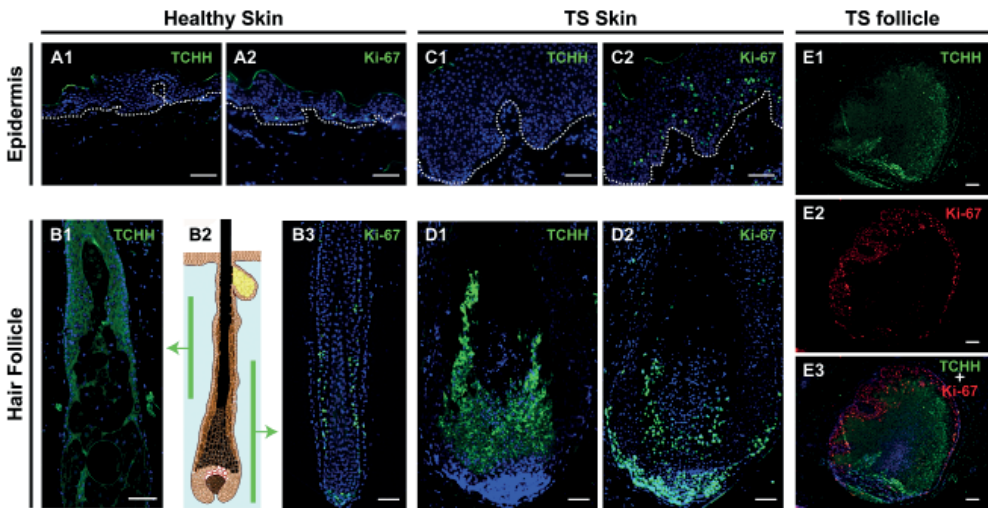
**Figure 1.** Histological features of trichodysplasia spinulosa. Left column illustrates H&E staining of a low power field of healthy skin (A1) and TS lesional skin (B1). High power fields of healthy (A2) and TS (B2) epidermis and hair follicles (A3 and B3). Note the hair follicle shown in B3 is enlarged and dysmorphic, containing eosinophilic granular protein deposits in the cytoplasm of the cells (arrowheads in inset) with abrupt cornification in the center of the follicle. Bars depict 100µm.



## Trichohyalin and Ki-67 staining

To detect the presence of IRS cells and to pinpoint areas of proliferation in the TS-affected tissue, the sections were stained for trichohyalin and Ki-67, respectively. As expected, in healthy skin trichohyalin staining was detected only along the IRS and absent in the epidermis (**Figure 2, A1 and B1**). Ki-67-staining in healthy skin was restricted to the epidermal basal layer, and to the follicle bulb and the suprabulbar (stem) area (**Figure 2, A2 and B3**). Positive staining of the top cornified layer of the epidermis observed in some of the stained sections in **Figures 1 and 2** was considered nonspecific.

In the TS sections, excessive amounts of trichohyalin were observed in the affected follicles, whereas the acanthotic epidermis did not stain for trichohyalin (**Figure 2, C1 and D1**). A substantial increase in Ki-67-positive nuclei was observed both in the follicles and in the TS epidermis, in the latter in basal as well as in suprabasal layers. At the follicle base, a significant increase in Ki-67 expression was evident (**Figure 2, C2 and D2**). Trichohyalin and Ki-67 costaining of a cross-sectioned TS follicle showed trichohyalin staining especially in the follicle center, whereas Ki-67 staining was primarily detected at the follicle margins (**Figure 2, E1-E3**).

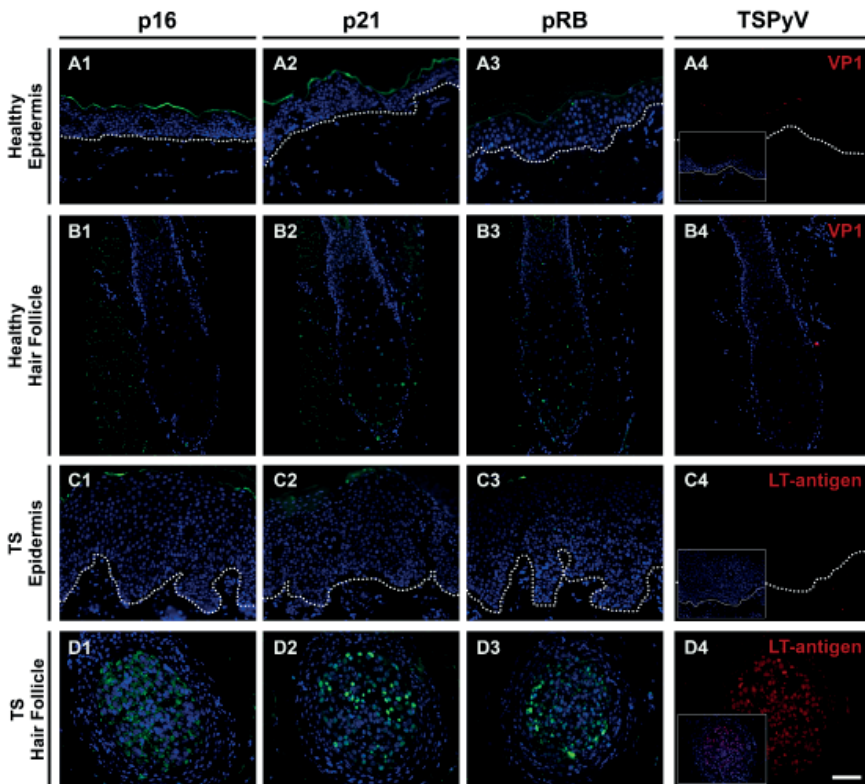


**Figure 2.** Trichohyalin and Ki-67 expression in healthy and lesional skin. This figure illustrates staining of trichohyalin (TCHH) and Ki-67 in healthy (**A and B**) and in TS skin (**C, D and E**). In the left panel, trichohyalin (**A1**) and Ki-67 (**A2**) staining in healthy epidermis and in healthy hair follicles (**B1** and **B2**) are shown, with the corresponding levels of the follicle illustrated in **B2**. In the middle panel, trichohyalin (TCHH) and Ki-67 staining in TS lesional skin are shown with TS epidermis on top (**C1** and **C2**) and vertical sections of TS hair follicle beneath (**D1** and **D2**). Dotted lines indicate the dermoepidermal junction. Costaining for trichohyalin (green) and Ki-67 (red) of a TS follicle cross-sectioned at the suprabulbar region is shown in the right panel (**E1-E3**). Bars depict 100µm.

### Cell cycle regulation markers and TSPyV LT-antigen expression

To explore the nature of the hyperproliferation in the TS-affected skin, as demonstrated by the increased Ki-67 staining, we investigated locally the expression of major cell cycle regulatory proteins p16<sup>ink4a</sup>, p21<sup>waf</sup> and pRB. Sections of HPV16 E6/E7-transformed raft cultures were used as positive staining controls (**Supplementary Figure S1**). Despite occasional faint suprabasal nuclear pRB staining in the TS epidermis, none of these markers were detected in the epidermis of healthy controls or TS cases (**Figure 3, A1-A3 and C1-C3**). A comparable staining pattern was observed in healthy hair follicles, although sometimes faint nuclear staining for p21<sup>waf</sup> and pRB was observed (**Figure 3, B1-B3**).

In the TS-affected hair follicles, expression of p16<sup>ink4a</sup>, p21<sup>waf</sup> and pRB was increased (**Figure 3, D1-D3**). For p16<sup>ink4a</sup>, especially cytoplasmic staining was observed and nuclear staining was seen for p21<sup>waf</sup> and pRB. These analyses were completed by determining the presence of TSPyV. LT-antigen was detected only in affected hair follicles (**Figure 3, D4**), which also stained positive for p16<sup>ink4a</sup>, p21<sup>waf</sup> and pRB, suggestive of colocalization of these markers.

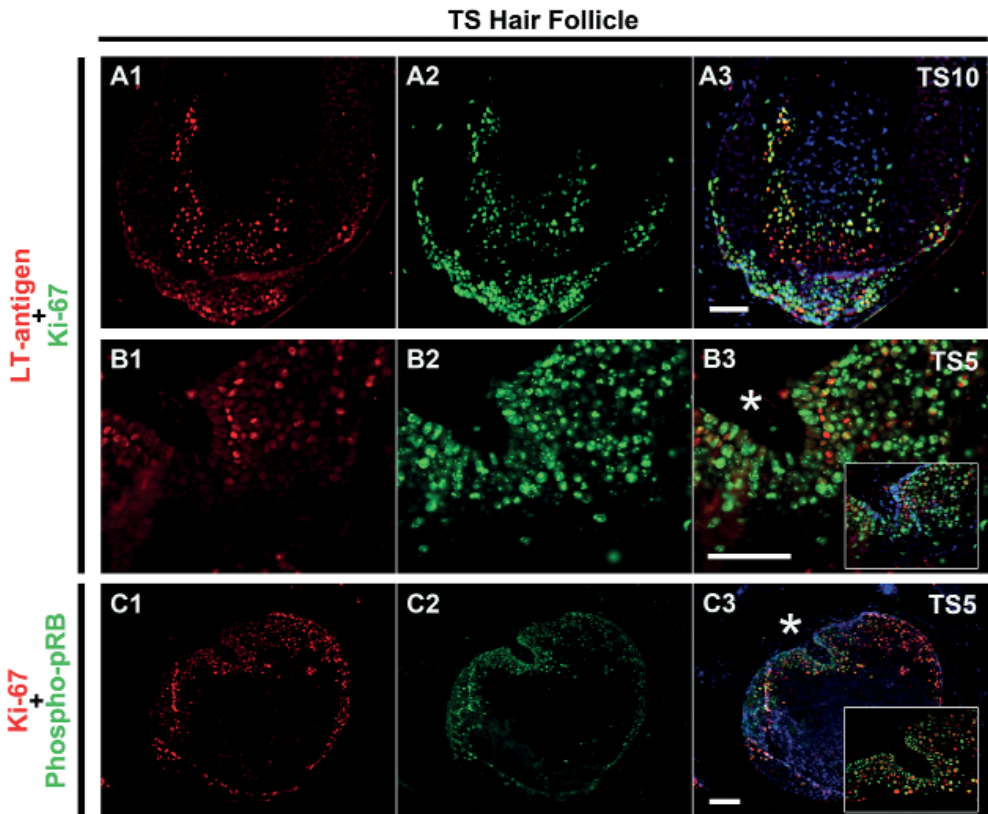


**Figure 3.** Cell cycle regulation markers and TSPyV LT-antigen expression. Sections of healthy epidermis (**A1-A4**), healthy hair follicle (**B1-B4**), TS epidermis (**C1-C4**) and TS follicle (**D1-D4**) are stained for p16<sup>ink4a</sup> (first panel), p21<sup>waf</sup> (second panel), pRB (third panel) and TSPyV (fourth panel). Insets in the fourth panel depict the same region with Hoechst DNA staining (blue). Dotted lines indicate the dermoepidermal junction. Bar depicts 100µm.

### Colocalization of TSPyV LT-antigen, Ki-67 and phosphorylated pRB

Finally, we investigated whether TSPyV LT-antigen expression in the TS sections colocalize with staining of Ki-67 and phosphorylated pRB, in order to explain proliferation induction.

Double staining for Ki-67 and TSPyV LT-antigen of a vertical-sectioned follicle illustrated that Ki-67-positive cells colocalized with TSPyV LT-antigen expression in the margins of the extended bulbar and suprabulbar region (**Figure 4, A1-A3**). The same colocalization was observed in a suprabulbar cross-sectioned hair follicle (**Figure 4, B1-B3**). When analyzing Ki-67 in combination with phosphospecific pRB, we observed colocalization of Ki-67 and phosphorylated pRB in TS hair follicle margins, indicating hyperphosphorylation of pRB in the proliferating cells (**Figure 4, C1-C3**). A summary of all our histological findings using these markers in individual TS-samples is shown in **Table 3**.



**Figure 4.** Colocalization of TSPyV LT-antigen, Ki-67 and phosphorylated pRB. TSPyV LT-antigen (red) (**A1**) with Ki-67 (green) (**A2**) and merge (yellow) (**A3**) in a vertical section of hair follicle of TS10 is shown in the upper row. A higher magnification of TSPyV LT-antigen (**B1**) with Ki-67 (green) (**B2**) and merge (yellow) (**B3**) in a suprabulbar cross-sectioned hair follicle region of TS5 is shown in the middle. **B3** inset depicts the same region with Hoechst DNA staining (blue). Ki-67 (red) (**C1**) with phosphospecific (Ser807/811) pRB (green) (**C2**) and merge (yellow) (**C3**) in a suprabulbar cross-sectioned hair follicle region of TS5 is shown in the last row. A higher magnification of TS5 margin (**C3**, asterisk) is shown in the inset. Dotted lines indicate the dermoepidermal junction. Bars depict 100µm.

**Table 3:** Overview of cellular and virus markers detected in TS lesions

Case ID	Cellular Markers						Virus Markers	
	TCHH	p16	p21	pRB	Phospho-pRB	Ki-67	TSPyV VP1 *	TSPyV LT
TS4	+	+	+	+	+	+	+	+
TS5	+	+	+	+	+	+	+	+
TS8	+	NA	+	NA	NA	+	+	+
TS10	+	NA	+	NA	+	+	+	+
TS11	+	NA	+	NA	NA	+	+	+
TS13	NA	NA	+	NA	NA	+	NA	+

\*, Reported by Kazem *et al.* [12]

-, negative; +, positive; NA, not available

## Discussion

In a previous study, in a group of 11 TS patients we established that presence and high load of TSPyV DNA was strongly associated with TS disease [12]. In the same sample-set we showed that viral capsid protein (VP1) expression was exclusively present in distended dysmorphic trichohyalin-positive TS hair follicles [12]. In this first systematic immunohistochemical study of its kind, we analyzed lesional sections from six of these TS patients, which is roughly one fifth of all TS cases reported worldwide [21]. Still, we were unable to perform every staining on all patients, because we were limited in the number of TS sections available for analysis.

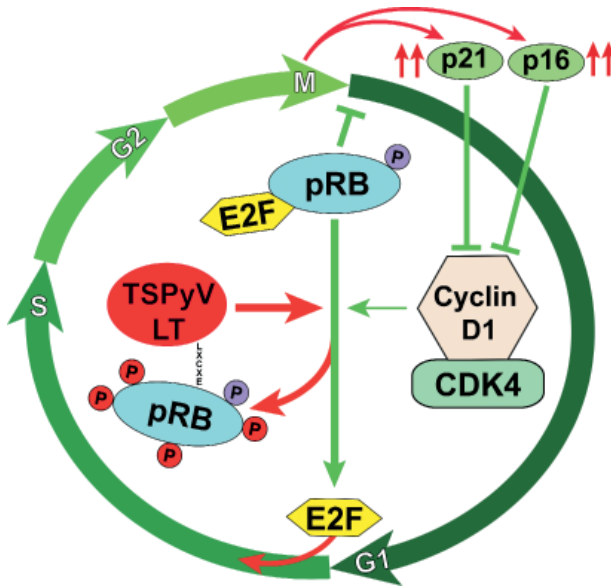
Intense Ki-67 staining was detected in the TS-affected hair follicles, especially in the bulbar and suprabulbar marginal regions, indicative of hyperproliferation in these areas. The observed pattern of Ki-67-rich and trichohyalin-poor follicle margins and trichohyalin-rich and Ki-67-poor follicle centers may suggest arrest of proliferation in ‘mature’ IRS cells along central, terminal differentiation of these cells. However, we cannot exclude that different (IRS) [22] cell types explain the difference in margin and center staining.

In addition to increased proliferation of follicular cells, we observed increased Ki-67-staining within the overlying acanthotic epidermis of every TS-patient analyzed. This was observed in particular in the suprabasal layers. To our knowledge, this observation has not been reported previously and possibly implies that TS is not confined to the hair follicles but involves other parts of the skin as well. Whether the observed epidermal hyperproliferation is seen only in the vicinity of a TS lesion, with visible papules and/or spicules, or represents a general feature of TS-patients is not known. In the affected epidermis, we could not detect TSPyV LT-antigen (and VP1-antigen, as analyzed previously [12] (and data not shown)). Therefore, the relationship between epidermal hyperproliferation and acanthosis, and TSPyV infection remains unclear.

Subsequent analyses of the proliferative hair follicles for p16<sup>ink4a</sup>, p21<sup>waf</sup> and pRB demonstrated a pattern that is known for tissues infected by small DNA viruses involved in cellular transformation, such as SV40 and HPV16 [23]. The staining pattern was indeed exemplified by the staining pattern of the HPV16 E6/E7-expressing organotypic skin cul-

tures included as a positive control, where increased Ki-67 staining was seen together with increased detection of p16<sup>ink4a</sup> and p21<sup>waf</sup> (**Supplementary Figure S1**). This observed association suggests that a comparable disruption of the pRB-dependent cell cycle regulation pathway may be involved in TSPyV infection and TS development.

Retinoblastoma family proteins (i.e., pRB, p107 and p130) are important regulators of the G1- (rest) to S-phase (DNA synthesis) transition of cells during cell cycle progression. For instance, hypophosphorylated pRB inhibits function of the E2F transcription factor that regulates gene expression required for DNA synthesis. Hyperphosphorylation of pRB by complexes of Cyclin-D and cyclin-dependent kinases (CDKs) results into pRB–E2F complex dissociation and cell-cycle entry, which is reverted by several inhibitory proteins such as p16<sup>ink4a</sup> and p21<sup>waf</sup> [24]. Demonstration of phosphorylated pRB colocalized with Ki-67 indicates pRB-inactivation in the TS lesions and suggests progression into the S-phase (**Figure 5**). The observed increased expression of p16<sup>ink4a</sup> and p21<sup>waf</sup>, factors that normally inhibit Cyclin-D1/CDK activity, is explained as a negative feedback mechanism to inhibit cell cycle progression [24]. Unfortunately, because of lack of additional TS lesional samples/sections, we were unable to look into other cell cycle-regulatory pathways, for example involving p53, which can be revoked by polyomaviruses as well [5, 25 - 27].



**Figure 5.** A hypothetical scenario of TSPyV LT-antigen interference in cell cycle regulation. An oversimplified cell cycling scenario is shown that envisions the TSPyV LT-antigen involvement in regulation of pRB pathway activity. In a normal physiological condition, hypophosphorylated pRB is complexed with transcription factor E2F during early G1 (rest) phase of the cell cycle. When pRB is hyperphosphorylated at specific residues by Cyclin-dependent kinases (CDK) coupled to Cyclin-D1, E2F is released that activates expression of growth stimulatory genes needed for the cells to enter the S (DNA synthesis) phase. pRB phosphorylation is under tight regulation of p16<sup>ink4a</sup> and p21<sup>waf</sup>. Hypothetically, through its conserved LXCXE motif TSPyV LT-antigen interacts with pRB/E2F complex to dissociate these proteins via pRB hyperphosphorylation, resulting into S phase entry and subsequent increased expression of p16<sup>ink4a</sup> and p21<sup>waf</sup> as a negative cell cycling feedback (red arrows).

The observed pattern of cell cycle deregulation and S-phase progression through hyperphosphorylation/inactivation of pRB is also seen in HPV16-induced cervical dysplasia and neoplasia [8]. In that case, phosphorylation of pRB is mediated by the E7 viral oncoprotein, comparable to the action of LT-antigen of SV40 [23]. Crucial to the inactivation of pRB is binding by LT-antigen and E7 through a conserved LXCXE motif found in these viral oncoproteins, and in TSPyV LT-antigen as well [3, 28]. Whether TSPyV LT-antigen interacts with pRB and hampers its function, for instance by hyperphosphorylation, requires experimental confirmation. Especially, since in our analyses technical limitations (shared origin of the antibody, **Table 2**) prevented discrimination between TSPyV LT-antigen and phosphorylated pRB, and therefore, the ability to demonstrate colocalization of both markers. Since the TS follicles were pRB-positive, it is unlikely that TSPyV LT-antigen promotes pRB degradation next to hyperphosphorylation, as is known for HPV16 (**Supplementary Figure S1**).

Taken together, our findings are compatible with a scenario of TSPyV LT-antigen-induced cell cycle progression through disruption of pRB-regulatory pathways, thereby creating a reservoir of proliferating IRS cells that enable viral DNA replication. Terminal differentiation of this large pool of IRS cells could explain the final accumulation of trichohyalin-positive cells and the formation of TS-characteristic spicules.

## Acknowledgements

The authors would like to thank Elsemieke Plasmeijer for her help with collection of healthy skin samples and Abdoel El Ghalbzouri for useful discussions regarding use of antibodies.

## References

1. Johne R, Buck CB, Allander T, Atwood WJ, Garcea RL, Imperiale MJ, Major EO, Ramqvist T, Norkin LC. Taxonomical developments in the family Polyomaviridae (2011) *Arch. Virol.* 156: 1627-1634.
2. Feltkamp MC, Kazem S, van der Meijden E, Lauber C, Gorbalenya AE. From Stockholm to Malawi: recent developments in studying human polyomaviruses (2012) *J. Gen. Virol.* 94: 482-496.
3. Topalis D, Andrei G, Snoeck R. The large tumor antigen: a “Swiss Army knife” protein possessing the functions required for the polyomavirus life cycle (2013) *Antiviral Res.* 97: 122-136.
4. Pipas JM. SV40: Cell transformation and tumorigenesis (2009) *Virology* 384: 294-303.
5. Decaprio JA, Ludlow JW, Figge J, Shew JY, Huang CM, Lee WH, Marsilio E, Paucha E, Livingston DM. SV40 large tumor antigen forms a specific complex with the product of the retinoblastoma susceptibility gene (1988) *Cell* 54: 275-283.
6. Dyson N, Bernards R, Friend SH, Gooding LR, Hassell JA, Major EO, Pipas JM, Vandyke T, Harlow E. Large T antigens of many polyomaviruses are able to form complexes with the retinoblastoma protein (1990) *J. Virol.* 64: 1353-1356.
7. Harris KF, Christensen JB, Radany EH, Imperiale MJ. Novel mechanisms of E2F induction by BK virus large-T antigen: requirement of both the pRb-binding and the J domains (1998) *Mol. Cell Biol.* 18: 1746-1756.
8. Todorovic B, Hung K, Massimi P, Avvakumov N, Dick FA, Shaw GS, Banks L, Mymryk JS. Conserved region 3 of human papillomavirus 16 E7 contributes to deregulation of the retinoblastoma tumor suppressor (2012) *J. Virol.* 86: 13313-13323.
9. Malanchi I, Accardi R, Diehl F, Smet A, Androphy E, Hoheisel J, Tommasino M. Human papillomavirus type 16 E6 promotes retinoblastoma protein phosphorylation and cell cycle progression (2004) *J. Virol.* 78: 13769-13778.
10. Portari EA, Russomano FB, de Camargo MJ, Machado Gayer CR, da Rocha Guillobel HC, Santos-Reboucas CB, Brito Macedo JM. Immunohistochemical expression of cyclin D1, p16Ink4a, p21WAF1, and Ki-67 correlates with the severity of cervical neoplasia (2013) *Int. J. Gynecol. Pathol.* 32: 501-508.
11. Rayess H, Wang MB, Srivatsan ES. Cellular senescence and tumor suppressor gene p16 (2012) *Int. J. Cancer* 130: 1715-1725.
12. Kazem S, van der Meijden E, Kooijman S, Rosenberg AS, Hughey LC, Browning JC, Sadler G, Busam K, Pope E, Benoit T, Fleckman P, de VE, Eekhof JA, Feltkamp MC. Trichodysplasia spinulosa is characterized by active polyomavirus infection (2012) *J. Clin. Virol.* 53: 225-230.
13. van Drongelen V, Danso MO, Mulder A, Mieremet A, van Smeden J., Bouwstra JA, El Ghalbzouri A. Barrier properties of an N/TERT based human skin equivalent (2014) *Tissue Eng. Part A.*

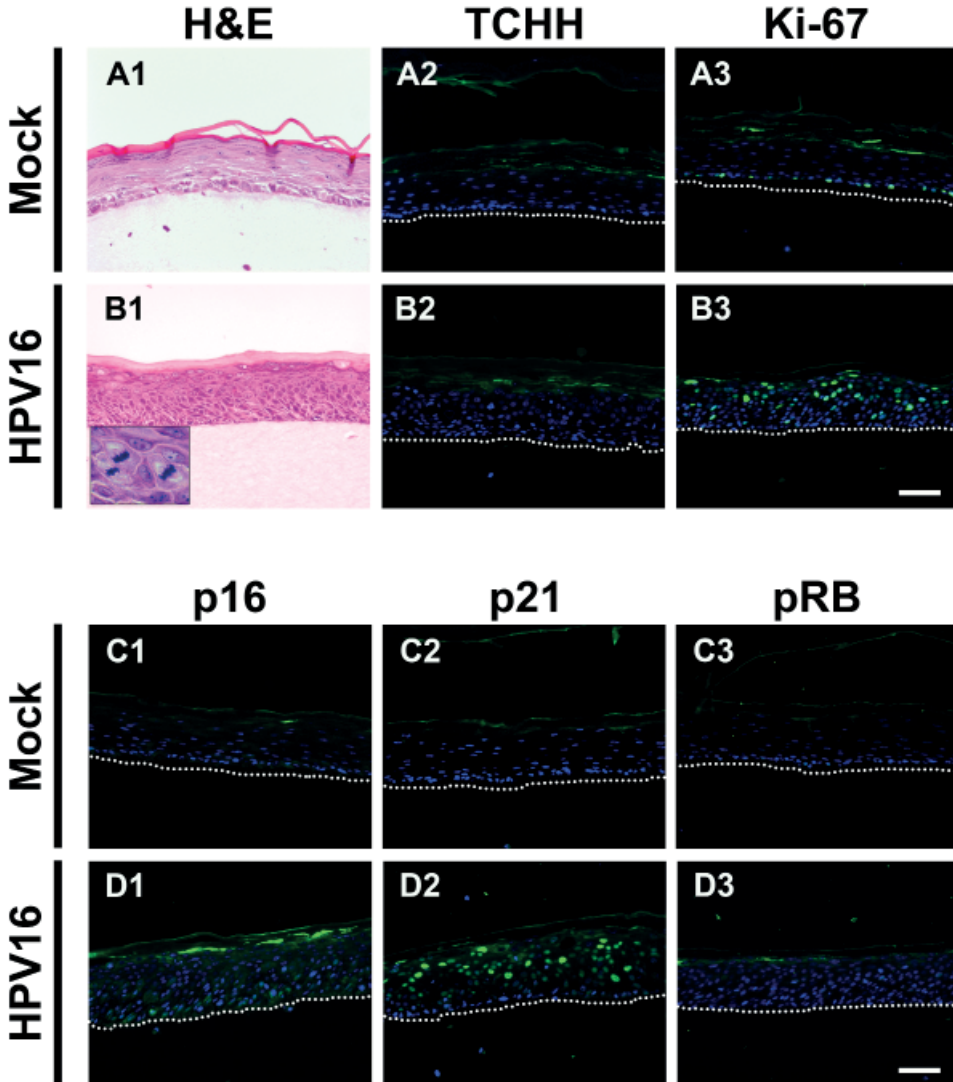
14. Struijk L, van der Meijden E, Kazem S, ter Schegget J, de Gruijl FR, Steenberg RD, Feltkamp MC. Specific betapapillomaviruses associated with squamous cell carcinoma of the skin inhibit UVB-induced apoptosis of primary human keratinocytes (2008) *J. Gen. Virol.* 89: 2303-2314.
15. Boxman IL, Mulder LH, Noya F, de Waard V, Gibbs S, Broker TR, ten Kate F, Chow LT, ter Schegget J. Transduction of the E6 and E7 genes of epidermodysplasia- verruciformis-associated human papillomaviruses alters human keratinocyte growth and differentiation in organotypic cultures (2001) *J. Invest. Dermatol.* 117: 1397-1404.
16. Kazem S, van der Meijden E, Struijk L, de Gruijl FR, Feltkamp MC. Human papillomavirus 8 E6 disrupts terminal skin differentiation and prevents pro-Caspase-14 cleavage (2012) *Virus Res.* 163: 609-616.
17. Schwieger-Briel A, Balma-Mena A, Ngan B, Dipchand A, Pope E. Trichodysplasia spinulosa--a rare complication in immunosuppressed patients (2010) *Pediatr. Dermatol.* 27: 509-513.
18. Benoit T, Bacelieri R, Morrell DS, Metcalf J. Viral-associated trichodysplasia of immunosuppression: report of a pediatric patient with response to oral valganciclovir (2010) *Arch. Dermatol.* 146: 871-874.
19. Haycox CL, Kim S, Fleckman P, Smith LT, Piepkorn M, Sundberg JP, Howell DN, Miller SE. Trichodysplasia spinulosa--a newly described folliculocentric viral infection in an immunocompromised host (1999) *J. Invest. Dermatol. Symp. Proc.* 4: 268-271.
20. Kanitakis J, Kazem S, van der Meijden E, Feltkamp M. Absence of the trichodysplasia spinulosa-associated polyomavirus in human pilomatrixomas (2011) *Eur. J. Dermatol.* 21: 453-454.
21. Kazem S, van der Meijden E, Feltkamp MC. The trichodysplasia spinulosa-associated polyomavirus: virological background and clinical implications (2013) *APMIS* 121: 770-782.
22. Langbein L, Rogers MA, Praetzel S, Winter H, Schweizer J. K6irs1, K6irs2, K6irs3, and K6irs4 represent the inner-root-sheath-specific type II epithelial keratins of the human hair follicle (2003) *J. Invest. Dermatol.* 120: 512-522.
23. Sadasivam S, Decaprio JA. The DREAM complex: master coordinator of cell cycle-dependent gene expression (2013) *Nat. Rev. Cancer* 13: 585-595.
24. Sherr CJ, McCormick F. The RB and p53 pathways in cancer (2002) *Cancer Cell* 2: 103-112.
25. Pipas JM, Levine AJ. Role of T antigen interactions with p53 in tumorigenesis (2001) *Semin. Cancer Biol.* 11: 23-30.
26. Frisque RJ, Hofstetter C, Tyagarajan SK. Transforming activities of JC virus early proteins (2006) *Adv. Exp. Med. Biol.* 577: 288-309.
27. Shivakumar CV, Das GC. Interaction of human polyomavirus BK with the tumor-suppressor protein p53 (1996) *Oncogene* 13: 323-332.



28. van der Meijden E, Janssens RW, Lauber C, Bouwes Bavinck JN, Gorbalenya AE, Feltkamp MC. Discovery of a new human polyomavirus associated with trichodysplasia spinulosa in an immunocompromized patient (2010) PLoS Pathog. 6: e1001024.

## Supplementary data

## Supplementary figure



**Supplementary Figure S1.** Organotypic raft cultures used as staining controls. H&E staining (A1 and B1), trichohyalin staining (TCHH) (A2 and B2) and Ki-67 staining (A3 and B3) in organotypic raft cultures expressing empty vector (pLZRS) (Mock) or HPV16 oncogenes E6/E7 are shown in the upper group of figures. Note many suprabasal mitotic cells in B1 (inset). In the lower group of figures, staining for cell cycle regulatory proteins, p16<sup>ink4a</sup> (C1 and D1), p21<sup>waf</sup> (C2 and D2) and pRB (C3 and D3) in Mock rafts and HPV16 rafts are shown. Some (secondary antibody) nonspecific staining of the cornified layer was present in all materials tested in this study. The dermoepidermal junction is indicated by dotted lines. Bar depicts 100µm.

VePyV1 CaPyV JCPyV *mPyV* SA12 RacPyV GHPyV BKPyV

*BatsgU* CPyV OrAPyV1 FPyV MptV APPyV1 SqPyV LPyV CSLPyV

EPyV PRPyV1 APP<sub>g</sub>V2 KIPyV **MIRgV** PtvPyV2c OtPyV1 **STLLPyV**

MFPyV1 *KSgU9* SV40 TSPyV CoPyV1 PPPyV CPPyV HPyV12

MXPyV PtvPyV1a PDPyV EIPyV1 AtPPyV1 HaPyV (TggPyV1

CdPyV DRPyV MWPyV APyV CAPyV1 HPyV7 CHPyV MasPyV

WUPyV *HSp<sub>g</sub>V6* BPyV MCPyV OrAPyV2 MMPyV SLPyV HPyV10

VePyV1 CaPyV JCPyV *mPyV* SA12

*BatPyV* CPyV OraPyV1 FPyV MptV APPy

EPyV PRPyV1 APPyV2 KIPyV **MPyV** P

MFPyV1 *HPyV9* SV40 TSPyV CoPyV1

MPyV PtvPyV1a PDPyV EiPyV1 At

CdPyV DRPyV MWPyV APyV CAPyV1

WUPyV *HPyV6* BPyV MCPyV **OraPyV2**

## Part III

### TSPyV host adaptation and evolution

PyV BKPyV

PyV CSLPyV

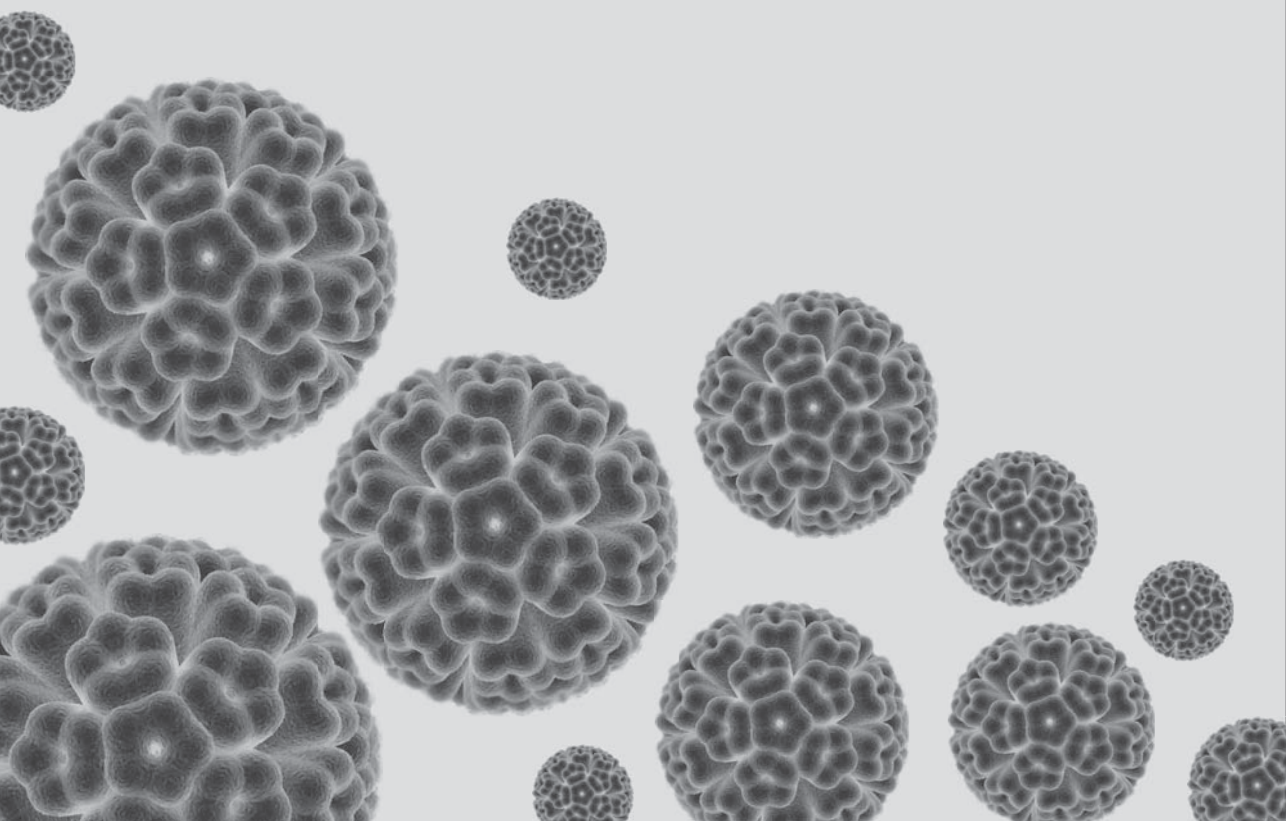
PyV1 STLPyV

PyV HPyV12

PyV GgPyV1

PyV MasPyV

PyV HPyV10



# Chapter 5

## Polyomavirus host adaptation

**Interspecific adaptation by binary choice at *de novo* polyomavirus T antigen site through accelerated codon-constrained Val-Ala toggling within an intrinsically disordered region**

*Authors:*

Chris Lauber<sup>1, 2\*</sup>

Siamaque Kazem<sup>1\*</sup>

Alexander Kravchenko<sup>3†</sup>

Mariet Feltkamp<sup>1</sup>

Alexander Gorbalenya<sup>1, 3, 4</sup>

*\* These authors contributed equally to the paper as first authors*

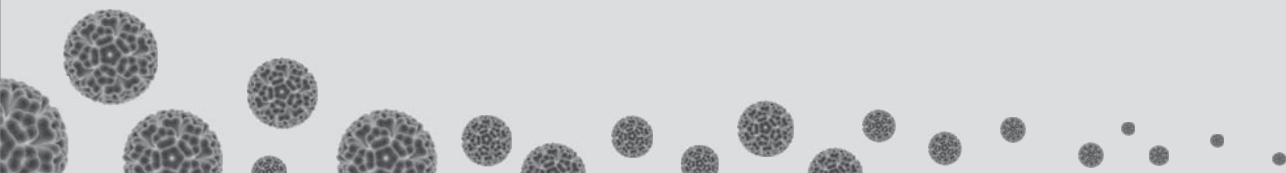
*† Deceased; of blessed memory*

*Affiliations:*

<sup>1</sup> Department of Medical Microbiology, Leiden University Medical Center, Leiden, The Netherlands, <sup>2</sup> Institute for Medical Informatics and Biometry, Technische Universität Dresden, Dresden, Germany, <sup>3</sup> Belozersky Institute of Physico-Chemical Biology, Lomonosov Moscow State University, Moscow, Russia and <sup>4</sup> Faculty of Bioengineering and Bioinformatics, Lomonosov Moscow State University, Moscow, Russia

*Published in:*

Nucleic Acids Research, 2015 (*in-press*)



## Abstract

It is common knowledge that conserved residues evolve slowly. We challenge generality of this central tenet of molecular biology by describing the fast evolution of a conserved nucleotide position that is located in the overlap of two open reading frames (ORFs) of polyomaviruses. The *de novo* ORF is expressed through either the ALTO protein or the Middle T antigen (MT/ALTO), while the ancestral ORF encodes the N-terminal domain of helicase-containing Large T (LT) antigen. In the latter domain the conserved Cys codon of the LXCXE pRB-binding motif constrains codon evolution in the overlapping MT/ALTO ORF to a binary choice between Val and Ala codons, termed here as codon-constrained Val-Ala (COCO-VA) toggling. We found the rate of COCO-VA toggling to approach the speciation rate and to be significantly accelerated compared to the baseline rate of chance substitution in a large monophyletic lineage including all viruses encoding MT/ALTO and three others. Importantly, the COCO-VA site is located in a short linear motif (SLiM) of an intrinsically disordered region, a typical characteristic of adaptive responders. These findings provide evidence that the COCO-VA toggling is under positive selection in many polyomaviruses, implying its critical role in interspecific adaptation, which is unprecedented for conserved residues.

Abstract

## Introduction

Intrinsically disordered regions (IDRs), which are either not structured or may become structured upon interaction with diverse partners [1], have been identified in many proteins and implicated in various biological processes as adaptive responders [2 - 5]. They have a biased amino acid residue composition and evolved faster than structured proteins [6, 7], with exception of very small islands of relative conservation, known as short linear motifs (SLiM), that mediate protein-protein interactions [8].

IDRs are frequently encoded by overlapping open reading frames (ORFs) that evolved *de novo* by overprinting the ancestral ORFs and are common in viruses [9 - 13]. This overlapping of ORFs is accompanied by suppression of synonymous substitution rate in the ancestral ORFs (negative or purifying selection) compared to that of non-overlapping ORFs, indicative of codon constraints in the *de novo* ORFs due to their expression. The observed phenomenon has been extensively used for *in silico* identification of functional *de novo* ORFs [12, 14, 15], which often led to the elucidation of non-canonical expression mechanisms of these ORFs (e.g., [16 - 19]). Suppression of synonymous substitution rate is also reciprocally imposed on *de novo* ORFs by the overlapping ancestral gene. These observations led to analysis of relative rate change of substitutions in the *de novo* genes compared to ancestral or non-overlapping genes [12, 20 - 23].

This ORF-wide analysis has not been extended to individual codons of the *de novo* ORFs due to formidable technical challenges. A common approach to characterize site-specific evolution is to estimate deviation from the substitution rate under a model of neutral evolution for each codon of an ORF. Suppression and acceleration of the substitution rate is attributed to negative and positive selection, respectively, with positive selection being seen as the hallmark signature of adaptation during intra-species evolution [24]. One particular pattern of variation under positive selection is the frequent exchange of residues with pervasive return to the wild-type state, dubbed residue toggling [25]. Identification of codons under selection, either negative or positive, is part of the established evolutionary-based pipeline that informs functional characterization of proteins encoded in non-overlapping ORFs [2, 27]. However, the available techniques were not developed to untangle selection forces acting on the overlapping ORFs, which constrain evolution of each other. This may explain the lack of identification of *de novo* codon(s) under positive selection, despite broad recognition of a prominent role that the overlapping ORFs play in adaptation of viruses to host [12].

One of the largest and poorly characterized pairs of proteins encoded by overlapping ORFs is expressed by members of the fast growing *Polyomaviridae* family (**Supplementary Table S1**). These viruses cause latent infections in diverse mammals and birds, and in humans, some of these viruses have been responsible for different pathologies in immunocompromised individuals [28, 29]. Polyomaviruses employ multi-ORF double-stranded DNA (dsDNA) genomes of approximately 5 kb [30, 31]. Genomes of a large subset of polyoma-



viruses include two overlapping ORFs [15, 32, 33], designated here ORF2 and ORF5 (**Figure 1A**; for other designations see **Supplementary Text S1** and **Supplementary Table S2**). ORF2 encodes the second exon of the large T antigen (LT) that includes a helicase domain [30, 34]. ORF5 is expressed as a separate protein (ALTO) in Merkel cell polyomavirus (MCPyV) [15]; while it encodes the second exon of Middle T antigen (MT) in murine and hamster polyomaviruses (MPyV and HaPyV) [33, 35, 36]. The ORF5-encoded part of MT antigen is implicated in control of cell transformation [33, 35, 36], enriched with Pro residues [37, 38] and includes a C-terminal transmembrane domain [35] that is essential for the oncogenic function of MT [39]. This function and interaction of MT with different cellular proteins may be modulated by phosphorylation at several Ser, Thr and Tyr residues in rodent polyomaviruses [36]. We will use ORF5-plus and ORF5-less to refer to respective subsets of polyomaviruses; ORF5-plus viruses are also known as *Almipolyomaviruses* [15]. Likewise, and purely for the sake of uniformity, hereafter we have designated the ORF5-encoding product as MT/ALTO for all ORF5-plus polyomaviruses. Because ORF5 is conserved in only ORF5-plus polyomaviruses, while the overlapping part of ORF2 is found in all mammalian polyomaviruses [15], these ORFs are defined as *de novo* (ORF5) and ancestral (ORF2), according to Sabath *et al.*, [12]. ORF5-plus viruses form a large monophyletic cluster in one of the main branches of polyomavirus tree [15], dubbed *Orthopolyomaviruses I* [Ortho-I]; with three other branches being *Orthopolyomaviruses II*, *Malawipolyomaviruses* and *Wukipolyomaviruses* [40], although branch delineation and designation may vary in different studies [15, 41].

To understand the evolution of overlapping ORFs, we studied ORF2 and ORF5 at codon resolution. We found that one of the most conserved ORF5 codons, located in a SLiM of ORF5, experienced an accelerated evolutionary rate despite being strongly constrained to two amino acids by the overlapping ancestral ORF2. Using available and specially developed evolutionary-based approaches we revealed an unprecedented frequent toggling between these two residues during large-scale multi-species evolution in the Ortho-I clade of polyomaviruses. This analysis is, to our knowledge, the first to identify a conserved position of *de novo* protein under positive selection. Its results suggest a new IDR-mediated adaptation mechanism employed by many mammalian polyomaviruses with potential relevance to understanding adaptation of other viruses and organisms.

## Materials and Methods

### Datasets: viruses, sequences and alignments

Full-length genome sequences of 55 polyomaviruses available in the Genbank/RefSeq database on February 2013 (**Supplementary Table S1**) were downloaded into the Viralis platform [42]. When several genomes per species were available, the RefSeq sequence was chosen for presentation. The Muscle program [43] and ClustalW [44] were used to generate family-wide multiple amino-acid alignments for viral capsid protein (VP)1 encoded in ORF3, VP2 (ORF4) and LT (ORF2), followed by manual curation. For each of the three protein

alignments, strongly conserved blocks [45] were extracted using the Blocks Accepting Gaps Generator (BAGG) tool ([www.genebee.msu.su/~antonov/bagg/cgi/bagg.cgi](http://www.genebee.msu.su/~antonov/bagg/cgi/bagg.cgi)) to produce a concatenated multiple sequence alignment used for phylogenetic reconstruction and other analyses (see below). The ORF2-wide alignment was also mapped on the genome sequences, which were then translated in the alternative reading frame (RF -3) encoding ORF5 in twenty-two viruses of the ORF5-plus group to produce an ORF5 alignment. ORF5 size varies from 441 nucleotides (nts) to 846 nts, and ORF5 sequence conservation was detectable only in some subsets of polyomaviruses (**Supplementary Table S3**; data not shown; [15]).

For analysis of site-specific evolutionary selection by Datamonkey programs, we used 10 alignments of selected positions of the ORF5 and ORF2 (datasets, D1-D10). These 10 alignments represented different groups of viruses, including all mammalian polyomaviruses (D1 and D2), Ortho-I viruses (D3 and D4), ORF5-plus viruses (D5 and D6), ORF5-less viruses (D7), and three non-overlapping lineages of ORF5-plus viruses (D8-D10), each analyzed separately (see **Supplementary Table S4** for details). Using conservation considerations, some codons of ORF5 and ORF2 were selected, so all datasets included ORF5 codons while D2, D4, D6 and D7 included also ORF2 codons. For ORF5 of D1-D7, those codons were chosen whose overlapping codon in ORF2 (-1 frame) was aligned with no gaps across mammalian polyomaviruses. For ORF5 of D8-D10 and ORF2, most conserved codons in respective alignments were used after manual pruning of weakly aligned codons.

Alignments of the conserved motifs in the N-terminal part of LT ORF2 and ORF5, partially described elsewhere [15], were produced and converted into logos. To produce alignments as input for the RNAz program, we converted codon ORF5-based alignments of four subsets of ORF5-plus and two subsets of ORF5-less polyomaviruses, into the respective nucleotide alignments (**Supplementary Table S3**).

## Phylogeny reconstruction

Phylogenetic analyses were performed by using a Bayesian approach implemented in BEAST version 1.7.4 [46] and the Whelan and Goldman (WAG) amino acid substitution matrix [47]. Rate heterogeneity among sites was modeled using a gamma distribution with four categories, and a relaxed molecular-clock approach was tested against the strict molecular-clock approach [48] and was found to be superior. Markov chain Monte Carlo (MCMC) chains were run for 2 million steps and the first 10% were discarded as burn-in. Convergence of the runs was verified using the Tracer tool (<http://beast.bio.ed.ac.uk/tracer>).

## Analysis of natural selection at codons

We have used Mixed Effects Model of Evolution (MEME) [49] and Fast, Unconstrained Bayesian AppRoximation (FUBAR) [27] at the Datamonkey website (<http://www.datamonkey.org>) [26] to test for natural selection at conserved ORF5 codons. In addition, we have screened for toggling at ORF5 residues using TOGGLE, an implementation of the residue toggling method developed for HIV-1 by Delpont *et al.*, [25]. We have analyzed in total ten different datasets, D1-D10 (see above), capturing different positions and virus diversities

(see **Supplementary Table S4**). For each analyzed data set, selection of evolutionary model was performed automatically at the Datamonkey web site using default parameters prior to the analysis.

### **Analysis of COCO-VA toggling by BayesTraits**

Evolution of non-synonymous replacements at the **Codon-Constrained Val-Ala** (COCO-VA) site of ORF5 was analyzed by BayesTraits package using the Multistate model (<http://www.evolution.rdg.ac.uk/BayesTraits.html>) [50]. This codon is constrained to encode either Ala or Val in all mammalian polyomaviruses due to the overlapping Cys codon of the LXCXE motif that is expressed in the LT ORF2 of these viruses. The analyzed polyomaviruses were divided into two groups based on whether or not they express ORF5: ORF5-plus and ORF5-less viruses, respectively. The COCO-VA site is expressed as part of ORF5 in ORF5-plus, but not in ORF5-less viruses.

To test whether Ala-Val trait transitions are statistically more frequent in the ORF5-plus lineage compared to ORF5-less viruses, we applied the BayesTraits multistate model using a single trait (Ala/Val). We ran the analysis for three virus datasets: the combined set of mammalian polyomaviruses as well as separately for ORF5-plus and ORF5-less viruses, with respective posterior tree samples obtained through independent BEAST analyses. We then compared the estimated Ala-to-Val and Val-to-Ala transition rates between the three datasets, including an average Ala-Val exchange rate (corresponding to the toggling rate) by plotting the distributions. Statistical significance of differences in Ala-Val exchange rates was assessed using log Bayes Factors that was calculated with the R package Bayes Factor (<http://bayesfactorppl.r-forge.r-project.org/>). As Ala-Val exchange is equivalent to T-C exchange at the second codon position of the COCO-VA codon (see “Results and Discussion”), we applied BayesTraits also to the third position of that codon as a control.

### **Statistical analyses of COCO-VA toggling using patristic distances**

For each virus the smallest pair-wise patristic distance (SPAT) to a virus encoding the same amino acid (monomorphic pairs: Ala $\leftrightarrow$ Ala and Val $\leftrightarrow$ Val; monoSPAT) and to that encoding the different amino acid (polymorphic pair: Ala $\leftrightarrow$ Val, polySPAT) was calculated. Patristic distances were extracted from the polyomavirus phylogeny using the package Analyses of Phylogenetics and Evolution (APE) in R language [51].

We estimated the rate of COCO-VA toggling as the ratio of monoSPAT to the sum of polySPAT and monoSPAT values; designated SPAT ratio hereafter. Due to limited virus sampling at the intra-species level, we applied a sliding window approach to compare SPAT ratios between ORF5-plus and ORF5-less viruses. A window size of 0.15 and a shift of 0.05 at the monoSPAT scale were used. A two-sample non-parametric Mann-Whitney U test was utilized to test for statistically significant differences between the two virus groups within a particular window. A deviation of distributions of SPAT ratios from the average toggling rate of 0.5 was assessed using the Wilcoxon rank-sum test.

To independently assess the partitioning of mammalian polyomaviruses into ORF5-plus and ORF5-less virus groups, we determined the ranking of a predefined two-set partitioning among all possible two-set partitioning of the same type for the 30 viruses with monoSPAT values smaller than the derived threshold of 0.35. These 30 viruses comprise 14 ORF5-plus and 16 ORF5-less viruses or 16 Ortho-I and 14 non-Ortho-I viruses. We calculated the difference of mean toggling rate values between the two groups in each of these partitionings and determined its ranking among the differences of mean toggling rates obtained for all other 14-16 or 16-14 partitioning of the 30 viruses, whose total was 145,422,675 possible partitionings (e.g. combinations).

## General bioinformatics analyses

For selected phylogenetic lineages, alignments of ORF5 were converted HMM profiles and compared to each other using HHsearch [52] in both local and global alignment modes.

Sequence logos of selected alignments were produced using the WebLogo server [53, 54].

Secondary structure and disorder prediction of protein sequences were generated using the Disorder Prediction MetaServer, which reports consensus results of eight protein disorder predictor tools: DISEMBL [55], DISOPRED [56], DISpro [57], FoldIndex [58], GlobPlot2 [59], IUPred [60], RONN [61], and VSL2 [62], and two protein secondary structure predictor tools: PROFsec [63] and PSIPred [64] (<http://wwwnmr.cabm.rutgers.edu/bioinformatics/disorder/>). The prediction of disorder was considered significant if at least four predictors gave a hit.

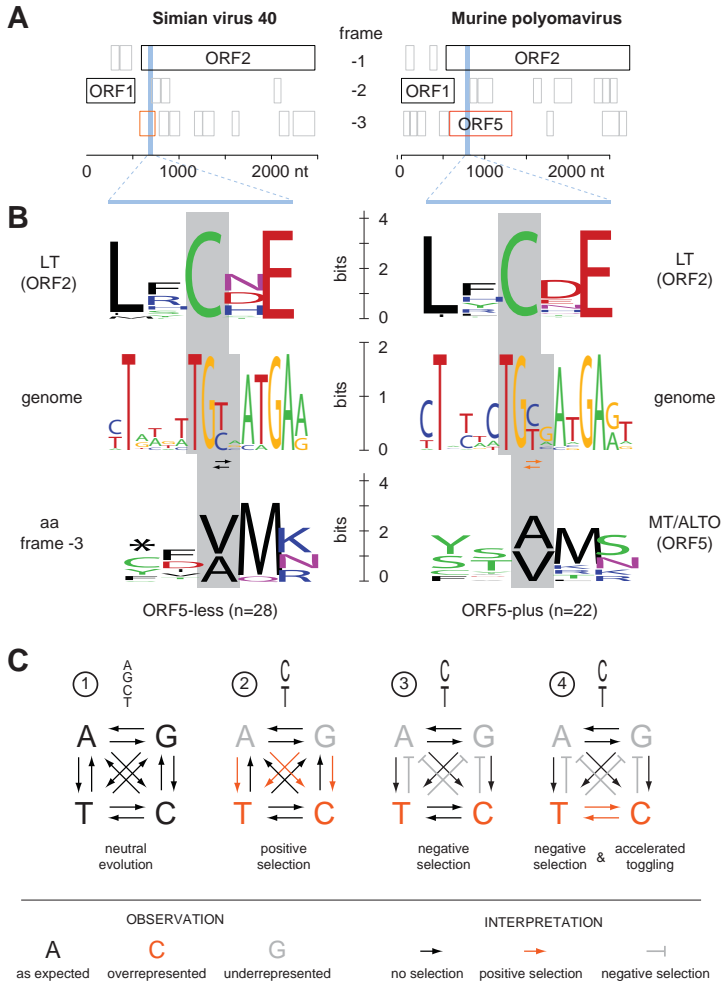
Secondary RNA structures in ORF2/ORF5 overlapping region were predicted with the program RNAz in a region of about 300-900 bp flanking the region encoding LXCXE motif sequence [65]. The server uses an algorithm that detects thermodynamically stable and evolutionarily conserved RNA secondary structures in multiple RNA-sequence alignments on both RNA-strands, with number of sequences in alignments not exceeding six. If subsets were larger than six, they were reduced to a combination of six virus sequences. For structure prediction the default RNAz parameters of "Standard Analysis" were utilized, which scored in the overlapping windows of 120 alignment columns with step-size of 40 nucleotides (<http://rna.tbi.univie.ac.at/cgi-bin/RNAz.cgi?PAGE=1&TYPE=5>).

Proline enrichment in putative ORF5-encoded protein sequences was analyzed by use of a custom R script that counts Proline residues and visualizes the counts with respect to location in the protein sequence and premature stop codons in the case of ORF5-less viruses ([www.R-project.org](http://www.R-project.org)) [66].

## Results and Discussion

### Discovery of Codon-Constrained Val-Ala (COCO-VA) toggling in MT/ALTO

We were interested in understanding the evolution and function of the *de novo* ORF5. Only



**Figure 1.** Toggling at the COCO-VA site in mammalian ORF5-plus and ORF5-less polyomaviruses. **(A)** ORF organization in three reading frames of the genomic region encoding the early genes is shown for Simian virus 40, SV-40 (left; NC\_001669) and Murine polyomavirus, MPyV (right; Genbank accession NC\_001515) representing ORF5-less and ORF5-plus polyomaviruses, respectively. The ORF2 frame was chosen as -1 frame for both viruses. ORF borders are defined here from stop to stop codon. Large expressed ORFs are boxed/outlined and named while other ORFs with a size of at least 75 nt are shown in grey. ORF5 of the ORF5-plus virus and one of its derivatives of the ORF5-less virus are highlighted in the -3 frame. The background highlighting indicates location of the LXCXE motif (an essential motif found in polyomaviruses and other viruses, and cellular proteins that mediates binding and inactivation of the cellular tumour-suppressor protein pRB [82 – 84]). **(B)** Shown are sequence logos of the LXCXE motif (top), the corresponding nucleotide sequence (middle) and the amino acid sequence translated from the ORF5 frame (bottom) for multiple alignment of the 28 ORF5-less (left) and 22 ORF5-plus viruses (right) analysed in this study using Viralis platform [42]. The asterisk indicates stop codons in the -3 frame of some ORF5-less viruses. See **M&M** section for other details. **(C)** Shown are four possible scenarios of evolution of a polynucleotide site under different selection regimes. In scenarios 2 to 4 different selection force(s) result in the same observed nucleotide diversity restricted to C or T. Scenarios 3 and 4 depict the COCO-VA toggling in ORF5-less and -plus viruses, respectively.

four short conserved motifs, designated ORF5m1 to ORF5m4, were evident in the ORF5-wide alignment (**Supplementary Figure S1**) due to an extremely high residue and two-fold size variation (see also below and Carter *et al.*, [15]). They are counterparts of four motifs of LT antigen in the overlapping part of ORF2. Remarkably, the most conserved 3<sup>rd</sup> aa residue of ORF5m2, identified in this study, has a restricted binary residue variation (Val/Ala) in both ORF5-plus and ORF5-less polyomaviruses (**Figure 1A** and **1B**). Val and Ala are encoded by eight **G(C/T)(A/G/C/T)** triplets which are the only codons compatible with the two **TG(C/T)** codons for conserved Cys of the LT LXCXE motif in the ancestral ORF2 (the two-nucleotide overlap between the Val/Ala and Cys codons is highlighted in bold). In other words, only variation at the second codon position of the COCO-VA codon (**C** or **T**) determines the encoded amino acid (Ala or Val) (**Figure 1B**). We named the observed phenomenon **Codon-Constrained Val-Ala** (COCO-VA) toggling.

The C/T variation represents only half of the full four-nucleotide variation possible at a polynucleotide position (**Figure 1C**). When each kind of nucleotide is equally frequent at a given position, it is likely to evolve at no selection (neutral evolution) (**Figure 1C1**), which may be found in the third codon positions of non-overlapping ORFs. In contrast, a restricted nucleotide variation, like C/T, may emerge as a result of selection, either positive (**Figure 1C2**) or negative (**Figure 1C3**), which is typically observed at the first and second positions of codons of non-overlapping ORFs. Evolutionary interpretation of the nucleotide variation is more complex in the overlapping ORFs, which may be subject to several evolutionary forces acting on each ORF. For instance, there is no doubt that the restricted C/T variation at the 2<sup>nd</sup> codon position of the COCO-VA site is due to negative selection in the alternative ORF2 to maintain the Cys residue. On the other hand, this restricted variation would be equally compatible with no selection or positive selection in ORF5, with the latter scenario leading to accelerated toggling between C and T (compare **Figure 1C3** and **Figure 1C4**). Therefore, we asked whether selection is involved in the COCO-VA toggling.

### Phylogeny suggests accelerated COCO-VA toggling in ORF5-encoding polyomaviruses

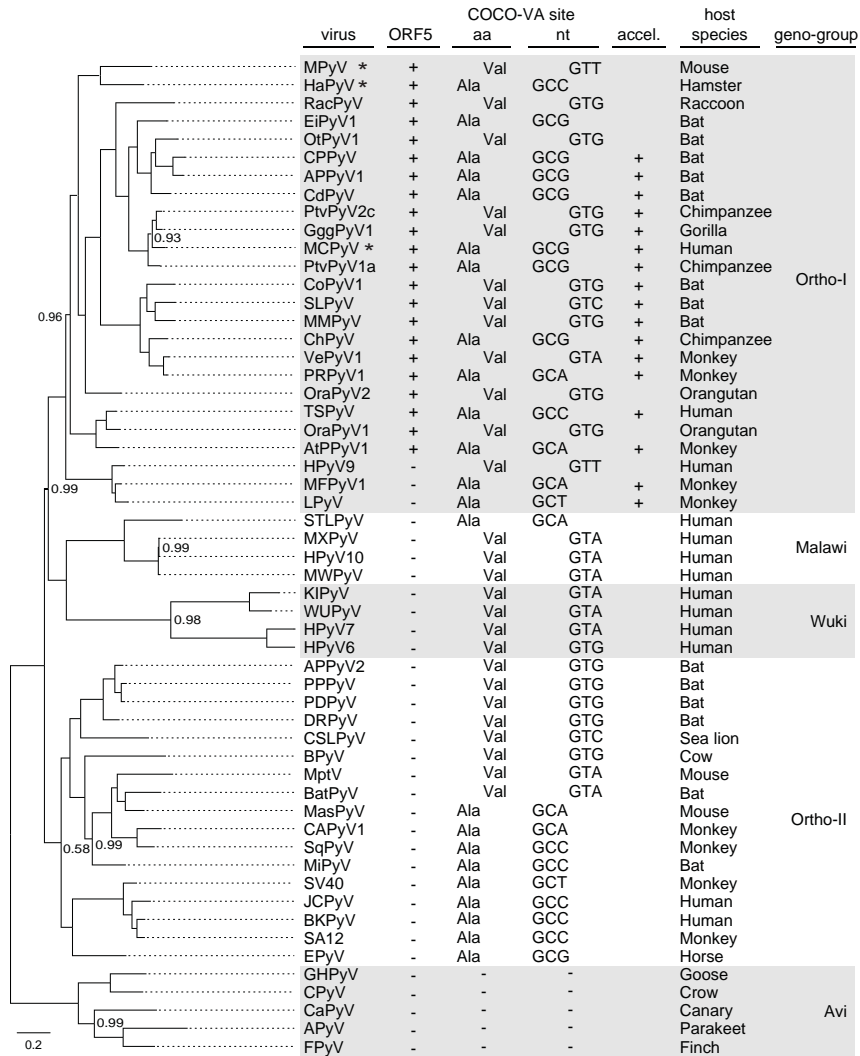
First and in line with general reasoning [67], we note that conservation of the LXCXE Cys residue may not constrain the COCO-VA toggling. Second, if C and T nucleotides at the third position of the Cys codon are utilized unevenly, additional *non-ORF2* selection pressure(s), for instance on RNA, must be taken into account when analyzing the toggling. Third, the ORF2 LXCXE conservation in *both* ORF5-plus and -less polyomaviruses provided us with two contrasting virus groups that differ in relation to the COCO-VA site expression through ORF5. Consequently, the COCO-VA site is not expected to be under selection pressure in ORF5-less viruses (**Figure 1C3** scenario), while its evolution in ORF5-plus viruses may or may not be driven by selection depending on the functional importance of these residues (either **Figure 1C3** or **Figure 1C4** scenario). Fourth, the restricted binary choice of aa residues at the COCO-VA site compared to the full 20 amino acid (aa) residue variation simplifies the evolutionary analysis of its residue variation.

Taking all these considerations into account, we reasoned that the relative abundance of *either Ala or Val* in ORF5-plus compared to ORF5-less polyomaviruses would be indicative of selection on residue *type*. Since Ala and Val are similarly and evenly abundant at the COCO-VA site in the known ORF5-plus and ORF5-less mammalian polyomaviruses: 11 vs. 11 and 12 vs. 14 (**Figure 2**), respectively, no indication for selection is apparent. This observation indicates also that the COCO-VA site may not have experienced other, non-ORF2-related selection favoring one of the two nucleotides. Accordingly, we have not found conserved RNA secondary structure elements in this region (see **Supplementary Text S2** and **Supplementary Table S3, Supplementary Figure S2** and **S3**), which, potentially, could have been an alternative source of constraint on the non-synonymous substitution in ORF5.

Next, we investigated the *frequency* of COCO-VA toggling among polyomaviruses. In this and subsequent analyses, switching between Ala and Val residues was accounted with no regard to its direction: from Ala to Val or from Val to Ala. The analysis was limited to the interspecies comparisons. The rate of the COCO-VA toggling in the ORF5-less polyomaviruses provided a baseline rate of COCO-VA toggling that can be expected by chance mutation (neutral evolution). Comparison of this rate with that of the ORF5-plus viruses informed us about directional selection at the COCO-VA codon in the latter viruses.

In the framework of this comparison, we have first mapped COCO-VA toggling on a Bayesian phylogenetic tree of polyomaviruses (**Figure 2**). Due to extreme sequence divergence of the ORF2/ORF5 overlap region in mammalian polyomaviruses (see above and Carter *et al.*, [15]), reliable alignment of this region is limited to four motifs of only ~30 residues in total (**Supplementary Figure S1**), which may not be sufficient for reliable phylogeny reconstruction. Therefore we choose to use a concatenated alignment of other conserved domains representing LT, VP1 and VP2 proteins and accounting for ~50% of genome for phylogeny inference. Large monophyletic groups on this tree were formed by viruses, which were recognized as similar in the ORF2/ORF5 overlapping region. Additionally, we have observed good agreement between topologies of separate branches of this tree, each representing closely related polyomaviruses, with trees of these same viruses using alignments of the ORF2/ORF5 overlap region (**Supplementary Figure S4**). These observations showed that the ORF2/ORF5 overlap region is likely to have coevolved with the LT, VP1 and VP2/3 proteins, whose tree was thus considered suitable for analysis of the COCO-VA toggling.

Subsequently, visual inspection of the tree revealed contrasting patterns of phylogenetic grouping for Ala- and Val-specific viruses in ORF5-plus and ORF5-less subsets of mammalian polyomaviruses, respectively (**Figure 2**). While Ala- and Val-specific viruses were largely intertwined in the first subset, they predominantly formed large residue-specific monophyletic groups in the second subset. This result was indicative of acceleration of the COCO-VA toggling in ORF5-plus viruses. To verify and extend this observation further, we have conducted additional evolutionary-based analyses using available and specially designed approaches.



**Figure 2.** Polyomavirus phylogeny and ORF5 characteristics. Shown is a Bayesian phylogeny using BEAST version 1.7.4 [46] for 55 polyomaviruses (listed in **Supplementary Table S1**) based on conserved regions in the LT, VP1 and VP2 proteins (see **M&M** section for details). The numbers plotted in the tree show posterior probability support values for internal branching events <1. The scale bar is in average number of amino acid substitutions. Asterisks in the virus column indicate viruses for which the ORF5 expression has been demonstrated experimentally. The ORF5 column indicates the presence (+) or absence (-) of ORF5 in polyomaviruses genomes. The COCO-VA site column depicts the residue (Ala or Val) and corresponding codon at the COCO-VA site that is constrained by the Cys codon of the LT LXCXE motif in mammalian polyomaviruses (see **Figure 1** and **Supplementary Figure S1**). The acceleration column (accel.) labels viruses that experienced selection-driven acceleration at the COCO-VA site. The geno-group column depicts the phylogenetic distribution of polyomaviruses according to Feltkamp *et al.*, [40]. Please note that Carter *et al.*, 2013 [15] divided all mammalian polyomaviruses into two groups, monophyletic *Almipolyomaviruses* and paraphyletic *non-Almipolyomaviruses*, which correspond to ORF5-plus and ORF5-less polyomaviruses, respectively. The tree was pseudorooted at the branch connecting mammalian and avian (Avi) polyomaviruses (see **M&M** section for other details).





## No evidence for positive selection at the COCO-VA site by conventional evolutionary analyses

We started with employing two most advanced and widely used programs, MEME [49] and FUBAR [27], developed for evolutionary analysis of residue variation. Also included was TOGGLE [25], which was specifically developed for analysis of residue toggling. These three programs are available through the Datamonkey website [26]. The employed programs differ in how they accommodate lineage- and site-specific variation in the analyzed dataset to infer patterns of evolution and deduce selection forces acting on individual codons. Since these tools were developed for the analysis of non-overlapping ORFs, only a single evolutionary force, if identified, is reported. Consequently, we did not expect that these programs could infer both purifying selection (due to Cys conservation of LXCXE) and positive selection (accelerated Val-Ala toggling) at the COCO-VA site of the ORF5-plus viruses as depicted in **Figure 1C4**. Rather, we asked whether the programs could provide evidence for either negative or positive selection at this site of ORF5-plus viruses and negative selection at this site of ORF5-less viruses. This type of inferences depends on the number and diversity of alignment positions under analysis. Due to the high sequence divergence of the ORF2/ORF5 overlapping region, we thus analyzed different subsets of mammalian polyomaviruses, in order to facilitate identification of selection forces. Specifically, the programs were applied to ten different alignments of ORF5, D1-D10, representing selected ORF5 codons, which may or may not be merged with ORF2 codons for different subsets (see **M&M** and **Supplementary Table S4**). In none of the thirty conducted analyses, the COCO-VA codon was identified to be under positive/diversifying selection, including toggling. Also the COCO-VA codon was not found to be negatively selected in analyses that included only ORF5-plus viruses, neither in the entire set nor its D5, D6, D8, D9 and D10 subsets. However, upon analysis of the other five virus sets by FUBAR, either including all ORF5-plus viruses along with other viruses (D1-D4) or including only ORF5-less viruses (D7), the COCO-VA codon was identified to be under purifying selection. In contrast, various other ORF5 codons were identified as being positively or negatively selected or be involved in toggling, in many of these analyses (**Supplementary Table S4**).

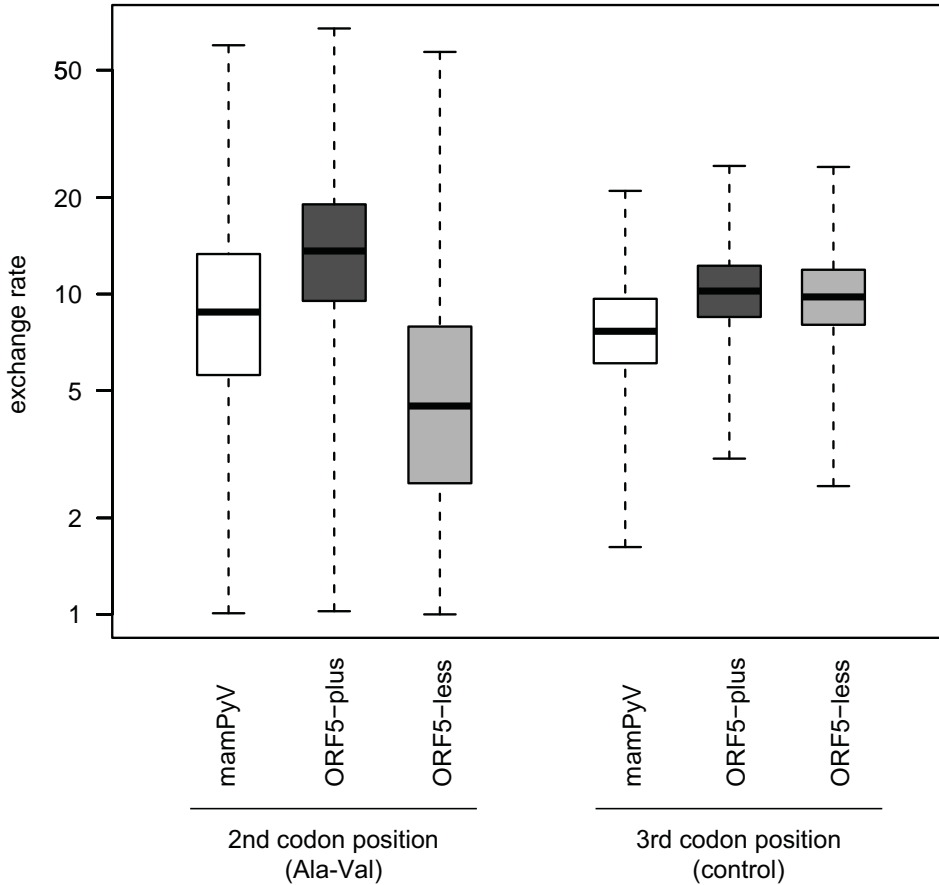
The lack of evidence for positive selection/toggling at the COCO-VA codon in ORF5-plus viruses according to these analyses could be either a true negative result (lack of the phenomenon) or a false negative result (failure to detect a signal due to systematic technical deficiency). As detailed below, we believe that the latter explanation is most likely. Indeed, the employed three programs operate under the assumption that the entire codon table of 61 varieties is available for evolution at every site in the analyzed alignments. Consequently, the eight different codons (four for Val and four for Ala) observed at the COCO-VA site were seen as severely restricted rather than representing the full spectrum allowed at this site (imposed by Cys conservation in the overlapping ORF2 codon). This misreading of the observed residue variation has profound implications for its evolutionary interpretation, since high diversity tends to be interpreted as a sign of positive selection, while restricted diversity is commonly associated with purifying selection during evolution of non-overlapping

ORFs. It is thus not surprising that the COCO-VA site evolution was qualified to be under negative selection in several tests by FUBAR. This result could be seen as evidence for the dominance of purifying selection at the COCO-VA site according to FUBAR.

### **Toggling at the COCO-VA site is significantly accelerated**

Due to the above considerations, we decided to continue our testing of the toggling by applying an approach that could be free from the limitations of standard evolutionary based programs developed for non-overlapping ORFs. First, we sought to verify the elevated frequency of Ala-Val exchange in ORF5-plus compared to ORF5-less viruses that is apparent from polyomavirus phylogeny (**Figure 2**). To this end, we have applied the Multistate method of the BayesTraits package [50] to compare the COCO-VA site variation in mammalian ORF5-plus and ORF5-less polyomaviruses. The program employs continuous-time Markov models to estimate the transition rates between multiple states for a single trait (Ala/Val for COCO-VA site in this case) while it traverses a tree. The produced estimates take into account the uncertainty associated with tree reconstruction as it utilizes the full posterior tree sample. The estimated transition rate distribution was plotted for three virus datasets (**Figure 3** left). From this plot it is evident that the estimated Ala-Val exchange rate is more than 3 times higher (13.6 vs 4.3) for ORF5-plus viruses compared to ORF5-less viruses, with a 25-75% interquartile range of 9.5-19.1 and 2.5-7.7, respectively. This striking difference between the two datasets is strongly supported by a log Bayes Factor (logBF) of 3352.2, which is astronomically large and dwarfs the significance threshold of 2. As expected, the estimate for mammalian polyomaviruses was intermediate between those two with a 25-75% interquartile range of 5.6-13.3 (**Figure 3** left). As Ala-Val exchange at the COCO-VA position is equivalent to C-T exchange at the second codon position (see **Figure 2** and **M&M**) we have compared its exchange rate to that at the third codon position (**Figure 3** right). This position accepts all four nucleotides and its variation is primarily driven by selection in the overlapping ORF2 in which it occupies the first codon position of the subsequent residue. As may be expected, the exchange rate at this position (now averaged over four instead of two nucleotides) is comparable for ORF5-plus and ORF5-less viruses (median and 25-75% interquartile range: 10.2 and 8.5-12.3 vs. 9.8 and 8.0-11.9, respectively). Of notice, these numbers are still and consistently smaller than those of the Ala-Val exchange rate for ORF5-plus viruses.

Importantly, the observed difference at the second codon position (i.e., the Ala-Val exchange) may not be attributed to differences in virus diversity of the compared two datasets, whose distributions of smallest pair-wise patristic distance (SPAT) values (median value and 25-75% inter-quartile range: 0.23 and 0.13-0.32 vs. 0.28 and 0.12-0.46) were not different at a statistically significant level (Mann-Whitney U test;  $p=0.42$ ). Consequently, we concluded that the COCO-VA toggling rate is significantly and genuinely accelerated in the ORF5-plus compared to the ORF5-less polyomaviruses. Since the COCO-VA site is expressed in ORF5-plus but not ORF5-less viruses (although see below), this result implies positive selection on the COCO-VA site in ORF5-plus viruses.



**Figure 3.** COCO-VA toggling is accelerated in ORF5-plus compared to ORF5-less viruses. Shown are results of BayesTraits multistate analysis of COCO-VA toggling rate in three groups of viruses. The distributions of estimated exchange rates at second (left side) and third (right side) codon position of the COCO-VA codon is shown. The exchange rate at the second codon position corresponds to the COCO-VA toggling while the rate at the third codon position serves as a control. The distributions are shown as Box-and-whisker graphs. The boxes span from the first to the third quartile and include the median (bold line), and the whiskers (dashed lines) extend to the extreme values.

### Ratio approach to study accelerated COCO-VA toggling

To study the accelerated COCO-VA toggling further, we have developed a ratio approach remotely similar to that of comparing the ratio of non-synonymous to synonymous substitutions. We used the ratio of monoSPAT/(monoSPAT+polySPAT) values as a normalized measure of the COCO-VA toggling rate relative to Ala/Ala or Val/Val persistence, with polySPAT and monoSPAT resembling estimations of non-synonymous and synonymous substitutions, respectively (for group designations see **M&M**). Only the C/T variation at the 2<sup>nd</sup> codon position that controls Val-Ala exchange, rather than the entire codon for Ala/Val as it would have

been the case upon analysis of a non-overlapping ORF by a conventional technique, was analyzed in our test. We thus avoided complications to the analysis that would otherwise be caused by the unaccounted evolutionary pressure on the third position of Ala/Val codons by the ORF2 overlapping codon, where it occupies the first position of the subsequent residue (**Figure 1B**). An SPAT ratio of 0.5 indicates that the Ala/Val exchange rate matches that of Ala/Ala or Val/Val persistence during evolution of a particular lineage (hereafter, matching rate). Since amino acid residue persistence at the COCO-VA site in a pair of viruses may involve either no genetic change or synonymous substitution, the matching rate for toggling under the model of neutral evolution could be expected only at sufficiently large evolutionary distances when chance mutation, either synonymous or non-synonymous, is highly probable. Accordingly, persistence would dominate over toggling at smaller distances under this model, resulting in SPAT ratios smaller than 0.5. If positive selection is involved in toggling, increase of SPAT ratios compared to those expected under neutral evolution could be observed at sufficiently small evolutionary distances.

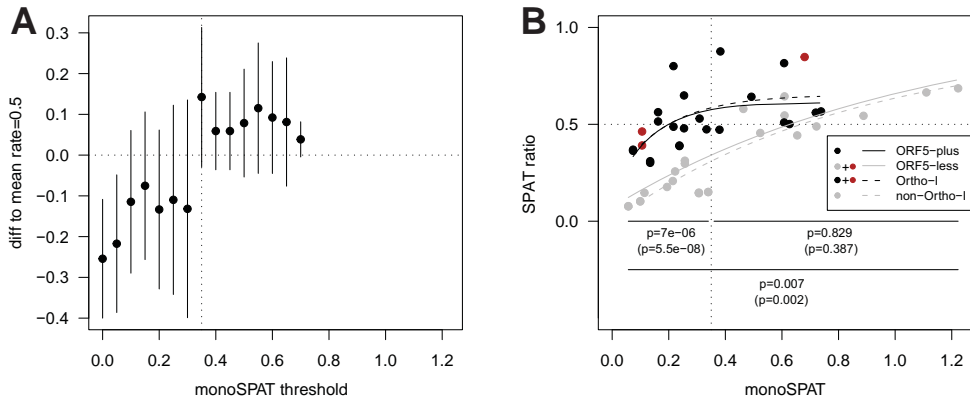
The above considerations indicate that under the model of neutral evolution we could expect different SPAT ratios at small and large evolutionary distances. To verify this and define ranges for small and large evolutionary distances separating pairs of monomorphic viruses, we analyzed the difference between the matching rate and within-window distributions of SPAT ratios involving all mammalian polyomaviruses, which were plotted against monoSPAT values. Due to stochastic reasons, the estimated toggling rate may deviate from the actual rate for a virus. To address this limitation, we pooled SPAT ratios within a predefined window that was slid along the monoSPAT axis. Our analysis revealed that the Ala/Val exchange rate of polyomaviruses varies considerably, with very different median values being observed in two monoSPAT ranges (**Figure 4A**). In the monoSPAT range of 0-0.35, median SPAT ratios were consistently smaller compared to the matching rate, while in the monoSPAT range of 0.35-0.75, they were consistently larger than the matching rate. Accordingly, the entire monoSPAT range was split into two sub-ranges in our subsequent analyses. For evolutionary interpretations of the ratio test in subsequent analyses, we used the results obtained for ORF5-less viruses as a base-line, since Val-Ala toggling in these viruses is expected to experience no selection (**Figure 1C3** scenario) over the entire evolutionary distance range.

### **Accelerated COCO-VA toggling is associated with interspecies diversification of ORF5-plus polyomaviruses**

Is the accelerated toggling a characteristic of the entire ORF5-plus viruses or its subsets? The difference between SPAT ratios in the distributions for ORF5-plus and ORF5-less viruses was statistically significant over the monoSPAT range of 0-0.35 (MWU test  $p$ -value=7e-06), but not over the 0.35-1.2 range (MWU test  $p$ -value=0.829) (**Figure 4B**). Importantly, this result may not be due to biases of the virus sampling which was comparable for ORF5-plus and -less viruses in the two distributions along the monoSPAT range (**Supplementary Figure S5**).

Consequently, the above observations indicate a selection-driven acceleration of COCO-VA toggling in the majority (fifteen out of twenty-two) of ORF5-plus polyomaviruses, each of which is separated from another monomorphic virus by a monoSPAT of 0.35 or smaller (**Figure 4**). For the remaining seven ORF5-plus viruses, each of which is separated from another monomorphic virus by a monoSPAT larger than 0.35, no accelerated COCO-VA toggling was observed. This could be either due to specifics of evolution or the unavailability of close monomorphic relatives of these viruses in the current sampling. If the former is true, the viruses with accelerated COCO-VA toggling may be expected to cluster in the tree, while a random phyletic distribution is likely otherwise. **Figure 2** shows that the fifteen viruses with accelerated COCO-VA toggling are scattered across the entire branch of ORF5-plus viruses. This observation implies that the accelerated COCO-VA toggling may involve *all* ORF5-plus viruses (all terminal nodes in the respective tree branch) thus presenting an extreme case of convergent evolution. An improved, much larger virus sampling, which includes closely related viruses for each analyzed virus species, will enable verification of this implication. Also, it may facilitate additional insights, including: a) refining the estimate of the monoSPAT threshold at which the COCO-VA toggling acceleration can be observed, and b) extending our analysis to poorly sampled intra-species diversity, in order to address the question whether COCO-VA toggling drives speciation or vice versa.

Could the observed difference between ORF5-plus and ORF5-less viruses in the monoSPAT range of 0-0.35 (**Figure 4B**) have emerged also under the evolutionary scenario that is alternative to that involving positive selection on ORF5-plus and no selection on ORF5-less viruses? If the COCO-VA site was under strong negative selection in ORF5-less viruses while being under either weak negative or no selection in ORF5-plus viruses, SPAT ratio of these viruses would differ. The following considerations make this scenario unlikely to be applicable to explain the data obtained in our study. First, this scenario implies that the COCO-VA site must be expressed in *all* ORF5-less viruses. These viruses include some of the most well characterized polyomaviruses, e.g. SV40, with no evidence for the expression of the COCO-VA site, although some of the poorly characterized ORF5-less viruses may indeed express this site (see below). We could also recall that ORF5-less viruses were defined as a group not having the property (ORF5) rather than having one, which would be required to link strong negative selection to the functional characteristic. Second, SPAT ratio of ORF5-plus viruses in the monoSPAT range of 0-0.35 is comparable to the matching rate ( $p=0.390$  in Wilcoxon rank sum test; see **Supplementary Table S5**). This result is in the excellent agreement with positive selection acting on the COCO-VA site in ORF5-plus viruses, while it may not be reconciled with the weak negative selection hypothesis. On the other hand, it would in principle be compatible with neutral evolution of the COCO-VA site in ORF5-plus viruses under the condition that polyomaviruses have very high mutation rate. The estimates of this rate vary greatly and generally this aspect has not been fully resolved [68]. However, we note that the Val-Ala variation is already observed in several monophyletic subsets of ORF5-plus viruses which otherwise diverged little or modestly. This observation indicates that the



**Figure 4.** Accelerated toggling at the COCO-VA site in ORF5-plus and Ortho-I viruses. For the purpose of this analysis a representative set of 50 mammalian polyomaviruses (see **Supplementary Table S1**) was studied. Due to the lack of species demarcation criteria for polyomaviruses, we chose to consider viruses with different names as representing different species (dots in the plot). The only exception was made for MX polyomavirus, Human polyomavirus 10 and MW polyomavirus, which were represented only by the latter because of the very small distances that separate these three viruses. Two pairs of virus partitioning (subsets) of the mammalian polyomaviruses, based on the ORF5 presence and phylogeny, were considered. They and their colour codes are defined in the inset of panel **B**. **(A)** The partitioning of the monoSPAT scale at 0.35 was derived based on the drop of the mean difference of SPAT ratios to the matching rate. Here, a sliding window (size 0.15, shift 0.05) starting at monoSPAT of 0.0 was moved along the monoSPAT range to calculate within-window mean differences (dots) and associated standard deviations (vertical lines). See also **M&M** section and **Supplementary Figure S5** for other details. **(B)** The curves show the fit of a 3-parameter logistic function to each of four different subsets. The numbers below show P-values of Mann–Whitney U tests comparing the SPAT ratio distributions between ORF5-plus and ORF5-less viruses (*Orthopolyomavirus-I* and non-*Orthopolyomavirus-I*) for two monoSPAT ranges (0–0.35, 0.35–1.25). A horizontal dotted line is drawn at the matching rate, whose evolutionary interpretation is defined in the text.

Val-Ala variation may be among most frequent rather than average as would be expected under the neutral evolution scenario. This aspect could be studied most closely with the improved virus sampling. In conclusion, based on the available data the accelerated COCO-VA toggling due to positive selection is the most likely evolutionary scenario.

### Accelerated COCO-VA toggling is most strongly associated with monophyletic Ortho-I viruses

The results described above provide evidence for accelerated COCO-VA toggling in the 0-0.35 monoSPAT range for ORF5-plus viruses. However, it is also evident that the distributions of ORF5-plus and –less SPAT ratios overlap with two ORF5-less viruses deviating considerably from their group-mates and instead fitting into the other group rather well (**Figure 4B**, red dots in the 0-0.35 monoSPAT range). This grouping with ORF5-plus viruses received strong statistical support when analyzing all of the 145,422,675 possible 16-by-14 combinations of the 30 viruses with monoSPAT values in the range of 0-0.35 (**Supplementary Text S3** and **Supplementary Figure S6**). Intriguingly, these two viruses along with another one for which no closely related monomorphic virus is available in the current virus sampling (**Figure 4B**,

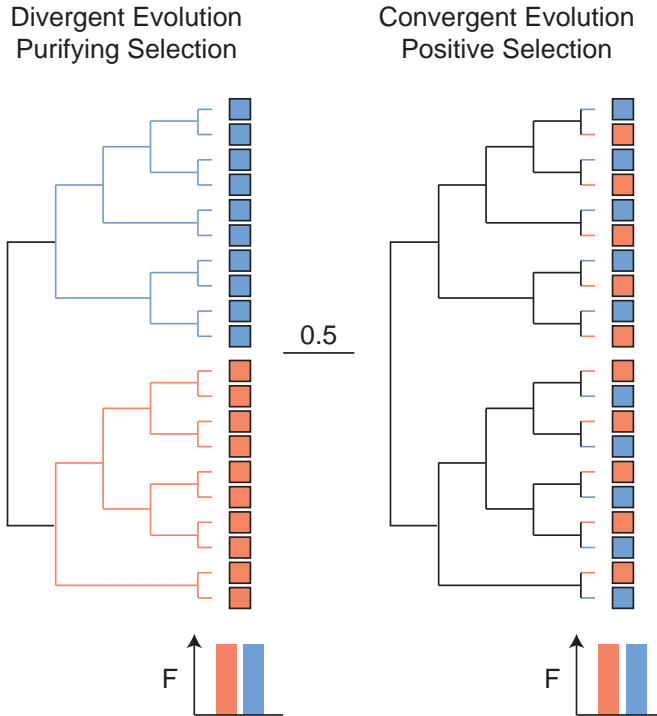
red dot in the 0.35-1.20 monoSPAT range) form a sister lineage of ORF5-plus polyomaviruses at the root of the Ortho-I monophyletic group (**Figure 2**) [15, 40]. These results suggest that the accelerated toggling is most strongly associated with the Ortho-I group. Since the observed accelerated toggling is indicative of positive selection that may be realized only upon expression of the COCO-VA site in the two (three) poorly characterized basal Ortho-I viruses, such hypothesis must be considered. It could be achieved using a mechanism other than expression of the entire ORF5, for instance, through alternative splicing of mRNA(s) [69] that could fuse the COCO-VA site with other ORF(s). In the evolutionary framework, such expression of the COCO-VA site would be ancestral to those used by ORF5-plus viruses, implying that the COCO-VA site and the associated sequence motif could be a nucleation site for the subsequent ORF5 origin by ORF expansion [15].

### **COCO-VA toggling is located in a SLiM of an intrinsically disordered region**

What could be the structural basis of COCO-VA toggling? Bioinformatics analyses indicate that MT/ALTO is a Pro residue rich IDR, whose motifs could form SLiMs (**Figures S1, S7, S8 and S9**) [15]. Thus, the interspecific toggling targets a SLiM, which is in line with the notion that IDRs evolving differently than structured protein regions [6, 7]. Since SLiMs promote protein folding in relatively flat energy landscapes [70] and mediate interactions with partners that are relatively weak [71, 72], difference between physico-chemical properties of the just two possible COCO-VA residues, Val and Ala, could be of significance. For instance, these residues have contrasting structural propensities, favoring the formation of either  $\alpha$ -helix (Ala) or  $\beta$ -sheet (Val) [73], which might be used to promote alternative folding of MT/ALTO upon interaction with partner(s). Unfortunately, this hypothesis may not be tested using the available computational approaches.

### **Concluding Remarks**

In complex protein networks, SLiMs are emerging as evolutionary adaptive transmitters of intracellular signals involving multiple interacting partners [2, 3, 13, 74]. Here we presented evidence for the evolutionary signature of adaptation in the otherwise uncharacterized SLiM of MT/ALTO. The effect of COCO-VA toggling on the SLiM may be similar to that of phosphorylation which could modulate SLiM activity considerably [75]. MT antigen of rodent polyomaviruses has been shown to interact through its ORF5-encoded part with numerous cellular targets involved in signal transduction [33, 35, 36, 39]. The function of ALTO, identified just recently, has not been resolved yet [15]. The described COCO-VA toggling is notable because of a unique combination of properties: it involves one of the just few conserved positions of the otherwise highly divergent MT/ALTO protein, and it may affect every species of Ortho-I polyomaviruses. These viruses are known to infect bats, rodents, monkeys, hominids and humans with apparently frequent host switching (**Figure 2**). Future studies should identify driving forces of the COCO-VA toggling to enable its comparison with intra-species residue toggling [25]. The latter is likely driven by the cellular immune



**Figure 5.** Contrasting modes of evolution at a conserved protein site accepting two residues. Shown are fictional examples of evolution of a conserved protein site with two-residue variation in families of structured (left panel) and unstructured (right) proteins, whose evolutionary scale of replacement at all sites was considerable (bar 0.5) and whose phylogeny is described by identical trees. In both cases, two residues are evenly distributed, each occupying 50% of terminal nodes. Residue type either clusters into two monophyletic groups (left) or is intertwined (right). The left panel depicts divergent evolution driven mostly by purifying selection as seen in many characterized structured proteins. The right panel depicts convergent evolution driven by positive selection as discovered at the COCO-VA site in the presented chapter and may be experienced at other sites in unstructured proteins.

response and occurs at much smaller time and divergence scales, and with the exchange of many residues. Practically, our study suggests that analysis of substitution rates can be applied to individual residues in overlapping ORFs. It extends the utility of the substitution rate analysis from mapping to dissecting functional elements in overlapping ORFs.

The described phenomenon also challenges common perception of conservation of proteins, which is believed to be inversely and universally correlated with the rate of evolution. Accordingly, sites accepting relatively few residues are classified conserved and evolving slowly under negative selection. Typically, such residues are critical for maintaining protein core and/or playing an essential role in the active site of structured proteins. Besides the sites that are strictly invariant, those that accept only two residues during large-scale evolution are among the most conserved. Exchange of these residues could happen due to either rare fixation of non-synonymous mutation that is driven by episodic positive selection



or residue drift. As a result, each of the two residues is likely to be associated with a large monophyletic clade in the tree (**Figure 5** left panel), the pattern that can be recognized by available programs (e.g. [76]). Examples of this type of evolution are plenty in many protein families. For instance, rare exchange of the catalytic nucleophile Cys and Ser residues in virus proteases with chymotrypsin-like fold [77] or phosphate-binding Ser and Thr residues in the Walker-box GKS/T motif of nucleotide-binding proteins [78], are notable. The above considerations indicate that in structured proteins, limited residue variation may largely be imposed by the molecular environment in which these proteins operate. In contrast, constraints on the genetic level is the chief factor determining residue variation in proteins encoded in overlapping ORFs. Consequently, this restricted residue variation in overlapping ORFs may not be linked to residue function in the manner described for structured proteins. Accordingly, overlapping ORFs predominantly encode unstructured proteins with their most conserved SLiMs mediating adaptation, a function that is commonly facilitated by the least conserved elements in structured proteins. Along the same line, we now provide evidence for the phylogenetic intertwining of viruses that employ, respectively, Val and Ala at the conserved COCO-VA site in the IDR of MT/ALTO. When depicted in a simplified form, this phylogenetic pattern can be contrasted with the clade-specific association of residues in a tree of structured proteins (compare right and left panels of **Figure 5**). This contrast is particularly striking since it is not evident in the cumulative frequency of residues at terminal nodes (bottom panels underneath of trees in **Figure 5**). Thus, this logos-style representation of residue conservation, which is very popular in functional studies, may not capture residue change and its role in adaptation. Only analysis in the context of phylogeny could do it, as demonstrated in this study.

Since Cys is one of the least frequent amino acid residues and none of the other residues can constrain evolution in the -3 RF (or +1 RF) to only two residues, the described codon-constrained accelerated toggling might be viewed as an extremely exotic phenomenon limited to polyomaviruses. We believe that this perception is biased for several reasons. First of all, the (unknown) diversity of the Virus Universe is expected to be many orders of magnitude larger than the number of currently recognized few thousand virus species [79, 80]. This implies a good chance of discovering COCO-VA toggling in other viruses in the future. Furthermore, accelerated toggling might involve more than two residues at a site that could still be considered conservative relative to many other sites. Such constraint could be imposed by conserved amino acids other than Cys in the overlapping ORF or, if non-overlapping ORF is involved, by a different genetic mechanism, e.g. RNA structure, or even by a partner or partners interacting with an IDR site. Thus, the described COCO-VA toggling may represent an extreme case of common evolution of individual residues in IDRs of proteins, making it potentially relevant to understanding biology and pathology of adaptation of many organisms.

## Acknowledgement

The authors thank Jelle J. Goeman, Sander Kooijman, Andrey M. Leontovich, and Els van der Meijden for useful discussions, and Igor Sidorov and Dmitry Samborskiy for help with Viralis. A.E.G. is member of the Netherlands Bioinformatics Center (NBIC) Faculty.

## Funding

This study was funded by the Leiden University Medical Center, The Netherlands, partially through the Collaborative Agreement in Bioinformatics between Leiden University Medical Center and Moscow State University (MoBiLe Program), and Leiden University Fund.

## References

1. Wright PE, Dyson HJ. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm (1999) *J Mol. Biol.* 293: 321-331.
2. Dyson HJ, Wright PE. Intrinsically unstructured proteins and their functions (2005) *Nat Rev Mol Cell Biol.* 6: 197-208.
3. Dunker AK, Silman I, Uversky VN, Sussman JL. Function and structure of inherently disordered proteins (2008) *Curr Opin Struct Biol.* 18: 756-764.
4. Tompa P. The interplay between structure and function in intrinsically unstructured proteins (2005) *FEBS Lett.* 579: 3346-3354.
5. Xue B, Dunker AK, Uversky VN. Orderly order in protein intrinsic disorder distribution: disorder in 3500 proteomes from viruses and the three domains of life (2012) *J Biomol. Struct. Dyn.* 30: 137-149.
6. Brown CJ, Johnson AK, Daughdrill GW. Comparing models of evolution for ordered and disordered proteins (2010) *Mol. Biol. Evol.* 27: 609-621.
7. Brown CJ, Johnson AK, Dunker AK, Daughdrill GW. Evolution and disorder (2011) *Curr. Opin. Struct. Biol.* 21: 441-446.
8. Davey NE, Van Roey K, Weatheritt RJ, Toedt G, Uyar B, Altenberg B, Budd A, Diella F, Dinkel H, Gibson TJ. Attributes of short linear motifs (2012) *Mol. Biosyst.* 8: 268-281.
9. Keese PK, Gibbs A. Origins of genes: "big bang" or continuous creation? (1992) *Proc Natl Acad Sci U S A* 89: 9489-9493.
10. Firth AE, Brown CM. Detecting overlapping coding sequences in virus genomes (2006) *BMC Bioinformatics.* 7: 75.
11. Belshaw R, Pybus OG, Rambaut A. The evolution of genome compression and genomic novelty in RNA viruses (2007) *Genome Res.* 17: 1496-1504.
12. Sabath N, Wagner A, Karlin D. Evolution of viral proteins originated de novo by overprinting (2012) *Mol. Biol. Evol.* 29: 3767-3780.
13. Pushker R, Mooney C, Davey NE, Jacque JM, Shields DC. Marked variability in the extent of protein disorder within and between viral families (2013) *PLoS. One.* 8: e60724.
14. Ling R, Pate AE, Carr JP, Firth AE. An essential fifth coding ORF in the sobemoviruses (2013) *Virology* 446: 397-408.
15. Carter JJ, Daugherty MD, Qi X, Bheda-Malge A, Wipf GC, Robinson K, Roman A, Malik HS, Galloway DA. Identification of an overprinting gene in Merkel cell polyomavirus provides evolutionary insight into the birth of viral genes (2013) *Proc Natl Acad Sci U. S. A.* 110: 12744-12749.
16. Firth AE, Atkins JF. Evidence for a novel coding sequence overlapping the 5'-terminal approximately 90 codons of the gill-associated and yellow head okavirus envelope glycoprotein gene (2009) *Virology* 6: 222.

17. Firth AE, Atkins JF. A case for a CUG-initiated coding sequence overlapping torovirus ORF1a and encoding a novel 30 kDa product (2009) *Virol J* 6: 136.
18. Loughran G, Firth AE, Atkins JF. Ribosomal frameshifting into an overlapping gene in the 2B-encoding region of the cardiovirus genome (2011) *Proc Natl Acad Sci U S A* 108: E1111-E1119.
19. Fang Y, Treffers EE, Li Y, Tas A, Sun Z, van der Meer Y, de Ru AH, van Veelen PA, Atkins JF, Snijder EJ, Firth AE. Efficient -2 frameshifting by mammalian ribosomes to synthesize an additional arterivirus protein (2012) *Proc Natl Acad Sci U S A* 109: E2920-E2928.
20. Mizokami M, Orito E, Ohba K, Ikeo K, Lau JY, Gojobori T. Constrained evolution with respect to gene overlap of hepatitis B virus (1997) *J Mol. Evol.* 44 Suppl 1: S83-S90.
21. Hughes AL, Hughes MA. Patterns of nucleotide difference in overlapping and non-overlapping reading frames of papillomavirus genomes (2005) *Virus Res.* 113: 81-88.
22. de Groot S, Mailund T, Hein J. Comparative annotation of viral genomes with non-conserved gene structure (2007) *Bioinformatics.* 23: 1080-1089.
23. Sabath N, Landan G, Graur D. A method for the simultaneous estimation of selection intensities in overlapping genes (2008) *PLoS ONE* 3: e3996.
24. Kryazhimskiy S, Plotkin JB. The population genetics of dN/dS (2008) *PLoS Genet.* 4: e1000304.
25. Delpont W, Scheffler K, Seoighe C. Frequent toggling between alternative amino acids is driven by selection in HIV-1 (2008) *PLoS Pathog.* 4: e1000242.
26. Kosakovsky Pond SL, Frost SD. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments (2005) *Bioinformatics* 21: 2531-2533.
27. Murrell B, Moola S, Mabona A, Weighill T, Sheward D, Kosakovsky Pond SL, Scheffler K. FUBAR: a fast, unconstrained bayesian approximation for inferring selection (2013) *Mol. Biol. Evol.* 30: 1196-1205.
28. Dalianis T, Hirsch HH. Human polyomaviruses in disease and cancer (2013) *Virology* 437: 63-72.
29. Kazem S, van der Meijden E, Feltkamp MC. The trichodysplasia spinulosa-associated polyomavirus: virological background and clinical implications (2013) *APMIS* 121: 770-782.
30. Decaprio JA, Garcea RL. A cornucopia of human polyomaviruses (2013) *Nat Rev Microbiol.* 11: 264-276.
31. DeCaprio JA, Imperiale MJ, Major EO. Polyomaviruses. In: *Fields VIROLOGY* (2013). Knipe DM, Howley PM (editors). Philadelphia: Wolters Kluwer / Lippincott Williams & Wilkins: 1633-1661.
32. Kazem S, Lauber C, van der Meijden E, Kooijman S, Bialasiewicz S, Wang RC, Gorbalenya AE, Feltkamp MC. Global circulation of slowly evolving trichodysplasia spinulosa-associated polyomavirus and its adaptation to the human population through alternative T antigens (2013) *J Neurovirol.* 19: 298-299.

33. Hutchinson MA, Hunter T, Eckhart W. Characterization of T antigens in polyoma-infected and transformed cells (1978) *Cell* 15: 65-77.
34. Topalis D, Andrei G, Snoeck R. The large tumor antigen: a “Swiss Army knife” protein possessing the functions required for the polyomavirus life cycle (2013) *Antiviral Res.* 97: 122-136.
35. Courtneidge SA, Goutebroze L, Cartwright A, Heber A, Scherneck S, Feunteun J. Identification and characterization of the hamster polyomavirus middle T antigen (1991) *J. Virol.* 65: 3301-3308.
36. Fluck MM, Schaffhausen BS. Lessons in signaling and tumorigenesis from polyomavirus middle T antigen (2009) *Microbiol. Mol. Biol. Rev.* 73: 542-563.
37. Magnusson G, Nilsson MG, Dilworth SM, Smolar N. Characterization of polyoma mutants with altered middle and large T-antigens (1981) *J. Virol.* 39: 673-683.
38. Yi X, Freund R. Deletion of proline-rich domain in polyomavirus T antigens results in virus partially defective in transformation and tumorigenesis (1998) *Virology* 248: 420-431.
39. Cheng J, DeCaprio JA, Fluck MM, Schaffhausen BS. Cellular transformation by Simian Virus 40 and Murine Polyoma Virus T antigens (2009) *Semin Cancer Biol.* 19: 218-228.
40. Feltkamp MC, Kazem S, van der Meijden E, Lauber C, Gorbalenya AE. From Stockholm to Malawi: recent developments in studying human polyomaviruses (2013) *J Gen Virol* 94: 482-496.
41. Johne R, Buck CB, Allander T, Atwood WJ, Garcea RL, Imperiale MJ, Major EO, Ramqvist T, Norkin LC. Taxonomical developments in the family Polyomaviridae (2011) *Arch. Virol.* 156: 1627-1634.
42. Gorbalenya AE, Lieutaud P, Harris MR, Coutard B, Canard B, Kleywegt GJ, Kravchenko AA, Samborskiy DV, Sidorov IA, Leontovich AM, Jones TA. Practical application of bioinformatics by the multidisciplinary VIZIER consortium (2010) *Antiviral Res.* 87: 95-110.
43. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput (2004) *Nucleic Acids Res.* 32: 1792-1797.
44. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. Clustal W and Clustal X version 2.0 (2007) *Bioinformatics.* 23: 2947-2948.
45. Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis (2000) *Mol. Biol. Evol.* 17: 540-552.
46. Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the BEAST 1.7 (2012) *Mol Biol Evol* 29: 1969-1973.
47. Whelan S, Goldman N. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach (2001) *Mol. Biol. Evol.* 18: 691-699.

48. Drummond AJ, Ho SY, Phillips MJ, Rambaut A. Relaxed phylogenetics and dating with confidence (2006) *PLoS Biol* 4: e88.
49. Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Kosakovsky Pond SL. Detecting individual sites subject to episodic diversifying selection (2012) *PLoS Genet.* 8: e1002764.
50. Pagel M, Meade A, Barker D. Bayesian estimation of ancestral character states on phylogenies (2004) *Syst. Biol* 53: 673-684.
51. Paradis E, Claude J, Strimmer K. APE: Analyses of Phylogenetics and Evolution in R language (2004) *Bioinformatics.* 20: 289-290.
52. Soding J. Protein homology detection by HMM-HMM comparison (2005) *Bioinformatics.* 21: 951-960.
53. Schneider TD, Stephens RM. Sequence logos: a new way to display consensus sequences (1990) *Nucleic Acids Res.* 18: 6097-6100.
54. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo generator (2004) *Genome Res.* 14: 1188-1190.
55. Linding R, Jensen LJ, Diella F, Bork P, Gibson TJ, Russell RB. Protein disorder prediction: implications for structural proteomics (2003) *Structure.* 11: 1453-1459.
56. Ward JJ, Sodhi JS, McGuffin LJ, Buxton BF, Jones DT. Prediction and functional analysis of native disorder in proteins from the three kingdoms of life (2004) *J Mol. Biol.* 337: 635-645.
57. Deng X, Eickholt J, Cheng J. PreDisorder: ab initio sequence-based prediction of protein disordered regions (2009) *BMC. Bioinformatics* 10: 436.
58. Prilusky J, Felder CE, Zeev-Ben-Mordehai T, Rydberg EH, Man O, Beckmann JS, Silman I, Sussman JL. FoldIndex: a simple tool to predict whether a given protein sequence is intrinsically unfolded (2005) *Bioinformatics* 21: 3435-3438.
59. Linding R, Russell RB, Neduva V, Gibson TJ. GlobPlot: Exploring protein sequences for globularity and disorder (2003) *Nucleic Acids Res.* 31: 3701-3708.
60. Dosztanyi Z, Csizmok V, Tompa P, Simon I. IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content (2005) *Bioinformatics* 21: 3433-3434.
61. Yang ZR, Thomson R, McNeil P, Esnouf RM. RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins (2005) *Bioinformatics* 21: 3369-3376.
62. Vucetic S, Brown CJ, Dunker AK, Obradovic Z. Flavors of protein disorder (2003) *Proteins* 52: 573-584.
63. Rost B, Yachdav G, Liu J. The PredictProtein server (2004) *Nucleic Acids Res.* 32: W321-W326.
64. Buchan DW, Ward SM, Lobley AE, Nugent TC, Bryson K, Jones DT. Protein annotation and modelling servers at University College London (2010) *Nucleic Acids Res.* 38: W563-W568.

65. Gruber AR, Neubock R, Hofacker IL, Washietl S. The RNAz web server: prediction of thermodynamically stable and evolutionarily conserved RNA structures (2007) *Nucleic Acids Res.* 35: W335-W338.
66. Goodman SN. Toward evidence-based medical statistics. 2: The Bayes factor (1999) *Ann Intern Med.* 130: 1005-1013.
67. Rogozin IB, Spiridonov AN, Sorokin AV, Wolf YI, Jordan IK, Tatusov RL, Koonin EV. Purifying and directional selection in overlapping prokaryotic genes (2002) *Trends Genet.* 18: 228-232.
68. Firth C, Kitchen A, Shapiro B, Suchard MA, Holmes EC, Rambaut A. Using time-structured data to estimate evolutionary rates of double-stranded DNA viruses (2010) *Molecular Biology and Evolution* 27: 2038-2051.
69. Zheng ZM. Viral oncogenes, noncoding RNAs, and RNA splicing in human tumor viruses (2010) *Int. J. Biol. Sci.* 6: 730-755.
70. Fisher CK, Stultz CM. Constructing ensembles for intrinsically disordered proteins (2011) *Curr Opin Struct Biol* 21: 426-431.
71. Van Roey K, Gibson TJ, Davey NE. Motif switches: decision-making in cell regulation (2012) *Curr Opin. Struct Biol* 22: 378-385.
72. Perkins JR, Diboun I, Dessailly BH, Lees JG, Orengo C. Transient protein-protein interactions: structural, functional, and network properties (2010) *Structure.* 18: 1233-1243.
73. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features (1983) *Biopolymers* 22: 2577-2637.
74. Kovacs E, Tompa P, Liliom K, Kalmar L. Dual coding in alternative reading frames correlates with intrinsic protein disorder (2010) *Proc Natl Acad Sci U. S. A.* 107: 5429-5434.
75. Deribe YL, Pawson T, Dikic I. Post-translational modifications in signal integration (2010) *Nat. Struct Mol. Biol* 17: 666-672.
76. Gu X, Zou Y, Su Z, Huang W, Zhou Z, Arendsee Z, Zeng Y. An update of DIVERGE software for functional divergence analysis of protein family (2013) *Mol. Biol Evol.* 30: 1713-1719.
77. Gorbalenya AE, Snijder EJ. Viral cysteine proteinases (1996) *Persp. Drug Discov. Design* 6: 64-86.
78. Koonin EV, Gorbalenya AE. Tale of two serines (1989) *Nature* 338: 467-468.
79. King AMQ, Adams MJ, Carstens EB, Lefkowitz EJ. *Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses.* London: Academic Press; 2012.
80. Breitbart M, Rohwer F. Here a virus, there a virus, everywhere the same virus? (2005) *Trends Microbiol.* 13: 278-284.
81. Komarova AV, Haenni AL, Ramirez BC. Virus versus host cell translation love and hate stories (2009) *Adv. Virus Res.* 73: 99-170.

82. Kazem S, van der Meijden E, Wang RC, Rosenberg AS, Pope E, Benoit T, Fleckman P, Feltkamp MC. Polyomavirus-associated trichodysplasia spinulosa involves hyperproliferation, pRB phosphorylation and upregulation of p16 and p21 (2014) PLoS ONE 9: e108947.
83. Decaprio JA. How the Rb tumor suppressor structure and function was revealed by the study of Adenovirus and SV40 (2009) Virology. 384: 274-284.
84. de Souza RF, Iyer LM, Aravind L. Diversity and evolution of chromatin proteins encoded by DNA viruses (2010) Biochim. Biophys. Acta 1799: 302-318.



## Supplementary Data

### Supplementary Text

#### **Supplementary Text S1. Nomenclature of open reading frames of mammalian polyomaviruses**

Polyomaviruses use genomes with two non-overlapping protein-coding regions that are expressed either early or late in infection, respectively [1]. Each region includes open reading frames (ORFs) that produce – for some through alternative splicing – early proteins Small, Middle and Large T antigen (ST, MT and LT) and the late capsid proteins (VP2/VP3 and VP1), respectively (**Supplementary Table S2**) [2]. ST and LT are ubiquitous in all mammalian polyomaviruses. MT, however, was known only for the mouse (MPyV) and hamster (HaPyV) polyomaviruses. Recently, a third early protein called ALTO was described for MCPyV that is homologous to the C-terminal domain of MT [3].

Polyomaviruses use alternative pre-messengerRNA (primary transcript) splicing to produce mRNAs that direct synthesis of proteins. The currently used nomenclature focuses on annotation of proteins and the respective transcripts. When ORFs are named, which happens only occasionally in literature, protein-based designation is used. The produced designation could be confusing for ORFs that encode more than one protein. For instance, the first ORF in the early region is commonly called ST ORF, while it also encodes the first exon of MT and LT protein. To address this complexity, we have used in this study a rational nomenclature of ORFs designations that is independent from names of proteins/transcripts. Its rational is similar to that used to design ORFs nomenclature in other virus families with similarly complex relations between ORFs, transcripts and proteins [4]. We defined regions flanked by two stop codons as ORFs in three different reading frames. Two pairs of ubiquitous early and late ORFs were designated ORF1 and ORF2, and ORF3 and ORF4, respectively, while an optional early ORF was designated ORF5. In this regard, the early primary transcript can be alternatively spliced to merge (parts of) ORF1 and ORF2 for directing the synthesis of LT, and (parts of) ORF1 and ORF5 for directing the synthesis of MT. Alternatively, ORF1 and ORF5 can be translated directly through internal start codons, respectively, resulting into synthesis of ST and ALTO (**Supplementary Table S2**). The developed ORF nomenclature can accommodate the identification of new alternatively spliced transcripts as well as proteins expressed by canonical and non-canonical mechanisms.

#### **Supplementary Text S2. No conserved RNA secondary structure elements are evident around COCO-VA genomic site**

Conserved RNA secondary structures constrain evolution of the respective genomic regions of viruses involved. To clarify whether toggling at the COCO-VA site was affected by the presence of conserved RNA secondary structures in polyomaviruses, we used the RNAz web-based program to predict functional RNA structures based on two criteria: i) evolutionary

conservation and ii) thermodynamic stability [5]. Due to overall poor conservation of ORF5 in mammalian polyomaviruses, the analysis was conducted in monophyletic subsets of polyomaviruses for which reliable alignments were possible to produce (see also [3]). In total, nucleotide alignments for six subsets of the ORF5-plus and ORF5-less polyomaviruses were generated (**Supplementary Table S3** and **Figure 3**) and analyzed in a genomic region around the LXCXE overlapping COCO-VA site (see **M&M** for details).

Several RNA secondary structures in the tested region with a probability higher than the cutoff value of  $p=0.5$  were predicted by the RNAz program in the subsets 1, 2 and 3 of the ORF5-plus viruses (**Supplementary Figure S2**). Majority of these secondary structures were predicted on the negative-strand of the input RNA-alignments. No conserved RNA secondary structures were identified on the positive or negative-stands of other subsets. Importantly, one of the predicted RNA structures closely corresponds to the experimentally validated pre-microRNA located on the negative-strand of MCPyV RNA [6, 7] (**Supplementary Figure S3**), lending further and independent support to the results of RNA structure analysis by RNAz. Since COCO-VA site is NOT base-paired in either of the predicted RNA structures, we concluded that the evolution of the COCO-VA site is not likely constrained by RNA secondary structure elements in mammalian polyomaviruses.

### **Supplementary Text S3. Accelerated COCO-VA toggling is most associated with monophyletic Ortho-I viruses**

In addition to assessing statistical significance of difference between SPAT ratio distributions of ORF5-plus and ORF5-less viruses, we sought to assess the combinatory scale of 14/16 partitioning that would give another, top-down perspective on the probability value obtained for the difference. To this end, we ranked the difference among differences for all possible two-side partitioning of this type (14 vs. 16) for 30 viruses in the 0-0.35 mono-SPAT range. In total, 145,422,675 values were obtained and plotted as a histogram revealing highly symmetrical bell-like distribution of values (**Supplementary Figure S6**), whose difference of group means of toggling rate ratios for all possible combinations was virtually zero ( $2.7e-17$ ). The ORF5-plus/ORF5-less partitioning was ranked #284 (counted from the left side) with a difference of means of -0.2726 between the compared datasets. This high rank corresponds to a  $p$ -value of  $1.95e-06$  that was close to the  $7e-06$  value obtained in MWU test. The high ranking of the ORF5-plus/ORF5-less partitioning strongly supported the prior conclusions.

In the view of such high ranking, why this partitioning did not outrank all other partitionings? Inspection of partitioning with higher ranks showed that, like ORF5-plus viruses, many of them and including the number 1 (difference of means of -0.3051), involved a large subset of phylogenetically compact Ortho-I viruses, which was contrasted against mainly non-Ortho-I viruses. This observation prompted us to compare Ortho-I vs. non-Ortho-I viruses. Since this partitioning compares 16 to 14 viruses, its ranking could be derived from the density distribution already used to rank ORF5-plus/ORF5-less partitioning

(14 vs. 16 viruses), now from its right tail (**Supplementary Figure S6**). Its analysis showed that the Ortho-I/non-Ortho-I partitioning had a difference of mean toggling rate values of 0.2993 that ranked number three. This ranking is a two-order improvement over the ranking of ORF5-plus/ORF5-less partitioning and corresponded to a p-value of  $2.1e-08$ , which was close to  $5.5e-08$  in the MWU test (**Figure 4B**). The difference in values of mean toggling rates for the Ortho-I/non-Ortho-I partitioning was only marginally worse than those for two top-ranking 16/14 partitioning, which had 0.3007 and 0.2996 values, respectively. This result showed that the accelerated toggling is most strongly associated with the tree-based Ortho-I/non-Ortho-I partitioning.

Sequence alignments and the polyomavirus phylogeny used in this study can be found at <https://github.com/chrartin/COCOVAatoggling>.

## Supplementary Tables

Supplementary Table S1. Polyomavirus representatives used in this study\*

Accession	Abbreviation	Virus name	Host
NC_001515	MPyV	Murine polyomavirus	Mouse
NC_001663	HaPyV	Hamster polyomavirus	Hamster
JQ178241	RacPyV	Raccoon polyomavirus	Raccoon
NC_020068	EiPyV1	Eidolon polyomavirus-1	Bat
NC_020071	OtPyV1	Otomops polyomavirus-1	Bat
JQ958889	CPPyV	Carollia perspicillata polyomavirus	Bat
JQ958887	APPyV1	Artibeus planirostris polyomavirus	Bat
NC_020067	CdPyV	Cardioderma polyomavirus	Bat
HQ385749	PtvPyV2c	Pan troglodytes verus polyomavirus-2c	Chimpanzee
HQ385752	GggPyV1	Gorilla gorilla gorilla polyomavirus-1	Gorilla
NC_010277	MCPyV	Merkel cell polyomavirus	Human
HQ385746	PtvPyV1a	Pan troglodytes verus polyomavirus1a	Chimpanzee
NC_020065	CoPyV1	Chaerephon polyomavirus-1	Bat
JQ958888	SLPyV	Sturnira lilium polyomavirus	Bat
JQ958893	MMPyV	Molossus molossus polyomavirus	Bat
NC_014743	ChPyV	Chimpanzee polyomavirus	Chimpanzee
NC_019844	VePyV1	Vervet monkey polyomavirus-1	Monkey
NC_019850	PRPyV1	Ptilocolobus rufomitratus polyomavirus-1	Monkey
FN356901	OraPyV2	Orangutan polyomavirus	Orangutan
NC_014361	TSPyV	Trichodysplasia spinulosa-associated polyomavirus	Human
FN356900	OraPyV1	Orangutan polyomavirus	Orangutan
NC_019853	AtPPyV1	Ateles paniscus polyomavirus-1	Monkey
NC_015150	HPyV9	Human polyomavirus-9	Human
NC_019851	MFPyV1	Macaca fascicularis polyomavirus-1	Monkey
NC_004763	LPyV	African green monkey polyomavirus	Monkey
NC_020106	STLPyV	STL polyomavirus	Human
JX259273	MXPyV	MX polyomavirus	Human
JX262162	HPyV10	Human polyomavirus-10	Human
JQ898291	MWPyV	MW polyomavirus	Human
NC_009238	KIPyV	KI polyomavirus	Human
NC_009539	WUPyV	WU Polyomavirus	Human
NC_014407	HPyV7	Human polyomavirus-7	Human
NC_014406	HPyV6	Human polyomavirus-6	Human
JQ958890	APPyV2	Artibeus planirostris polyomavirus	Bat
JQ958891	PPPyV	Pteronotus parnellii polyomavirus	Bat
NC_020070	PDPyV	Pteronotus polyomavirus	Bat
JQ958892	DRPyV	Desmodus rotundus polyomavirus	Bat
NC_013796	CSLPyV	California sea lion polyomavirus	Sea lion
NC_001442	BPyV	Bovine polyomavirus	Cow
NC_001505	MptV	Murine pneumotropic virus	Mouse
NC_011310	BatPyV	Myotis polyomavirus VM-2008	Bat
AB588640	MasPyV	Mastomys polyomavirus	Mouse
NC_019854	CAPyV1	Cebus albifrons polyomavirus-1	Monkey
NC_009951	SqPyV	Squirrel monkey polyomavirus	Monkey
NC_020069	MiPyV	Miniapterus polyomavirus	Bat
NC_001669	SV40	Simian virus-40	Monkey
NC_001699	JCPyV	JC polyomavirus	Human
NC_001538	BKPyV	BK polyomavirus	Human
NC_007611	SA12	Simian virus-12	Monkey
JQ412134	EPyV	Equine polyomavirus	Horse
NC_004800	GHPyV	Goose hemorrhagic polyomavirus	Goose
NC_007922	CPyV	Crow polyomavirus	Crow
GU345044	CaPyV	Canary polyomavirus	Canary
NC_004764	APyV	Budgerigar fledgling disease virus-1	Parakeet
NC_007923	FPyV	Finch polyomavirus	Finch

\*Viruses are listed according to phylogenetic tree in Figure 2.

**Supplementary Table S2.** ORF designations used in this study and by others

Open reading frame nomenclature			
This study	By others	Encoding	Reference
ORF1	ST-antigen	Small T antigen	[8]
	MT-antigen	Middle T antigen 1 <sup>st</sup> exon	[9]
	LT-antigen	Large T antigen 1 <sup>st</sup> exon	[10]
ORF2	LT-antigen	Large T antigen 2 <sup>nd</sup> exon	[10]
ORF3	VP2	VP2	[11]
	VP3	VP3	[12]
ORF4	VP1	VP1	[12]
ORF5	MT-antigen	Middle T antigen 2 <sup>nd</sup> exon	[9]
	ALTO protein	ALTO protein	[3]

**Supplementary Table S3.** Virus subset compositions whose alignments were used for RNA secondary structure prediction

Subset	Polyomavirus	ORF5
1	GggPyV1, PtvPyV2c, PtvPyV1a, MCPyV APPyV1, CPPyV, CdPyV, OtPyV1, EiPyV1, RacPyV	plus
2	VePyV1, ChPyV, PRPyV1, MMPyV, SLPyV, CoPyV1	plus
3	TSPyV, AtPPyV1, OraPyV1	plus
4	MPyV, HaPyV	plus
5	HPyV9, MFPyV1, LPyV	plus
6	KIPyV, WUPyV, HPyV6, HPyV7, STLPyV, HPyV10, MXPyV, MWPyV, JCPyV, SA12, SV40, BKPyV, APPyV2, PPPyV, DRPyV, PDPyV, CAPyV1, SqPyV, MiPyV, MptV, MasPyV, BatPyV	less

**Supplementary Table S4.** W Results of natural selection analysis of conserved ORF5 positions using Datamonkey

Dataset	Genome region <sup>1</sup>	Viruses <sup>2</sup>	Codons <sup>3</sup>		MEME <sup>4</sup>		FUBAR <sup>4</sup>		TOGGLE <sup>4</sup>		
			ORF5	ORF2	#pos	COCOVA	#pos	COCOVA	#neg	COCOVA	#toggling
D1	ORF5	mamPyV	34	0	12	none	17	1	neg	15	none
D2	ORF5+ORF2	mamPyV	34	474	15	none	14	2	neg	14	none
D3	ORF5	Ortho-I	34	0	8	none	10	1	neg	7	none
D4	ORF5+ORF2	Ortho-I	34	473	6	none	5	3	neg	7	none
D5	ORF5	ORF5+	34	0	5	none	8	0	none	6	none
D6	ORF5+ORF2	ORF5+	34	473	7	none	7	2	none	5	none
D7	ORF5+ORF2	ORF5-	34	474	15	none	11	4	neg	11	none
D8	ORF5	lineageA	121	0	30	none	15	20	none	2	none
D9	ORF5	lineageB	183	0	14	none	4	15	none	1	none
D10	ORF5	lineageC	314	0	2	none	1	2	none	n.a. <sup>5</sup>	n.a. <sup>5</sup>

<sup>1</sup> ORF2: C-terminal part of ORF2 not overlapping with ORF5

<sup>2</sup> mamPyV: mammalian polyomaviruses; ORF5+: ORF5-plus viruses; ORF5-: ORF5-less viruses; lineageA: MCPyV, GgPyV1, PtvPyV1a, EIPyV, OtPyV1, CPPyV1, APPyV1, CdPyV, RacPyV; lineageB: VePyV1, PRPyV1, ChPyV, MIMPyV1, SLPyV, CoPyV1; lineageC: TSPyV (3 isolates), OraPyV1, AtPyV1

<sup>3</sup> The number and origin of codons used to infer phylogeny. For ORF5 of D1-D7, only codons whose overlapping codon in ORF2 (-1 frame) was aligned with no gaps across all mammalian polyomaviruses were chosen. For ORF2 of D2, D4, D6 and D7, and ORF5 of D8-D10, weakly aligned columns were removed manually from the alignment.

<sup>4</sup> #: number of ORF5 codons; pos: positive/diversifying selection; neg: negative/purifying selection; none: no selection/toggling detected; COCOVA: COCO-VA codon in ORF5. When using TOGGLE, only ORF5 codons have been screened for toggling

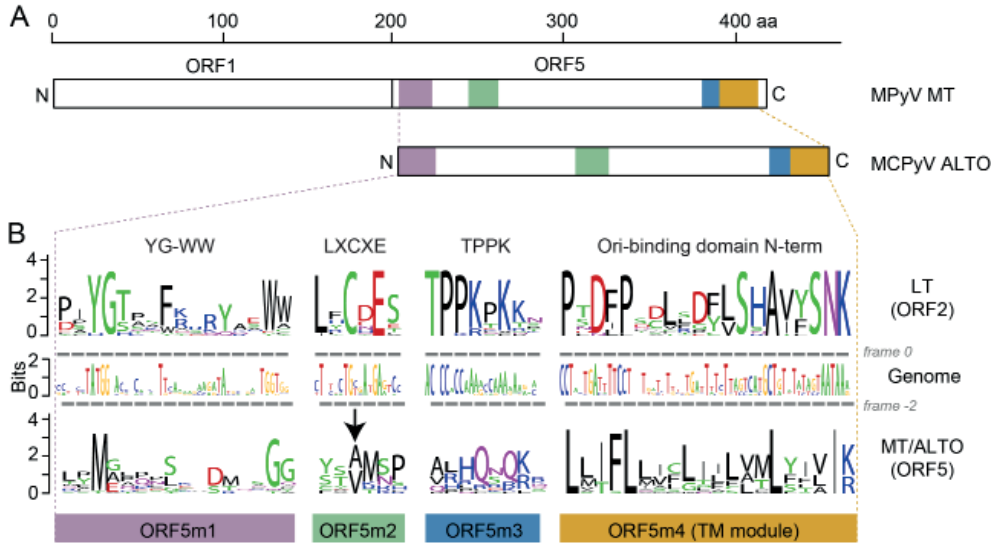
<sup>5</sup> n.a., non-available: no results were obtained due to small (<10) number of tree branches

**Supplementary Table S5.** Shown are p-values of a Wilcoxon rank sum test for deviation of the mean of a distribution of SPAT ratios from the matching rate

Virus group	MonoSPAT range		
	0 - 0.35	0.35 - 1.25	0 - 1.25
ORF5-plus	0.390	<b>0.039</b>	0.656
ORF5-less	<b>3.1e-5</b>	0.064	<b>0.003</b>
Ortho-I	0.211	<b>0.020</b>	0.672
non-Ortho-I	<b>1.2e-4</b>	0.129	<b>0.001</b>

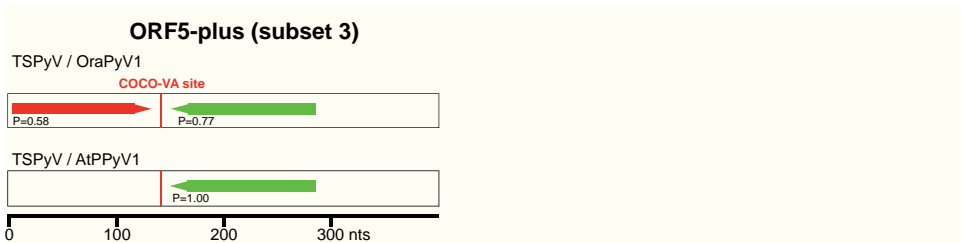
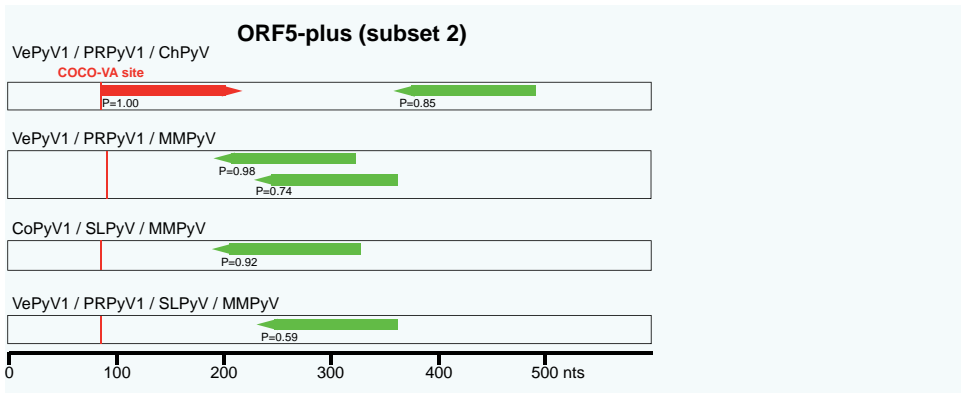
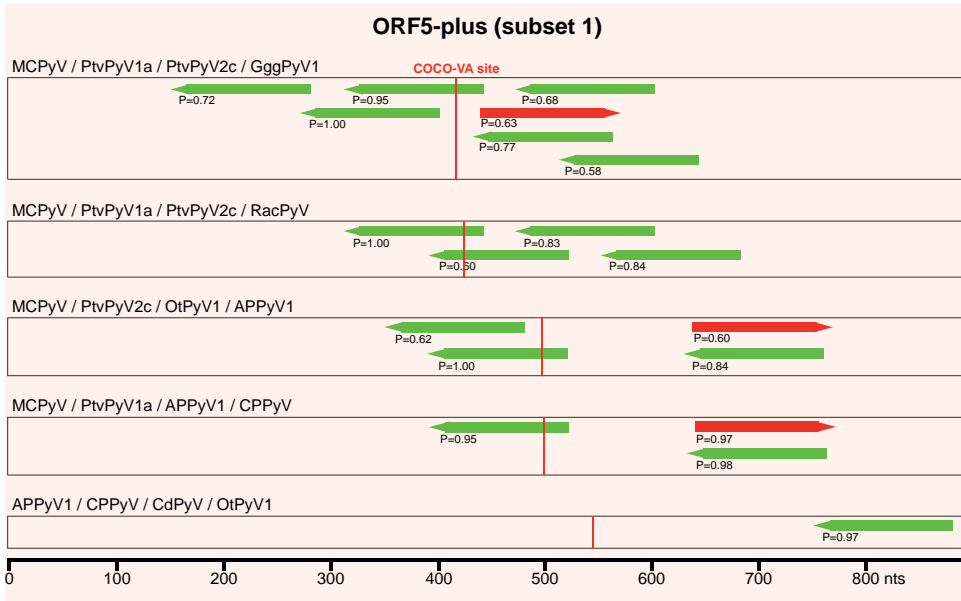
Bold, statistically significant p-values ( $\alpha=0.05$ )

## Supplementary Figures



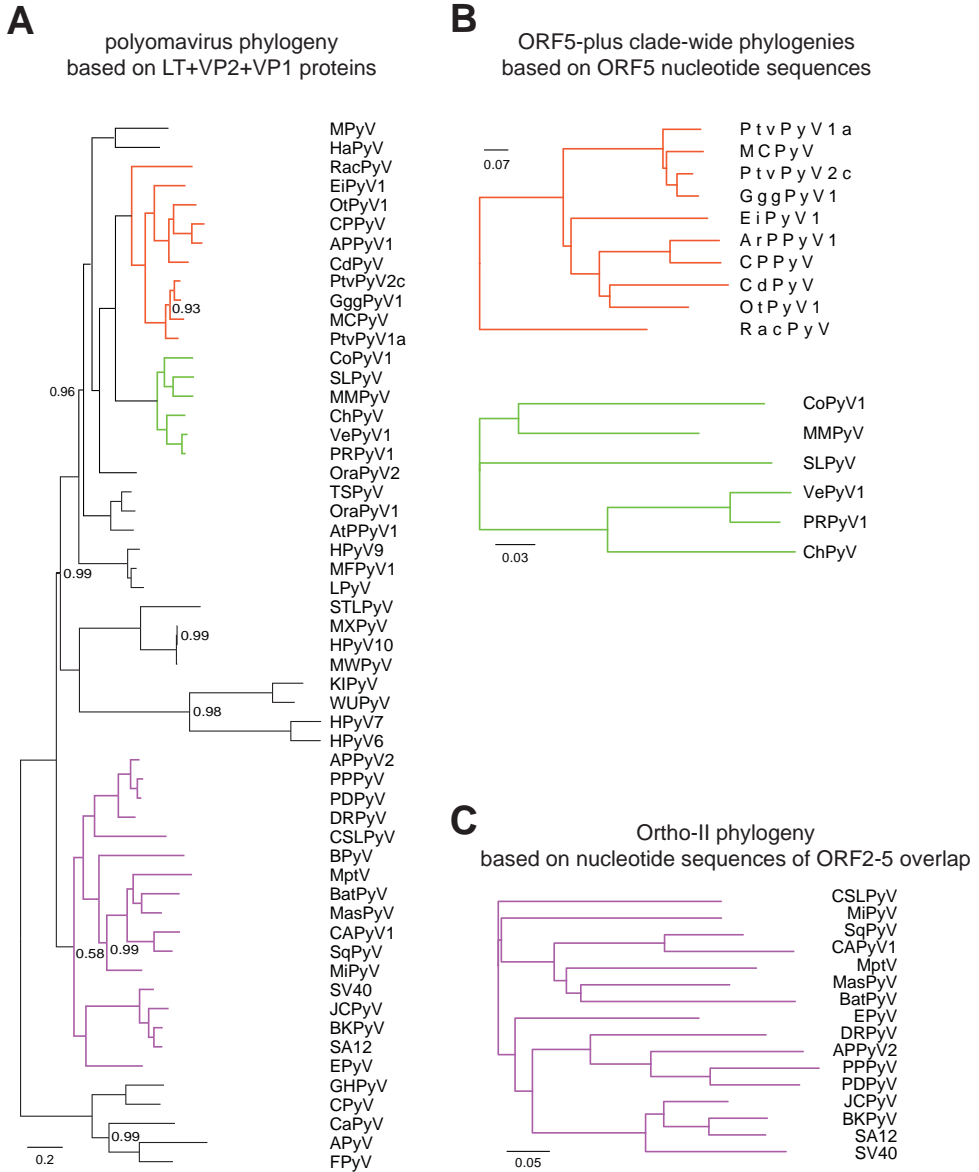
**Supplementary Figure S1.** Conserved features of ORF5 of the Ortho-I group. **(A)** domain organization of two different proteins, partially or fully encoded in ORF5, that were experimentally characterized in MPyV and MCPyV. The ORF1-encoded MT domain is not further specified. The ORF5-encoded domain includes four conserved motifs, ORF5m1 - ORF5m4, colored differently and detailed in panel **B**. The ORF5 remains open upstream of the start codon for MCPyV ALTO. **(B)** Conserved motifs in the ORF5/ORF2 overlap. Presented are sequence logos of four conserved motifs in LT ORF2 (bottom), their counterparts in MT/ALTO ORF5 (top), and the corresponding genome region (middle). The logos are based on alignments of the 22 viruses of the Ortho-I group (see **M&M**). Gray bars indicate the codon structure of the LT ORF2 and MT/ALTO ORF5 reading frames. The arrow indicates the COCO-VA toggling position.



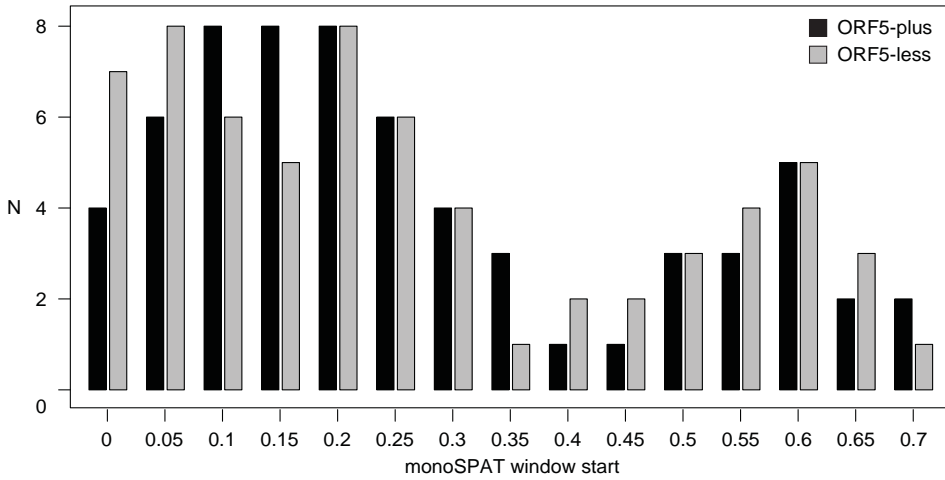


**Supplementary Figure S2.** Prediction of RNA secondary structure consensus using RNAz. ORF5-plus viruses from subset 1 (top panel), subset 2 (middle panel) and subset 3 (bottom panel) alignment demonstrated regions of RNA secondary structures in several combination of viruses. The predicted secondary structures from the negative-strand (green bars) and positive-strand (red bars), within the given alignment of indicated viruses with a probability value higher than 0.5, are shown. Position of the COCO-VA in the given alignment is illustrated by horizontal red line.

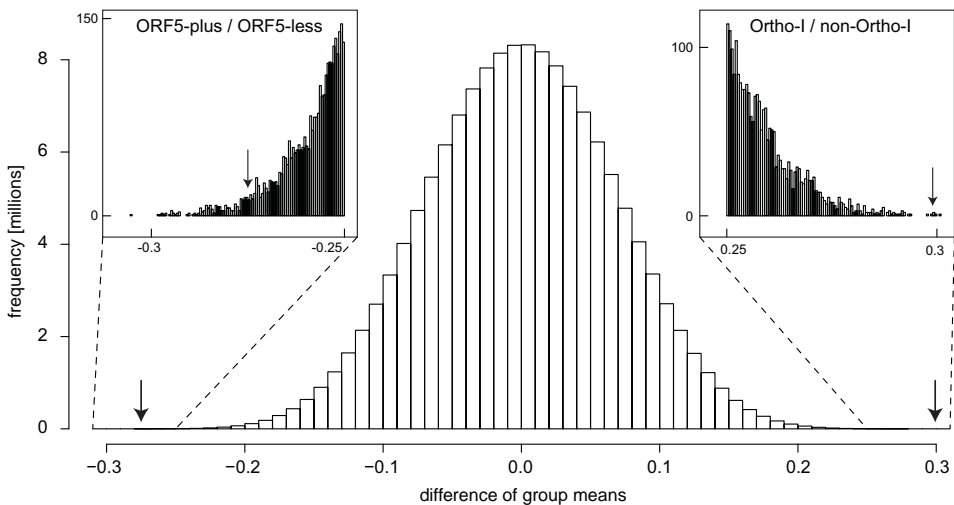




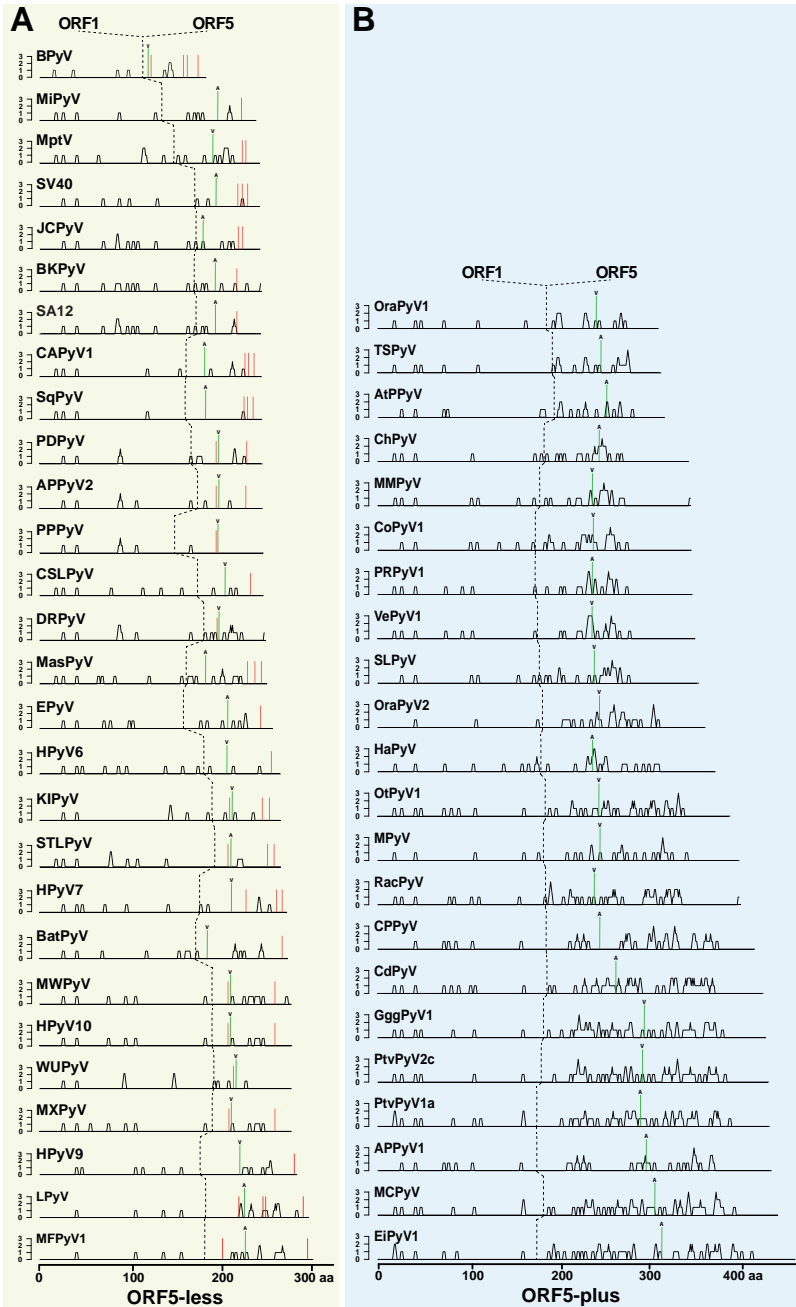
**Supplementary Figure S4.** Congruence of lineage-specific phylogenies with the polyomavirus tree. Compared is the topology of the polyomavirus tree estimated using LT, VP1, and VP2 proteins (**A**) with that of two ORF5-plus monophyletic lineages independently estimated using ORF5 codon alignments (**B**) and of Ortho-II viruses estimated using nucleotide alignments of the ORF2-ORF5 overlapping genomic region (**C**). The lineages are indicated by color. The trees in **B** and **C** are neighbor joining trees used in the Datamonkey analyses.



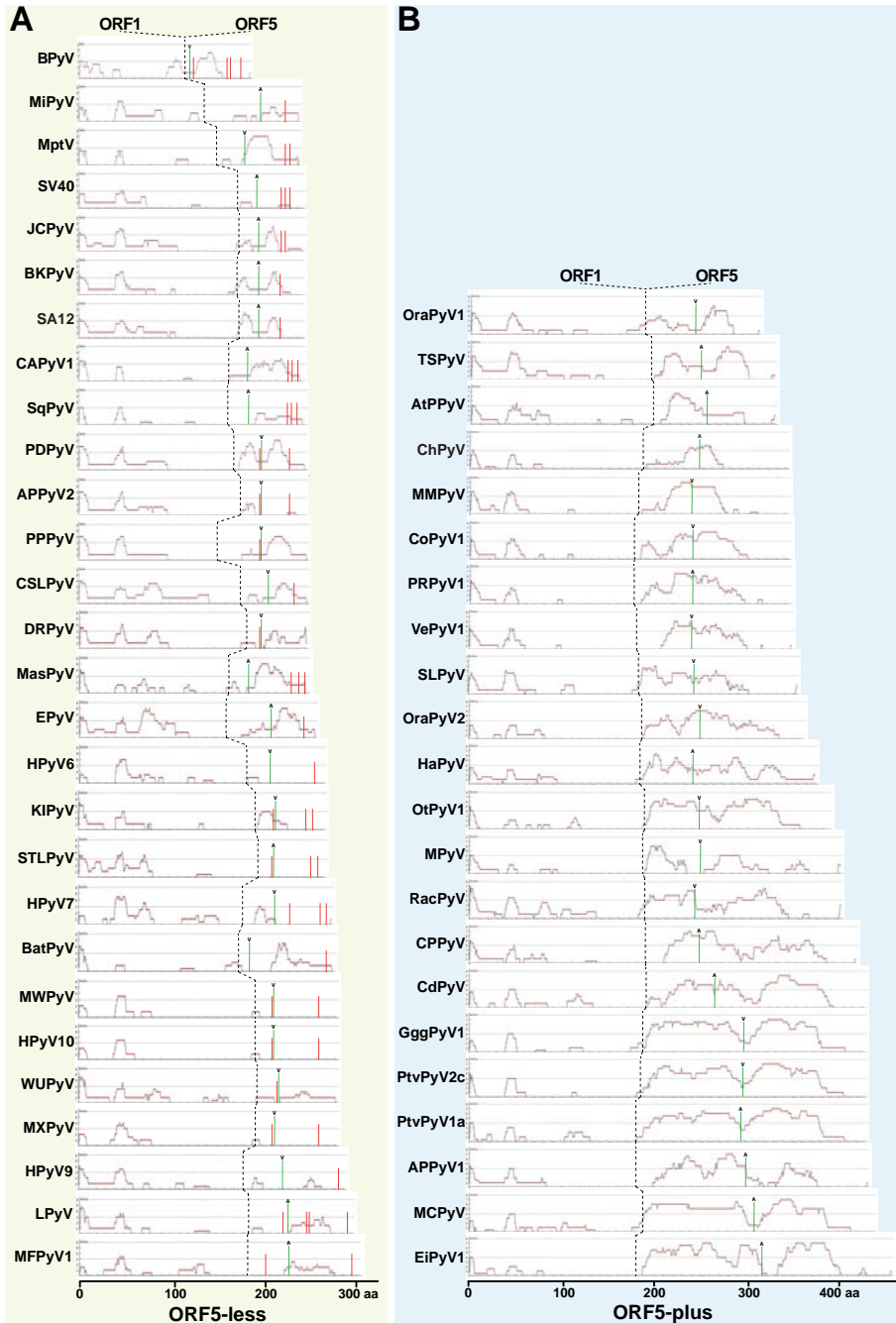
**Supplementary Figure S5.** Virus sampling in relation to monoSPAT. The number of ORF5-plus and ORF5-less viruses per sliding window (size: 0.15, shift: 0.05) along the monoSPAT scale is shown.



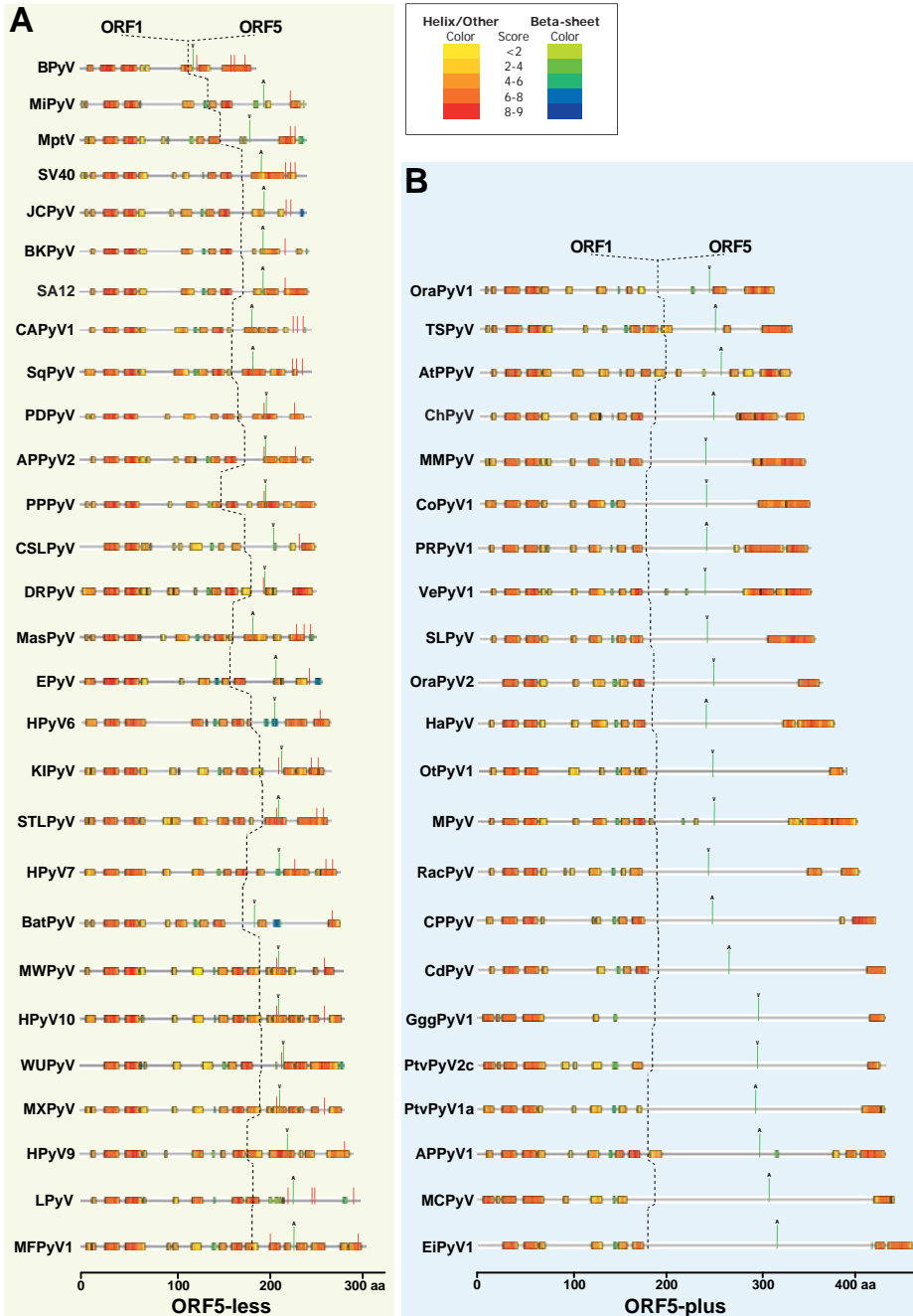
**Supplementary Figure S6.** Ranking test. Shown are differences of mean SPAT ratios between groups for all possible 14-16 partitionings of the 30 viruses with monoSPAT values in the range of 0-0.35. These are 145,422,675 combinations in total. Values for the Ortho-I/non-Ortho-I and for the ORF5-plus/ORF5-less partitionings are indicated by arrows and insets.



**Supplementary Figure S7.** Enrichment of Proline residues near the COCO-VA site in ORF5-less (A) and ORF5-plus (B) viruses. Shown is the number and location of Proline residues along the translated putative MT polypeptide sequence of the mammalian polyomaviruses analyzed in this study. The green and red bars indicate the COCO-VA site and termination codons, respectively. The dashed line indicates the ORF1-ORF5 splice junction. ORF5-plus viruses show an enrichment of Proline residues at the ORF5-encoded part of MT, which is a characteristic of intrinsically disordered protein regions.



**Supplementary Figure S8.** The COCO-VA site is embedded in an intrinsically disordered protein region in ORF5-less (A) and ORF5-plus (B) viruses. Shown is the prediction of protein disorder along the translated putative MT polypeptide sequence of the mammalian polyomaviruses analyzed in this study. The green and red bars indicate the COCO-VA site and termination codons, respectively. The dashed line indicates the ORF1-ORF5 splice junction. Large parts of the ORF5-encoded part of MT are predicted to be disordered for ORF5-plus viruses.

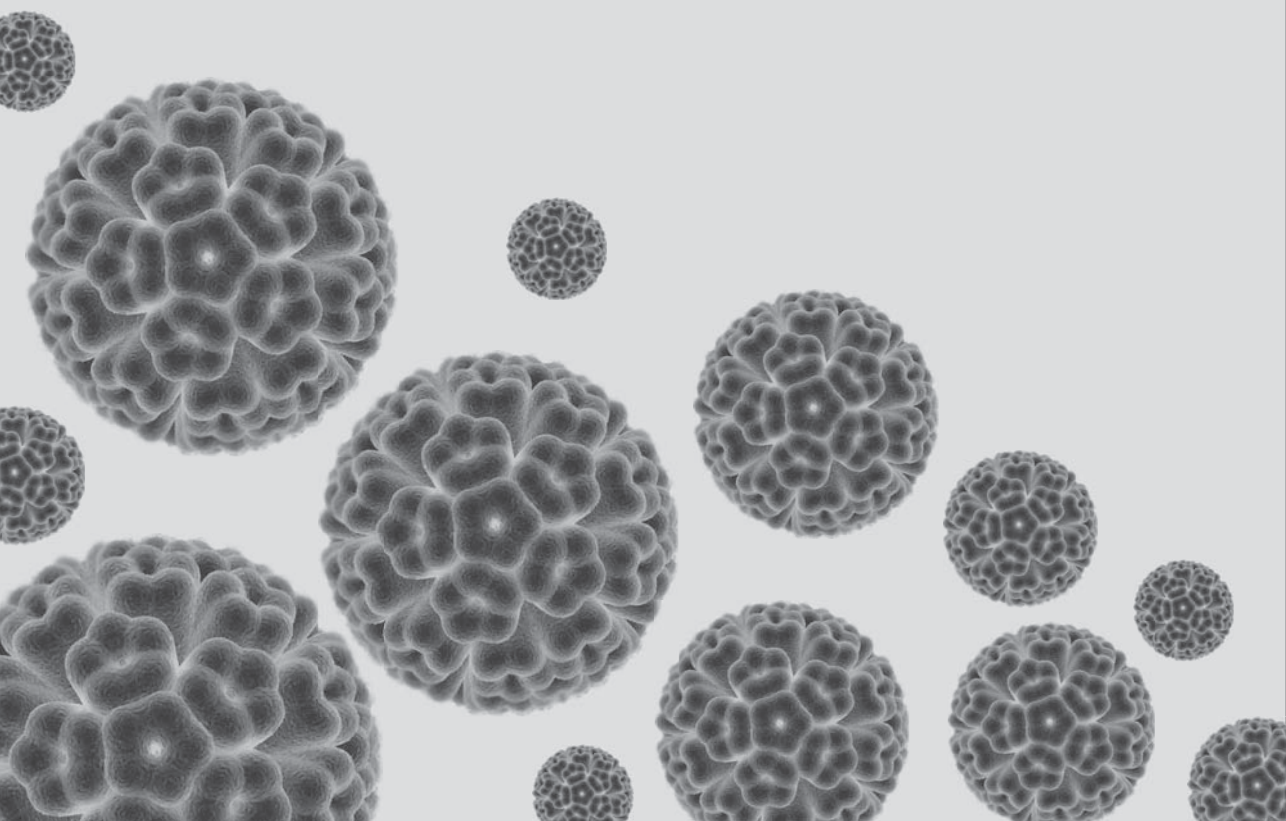


**Supplementary Figure S9.** Protein secondary structure and COCO-VA site in ORF5-less (A) and ORF5-plus (B) viruses. Consensus of two protein secondary structure predictors is shown along the translated putative MT polypeptide sequence of the mammalian polyomaviruses. Red and green bars indicate termination codons and COCO-VA site, respectively, and dashed lines suggest the putative ORF1-ORF5 splice junction. On top, a color-coded scoring table for Helices and Sheets is shown.

## Supplementary References

1. Broekema NM, Imperiale MJ. miRNA regulation of BK polyomavirus replication during early infection (2013) *Proc. Natl. Acad. Sci. U. S. A* 110: 8200-8205.
2. Decaprio JA, Garcea RL. A cornucopia of human polyomaviruses (2013) *Nat Rev Microbiol.* 11: 264-276.
3. Carter JJ, Daugherty MD, Qi X, Bheda-Malge A, Wipf GC, Robinson K, Roman A, Malik HS, Galloway DA. Identification of an overprinting gene in Merkel cell polyomavirus provides evolutionary insight into the birth of viral genes (2013) *Proc Natl Acad Sci U. S. A.* 110: 12744-12749.
4. Komarova AV, Haenni AL, Ramirez BC. Virus versus host cell translation love and hate stories (2009) *Adv. Virus Res.* 73: 99-170.
5. Gruber AR, Neubock R, Hofacker IL, Washietl S. The RNAz web server: prediction of thermodynamically stable and evolutionarily conserved RNA structures (2007) *Nucleic Acids Res.* 35: W335-W338.
6. Seo GJ, Chen CJ, Sullivan CS. Merkel cell polyomavirus encodes a microRNA with the ability to autoregulate viral gene expression (2009) *Virology* 383: 183-187.
7. Lee S, Paulson KG, Murchison EP, Afanasiev OK, Alkan C, Leonard JH, Byrd DR, Hannon GJ, Nghiem P. Identification and validation of a novel mature microRNA encoded by the Merkel cell polyomavirus in human Merkel cell carcinomas (2011) *J. Clin. Virol.* 52: 272-275.
8. Cho US, Morrone S, Sablina AA, Arroyo JD, Hahn WC, Xu W. Structural basis of PP2A inhibition by small t antigen (2007) *PLoS Biol.* 5: e202.
9. Fluck MM, Schaffhausen BS. Lessons in signaling and tumorigenesis from polyomavirus middle T antigen (2009) *Microbiol. Mol. Biol. Rev.* 73: 542-563.
10. Topalis D, Andrei G, Snoeck R. The large tumor antigen: a “Swiss Army knife” protein possessing the functions required for the polyomavirus life cycle (2013) *Antiviral Res.* 97: 122-136.
11. Stehle T, Yan Y, Benjamin TL, Harrison SC. Structure of murine polyomavirus complexed with an oligosaccharide receptor fragment (1994) *Nature* 369: 160-163.
12. Griffith JP, Griffith DL, Rayment I, Murakami WT, Caspar DL. Inside polyomavirus at 25-A resolution (1992) *Nature* 355: 652-654.





# Chapter 6

## Constrained TSPyV evolution in human population

**Adaptation of trichodysplasia spinulosa-associated polyomavirus to the human population is mediated by middle T antigen and involves COCO-VA toggling**

*Authors:*

Siamaque Kazem<sup>1\*</sup>

Chris Lauber<sup>1, 2\*</sup>

Els van der Meijden<sup>1</sup>

Sander Kooijman<sup>1</sup>

Alexander Kravchenko<sup>3†</sup>

Mariet Feltkamp<sup>1#</sup>

Alexander Gorbalenya<sup>1, 3, 4#</sup>

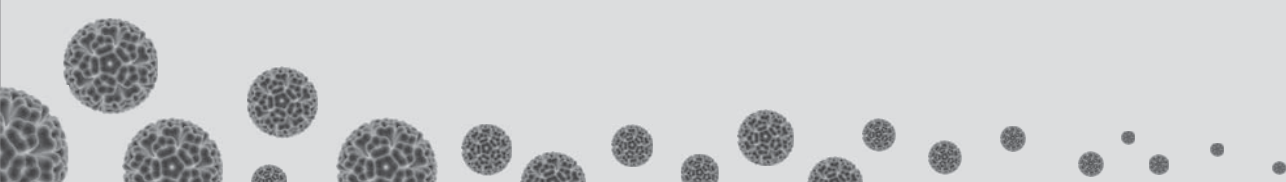
*\* These first authors and # last authors contributed equally to the paper*

*† Deceased; of blessed memory*

*Affiliations:*

<sup>1</sup>Department of Medical Microbiology, Leiden University Medical Center, Leiden, The Netherlands, <sup>2</sup> Institute for Medical Informatics and Biometry, Technische Universität Dresden, Dresden, Germany, <sup>3</sup>Belozersky Institute of Physico-Chemical Biology, Lomonosov Moscow State University, Moscow, Russia and <sup>4</sup> Faculty of Bioengineering and Bioinformatics, Lomonosov Moscow State University, Moscow, Russia

Submitted for publication



## Abstract

Trichodysplasia spinulosa-associated polyomavirus (TSPyV) causes a rare disease in immunocompromised individuals, TS, while infecting up to 70% of the human population. To gain insight into the evolution of TSPyV, we have analyzed thirteen mostly newly sequenced TSPyV genomes representing ~40% of reported TS cases world-wide. In TSPyV, genome variation was limited ( $\leq 0.6\%$ ) and accumulated along three phylogenetic lineages, whose topology correlated poorly with host age and residence, and the elapsed time since sample collection. Consequently, the rate and recent timeframe of the TSPyV evolution were estimated with large uncertainty. The observed substitutions were disproportionately frequent in the non-coding control region, with a substantial fraction being observed in networks of co-evolving sites across all major genome regions. In the coding region, non-synonymous substitutions (NSS) were only present in the open reading frames (ORFs) encoding small T (ST) and middle T (MT) antigens. They included ORF1 (one NSS) and the newly identified ORF5 (three NSS), which were under positive selection. One NSS fixed most early in TSPyV evolution involves a Val/Ala codon in a short linear motif of an intrinsically disordered, Proline-rich domain of MT-antigen. In **Chapter 5**, we identified this codon to be under strong positive selection in ORF5-encoding polyomaviruses and named its variation Codon-Constrained Val-Ala (COCO-VA) toggling. Thus the combined results of this and **Chapter 5** show that the COCO-VA toggling within MT-antigen may drive both intra- and inter-host adaptation of a large, monophyletic group of polyomaviruses that includes TSPyV.

## Introduction

The fast growing *Polyomaviridae* family includes viruses infecting diverse vertebrate species [1 - 3]. Humans are commonly infected with polyomaviruses at young age after which a persistent, latent infection sets in before virus reactivates, typically when the immune system is compromised [4]. Polyomaviruses use small circular double stranded DNA genomes with two non-overlapping protein-coding regions that are expressed either early or late in infection [5, 6]. Each region includes two ubiquitous open reading frames (ORFs), ORF1-ORF2, and ORF3-ORF4, respectively, that are expressed through alternative splicing to produce the early small and large T antigens (ST and LT) [5] and the late capsid proteins (VP1 and VP2/VP3), respectively [7]. Further details regarding polyomavirus ORFs and their expression can be found in **Chapter 5** [8].

As pointed out in **Chapter 5** [8], mouse (MPyV) and hamster (HaPyV) polyomaviruses encode an extra ORF, called ORF5, that largely overlaps with the 5'-end of ORF2. It is merged with the ORF1 in a spliced transcript that directs the synthesis of middle T (MT) antigen [9 - 12]. The ORF5-encoded part of MT-antigen is enriched with Proline residues [13, 14] and includes a C-terminal transmembrane domain [10] that is essential for the oncogenic function of MT-antigen [11]. This function and interaction of MT-antigen with different cellular proteins may be modulated by phosphorylation at several Serine, Threonine and Tyrosine residues in MPyV [12]. The ORF5 (named also ALTO, [15]) was recently identified to be conserved within a monophyletic cluster of mammalian polyomaviruses that includes the Merkel cell (MCPyV) and trichodysplasia spinulosa-associated polyomaviruses (TSPyV) [8]. This large cluster comprises most of the *Orthopolyomavirus* genogroup I [16].

The human polyomavirus genome could diverge by as much as 5.3% upon virus circulation in a host population, as was shown for Malawi polyomavirus (MWPyV) [17], and 0.55% within an individual, as was shown for BK polyomavirus (BKPv) [18], with mutations preferentially accumulating in the NCCR [19, 20]. Depending on the method and dataset used, estimated evolutionary rates of human polyomaviruses varied broadly from  $10^{-5}$  to  $10^{-8}$  substitutions/site/year. The slow rate estimates were originally obtained under the model of virus-host codivergence [21, 22]. On the other hand, application of Bayesian approach to this problem using time-structured data favors high evolutionary rates [18, 22, 23].

Among the recently discovered human polyomaviruses, TSPyV is associated with a rare skin disease, TS, exclusively seen in severely immunocompromized patients (**Chapters 2 and 3**) [24 - 32]. Etiologic involvement of TSPyV in TS is evident from high viral load, antigen expression and cell cycle deregulation in affected skin (**Chapters 3 and 4**) [30]. TSPyV is globally circulating with seroprevalences of approximately 70% in the general adult population, as demonstrated in Europe [33, 34] and Australia [35]. In asymptomatic individuals, healthy and immunocompromized, TSPyV DNA is seldom detected [30, 31, 36].

Based on phylogenetic analysis of the conserved parts of VP1, VP2 and LT-antigen, TSPyV as well as MPyV, HaPV, and MCPyV belong to the large *Orthopolyomavirus-I* (Ortho-I) genogroup [1, 16]. Most other studied mammalian polyomaviruses, including BKPyV and JC polyomavirus (JCPyV) [37] and Simian virus 40 (SV40), belong to the *Orthopolyomavirus-II* (Ortho-II) genogroup, while majority of recently identified human polyomaviruses [36, 38 - 47] belong to either the *Malawipolyomavirus* (Malawi) or the *Wukipolyomavirus* (Wuki) genogroups (**Chapter 1**) [1, 16].

In this chapter, we describe evolutionary analysis of 13 TSPyV genomes, most newly sequenced and representing ~40% of the reported TS cases world-wide. We demonstrate that the newly identified ORF5 [8, 15] is the major site of adaptive non-synonymous substitutions during highly constrained evolution of TSPyV. The non-synonymous substitution fixed most early in the TSPyV evolution involves codon-constrained Val-Ala (COCO-VA) toggling at the site of positive selection in the inter-species evolution of the Ortho-I polyomaviruses, as explained in **Chapter 5** [8]. Thus, the combined results reveal a connection between micro- and macro-evolution of polyomaviruses and establish TSPyV as a virus to study this connection.

## Material and Methods

### Samples, DNA extraction and PCR amplification

As described previously [30], DNA was extracted from ten TS samples in our laboratory, and extracted DNA was received from three TS samples. An overview of all TSPyV isolates analyzed in this study and their whereabouts is given in **Table 1**. The DNA isolates were amplified by PCR with the help of 18 TSPyV-specific primer-pairs (**Supplementary Table S1**). This resulted in generation of eighteen overlapping TSPyV DNA fragments of 266-613 base-pairs in length that cover the whole TSPyV genome (**Supplementary Figure S1A**). Ten nanogram input DNA was used in a total volume of 50  $\mu$ l containing 15 pmol/ $\mu$ l of each primer, 2.5 U of *Pfu* DNA polymerase, 10x *Pfu* buffer and 15 mM of dNTP's. A 3-step PCR amplification was performed; starting with 5 minutes at 95°C, followed by 30 cycles of 1 minute at 95°C, 1 minute at 56°C and 2 minutes at 72°C, and finalized with 10 minutes at 72°C. Subsequently, one unit of GoTaq DNA polymerase was added to the PCR reaction, to generate an adenine-base overhang for the subsequent TOPO(TA)-cloning (Invitrogen, California, USA), and the reaction was incubated for an additional 10 minutes at 72°C. PCR products were assessed for length by gel-electrophoresis (**Supplementary Figure S1B**) and cloned directly. In case of multiple, potentially nonspecific products detected on gel, bands of the expected length were isolated from gel using the QIAEX II gel extraction kit (QIAGEN, California, USA) and then cloned into TOPO(TA) vector. The amplicons generated from samples TSPyV-1211, 1312 and 1412 were purified with GeneJET PCR Purification Kit (Fermentas) and subsequently sequenced directly in both directions.

**Table 1.** Overview of all TSPyV genomes analyzed in this study

Name isolate <sup>1</sup>	Case origin	Sample treatment	TSPyV DNA load <sup>2</sup>	GenBank Ref. Seq.	Reference
TSPyV-0109	NLD <sup>3</sup> , 's Hertogenbosch (N-Brabant)	FRFR <sup>3</sup>	7.2x10 <sup>4</sup>	NC_014361	[25]
TSPyV-0203	AUS, Nedlands (Western Australia)	FFPE	7.6x10 <sup>5</sup>	KF444091	[48]
TSPyV-0304	USA, New York (New York)	FFPE	4.5x10 <sup>6</sup>	KF444092	[49]
TSPyV-0409	USA, Cleveland (Ohio)	FFPE	1.2x10 <sup>5</sup>	KF444093	[30]
TSPyV-0510	USA, Cleveland (Ohio)	FFPE	4.4x10 <sup>6</sup>	KF444094	[30]
TSPyV-0602	USA, Birmingham (Alabama)	FFPE	6.3x10 <sup>4</sup>	KF444095	[30]
TSPyV-0797	USA, Birmingham (Alabama)	FFPE	7.8x10 <sup>1</sup>	- <sup>4</sup>	[50]
TSPyV-0807	CAN, Toronto (Ontario)	FFPE	5.1x10 <sup>4</sup>	- <sup>4</sup>	[29]
TSPyV-0910	USA, Dallas (Texas)	FFPE	1.4x10 <sup>6</sup>	KF444096	[26]
TSPyV-1008	USA, Charleston (South Carolina)	FFPE	5.6x10 <sup>5</sup>	KF444097	[51]
TSPyV-1197	USA, Seattle (Washington)	FFPE	1.7x10 <sup>6</sup>	KF444098	[24]
TSPyV-1211	NLD, Rotterdam (Zuid-Holland)	FFPE	2.8x10 <sup>4</sup>	KF444099	This study
TSPyV-1312	USA, Dallas (Texas)	FFPE	2.1x10 <sup>4</sup>	KF444100	This study
TSPyV-1412	AUS, Melbourne (Victoria)	FRFR	9.4x10 <sup>4</sup>	KF444101	This study
TSPyV-WK164	USA, St. Louis (Missouri)	-	2.4x10 <sup>4</sup>	JQ723730	[17]

<sup>1</sup> The first two digits correspond to the trichodysplasia spinulosa (TS) sample number used to describe the TS patients in Kazem *et al.* [30]. The last two digits indicate the year of sampling ranging from 1997 (xx97) to 2012 (xx12). TSPyV-WK164 was isolated in 2009 [17].

<sup>2</sup> TSPyV-load is shown in virus copies/cell, as was calculated and reported previously [25, 30]. For TSPyV-WK164 the virus copies was calculated per mL [17].

<sup>3</sup> Abbreviations: NLD, The Netherlands; AUS, Australia; USA, United States of America; CAN, Canada; FRFR, fresh frozen; FFPE, formalin-fixed paraffin embedded

<sup>4</sup> Attempts to amplify sequences failed, possibly due to either low viral load (TSPyV-0797) or poor DNA quality (TSPyV-0807).

## Cloning, sequencing and SNP analysis

TOPO-cloned plasmid DNA was isolated from at least three independently grown cultures and sequenced using the BigDye Terminator v1.1 cycle sequencing kit (Applied Biosystems, California, USA) in the ABI Prism 3130 Genetic analyzer, according to the manufacturer's instructions. M13-forward or M13-reverse primers were used to amplify the flanking TSPyV sequences inserted into the pCRII-TOPO vector. Obtained sequences were assembled into one contig with CLC Main Workbench 6 (CLC Bio; Aarhus, Denmark). All generated complete contigs, including the sequence of TSPyV-0109 (NC\_014361) (also serving as the reference TSPyV genome) [25] and TSPyV-WK164 (JQ723730) [17], were aligned using CLC Main Workbench 6 (CLC Bio; Aarhus, Denmark) and SNPs were analyzed. All viral sequences obtained in this study were deposited into GenBank and the corresponding assigned accession numbers are documented in **Table 1**.

## Splice-donor and splice-acceptor site prediction

Putative donor-acceptor splice sites were calculated under default settings using the web-based Human Splice Finder program (<http://www.umd.be/HSF>) [52]. Sites were searched in the ~120 nt sequence separating ORF1 and ORF5. The experimentally identified donor-acceptor sites for the MT mRNAs of HaPyV (NC\_001663) (HSF scores: donor, 76% and acceptor, 92%) and MPyV (NC\_001515) (HSF scores: donor, 89% and acceptor, 86%) were predicted by

this program as best in-frame sites. For TSPyV (NC\_014361), only one in-frame donor-acceptor coordinate was observed (HSF scores: donor, 93% and acceptor, 74%). Illustration of the ORF1 and ORF5 splicing was generated with Adobe Illustrator CS5 software (**Supplementary Figure S2**).

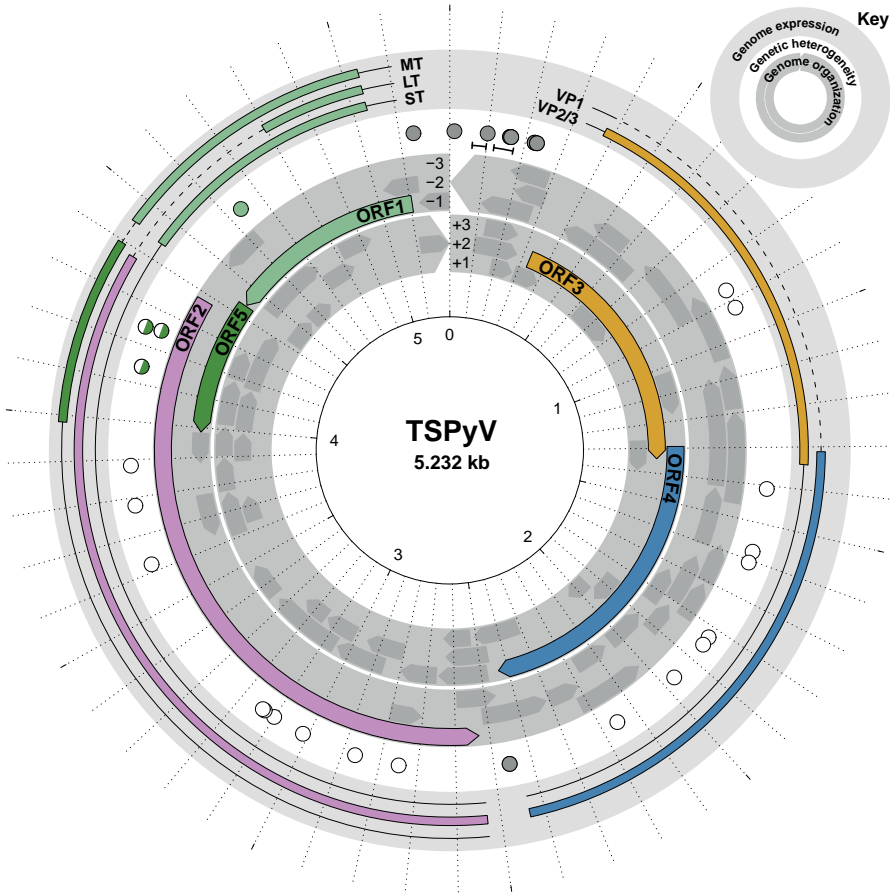
## Visualization of ORF organization, transcriptome and genetic variation in TSPyV

To facilitate analyses described in this chapter and communication of the obtained results, we have developed and used a new format for the presentation of genomic-based data. The currently used formats focus on presenting transcripts and differ primarily in choice of the direction for the early and late transcripts, either of which can be used as running clock- or counter clockwise. On the other hand, presentation of ORFs (genomic organization) is given less attention in polyomaviruses. In our format, ORFs and transcripts are presented in separate layers (**Figure 1**) that reveal their complexity and relation in detail (see also **Chapter 5**). Particularly, ORFs are depicted in three positive (+1, +2, +3) and three negative (-1, -2, -3) RFs that encode late and early transcripts, respectively. For the purpose of this study, ORF is defined as a region of a RF that is flanked by two stop codons; only a portion of this may be actually expressed. As we show elsewhere in this chapter, these details are crucial for understanding of the biology of TSPyV and other polyomaviruses, particularly when combined with genetic variation that is presented in another layer separating those for genome and transcriptome (**Figure 1**).

## Evolutionary analyses of TSPyV genomes

Genomes of 13 TSPyV and other polyomaviruses available at the Genbank/RefSeq database on February 2013 were downloaded into the Viroal platform [53]. The Muscle program [54] was used to generate family-wide multiple amino-acid alignments for VP1, VP2 and LT, followed by manual curation [8]. The alignment of TSPyV isolates and the closely related Bornean orang-utan polyomavirus (OraPyV1) was extended genome-wide at the nucleotide level. Multiple sequence alignments of three datasets were submitted to phylogeny reconstruction and other evolutionary analyses: (i) a complete genome alignment of the 13 TSPyVs (DS1), (ii) a complete genome alignment of the 13 TSPyVs and OraPyV1 (DS2), and (iii) a concatenated codon alignment that included the five ORFs of the 13 TSPyV genomes (DS3).

Phylogenetic analyses were performed by using a Bayesian approach implemented in BEAST version 1.7.4 [55]. For DS2 the dates of virus isolation were incorporated in the analysis in order to estimate the time of divergence of ancestral nodes in the tree. The dates for TSPyV isolates are shown in **Table 1**; the OraPyV1 isolate originates from 1996 (Ernst Verschoor, personal communication). The HKY nucleotide substitution model [56] (DS1 and DS2) was used; for each dataset, rate heterogeneity among sites was modeled using a gamma distribution with four categories, and a relaxed molecular-clock approach was tested



**Figure 1.** Genome organization and variation, and putative transcriptome of TSPyV. This figure has three circular layers, inner, intermediate, and outer, respectively, as detailed in the key inset at top right. The inner layer depicts the six reading frames of the TSPyV genome, frame +1 starts at the first nucleotide of the linearized genome clock-wise and frame -1 at the last nucleotide in reverse direction; the five largest ORFs and other ORFs of at least 100nts in length are indicated by colors and in gray, respectively. The intermediate layer depicts SNPs as circles of three shades: gray, SNPs in non-coding regions, NCCR and IR; white, synonymous substitutions; colored (according to the respective ORF), non-synonymous substitutions. Note the duality of the substitutions in the ORF2/ORF5 overlapping region that were colored accordingly. Deletions in the NCCR are shown as range symbols. The outer layer depicts predicted spliced transcripts; their coding parts are colored according to the respective ORFs, while untranslated regions and predicted introns are shown as solid and dashed lines, respectively.

against the strict molecular-clock approach [57]. The latter was favored for DS1 [58], whereas the relaxed molecular-clock approach was superior for DS2. MCMC chains were run for 20 million steps (DS1 and DS2) and the first 10% were discarded as burn-in. Convergence of the runs was verified and parameter estimates, including evolutionary rate and tMRCA estimates, were obtained using Tracer [59].

For calculation of evolutionary rates and tMRCA, time of virus isolation from patient is used to time-calibrate changes in genome sequences. The involved analyses saliently



assume that the time of virus collection is very close to the time of the virus infection. That link does exist for acute infections, but it may be broken for latent infections in which a virus may stay in a patient without replicating genome and accepting mutations for prolonged time (years), possibly since the initial infection. Polyomaviruses are known to go into latent infection and TSPyV is not likely to be an exception. Consequently, dates of the TSPyV isolation from patients may misinform time-dependent evolutionary analyses. To address this complication we used DS1 to (re-)estimate the date of patient primary infection for each of the 13 TSPyVs and compare it to the respective date of virus isolation. We treated the dates of the patient birth and the virus isolation as the possible lower and upper limits, respectively, of the date of the original infection for this patient. When the age information was not available at the time of analyses (one patient: TSPyV-1412), we set the patient's age to 100 years. These ranges (one for each of the 13 TSPyVs) were used to define uniform priors for the dates of the tips in the TSPyV phylogeny and to subsequently estimate the posterior distribution of these dates using BEAST as described above.

Sites under positive and negative selection were identified using the Datamonkey webserver [60]. The SLAC, FEL, IFEL, REL, MEME, and FUBAR methods were applied under default parameters on DS3. These methods implement different approaches for the detection of positive and negative selection, as well as episodic selection. For a detailed discussion on the different approaches see publications [60 - 63]. Coevolution of genomic positions was assessed by using the Spidermonkey method [64] at the Datamonkey webserver using DS1.

### **Phylogenetic association of amino-acid variations and place of TSPyV origin**

We used the BaTS tool [65] to test whether there is any association of the observed amino-acid variation at a particular site with lineages in the underlying phylogeny. To do so, we treated the amino acids observed at a particular site as the varying trait. Briefly, BaTS compares the observed distribution of the trait values (amino acid residues in this case) among the lineages in the phylogeny against the null hypothesis of randomly distributed trait values using the following three test statistics: parsimony score (PS) [66], association index (AI) [67], and maximum monophyletic clade size (MC) [65]. We also used this BaTS-based approach to analyze the association of the country of residence and gender of patients with lineages in the tree.

### **Bioinformatics analyses of the putative ST and MT antigens**

Sequence logos were compiled using the WebLogo server [68, 69]. Multiple sequence alignment of ST antigen was generated with the CLC Main Workbench (CLC Bio v6.6; Aarhus, Denmark) using the programs default options. Secondary structure, transmembrane helix, and disorder prediction of MT protein sequences were generated using the Disorder Prediction MetaServer [70], which reports consensus results of the following eight predictors: DIS-EMBL [71], DISOPRED [72], DISpro [73], FoldIndex [74], GlobPlot2 [75], IUPred [76], RONN

[77], and VSL2 [78]. The prediction of disorder was considered significant if at least four predictors gave a hit. Amino acid composition and other sequence features of MT and LT proteins were analyzed and plotted using custom R scripts [58, 79].

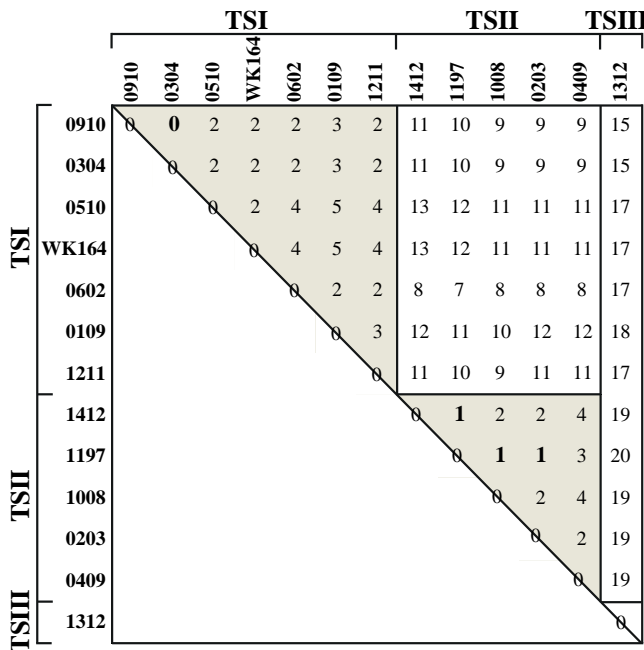
### Accession numbers

Accession numbers of virus genomes utilized in the study are shown in **Table 1**.

## Results

### TSPyV diverged by substitution across the entire genome and deletion in NCCR

To gain insight into the role of different genome regions in polyomavirus adaptation, we sequenced and analyzed genomes of a representative set of TSPyV. The viral genomic DNA sequences were obtained from thirteen lesional samples, which were collected over 16 years from largely previously described TS patients residing in the Northern America, the Netherlands, and Australia (**Table 1**). Sample DNA was subjected to 18 different TSPyV-specific PCR reactions which created amplicons that covered the entire TSPyV genome. The target sequences were amplified and the products visualized on gel (**Supplementary Figure S1**). Subsequent cloning, sequencing and assembling of the amplicons of every isolate resulted



**Figure 2.** Cumulative number of single nucleotide substitutions in TSPyV pairs. For brevity, the “TSPyV” label was omitted from the virus names (see **Table1**). Deletions were excluded upon the distance calculation. Due to diagonal symmetry, distances in the lower triangle of matrix are not shown. Intra-lineage values for the three TSPyV lineages are highlighted by grey background. Pairs with none or one substitution are in bold.

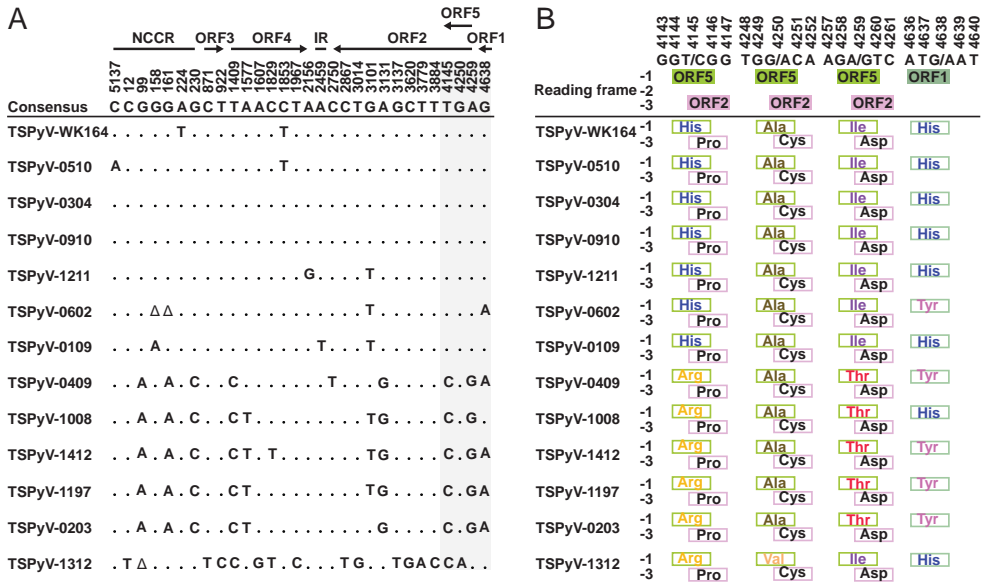
in eleven complete circular contigs, each representing a TSPyV genome. These genomes, together with the previously reported TSPyV-WK164 [17] and TSPyV-0109 [25] formed dataset 1 (DS1, see **M&M**). Analysis of these genomes revealed 30 single (di)nucleotide polymorphism (SNP) sites across the entire 5232 nts genome ( $\leq 0.6\%$  variation). Furthermore, genomes of TSPyV-0602 and TSPyV-1312 are 39 and 54 nts, respectively, shorter than genomes of the other viruses; they are thereafter indicated as deletions. The total number of substitutions separating TSPyVs was small, in the range from zero to twenty in different pairs (**Figure 2**).

All SNPs and deletions were mapped on the TSPyV circular genome whose open reading frame (ORF) organization has been updated to include five ORFs (ORF1-ORF5) (**Figure 1**, internal and intermediate circles), see also **Chapter 5** for additional detail [8]. Compared to the genome organization established in our prior analysis [25], it includes a previously non-recognized, fifth large ORF with a length of 450nts in the early genome region immediately downstream of and in-frame (-1 RF) with the ORF1 (**Figure 1**, inner circle). This ORF5 overlaps  $\sim 90\%$  with the 5'-terminal part of the LT-specific ORF2 located in the -3 RF (**Figure 1**). The sixth largest ORF (next after ORF5) consists of  $\sim 300$ nts.

Like with other human polyomaviruses (e.g., JCPyV and BKPyV) [19, 20]), a disproportionately large fraction of the SNPs (7 of 30;  $\sim 23\%$ ) was located in the NCCR that comprises only  $\sim 12\%$  of the genome (603 of 5232 nts). Three of these SNPs are also part of the non-overlapping deletions found in TSPyV-0602 (two SNPs) and TSPyV-1312 (one SNP). The identified SNPs and deletions are located outside of counterparts of the known essential functional signals, such as the pentanucleotide enhancer elements GAGGC and reversed GCCTC (**Supplementary Figure S3**). The other non-coding region, the intergenic region (IR) of  $\sim 100$ nts around genomic position 2500 separating the 3'-termini of the early and late transcripts, contained one SNP (**Figures 1** and **3A**). Nine SNPs were identified in the late genomic region of 2,032 bp: seven in ORF4 and two in ORF3, respectively. Further twelve SNPs were found in the early genomic region of 2,507 bp: one in ORF1, three in the overlap between ORF2 and ORF5, and eight in the unique part of ORF2. All these variations are expected to be expressed from one or more transcripts. All twenty SNPs in LT ORF2 and in the VP ORFs represented synonymous substitutions. The three SNPs in the ORF2/ORF5 overlap were synonymous for the LT ORF2 but non-synonymous for the MT ORF5. The one SNP in the ST/MT-specific part of ORF1 was also non-synonymous (**Figures 1** and **3B**).

## The timeframe and rate of evolution of TSPyV are estimated with high uncertainty

Using a Bayesian approach, we found that TSPyV clustered into three monophyletic lineages (TSI, TSII, and TSIII) with high probability support values of 0.9 or higher, with TSPyV-1312 being the sole representative of TSIII (**Figure 4A**). No significant association between location (country) and phylogenetic lineage was observed (**Table 2**). We also noted considerable overlap between the timeframes of sample collection for viruses that comprise the best

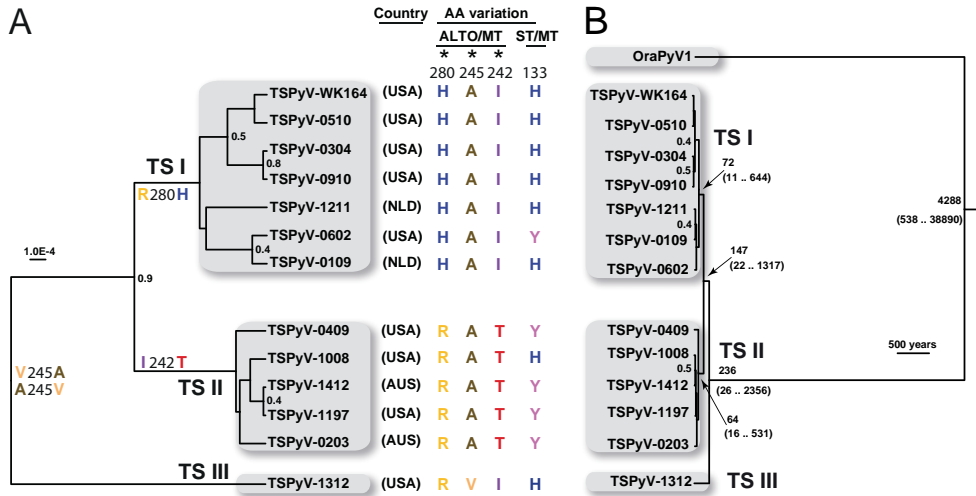


**Figure 3.** Nucleotide and amino acid variation in TSPyV. (A) Alignment of SNPs and deletions in TSPyV genomes. Top, SNP genome location, its respective region and consensus nucleotide. Bottom, nucleotide variation in TSPyV isolates: dots, nucleotides identical to the consensus; triangles, multi nucleotide deletions of 54nts (120-173; TSPyV-0602) and 39nts (61-99; TSPyV-1312). Gray shaded part, SNPs resulting in aa variations. (B) Amino acid variations in TSPyV. Top, SNPs gray-shaded in panel A are shown with flanking dinucleotides along with the affected ORF-specific codons. Bottom, the SNP-containing codons are translated according to respective ORF and colored according to **Figure 1**.

sampled TSI and TSII lineages, from 2002 to 2011 vs. from 1997 to 2012, respectively, indicative of co-circulation of the two lineages.

To estimate the time of divergence of the most recent common ancestor (tMRCA) of TSPyV, the 13 TSPyV variants were analyzed together with their closest relative, OraPyV1, which was isolated in 1996 [80], in dataset 2 (DS2, see **M&M**). This analysis resulted in a rooted tree with three distinct TSPyV lineages (TSI, TSII, and TSIII) (**Figure 4B**), similar to the topology of the DS1-based tree (**Figure 4A**). The tMRCA of TSPyV and OraPyV1 was estimated to be in the range of 538-38890 years (95% confidence interval, 95% CI) with a median of about 4288 years before present (ybp) with respect to the latest TSPyV genome sampled in 2012. Divergence of TSI, TSII, and TSIII from their common ancestor was estimated to have occurred in the 95% CI period of 26-2356 ybp (median: 236 ybp). Subsequently, the TSI and TSII lineages separated from each other in the 95% CI period of 22-1317 ybp (median: 147 ybp), while members of the TSI and TSII lineage shared a MRCA in the 95% CI periods of 11-644 and 16-531 ybp, respectively, (median: 72 and 64 ybp, respectively). Using the times of isolation of the 13 TSPyVs, we estimated the evolutionary rate of TSPyV to be in the 95% CI range of  $1.2 \times 10^{-7}$  -  $2.0 \times 10^{-5}$  subs/site/year (median:  $9.2 \times 10^{-6}$  subs/site/year).

Our estimates of the tMRCA and evolutionary rate are associated with ~2 orders of magnitude confidence intervals. One possible source of this large uncertainty is that the



**Figure 4.** Phylogeny and evolution of TSPyV. (A) TSPyV phylogeny. Depicted is the pseudo-rooted tree based on the complete TSPyV genome sequences, with three groups highlighted. The country of patient residence and four variable aa residues in MT and ST/MT are shown next to the virus names (colored according to **Figure 3B**). Traits that are significantly associated ( $\alpha \leq 0.05$ ; **Table 2**) with lineages in the tree are indicated by asterisks (top). The scale bar, the average number of substitutions per site. Probability support values for internal nodes smaller than 1 are shown. (B) TSPyV evolution. Presented is the rooted tree; tree branches represent time (years), which were estimated using the sampling dates of the TS materials (**Table 1**) and that of OraPyV1 (1996). Numbers to the left of (some) branching events represent estimated median time to the most recent common ancestor (tMRCA) of the respective ancestors, with confidence intervals (95% highest posterior density) depicted in parentheses. The scale bar indicates time. For other details see (A) and text.

dates of sample collection at time of symptomatic viral (re)activation may deviate considerably from the dates when the patient was primarily infected with the polyomavirus. Using BEAST, we estimated the dates of the patient infection for each TSPyV isolate (see **M&M** for details; **Supplementary Figure S4**). In most cases, the obtained estimates remained associated with large confidence intervals that often spanned the full range of possible dates. On the other hand, median numbers deviated considerably from the collection dates and for some isolates (see for example TSPyV-1211 and 1312; **Supplementary Figure S4**) this interval did not include the date of sample collection. Collectively, these data support the notion that the date of sample collection deviates from the date of primary infection.

### Non-synonymous SNPs in MT ORF5 are strongly associated with three phylogenetic clusters of TSPyV and co-evolve with SNPs in other regions

Next, we asked whether amino acid substitutions at the four positions in the TSPyV variants (**Figure 3B**) are the result of adaptive evolution of the virus to humans. Positive selection was favored by analyses using random effect likelihood method (REL,  $p=0.99$ ) and fast unconstrained Bayesian approximation method (FUBAR,  $p=0.95$ ) [60 - 62] for the G4638A variation in ORF1 that results in the H133Y substitution in the common ST/MT domain. How-

**Table 2.** Support values for phylogenetic associations of amino acid variations observed among the 13 TSPyV isolates.

Variation <sup>1</sup>	Genomic position <sup>2</sup>	Affected transcripts	Support by criterion <sup>3</sup>		
			AI	PS	MC
USA-NLD-AUS	/	/	0.143	1	0.121
H-Y	4638	ST, MT	0.124	0.072	0.051
I-T	4259	MT	<b>0</b>	<b>0.001</b>	<b>0.002</b>
A-V	4250	MT	<b>0</b>	1	<b>0.001</b>
H-R	4145	MT	<b>0</b>	<b>0</b>	<b>0.001</b>
Negative control	/	/	0.503	0.642	0.383
Positive control	/	/	<b>0</b>	<b>0</b>	<b>0.001</b>

<sup>1</sup> Country of origin: the home country of the patient; observed amino acid variations: The respective amino acids are shown, separated by a dash; negative control: To each tip of the phylogeny one of two arbitrary amino acids was assigned randomly. The average p-value among 10 iterations is shown; positive control: Each of the three TSPyV lineages was assigned a different amino acid, and each member of the same lineage was assigned the same amino acid.

<sup>2</sup> Coordinate of the SNP causing the amino acid variation in the genome of TSPyV-0109.

<sup>3</sup> P-values of the null hypothesis (random association of the amino acid variation with the phylogeny) are shown for the following test statistics: association index (AI), parsimony score (PS), maximum monophyletic clade size (MC). For MC, the minimum value over the MC values of the single amino acid states is shown. Significant values under the confidence level of  $\alpha=0.05$  are in bold.

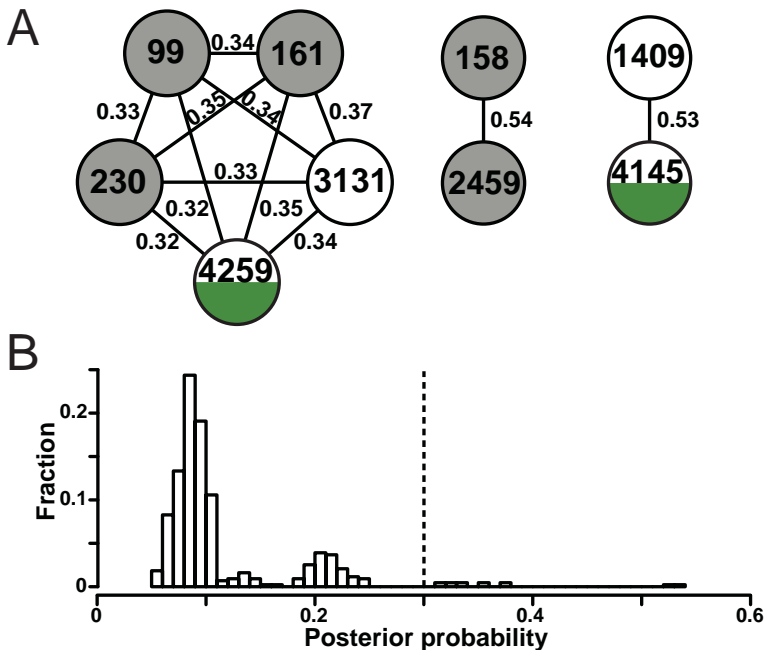
ever, the genetic diversity and the size of our dataset may be too small to assess selection accurately at *all* sites using these methods [63]. Therefore, we used three statistical tests (see **M&M**) to estimate the association between a trait (the amino acid variations in this case) and the phylogeny. We reasoned that statistically significant association would be indicative for the analyzed sites being constrained in a lineage-specific fashion (selected) [81]. Before conducting these analyses, the three TSPyV lineages were reproduced with a modified DS1 that did *not* include the SNPs with non-synonymous substitutions (data not shown). This observation implied that the observed lineages are independent from the MT codons encoding the I242T, A245V and H280R variation. Consequently, the tree (**Figure 4A**) could be used in the association test.

All three amino acid variations in the ORF5 (but not that of ORF1) were found to be significantly associated ( $p \leq 0.002$ ) with the three major phylogenetic lineages of TSPyV (**Figure 4A**): two variations by the three test statistics and one variation by two statistics (**Table 2**). Given this strong association, we then asked whether the non-synonymous ORF5 SNPs coevolve with other SNPs. To this end, we tested all pairs of the 30 SNPs (435 pairs in total) for coevolution using a Bayesian evolutionary framework as implemented in the Spider-monkey tool [60, 64]. Strikingly, T4145C (H280R) and A4259G (I242T) were among two out of the three networks (**Figure 5A**) that were supported with the top 2.7% of the posterior probabilities values for network edges (**Figure 5B**). For T4145C, the coevolving synonymous SNP T1409C is located in the VP1 ORF4. For T4259G, three coevolving SNPs (G99A, G161A, and G230C) are located in the NCCR and one (synonymous A3131G) in the LT ORF2. Accord-

ing to parsimonious reasoning, amino acid changes at the three positions in the ORF5 (and the coevolving nucleotides outside ORF5) were most likely fixed in the MRCA of TSI (R280H), TSII (I242T), and either TSI/TSII (A245V) or TSIII (V245A). It is noteworthy that the ancestral and the newly acquired amino acids (nucleotides) persisted in *all* characterized variants of the respective two lineages, TSI and TSII, for which more than one sequence was available (**Figures 3** and **4A**). The third network involves the NCCR G158A and the IR A2459T (**Figure 5A**) that is observed only in TSPyV-0109 (**Figures 3A** and **4A**) suggestive of a recent fixation. These results reveal highly constrained evolution of the TSPyV variants indicative of strong selection.

### Genetic constraints on the non-synonymous substitutions in ST and MT antigens of TSPyV

Next we analyzed the nature of these constraints. The ORF1-encoded H133Y variation is observed at the most variable position of the highly conserved region in polyomaviruses, in which it is located between the two CXCXC Zinc-finger motifs and adjacent to the invariant His residue (**Supplementary Figure S5**). Thus, the 133 position is constrained differently and less compared to other neighboring positions.



**Figure 5.** The genome-wide networks of coevolving nucleotides. (A) The 12 top scoring pairs (top 2.7%; see panel B) form three networks involving different types of SNPs (nodes) defined by color in **Figure 1** and numbered according to **Figure 3A**. Each edge is provided with its support value. (B) Distribution of the marginal posterior probability of network edges. The histogram depicts the normalized distribution of probability values for 435 pairs formed by the 30 SNPs using the 0.01 bin size. A vertical line is the arbitrary cutoff value of 0.97.

The ORF5-encoded H280R variation is located in the most divergent region of TSPyV and its two most closely related viruses, OraPyV1 and AtPPyV1 (**Supplementary Figure S6**). In contrast, the I242T variation is adjacent to the A245V variation within the ORF5m2, one of the four most conserved motifs of ORF5 (from ORF5m1 to ORF5m4) (**Chapter 5**) [8]. The evolution of this motif is highly constrained by the LT LXCXE motif in the overlapping ORF2. Particularly, the A245V variation represents the full spectrum of residue variation possible at this position, which was identified under positive selection in the inter-species evolution of ORF5-encoding polyomaviruses and was named the COCO-VA toggling [8] (see also “Discussion”).

### Structural context of the non-synonymous substitutions in ST and MT antigen of TSPyV

Tertiary structure of the H133Y variants of the TSPyV ST/MT domain was modeled using the solved structure of this domain for SV40, but this did not reveal any notable differences that could be attributed to the Tyr/His variation (data not shown).

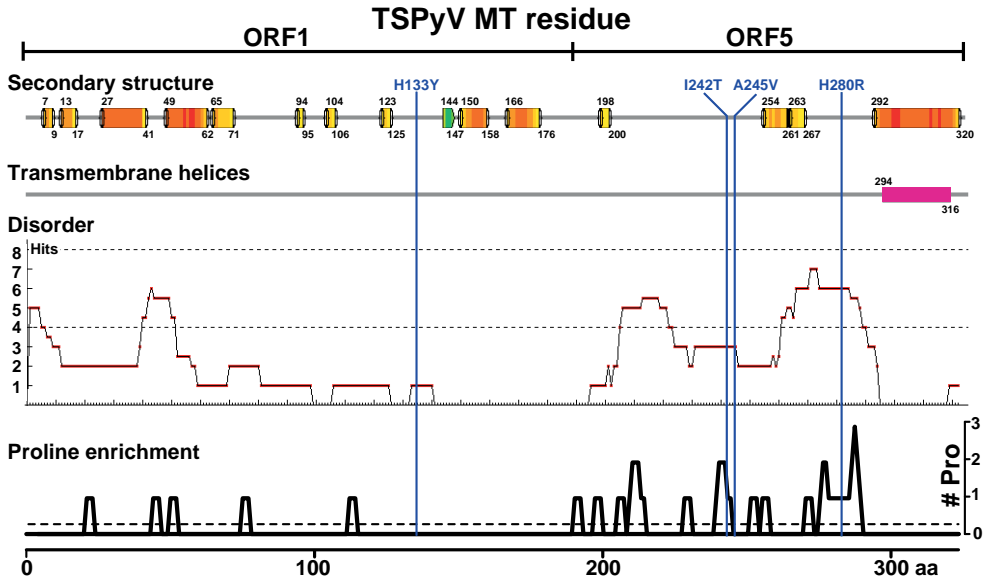
The amino-acid variations I242T, A245V and H280R were found in the Proline-rich and intrinsically disordered region of the MT antigen (**Figure 6**). Also, this domain contains a large number of Serine, Threonine and Tyrosine residues partially conserved in TSPyV, OraPyV1 and AtPPyV1 (**Supplementary Figure S6**), which could be targets for regulatory phosphorylation as was demonstrated for MPyV [12]. Because of this characteristic sequence context and a relatively high local sequence conservation of ORF5m2, the latter may be regarded as a short linear motif (SLiM) [82]. Consequently, the I242T and A245V substitutions could affect MT antigen through the ORF5m2 SLiM (see “Discussion”).

## Discussion

In this Chapter, we provide genetic evidence for strong and multi-directional selection pressure acting at the MT-specific ORF5 in TSPyV. Our results implicate MT antigen in the adaptation of TSPyV to humans by a new mechanism that includes COCO-VA toggling in a SLiM of a disordered protein region [8].

TSPyV is one of only four polyomaviruses with either established or highly probable link to pathology in humans. In total, we have analyzed TSPyV genome sequences from 12 TS patients that represent ~40% of the reported TS cases worldwide, making this dataset highly inclusive for individuals suffering from TS, despite it being small compared to those of other characterized polyomaviruses [23, 83, 84]. We also believe that the obtained results are relevant beyond this patient cohort. Indeed, another sequence (TSPyV-WK164) was reported from a TS-*asymptomatic* immunocompromised patient [17] carrying only one unique A224T substitution in the NCCR region, and otherwise differing just by two substitutions from TSPyV-0510 (**Figures 2 and 3**). Furthermore, unless the immunocompromised (TS) patients are epidemiologically isolated from the rest of human population known to





**Figure 6.** Structural context of non-synonymous mutations in putative MT of TSPyV. The position of four non-synonymous mutations in ORF1 and ORF5 are shown against plots for the MT of TSPyV-0109 (from top to bottom and drawn to scale) predicted secondary structure score ranges: from yellow (low) to red (high) for helix, and from green (low) to blue (high) for beta-sheet; transmembrane helices (magenta); intrinsic structural disorder, scores for disordered regions were calculated as a consensus over eight predictors; and Pro residue density distribution. The top three assignments were assisted with the Disorder Prediction MetaServer. The Pro distribution was calculated using a three-amino-acid window along the MT sequence of TSPyV. Dashed line, the average number of Pro residues according to ST, LT, MT, VP1, and VP2 sequences. For more details see **M&M** and **Chapter 5**.

be up to 70% TSPyV-seropositive [33, 34], immunocompetent individuals must have been involved in virus transmission between the surveyed patients. In this respect we note that only viruses forming a monophyletic cluster could be implicated in the direct transmission between two patients from whom they were isolated. This transmission is limited by the time-window opportunity, which is defined by the smallest difference between the time of the sample collection from one patient and the date of birth of the other patient. Strikingly, this window was just one year for the two patients who resided in different USA states and carried monophyletic TSPyV-0910 and TSPyV-0304, respectively (**Supplementary Figure S4**). If other individuals were involved in the transmission, this strict window limit would no longer apply. On the other hand, the more people involved in the transmission the more likely that the virus must have evolved slowly to fit the observed small genetic variation, which is *none* (zero) for the above two TSPyV isolates.

Although our samples were collected over 16 years, the actual time-frame of substitution accumulation due to polyomavirus biology may be as large as the time difference between the years of the most recent sample collection (2012) and the birth of the oldest patient (1940), which is 72 years (**Supplementary Figure S4**). The large difference between

the above two numbers (16 and 72) is likely to be a major source of uncertainty for our estimates of the high evolutionary rate and the recent tMRCA of TSPyV (~200 years ago) by a relaxed molecular-clock framework, which used the dates of sample collection. Without the assumption of virus-host codivergence, the dominant evolutionary model until recently [83, 85, 86], similar estimates for recent emergence were reported for other human polyomaviruses, including JCPyV [23] and BKPyV [18, 84] of the *Orthopolyomavirus II* genogroup [16].

Our estimates of mean and upper limit values of the evolutionary rate are relatively close to those published for JCPyV in a similarly executed study [23]:  $9.2 \times 10^{-6}$  and  $2.0 \times 10^{-5}$  subs/site/year vs.  $1.7 \times 10^{-5}$  and  $3.1 \times 10^{-5}$  subs/site/year, respectively. A similar median estimate was also reported for BKPyV [84]. However, a difference between these viruses for the lower limit of the 95% CI for evolutionary rate was considerable:  $1.2 \times 10^{-7}$  vs  $2.1 \times 10^{-6}$  subs/site/year for TSPyV and JCPyV, respectively; in fact, the lower CI limit estimate approximates the mean evolutionary rate estimates for JCPyV and BKPyV calculated under the assumption of virus-host codivergence. The extremely large uncertainty about the obtained median values in our study could be due to the small size of the TSPyV dataset. On the other hand, the especially large difference between the low-end CI values of the evolutionary rate for TSPyV and JCPyV may be explained by the low rate of evolution of TSPyV. Indeed, the scale of genome variation in TSPyV isolated from 13 individuals in our study was an order of magnitude smaller than those obtained for other polyomaviruses isolated from human population, while being remarkably close to that in BKPyV isolated from a single adult individual [17, 18]. Furthermore, genomes of the two TSPyV pairs, 0910-0304 and 1197-1412, are identical or differ by just a single nucleotide substitution, respectively, while being collected from infected patients six and fifteen years apart. Similarly, striking examples of extraordinary pairwise similarity could be identified for other human polyomaviruses, including BKPyV [87], MCPyV [41], WU polyomavirus [88], and KI polyomaviruses [89], which were isolated from individuals residing in different countries and/or at different time points. These observations could be explained by very low mutation rate, prolonged virus latency, strong purifying selection, or a combination of either of these factors [84, 90].

Reliable estimates of the background mutation rate for polyomaviruses are lacking, although the host DNA polymerase mediating polyomavirus genome replication is of high fidelity. Also unknown is the replication rate (generation time) of polyomaviruses during natural infection that may be primarily determined by the duration of virus latency in the infected individual. However, virus latency alone may not explain the observed extraordinary similarity between some TSPyV variants. Based on the strong co-evolution of many SNPs, including two out of four non-synonymous SNPs, we implicate purifying selection in the evolution of TSPyV.

We observed very limited amino acid variations in TSPyV. The ones identified affected only ORF1 and ORF5, which are known to jointly encode MT antigen in MPyV and HaPyV [9 - 12]. To identify ORF5 in TSPyV, unknown at the start of this study and dissertation, and explore its implications we developed a new format of depicting genome-related

information of polyomaviruses in a plot that separated genome organization from genome expression (**Figure 1**). Adopting this format broadly could facilitate further exploration of polyomaviruses, since even very small ORFs (below the ORF5-size threshold) could be expressed jointly with other ORF(s) using alternative splicing. Our preliminary experiments reveal mRNA and protein expression from ORF5 in primary and immortalized human keratinocytes transfected or transduced with TSPyV early region, as well as ORF5-encoding mRNA expression in a TS patient sample (Van der Meijden *et al.*, unpublished results).

ORF5 is the predominant place of non-synonymous substitutions during evolution of TSPyV as well as of most profound genomic changes that were accumulated during evolution of polyomaviruses. Despite this exceptional fast pace, the evolution of ORF5 is evidently coordinated with that of other regions. We identified sets of coevolving sites representing different regions in TSPyV, as well as the association of ORF5 with the phylogenetically compact majority of Ortho-I polyomavirus cluster that was delineated using other genome regions [8]; the latter result was also reported by Carter and coworkers [15]. This agreement at the levels of micro- and macro-evolution reveals ORF5 as a major locus mediating host adaptation and driving speciation in a large subset of polyomaviruses.

According to our phylogenetic analysis, the H133Y variation in ST and MT of TSPyV resulted from recurrent fourth/back mutation in the ORF1-encoded position producing a typical signature of positive selection. In contrast, the observed amino acid diversity in ORF5 of the TSPyV isolates could have been generated with just a single substitution at the MT position 242, 245, and 280 in three respective ancestors. The evolutionary origin of this pattern is opened to two interpretations. These replacements could have been fixed as a result of three bottle-neck events during neutral evolution. Alternatively, they could have been driven by episodic positive selection. The latter scenario is supported by considerations listed below. First, these substitutions were accepted in the most prominent intermediate ancestors leading to the three major contemporary lineages of TSPyV. Second, they are clustered within a small region of 114nts (~2.5% of the genome) rather than scattered over the genome. Third, they are observed in a single ORF, ORF5, while three major ORFs, ORF2-ORF4, remain unaffected. Fourth, under the model of neutral evolution and with the constraints by ORF5 codons, the frequency of synonymous substitution in the ORF2 region overlapping with ORF5 is expected to be lower than that in the downstream part of ORF2, according to prior research with overlapping genes [91, 92]. However these frequencies in the two regions are rather similar in TSPyV: three substitutions per 450nts of the ORF2/ORF5 overlap vs. nine substitutions per 1450nts of the ORF2-only. Fifth, one of the three ORF5 aa variations, A245V, occurred in one of the most constrained position of the ORF2/ORF5 overlap that encodes the conserved Cys of the LT pRB-interacting site and subject to positive selection in inter-species evolution (COCO-VA toggling) [8]. Thus, our data strongly suggest that all four non-synonymous replacements observed during the TSPyV evolution were driven by (episodic) positive selection that promoted adaptation of TSPyV to the human population.

TSI and TSII viruses, the two most populous lineages of TSPyV according to its phylogeny, use the Ala allele at MT245. Thus, the acquisition of the Ala allele may have provided a fitness gain that facilitated further directional evolution elsewhere in the genome observed in these lineages. Because of the availability of viruses with each allele, Ala and Val, at MT245 and the overall low sequence diversity in the TSPyV isolates, TSPyV offers a unique model to study the molecular basis of the COCO-VA toggling for host adaptation of polyomaviruses in general.

## Acknowledgements

The authors thank Ernst Verschoor for providing the dates of OraPyV isolations, Hans van Leeuwen for the useful discussions, and Igor Sidorov and Dmitry Samborskiy for help with Viralis. AEG is member of the Netherlands Bioinformatics Center (NBIC) Faculty.

## References

1. Johne R, Buck CB, Allander T, Atwood WJ, Garcea RL, Imperiale MJ, Major EO, Ramqvist T, Norkin LC. Taxonomical developments in the family Polyomaviridae (2011) *Arch. Virol.* 156: 1627-1634.
2. Feltkamp MC, Kazem S, van der Meijden E, Lauber C, Gorbalenya AE. From Stockholm to Malawi: recent developments in studying human polyomaviruses (2012) *J. Gen. Virol.* 94: 482-496.
3. Decaprio JA, Garcea RL. A cornucopia of human polyomaviruses (2013) *Nat Rev Microbiol.* 11: 264-276.
4. Imperiale MJ, Major EO. Polyomaviruses. In: *Fields VIROLOGY* (2007). Knipe DM, Howley PM (editors). Philadelphia: Wolters Kluwer / Lippincott Williams & Wilkins: 2263-2299.
5. White MK, Safak M, Khalili K. Regulation of Gene Expression in Primate Polyomaviruses (2009) *J. Virol.* 83: 10846-10856.
6. Jiang M, Abend JR, Johnson SF, Imperiale MJ. The role of polyomaviruses in human disease (2009) *Virology* 384: 266-273.
7. Shishido-Hara Y, Hara Y, Larson T, Yasui K, Nagashima K, Stoner GL. Analysis of capsid formation of human polyomavirus JC (Tokyo-1 strain) by a eukaryotic expression system: splicing of late RNAs, translation and nuclear transport of major capsid protein VP1, and capsid assembly (2000) *J. Virol.* 74: 1840-1853.
8. Lauber C, Kazem S, Kravchenko AA, Feltkamp MC, Gorbalenya AE. Interspecific adaptation by binary choice at de novo polyomavirus T antigen site through accelerated codon-constrained Val-Ala toggling within an intrinsically disordered region (2015) *Nucleic Acids Res.* doi: 10.1093/nar/gkv378
9. Hutchinson MA, Hunter T, Eckhart W. Characterization of T antigens in polyoma-infected and transformed cells (1978) *Cell* 15: 65-77.
10. Courtneidge SA, Goutebroze L, Cartwright A, Heber A, Scherneck S, Feunteun J. Identification and characterization of the hamster polyomavirus middle T antigen (1991) *J. Virol.* 65: 3301-3308.
11. Cheng J, DeCaprio JA, Fluck MM, Schaffhausen BS. Cellular transformation by Simian Virus 40 and Murine Polyoma Virus T antigens (2009) *Semin. Cancer Biol.* 19: 218-228.
12. Fluck MM, Schaffhausen BS. Lessons in signaling and tumorigenesis from polyomavirus middle T antigen (2009) *Microbiol. Mol. Biol. Rev.* 73: 542-563.
13. Magnusson G, Nilsson MG, Dilworth SM, Smolar N. Characterization of polyoma mutants with altered middle and large T-antigens (1981) *J. Virol.* 39: 673-683.
14. Yi X, Freund R. Deletion of proline-rich domain in polyomavirus T antigens results in virus partially defective in transformation and tumorigenesis (1998) *Virology* 248: 420-431.

15. Carter JJ, Daugherty MD, Qi X, Bheda-Malge A, Wipf GC, Robinson K, Roman A, Malik HS, Galloway DA. Identification of an overprinting gene in Merkel cell polyomavirus provides evolutionary insight into the birth of viral genes (2013) *Proc. Natl. Acad. Sci. U. S. A.* 110: 12744-12749.
16. Feltkamp MC, Kazem S, van der Meijden E, Lauber C, Gorbalenya AE. From Stockholm to Malawi: recent developments in studying human polyomaviruses (2013) *J. Gen. Virol.* 94: 482-496.
17. Siebrasse EA, Bauer I, Holtz LR, Le B, Lassa-Claxton S, Canter C, Hmiel P, Shenoy S, Sweet S, Turmelle Y, Shepherd R, Wang D. Human polyomaviruses in children undergoing transplantation, United States, 2008–2010 (2012) *Emerg. Infect. Dis.* 18: 1676-1679.
18. Chen Y, Sharp PM, Fowkes M, Kocher O, Joseph JT, Koralnik IJ. Analysis of 15 novel full-length BK virus sequences from three individuals: evidence of a high intra-strain genetic diversity (2004) *J. Gen. Virol.* 85: 2651-2663.
19. Yogo Y, Zhong S, Shibuya A, Kitamura T, Homma Y. Transcriptional control region rearrangements associated with the evolution of JC polyomavirus (2008) *Virology* 380: 118-123.
20. Gosert R, Rinaldo CH, Funk GA, Egli A, Ramos E, Drachenberg CB, Hirsch HH. Polyomavirus BK with rearranged noncoding control region emerge in vivo in renal transplant patients and increase viral replication and cytopathology (2008) *J. Exp. Med.* 205: 841-852.
21. Shadan FF, Villarreal LP. The evolution of small DNA viruses of eukaryotes: past and present considerations (1995) *Virus Genes* 11: 239-257.
22. Perez-Losada M, Christensen RG, McClellan DA, Adams BJ, Viscidi RP, Demma JC, Crandall KA. Comparing Phylogenetic Codivergence between Polyomaviruses and Their Hosts (2006) *J. Virol.* 80: 5663-5669.
23. Shackelton LA, Rambaut A, Pybus OG, Holmes EC. JC virus evolution and its association with human populations (2006) *J. Virol.* 80: 9928-9933.
24. Haycox CL, Kim S, Fleckman P, Smith LT, Piepkorn M, Sundberg JP, Howell DN, Miller SE. Trichodysplasia spinulosa—a newly described folliculocentric viral infection in an immunocompromised host (1999) *J. Investig. Dermatol. Symp. Proc.* 4: 268-271.
25. van der Meijden E, Janssens RW, Lauber C, Bouwes Bavinck JN, Gorbalenya AE, Feltkamp MC. Discovery of a new human polyomavirus associated with trichodysplasia spinulosa in an immunocompromized patient (2010) *PLoS Pathog.* 6: e1001024.
26. Matthews MR, Wang RC, Reddick RL, Saldivar VA, Browning JC. Viral-associated trichodysplasia spinulosa: a case with electron microscopic and molecular detection of the trichodysplasia spinulosa-associated human polyomavirus (2011) *J. Cutan. Pathol.* 38: 420-431.

27. Blake BP, Marathe KS, Mohr MR, Jones N, Novosel TA. Viral-Associated Trichodysplasia of Immunosuppression in a Renal Transplant Patient (2011) *J. Drugs Dermatol.* 10: 422-424.
28. Burns A, Arnason T, Fraser R, Murray S, Walsh N. Keratotic “spiny” papules in an immunosuppressed child. Trichodysplasia spinulosa (TS) (2011) *Arch. Dermatol.* 147: 1215-1220.
29. Schwieger-Briel A, Balma-Mena A, Ngan B, Dipchand A, Pope E. Trichodysplasia spinulosa--a rare complication in immunosuppressed patients (2010) *Pediatr. Dermatol.* 27: 509-513.
30. Kazem S, van der Meijden E, Kooijman S, Rosenberg AS, Hughey LC, Browning JC, Sadler G, Busam K, Pope E, Benoit T, Fleckman P, de VE, Eekhof JA, Feltkamp MC. Trichodysplasia spinulosa is characterized by active polyomavirus infection (2011) *J. Clin. Virol.* 5: 225-230.
31. Kazem S, van der Meijden E, Feltkamp MC. The trichodysplasia spinulosa-associated polyomavirus: virological background and clinical implications (2013) *APMIS* 121: 770-782.
32. Kazem S, van der Meijden E, Wang RC, Rosenberg AS, Pope E, Benoit T, Fleckman P, Feltkamp MC. Polyomavirus-associated trichodysplasia spinulosa involves hyperproliferation, pRB phosphorylation and upregulation of p16 and p21 (2014) *PLoS One* 9: e108947.
33. van der Meijden E, Kazem S, Burgers MM, Janssens R, Bouwes Bavinck JN, de Melker H, Feltkamp MC. Seroprevalence of Trichodysplasia Spinulosa-associated Polyomavirus (2011) *Emerg. Infect. Dis.* 17: 1355-1363.
34. Chen T, Mattila PS, Jartti T, Ruuskanen O, Soderlund-Venermo M, Hedman K. Seroepidemiology of the Newly Found Trichodysplasia Spinulosa-Associated Polyomavirus (2011) *J. Infect. Dis.* 204: 1523-1526.
35. van der Meijden E, Bialasiewicz S, Rockett RJ, Tozer SJ, Sloots TP, Feltkamp MC. Different serologic behavior of MCPyV, TSPyV, HPyV6, HPyV7 and HPyV9 polyomaviruses found on the skin (2013) *PLoS ONE* 8: e81078.
36. Scuda N, Hofmann J, Calvignac-Spencer S, Ruprecht K, Liman P, Kuhn J, Hengel H, Ehlers B. A novel human polyomavirus closely related to the african green monkey-derived lymphotropic polyomavirus (2011) *J. Virol.* 85: 4586-4590.
37. Imperiale MJ. The human polyomaviruses, BKV and JCV: molecular pathogenesis of acute disease and potential role in cancer (2000) *Virology* 267: 1-7.
38. Allander T, Andreasson K, Gupta S, Bjerkner A, Bogdanovic G, Persson MA, Dalianis T, Ramqvist T, Andersson B. Identification of a third human polyomavirus (2007) *J. Virol.* 81: 4130-4136.
39. Gaynor AM, Nissen MD, Whiley DM, Mackay IM, Lambert SB, Wu G, Brennan DC, Storch GA, Sloots TP, Wang D. Identification of a novel polyomavirus from patients with acute respiratory tract infections (2007) *PLoS Pathog.* 3: e64.

40. Feng H, Shuda M, Chang Y, Moore PS. Clonal integration of a polyomavirus in human Merkel cell carcinoma (2008) *Science* 319: 1096-1100.
41. Schowalter RM, Pastrana DV, Pumphrey KA, Moyer AL, Buck CB. Merkel cell polyomavirus and two previously unknown polyomaviruses are chronically shed from human skin (2010) *Cell Host Microbe* 7: 509-515.
42. Sauvage V, Foulongne V, Cheval J, Ar Gouilh M, Pariente K, Dereure O, Manuguerra JC, Richardson J, Lecuit M, Burguiere A, Caro V, Eloit M. Human polyomavirus related to african green monkey lymphotropic polyomavirus (2011) *Emerg. Infect. Dis.* 17: 1364-1370.
43. Siebrasse EA, Reyes A, Lim ES, Zhao G, Mkakosya RS, Manary MJ, Gordon JI, Wang D. Identification of MW Polyomavirus, a Novel Polyomavirus in Human Stool (2012) *J. Virol.* 86: 10321-10326.
44. Buck CB, Phan GQ, Raiji MT, Murphy PM, McDermott DH, McBride AA. Complete genome sequence of a tenth human polyomavirus (2012) *J. Virol.* 86: 10887.
45. Yu G, Greninger AL, Isa P, Phan TG, Martinez MA, de la Luz SM, Contreras JF, Santos-Preciado JI, Parsonnet J, Miller S, Derisi JL, Delwart E, Arias CF, Chiu CY. Discovery of a novel polyomavirus in acute diarrheal samples from children (2012) *PLoS ONE* 7: e49449.
46. Lim ES, Reyes A, Antonio M, Saha D, Ikumapayi UN, Adeyemi M, Stine OC, Skelton R, Brennan DC, Mkakosya RS, Manary MJ, Gordon JI, Wang D. Discovery of STL polyomavirus, a polyomavirus of ancestral recombinant origin that encodes a unique T antigen by alternative splicing (2013) *Virology* 436: 295-303.
47. Korup S, Rietscher J, Calvignac-Spencer S, Trusch F, Hofmann J, Moens U, Sauer I, Voigt S, Schmuck R, Ehlers B. Identification of a novel human polyomavirus in organs of the gastrointestinal tract (2013) *PLoS ONE* 8: e58021.
48. Sadler GM, Halbert AR, Smith N, Rogers M. Trichodysplasia spinulosa associated with chemotherapy for acute lymphocytic leukaemia (2007) *Australas. J. Dermatol.* 48: 110-114.
49. Wyatt AJ, Sachs DL, Shia J, Delgado R, Busam KJ. Virus-associated trichodysplasia spinulosa (2005) *Am. J. Surg. Pathol.* 29: 241-246.
50. Holzer AM, Hughey LC. Trichodysplasia of immunosuppression treated with oral valganciclovir (2009) *J. Am. Acad. Dermatol.* 60: 169-172.
51. Benoit T, Bacelieri R, Morrell DS, Metcalf J. Viral-associated trichodysplasia of immunosuppression: report of a pediatric patient with response to oral valganciclovir (2010) *Arch. Dermatol.* 146: 871-874.
52. Desmet FO, Hamroun D, Lalande M, Collod-Beroud G, Claustres M, Beroud C. Human Splicing Finder: an online bioinformatics tool to predict splicing signals (2009) *Nucleic Acids Res.* 37: e67.



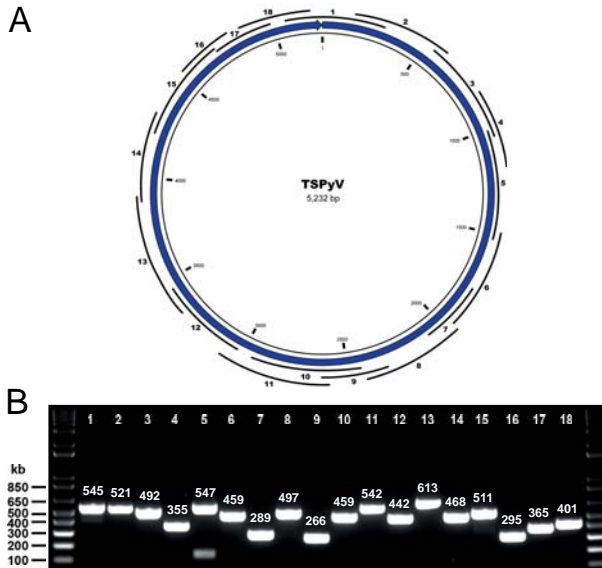
53. Gorbalenya AE, Lieutaud P, Harris MR, Coutard B, Canard B, Kleywegt GJ, Kravchenko AA, Samborskiy DV, Sidorov IA, Leontovich AM, Jones TA. Practical application of bioinformatics by the multidisciplinary VIZIER consortium (2010) *Antiviral Res.* 87: 95-110.
54. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput (2004) *Nucleic Acids Res.* 32: 1792-1797.
55. Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees (2007) *BMC Evol. Biol.* 7: 214.
56. Hasegawa M, Kishino H, Yano T. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA (1985) *J. Mol. Evol.* 22: 160-174.
57. Drummond AJ, Ho SY, Phillips MJ, Rambaut A. Relaxed phylogenetics and dating with confidence (2006) *PLoS Biol.* 4: e88.
58. Goodman SN. Toward evidence-based medical statistics. 2: The Bayes factor (1999) *Ann. Intern. Med.* 130: 1005-1013.
59. Rambaut A, Drummond AJ. Tracer v1.4. In: 2007.
60. Kosakovsky Pond SL, Frost SD. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments (2005) *Bioinformatics* 21: 2531-2533.
61. Murrell B, Moola S, Mabona A, Weighill T, Sheward D, Kosakovsky Pond SL, Scheffler K. FUBAR: a fast, unconstrained bayesian approximation for inferring selection (2013) *Mol. Biol. Evol.* 30: 1196-1205.
62. Kosakovsky Pond SL, Muse SV. Site-to-site variation of synonymous substitution rates (2005) *Mol. Biol. Evol.* 22: 2375-2385.
63. Kosakovsky Pond SL, Frost SD. Not so different after all: a comparison of methods for detecting amino acid sites under selection (2005) *Mol. Biol. Evol.* 22: 1208-1222.
64. Poon AF, Lewis FI, Pond SLK, Frost SD. An evolutionary-network model reveals stratified interactions in the V3 loop of the HIV-1 envelope (2007) *PLoS Comput. Biol.* 3: e231.
65. Parker J, Rambaut A, Pybus OG. Correlating viral phenotypes with phylogeny: accounting for phylogenetic uncertainty (2008) *Infect. Genet. Evol.* 8: 239-246.
66. Slatkin M, Maddison WP. A clastic measure of gene flow inferred from the phylogenies of alleles (1989) *Genetics* 123: 603-613.
67. Wang TH, Donaldson YK, Brettell RP, Bell JE, Simmonds P. Identification of shared populations of human immunodeficiency virus type 1 infecting microglia and tissue macrophages outside the central nervous system (2001) *J. Virol.* 75: 11686-11699.
68. Schneider TD, Stephens RM. Sequence logos: a new way to display consensus sequences (1990) *Nucleic Acids Res.* 18: 6097-6100.
69. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo generator (2004) *Genome Res.* 14: 1188-1190.
70. NESG. Disorder Prediction MetaServer. In: 2013.

71. Linding R, Jensen LJ, Diella F, Bork P, Gibson TJ, Russell RB. Protein disorder prediction: implications for structural proteomics (2003) *Structure* 11: 1453-1459.
72. Ward JJ, Sodhi JS, McGuffin LJ, Buxton BF, Jones DT. Prediction and functional analysis of native disorder in proteins from the three kingdoms of life (2004) *J. Mol. Biol.* 337: 635-645.
73. Deng X, Eickholt J, Cheng J. PreDisorder: ab initio sequence-based prediction of protein disordered regions (2009) *BMC. Bioinformatics* 10: 436.
74. Prilusky J, Felder CE, Zeev-Ben-Mordehai T, Rydberg EH, Man O, Beckmann JS, Silman I, Sussman JL. FoldIndex: a simple tool to predict whether a given protein sequence is intrinsically unfolded (2005) *Bioinformatics* 21: 3435-3438.
75. Linding R, Russell RB, Neduva V, Gibson TJ. GlobPlot: Exploring protein sequences for globularity and disorder (2003) *Nucleic Acids Res.* 31: 3701-3708.
76. Dosztanyi Z, Csizmok V, Tompa P, Simon I. IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content (2005) *Bioinformatics* 21: 3433-3434.
77. Yang ZR, Thomson R, McNeil P, Esnouf RM. RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins (2005) *Bioinformatics* 21: 3369-3376.
78. Vucetic S, Brown CJ, Dunker AK, Obradovic Z. Flavors of protein disorder (2003) *Proteins* 52: 573-584.
79. R Development Core Team. A language and environment for statistical computing. In: 2009.
80. Groenewoud MJ, Fagrouch Z, Van Gessel S, Niphuis H, Bulavaite A, Warren KS, Heeney JL, Verschoor EJ. Characterization of novel polyomaviruses from Bornean and Sumatran orang-utans (2010) *J. Gen. Virol.* 91: 653-658.
81. Kumar S, Filipinski AJ, Battistuzzi FU, Pond SLK, Tamura K. Statistics and truth in phylogenomics (2012) *Mol. Biol. Evol.* 29: 457-472.
82. Davey NE, Van Roey K, Weatheritt RJ, Toedt G, Uyar B, Altenberg B, Budd A, Diella F, Dinkel H, Gibson TJ. Attributes of short linear motifs (2012) *Mol. Biosyst.* 8: 268-281.
83. Krumbholz A, Bininda-Emonds OR, Wutzler P, Zell R. Phylogenetics, evolution, and medical importance of polyomaviruses (2009) *Infect. Genet. Evol.* 9: 784-799.
84. Firth C, Kitchen A, Shapiro B, Suchard MA, Holmes EC, Rambaut A. Using time-structured data to estimate evolutionary rates of double-stranded DNA viruses (2010) *Mol. Biol. Evol.* 27: 2038-2051.
85. Sugimoto C, Hasegawa M, Kato A, Zheng HY, Ebihara H, Taguchi F, Kitamura T, Yogo Y. Evolution of human Polyomavirus JC: implications for the population history of humans (2002) *J. Mol. Evol.* 54: 285-297.
86. Hatwell JN, Sharp PM. Evolution of human polyomavirus JC (2000) *J. Gen. Virol.* 81: 1191-1200.

87. Zheng HY, Nishimoto Y, Chen Q, Hasegawa M, Zhong S, Ikegaya H, Ohno N, Sugimoto C, Takasaka T, Kitamura T, Yogo Y. Relationships between BK virus lineages and human populations (2007) *Microbes Infect.* 9: 204-213.
88. Bialasiewicz S, Rockett R, Whiley DW, Abed Y, Allander T, Binks M, Boivin G, Cheng AC, Chung JY, Ferguson PE, Gilroy NM, Leach AJ, Lindau C, Rossen JW, Sorrell TC, Nissen MD, Sloots TP. Whole-genome characterization and genotyping of global WU polyomavirus strains (2010) *J. Virol.* 84: 6229-6234.
89. Bialasiewicz S, Whiley DM, Lambert SB, Wang D, Nissen MD, Sloots TP. A newly reported human polyomavirus, KI virus, is present in the respiratory tract of Australian children (2007) *J. Clin. Virol.* 40: 15-18.
90. Duffy S, Shackelton LA, Holmes EC. Rates of evolutionary change in viruses: patterns and determinants (2008) *Nat. Rev. Genet.* 9: 267-276.
91. Miyata T, Yasunaga T. Evolution of overlapping genes (1978) *Nature* 272: 532-535.
92. Sabath N, Wagner A, Karlin D. Evolution of viral proteins originated de novo by overprinting (2012) *Mol. Biol. Evol.* 29: 3767-3780.

## Supplementary Data

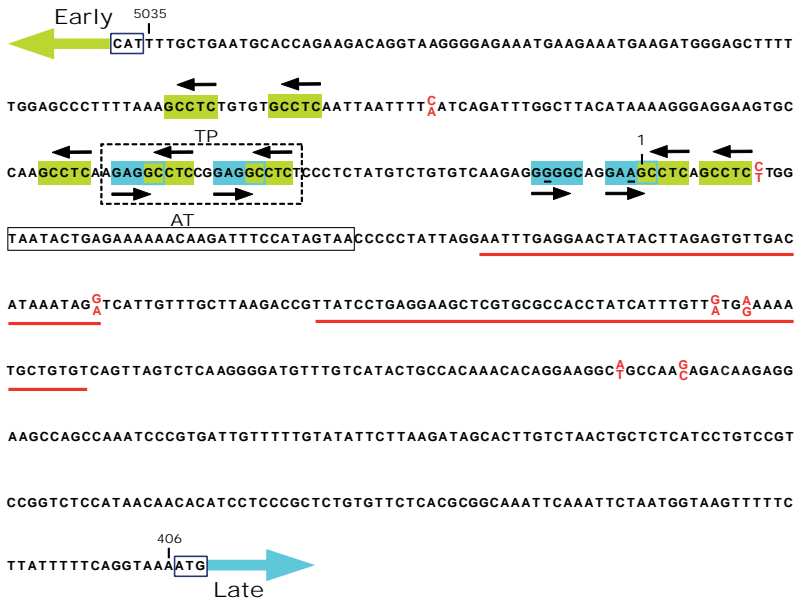
### Supplementary figures



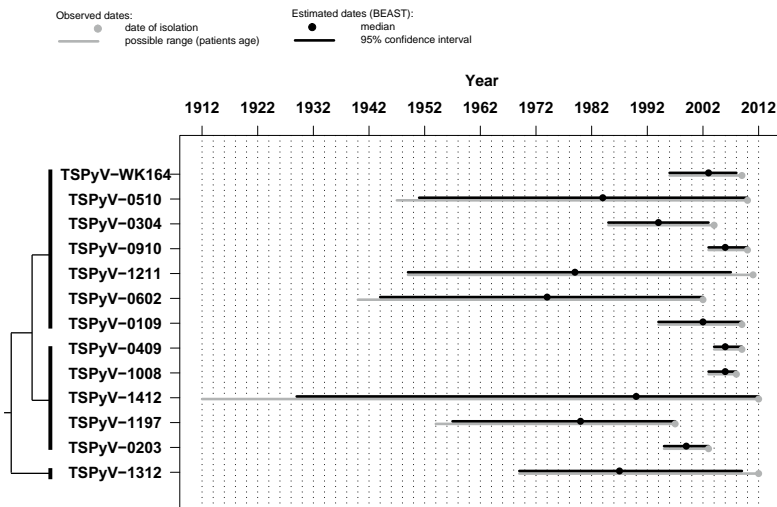
**Supplementary Figure S1.** Amplicons used to sequence TSPyV genomes. **(A)** Location of the eighteen (1 to 18) overlapping amplicons that were generated with primer-pairs specified in **Supplementary Table S1** and used for sequencing of TSPyV genomes. **(B)** Gel-electrophoresis of amplicons. The amplicons size is indicated, including a ladder at left and right side with size indication at left.

	ORF1 Exon	Donor	Intron	Acceptor	ORF5 Exon
<b>MPyV</b>	RF -1	<b>4550</b>		<b>4488</b>	RF -3
Predicted Splice sites	CACAGCGUGUAUAAUCCA	<b>A</b>	<b>GUAAGUAUCA</b> // CUCCCCUAG	<b>A</b>	ACGCGGAGCGAGGAACUG
	H S V Y N P				K R R S E E L
<b>HaPyV</b>	RF -3	<b>4595</b>		<b>4536</b>	RF -2
Predicted Splice sites	UAACAACCGCCUCAGGUAU	<b>A</b>	<b>GUAUGAAUAU</b> // CUUUUUUCAG	<b>G</b>	ACCCUAAUGCUUCCACCU
	Y N P A S G M				T L M L P P
<b>TSPyV</b>	RF -1	<b>4446</b>		<b>4395</b>	RF -1
Predicted Splice sites	AAUCAAGCCUUUAACUGGG	<b>A</b>	<b>GUAAGUAAAA</b> // AUUAAUUUAG	<b>A</b>	UUUUUCCUUUAGCAUGAUG
	N Q G F N W				V F P F S M M

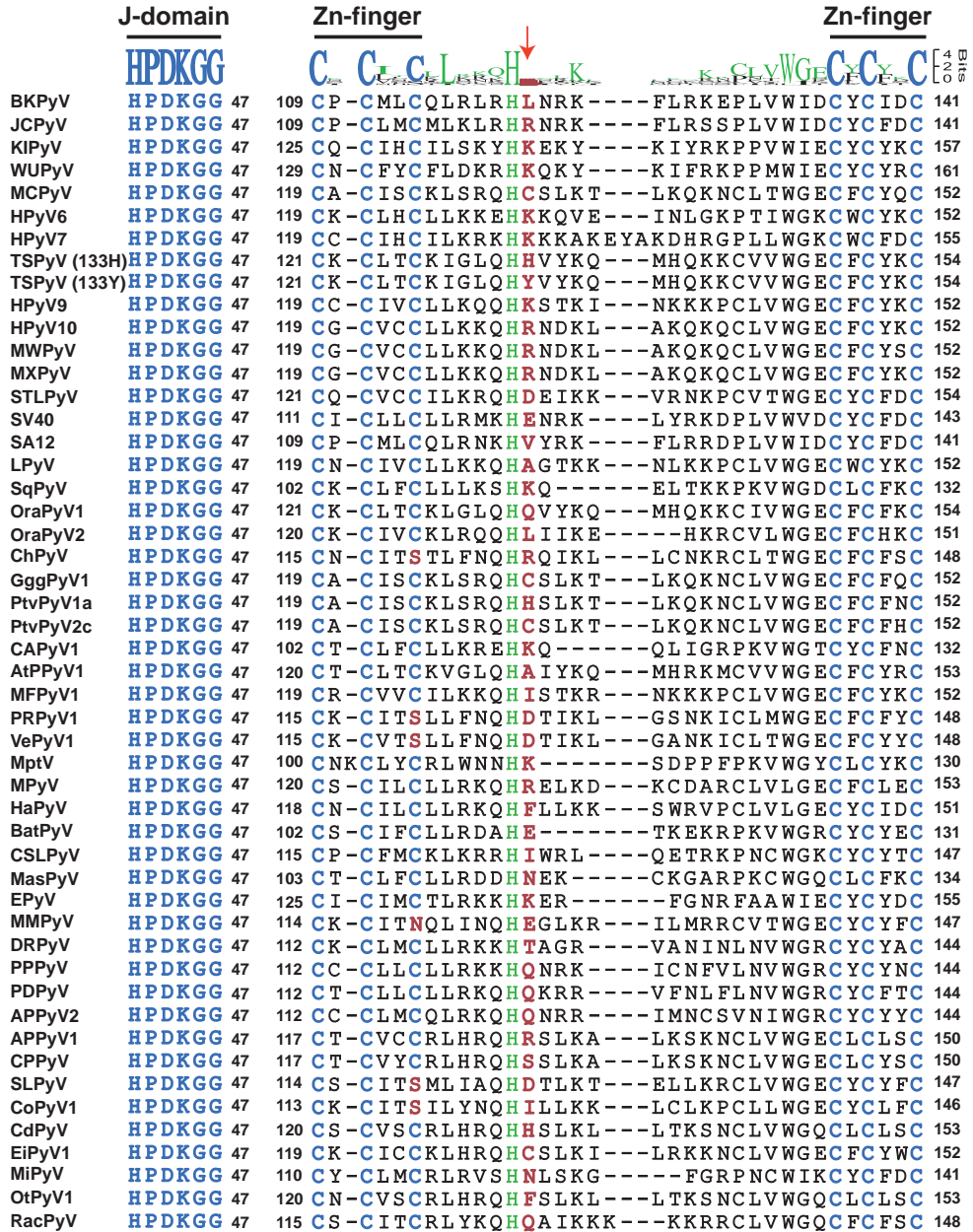
**Supplementary Figure S2.** Donor-acceptor splice sites predicted for TSPyV MT mRNA. Shown are prediction of the donor-acceptor splice sites for MPyV, HaPyV and TSPyV using the web-based Human Splice Finder prediction program (see **M&M**). Splice-donor, GU and AG signals, and splice-acceptor coordinates are depicted in bold. Sequences indicated by // are not shown. Reading frames (RF) are shown at the top of the mRNA sequence and the AA residues below the mRNA sequences. The predicted sites matched the experimentally identified sites for MPyV and HaPyV.



**Supplementary Figure S3.** SNPs and deletions located in the NCCR. Linear projection of the TSPyV NCCR positive strand from the start-codon of the early T-genes (green arrow, position 5035) until the start-codon of the late VP2 gene (turquoise arrow, position 406). The SNPs are depicted in red. The 54nts deletion found in TSPyV-0602 and the 39nts deletion found in TSPyV-1312 are red underlined. Highlighted are TSPyV conserved pentamer enhancer elements. The putative early and late gene enhancers GCC(T/C)C and G(A/G)GGC are highlighted in green and turquoise, respectively. The closed black box depicts the poly(A/T) track where the TATA-box may be located. Downstream of the putative ori region (position 5180 -5200) is a perfect 20-nucleotide true-palindrome (TP) located (dotted black box).



**Supplementary Figure S4.** Dates of sample collection and primary TSPyV infection. The date of virus isolation (grey dots), the possible range of the date of primary infection (grey lines), and the estimated date of primary infection (black dot) as well as its confidence interval (black lines) are shown for each of the 13 TSPyVs.



**Supplementary Figure S5.** Local sequence conservation around the H133Y variation in the ST/MT ORF1-encoded domain. Shown is a multiple sequence alignment around the TSPyV H133Y variation (red arrow) located in between two Zink-fingers (CXCXC motif) in mammalian polyomaviruses, and its conservation profile. Alignment of the upstream highly conserved J-domain (HPDKGG motif) is shown for comparison at left. The selection of polyomaviruses is based on the genomes available before February 1, 2013. A list of acronyms is given in Chapter 5, Supplementary Figure S1.

```

TSPyV   -KVGK-LQVLCTYINLVFFSMMFQPRMEEIYLPMGTPPGPAGGKASIKNGTTCLTPCRT
OraPyV1 VRDGGDTNIVKHVIVVFFSMMFQPRTEETYLPMGTPPGPPGGKASIGIGTSLKTSKT
AtPPyV1 -NGGRGGKAVSLHLIIAFFFSMMFQAGAQEMYLHMAPESGMHGGQFSTQSGTPCSITLMT
      . * : : : :.*****. : * * * . . * * : . * * * . . *

TSPyV   QLSAMPPFPLMNLDLQAPLRDPLLNLARRIQEEEELPHQRTPPAAPRAPSLPPPQSQK
OraPyV1 QTFSVMNPPFPLMNLDLQAPLRAPLHNLARRIQEEEEVR---TPRLAASPPSQQPHQNQR
AtPPyV1 PTSSAMSPPFPLIHLDQTAQLRPSLEILARRIQESEDLLTLTAARVR----SLQPPPSLK
      * . . ***** : ** * * * . * ***** . * : : : . * * . :
      =====
TSPyV   NLSMTLSLMIFLICGLFFLMLSIVIKLYHLF (4437-3988)
OraPyV1 NLSMTLSLMIFLMSCGLFFLLLSIVIRLYHLS (4574-4128)
AtPPyV1 SLKMIMCLAIFLMVLGASCLVLSIVIRLYLHS (4679-4239)
      . * . * : . * * * : * * : * * * * : * *
      =====
TM

```

**Supplementary Figure S6.** Amino acid conservation of the ORF5 in TSPyV, OraPyV1 and AtPPyV1. Shown is a multiple amino-acid alignment of the ORF5 of the three viruses. Genomic coordinates are indicated. Asterisk, conserved amino-acids; single and double dots, similar amino-acids; arrowhead, location of the predicted splice donor site; gray highlight, Serine, Threonine and Tyrosine residues conserved among the three viruses as possible phosphorylation sites; green, amino acid residues varying in TSPyV isolates (see **Figures 1** and **3**); bold, Proline residues. Location of the putative transmembrane domain (TM) is indicated.





VePyV1 CaPyV JCPyV *mPyV* SA12 RacPyV GHPyV BKPyV

*BatsyV* CPyV OrAPyV1 FPyV MptV APPyV1 SqPyV LPyV CSLPyV

EPyV PRPyV1 APP<sub>g</sub>V2 KIPyV **MIRgV** PtvPyV2c OtPyV1 **STLLPyV**

MFPyV1 *KSyV09* SV40 TSPyV CoPyV1 PPPyV CPPyV HPyV12

MXPyV PtvPyV1a PDPyV EIPyV1 AtPPyV1 HaPyV (TggPyV1

CdPyV DRPyV MWPyV APyV CAPyV1 HPyV7 CHPyV MasPyV

WUPyV *HSPyVc* BPyV MCPyV OrAPyV2 MMPyV SLPyV HPyV10

VePyV1 CaPyV JCPyV mPyV SA12

BatPyV CPyV OraPyV1 FPyV MptV APPy

EPyV PRPyV1 APPyV2 KIPyV MiPyV P

MFPyV1 HSPyV9 SV40 TSPyV CoPyV1

MPyV PtvPyV1a PDPyV EiPyV1 At

CdPyV DRPyV MWPyV APyV CAPyV1

WUPyV HSPyV6 BPyV MCPyV OraPyV2

## Part IV

### Discussion and Summary

yV BKPyV

V CSLPyV

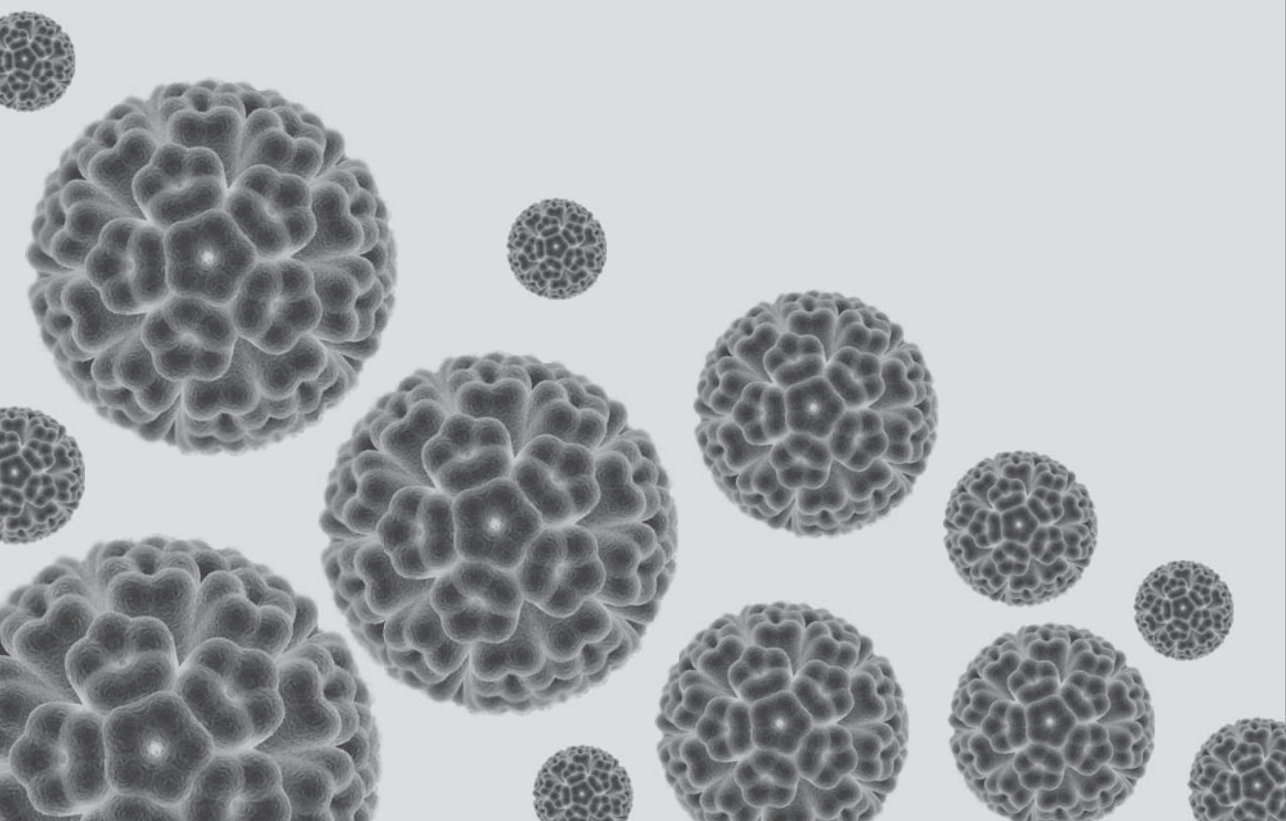
V1 STLPyV

V HPyV12

V GgPyV1

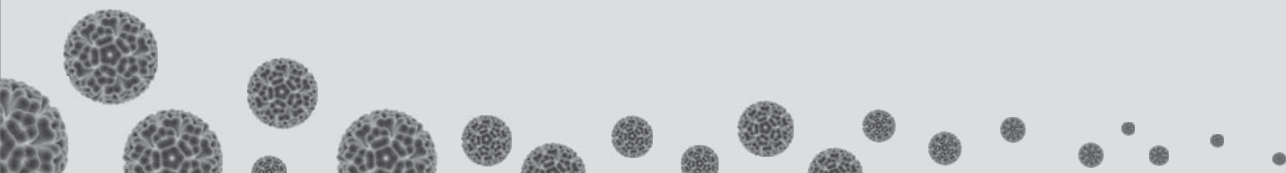
V MasPyV

V HPyV10



# Chapter 7

Summarizing discussion



## General aspects of TSPyV infection

**D**iscovery of the trichodysplasia spinulosa-associated polyomavirus (TSPyV) in 2010 paved the way for research focusing on the role of this virus in the pathogenesis of trichodysplasia spinulosa (TS) [1, 2]. When one considers a causal relationship between a pathogen and a certain (proliferative) disease, as proposed by Fredericks and Relman for instance [3], the prevalence of the putative pathogen is expected to be higher in the specified patient group compared to a healthy population (**Chapters 2 and 3**) [2, 4], and the pathogen (viral)-load of symptomatic tissue is expected to significantly exceed that of healthy tissue (**Chapter 3**) [4]. Equally important, by action of its encoded proteins, the pathogen should possess specific pathogenic (transforming) properties that render the infected cell, tissue or organ in a state reminiscent of the indicated disease (**Chapter 4**) [5]. In the process, these interacting viral proteins responsible for adaptation in the host may mutate (**Chapter 6**). Some of these mutations may be recurrent in the virus family arguing for a conserved mechanism of virus adaptation (**Chapters 5**).

In this summarizing chapter, most of these aspects of TSPyV infection, including its prevalence, localization, latency, and persistence, as well as viral pathogenesis and TSPyV host adaptation and evolution will be discussed in the light of available literature to establish a causative role for TSPyV in the development of TS.

### Virus reservoir for latency/persistence

TSPyV is a ubiquitous virus similarly to most other human polyomaviruses (HPyVs) as can be concluded from its high seroprevalence (**Table 1**). In addition, the intensity of measured TSPyV antibody responses is high [8 - 10]. Apparently, TSPyV infection creates sufficient viral antigen to promote strong, possibly lifelong, antibody responses, suggesting that efficient TSPyV replication takes place at compartment(s) in the body where virus progeny is produced. Whether replication takes place only in the skin, or elsewhere in the body as well, is unclear at the moment. The intensity of the measured seroresponses, to some extent comparable with BKPyV-seroresponses, might indicate circulation of TSPyV in blood (viremia).

In contrast to its high seroprevalence (~75%), the TSPyV DNA prevalence analyzed with PCR in skin-derived samples of groups of asymptomatic (healthy and immunocompromized) individuals is much lower (<5%) (**Table 1** and **Chapter 2**) [2, 11 - 14]. When TSPyV is detected on the skin, its load is usually very low (**Chapters 2 and 3**). The discrepancy between DNA-prevalence and seroprevalence is evident for many polyomaviruses, for instance KIPyV and WUPyV, and HPyV9 (**Table 1**).

Despite explanations of a technical nature, such as differences in sensitivity to detect TSPyV infection, this discrepancy might be explained by persistent infection of body site(s) other than the skin, for instance of internal organs after dissemination. In this regard, it is worthwhile to note that for instance JCPyV persistently infects lymphocytes and urothelial cells, whereas upon reactivation it invades central nerve cells and causes encephalitis

in immunocompromized hosts [16, 41]. A recent finnish study suggested that TSPyV might persist in lymphoid tissues, specifically the tonsils, of healthy individuals [11], and an american study showed the presence of TSPyV in a kidney biopsy [42]. Nevertheless, the detected percentage in these tissues did not exceed the prevalence found on the skin. Previous studies that looked into lymphoid systems or mucosa (e.g., tonsils) as a residence or reservoir for HPyV infections suggested that these body compartments might play a role in polyomavirus latency/persistence [43 - 46], which might be the case for TSPyV as well. Alternatively, these findings may simply imply a respiratory transmission route for these viruses, which results in transient detection in these lymphoid systems or mucosal tissues (e.g., tonsils).

### A possible mechanism of polyomavirus latency/persistence

In general, it is believed that particular virus infections may reside in two distinct non-pathological or asymptomatic states, namely the (non-replicative) latency and low-level persistent replication state, together designated latency/persistence. For this state of latency/persistence, viruses have evolved mechanisms that focus on evading detection and clearance by the host immune system. For instance, viruses encode microRNAs that interfere with host cellular processes [47, 48]. These small non-coding RNAs [49] are normally utilized by polyomaviruses to regulate their own transcriptome, but can also target the host transcriptome, for example involved in cellular innate and adaptive immune responses (**Figure 1**) [50]. Thus, next to virus transcriptome auto-regulation – shown for MCPyV [51, 52], BKPyV

**Table 1.** Prevalence of human polyomavirus infection in immunocompetent individuals measured by PCR and serology

Polyomavirus*	Prevalence of polyomavirus infection	
	Viral DNA	Viral serum antibodies
	Range in % (major detection site) [Ref]	Range in % [Ref]
BKPyV	30 - 50 (Kidney) [15]	80 - 100 [16 - 19]
JCPyV	10 - 40 (Kidney) [20]	40 - 70 [16 - 19]
KIPyV	1 - 12 (Respiratory tract) [21 - 23]	55 - 90 [17, 18, 24, 25]
WUPyV	1 - 16 (Respiratory tract) [22, 26]	70 - 100 [17, 18, 24, 25]
MCPyV	60 - 80 (Skin) [27 - 29]	40 - 80 [9, 17, 18, 30 - 32]
HPyV6	14 - 50 (Skin) [30, 33]	50 - 80 [9, 30]
HPyV7	11 - 17 (Skin) [30, 33]	35 - 60 [9, 30]
TSPyV	2 - 4 (Skin/Tonsil) [2, 11 - 14]	70 - 80 [8 - 10]
HPyV9	1 - 17 (Skin) [33, 34]	30 - 50 [9, 34]
HPyV10**	1 - 5 (Stool) [35 - 37]	N/A
STLPyV	0 - 1 (Stool) [35]	60 - 70 [38]
HPyV12	1 - 11 (GI tract) [39]	17 - 23 [39]
NJPyV	N/A (Endothelial cells) [40]	N/A

\*, For polyomavirus name abbreviation see **Chapter 1, Table 1**

\*\* ,HPyV10 includes MWPyV and MXPyV

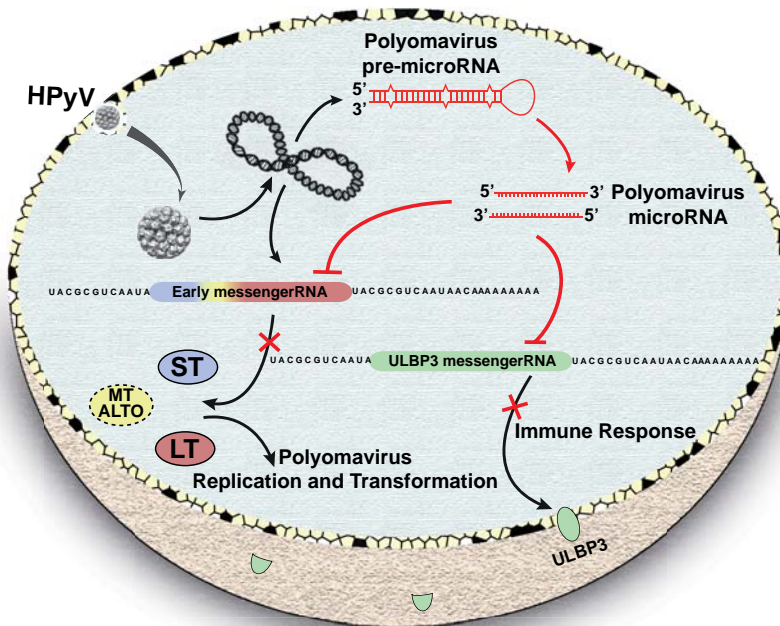
GI-tract, Gastro-intestinal tract; N/A, No data available so far

[53] and JCPyV [50, 54] – microRNAs play most probably key roles in controlling specific host factors to evade the immune system, which could promote latency/persistence (**Figure 1**) [50, 52, 55]. In this context, preliminary genome analysis of TSPyV suggests that it might express a microRNA compatible with regulation of its early transcripts (**Chapter 2, Figure 2; Chapter 5, Supplementary Figure S2**). Whether this microRNA also targets host cell factors, specifically those involved in immune recognition, is unknown.

## TSPyV active infection and trichodysplasia spinulosa

### From immunosuppression to disease development; a causal relationship

It seems apparent that all other established pathogenic HPyVs described so far, i.e., JCPyV, BKPyV and MCPyV, follow similar rules; i) an asymptomatic primary initial infection at an early age, ii) widespread prevalence in human populations, and iii) upon reactivation pathogenic consequences observed only in the elderly and/or immunocompromized individuals. TSPyV, with its described (sero)prevalence and its pathogenic consequences in severely immunocompromized hosts, seems to fit these rules.



**Figure 1.** Schematic overview of putative function of polyomavirus microRNA related to infection. Polyomavirus-encoded microRNAs exert an auto-regulatory role by targeting their own early messengerRNAs. Upon viral infection, the innate immune response produces factors, such as the stress-induced ligand ULBP3, which are essential for recognition of infected cells by the immune system (e.g., NK cells). This viral microRNA plays also a putative role in regulation of this host immune response by targeting the ULBP3 messengerRNA. Figure adapted from [48].

One of the aims described in this dissertation was to study and describe the pathogenic consequences of TSPyV infection. In **Chapter 3** it was reported that high viral loads were exclusively measured in lesional samples originating from TS patients, which indicates that TSPyV is actively replicating in these lesions [4]. In healthy subjects and tissues, TSPyV was almost undetectable (**Chapters 2 and 3**). Another indication that active TSPyV infection is involved in the development of TS was the detection of viral proteins VP1 solely in the affected hair follicles (**Chapter 3**). Furthermore, TSPyV LT-antigen was expressed in all lesional TS samples analyzed, lesions which are known to be hyperproliferative (**Chapter 4**) (discussed in next paragraph) [5, 56].

For BKPyV and JCPyV it is well established that when higher viral loads are reached in affected organs, this could have pathogenic consequences [57 - 59]. According to Rothman *et al.*, this could fit the complete causal effect, which states that “an agent that causes disease should be both necessary and sufficient for the disease to occur and should be reproducible and consistent in different settings, and in studies performed by different investigators” [60]. In this regard, hypothetically speaking, it could be stated that, when i) the correlation is strong – all TS cases tested are positive for TSPyV, compared to healthy subjects, ii) reproducible – other research groups reporting TSPyV presence in TS lesions, and iii) predictive – when detection of high TSPyV loads in TS lesions is evident, then a causative conclusion is highly probable [4, 42, 61, 62]. Equally important iv) would be the molecular evidence of active infection in this regard (discussed next). Furthermore, other criteria’s could be considered for testing to establish a causal relationship between a microbe and a disease, which in this case would also be in line with the Koch’s postulates, by investigating the TSPyV infection and pathogenesis in animal models.

Altogether, despite the limitations in number of TS cases described and available worldwide, and the number of tissue samples analyzed, the studies described in **Chapter 2** and particularly in **Chapter 3** provide strong evidence that active TSPyV infection is associated with TS, and TSPyV is most probably the causative agent of this disease. Obviously, more studies are needed from independent laboratories that further confirm the epidemiological and experimental studies to underscore causality.

## Molecular pathology of trichodysplasia spinulosa

### Possible pathogenic events in polyomavirus-associated disease development

During polyomavirus-associated human disease development, a complex interplay takes place between multiple factors, like host immunity, ageing and genetics, and environmental exposure. In addition, virus-host molecular interactions occur, which can induce changes in viral and cellular expression profiles and potentially disrupt physiological cellular pathways. This issue is best illustrated by MCPyV, the HPyV associated with Merkel cell carcinoma (MCC) [63, 64], a skin tumor already introduced in **Chapter 1** [65]. While MCPyV infects



humans early in life in a fashion comparable to many others, including non-pathogenic polyomaviruses, several successive events (although very rare) can result in neoplasm development. At first, an exogenous mutagenic exposure such as UV-radiation causes DNA damage, for instance DNA double-strand breaks. The subsequent repair of this damage could facilitate virus genome integration into the host DNA [66] (**Figure 2A**). Additional successive mutations, for instance also as a result of UV-radiation, that abrogate the LT-antigen function involved in viral DNA replication without disturbing the pathogenic (transforming) domains (e.g., the LXCXE pRB binding site), is an important event in MCC cell survival and proliferation (**Figure 2A**).

Comparable to MCPyV, TSPyV induces robust proliferation of its host cells located in the hair follicles (see **Chapter 4** and discussion next) (**Figure 2B**) [5]. Whether TSPyV can integrate into the host genome is unknown so far. In symptomatic TS disease, however, the measured high TSPyV loads argue against integration in the productive stage of infection (**Chapter 3**).

### Viral-host molecular interplay

Historically, polyomavirus research has repeatedly revealed identification of specific viral-protein and host-protein interactions that lead to important insights into mammalian cell biology [70, 71]. Polyomaviruses have evolved mechanisms that enable them to hijack host regulatory cell factors for the sake of fulfilling their life cycle (**Figure 3**). The ability of a polyomavirus to oncogenically transform its host cells was first demonstrated for SV40. Inoculation of SV40 DNA into rat and mouse cells resulted in uncontrolled cell growth and proliferation [72]. Ever since, large tumor (LT), middle tumor (MT) and small tumor (ST) antigen of several polyomaviruses have been investigated and were postulated as prototype virus oncoproteins that can abrogate several host cell regulatory factors (**Figure 3**). Until recently, MT-antigen expression was demonstrated only for rodent polyomaviruses [71]. New data shows that MT-antigen and/or its derivative ALTO are expressed also by MCPyV [73] and TSPyV (Van der Meijden and Feltkamp *et al.*, unpublished observations), as briefly mentioned in **Chapters 5** and **6**.

---

**Figure 2.** Molecular events in the (putative) pathogenesis of MCC and TS. **(A)** (1) In general, MCPyV infection (red particles) is acquired at early childhood. (2) In a state of decreased immunity (either iatrogenic or age-related), reactivation and replication of MCPyV may occur. (3) At some stage, such a burst of infective virus production could facilitate viral integration in susceptible cells. (4) Mutagenic events could result into replication-competent virus infection. (5) Expression of MCPyV mutation-truncated LT-antigen that provides pro-proliferative signals (e.g., phosphorylation of pRB) leading to clonal expansion and MCC development [67]. Figure adapted from [68]. **(B)** (1) Also TSPyV infection (blue particles) is acquired at early childhood. (2) During decreased immunity, TSPyV reactivation and replication may occur. (3) Alongside high viral replication, tumor antigens are expressed that provide pro-proliferative signals (e.g., phosphorylation of pRB) [5]. (4) Hair follicles accumulate with viral particles – which are also shed from the skin – and trichohyalin-protein aggregates resulting into keratinized, spiny structures protruding the skin [4, 69].

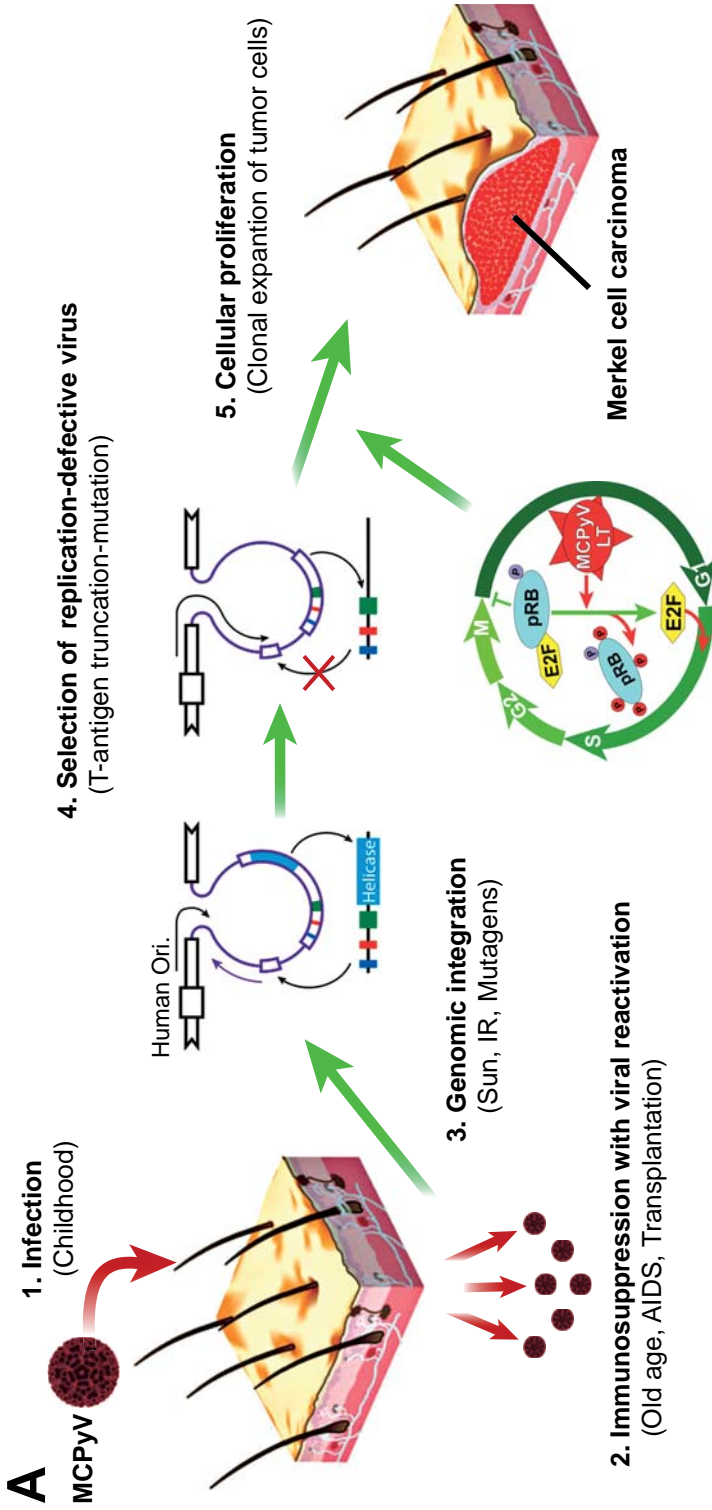


Figure 2. For legend text panel A see previous page. For panel B see next page.

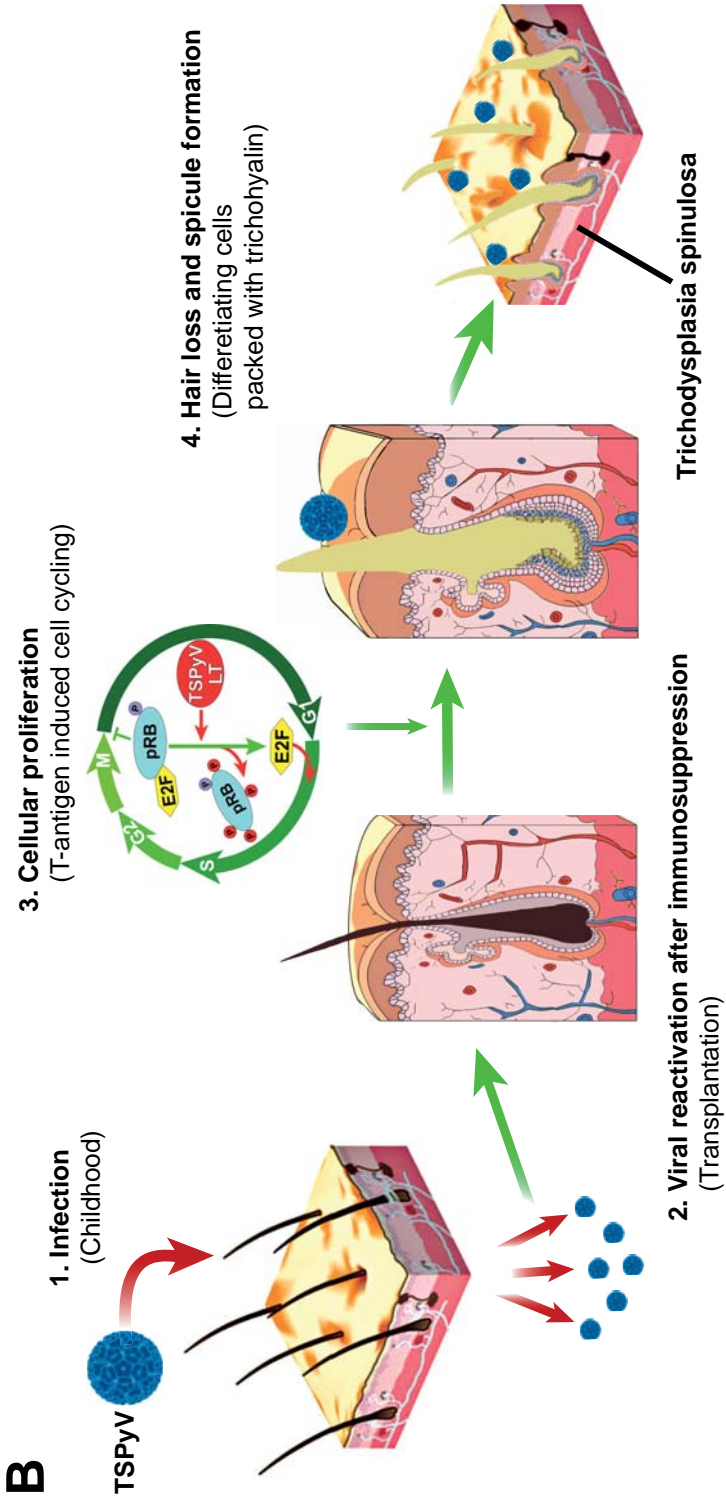


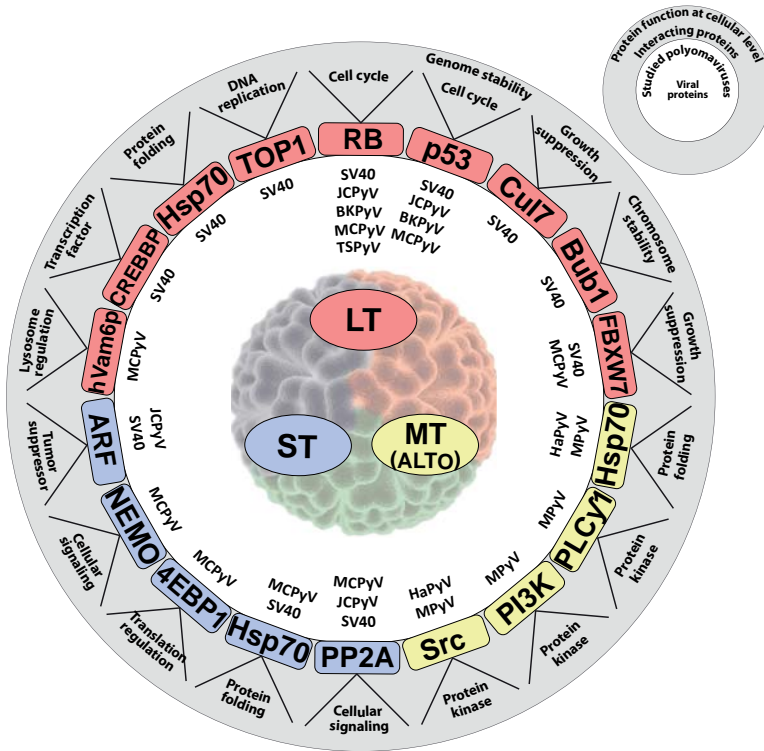
Figure 2. For legend text panel B see page 176.

The patho/oncogenicity of the polyomavirus-encoded transforming proteins, most notably the LT-antigen, in cell culture and in animal models has been known for decades. Cellular interacting proteins in this regard are the retinoblastoma tumor suppressor protein (pRB) and its family members (p107 and p130) (discussed next) [74]. The poly-functional LT-antigen is expressed at early stages of the viral life cycle and is involved in both viral replication (ori-binding and helicase activity [75]) and virus-induced cellular transformation and growth [74]. In addition to the full-length LT-antigen, alternative splicing of the LT-antigen messengerRNA can occur, as well as premature termination of translation because of mutations (discussed before). Truncation of LT-antigen usually causes removal of the helicase and putative p53-binding domains, but preserves most of its other domains, for instance the LXCXE domain important for interaction with the RB family members [76]. Preliminary findings suggest that the LT-antigen of TSPyV, comparable to other *in vitro* studied polyomaviruses, demonstrates comparable versatility which expression is regulated through the alternative splicing of TSPyV pre-T-antigen messengerRNA (Van der Meijden and Felkamp, personal communication). The regulation of (L)T-antigen expression emphasizes that TSPyV, just like other polyomaviruses, has developed an evolutionary feature that enables switching between roles in both viral replication and cellular transformation using different versions of the same protein (other polyomavirus evolutionary aspects are discussed next).

Probably one of the most described interactions and a key mechanism in cellular transformation caused by polyomaviruses is the LT-antigen interaction with the RB family proteins (**Figure 3**). RB family proteins (i.e., pRB, p107 and p130) are important regulators of the G1- (rest) to S-phase (DNA synthesis) transition of cells during cell cycling initiation [99]. Hypophosphorylated pRB, for instance, inhibits function of the E2F transcription factor that regulates gene expression required for host DNA synthesis, whereas hyperphosphorylated pRB induces pRB–E2F complex dissociation and cell-cycle entry. The latter can be reverted by several inhibitory proteins such as p16<sup>ink4a</sup> and p21<sup>waf</sup> [5, 99]. For MCPyV it was shown that ST- and LT-antigen, including its derivative 57kT that bears also the pRB-binding LXCXE motif, are required for the proliferation of MCC cells [91, 96, 100 - 102]. The TSPyV LT-antigen also contains the conserved LXCXE motif identified *in silico* (**Chapter 5**) [1, 6]. The colocalization of phosphorylated pRB with proliferation marker Ki-67 observed in TS lesions could indicate that pRB is inactivated, which explains the hyperproliferative nature of these lesions (**Chapter 4**). Thus, through its LT-antigen or one of its truncated derivatives, TSPyV might induce cell cycle progression through disruption of the RB-regulatory pathway, and create a reservoir of proliferating cells that permit viral DNA replication. This putative mechanism could be assessed in future *in vitro* studies – aside other factors indicated in **Figure 3** – that may be involved in the pathogenic process of cellular transformation.

The characteristic TS phenotype observed repeatedly and most notable on the face, remains unexplained so far. However, terminal differentiation of the large proliferating pool of cells into inner root sheath (IRS) might explain the accumulation of trichohyalin-positive cells that facilitate formation of the characteristic spicules (**Figure 2B, Chapters 3 and 4**). IRS

cells are known to produce keratins and trichohyalin that serve as an intracellular “cement”, which gives strength to the IRS layer of a hair follicle to support and mold the growing and protruding hair shaft [103]. Hypothetically, the excess amounts of keratins and trichohyalin produced could shape the spicule on the skin when the cells have undergone differentiation. To investigate spicule formation and involvement of TSPyV T-antigens in more detail, TSPyV T-antigen-expressing organotypic raft culture systems could be exploited, which mimic skin growth in a sophisticated 3D-Skin culture system [104].



**Figure 3.** A selection of host cellular factors targeted by polyomavirus early proteins. So far known, polyomavirus LT- (red), ST- (blue) and MT-antigen (yellow), as shown in the inner circle for each virus protein, interact with many cellular proteins (colored according viral proteins), as depicted in the outer circle with their function at cellular level on top (see also inset, top-right). For LT- and ST-antigen, most experiments that have identified these interactions have been shown around SV40 polyomavirus [74, 77]. The J-domain, shared by all early T-antigens, recruits Hsp70 protein homologues that aids in protein folding [78 - 80]. Other regions inside LT-antigen have been shown to bind other important factors, like CREB-binding protein (CREBBP) [81], TOP1 [82], p53 [74, 76, 83], Cul7 [84], Bub1 [85] and FBXW7 [86]. MCPyV LT-antigen, in addition, can probably alter hVam6p protein function that is important in lysosome regulation [87]. LT-antigen of the human pathogenic polyomaviruses (JCPyV, BKPyV and MCPyV) has been shown to abrogate functions of RB proteins (i.e., pRB, p107 and p130), which probably includes TSPyV LT-antigen (Chapter 4) [5, 88 - 91]. A major function of ST-antigen is ascribed to its ability to modulate PP2A cellular signaling protein [92 - 94]. Aside that, recently NEMO [95] and 4EBP1 [96] modulation was identified for MCPyV ST-antigen. MT/ALTO-antigen, expressed by MPyV, HaPyV, MCPyV, TSPyV and likely other members of the *Orthopolyomavirus-l* (Chapter 5), can regulate cellular signal transduction through modulation of protein kinases Src, PI3K and PLCγ1 [71, 97, 98].

## TSPyV evolution and host adaptation

### Molecular signatures/footprints of polyomavirus host-adaptation

Virus host adaptation facilitates virus dissemination, which could also promote virus speciation. Virus adaptation to the host involves fixation of beneficial mutations under the selection pressure imposed by the host. Successful adaptations may eventually lead to establishment of a new virus species that is genetically fit to occupy a separate niche in the host or multiple hosts. Virus evolution is constrained by host environment and many structural and genetic factors [105], including overlapping genes [106, 107]. Overlap regards to the expressible nucleotide sequence of one gene that can be part of expressible nucleotide sequence of another gene in an alternate reading frame, which may contribute to the expression and function of one or more gene products. The overlapping gene is always *de novo*, in case of polyomaviruses the ALTO/MT-antigen, which overlaps an ancestral gene, i.e., the LT-antigen, and the evolutionary forces active on these genes may differ [108]. When the ancestral gene product regions are evolutionary conserved - called the short linear motifs (SLiMs) - because of their critical role in regulatory function, protein-protein interaction or signal transduction [109, 110], selection pressure may act more meticulous on that genomic region.

In **Chapter 5**, we studied an overlapping open reading frame (ORF), called ORF5, that is conserved in a large monophyletic lineage of polyomaviruses designated the Ortho-I genogroup of the *Polyomaviridae* family (**Chapter 1, Figure 8**) [111], which includes four human viruses, i.e., MCPyV, TSPyV, HPyV12 and NJPyV. We have discovered this conservation independently from Carter *et al.* [73], who designed and called this overlapping ORF ALTO in their study on MCPyV. Besides the above viruses, Ortho-I genogroup includes two rodent polyomaviruses whose ORF5 encodes the second exon of MT-antigen (**Figure 3**) [71]. Sequence region of ORF2 that encodes the N-terminal part of LT-antigen, and the overlapping ORF5 encoding the ALTO or second exon of MT-antigen, encompasses the LXCXE domain and the ORF5m2 motif, respectively, both of which are considered SLiMs. Accelerated switching between Valine and Alanine was shown in this SLiM of ORF5 in ORF5-plus viruses as a result of the Cysteine residue conservation in LXCXE domain (**Chapter 5**). This was also the predominant site of amino-acid-changing mutations in TSPyV (**Chapter 6**). This switching between residues – called the COCO-VA toggling – implied a positive (purifying) selection on this site of ORF5m2, indicating the importance of this SLiM to preserve the overall function of the ALTO/MT-antigen. This observation lends support to ORF5 expression – either “directly” through the synthesis of MCPyV ALTO protein [73] or upon alternative splicing events directing the synthesis of TSPyV MT-antigen [71] (Van der Meijden and Feltkamp, personal communication).

The MPyV and HaPyV MT-antigens have been investigated for decades now. It is generally accepted that MT-antigen is involved in cellular transformation through modulation of several cellular protein kinases and signal transduction mediators (**Figure 3**) [71].

The observed toggling in ALTO/MT-antigen could emphasize a site-reversion mutation that involves, for instance, escape from immunological pressure facilitating adaptation of the protein to its niche [112]. Still, many questions remain unanswered. How common is CO-CO-VA toggling in other (overlapping) SLiMs? Is it one of many only in polyomaviruses or also evident even beyond this virus family? These are just few questions for future research. Finally, as an extreme case of adaptive evolution, CO-CO-VA toggling may enlighten evolutionary thinking and stimulate development of applications to extend the research on other (polyoma)virus (overlapping) SLiMs.

### TSPyV evolution as an example of polyomavirus host-adaptation

Polyomaviruses do not encode viral DNA polymerase enzymes, but rather rely on host-cell enzymes for their replication [75, 82]. Only during cellular DNA synthesis-phase, an optimal condition is created that facilitates viral replication because of excess supply of host polymerases and deoxyribonucleotides [113]. Human DNA polymerase enzymes are of high fidelity and therefore HPyV DNA mutation rates are expected similar to the host mutation rates. For TSPyV was calculated that it indeed has low intra-host viral DNA substitutions per site per year [114], comparable to those established for BKPyV and JCPyV [115, 116]. Taken into account the host DNA substitution rate and the time-span over which these viral substitutions could have taken place, it was calculated that BKPyV and JCPyV may have diverged recently from a common ancestor – less than 1000 and 350 years, respectively, which on the evolutionary scale is comparable to what we found for TSPyV – less than 4300 years. Still, these data on TSPyV estimation of divergence (**Chapter 6**), and notably also on BKPyV and JCPyV [115, 116], should be sought with caution because of the large uncertainty as a result of extremely large confidence intervals of the calculated evolutionary rates.

The dominant evolutionary model of polyomavirus divergence until recently was virus-host co-evolution [117 - 119]. On the contrary, recent studies showed that polyomavirus evolution could have taken place at least two orders of magnitude faster making the assumptions of a zoonotic transmission of primate polyomaviruses more probable. The calculated time of divergence of the most recent common ancestor (tMRCA) of TSPyV and its closest polyomavirus relative OraPyV1, provided that it may be valid to assume that these two viruses emerged recently from a common ancestor [114]. Similar assumptions exist for JCPyV [116] and BKPyV [115, 120], viruses that most probably have emerged from a common ancestor closely related to SV40 and SA12 – both baboon viruses – with a phyletic position in the Ortho-II polyomavirus genogroup (**Chapter 1, Figure 8**) [111]. Although similar estimates are lacking for most of the recently identified HPyVs, when analysing the polyomavirus phylogenetic tree and looking at the Ortho-I genogroup, one can note that many HPyVs appear to be closely related to viruses isolated from non-human primates (**Chapter 1, Figure 8**) [111]. In this regard, HPyV9 is very similar to the B-lymphotropic polyomavirus (LPyV, also known as the African green monkey polyomavirus, AGMPyV). Furthermore, MCPyV resembles the Gorilla gorilla gorilla polyomavirus 1 (GggPyV1), as well as the Pan

troglodytes versus polyomaviruses (PtvPyV). As indicated above and assessed in **Chapter 6**, TSPyV resembles Bornean orangutan polyomavirus 1 (OraPyV1). Finally, the recently identified NJPyV shows close relationship to the Chimpanzee polyomavirus (ChPyV) [40]. Thus, both virus-host co-evolution as well as a zoonotic non-human to human jump of most (if not all) primate polyomaviruses should be considered. Which of these hypothetical scenarios may be true for TSPyV and new HPyVs, awaits future estimations with the help of larger cohorts of virus genome isolates and sophisticated polyomavirus evolutionary estimation-models.

As discussed in the previous paragraph, genome and proteome adaptation of viruses to the host involve mutations because of the selection pressure imposed by the host. In **Chapter 5**, we described COCO-VA toggling as a probable evolutionary selection mechanism. In **Chapter 6**, we provided genetic evidence for this mechanism with ORF5 being the predominant place of non-synonymous substitutions during TSPyV evolution. The four identified motifs in this ORF5 are under selection pressure because they overlap with the conserved ORF2 motifs. The motif involved in COCO-VA mechanism was located in the ORF5m2 motif. Out of three TSPyV lineages (TSI, TSII and TSIII), the TSI and TSII viruses used the Alanine residue most frequently at the COCO-VA site, implying that the Alanine residue provided a fitness-gain for the virus to adapt to its human host and may have facilitated additional TSPyV speciation (see **Chapter 6** and previous discussion). Experimental studies on TSPyV ALTO/MT-antigen could reveal molecular forces that drive toggling in this protein, in particular in the concerning SLiMs that facilitate in protein function by playing critical roles in protein regulation, protein-protein interaction or signal transduction. Taken together, without restraint one could hypothesize that the putative ALTO/MT-antigen of TSPyV is acting as a cellular protein kinase and signal transduction protein to stimulate cell growth and cellular transformation and future *in vitro* studies could prove this proposition.

## Concluding remarks and future perspectives

We are only beginning to comprehend the basics regarding TSPyV and many knowledge gaps remain. One of the unknowns in TSPyV research is the mechanism(s) underlying latency/persistence and reactivation. In this regard, it is still unknown whether TS disease is the result of a primary TSPyV infection in the midst of immunosuppression or whether it concerns a reactivation from a latent site, be it skin or another not yet identified body location. Especially, limitations in the number of TS patients and availability of follow-up samples obtained after primary infection from different organs and sites have prevented further insights in this matter. *In vitro* study of (putative) TSPyV microRNA regulation, which is known to play a key role in latency/persistence and reactivation in other (polyoma)viruses, might offer additional opportunities in this regard.

As indicated in **Figure 3**, the knowledge about (putative) molecular interactions between T-antigens of polyomaviruses and cellular pathways is fragmented. Many factors have



been analyzed only for SV40 and needs to be validated for other (human) polyomaviruses as well. The lack of an experimental model for TSPyV makes answering these questions difficult. Skin-mimicking organotypic culture models could provide means for investigating the TSPyV and other cutaneous polyomaviruses. Knowledge about the cell type(s) that serve as the viral reservoir(s) could help to study specific mechanisms and pathways locally in monolayers, for instance IRS cells for TSPyV. With these specific and tailored *in vitro* culture systems at hand for research, one could identify interacting host molecular partners of TSPyV T-antigens *in vivo*.

Bioinformatics analysis of the polyomavirus evolution is an area of investigation that is gradually evolving and recent discoveries highlight that a broader taxonomic range of hosts may exist. The *Polyomaviridae* family tree may grow beyond what is known. The close relationship of primate polyomaviruses may provide interesting leads for further studies into inter- and intra-specific evolution of this virus family, as well as virus adaptation and speciation. First, with regard to virus-host evolution, TSPyV is probably a good example because of its possible recent zoonotic ancestry. Together with OraPyV1 they could be used as a model to elucidate intra-host evolution and adaptation of other HPyVs that also closely resemble non-human primate polyomaviruses. Second, with regard to host adaptation and speciation, the novel evolutionary mechanism called COCO-VA toggling described in this dissertation, could be an excellent starting point for studies investigating the potential effect of protein-residue toggling on mechanisms that circumvent host defenses. *In silico*, one could try to identify how common this COCO-VA toggling mechanism is in other (viral) proteins that show overlapping ORFs, whether it involves SLiMs and what it means for virus adaptation and speciation. These are just few experimental and bioinformatics questions for future research that have surfaced from research described in this dissertation.

Altogether, the research described in this dissertation provided many new insights regarding the prevalence, pathogenesis, evolution and host adaptation of TSPyV, which is one of the newer human polyomaviruses. These insights have substantially increased our understanding of TSPyV and the etiology of trichodysplasia spinulosa. Furthermore, they will allow us to better comprehend the clinical impact of polyomavirus infections in general, and provide new leads for future studies devoted to the identification of (antiviral) options to prevent or treat polyomavirus-associated diseases.

## References

1. Van der Meijden E, Janssens RW, Lauber C, Bouwes Bavinck JN, Gorbalenya AE, Feltkamp MC. Discovery of a new human polyomavirus associated with trichodysplasia spinulosa in an immunocompromized patient (2010) *PLoS Pathog.* 6: e1001024.
2. Kazem S, van der Meijden E, Feltkamp MC. The trichodysplasia spinulosa-associated polyomavirus: virological background and clinical implications (2013) *APMIS* 121: 770-782.
3. Fredericks DN, Relman DA. Sequence-based identification of microbial pathogens: a reconsideration of Koch's postulates (1996) *Clin. Microbiol. Rev.* 9: 18-33.
4. Kazem S, van der Meijden E, Kooijman S, Rosenberg AS, Hughey LC, Browning JC, Sadler G, Busam K, Pope E, Benoit T, Fleckman P, de VE, Eekhof JA, Feltkamp MC. Trichodysplasia spinulosa is characterized by active polyomavirus infection (2012) *J. Clin. Virol.* 53: 225-230.
5. Kazem S, van der Meijden E, Wang RC, Rosenberg AS, Pope E, Benoit T, Fleckman P, Feltkamp MC. Polyomavirus-associated trichodysplasia spinulosa involves hyperproliferation, pRB phosphorylation and upregulation of p16 and p21 (2014) *PLoS One* 9: e108947.
6. Lauber C, Kazem S, Kravchenko AA, Feltkamp MC, Gorbalenya AE. Interspecific adaptation by binary choice at de novo polyomavirus T antigen site through accelerated codon-constrained Val-Ala toggling within an intrinsically disordered region (2015) *Nucleic Acids Res.* doi: 10.1093/nar/gkv378
7. Kazem S, Lauber C, van der Meijden E, Kooijman S, Kravchenko AA, The TrichSpin Network, Feltkamp MC, Gorbalenya AE. Adaptation of trichodysplasia spinulosa-associated polyomavirus to the human population is mediated by middle T antigen and involves COCO-VA toggling (2014) [Submitted]
8. van der Meijden E, Kazem S, Burgers MM, Janssens R, Bouwes Bavinck JN, de Melker H, Feltkamp MC. Seroprevalence of Trichodysplasia Spinulosa-associated Polyomavirus (2011) *Emerg. Infect. Dis.* 17: 1355-1363.
9. van der Meijden E, Bialasiewicz S, Rockett RJ, Tozer SJ, Sloots TP, Feltkamp MC. Different serologic behavior of MCPyV, TSPyV, HPyV6, HPyV7 and HPyV9 polyomaviruses found on the skin (2013) *PLoS One* 8: e81078.
10. Chen T, Mattila PS, Jartti T, Ruuskanen O, Soderlund-Venermo M, Hedman K. Seroepidemiology of the Newly Found Trichodysplasia Spinulosa-Associated Polyomavirus (2011) *J. Infect. Dis.* 204: 1523-1526.
11. Sadeghi M, Aaltonen LM, Hedman L, Chen T, Soderlund-Venermo M, Hedman K. Detection of TS polyomavirus DNA in tonsillar tissues of children and adults: Evidence for site of viral latency (2014) *J. Clin. Virol.* 59: 55-58.

12. Siebrasse EA, Bauer I, Holtz LR, Le B, Lassa-Claxton S, Canter C, Hmiel P, Shenoy S, Sweet S, Turmelle Y, Shepherd R, Wang D. Human polyomaviruses in children undergoing transplantation, United States, 2008–2010 (2012) *Emerg. Infect. Dis.* 18: 1676-1679.
13. Scuda N, Hofmann J, Calvignac-Spencer S, Ruprecht K, Liman P, Kuhn J, Hengel H, Ehlers B. A novel human polyomavirus closely related to the african green monkey-derived lymphotropic polyomavirus (2011) *J. Virol.* 85: 4586-4590.
14. Wieland U, Silling S, Hellmich M, Potthoff A, Pfister H, Kreuter A. Human polyomaviruses 6, 7, 9, 10 and Trichodysplasia spinulosa-associated polyomavirus in HIV-infected men (2014) *J. Gen. Virol.* 95: 928-932.
15. Nিকেleit V, Singh HK, Mihatsch MJ. Polyomavirus nephropathy: morphology, pathophysiology, and clinical management (2003) *Curr. Opin. Nephrol. Hypertens.* 12: 599-605.
16. Egli A, Infanti L, Dumoulin A, Buser A, Samaridis J, Stebler C, Gosert R, Hirsch HH. Prevalence of polyomavirus BK and JC infection and replication in 400 healthy blood donors (2009) *J. Infect. Dis.* 199: 837-846.
17. Kean JM, Rao S, Wang M, Garcea RL. Seroepidemiology of Human Polyomaviruses (2009) *PLoS Pathog.* 5: e1000363.
18. Carter JJ, Paulson KG, Wipf GC, Miranda D, Madeleine MM, Johnson LG, Lemos BD, Lee S, Warcola AH, Iyer JG, Nghiem P, Galloway DA. Association of Merkel Cell Polyomavirus-Specific Antibodies With Merkel Cell Carcinoma (2009) *J. Natl. Cancer Inst.* 4: 1510-1522.
19. Kjaerheim K, Roe OD, Waterboer T, Sehr P, Rizk R, Dai HY, Sandeck H, Larsson E, Andersen A, Boffetta P, Pawlita M. Absence of SV40 antibodies or DNA fragments in prediagnostic mesothelioma serum samples (2007) *Int. J. Cancer* 120: 2459-2465.
20. Aoki N, Kitamura T, Tominaga T, Fukumori N, Sakamoto Y, Kato K, Mori M. Immunohistochemical detection of JC virus in nontumorous renal tissue of a patient with renal cancer but without progressive multifocal leukoencephalopathy (1999) *J. Clin. Microbiol.* 37: 1165-1167.
21. Mourez T, Bergeron A, Ribaud P, Scieux C, de Latour RP, Tazi A, Socie G, Simon F, LeGoff J. Polyomaviruses KI and WU in immunocompromised patients with respiratory disease (2009) *Emerg. Infect. Dis.* 15: 107-109.
22. Babakir-Mina M, Ciccozzi M, Perno CF, Ciotti M. The human polyomaviruses KI and WU: virological background and clinical implications (2013) *APMIS* 121: 746-754.
23. Babakir-Mina M, Ciccozzi M, Bonifacio D, Bergallo M, Costa C, Cavallo R, Di BL, Perno CF, Ciotti M. Identification of the novel KI and WU polyomaviruses in human tonsils (2009) *J. Clin. Virol.* 46: 75-79.
24. Neske F, Prifert C, Scheiner B, Ewald M, Schubert J, Opitz A, Weissbrich B. High prevalence of antibodies against polyomavirus WU, polyomavirus KI, and human bocavirus in German blood donors (2010) *BMC Infect. Dis.* 10: 215.

25. Nguyen NL, Le BM, Wang D. Serologic evidence of frequent human infection with WU and KI polyomaviruses (2009) *Emerg. Infect. Dis.* 15: 1199-1205.
26. Teramoto S, Kaiho M, Takano Y, Endo R, Kikuta H, Sawa H, Ariga T, Ishiguro N. Detection of KI polyomavirus and WU polyomavirus DNA by real-time polymerase chain reaction in nasopharyngeal swabs and in normal lung and lung adenocarcinoma tissues (2011) *Microbiol. Immunol.* 55: 525-530.
27. Wieland U, Mauch C, Kreuter A, Krieg T, Pfister H. Merkel cell polyomavirus DNA in persons without merkel cell carcinoma (2009) *Emerg. Infect. Dis.* 15: 1496-1498.
28. Foulongne V, Kluger N, Dereure O, Mercier G, Moles JP, Guillot B, Segondy M. Merkel cell polyomavirus in cutaneous swabs (2010) *Emerg. Infect. Dis.* 16: 685-687.
29. Coursaget P, Samimi M, Nicol JT, Gardair C, Touze A. Human Merkel cell polyomavirus: virological background and clinical implications (2013) *APMIS* 121: 755-769.
30. Schowalter RM, Pastrana DV, Pumphrey KA, Moyer AL, Buck CB. Merkel cell polyomavirus and two previously unknown polyomaviruses are chronically shed from human skin (2010) *Cell Host Microbe* 7: 509-515.
31. Tolstov YL, Pastrana DV, Feng H, Becker JC, Jenkins FJ, Moschos S, Chang Y, Buck CB, Moore PS. Human Merkel cell polyomavirus infection II. MCV is a common human infection that can be detected by conformational capsid epitope immunoassays (2009) *Int. J. Cancer* 125: 1250-1256.
32. Touze A, Gaitan J, Arnold F, Cazal R, Fleury MJ, Combelas N, Sizaret PY, Guyétant S, Maruani A, Baay M, Tognon M, Coursaget P. Generation of Merkel cell polyomavirus (MCV)-like particles and their application to detection of MCV antibodies (2010) *J. Clin. Microbiol.* 48: 1767-1770.
33. Foulongne V, Sauvage V, Hebert C, Dereure O, Cheval J, Gouilh MA, Pariente K, Segondy M, Burguiere A, Manuguerra JC, Caro V, Eloit M. Human skin microbiota: high diversity of DNA viruses identified on the human skin by high throughput sequencing (2012) *PLoS One* 7: e38499.
34. van der Meijden E, Wunderink HF, van der Blij-de Brouwer CS, Zaaier HL, Rotmans JI, Bavinck JN, Feltkamp MC. Human polyomavirus 9 infection in kidney transplant patients (2014) *Emerg. Infect. Dis.* 20: 991-999.
35. Lim ES, Reyes A, Antonio M, Saha D, Ikumapayi UN, Adeyemi M, Stine OC, Skelton R, Brennan DC, Mkakosya RS, Manary MJ, Gordon JI, Wang D. Discovery of STL polyomavirus, a polyomavirus of ancestral recombinant origin that encodes a unique T antigen by alternative splicing (2013) *Virology* 436: 295-303.
36. Siebrasse EA, Reyes A, Lim ES, Zhao G, Mkakosya RS, Manary MJ, Gordon JI, Wang D. Identification of MW Polyomavirus, a Novel Polyomavirus in Human Stool (2012) *J. Virol.* 86: 10321-10326.

37. Yu G, Greninger AL, Isa P, Phan TG, Martinez MA, de la Luz SM, Contreras JF, Santos-Preciado JI, Parsonnet J, Miller S, Derisi JL, Delwart E, Arias CF, Chiu CY. Discovery of a novel polyomavirus in acute diarrheal samples from children (2012) *PLoS One* 7: e49449.
38. Lim ES, Meinerz NM, Primi B, Wang D, Garcea RL. Common Exposure to STL Polyomavirus During Childhood (2014) *Emerg. Infect. Dis.* 20: 1559-1561.
39. Korup S, Rietscher J, Calvignac-Spencer S, Trusch F, Hofmann J, Moens U, Sauer I, Voigt S, Schmuck R, Ehlers B. Identification of a novel human polyomavirus in organs of the gastrointestinal tract (2013) *PLoS One* 8: e58021.
40. Mishra N, Pereira M, Rhodes RH, An P, Pipas JM, Jain K, Kapoor A, Briese T, Faust PL, Lipkin WI. Identification of a Novel Polyomavirus in a Pancreatic Transplant Recipient With Retinal Blindness and Vasculitic Myopathy (2014) *J. Infect. Dis.* 210: 1595-1599.
41. Cinque P, Koralik IJ, Gerevini S, Miro JM, Price RW. Progressive multifocal leukoencephalopathy in HIV-1 infection (2009) *Lancet Infect. Dis.* 9: 625-636.
42. Fischer MK, Kao GF, Nguyen HP, Drachenberg CB, Rady PL, Tyring SK, Gaspari AA. Specific detection of trichodysplasia spinulosa-associated polyomavirus DNA in skin and renal allograft tissues in a patient with trichodysplasia spinulosa (2012) *Arch. Dermatol.* 148: 726-733.
43. Goudsmit J, Wertheim-van DP, van SA, van der Noordaa J. The role of BK virus in acute respiratory tract disease and the presence of BKV DNA in tonsils (1982) *J. Med. Virol.* 10: 91-99.
44. Kato A, Kitamura T, Takasaka T, Tominaga T, Ishikawa A, Zheng HY, Yogo Y. Detection of the archetypal regulatory region of JC virus from the tonsil tissue of patients with tonsillitis and tonsillar hypertrophy (2004) *J. Neurovirol.* 10: 244-249.
45. Monaco MC, Atwood WJ, Gravell M, Tornatore CS, Major EO. JC virus infection of hematopoietic progenitor cells, primary B lymphocytes, and tonsillar stromal cells: implications for viral latency (1996) *J. Virol.* 70: 7004-7012.
46. Monaco MC, Jensen PN, Hou J, Durham LC, Major EO. Detection of JC virus DNA in human tonsil tissue: evidence for site of initial viral infection (1998) *J. Virol.* 72: 9918-9923.
47. Sullivan CS, Ganem D. MicroRNAs and viral infection (2005) *Mol. Cell* 20: 3-7.
48. Lagatie O, Tritsmans L, Stuyver LJ. The miRNA world of polyomaviruses (2013) *Virol. J.* 10: 268.
49. Carthew RW, Sontheimer EJ. Origins and Mechanisms of miRNAs and siRNAs (2009) *Cell* 136: 642-655.
50. Bauman Y, Nachmani D, Vitenshtein A, Tsukerman P, Drayman N, Stern-Ginossar N, Lankry D, Gruda R, Mandelboim O. An identical miRNA of the human JC and BK polyoma viruses targets the stress-induced ligand ULBP3 to escape immune elimination (2011) *Cell Host Microbe* 9: 93-102.

51. Seo GJ, Chen CJ, Sullivan CS. Merkel cell polyomavirus encodes a microRNA with the ability to autoregulate viral gene expression (2009) *Virology* 383: 183-187.
52. Lee S, Paulson KG, Murchison EP, Afanasiev OK, Alkan C, Leonard JH, Byrd DR, Hannon GJ, Nghiem P. Identification and validation of a novel mature microRNA encoded by the Merkel cell polyomavirus in human Merkel cell carcinomas (2011) *J. Clin. Virol.* 52: 272-275.
53. Broekema NM, Imperiale MJ. miRNA regulation of BK polyomavirus replication during early infection (2013) *Proc. Natl. Acad. Sci. U. S. A* 110: 8200-8205.
54. Seo GJ, Fink LH, O'Hara B, Atwood WJ, Sullivan CS. Evolutionarily conserved function of a viral microRNA (2008) *J. Virol.* 82: 9823-9828.
55. Boss IW, Renne R. Viral miRNAs and immune evasion (2011) *Biochim. Biophys. Acta* 1809: 708-714.
56. Haycox CL, Kim S, Fleckman P, Smith LT, Piepkorn M, Sundberg JP, Howell DN, Miller SE. Trichodysplasia spinulosa--a newly described folliculocentric viral infection in an immunocompromised host (1999) *J. Investig. Dermatol. Symp. Proc.* 4: 268-271.
57. Bossolasco S, Calori G, Moretti F, Boschini A, Bertelli D, Mena M, Gerevini S, Bestetti A, Pedale R, Sala S, Sala S, Lazzarin A, Cinque P. Prognostic significance of JC virus DNA levels in cerebrospinal fluid of patients with HIV-associated progressive multifocal leukoencephalopathy (2005) *Clin. Infect. Dis.* 40: 738-744.
58. Ramos E, Drachenberg CB, Papadimitriou JC, Hamze O, Fink JC, Klassen DK, Drachenberg RC, Wiland A, Wali R, Cangro CB, Schweitzer E, Bartlett ST, Weir MR. Clinical course of polyoma virus nephropathy in 67 renal transplant patients (2002) *J. Am. Soc. Nephrol.* 13: 2145-2151.
59. Bayliss J, Karasoulos T, Bowden S, Glogowski I, McLean CA. Immunosuppression increases latent infection of brain by JC polyomavirus (2011) *Pathology* 43: 362-367.
60. Rothman KJ, Greenland S. Causation and causal inference in epidemiology (2005) *Am. J. Public Health* 95 Suppl 1: S144-S150.
61. Wanat KA, Holler PD, Dentchev T, Simbiri K, Robertson E, Seykora JT, Rosenbach M. Viral-associated trichodysplasia: characterization of a novel polyomavirus infection with therapeutic insights (2012) *Arch. Dermatol.* 148: 219-223.
62. Lee YY, Tucker SC, Prow NA, Setoh YX, Banney LA. Trichodysplasia spinulosa: A benign adnexal proliferation with follicular differentiation associated with polyomavirus (2013) *Australas. J. Dermatol.*
63. Feng H, Shuda M, Chang Y, Moore PS. Clonal integration of a polyomavirus in human Merkel cell carcinoma (2008) *Science* 319: 1096-1100.
64. Arora R, Chang Y, Moore PS. MCV and Merkel cell carcinoma: a molecular success story (2012) *Curr. Opin. Virol.* 2: 489-498.
65. Agelli M, Clegg LX. Epidemiology of primary Merkel cell carcinoma in the United States (2003) *J. Am. Acad. Dermatol.* 49: 832-841.

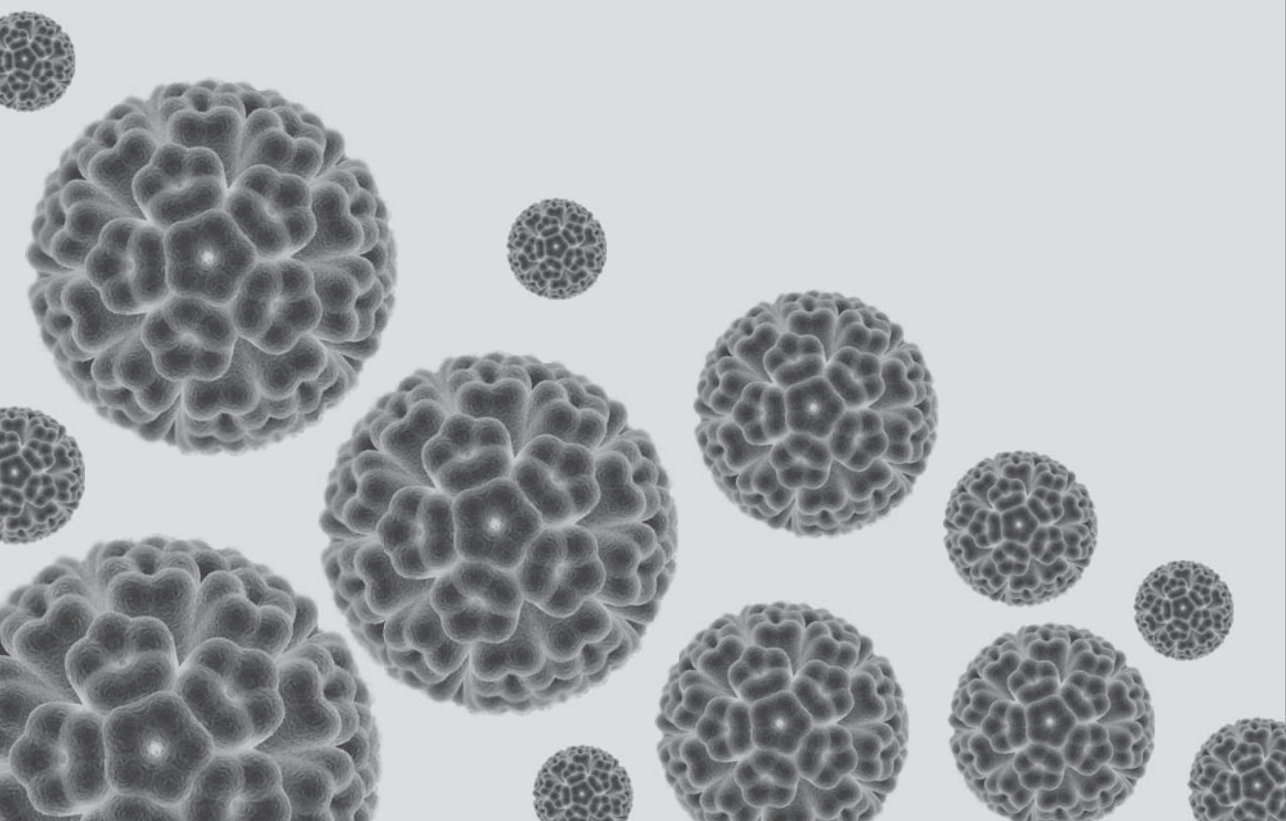
66. Molho-Pessach V, Lotem M. Ultraviolet radiation and cutaneous carcinogenesis (2007) *Curr. Probl. Dermatol.* 35: 14-27.
67. Sihto H, Kukko H, Koljonen V, Sankila R, Bohling T, Joensuu H. Merkel cell polyomavirus infection, large T antigen, retinoblastoma protein and outcome in Merkel cell carcinoma (2011) *Clin. Cancer Res.* 17: 4806-4813.
68. Chang Y, Moore PS. Merkel cell carcinoma: a virus-induced human cancer (2012) *Annu. Rev. Pathol.* 7: 123-144.
69. Schwieger-Briel A, Balma-Mena A, Ngan B, Dipchand A, Pope E. Trichodysplasia spinulosa--a rare complication in immunosuppressed patients (2010) *Pediatr. Dermatol.* 27: 509-513.
70. Pipas JM. SV40: Cell transformation and tumorigenesis (2009) *Virology* 384: 294-303.
71. Fluck MM, Schaffhausen BS. Lessons in signaling and tumorigenesis from polyomavirus middle T antigen (2009) *Microbiol. Mol. Biol. Rev.* 73: 542-563.
72. Abrahams PJ, van der Eb AJ. In vitro transformation of rat and mouse cells by DNA from simian virus 40 (1975) *J. Virol.* 16: 206-209.
73. Carter JJ, Daugherty MD, Qi X, Bheda-Malge A, Wipf GC, Robinson K, Roman A, Malik HS, Galloway DA. Identification of an overprinting gene in Merkel cell polyomavirus provides evolutionary insight into the birth of viral genes (2013) *Proc. Natl. Acad. Sci. U. S. A.* 110: 12744-12749.
74. Topalis D, Andrei G, Snoeck R. The large tumor antigen: a "Swiss Army knife" protein possessing the functions required for the polyomavirus life cycle (2013) *Antiviral Res.* 97: 122-136.
75. Smelkova NV, Borowiec JA. Dimerization of simian virus 40 T-antigen hexamers activates T-antigen DNA helicase activity (1997) *J. Virol.* 71: 8766-8773.
76. Borchert S, Czech-Sioli M, Neumann F, Schmidt C, Wimmer P, Dobner T, Grundhoff A, Fischer N. High-affinity Rb-binding, p53 inhibition, subcellular localization and transformation by wild type or tumor-derived shortened Merkel Cell Polyomavirus Large T-antigens (2013) *J. Virol.* 88: 3144-3160.
77. Cheng J, DeCaprio JA, Fluck MM, Schaffhausen BS. Cellular transformation by Simian Virus 40 and Murine Polyoma Virus T antigens (2009) *Semin. Cancer Biol.* 19: 218-228.
78. Decaprio JA. The role of the J domain of SV40 large T in cellular transformation (1999) *Biologicals* 27: 23-28.
79. Campbell KS, Mullane KP, Aksoy IA, Stubdal H, Zalvide J, Pipas JM, Silver PA, Roberts TM, Schaffhausen BS, Decaprio JA. DnaJ/hsp40 chaperone domain of SV40 large T antigen promotes efficient viral DNA replication (1997) *Genes Dev.* 11: 1098-1110.
80. Pipas JM. Molecular chaperone function of the SV40 large T antigen (1998) *Dev. Biol. Stand.* 94: 313-319.
81. Ali SH, Decaprio JA. Cellular transformation by SV40 large T antigen: interaction with host proteins (2001) *Semin Cancer Biol* 11: 15-23.

82. Khopde S, Roy R, Simmons DT. The binding of topoisomerase I to T antigen enhances the synthesis of RNA-DNA primers during simian virus 40 DNA replication (2008) *Biochemistry* 47: 9653-9660.
83. Wiesend WN, Parasuraman R, Li W, Farinola MA, Rooney MT, Hick SK, Samarapungavan D, Cohn SR, Reddy GH, Rocher LL, Dumler F, Lin F, Zhang PL. Adjuvant role of p53 immunostaining in detecting BK viral infection in renal allograft biopsies (2010) *Ann. Clin. Lab. Sci.* 40: 324-329.
84. Kasper JS, Kuwabara H, Arai T, Ali SH, Decaprio JA. Simian virus 40 large T antigen's association with the CUL7 SCF complex contributes to cellular transformation (2005) *J. Virol.* 79: 11685-11692.
85. Cotsiki M, Lock RL, Cheng Y, Williams GL, Zhao J, Perera D, Freire R, Entwistle A, Golemis EA, Roberts TM, Jat PS, Gjoerup OV. Simian virus 40 large T antigen targets the spindle assembly checkpoint protein Bub1 (2004) *Proc. Natl. Acad. Sci. U. S. A.* 101: 947-952.
86. Kwun HJ, Shuda M, Feng H, Camacho CJ, Moore PS, Chang Y. Merkel cell polyomavirus small T antigen controls viral replication and oncoprotein expression by targeting the cellular ubiquitin ligase SCFFbw7 (2013) *Cell Host Microbe* 14: 125-135.
87. Liu X, Hein J, Richardson SC, Basse PH, Toptan T, Moore PS, Gjoerup OV, Chang Y. Merkel cell polyomavirus large T antigen disrupts lysosome clustering by translocating human Vam6p from the cytoplasm to the nucleus (2011) *J. Biol. Chem.* 286: 17079-17090.
88. Decaprio JA, Ludlow JW, Figge J, Shew JY, Huang CM, Lee WH, Marsilio E, Paucha E, Livingston DM. SV40 large tumor antigen forms a specific complex with the product of the retinoblastoma susceptibility gene (1988) *Cell* 54: 275-283.
89. Harris KF, Christensen JB, Radany EH, Imperiale MJ. Novel mechanisms of E2F induction by BK virus large-T antigen: requirement of both the pRb-binding and the J domains (1998) *Mol. Cell Biol.* 18: 1746-1756.
90. Bollag B, Kilpatrick LH, Tyagarajan SK, Tevethia MJ, Frisque RJ. JC virus T'135, T'136 and T'165 proteins interact with cellular p107 and p130 in vivo and influence viral transformation potential (2006) *J. Neurovirol.* 12: 428-442.
91. Houben R, Adam C, Baeurle A, Hesbacher S, Grimm J, Angermeyer S, Henzel K, Hauser S, Elling R, Brocker EB, Gaubatz S, Becker JC, Schrama D. An intact retinoblastoma protein-binding site in Merkel cell polyomavirus large T antigen is required for promoting growth of Merkel cell carcinoma cells (2012) *Int. J. Cancer* 130: 847-856.
92. Arroyo JD, Hahn WC. Involvement of PP2A in viral and cellular transformation (2005) *Oncogene* 24: 7746-7755.
93. Cho US, Morrone S, Sablina AA, Arroyo JD, Hahn WC, Xu W. Structural basis of PP2A inhibition by small t antigen (2007) *PLoS Biol.* 5: e202.



94. Bollag B, Hofstetter CA, Reviriego-Mendoza MM, Frisque RJ. JC virus small T antigen binds phosphatase PP2A and Rb family proteins and is required for efficient viral DNA replication activity (2010) *PLoS One* 5: e10606.
95. Griffiths DA, Abdul-Sada H, Knight LM, Jackson BR, Richards K, Prescott EL, Peach AH, Blair GE, Macdonald A, Whitehouse A. Merkel cell polyomavirus small T antigen targets the NEMO adaptor protein to disrupt inflammatory signaling (2013) *J. Virol.* 87: 13853-13867.
96. Shuda M, Kwun HJ, Feng H, Chang Y, Moore PS. Human Merkel cell polyomavirus small T antigen is an oncoprotein targeting the 4E-BP1 translation regulator (2011) *J. Clin. Invest.* 121: 3623-3634.
97. Su W, Liu W, Schaffhausen BS, Roberts TM. Association of Polyomavirus middle tumor antigen with phospholipase C-gamma 1 (1995) *J. Biol. Chem.* 270: 12331-12334.
98. Zhou AY, Ichaso N, Adamarek A, Zila V, Forstova J, Dibb NJ, Dilworth SM. Polyomavirus middle T-antigen is a transmembrane protein that binds signaling proteins in discrete subcellular membrane sites (2011) *J. Virol.* 85: 3046-3054.
99. Sherr CJ, McCormick F. The RB and p53 pathways in cancer (2002) *Cancer Cell* 2: 103-112.
100. Shuda M, Feng H, Kwun HJ, Rosen ST, Gjoerup O, Moore PS, Chang Y. T antigen mutations are a human tumor-specific signature for Merkel cell polyomavirus (2008) *Proc. Natl. Acad. Sci. U. S. A.* 105: 16272-16277.
101. Houben R, Shuda M, Weinkam R, Schrama D, Feng H, Chang Y, Moore PS, Becker JC. Merkel cell polyomavirus-infected Merkel cell carcinoma cells require expression of viral T antigens (2010) *J. Virol.* 84: 7064-7072.
102. Shuda M, Chang Y, Moore PS. Merkel cell polyomavirus-positive Merkel cell carcinoma requires viral small T-antigen for cell proliferation (2014) *J. Invest. Dermatol.* 134: 1479-1481.
103. Steinert PM, Parry DA, Marekov LN. Trichohyalin mechanically strengthens the hair follicle: multiple cross-bridging roles in the inner root sheath (2003) *J. Biol. Chem.* 278: 41409-41419.
104. Kazem S, van der Meijden E, Struijk L, de Gruijl FR, Feltkamp MC. Human papillomavirus 8 E6 disrupts terminal skin differentiation and prevents pro-Caspase-14 cleavage (2012) *Virus Res.* 163: 609-616.
105. Crick FH, Barnett L, Brenner S, Watts S-Tobin RJ. General nature of the genetic code for proteins (1961) *Nature* 192: 1227-1232.
106. Sabath N, Wagner A, Karlin D. Evolution of viral proteins originated de novo by overprinting (2012) *Mol. Biol. Evol.* 29: 3767-3780.
107. Pushker R, Mooney C, Davey NE, Jacque JM, Shields DC. Marked variability in the extent of protein disorder within and between viral families (2013) *PLoS One* 8: e60724.

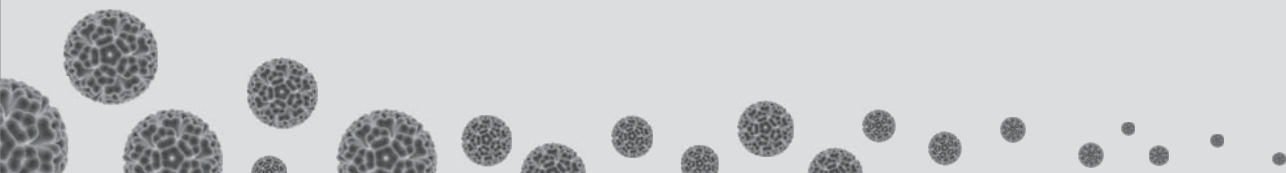
108. Rogozin IB, Spiridonov AN, Sorokin AV, Wolf YI, Jordan IK, Tatusov RL, Koonin EV. Purifying and directional selection in overlapping prokaryotic genes (2002) *Trends Genet.* 18: 228-232.
109. Davey NE, Van Roey K, Weatheritt RJ, Toedt G, Uyar B, Altenberg B, Budd A, Diella F, Dinkel H, Gibson TJ. Attributes of short linear motifs (2012) *Mol. Biosyst.* 8: 268-281.
110. Neduva V, Russell RB. Linear motifs: evolutionary interaction switches (2005) *FEBS Lett.* 579: 3342-3345.
111. Feltkamp MC, Kazem S, van der Meijden E, Lauber C, Gorbalenya AE. From Stockholm to Malawi: recent developments in studying human polyomaviruses (2013) *J. Gen. Virol.* 94: 482-496.
112. Delpont W, Scheffler K, Seoighe C. Frequent toggling between alternative amino acids is driven by selection in HIV-1 (2008) *PLoS Pathog.* 4: e1000242.
113. Bernhoff E, Gutteberg TJ, Sandvik K, Hirsch HH, Rinaldo CH. Cidofovir inhibits polyomavirus BK replication in human renal tubular cells downstream of viral early gene expression (2008) *Am. J. Transplant.* 8: 1413-1422.
114. Kazem S, Lauber C, van der Meijden E, Kooijman S, Kravchenko AA, The TrichSpin Network, Feltkamp MC, Gorbalenya AE. Global circulation of slowly evolving trichodysplasia spinulosa-associated polyomavirus and its adaptation to the human population through alternative T antigens (2013) *J. Neurovirol.* 19: 298-299.
115. Chen Y, Sharp PM, Fowkes M, Kocher O, Joseph JT, Koralnik IJ. Analysis of 15 novel full-length BK virus sequences from three individuals: evidence of a high intra-strain genetic diversity (2004) *J. Gen. Virol.* 85: 2651-2663.
116. Shackelton LA, Rambaut A, Pybus OG, Holmes EC. JC virus evolution and its association with human populations (2006) *J. Virol.* 80: 9928-9933.
117. Hatwell JN, Sharp PM. Evolution of human polyomavirus JC (2000) *J. Gen. Virol.* 81: 1191-1200.
118. Sugimoto C, Hasegawa M, Kato A, Zheng HY, Ebihara H, Taguchi F, Kitamura T, Yogo Y. Evolution of human Polyomavirus JC: implications for the population history of humans (2002) *J. Mol. Evol.* 54: 285-297.
119. Krumbholz A, Bininda-Emonds OR, Wutzler P, Zell R. Phylogenetics, evolution, and medical importance of polyomaviruses (2009) *Infect. Genet. Evol.* 9: 784-799.
120. Firth C, Kitchen A, Shapiro B, Suchard MA, Holmes EC, Rambaut A. Using time-structured data to estimate evolutionary rates of double-stranded DNA viruses (2010) *Mol. Biol. Evol.* 27: 2038-2051.



# Chapter 8

Samenvatting (Dutch)

(Persian/Dari) مختصر



Leden van de *Polyomaviridae* familie, waarvan één ervan ten grondslag lag aan dit doctoraalonderzoek, behoren tot de kleinste virussen en zijn ongeveer 45 nanometer groot. Humane polyomavirussen komen wereldwijd voor. Sinds 2007 zijn er 11 nieuwe soorten ontdekt. Hun erfelijke materiaal bestaat uit DNA<sup>1</sup> dat ongeveer 5000 basenparen<sup>2</sup> omvat. Dit erfelijke materiaal codeert voor twee soorten eiwitten die te eniger tijd tijdens een infectie tot expressie kunnen komen. De eiwitten die vroeg tijdens een polyomavirus infectie worden aangemaakt, worden tumor (T)-eiwitten genoemd. Hier zijn er drie van: groot (LT), medium (MT) en kleine (ST). Deze eiwitten zijn noodzakelijk voor de levenscyclus van het virus. De eiwitten die later tijdens een infectie worden aangemaakt, zijn de “structurele” virale eiwitten (zoals VP1 en VP2) die de virale eiwitmantel vormen.

Polyomavirussen zijn over het algemeen goedaardige infectieuze virussen waar naar schatting 50-100% van de mensen mee geïnfecteerd is. De eerste twee humane polyomavirussen, die al in 1971 waren ontdekt, zijn JCPyV en BKPyV. Epidemiologische en klinische studies tonen aan dat infecties met deze twee virussen meestal vóór het 4<sup>de</sup> levensjaar optreden. Hierna kunnen de virussen zich levenslang in het lichaam nestelen en voor aanhoudende infecties zorgen zonder dat er ziekte te constateren valt. Echter, bij transplantatiepatiënten (bij wie het afweersysteem bewust wordt verminderd om afstoting van het getransplanteerde orgaan te voorkomen), bij patiënten met AIDS (bij wie het afweersysteem verzwakt is door het HIV virus), of bij Multiple Sclerose patiënten die worden behandeld met Natalizumab (een afweermodulerend geneesmiddel), kan een reactivatie van BKPyV en JCPyV tot ernstige en soms levensbedreigende complicaties leiden. Dus een goed functionerend afweersysteem is een sterk wapen tegen polyomavirus reactivatie en het ontstaan van ziekte. Hierom worden de polyomavirussen ook wel opportunistisch genoemd.

De recente ontdekkingen van nieuwe humane polyomavirussen hebben ervoor gezorgd dat deze groep van DNA (tumor) virussen en hun potentiële bijdragen aan ziektes bij mensen, meer onder de aandacht is gekomen. In 2008, bijvoorbeeld, werd voor het eerst aangegeven dat MCPyV<sup>3</sup> polyomavirus de oorzaak is van de meerderheid van Merkelcelcarcinomen. Merkelcelcarcinomen zijn agressieve huidtumoren die een stijgende incidentie<sup>4</sup> vertonen als gevolg van bijvoorbeeld veroudering van mensen (bij wie het afweersysteem geleidelijk met de leeftijd is afgezwakt) of door de toegenomen frequentie aan orgaantransplantaties in de ontwikkelde landen. Het virus dat door ons werd ontdekt in 2010 binnen de groep van Dr. Feltkamp, was ook een polyomavirus dat betrekking had op een huidaandoening. Dit virus dat trichodysplasia spinulosa-geassocieerde polyomavirus (TSPyV) werd genoemd, gedraagt zich eveneens goedaardig bij een goed functionerend afweersysteem, maar kan ziekteverschijnselen veroorzaken bij patiënten met een ernstig verzwakt immu-

<sup>1</sup> Desoxyribonucleïnezuur, de blauwdruk voor virale eiwitten, ook wel erfelijk materiaal genoemd

<sup>2</sup> DNA bestaat uit een combinatie van vier verschillende nucleobasen: Adenine, Thymine, Guanine en Cytosine

<sup>3</sup> Een polyomavirus dat kwaadaardige veranderingen van Merkel cellen teweeg brengt

<sup>4</sup> Het aantal nieuwe gevallen per tijdseenheid, per aantal mensen binnen een bevolking

unsysteem. Deze aandoeningen veroorzaakt door deze twee nieuwe virussen zijn zeer zeldzaam.

In **Hoofdstuk 1** van dit proefschrift, werden de ontwikkelingen in het kader van recentelijk geïdentificeerde humane polyomavirussen samengevat, hetgeen de start betekende voor dit promotieonderzoek naar TSPyV infecties, pathogenese<sup>5</sup> en virusgastheer aanpassing en virusevolutie<sup>6</sup>. Vervolgens werd er in **Hoofdstuk 2** nader ingegaan op de klinische en virologische eigenschappen van TSPyV, waar ook epidemiologische, diagnostische en therapeutische aspecten van de infectie met TSPyV werden belicht. Kort samengevat, TSPyV is betrokken bij het ontstaan van een zeer zeldzame huidaandoening genaamd trichodysplasia spinulosa (TS). TS is een huidziekte van patiënten met een ernstig verzwakt afweersysteem, bij wie kenmerken van haarfollikel vergroting en harde keratotische huiduitstulpingen (vooral op het gezicht) waarneembaar zijn. Ongeveer 75% van de mensen zijn met TSPyV besmet op jonge leeftijd, maar slechts een heel klein percentage ervan vertoont de symptomen ervan na afweersysteem verzwakking.

In het onderzoek beschreven in **Hoofdstuk 3**, werd geprobeerd om de relatie tussen TSPyV infectie en de huidziekte TS uitvoeriger te beschrijven door een analyse van de aanwezigheid en de hoeveelheid van viraal DNA, en de precieze lokalisatie van TSPyV eiwitten in huidcellen van TS-patiënten. Daartoe is er gebruik gemaakt van gearchiveerde TS laesionale<sup>7</sup> en niet-laesionale huidbiopten. De verkregen resultaten toonden aan dat het detecteren van viraal DNA een nauw verband laat zien met de ziekte TS, aangezien 100% van de aangedane- en maar 2% van de controle-monsters een positieve testuitkomst boden. Kwantificering van TSPyV DNA liet hoge virale aanwezigheid zien in de laesionale huidbiopten, met enkele miljoenen virale DNA kopieën per cel. Ook de virale deeltjes (geanalyseerd door elektronenmicroscopie<sup>8</sup>) en eiwitexpressie (geanalyseerd door immunofluorescentie<sup>9</sup>) werden slechts in de laesionale huidbiopten aangetroffen, welk beperkt bleven tot bepaalde soort haarfollikel cellen. Beslissend voor dit hoofdstuk is het feit dat de aanwezigheid van veel TSPyV DNA, en de overvloedige expressie van TSPyV eiwit in de getroffen haarfollikel cellen, aantonen dat er een sterke relatie bestaat tussen een actieve TSPyV infectie en het ontstaan van deze ontsierende TS huidziekte.

In voorafgaande onderzoeken door anderen werd gesuggereerd dat in TS laesies de haarfollikel cellen sterke proliferatie<sup>10</sup> vertonen. Conform de mechanismen van andere polyomavirussen, zoals MCPyV die de veroorzaker is van Merkelcelcarcinomen, werd er in **Hoofdstuk 4** verondersteld dat ook TSPyV soortgelijke mechanismen hanteert om proliferatie te stimuleren.

---

<sup>5</sup> Bestuderen van ontstaan, ontwikkelen en verloop van een aandoening of ziekte

<sup>6</sup> De geleidelijke verandering van een soort of populatie door genetische overerving

<sup>7</sup> Weefsels die kenmerken van ziekte vertonen

<sup>8</sup> Een vorm van microscopie waarbij tot aan nanometer grootte gevisualiseerd kan worden

<sup>9</sup> Een methode waarbij virale eiwitten gevisualiseerd kunnen worden met kleurgekoppelde antilichamen

<sup>10</sup> Vermenigvuldiging of deling van cellen

eratie van haarfollikel cellen te induceren. Daarom hebben we TS laesies op verschillende cellulaire markers (zoals Ki-67), die een indicatie kunnen geven van celproliferatie, met behulp van immunofluorescentie onderzocht. Daarnaast hebben we naar de betrokkenheid van het (vermoedelijk) celtransformerende LT eiwit van TSPyV in dit proces gekeken. We toonden aan dat het expressie van TSPyV LT eiwit inderdaad samenkomt met de toegenomen proliferatie markers in haarfollikel cellen, ondanks de toegenomen cellulaire defensie eiwitten die dit proces van proliferatie juist tegen horen te gaan. Kortom, de verkregen resultaten uit dit hoofdstuk suggereren dat er een scenario mogelijk is waarbij TSPyV LT eiwit proliferatie induceert, mogelijk om een populatie van delende cellen te creëren die de vermenigvuldiging van het viraal DNA kan ondersteunen. Door het ontregelen van dit cellulaire proces door dit virus voor zijn eigen belang, ontstaan er bij TS patiënten de vaak waargenomen symptomen.

Een ander doel van dit promotieonderzoek was om meer inzicht te krijgen in virus-gastheer aanpassing en virusevolutie van dit snel groeiende *Polyomaviridae* familie. Daar toe hebben we in **Hoofdstuk 5** het genoom van alle polyomavirussen bij elkaar genomen en deze softwarematig geanalyseerd. Uit ons onderzoek bleek, wat tot dusver onbekend was voor de meeste polyomavirussen, dat het genoom van een grote groep polyomavirussen een extra coderend stuk bevat dat de aanmaak van MT<sup>11</sup> (of ALTO) eiwit verzorgt. Expressie van dit eiwit was tot nu toe alleen bekend bij twee knaagdier polyomavirussen. Nu blijkt dus dat dit coderende stuk geconserveerd is in het genoom van tweeëntwintig van de totaal vijfenvijftig polyomavirussen, waaronder twee humane soorten. Uit een modelberekening komt tot uiting dat een deel van dit eiwit hoogstwaarschijnlijk een belangrijke rol speelt (of heeft gespeeld) bij de aanpassing van het virus tot zijn gastheer. Dit nieuwe mechanisme genaamd COCO-VA, zoals beschreven in dit hoofdstuk, zou de aanpassing van deze polyomavirussen tot hun gastheren bevorderen en zou dus een belangrijke rol kunnen spelen in de evolutie van polyomavirussen.

In **Hoofdstuk 6** is er vervolgens onderzoek verricht naar de evolutie van TSPyV. Hier werden de door ons dertien nieuw verkregen TSPyV genomen uit **Hoofdstuk 3** geanalyseerd die afkomstig waren van TS patiënten over de hele wereld en die ongeveer 40% van alle gerapporteerde TS gevallen vertegenwoordigen. Uit onderlinge genoomvergelijkingen bleek dat deze TSPyV genomen een beperkte variatie vertonen. Toch konden we door die minieme variaties het geheel van TSPyV genomen fylogenetisch<sup>12</sup> in drie aparte groepen verdelen. Genoom variaties in eiwitcoderende stukken die voor eiwit veranderingen zouden kunnen zorgen, waren voornamelijk gesitueerd in het gedeelte dat codeert voor MT eiwit. Zoals uit het onderzoek in **Hoofdstuk 5** naar voren kwam, bleek ook in **Hoofdstuk 6** dat het MT eiwit van TSPyV betrokken is bij virusgastheer aanpassing en dat het een belangrijke rol

---

<sup>11</sup> Een eiwit dat tumoren indiceert door bijvoorbeeld groei factoren na te bootsen in experimentele muizen

<sup>12</sup> Een schema dat de evolutionaire geschiedenis of biologische verwantschap weergeeft

zou kunnen spelen in de evolutie van dit virus. Gecombineerde resultaten van beide studies tonen aan dat de evolutie van een groep polyomavirussen, inclusief TSPyV, hoogstwaarschijnlijk het beste te traceren valt via hun MT eiwit.

Tot slot, middels in dit proefschrift beschreven promotieonderzoek, hebben we veel inzichten kunnen verkrijgen met betrekking tot de prevalentie, pathogenese en evolutie van het polyomavirus TSPyV. Eventueel gecombineerd met toekomstige onderzoeken aan dit virus, draagt deze informatie substantieel bij aan ons kennis en begrip over deze ziekteverwekker. Ontdekkingen van nieuwe aanknopingspunten betreffend virusmechanismen die betrekking hebben tot ziekteprocessen, zullen ons in staat stellen om klinische aspecten beter te kunnen begrijpen. Deze aanknopingspunten zouden in latere studies ook gebruikt kunnen worden om antivirale middelen te ontwikkelen ter voorkoming of genezing van polyomavirus geassocieerde ziektes bij mensen.

---

<sup>13</sup> Aantal infectiegevallen op een specifiek moment in een specifiek populatie



## مختصر تیز دکترا برای مطالعه خوانندگان غیر علوم طبی

اعضای فامیل پولیوماویرییدی (*Polyomaviridae*)، که یکی از آنها در بنا و اساس این تحقیق تیز دکترا بوده، و عبارت از کوچکترین ویروسها هستند که اندازه آنها 45 نانومتر (nanometer) میباشد. ویروس های پولیوما در سراسر جهان موجود است و از آغاز سال 2007 به بعد 11 نوع آنان کشف گردیده اند. مواد ژنتیک این ویروسها DNA است که در حدود 5000 جوره اسیدی میباشد. این مواد ژنتیک دو نوع پروتئین را تدوین میکند که ممکن است در هر زمان بعد از ابتلا توضیح گردد. پروتئین های که بعد از ابتلا اولی بوجود می آیند، عبارت اند از تومور کوچک (ST)، تومور متوسط (MT) و تومور بزرگ (LT) که مجموع این پروتئین ها در دوام زندگی این ویروس ضروری هستند. در زمان بعدی ابتلا، پروتئین های بوجود می آیند که پوش ساختاری ویروس را تشکیل خواهند داد و عبارت اند از VP1, VP2 و VP3.

ویروسهای پولیوما به طور کلی ویروسهای عفونی سلیم هستند و در زمان مطالعه سیرولوژیک در 75٪ تا 100٪ افراد ابتلای آن دیده میشوند. دو پولیوما ویروسهای انسانی، که در سال 1971 کشف گردیدند، عبارت اند از BKPyV و JCPyV. مطالعات بالینی اپیدمیولوژیک نشان میدهد که ابتلا با این دو ویروس معمولاً قبل از سن چهار سالگی رخ میدهد. بعد از ابتلا، این ویروسها میتوانند در تمام عمر خود را در بدن انسان نگهدارند و باعث ابتلای مداوم شوند، بدون آنکه مستقیماً باعث بیماری گردند. با این حال، در بیماران ترانسپلانت شده عضوی (که سیستم معافیت بدن آنان به منظور جلوگیری از رد ترانسپلانت شده عضوی کاهش می یابد)، در بیماران مبتلا به AIDS (که در آنها سیستم معافیت بدن توسط ویروس HIV ضعیف گردیده)، و یا در بیماران مولتیپل اسکلروزیس تحت تداوی با دوی Natalizumab (که این یک دوا برای تضعیف سیستم معافیت بدن است)، فعال شدن مجدد BKPyV یا JCPyV ویروسها می توانند باعث عوارض جدی و گاهی تهدید کننده حیات شوند. بنابراین درستی عملکرد سیستم معافیت بدن یک سلاح قوی در برابر فعال شدن مجدد این ویروسها است که پیشگیری از شروع بیماری میکنند.

اکتشافات اخیر از ویروسهای پولیوما انسانی نشان داده شد که این گروه از ویروسهای توموری و توانایی های بالقوه آنان به منظور بیماری در انسان، باز هم مورد توجه تحقیقی بیشتر قرار گیرند. به طور مثال، در سال 2008 برای اولین بار نشان داده شد که ویروس پولیوما به نام MCPyV علت اکثر سرطانهای Merkel-cell-carcinoma است. این سرطانهای تهاجمی باعث بیماریهای جدی هستند که به طور مثال بنا به پیشرفت سن انسانها (در انهای که سیستم معافیتی بدن شان ضعیف شده)،

و یا به دلیل ترانسپلانت شده عضوی که باز هم سیستم معافیتی بدن ضعیف میشود. یک ویروس پولیومای دیگر که در سال 2010 کشف گردیده است در تحقیقات گروه مان بود (گروه دکتر Feltkamp) که ارتباط آن با یک بیماری جلدی دیگر میباشد. این ویروس که TSPyV نامیده میشود، نیز خاصیت عفونی سلیم دارد. چنانچه سیستم معافیتی بدن انسان به شدت ضعیف گردد، که باعث نشانه بیماری میشود. به هر حال، این بیماری های ناشی از این دو ویروس تازه کشف شده در تمام جهان بسیار نادر هستند.

در **فصل 1** این پایان نامه، اکتشاف مختصر در زمینه ویروسهای پولیومای تازه کشف شده انسانی بود که، آغاز این تحقیق دکتر در مورد عفونت، اسبب شناسی، سازگاری و تکامل ویروسی TSPyV گردید. پس از آن فصل، در **فصل 2** بحث در مورد ویژگی های کلینیک و ویروسی در مورد TSPyV و همچنین اپیدمیولوژی، تشخیص و تداوی عفونت این ویروس تعیین گردید. به طور مختصر از این دو فصل، TSPyV در تشکل یک بیماری بسیار نادر جلدی به نام Trichodysplasia-spinulosa (TS) به دست می آید. مانند یاد داشت قبل، TS یک بیماری جلدی افراد که سیستم معافیت بدن آنان به شدت تضعیف گردیده است، که آنها ویژگی های بزرگ شدن فولیکول و تفرن سخت مو (خصوصاً در صورت) را دارد. هنگامیکه DNA این ویروس مشاهده و اندازه گردید، معلوم شد که این ویروس در افراد سالم بسیار نادر است. اما زمانی که سرولوژی مورد بررسی قرار گرفت، معلوم شد که حدود 75٪ مردم از سن طفولیت به این ویروس مبتلا بوده اند.

در مطالعه که شرح داده شد در **فصل 3** این پایان نامه، تلاش گردید که رابطه عفونت با TSPyV و بیماری های جلدی TS توصیف گردد. این توصیف از طریق تجزیه و تحلیل موجودیت و مقدار DNA و محل دقیق پروتئین ویروسی در سلولهای جلدی ماوف TS بود. نتایجی به دست آمده این تحلیل نشان داد که تشخیص DNA ویروسی در 100٪ از نمونه مریضی TS و 2٪ از نمونه شاهد سالم، نتیجه مثبت داشتند. اندازه گیری مقدار موجودیت DNA این ویروس در نمونه مریضی TS نتیجتاً نشان داد که DNA ویروسی با چند میلیون دفعه در هر سلول یافت میشوند. همچنین، اعداد ویروسی (که توسط میکروسکوپ الکترونیک تحلیل و شناخته شده) و توصیف پروتئین ویروسی (که توسط تکنیک فلورسانس تحلیل و شناخته شده) تنها در بیوپسی جلدی TS دیده شدند که محدود به گونه های خاص سلولهای فولیکول مو بودند. برای **فصل 3** این واقعیتها قطعی استند که رابطه بسیار قوی بین عفونت به TSPyV و شروع این بیماری TS به شکل جلدی وجود دارد.

در مطالعات قبلی توسط بعضی ها پیشنهاد شد که در بیوپسی جلدی TS سلولهای فولیکولی مو تکثر قوی را نشان دهند. بنابراین به عنوان مکانیسم پولیوما ویروسهای دیگر، و به عنوان MCPyV

که علت است به Merkel-cell-carcinoma, همچنین در فصل 4 این پایان نامه در نظر گرفته شد که TSPyV با استفاده از مکانیسم مشابه، تکثیر سلولهای فولیکول مو تسریع میابد. به این دلیل ما بیوپسی جلدی TS را در شناخت پروتئینهای مختلف سلولی (مانند Ki-67) که ممکن است معلومات از تکثیر سلولها دهند. علاوه براین، ما باید دخالت پروتئین LT از TSPyV را در دوام این نقیصه سلولی تحلیل نمایم. ما توصیفاً یادآور شدیم که تعبیر این پروتئین LT در واقع افزایش یافته است و محلی با توصیف پروتئینهای مختلف سلولی در سلولهای فولیکول مو بود. این دوام را افزایش پروتئینی دفاع سلولی نمیتواند جلوگیری کند. به طور خلاصه، نتایج به دست آمده از این فصل نشان میدهد که ممکن است یک سناریوی که در آن موجودیت پروتئین LT از TSPyV باعث تکثیر سلولی میشود، که مجموعه این سلولها امکانات را برای تکثیر DNA ویروسی قادر بسازند. پس از ایجاد اختلال در دورنمای سلولی توسط این ویروس برای منافع خود، علائم TS در بیماران مشاهده میشوند.

هدف دیگر این تحقیق دکترا به دست آوردن دانش بیشتر در مورد تطابق تکامل اعضای فامیل پولیوماویریدی با انسان و حیوانات بود. برای این منظور، در فصل 5، ما مجموعه تمام ژنوم پولیوما ویروسها را توسط کامپیوتر تجزیه و تحلیل کردیم. چیزی که ناشناخته بود برای اکثر پولیوما ویروسها، در تحقیقات ما نشان داده شد که ژنوم یک گروه بزرگی از این ویروسها یک ژن زیادتر دارند که پروتئین MT یا ALTO را تدوین میکند. موجودیت پروتئین MT تا کنون تنها به دو ویروس حیوانات چونده شناخته شده بود. در حال حاضر، از تحقیق ما بر می آید که این ژن MT از مجموع 55 پولیوما ویروسها در ژنوم 22 آنها و از جمله در دو ویروس انسانی حفظ شده است. ضمن محاسبه طراحی کامپیوتری درک شد که یک بخشی این پروتئین MT بسیار احتمال دارد که در توافق ویروس با انسان یا حیوان نقش مهمی بازی کند. این مکانیزم جدید را که ما کشف کرده ایم به نام CoCo-VA یاد میکنیم. در نهایت، این مکانیسم میتواند در ترویج تکاملی پولیوما ویروس نقش مهمی را بازی کند.

در فصل 6 این پایان نامه، تحقیقات تنها در مورد مطالعه تکامل ویروس TSPyV بود. در این فصل، ما سیزده ژنومی جدید را به دست آوردیم که تجزیه و تحلیل شدند، و آینده به 40٪ از تمام موارد گزارش شده بیماران مبتلا به TS خواهند بود که از سراسر جهان میباشند. مقایسه متقابل نشان داد که ژنوم این TSPyV ویروس تنوع محدود را نشان می دهند. در عین حال، ما توانستیم که توسط این تغییرات حداقل تمامیت مواد ژنتیک TSPyV را به سه گروه تکامل نژادی مجزا تقسیم کنیم. تنوع ژنوم در مناطق تدوین گذاری پروتئین که به تغییراتی پروتئینی بتوانند نتیجه بدهند، به طور عمده در بخش پروتئین MT یا ALTO بودند. همانطور که در مطالعه فصل 5 درک شد، همچنین در مطالعه

فصل 6 به این نتیجه رسیدیم که MT از TSPyV مربوط به انطباق ویروس با انسان میباشد و بنابراین نقش مهمی در تکامل این ویروس دارد. نتایج ترکیبی از هر دو مطالعه نشان میدهند که تکامل گروه ویروسهای پولیوما، از جمله TSPyV، احتمالاً از طریق پروتئین MT ردیابی شوند.

نهایتاً، از طریق تیز دکترا که در این پایان نامه شرح داده شد، ما توانستیم در مورد شیوع، آسیب شناسی و تکاملی ژنتیک ویروس TSPyV بینش زیاد به دست آوریم. همراه با مطالعات آینده در مورد این ویروس، این اطلاعات قابل ملاحظه باعث درک ما در باره این پتوزن میشود. اکتشافات نوآوری جدید در نقاط وصل مکانیسم ویروسی که مربوط به دورنما های بیماری هستند، ما را قادر به درک جنبه های کلینیکی بهتر خواهد ساخت. این نقاط وصل مکانیک ویروسی، همچنان میتوانند در تحقیقات علمی آینده مورد استفاده قرار گیرند که به منظور توسعه ادویه ضد ویروسهای پولیوما برای وقایه یا درمان بیماری باشند.

VePyV1 CaPyV JCPyV *mPyV* SA12 RacPyV GHPyV BKPyV

*BatsgU* CPyV OrAPyV1 FPyV MptV APPyV1 SqPyV LPyV CSLPyV

EPyV PRPyV1 APP<sub>g</sub>V2 KIPyV **MIRgV** PtvPyV2c OtPyV1 **STLLPyV**

MFPyV1 *KSgU9* SV40 TSPyV CoPyV1 PPPyV CPPyV HPyV12

MXPyV PtvPyV1a PDPyV EIPyV1 AtPPyV1 HaPyV (TggPyV1

CdPyV DRPyV MWPyV APyV CAPyV1 HPyV7 CHPyV MasPyV

WUPyV *HgPyVc* BPyV MCPyV OrAPyV2 MMPyV SLPyV HPyV10

**VePyV1** *CaPyV* **JCPyV** *mPyV* **SA12**

*BatPyV* **CPyV** **OraPyV1** **FPyV** **MptV** **APPy**

**EPyV** **PRPyV1** **APPyV2** **KIPyV** **MiPyV** **P**

**MFPyV1** *HPyV9* **SV40** **TSPyV** **CoPyV1**

**MXPyV** **PtvPyV1a** **PDPyV** **EiPyV1** **At**

**CdPyV** **DRPyV** **MWPyV** **APyV** **CAPyV1**

**WUPyV** *HPyV6* **BPyV** **MCPyV** **OraPyV2**

## Part V

### Appendix

**yV** **BKPyV**

**V** **CSLPyV**

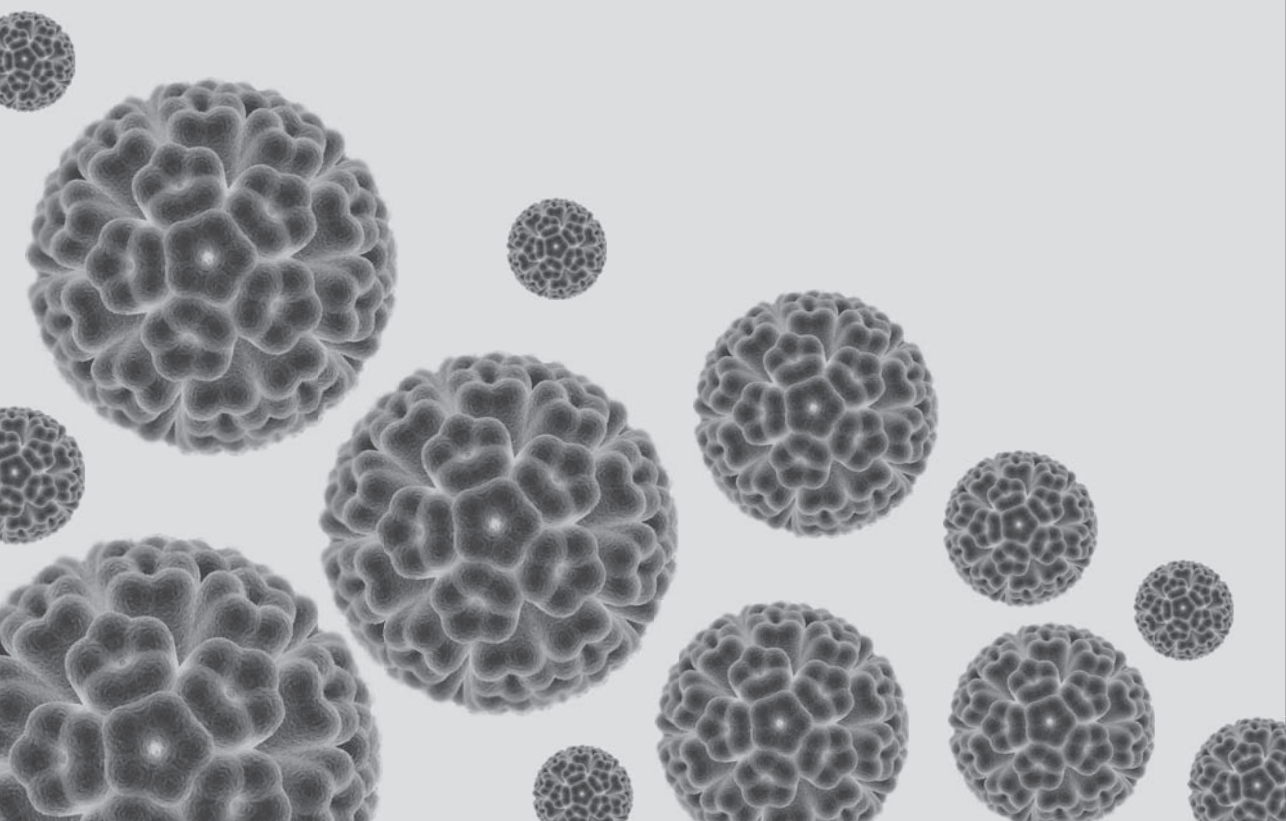
**V1** **STLPyV**

**V** **HPyV12**

**V** **GgPyV1**

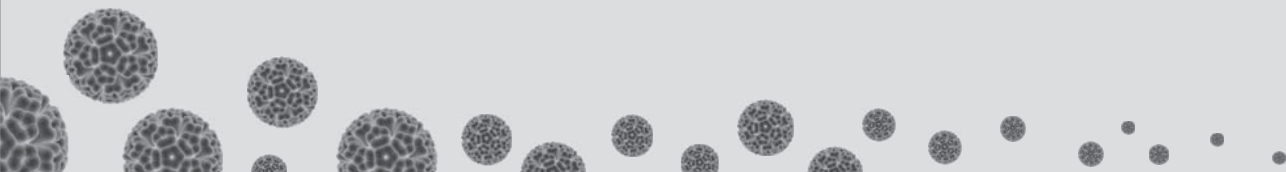
**V** **MasPyV**

**V** **HPyV10**



# Appendix 1

List of Abbreviations





## List of Abbreviations

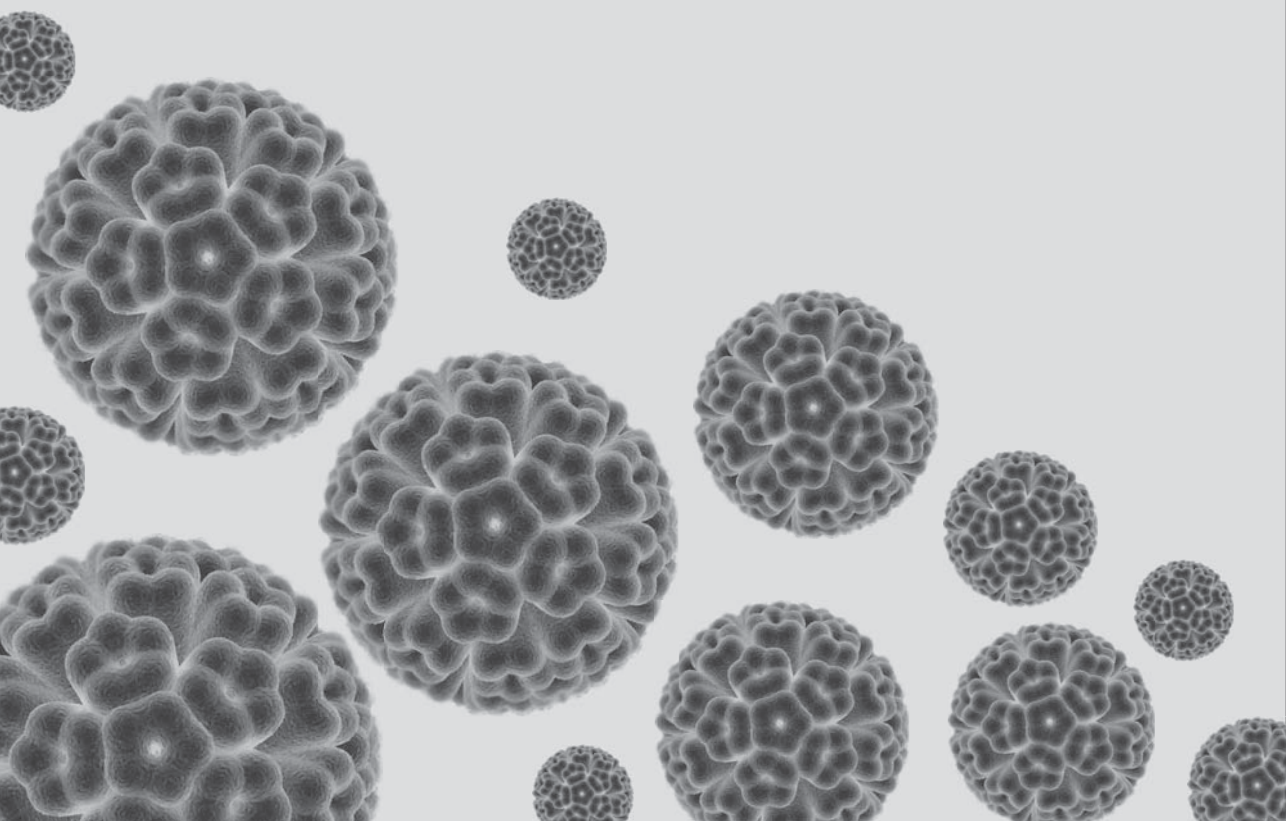
<b>A</b>	AelPyV	African elephant polyomavirus
	AI	Association index
	ALL	Acute lymphocytic leukemia
	ALTO	Alternate frame of the large T open reading frame
	APPyV1	Artibeus planirostris polyomavirus
	APPyV2	Artibeus planirostris polyomavirus
	APyV	Avian polyomavirus
	AtPPyV1	Ateles paniscus polyomavirus
<b>B</b>	BatPyV	Bat polyomavirus
	BKPyV	BK polyomavirus
	BPyV	Bovine polyomavirus
<b>C</b>	CAPyV1	Cebus albifrons polyomavirus
	CaPyV	Canary polyomavirus isolate
	CDK	Cyclin dependent kinases
	CdPyV	Cardioderma polyomavirus
	ChPyV	Chimpanzee polyomavirus
	CLL	Chronic lymphocytic leukemia
	COCO-VA	Codon-constrained Valine-Alanine
	CoPyV1	Chaerephon polyomavirus
	CPPyV	Carollia perspicillata polyomavirus
	CPyV	Crow polyomavirus
	CSLPyV	California sea lion polyomavirus
<b>D</b>	dsDNA	Double-stranded DNA
	DRPyV	Desmodus rotundus polyomavirus
	DS	Dataset
<b>E</b>	EiPyV1	Eidolon polyomavirus
	EM	Electron microscopy
	EPyV	Equine polyomavirus
<b>F</b>	FFPE	Formalin-fixed paraffin-embedded
	FPyV	Finch polyomavirus
	FRFR	Fresh frozen
	FUBAR	Fast unconstrained bayesian approximation

<b>G</b>	GE	Gastroenteritis
	GggPyV1	Gorilla gorilla gorilla polyomavirus
	GHPyV	Goose hemorrhagic polyomavirus
	GI-tract	Gastro-intestinal tract
	GST	Glutathione-S-transferase
<b>H</b>	HaPyV	Hamster polyomavirus
	H&E	Hematoxylin and eosin
	HPV	Human papillomavirus
	HPV16	Human papillomavirus type 16
	HPyV6	Human polyomavirus type 6
	HPyV7	Human polyomavirus type 7
	HPyV9	Human polyomavirus type 9
	HPyV10	Human polyomavirus type 10
	HPyV12	Human polyomavirus type 12
	HPyVs	Human polyomaviruses
HSF	Human Splice Finder	
<b>I</b>	ICTV	International committee on taxonomy of viruses
	IDRs	Intrinsically disordered regions
	IFA	Immunofluorescence assay
	IHC	Immunohistochemistry
	IR	Intergenic region
	IRB	Institutional review board
	IRS	Inner root sheath
<b>J</b>	JCPyV	JC polyomavirus
<b>K</b>	KIPyV	KI polyomavirus
<b>L</b>	LGN	Lupus glomerulonephritis
	LPyV	B-lymphotropic polyomavirus
	LT-antigen	Large tumor antigen
<b>M</b>	MasPyV	Mastomys polyomavirus
	MC	Maximum monophyletic clade size
	MCC	Merkel cell carcinoma
	MCMC	Markov chain monte carlo
	MCPyV	Merkel cell polyomavirus
	MDA	Multiple displacement amplification

	MFPyV1	Macaca fascicularis polyomavirus
	MiPyV	Miniopterus polyomavirus
	M&M	Material and Methods
	MMPyV	Molossus molossus polyomavirus
	MptV	Murine pneumotropic virus
	MPyV	Murine polyomavirus
	MT-antigen	Middle tumor antigen
	MWPyV	Malawi polyomavirus
	MXPyV	Mexico polyomavirus
<b>N</b>	NCCR	Non-coding control region
	NHL	Non-Hodgkin's lymphoma
	NJPyV	New Jersey polyomavirus
	NLS	Non-lesional skin
	nts	Nucleotides
<b>O</b>	OPTN	Organ procurement and transplantation network
	OraPyV1	Orang-utan polyomavirus Bornean
	OraPyV2	Orang-utan polyomavirus Sumatran
	ORF	Open reading frame
	Ortho-I	Orthopolyomavirus-I
	Ortho-II	Orthopolyomavirus-II
	OtPyV1	Otomops polyomavirus
<b>P</b>	PDPyV	Pteronotus davi polyomavirus
	PED	Pair-wise evolutionary distance
	PHK	Primary human keratinocytes
	PML	Progressive multifocal leukoencephalopathy
	PPPyV	Pteronotus parnellii polyomavirus
	PRPyV1	Piliocolobus rufomitratu polyomavirus
	PS	Parsimony score
	PtvPyV1a	Pan troglodytes verus polyomavirus
	PtvPyV2c	Pan troglodytes verus polyomavirus
<b>R</b>	RacPyV	Raccoon polyomavirus
	RB	Retinoblastoma
	RCA	Rolling-circle amplification
	RefSeq	Reference sequence
	REL	Random effect likelihood
	RF	Reading frames
	RTR	Renal transplant recipients

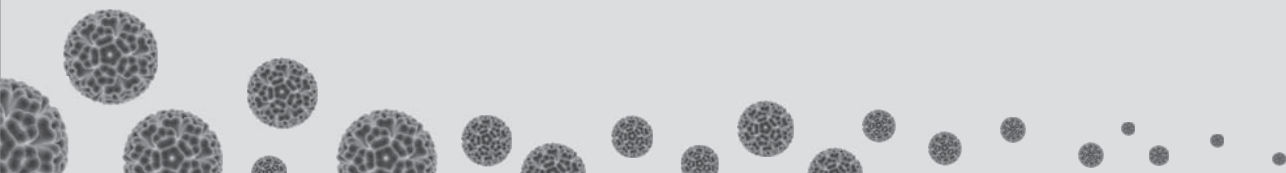
---

<b>S</b>	SA12	Baboon polyomavirus 1
	SLiMs	Short linear motifs
	SLPyV	Sturnira lilium polyomavirus
	SNP	Single (di)nucleotide polymorphism
	SPED	Smallest pair-wise evolutionary distance
	SqPyV	Squirrel monkey polyomavirus
	ST-antigen	Small tumor antigen
	STLPyV	St. Louis polyomavirus
SV40	Simian virus 40	
<b>T</b>	TCHH	Trichohyalin
	TM	Transmembrane domain
	tMRCA	Time to the most recent common ancestor
	TOI	Trichodysplasia of immunosuppression
	TP	True-palindrome
	TS	Trichodysplasia spinulosa
	TSPyV	Trichodysplasia spinulosa-associated polyomavirus
	TX	Transplant patient
<b>V</b>	VATD	Viral-associated trichodysplasia of immunosuppression
	VATS	Virus associated trichodysplasia spinulosa
	VePyV1	Vervet monkey polyomavirus
	VP1	Viral capsid protein 1
	VP2	Viral capsid protein 2
	VP3	Viral capsid protein 3
<b>W</b>	WUPyV	WU polyomavirus
<b>Y</b>	ybp	Years before present



# Appendix 2

Acknowledgement



## Acknowledgements

This dissertation would obviously have not been realized without the many contributions of my fellow researchers **Els, Chris, Sander, Richard, Alexander K., Arlene, Philip, Klaus, Elena, Taylor, Lauren, Esther, John, Just, Genevieve, Louis, Sasha** and **Mariet**, who co-authored with me the published papers because of their sample collection/collaboration, technical assistance, data analyses and/or manuscript writing and editing. In addition, I am also much gratitude indebted to my two promotors, Prof. dr. **Alexander Gorbalenya** and Prof. dr. **Louis Kroes**, for the time and continuous support they have invested in me and for making this PhD candidacy possible at the department of Medical Microbiology of the Leiden University Medical Center.

My co-promotor Dr. **Mariet Feltkamp** is sincerely thanked for the freedom she gave me to explore and pursue ideas and experiments of my own, but still be critically involved, despite her activities at the department as an associate professor and clinical virologist. She always remained accessible for questions and help. After all, it was her enthusiasm, openness and clarity that ensured me to develop a preference for DNA virus research and enrolled for this PhD candidacy.

I also owe my (past-)colleagues at the department of Medical Microbiology a lot of admiration, even if it was for the many great tips, chats, notes and ideas we shared. Indispensable were also the-always-helpful secretaries and the lab managers. **Many thanks to all of you!** My gratitude will not end with these words. I will thank you repeatedly in person when we meet again.

My (present-)colleagues at the **DDL Diagnostic Laboratory** are thanked for their support and help to let me adjust to a new research environment. Dr. **Wim Quint** is especially thanked for making my current position at the DDL possible.

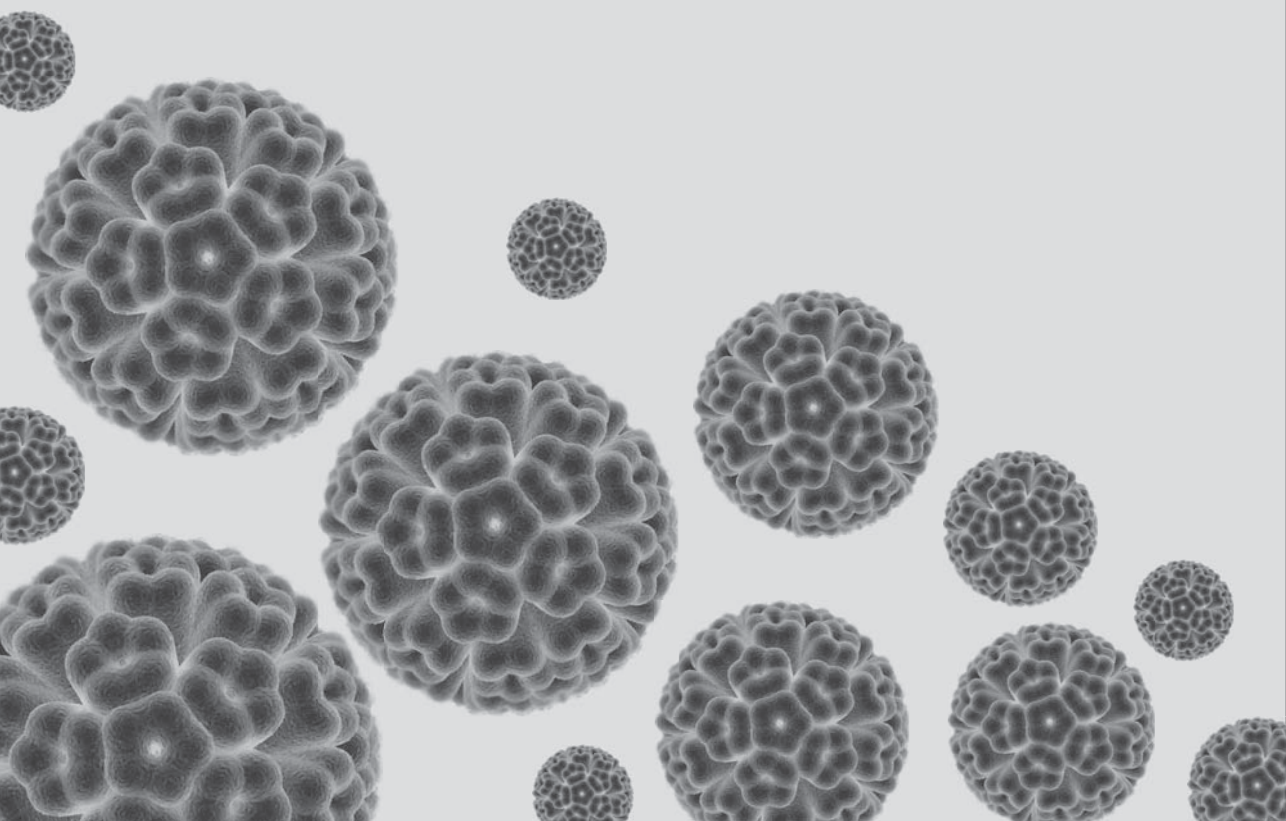
I am all my **friends** and **family** thankful for the pleasant moments that make life enjoyable. Your unconditional support and advice have always motivated me in my career choices.

In this part, first I would like to express my sincere gratitude towards my lovely parents. Words are powerless to express my feeling for all the care and love the two of you (**Mami** and **Baba**) have given us “The Four Brothers”! Even if it meant to risk everything you had and owned, just to create and ensure a perfect environment for us to live in. You succeeded! I wish I could do this much for the two of you. You are the core of my existence. My lovely and always supportive brother **Rosbe (Rob)** is thanked unconditionally for his support and helpfulness in any way. Big-bro, you are and will always be an excellent example in my life.

**Babak**, it is hard to find words to express how much I miss you my lovely brother. I wish we had more time to spend together and share thoughts about science. More importantly, to play again many matches of football together, the way you always won. The memories of you will always be with me, and your place in my hearth will be kept warm! Lil-bro **Dariush**, thanks for being such a wise and great brother by always helping others. Seeing your smile makes always my day. Great to always have your support at my side. In addition, special thanks are reserved for my **in-laws** for all the nice family occasions we had together and the discussions we shared about science and technology in relation to human believe in which everyone is always fully interested. Thank you for accepting me in the family!

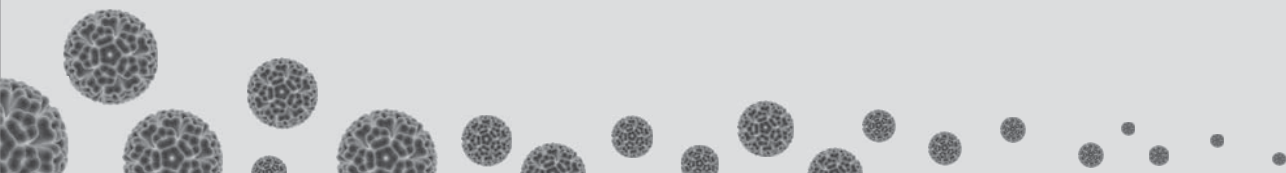
**Beheshta (BeBe)**, my love, my better half! The moment is here. We have finalized it! I am grateful for all your support in the past years. Thank you for being at my side in the moments of happiness and in the moments of sadness. Without you, there would be less love and happiness in my life. I am also grateful for all the beautiful moments we shared and I am already thrilled about the many many many more magnificent moments that will follow. With you by my side, we can conquer the world. I Love You!





# Appendix 3

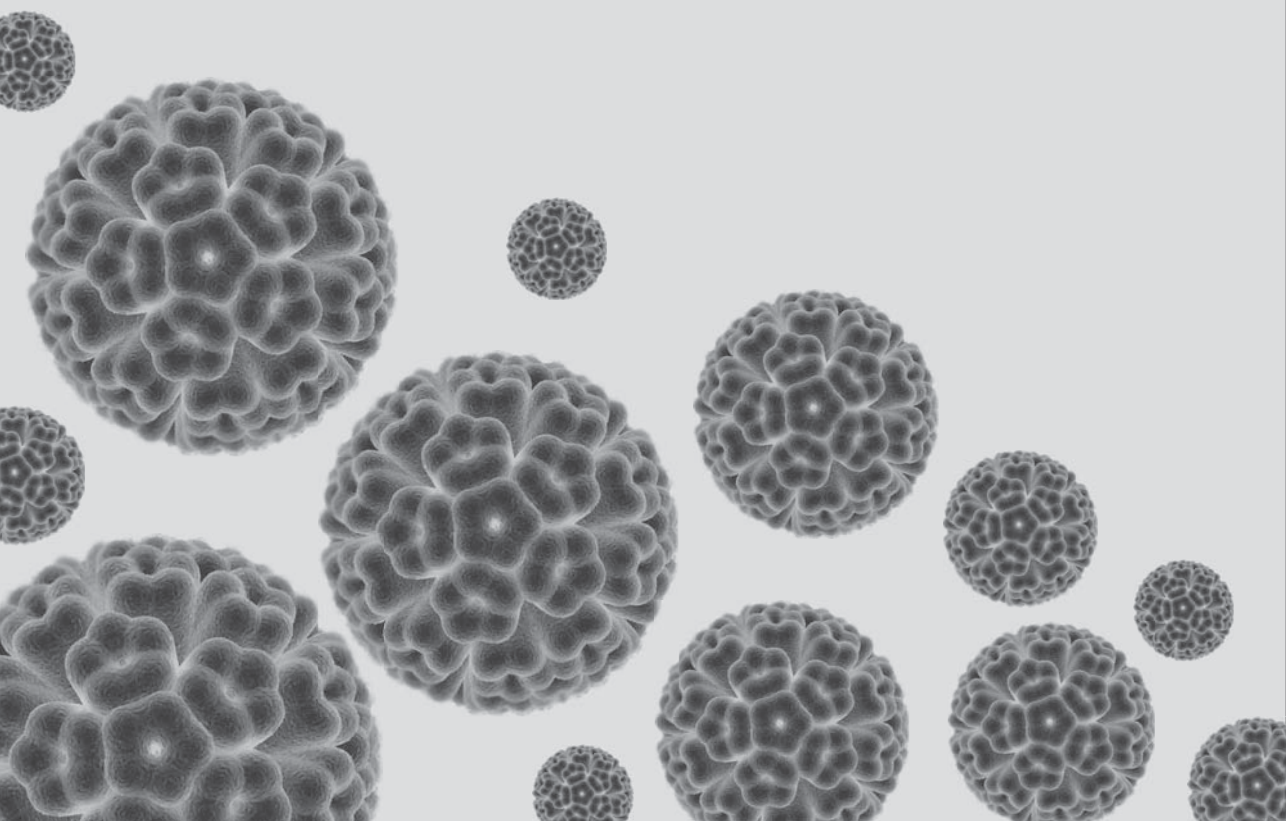
Curriculum Vitae





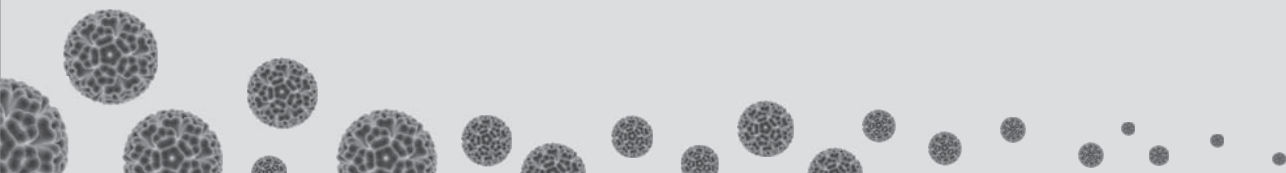
## Curriculum Vitae

**S**iamaque Kazem was born on December 23, 1982 in Kabul, Afghanistan. At the age of twelve, he arrived with his family in The Netherlands, embraced the opportunity to study and obtained his high school diploma at the Atlas College OSG West-Friesland in Hoorn, in the year 2000. One year later, he was enrolled for his Bachelor of Science (BSc) degree at the University of Applied Sciences (HSL) in Leiden, The Netherlands, where he became committed to specialize in the field of Medical Microbiology. In the final year before obtaining his BSc degree, he was trained for almost one year as an intern under supervision of Dr. Mariet Feltkamp and Dr. Linda Struijk working on the human papillomaviruses in relation to skin cancer at the Leiden University Medical Center department of Medical Microbiology. Upon graduating *cum laude* (with honor) for his BSc degree in 2006, in the same year he was admitted for his Master of Science (MSc) degree at the VU University Amsterdam aimed to specialize in (molecular) Oncology. As part of his MSc curriculum, for his minor-placement he was trained as an intern under supervision of Dr. David Noske, Dr. Gerrit Schuurhuis and Dr. Jacqueline Cloos working on Glioblastoma multiforme tumors at the VU University Medical Center departments of Neurosurgery, Hematology and Child oncology. For his major-placement, he was trained as an intern under supervision of Prof. dr. Paul van Diest and Dr. Petra van der Groep working again on brain tumors, now at the University Medical Center Utrecht department of Pathology. He graduated for his MSc degree in 2008 and in the same year, he joined again the group of Dr. Mariet Feltkamp as a junior researcher working on the human papillomaviruses in relation to skin cancer. Upon discovery of the polyomavirus TSPyV in 2010, in the same year he started his PhD studies on this virus. To obtain his PhD degree at the University of Leiden under supervision of Dr. Mariet Feltkamp, Prof. dr. Alexander Gorbalenya and Prof. dr. Louis Kroes at the Leiden University Medical Center department of Medical Microbiology section Experimental Virology, the results of his studies are put together in this dissertation. Since December 2014, Siamaque Kazem is working as a Postdoctoral Scientist in the Research and Development department of DDL Diagnostic Laboratory in Rijswijk (The Netherlands), to continue his career in the field of Molecular Oncology and Next Generation Sequencing.



# Appendix 4

List of Publications



## List of Publications

1. Van der Meijden E, **Kazem S**, Dargel CA, van Vuren N, Hensbergen PJ, Feltkamp MCW. Characterization of T-antigens, including Middle T and ALTO, expressed by the human polyomavirus associated with trichodysplasia spinulosa.  
*Submitted for publication*
2. **Kazem S**,\* Lauber C,\* van der Meijden E, Kooijman S, Kravchenko AA, Feltkamp MCW# and Gorbalenya AE.#  
Adaptation of trichodysplasia spinulosa-associated polyomavirus to the human population is mediated by middle T antigen and involves COCO-VA toggling.  
*Submitted for publication*  
*\*and # Authors than contributed equally*
3. Lauber C,\* **Kazem S**,\* Kravchenko AA, Feltkamp MCW and Gorbalenya AE.  
Interspecific adaptation by binary choice at de novo polyomavirus T antigen site through accelerated codon-constrained Val-Ala toggling within an intrinsically disordered region.  
*Nucleic Acids Research (10.1093/nar/gkv378), 2015*  
*\*Both authors contributed equally*
4. **Kazem S**, van der Meijden E, Wang RC, Rosenberg AS, Pope E, Benoit T, Fleckman P and Feltkamp MCW.  
Polyomavirus-associated trichodysplasia spinulosa involves hyperproliferation, phosphorylation of pRB and upregulation of p16 and p21.  
*PLoS ONE (9: e108947), 2014*
5. **Kazem S**, van der Meijden E and Feltkamp MCW.  
The trichodysplasia spinulosa-associated polyomavirus: virological background and clinical implications.  
*APMIS (121: 770 - 782), 2013*
6. Feltkamp MCW, **Kazem S**, van der Meijden E, Lauber C and Gorbalenya AE.  
From Stockholm to Malawi: recent developments in studying human polyomaviruses.  
*Journal of General Virology (94: 482 - 496), 2013*
7. **Kazem S**, van der Meijden E, Kooijman S, Rosenberg AS, Hughey LC, Browning JC, Sadler G, Busam K, Pope E, Benoit T, Fleckman P, de Vries E, Eekhof JA and Feltkamp MCW.  
Trichodysplasia spinulosa is characterized by active polyomavirus infection.  
*Journal of Clinical Virology (53: 225 - 230), 2012*
8. **Kazem S**, van der Meijden E, Struijk L, de Gruijl FR and Feltkamp MCW.  
Human papillomavirus 8 E6 disrupts terminal skin differentiation and prevents pro-Caspase-14 cleavage.  
*Virus Research (163: 609-616), 2012*

9. Lazic D, Hufbauer M, Zigrino P, Buchholz S, **Kazem S**, Feltkamp MCW, Mauch C, Steger G, Pfister H and Akgül B.  
Human Papillomavirus Type 8 E6 Oncogene Inhibits Transcription of the PDZ Protein Syntenin-2.  
*Journal of Virology* (86: 7943-7952), 2012
10. Kanitakis J, **Kazem S**, van der Meijden E and Feltkamp MCW.  
Absence of the trichodysplasia spinulosa-associated polyomavirus in human pilomatrixomas.  
*European Journal of Dermatology* (21: 453-454), 2011
11. Van der Meijden E, **Kazem S**, Burgers MM, Janssens R, Bouwes Bavinck JN, de Melker H and Feltkamp MCW.  
Seroprevalence of Trichodysplasia spinulosa-associated Polyomavirus.  
*Emerging Infectious Diseases* (17: 1355-1363), 2011
12. Struijk L, van der Meijden E, **Kazem S**, ter Schegget J, de Gruijl FR, Steenbergen RD and Feltkamp MCW.  
Specific betapapillomaviruses associated with squamous cell carcinoma of the skin inhibit UVB-induced apoptosis of primary human keratinocytes.  
*Journal of General Virology* (89: 2303-2314), 2008





*Fly me to the moon  
Let me play among the stars  
Let me see what spring is like on  
Jupiter and Mars*

*In other words, hold my hand  
In other words, baby, kiss me*

*Fill my heart with song  
And let me sing forever more  
You are all I long for  
All I worship and adore*

*In other words, please be true  
In other words, I love you ~ BeBe ~*

*Frank Sinatra (1964)*

