Expression

6

Els van der Meijden
Siamaque Kazem
Christina Dargel
Nick van Vuren
Paul Hensbergen
Mariet Feltkamp

Journal of Virology 2015;89(18):9427-9439

CHARACTERIZATION OF T-ANTIGENS, INCLUDING MIDDLE T AND ALTERNATIVE T, EXPRESSED BY THE HUMAN POLYOMA-VIRUS ASSOCIATED WITH TRICHODYSPLASIA SPINULOSA

Abstract

The polyomavirus tumor (T) antigens play crucial roles in viral replication, transcription, and cellular transformation. They are encoded by partially overlapping open reading frames (ORFs) located in the early region through alternative mRNA splicing. The T expression pattern of the trichodysplasia spinulosa-associated polyomavirus (TSPyV) has not been established yet, hampering further study of its pathogenic mechanisms and taxonomic relationship. Here, we characterized TSPyV T-antigen expression in human cell lines transfected with the TSPyV early region. Sequencing of T-antigen-encoded reverse transcription-PCR (RT-PCR) products revealed three splice donor and acceptor sites creating six mRNA splice products that potentially encode the antigens small T (ST), middle T (MT), large T (LT), tiny T, 21kT, and alternative T (ALTO). Except for 21kT, these splice products were also detected in skin of TSPyV-infected patients. At least three splice products were confirmed by Northern blotting, likely encoding LT, MT, ST, 21kT, and ALTO. Protein expression was demonstrated for LT, ALTO, and possibly MT, with LT detected in the nucleus and ALTO in the cytoplasm of transfected cells. Splice site and start codon mutations indicated that ALTO is encoded by the same splice product that encodes LT and uses internal start codons for initiation. The genuineness of ALTO was indicated by the identification of acetylated N-terminal ALTO peptides by mass spectrometry. Summarizing, TSPyV exhibits an expression pattern characterized by both MT and ALTO expression, combining features of rodent and human polyomaviruses. This unique expression pattern provides important leads for further study of polyomavirus-related disease and for an understanding of polyomavirus evolution.

Importance

The human trichodysplasia spinulosa-associated polyomavirus (TSPyV) is distinguished among polyomaviruses for combining productive infection with cell-transforming properties. In the research presented here, we further substantiate this unique position by indicating expression of both middle T-antigen (MT) and alternative T-antigen (ALTO) in TSPyV. So far, none of the human polyomaviruses was shown to express MT, which is considered the most important viral oncoprotein of rodent polyomaviruses. Coexpression of ALTO and MT, which involves a conserved, recently recognized overlapping ORF subject to positive selection, has not been observed before for any polyomavirus. As a result of our findings, this study provides valuable new insights into polyomavirus T gene use and expression. Obviously, these insights will be instrumental in further study and gaining an understanding of TSPyV pathogenicity. More importantly, however, they provide important leads with regard to the interrelationship, functionality, and evolution of polyomaviruses as a whole, indicating that TSPyV is a suitable model virus to study these entities further.

Introduction

Human polyomaviruses represent a rapidly expanding group of small, circular double-stranded DNA viruses that persistently infect the general population, usually without causing symptoms.[48,50,61,62,192,214] Seven of 13 described human polyomaviruses have been associated with disease in immunocompromised individuals, i.e., JC polyomavirus (JCPyV),[207] BK polyomavirus (BKPyV),[215] Merkel cell polyomavirus (MCPyV),[59] trichodysplasia spinulosa-associated polyomavirus (TSPyV),[28,65] human polyomavirus 6 (HPyV6),[124] human polyomavirus 7 (HPyV7),[63] and New Jersey polyomavirus (NJPyV).[64,80] MCPyV and TSPyV, which belong to *orthopolyomavirus lineage I*,[44,192] are associated with a malignant and a benign hyperproliferative skin disease called Merkel cell carcinoma (MCC) and trichodysplasia spinulosa (TS), respectively. The hyperproliferative phenotype of both diseases indicates involvement of the viral tumor (T) antigens in the patho(onco) genesis.

T-antigens that play a coordinating role in viral transcription and replication are generally known for their ability to disrupt cellular pathways involved in cell cycle regulation and signaling.[109,198] The large T-antigen (LT), for example, promotes cell cycle entry through inactivation (hyperphosphorylation) of pRB, as demonstrated, for example, for simian virus 40 (SV40).[109,115] Moreover, it disrupts cell cycle control by hampering DNA repair and apoptosis pathways, for example, through inactivation of p53.[109,115] The small T-antigen

(ST) can bind protein phosphatase 2A (PP2A) and deregulate cellular pathways that include c-myc, phosphatidylinositol 3-kinase (PI3K), Akt, Rac, mitogen-activated protein kinase (MAPK), and 4E-binding protein 1 (4E-BP1),[198,216] thereby potentially inducing cellular transformation. The membrane-associated middle T-antigen (MT) mimics activated transmembrane growth factor receptors with associated kinase activity (MAPK and PI3K) and, as such, contributes to cellular transformation, as shown for murine polyomavirus (MPyV), for example.[115-117]

For MCPyV, the role of the T-antigens in MCC pathogenesis has been largely resolved.[122,144,216,217] For TSPyV and TS, this piece of information is still lacking although we recently provided evidence of involvement of LT in the induction of pRB hyper-phosphorylation and hyperproliferation of follicular skin cells,[31] and Tyring and coworkers showed PP2A binding and hyperphosphorylation of cellular factors involved in the MAPK pathway by TSPyV ST.[218,219] In order to study the role of TSPyV in pathogenesis and in cellular transformation in more detail and to position TSPyV among its virus family members, further knowledge is required regarding the presence and pattern of expression of the TSPyV-encoded T-antigens. To this end, we study and describe TSPyV T-antigen expression here in detail.

T-antigen expression involves generation of different mRNA products encoded by at least two open reading frames (ORFs) through alternative splicing of a common pre-mRNA primary T-antigen transcript. Expression of several T-antigens has been demonstrated, including LT, MT, ST, 17kT, TruncT, T=165, T=136, T=135, 57kT, and tiny T.[87-90,92] Detailed genome analysis recently revealed that, in addition to the known ORFs putatively encoding ST and LT (ORF1 and ORF2) and the structural proteins VP2/3 and VP1 (ORF3 and ORF4), TSPyV contains a fifth ORF (ORF5) (**FIGURE 1A**),[220] which is present in most lineage I *orthopolyomaviruses*.[105] In rodent polyomaviruses ORF5 encodes the second exon of MT,[221,222] in MCPyV it encodes the alternative T-antigen (ALTO).[93]

Based on the TSPyV ORF organization shown in **FIGURE 1A** and the T-antigen expression described for other polyomaviruses and summarized in **FIGURE 1B**, we hypothesized that TSPyV expresses at least ST and LT and probably MT and/or ALTO as well. To study this hypothesis, in the absence of a TSPyV virus culturing system, we characterized human cell lines expressing the entire TSPyV T-antigen coding region and TS clinical samples with respect to T mRNA and protein expression using reverse transcription-PCR (RT-PCR), Northern blotting, Western blotting, immunoprecipitation, immunohistochemistry, and/or immunofluorescence. In this way we could confirm the presence of mRNAs encoding TSPyV ST, MT, LT, 21kT, tiny T, and ALTO and of LT, ALTO, and possibly MT proteins. Furthermore, we could substantiate the transcript used for ALTO expression and its internal initiation site.

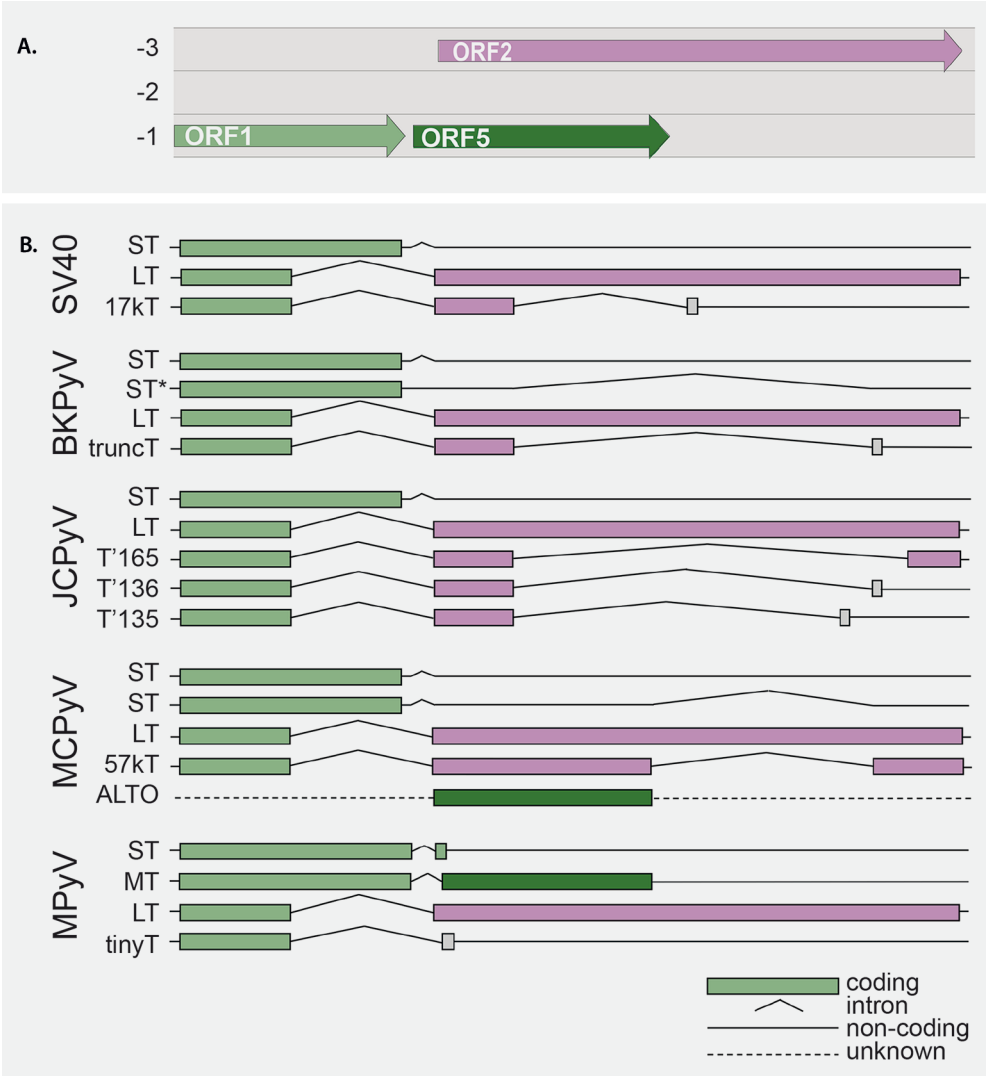


FIGURE 1. POLYOMAVIRUS EARLY-REGION ORGANIZATION AND MAIN SPLICE PRODUCTS.
A. Early region of the TSPyV genome. The three largest ORFs found in this region, on the genome minus strand, are defined between two stop codons and shown in color. -1, -2, - 3 indicate the reading frames. **B.** Overview of described T splice products of the human and animal polyomaviruses SV40, BKPyV, JCPyV, MCPyV, and MPyV. The colored rectangles refer to the distinct coding areas that likely correspond to the TSPyV ORFs shown in panel A.

Materials and methods

Constructs and site-directed mutagenesis.

For the characterization of TSPyV T mRNA and proteins, the complete early region (nucleotides [nt] 5046 to 2477) was cloned into the mammalian expression plasmid pcDNA3 (Invitrogen). The vector's leader and poly(A) termination signal were used for mRNA expression. For additional studies on ALTO expression, ORF5 (nt 4385 to 3985) of the TSPyV genome was cloned into pcDNA3. This was achieved by performing PCRs with pUC19-TSPyV containing the complete genome of TSPyV (GenBank accession number NC_014361)[28] as input material with primers TSPyV001 and TSPyV002 and primers TSPyV141 and TSPyV143, respectively, for the complete early region and ORF5. The PCR products were subcloned by Topo TA cloning (Invitrogen) and, after sequence confirmation, were subsequently cloned into the pcDNA3 plasmid by using BamHI and NotI restriction sites, resulting in the pcDNA3-TSPyV T and pcDNA3-TSPyV ORF5 constructs. The pcDNA3-TSPyV T construct was also used for the generation of splice acceptor mutants and for start codon mutants by performing QuikChange site-directed mutagenesis according to the manufacturer's instructions (Stratagene).

The retroviral expression construct pLZRS containing the early region (nt 5031 to 2528) was produced by Topo TA cloning of the PCR product which was generated by using primers TSPyV045 and TSPyV054 with pUC19-TSPyV as input DNA, followed by pLZRS cloning within the XhoI and NotI restriction sites (pLZRS-TSPyV T). Sequences of primers used for cloning, mutagenesis, and RT-PCR are available upon request.

Cell culture and clinical samples.

The human cervical carcinoma cell line HeLa (ATCC) and the human embryonic kidney cell line 293T were grown in Dulbecco's modified Eagle's medium (Lonza) supplemented with 10% fetal bovine serum (Bodinco BV), 100 units/ml of penicillin and streptomycin (Lonza), and 2 mM L-glutamine (Lonza). Most experiments were performed in both cell lines with similar results. Only the results obtained with HeLa are shown. Previously described lesional TS biopsy specimens, stored at -80°C, were used for RT-PCR and immunohistochemistry.[65]

Transfection and transduction.

For expression of TSPyV mRNA and proteins, cells were seeded into six-well plates and transiently transfected at 70 to 90% confluence with either pcDNA3-TSPyV T, pcDNA3-TSPyV ORF5, or empty pcDNA3 plasmid as a control. Lipofectamine 2000 was used as a transfection reagent (Invitrogen) according to the manufacturer's instructions, with a DNA-to-transfection reagent ratio of 1:2.5. Transfection complexes were removed at 4 to 6 h after transfection, and fresh medium was added to the cells. Total RNA and protein lysates

were isolated 18 to 24 h after transfection. In addition, a retroviral expression system was used for the expression of TSPyV T-antigens. Transduction was performed as described by Struijk et al..[223] In brief, retrovirus containing supernatant of Phoenix packaging cells transfected with either pLZRS-TSPyV T or empty pLZRS plasmid as a control was added to 30% confluent cells. Two rounds of transduction were performed on consecutive days, followed by at least 1 week of G418 selection before protein lysates were isolated from the transduced cells.

RNA and protein isolation.

Total cellular RNA was isolated from transfected cells using TRIzol reagent (Invitrogen) according to the manufacturer's instructions. From skin biopsy specimens of TSPyV DNA-positive TS patients, total RNA was isolated (~10 mg) by cutting the fresh-frozen biopsy specimen into small pieces and adding 100 µl of RNase free MilliQ; the isolation was performed with a QIAamp Viral RNA minikit (Qiagen) according to the manufacturer's instructions with a few adjustments. Lysis in AVL buffer (Qiagen), containing 10 ng/µl carrier RNA, was performed for 2 h at room temperature, and RNA was finally eluted from the column with 40 µl of AVE elution buffer (Qiagen). The RNA sample was stored at -80°C. Cellular proteins were isolated from transfected cells, grown in a well of a six-well plate, by lysing the cells in 300 µl of radioimmunoprecipitation assay (RIPA) buffer (1% NP-40, 0.5% sodium deoxycholate, and 0.1% SDS in phosphate-buffered saline [PBS]) supplemented with protease inhibitors (1 x complete EDTA-free protease inhibitor cocktail; Roche). Cell lysates were stored at -80°C.

Reverse transcriptase PCR (RT-PCR).

For identification of splice donor and acceptor sites in the TSPyV early region, several RT-PCRs were performed. One microgram of total RNA was treated with 1 U of DNase I (Ambion) for 20 min at 37°C, followed by inhibition of the enzyme by addition of 1 µl of EDTA (25 mM) and incubation at 65°C for 15 min. cDNA was synthesized by RT-PCR using Moloney murine leukemia virus reverse transcriptase (Fermentas) and random hexamer primers at 42°C for 1 h, followed by specific TSPyV PCRs with different primer sets under the following conditions. The 50-µl PCR mixture consisted of 1 x GeneAmp PCR buffer (15 mM Tris-HCl [pH 8,0], 50 mM KCl, 3.6 mM MgCl), 0.3 mM each deoxynucleoside triphosphate (dNTP), 15 pmol of each primer, and 2 U of AmpliTaq Gold polymerase (Applied Biosystems); cycle conditions were 5 min at 95°C, followed by 40 cycles of amplification (94°C for 1 min and 65°C for 1 min.). In addition to the reverse transcriptase reaction (+RT), in each RT-PCR the following controls were used: without reverse transcriptase (-RT), H_2O control (H_2O), and TSPyV plasmid DNA control (+C). A 1-kb Plus DNA Ladder (Life Technologies) was included in the gel electrophoresis experiments for interpretation of the RT-PCR results and consisted of the following bands: 100, 200, 300, 400, 500, 650, 850, 1,000, 1,650, 2,000,

3,000, 4,000, 5,000, 6,000, 7,000, 8,000, 9,000, 10,000, 11,000, and 12,000 bp. The RT-PCR products were cloned and subsequently sequenced.

Northern blotting and hybridization.

TSPyV early mRNA products were confirmed by Northern blotting. Samples containing 10 to 15 μg of total RNA of transfected cells were denatured by formaldehyde, electrophoresed on 1% formaldehyde-agarose gels, and transferred onto Hybond-N+ membranes (GE Healthcare) using 10 x SSC (1 x SSC is 0.15 M NaCl plus 0.015 M sodium citrate). TSPyV-specific primers located on different positions in the early region (see **FIGURE 3B**) were used to generate [γ-^{32}P]ATP-labeled probes using T4 polynucleotide kinase (Invitrogen). The following probes were used: probe A (nt 4409 to 4434), 5′-GTACATAACACCTGCAGTTTCCCCAC-3′; probe B (nt 4683 to 4710), 5′-AAGCTTTTTCTGCACACTGAGGCCAATG-3′; probe C (nt 3849 to 3873), 5′-GCTGAAACTCTATGCTTGCCAGGAG-3′; probe D (nt 2477 to 2494), 5′-GCTACG TAGGCTTGCTTTCCATGCTG-3′.

Blots were hybridized with the probes in 5 x SSPE (1 x SSPE is 0.18 M NaCl, 10 mM NaH_2PO_4, and 1 mM EDTA, pH 7.7), 5 x Denhardt's solution, 0.05% SDS, 35 mM urea, and 0.1 mg/ml of yeast RNA (catalog no. 109223; Boehringer Mannheim) with incubation overnight at 55°C. After hybridization, the blots were washed once in 2 x SSC for 10 min and once in 2 x SSC–0.1% SDS for 30 min at 65°C, followed by Typhoon phosphorimaging (Typhoon-9410).

Western blotting and immunoprecipitation.

T-antigen expression was analyzed by Western blotting and by metabolic labeling of transfected or transduced cells, followed by immunoprecipitation (IP) with specific rabbit polyclonal antibodies (pAbs) raised against TSPyV T-antigens (GenScript). pAbPanT was raised against peptide 41-YHPDKGGDPEKMSR-54 located in the first exon (in ORF1), pAbLT was raised against peptide 77-FSSQHDVPTQDGRD-90 located in the second LT exon (in ORF2), and pAbMT was raised against peptides 218-PPGPAGGKASIKNG-231 and 278-LPHQRTPPAAPRAP-291 located in the second MT exon (in ORF5). Western blotting was performed on transfected cell lysates by incubation of the blot with the specific TSPyV polyclonals (1:10,000), followed by secondary biotinylated anti-rabbit antibody (1:2,000; Dako) and horseradish peroxidase (HRP)-conjugated anti-biotin antibody (1:5,000; Jackson) incubation for detection.

Metabolic labeling was achieved by trypsinizing 5 x 10^6 to 10 x 10^6 transfected cells, followed by incubation of the cells in starvation medium (Dulbecco's modified Eagle's medium [DMEM] without methionine and cysteine; Lonza), for approximately 1 h at 37°C. Cells were metabolically labeled for 30 to 60 min at 37°C by the addition of 220 μCi of ^{35}S protein labeling mix (EasyTag EXPRESS^{35}S protein labeling mix; PerkinElmer). After the labeling step, the cells were lysed in 1 ml of Nonidet P-40 (NP-40)-containing lysis buffer

(150 mM sodium chloride, 50 mM Tris, pH 8.0, 1.0% NP-40) supplemented with protease inhibitors (1 x complete EDTA-free protease inhibitor cocktail; Roche).
Before immunoprecipitation the metabolically labeled samples were precleared with normal rabbit serum (obtained from the animal used for raising antibodies) coupled to protein A-Sepharose beads (GE Healthcare). Immunoprecipitation was performed on the precleared lysate for 2 to 4 h at 4°C by the addition of 5 µl of specific TSPyV T-antigen polyclonal antibody and fresh protein A-Sepharose beads. The beads were subsequently washed four times with NET buffer containing 0.1% SDS (20 mM Tris-HCl, pH 8.0, 200 mM NaCl, 1 mM EDTA, 0.1% NP-40, 10% glycerol, 0.1% SDS), resuspended in 1 x Laemmli buffer, and finally subjected to 10 to 15% SDS-PAGE gels.

Immunofluorescence and immunohistochemistry.
T-antigen localization in the transfected cells was assessed by an immunofluorescence assay (IFA). Cells were seeded on coverslips 2 days prior to transfection to reach 90% confluence; at 24 h posttransfection the cells were fixed with 3% paraformaldehyde in PBS and subsequently washed with PBS–10 mM glycine. Before the antibody staining, cells were permeabilized with 0.1% Triton-X in PBS, followed by blocking with 2% normal goat serum in PBS–10 mM glycine. Cells were subsequently stained by pAbLT or by pAbMT (1:1,000) for 1 h at room temperature. Cy3-conjugated secondary anti-rabbit antibody (1:1,000; Jackson) supplemented with Hoechst for nuclear DNA staining (1:200) was used for detection. Representative pictures were taken at magnifications of x200 and x400 and analyzed using AxioVision software (Carl Zeiss Vision).
Sections of a lesional skin biopsy specimen of a TS patient were stained with hematoxylin and eosin and the pAbLT antibody as described by Kazem et al..[31]

Mass spectrometry.
In-gel trypsin digestion and liquid chromatography (LC)-ion trap tandem mass spectrometry (MS/MS) analysis were performed as described previously.[224] Peak lists were generated using Data Analysis, version 4.0 (Bruker Daltonics, Bremen, Germany), with default settings and exported as Mascot generic files. TSPyV T-antigen sequences were added to the human UniProt database, and the MS/MS spectra were searched against this combined database using the Mascot algorithm (Mascot, version 2.4.1; Matrix Science, London, United Kingdom). An MS tolerance of 0.2 Da (# 13C = 1) and an MS/MS tolerance of 0.5 Da were used. For enzyme specificity, semiTrypsin was designated, and up to one missed cleavage site was allowed. Carbamidomethylcysteine was selected as a fixed modification, and oxidation of methionine and N-terminal peptide acetylation were set as variable modifications. All identified peptides were manually checked for the correct assignment.

Matrix-assisted laser desorption ionization-time of flight (MALDITOF) mass analyses were performed on an UltrafleXtreme time of flight mass spectrometer controlled by the FlexControl, version 3.3, software package (Bruker Daltonics, Bremen, Germany). The mass spectrometer was used in the positive-ion reflectron mode.

Results

Identification of TSPyV T mRNA products by RT-PCR.

To identify its putative mRNA products and locate its major splice donor and acceptor sites, the complete early region of TSPyV was transiently expressed in HeLa and 293T cells. RNA from these cells was analyzed by RT-PCR using primer sets A to H designed around donor and acceptor splice sites, either predicted by the Human Splicing Finder software (HSF; http://www.umd.be/HSF/)[225] or described for other polyomavirus T-splice products, as shown in **FIGURE 1B**. Primer sets I to K with antisense primers spanning an exon-intron boundary were generated to confirm the presence of observed or predicted splice donor and acceptor sites.

Sequencing of the detected RT-PCR products revealed the presence of three splice donor (SD1 to SD3) and three splice acceptor (SA1 to SA3) sites and five introns (I to V). The identification and positioning of each identified splice site and intron are shown in **FIGURE 2 and FIGURE 3A**. All experimentally identified SD and SA sites scored an HSF consensus value above 70 (minimal 1 to maximal 100).

Based on the identified splices, we could identify at least six different TSPyV T mRNA splice products (**FIGURE 3B**, **T1 to T6**) next to the full-length pre-mRNA. Splice acceptor sites SA1 and SA2 are located in close proximity to each other and create comparably sized introns I and II and introns III and IV, respectively (**FIGURE 3A**). However, since a frameshift is introduced between SA1 and SA2, the splice products T1 and T2 and splice products T3 and T4, respectively, generate different protein products (see below).

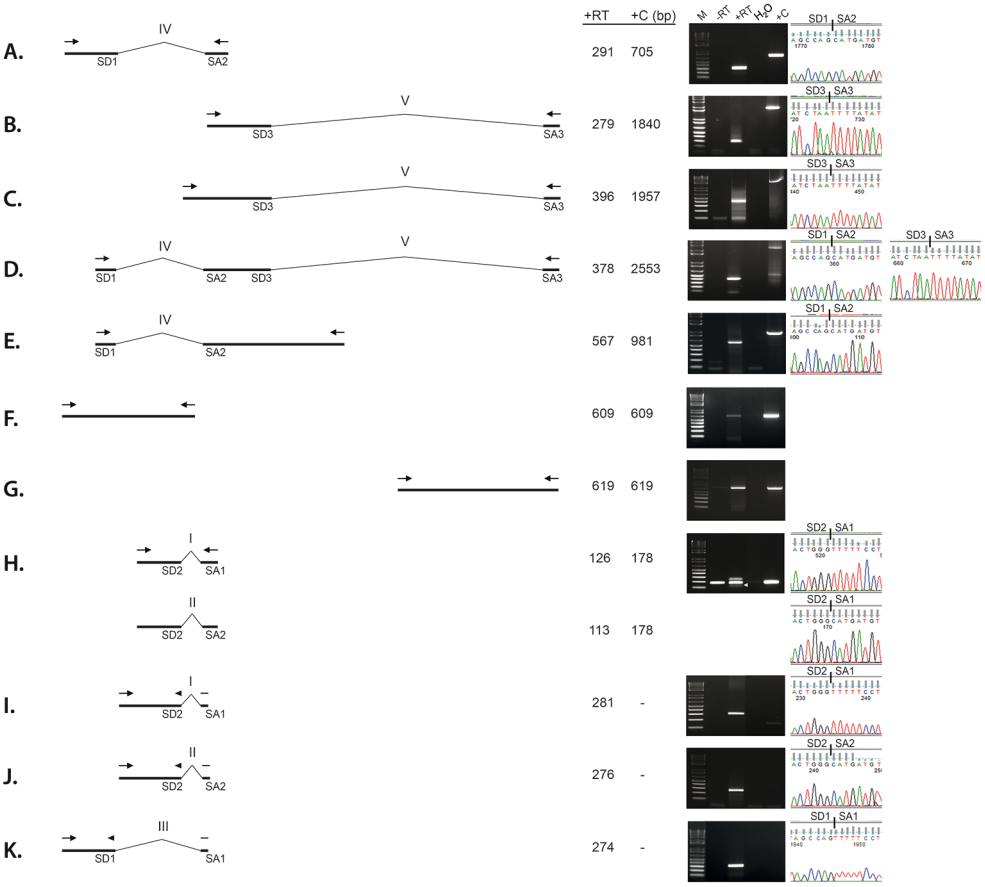


FIGURE 2. RT-PCR RESULTS.

In panels **A** to **K** the RT-PCR results are depicted for primer sets A to K. For each primer set the following results are depicted: (i) the RT-PCR product with the identified intron(s), (ii) the predicted mRNA size (+RT) and the TSPyV DNA control size (+C), (iii) the gel electrophoresis result for the 1-kb Plus DNA Ladder (M), the reaction with (+) and without (-) reverse transcriptase (RT), the H_2O control, and the TSPyV DNA control (+C), and (iv) the cDNA sequence at the splice position except for primer sets F and G (detection of unspliced products). In the RT-PCR performed with primer set H, clones were identified containing either intron I (SD2^SA1) or intron II (SD2^SA2). The antisense primers of primer sets I to K are located over the SD2^SA1, SD2^SA2, and SD1^SA1 sites, respectively.

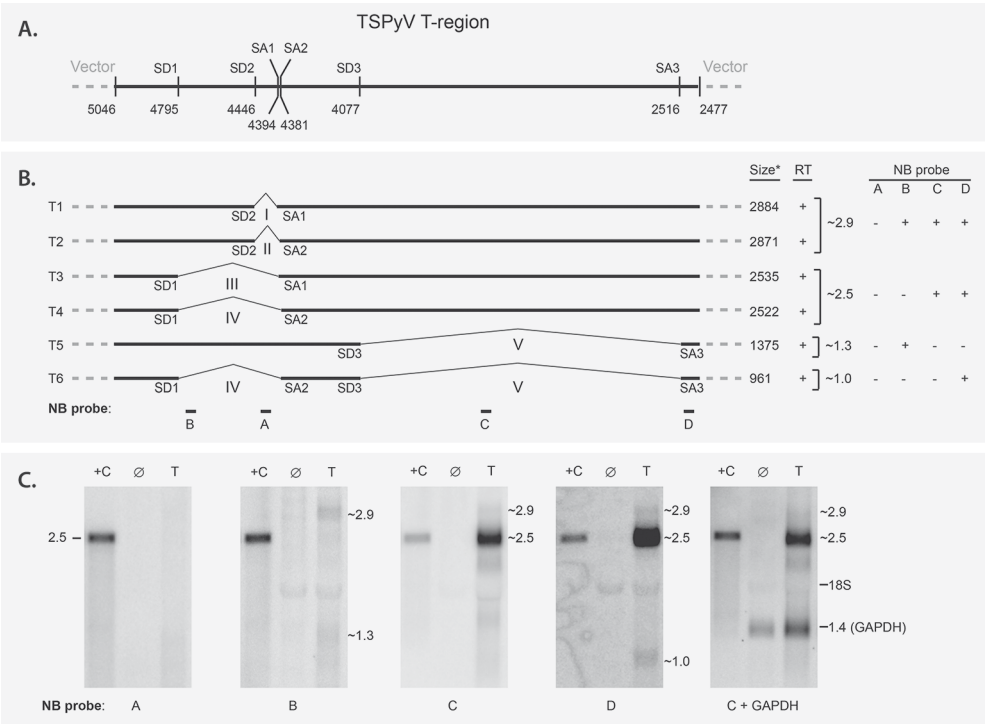


FIGURE 3. TSPyV EARLY REGION TRANSCRIPTION IDENTIFIED BY RT-PCR AND NORTHERN BLOTTING. **A.** Identified splice donor (SD) and acceptor (SA) site nucleotide positions are illustrated within the TSPyV early region. **B.** TSPyV T splice products (T1 to T6) are shown based on the RT-PCR results shown in **FIGURE 2**. For each product, the intron(s) between the used splice donor and acceptor sites is indicated and numbered I to V. The anticipated size of each product and detection by RT-PCR (RT) and Northern blotting (NB probe) are given on the right. The calculated size (*) of each splice product includes 400 nucleotides that originate from the leader and poly(A) termination signal found within the vector. **C.** Northern blot results are shown for early-region-transfected cells (T) and their mock-transfected counterparts (ø). An *in vitro*-transcribed TSPyV T product of 2.5 kb in size is used as a positive control (+C). Northern blots were hybridized with probe A (located within introns I to IV), probe B (located within introns III and IV), probe C (located within intron V), and probe D (located at the 3′ coding region found in all splice products). Detection of glyceraldehyde-3-phosphate dehydrogenase (GAPDH) was used as a positive control.

Confirmation of TSPyV T mRNA products by Northern blotting.

Subsequently, Northern blot experiments were performed on RNA isolated from TSPyV early-region-transfected cells in order to confirm the presence of the splice products identified by RT-PCR. To this purpose, four-oligonucleotide DNA probes were designed, located within introns I to IV (probe A), III and IV (probe B), and V (probe C) and in the 3′ coding region (probe D). The location of each probe and the expected size of each splice product are indicated in **FIGURE 3B**.

Hybridization with probe A, that should detect T5 and pre mRNA, revealed no clear bands (**FIGURE 3C**). Probe B, designed to detect T1, T2, and T5, showed a band of approximately 2.9 kb and possibly one of 1.3 kb. Probe C specific for T1 to T4 revealed a prominent band of approximately 2.5 kb, next to the 2.9-kb band, and possibly some weaker, smaller bands. Probe D, potentially detecting all anticipated splice products, revealed a band of approximately 1.0 kb, in addition to the 2.9- and 2.5-kb bands detected with probes B and C, respectively.

As indicated in the right part of **FIGURE 3B**, the 2.9-kb band could represent both T1 and T2. The 2.5-kb band compatible with T3 and T4 probably represents T4 since T4 was easily detected by RT-PCR, whereas an antisense primer spanning the exon-intron boundary SD1^SA1 (where the caret indicates the splice) was needed to detect T3 (**FIGURE 2A**, **D**, **E**, **and K**). The 1.0-kb band probably represents the double-spliced T6. The 1.3-kb band observed with probe B might represent T5. However, this product was hardly detected or not detected at all with probes A and D, respectively, and therefore remains difficult to attribute. The weak bands smaller than 2.5 kb found with probes B and C are not compatible with any of the splice products identified by RT-PCR.

Identification of TSPyV T-antigen protein products.

In **FIGURE 4A**, the splice products T1 to T6 identified by RT-PCR are shown, including the sizes of the potentially encoded protein products. Precise location of the putatively used start and stop codons and the expected amino acid length of the putatively encoded proteins are given in **TABLE 1**. Based on present open reading frames and identified splice sites and by analogy with T-antigens expressed by other polyomaviruses (**FIGURE 1**), T1 and T2 probably encode MT (332 amino acids [aa]) and ST* (197 aa), respectively, sharing ORF1 and SD2 and continuing either in the -1 reading frame in ORF5 or in the -2 reading frame, where a stop codon is located shortly after SA2. T3, T4, and T6 that share the first part of ORF1 and SD1 probably encode tiny T (85 aa), LT (697 aa), and 21kT (184 aa), respectively. In T3, immediately after SA1, a stop codon is found in the -2 reading frame creating tiny T, whereas T4 and T6 that use SA2 continue in ORF2 in the -3 reading frame to create LT and 21kT, respectively. T5, for which there is only weak evidence of its presence (see paragraphs above), possibly encodes ST (198 aa), which is almost identical to ST* introduced above. Except for T5 and T6, all identified splice products can potentially encode ALTO since ORF5 is found in all of them, as indicated in **FIGURE 4A**.

To assess the existence of MT, ST*, tiny T, LT, ST, 21kT, and ALTO potentially encoded by T1 to T6, we analyzed the early region-transfected cells by Western blotting and by immunoprecipitation of metabolically labeled cells with the help of polyclonal antibodies

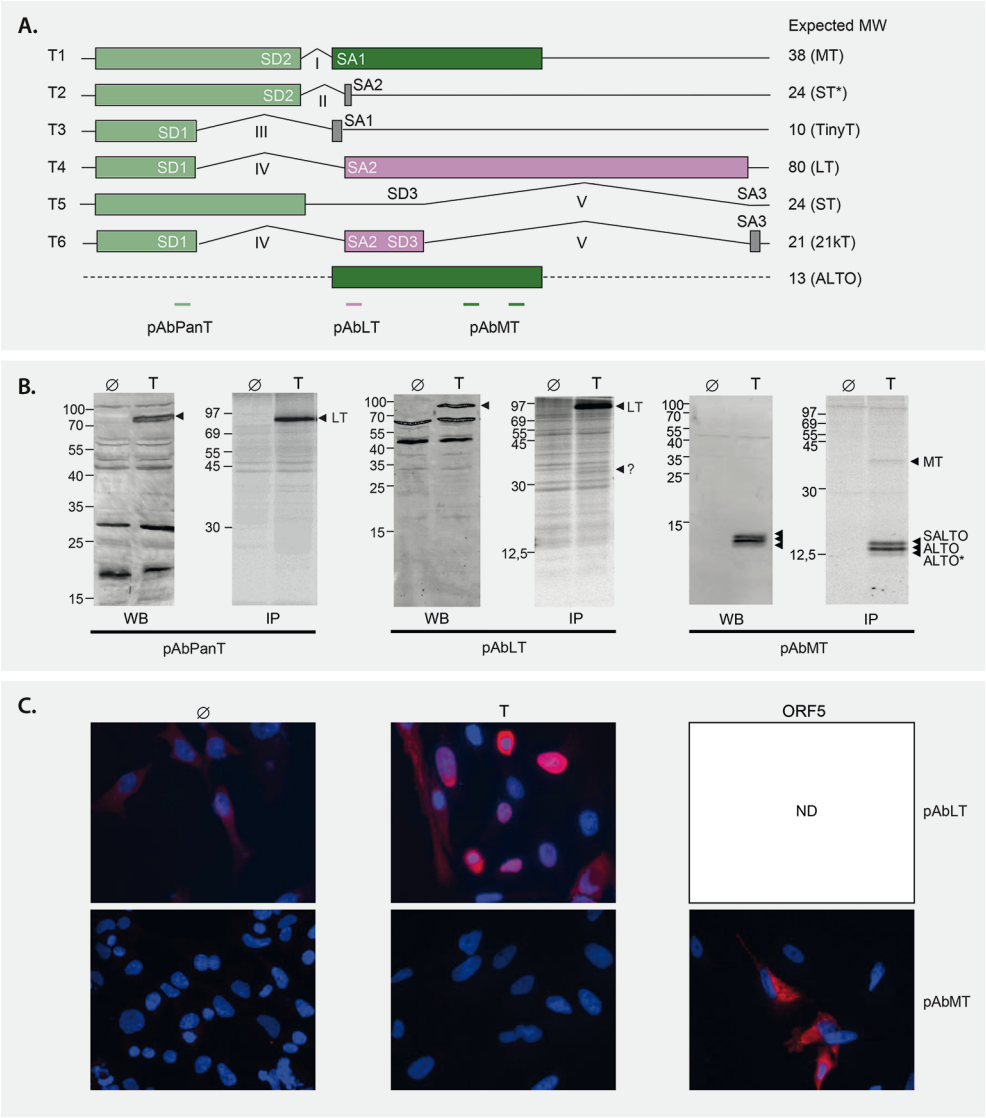


FIGURE 4. TSPyV T-ANTIGEN EXPRESSION.

A. Schematic overview of TSPyV splice products T1 to T6 and their putatively encoded T-antigens including expected molecular weights (in thousands). The colored boxes correspond to the ORFs depicted in **FIGURE 1A**. Epitope locations are given for the TSPyV polyclonal antibodies pAbPanT, pAbLT, and pAbMT. **B.** Western blotting (WB) and immunoprecipitation (IP) of metabolically labeled lysates of mock-transfected (ø) and TSPyV T (T)-transfected cells showing LT expression with pAbPanT and pAbLT along with MT and SALTO, ALTO, and ALTO* expression with pAbMT. **C.** Immunofluorescence of mock-transfected (ø) and TSPyV T-region (T)- and ORF5-transfected cells stained with pAbLT and pAbMT (red) and Hoechst nuclear staining (blue).

raised against peptides predicted to represent TSPyV T-antigen epitopes. Western blotting and immunoprecipitation with a PanT antibody that should detect all T-antigens except ALTO showed one prominent band of approximately 80 kDa compatible with the size of LT (**FIGURE 4B, pAbPanT**). The nature of this band was confirmed by staining with a specific LT antibody by both Western blotting and immunoprecipitation (**FIGURE 4B, pAbLT**). In addition and only by immunoprecipitation, which is considered more sensitive than Western blotting, a weak band of approximately 35 kDa was observed, considerably larger than the expected 21kT band that was not detected. Detection with an MT-directed antibody revealed a weak band of 38 kDa, the expected size of MT (**FIGURE 4B, pAbMT**), which was observed only by immunoprecipitation. In addition, three bands compatible in size with ALTO were observed by both Western blotting and immunoprecipitation (**FIGURE 4B, pAbMT**). The two upper ALTO products, spliced-ALTO (SALTO) and ALTO, were more or less equally strongly expressed, whereas the smallest product, called ALTO*, was sometimes hard to detect. To summarize, abundant protein expression was observed for LT and ALTO, whereas MT protein expression was considered weak. Protein expression of the other T-antigens was either too low for detection or absent.

TABLE 1. CHARACTERISTICS OF TSPyV T-ANTIGEN SPLICE PRODUCTS

Splice product	Start[a]	Spice donor[a]	Splice acceptor[a]	Stop[a]	Putative protein	No. of amino acids	Expected size (kDa)
T1	5034	SD2 4446	SA1 4394	3985	Middle T	332	38
T2	5034	SD2 4446	SA2 4381	4377	Small T*	197	24
T3	5034	SD1 4795	SA1 4394	4377	Tiny T	85	10
T4	5034	SD1 4795	SA2 4381	2528	Large T	697	80
T5	5034	SD3 4077	SA3 2516	4438	Small T	198	24
T6[b]	5034	SD1 4795	SA2 4381	2507	21kT	184	21
		SD3 4077	SA3 2516				

[a] Nucleotide position
[b] T6 is double spliced

An immunofluorescence assay was performed to study the subcellular localization of the most prominent TSPyV T-antigens. LT expression was detected in the nuclei of cells transfected with the complete early region (**FIGURE 4C**). MT and/or ALTO expression was not detected in these cells. However, when cells were transfected with an ORF5-only construct, which resulted in much higher expression of ALTO (data not shown), ALTO staining was observed in the cytosol (**FIGURE 4C**).

TSPyV T mRNA and protein expression in TSPyV-infected tissue.

To demonstrate TSPyV T expression *in vivo*, we analyzed skin sections of TS patients for the presence of T mRNA and proteins. With the help of RT-PCR and product sequencing, we detected the presence of mRNA compatible with splice products T1 to T4 putatively encoding MT, ST*, tiny T, and LT, respectively (**FIGURE 5A**), and ALTO. No evidence of T5 or T6 mRNA expression was obtained. Northern and Western blot experiments could not be performed on the clinical samples because of insufficient amounts of RNA and protein. LT protein expression was detected in the nuclei of inner root sheath cells of a TS-affected hair follicle by immunohistochemistry, whereas LT was not detected in the epidermis of this TS patient (**FIGURE 5B, pAbLT**). Staining with pAbMT did not show detectable MT or ALTO protein (data not shown); however, high background staining hampered the interpretation of the results.

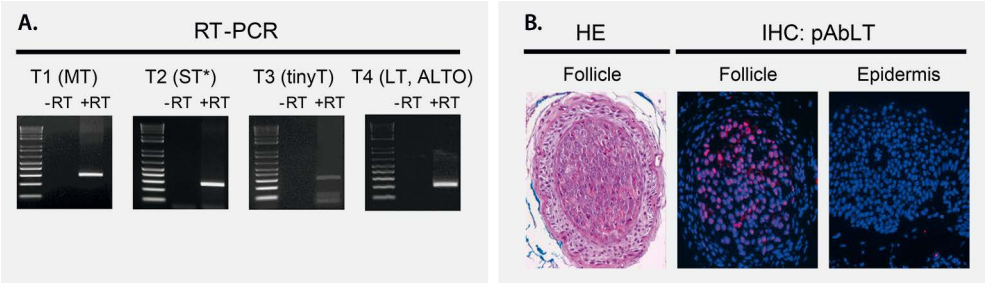


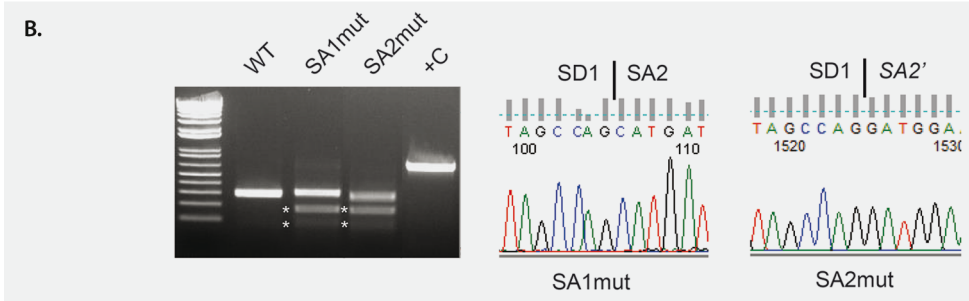
FIGURE 5. TSPyV T mRNA AND LT PROTEIN EXPRESSION IN TS-AFFECTED SKIN DETECTED BY RT-PCR. **A.** Detection by RT-PCR of T1 encoding MT with primer set I (281 bp), T2 encoding ST* with primer set J (276 bp), T3 encoding tiny T with primer set K (274 bp), and T4 encoding LT with primer set L (219 bp). The 1-kb Plus DNA Ladder was used as a marker in the gel electrophoresis experiment. **B.** Hematoxylin and eosin (HE) staining of a cross-section of a hair follicle and LT staining of nuclei of inner root sheath cells found in the hair follicle but not in the epidermis of a TS-affected skin section (magnification, x200). IHC, immunohistochemistry.

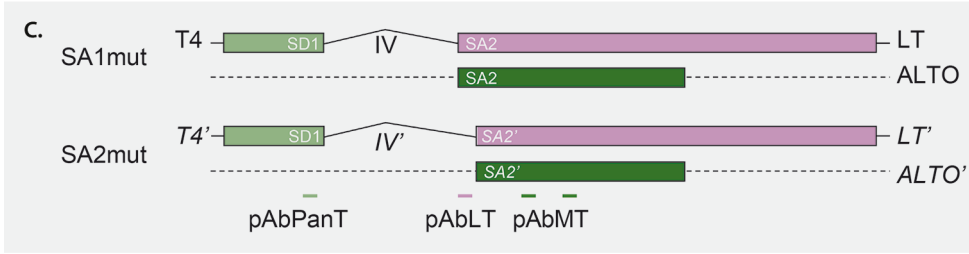
Characterization of TSPyV ALTO expression.

To determine which of the T1 to T4 splice products that each contain the entire ORF5 encodes ALTO (**FIGURE 4A**), mutations were introduced in splice acceptor sites SA1 and SA2 (**FIGURE 6A**). First, we analyzed the effect of these mutations on expression of the wild-type T splice products indicated in **FIGURE 3**. In the SA1 mutant, only product T4 could be detected by RT-PCR and subsequent sequencing (**FIGURE 6B**). In the SA2 mutant, next to T3, a cryptic splice product was detected (**FIGURE 6B**). This product, called *T4'*, used a cryptic splice acceptor site called *SA2'* located just downstream of the original SA2 site (**FIGURE 6A and C**). Expression of T2 and T1 in the SA1 and SA2 mutants, respectively, was not detected, even with specific primers spanning the exon-intron boundary (data not shown).

A.

Name	Sequence	nt position	Cryptic site	nt position	Deleted product	Cryptic product	Cryptic protein
SA1	atttag/TTT	4394					
SA1mut	atttgc/TTT		-		T1 and T3		
SA2	ttttag/CAT	4381					
SA2mut	ttttga/CAT		*SA2'*	4363	T2 and T4	*T4'*	*LT'*





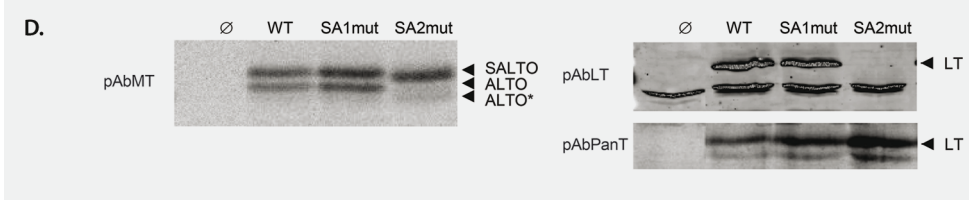


FIGURE 6. IDENTIFICATION OF SPLICE PRODUCTS USED FOR ALTO EXPRESSION.

A. Table showing characteristics of splice acceptor mutants SA1mut and SA2mut. **B.** RT-PCR results using primer set A for wild-type pcDNA3-TSPyV T (WT) (291 bp), splice acceptor SA1-mutant (SA1mut) (291 bp), SA2-mutant (SA2mut) (273 bp), and a TSPyV DNA control (+C) (705 bp). Nonspecific, human genome-derived RT-PCR products are indicated with a white asterisk, as identified by sequencing. A 1-kb Plus DNA Ladder was used as a marker in the gel electrophoresis experiment. Sequencing results of the specific RT-PCR products at the splice location are depicted for the SA1 and SA2 mutants, with detection of a cryptic splice acceptor (*SA2'*) in SA2mut. **C.** Schematic representation of T4 encoding LT and cryptic *T4'* encoding *LT'* detected in the SA1 (SA1mut) and SA2 (SA2mut) mutants, respectively, and their corresponding ALTO and *ALTO'* products. Locations of the antibody epitopes used are given for TSPyV polyclonal antibody pAbPanT, pAbLT, and pAbMT. **D.** ALTO expression of metabolically labeled cells of mock (ø), TSPyV T-antigen wild-type (WT), SA1mut, and SA2mut samples detected by immunoprecipitation (IP) with antibody pAbMT and LT expression detected by Western blotting with antibody pAbLT and by immunoprecipitation with antibody pAbPanT.

The effect of the splice acceptor mutants on ALTO expression was analyzed by immunoprecipitation. In the SA1 mutant with intact T4 expression, no difference in expression of the ALTO products was observed compared to the wild-type levels (**FIGURE 6D**), suggesting that T4 is used to encode the ALTO products. This was confirmed by the SA2 mutant, which created *T4'*, the cryptic, slightly shorter version of T4, and caused a downshift of SALTO and disappearance of ALTO expression. A related change in expression was observed for LT: LT was undetected when it was assessed with the LT-specific antibody pAbLT directed against the LT peptide epitope spliced out in the SA2 mutant (**FIGURE 6C and D**). Cryptic *LT'* expression was detected, however, with the PanT antibody directed at the N-terminal part of LT. The consequent LT size shift from 80.2 kDa to the 79.5 kDa of *LT'* was too small to observe. Based on the concomitant changes in LT and ALTO expression, we conclude that ALTO, just like LT, is derived from T4, which in the previous experiments proved to be the most prominent T-antigen splice product.

Localization of the TSPyV ALTO initiation sites.

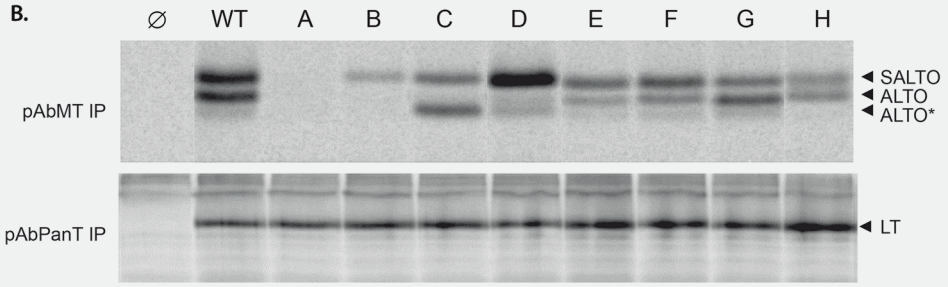
Based on the observation by Carter and coworkers that ALTO expression is initiated by an internal start codon located at the 5' end of the ALTO ORF,[93] we reasoned that an internal start codon present at the 5' end of ORF5 is used to initiate TSPyV ALTO expression. The results obtained with the splice acceptor mutants already provided some indication about the localization of the start codons used for the ALTO products. Given the observed size shift in the SA2 mutant, we assumed that expression of the upper product SALTO (for spliced-ALTO) is initiated by a start codon upstream of intron IV in T4. The smaller ALTO and ALTO* products are likely initiated by start codons at the 5' end of ORF5.

To pinpoint the origin of each ALTO product, we introduced mutations in the putatively relevant start codons within ORF5 and in a start codon located just upstream of intron IV (**FIGURE 7A**) and analyzed expression of the ALTO products by immunoprecipitation. Replacing all potentially relevant methionines by leucines completely abolished ALTO product expression (**FIGURE 7A and B, mutant A**). Restoring the wild-type methionine upstream of intron IV (mutant B) rescued SALTO expression, thereby confirming the spliced status of SALTO and initiation at the M residue at position -13 (M-13). In mutant C, next to SALTO, ALTO* expression was also restored, indicating the use of M14 for this product. This was confirmed by the M14-only mutant H, where no expression of ALTO* was detected. Expression of the intermediate ALTO product was rescued upon restoring M1 and/or M2 in mutants E to G. M7 mutation did not seem to affect expression of any ALTO product. We have no obvious explanation for the repeatedly observed increased expression of SALTO in mutant D, other than preferred translation of SALTO in the context of the M1 and M2 mutations. Since no effect on LT expression was observed by any of the substitutions, we believe that the SD1^SA2-containing T4 mRNA remains intact in these mutants (**FIGURE 7B, pAbPanT**). In **FIGURE 7C**, the ALTO products are schematically presented with their start codons and expected molecular masses.

A.

Mutant	Methionine position -13 1 2 7 14
WT	M Q D K S F Q L A S V A S ^ M M F Q P R M E E I Y L P M
A	L Q D K S F Q L A S V A S ^ L L F Q P R L E E I Y L P L
B	M Q D K S F Q L A S V A S ^ L L F Q P R L E E I Y L P L
C	M Q D K S F Q L A S V A S ^ L L F Q P R L E E I Y L P M
D	M Q D K S F Q L A S V A S ^ L L F Q P R M E E I Y L P M
E	M Q D K S F Q L A S V A S ^ L M F Q P R M E E I Y L P M
F	M Q D K S F Q L A S V A S ^ M L F Q P R M E E I Y L P M
G	M Q D K S F Q L A S V A S ^ M M F Q P R L E E I Y L P M
H	M Q D K S F Q L A S V A S ^ M M F Q P R M E E I Y L P L

Methionines are numbered according to the amino acid position relative to the SD1^SA2 splice



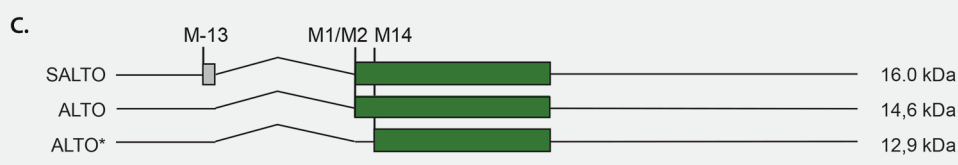


FIGURE 7. IDENTIFICATION OF START CODONS USED FOR ALTO EXPRESSION.

A. Table showing mutants of putative ALTO start codons. At top, the amino acid position(s) of the mutated methionines (M) are indicated in relation to intron SD1^SA2. In red the substituted leucines (L) are shown. The caret represents the SD1^SA2 splice creating intron IV. **B.** SALTO, ALTO, and ALTO* expression observed in metabolically labeled mock (ø), TSPyV T-antigen-expressing (WT), and start codon mutant-expressing cells detected by pAbMT immunoprecipitation. As a (loading) control, LT expression is shown detected by pAbPanT immunoprecipitation. **C.** Schematic overview of SALTO, ALTO, and ALTO* proteins using T4 with their expected molecular masses and methionine positions.

Confirmation of ALTO expression and initiation by mass spectrometry.

In addition to the start codon mutant experiments, the ALTO products expressed in the pcDNA3-TSPyV T transfected cells were analyzed by mass spectrometry. For this purpose, immunoprecipitation using the pAbMT Ab was performed. Following SDS-PAGE and Coomassie staining, no protein bands were visible (data not shown), and no guided excision could be performed. Therefore, four gel slices (slices 1 to 4, from high to low molecular mass) from the region between molecular masses of 10 and 15 kDa were excised, treated with trypsin, and analyzed by mass spectrometry. In the upper three gel slices, ALTO peptides were identified. The highest number of ALTO peptides were identified in gel slice 2 (**FIGURE 8A**). In addition to internal tryptic peptides, this spectrum also revealed peaks corresponding to putatively acetylated peptides of ALTO, with either M1 or M2 as the N-terminal amino acid. These peptides can only have been generated by trypsin when they belonged to the N terminus of the ALTO protein.

To confirm the assignment of these peptides, an LC-ion trap MS/MS experiment was performed. With this analysis the two acetylated peptides were clearly identified (**FIGURE 8B**). These N-terminal peptides were not identified in the analyses of gel parts 1 and 3. Moreover, neither of the mass spectrometry analyses resulted in the identification of peptides corresponding to the unique N terminus of SALTO and/or ALTO* in any of the three gel slices containing ALTO peptides.

Finally, to confirm ALTO as the most prominent product, we expressed the TSPyV early region through another (retroviral) expression system. After immunoprecipitation with the pAbMT antibody, we observed only one band comparable in size with the middle ALTO product obtained with pcDNA3 transfection of the T region (**FIGURE 8C**). Bands indicative of the presence of SALTO and ALTO* were not observed using retroviral TSPyV T-antigen expression.

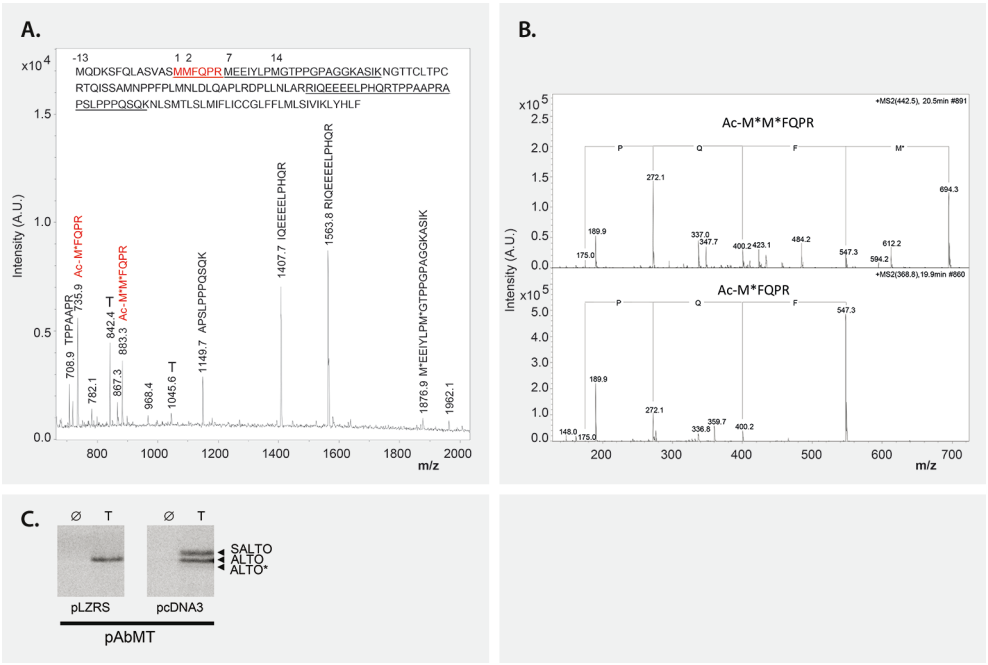


FIGURE 8. ALTO ANALYSIS BY MASS SPECTROMETRY AND BY USING A RETROVIRAL EXPRESSION SYSTEM.

A. MALDI-TOF MS identification of M1 and M2 as alternative translation starts of ALTO. M*, oxidized methionine; T, endoproteolytic fragment of trypsin; Ac, acetylated. Underlined regions in the inset were covered by tryptic peptides within the spectrum. **B.** Ion trap MS/MS spectra from ALTO (N-terminal) peptides. M*, oxidized methionine; Ac, acetylated. **C.** ALTO expression detected by immunoprecipitation of metabolically labeled cells with pAbMT. pLZRS, ALTO expression using the pLZRS retroviral expression system; pcDNA3, ALTO expression using the pcDNA3 expression system for mock (ø)- and TSPyV T-antigen (T)-transduced/transfected cells.

Discussion

In this study, we mapped the T transcriptome and proteome of TSPyV, one of the newer human polyomaviruses known to cause TS. A total of six TSPyV T splice products (T1 to T6) were identified, potentially coding for at least six T-antigens, including LT, MT, ST, tiny T, 21kT and ALTO. In TSPyV-infected tissue, mRNA expression of T1 to T4 was observed, suggesting that MT, ST, tiny T, LT, and ALTO are expressed during natural infection. For LT this was confirmed by immunohistochemistry. Whether the other proteins are expressed in a natural setting, where TSPyV transcription is controlled by cognate promoters and enhancers, remains to be shown. Of all T-antigens identified, LT was the most abundant, as indicated by RT-PCR, Northern and Western blotting, immunoprecipitation, and

immunohistochemistry. In a previous study we already provided evidence of LT involvement in inducing hyperproliferation of TSPyV-infected cells by hyperphosphorylation of the tumor suppressor protein pRB.[31]

As far as we know, expression of MT is unique among human polyomaviruses, whereas it is common among rodent polyomaviruses, such as MPyV.[115-117] Interestingly, TSPyV resembles MPyV in more ways. For example, TSPyV ST seems to be encoded by two exons (**FIGURE 3A, ST***) rather than by one (**FIGURE 3A, ST**) since T2, unlike T5, was easy to detect in the RT-PCR experiments and was found in TS lesional tissue. Furthermore, the putative presence of tiny T is reminiscent of an MPyV-like expression pattern. Resemblance particularly to MPyV (and hamster polyomavirus) is not entirely unexpected since phylogenetically these viruses are more closely related to TSPyV than most, if not all, of the human polyomaviruses analyzed so far.[192]

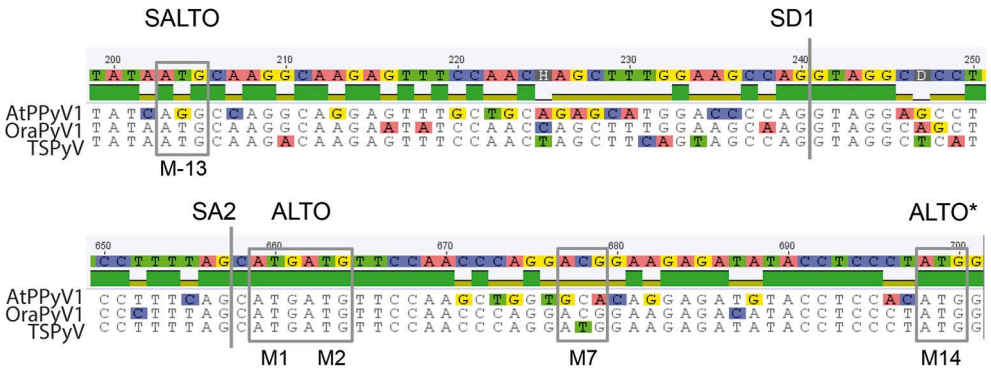


FIGURE 9. ALIGNMENT OF THE EARLY REGION.

ClustalW alignment of two parts (nt 200 to 250 and nt 650 to 700) of the early region of TSPyV with its closest relatives, Ateles paniscus polyomavirus 1 (AtPPyV1) and orangutan polyomavirus (OraPPyV1). Boxes show the start codon positions for SALTO (M-13), ALTO (M1/M2), ALTO* (M14), and M7. In addition, SD1 and SA2 positions are shown.

Expression of ALTO is clearly an aspect of TSPyV that relates to MCPyV.[93] Our study indicates that a number of start codons can be used for ALTO initiation, at least in the transfection model; nevertheless, it remains uncertain whether all of these start codons are used during a natural infection. To our knowledge, expression of a spliced ALTO product, SALTO, was not described before. The mass spectrometry analysis, as well as our findings with the retroviral expression system, points to the M1 and M2 start codons as the prime initiation site(s) of ALTO. Since alignment of TSPyV with its closest relatives, the Bornean orangutan polyomavirus and Ateles paniscus polyomavirus 1, shows

conservation of both M1 and M2 (**FIGURE 9**), we believe that (one of) these sites represent the natural ALTO initiation site. The identification of acetylated M1 and M2 N-terminal peptides underscores this idea and demonstrates that ALTO is a genuine protein and not, for example, a result of proteolytic degradation of MT. This is further substantiated by our splice acceptor site mutants, which showed that loss of the MT-encoding splice product T1 did not influence ALTO expression. The reason why we have not identified N-terminal peptides corresponding to SALTO (M-13 as start codon) or ALTO* (M14 as start codon) may be related to technical difficulties.

How ALTO expression is regulated remains elusive. Since the splice product used for ALTO expression, T4, contains a number of start codons upstream of ORF5, including the one used for the prominent LT product, it seems likely that ALTO is expressed through an internal ribosome entry site (IRES) located upstream of ORF5, as was described, for instance, for the VP3 structural protein of SV40.[99] Alternatively, as postulated for the translation initiation of early proteins of the related family of papillomaviruses, ribosomal scanning might play a role, e.g., leaky scanning, discontinuous scanning, or translation reinitiation.[226] Translation reinitiation might be the case for ALTO expression in the context of the pcDNA3-mediated expression, where we found three ALTO products present at the same time. Of note, the Kozak sequence[227] of the identified start codons M-13 (SALTO), M1, and M2 (ALTO) for translation are considered equally strong with the highly conserved adenine or guanine present at nucleotide position -3 relative to the start codon. For M14 (ALTO*) a cytosine is observed at this position. However, this start codon does contain the conserved guanine at position +4 relative to the start codon, which is not observed for the other ALTO start codons.

The ALTO protein was detected in the cytoplasm of the transfected cells, similar to results shown for MCPyV.[93] At the C-terminal end of TSPyV ALTO and of MT, the second exon of which is also encoded by ORF5, a transmembrane domain is found.[93,105] This domain was shown to be important for cellular transformation activity exploited by MPyV.[115-117] Whether TSPyV MT and ALTO anchor in the outer and/or endoplasmic reticulum (ER) membrane needs to be confirmed in future experiments. Since ORF5 was found to be largely conserved within the monophyletic cluster of lineage I *orthopolyomaviruses*,[105,192] also referred to as *almipolyomaviruses*,[93] it is likely that ALTO and/or MT plays an important role in the evolution of this cluster of polyomaviruses, probably by being the subject of positive selection, as we have recently shown.[105] Further studies have to resolve whether ORF5 is generally expressed as MT, ALTO, or both.

Despite the comprehensiveness of our data, there are still some findings that we cannot explain, for example, the detection by immunoprecipitation of a protein of approximately 35 kDa in size when the pAbLT antibody was used for staining. Such a product was not expected based on the predicted molecular mass of the proteins encoded by T1 to T6. Since the 35-kDa product is not detectable with the PanT antibody, if genuine, it likely

consists of a spliced or truncated version of the second LT exon that was not detected by RT-PCR. Perhaps one of the faint, undefined Northern blotting bands corresponds to a splice product encoding such a protein. Unfortunately, the expression of this product was too weak to be analyzed by mass spectrometry. Furthermore, the increased expression of SALTO in the ALTO start codon mutant D remains enigmatic. Perhaps the ORF5 ALTO start codon mutants interfere with the putative ALTO IRES, but this is difficult to assess at this point.

In sum, by describing the T-antigenome of TSPyV in detail, we have been able to add valuable new information to the existing body of knowledge regarding (human) polyomavirus T-antigen expression. As a result, specific aspects of TS disease, in which highly productive virus infection is found combined with a hyperproliferative cell phenotype, can now be correlated with a unique T-antigen expression pattern that combines aspects of both human and rodent polyomaviruses. In this regard, in particular the roles of ALTO and MT deserve further study as they represent proteins with potentially cell-transforming properties,[93,115-117] about which very little is known in the human context. The recently published observations which indicate that the overlapping (alternative) ORF5 that encodes ALTO and the second exon of MT is highly conserved among the lineage I *orthopolyomaviruses*[105,192] and *almipolyomaviruses*[93] underscore the importance of these products for polyomavirus host adaptation and evolution.

Acknowledgements

We thank Alexander Gorbalenya, Chris Lauber, and Hans van Leeuwen for fruitful discussions.