Journal of NeuroVirology (1998) 4, 182–193

http://www.jneurovirol.com

Sequence analyses of human tumor-associated SV40 DNAs and SV40 viral isolates from monkeys and humans

A Renee Stewart, John A Lednicky and Janet S Butel

Division of Molecular Virology, Baylor College of Medicine, One Baylor Plaza, Houston, Texas 77030, USA

SV40 DNA has been found associated with several types of human tumors. We now report a sequence comparison of SV40 DNAs from pediatric brain tumors and from osteosarcomas with viral isolates from monkeys and from humans. We analyzed the entire genomic sequences of five isolates, Baylor and VA45-54 strains from monkeys and SVCPC, SVMEN, and SVPML-1 recovered from humans, and compared them to the reference virus SV40-776. The viral sequences were highly conserved, but isolates could be distinguished by variations in the structure of the viral regulatory region and in the nucleotide sequence of the variable domain at the C-terminus of the large T-antigen gene. We conclude that multiple strains of SV40 exist that can be identified on the basis of sequences in these regions of the viral genome. The isolates were more similar to each other and to the Baylor strain than to the reference strain SV40-776. Human isolates SVCPC and SVMEN were found to be identical. The DNAs present in some human brain and bone tumors were authentic SV40 sequences. Many of the C-terminal T-ag sequences associated with human tumors were unique, but some sequences were shared by independent sources. There was no compelling evidence for human-specific strains of SV40 or for tumor type-specific associations, suggesting that SV40 has a relatively broad host range. The source of the viral DNA found in human tumors remains unknown.

Keywords: SV40; T-antigen; human tumors; viral isolates; regulatory region; genome sequences; polyomavirus

Introduction

SV40 is a small DNA tumor virus that is a member of the *Polyomavirus* genus in the Papovaviridae family. It has been well-characterized for its ability to immortalize and transform cells in culture and to induce tumors in experimental animals (Butel, 1994; Cole, 1996). SV40 DNA sequences have recently been found associated with some human tumors of brain, lung, and bone origin (Bergsagel *et al*, 1992; Carbone *et al*, 1994, 1996; Cristaudo *et al*, 1995; Lednicky *et al*, 1995a, 1997; Woloschak *et al*, 1995; Martini *et al*, 1996; Pepper *et al*, 1996), supporting older reports of detection of SV40 in human cancers (Soriano *et al*, 1974; Meinke *et al*, 1979; Scherneck *et al*, 1979; Krieg *et al*, 1981; Krieg and Scherer, 1984; reviewed in Geissler, 1990).

Correspondence: JS Butel
Received 11 July 1997; revised 22 August 1997; accepted 12 September 1997

The SV40 genome is small in size (5.2 kbp) and encodes only six genes. The early region encodes the large tumor antigen (T-ag) and the small tumor antigen (t-ag). Large T-ag mediates the immortalization and transformation of cells and included among its important functions are binding of the p53 and pRB tumor suppressor proteins, which effectively negates cellular growth control mechanisms (Butel, 1994; Cole, 1996). Large T-ag is absolutely required for virus replication, as it is necessary for the initiation of copying of the viral DNA (Fanning, 1992; Cole, 1996). The late region encodes the major capsid protein VP1, the minor capsid proteins VP2 and VP3, and the agnoprotein, a scaffolding, maturation protein. The capsid is a determinant of the host range of the virus, as it mediates attachment of virus particles to cellular receptors.

SV40 was initially isolated from contaminated virus vaccines that had been produced unknowingly in kidney cultures derived from naturally infected

monkeys, as well as from the uninoculated monkey kidney cells (Sweet and Hilleman, 1960). Millions of people were inadvertently exposed to SV40 when they were given SV40-contaminated poliovaccine in the late 1950's and early 1960's (Shah and Nathanson, 1976). A few of those early SV40 isolates have been established as common laboratory strains, and they have been the subject of the majority of studies of the biology of SV40 and the structure and function of its genome and gene products. The strain designated SV40-776 was the first, and until now, the only one to be fully sequenced (Fiers *et al*, 1978; Reddy *et al*, 1978), and it has been considered to be the reference strain of SV40. Several isolates of SV40 have been recovered from humans, including from a meningioma (SVMEN), from a choroid plexus carcinoma (SVCPC), and from the brain of a patient with progressive multifocal leukoencephalopathy (SVPML-1). In addition, as noted above, SV40 DNA sequences have been found associated with a number of human brain tumors, osteosarcomas, mesotheliomas, and other tumor types.

These findings raised several important questions. Does strain variation exist among different isolates of SV40? We have reported previously that not all SV40 isolates are identical, that different isolates contain a number of sequence differences in the early region of the viral genome, and that T-ag contains a variable domain at its C-terminus (Stewart *et al*, 1996). Are there human-specific strains of SV40? The discovery of SV40 DNA in association with human tumors raised the possibility that genetic changes might have produced a virus more tumorigenic for humans. Finally, are there tissue-tropic strains of SV40? As SV40 sequences have been found in tumors arising in different tissues (such as brain, bone, lung), the possibility was considered that subtle genetic changes might produce a virus strain more adapted to a particular target tissue. To address those questions, we performed an extensive sequence analysis of SV40 DNA from various sources. We report here the complete sequence for several laboratory strains and human isolates of SV40 and compare those sequences with the regulatory region sequences and the large T-ag variable domain sequences recovered from human brain tumors and osteosarcomas. We found the sequence of SV40 to be highly conserved among isolates, although a few differences in the late region were observed in addition to those changes already noted to occur in the early region. There are clearly different strains of SV40 which can be differentiated on the basis of nucleotide sequences in their regulatory region and the T-ag variable domain. Most of the viruses were more similar to each other and to the Baylor strain of SV40 than to the reference virus SV40-776. None of the inter-strain variations were diagnostic of a specific host or tissue type of origin.

Results

Complete sequences of SV40 isolates

Partial genomic sequences have been reported for human isolates of SV40, SVCPC, SVMEN and SVPML-1 and for the laboratory isolates Baylor and VA45-54 (see Materials and methods). The genomic sequences of SVCPC, SVMEN, Baylor, and VA45-54 are now complete and that of SVPML-1 is nearly complete. A listing of the nucleotide differences detected among these isolates is given in Table 1 As shown, only a few of the observed nucleotide changes would result in predicted amino acid substitutions. A schematic comparison of the nucleotide and amino acid changes throughout the entire genomes of these isolates is shown (Figure 1). Strain 776 is the prototype for SV40, and the numbering of nucleotides and amino acid residues described here is based on it. 'Changes' in a given isolate were determined by comparison to the sequence of SV40-776 (Figure 1a). However, as we analyzed the sequence data in detail, it became apparent that most of the isolates were more similar to each other than they were to reference strain SV40-776. This is illustrated by a comparison to the sequence of the Baylor laboratory strain (Figure 1b). This comparison results in less-apparent sequence diversity among natural isolates, relative to the strain 776 comparison, and emphasizes the high degree of sequence conservation among SV40 isolates. In fact, the sequences of SVCPC and SVMEN are exactly the same, and these viruses are hereafter referred to as SVCPC/SVMEN. This identity is noteworthy, as the two viruses were isolated from two different patients with two different types of brain tumors more than a decade apart.

Structural proteins of SV40 isolates

The differences noted among viral isolates were distributed over the regulatory region, the unique region of small t-ag, the variable region at the C-terminus of large T-ag, and the structural proteins VP1, VP2, and VP3 (Table 1, Figure 1). SVPML-1 contained a difference at nucleotide 623, resulting in a predicted change of amino acid 21 from Ala to Asp in VP2. VA45-54 had a difference at nucleotide 1300, resulting in a change of amino acid 247 of VP2 (129 of VP3) from Asp to Asn in the common region of VP2 and VP3. The Baylor, VA45-54, SVCPC/SVMEN and SVPML-1 strains all differed from SV40-776 at nucleotide 1756, resulting in a change of amino acid 86 from Glu to Asp in VP1. In addition, VA45-54 had a deletion of three nucleotides in the intergenic region between the VP1 and large T-ag genes. The regions coding for the agnoprotein of SVCPC/SVMEN, Baylor, and VA45-54 were absolutely conserved (the sequence of the agnoprotein-coding region of SVPML-1 has not been completed).

One region of the VP1 coding sequence has been analyzed in the SV40 DNAs associated with human tumors (nucleotides 2288–2450) (Lednicky *et al*, 1995a, 1997); samples included six brain tumors and five osteosarcomas. In each case, there was a nearly exact match with the known sequence of SV40. All the samples had an A, rather than a C, at nucleotide position 2384, as exemplified by

Table 1 Nucleotide differences among SV40 isolates

Region	*Nucleotide Number*	*SV40-776*	*Baylor*	*VA45-54*	*SVCPC/ SVMEN*	*SVPML-1*
Regulatory	82	A	del 1 nt	A	A	A
	144–230					rearranged
	145	C	T	C	C	C
	178	C	AT	ins 21 nt	C	na
	179–250				del 72 nt	
	217	C	T	C	C	na
VP2	623	C	C	C	C	***A***
	732	T	A	A	A	A
	768	G	G	C	G	G
	849	G	G	G	A	G
VP2/3	948	C	T	T	T	T
	1098	G	G	A	G	G
	1300	G	G	***A***	G	G
	1341	G	G	G	G	A
VP1	1756	A	***C***	***C***	***C***	***C***
	1831	G	A	G	G	G
	1939	T	A	A	A	A
	1951	T	C	C	C	C
	2044	A	A	C	A	A
	2209	G	A	G	A	A
	2239	T	G	G	G	G
	2384	C	A	C	A	A
Intragenic	2669–71	AAC	AAC	***del 3 nt***	AAC	AAC
T-ag	2716–2721	GTGGGA	GTGGGA	GTGGGA	GTGGGA	del 6 nt
	2751	A	G	A	G	G
	2757	G	A	A	A	A
	2764–69	TGATTA	TGATTA	***del 6 nt***	TGATTA	TGATTA
	2768–73	TATGAT	TATGAT	TATGAT	TATGAT	***del 6 nt***
	2794	T	***ins 9 nt***	***ins 9 nt***	***ins 9 nt***	***ins 9 nt***
	2817	G	A	A	A	G
	2900–11					***del 12 nt***
	2907	T	A	T	A	T
	2912	C	***T***	C	C	C
	2918	C	C	***T***	C	C
	2939	C	C	C	C	***T***
	2950	A	A	***G***	A	***G***
	2951	T	***C***	T	***C***	T
	3117	T	C	C	C	C
	3120	C	C	C	C	T
	3755	A	G	G	G	G
	3761	G	G	G	G	A
	3930	C	C	T	C	C
	4071	T	A	T	A	A
	4110	C	T	C	T	C
	4299	C	T	C	T	C
	4642	G	T	G	T	G
t-ag	4750	G	G	A	G	G
	4839	C	***T***	***T***	***T***	***T***
	4849	C	T	C	T	C
	4853	G	G	***A***	G	G
	4879	C	***T***	***T***	***T***	***T***
	5164	C	A	C	C	C
	5209	C	T	C	C	C

Nucleotide differences among SV40 isolates SV40-776, Baylor, VA45-54, SVCPC/SVMEN, and SVPML-1. The noncoding regions listed are the regulatory region and the intragenic space between VP1 and large T-ag. The coding regions include VP1, the unique region of VP2 (VP2), the common region of VP2 and VP3 (VP2/3), and the unique regions of large T-ag (T-ag) and small t-ag (t-ag). All numbers given are according to the SV40-776 numbering system. Nucleotide changes shown in bold and underscored would result in a predicted amino acid change. The sequence of SVPML-1 is not complete; the region between nucleotides 334 and 619, which encodes a portion of the agnoprotein, and the nucleotide at position 2163 are ambiguous.

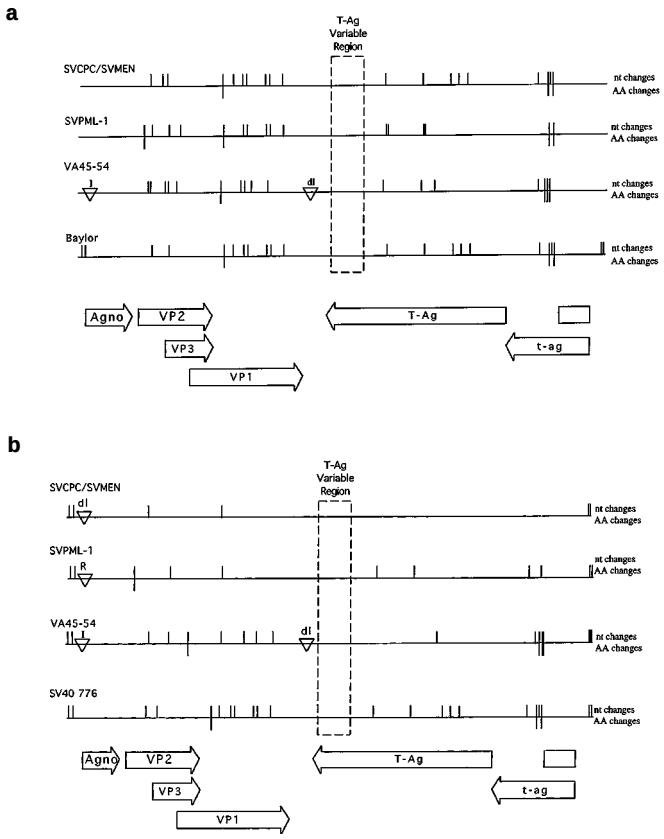


Figure 1 Schematic comparison of the genomes of SVCPC/SVMEN, SVPML-1, and VA45-54 to (a) SV40-776 or (b) Baylor strains of SV40. The horizontal lines represent the nucleotide sequences. Vertical lines above the horizontal lines represent nucleotide changes, and those below represent predicted amino acid changes. Inverted triangles represent either an insertion (I), a rearrangement (R), or a deletion (dl). The dashed line rectangle indicates the variable domain at the C-terminus of the large T-ag; changes in that region are not shown. The arrows at the bottom of the diagram represent the proteins encoded by the SV40 genome. The direction of the arrow indicates the direction of transcription.

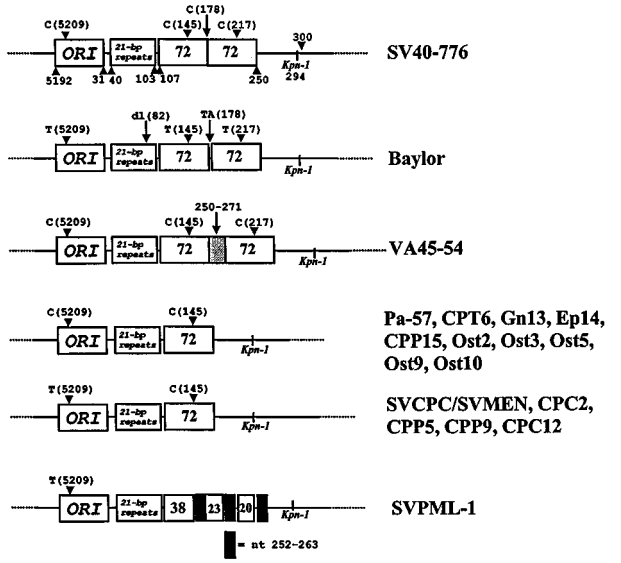


Figure 2 Regulatory region DNA sequence profiles of SV40 viral isolates and tumor-associated DNAs. The viral origin of DNA replication spanning nucleotides 5192 to 31 is represented by the box labeled ORI; the 21 bp repeat region between nucleotides 40 and 103 is shown; the boxed number 72 is the 72 bp sequence within the enhancer region that is duplicated in some laboratory-adapted strains, with the rearranged enhancer region of SVPML-1 represented with boxes labeled 38, 23, 20 and a shaded box referring to nucleotides 252–263. Nucleotide numbers are based on that of SV40-776. The identification of each viral isolate and tumor-associated DNA is given in Table 2.

SVCPC/SVMEN (Table 1). This nucleotide change, relative to SV40-776, would not result in a predicted change at the amino acid level.

Regulatory regions of SV40 isolates and tumor-associated DNAs

The sequence and organization of the regulatory regions of several viral isolates and tumor-associated DNAs are summarized schematically (Figure 2). Laboratory-adapted strains of SV40 typically contain two 72-basepair (bp) enhancer elements, as shown for strains SV40-776 (Fiers *et al*, 1978; Reddy *et al*, 1978), Baylor (Lednicky *et al*, 1995b; Lednicky and Butel, 1997), and VA45-54 (Lednicky and Butel, 1997). [The enhancer elements of SV40 are considered to be between nucleotides 107 and 300 of SV40-776 (Salzman *et al*, 1986).] There are slight sequence differences at the joint between the two 72 bp elements that distinguish each strain. Isolate SVPML-1 contains a more complex rearrangement in the enhancer region (Martin, 1989; Martin and Li, 1991). These isolates display a polymorphism at nucleotide 5209 (C or T) in the SV40 *ori*. Although human isolates SVPML-1, SVCPC and SVMEN contain the T at this position, so does the Baylor strain, recovered from monkeys, so this change relative to strain 776 is not a marker of human-specific strains of SV40.

In contrast to the duplication present in the regulatory region of the laboratory strains, natural isolates of SV40 from both monkeys (Ilyinskii *et al*, 1992) and humans (Lednicky *et al*, 1995a, 1997) usually have only one 72 bp element, an arrangement referred to as 'archetypal'. We have recently demonstrated that in low-passage, uncloned virus stocks of two laboratory strains (Baylor, VA45-54) both the laboratory-type and archetypal versions of the regulatory region could be detected (Lednicky and Butel, 1997). Of eight brain tumor- and five bone tumor-associated viral DNAs analyzed, all contained archetypal-length regulatory regions (Figure 2). Both polymorphisms at nucleotide 5209 were represented among the human tumor sequences. All five osteosarcoma-associated DNAs

contained a C at this position, but the sample size studied was too small to ascribe any tissue-specific significance to the observation.

T-ag variable regions of SV40 isolates and tumor-associated DNAs

Most of the large T-ag sequence was absolutely conserved among viral isolates, with the few nucleotide differences noted not affecting the predicted amino acid sequence of the protein (Table 1, Figure 1). However, we have previously identified a variable domain at the extreme C-terminus of SV40 T-ag, defined as the region from amino acid 622 through the end of the molecule (Stewart *et al*, 1996), and this region did show changes among viral isolates. The sequences in this region were compared for SV40 laboratory strains of monkey origin, SV40 isolates from humans, and SV40 DNAs associated with human brain tumors and human osteosarcomas (Figure 3). These comparisons were

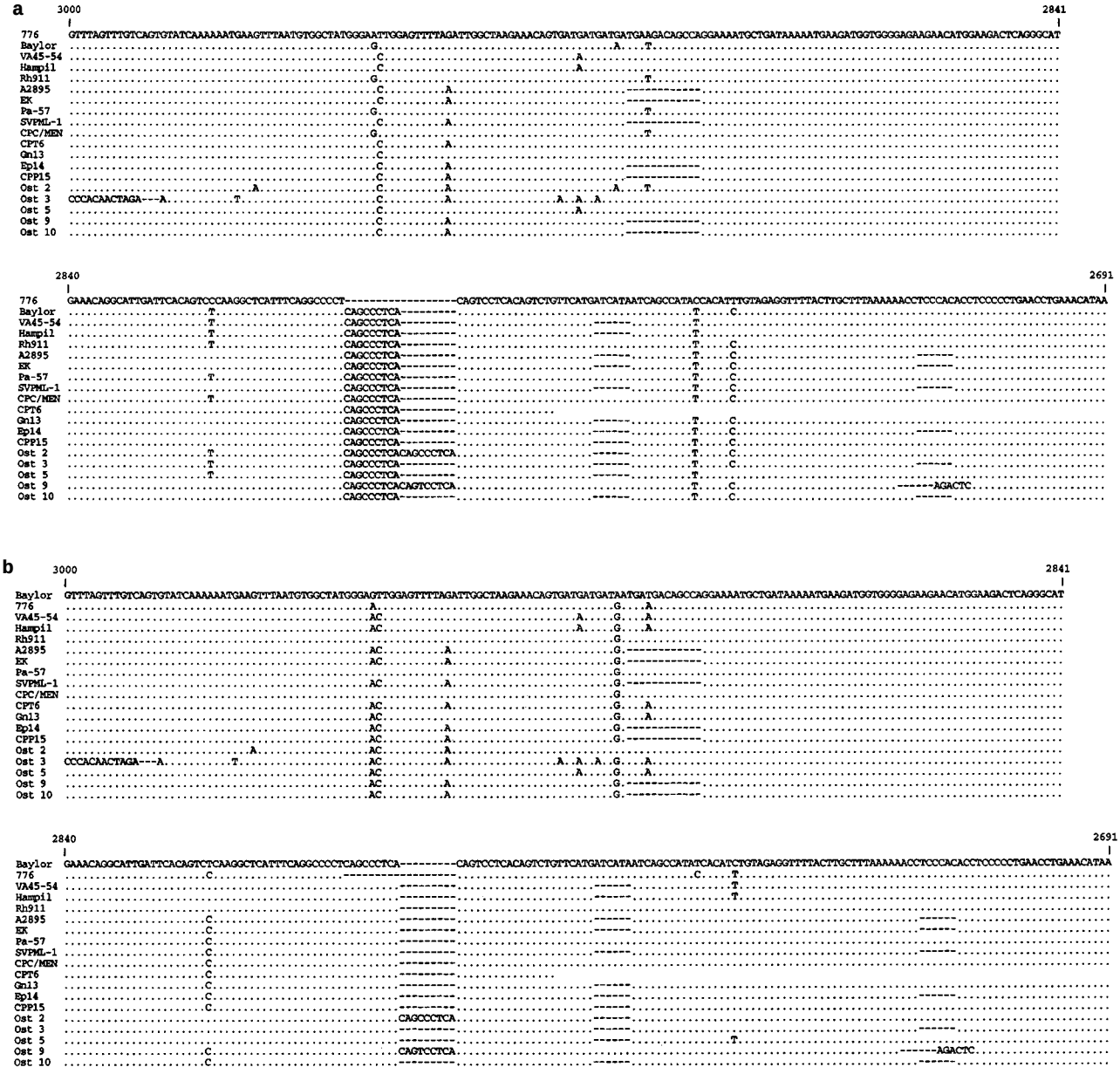


Figure 3 Nucleotide sequences of T-ag C-terminal variable domains of SV40 isolates and tumor-associated sequences compared to (a) SV40-776 or (b) Baylor strain. The sequence given is for the coding strand. The numbering is according to the system for SV40-776. Dots indicate identity with the top-most sequence given. Dashes indicate a deletion. Sources for all virus sequences are given in Table 2. Two different SV40 strains were recovered from a low-passage stock of Rh911, only one of which is reported here. The details of these isolations will be reported elsewhere.

made both to the reference virus SV40-776 (Figure 3a) and to the Baylor laboratory strain (Figure 3b) because, as noted above, many isolates were more similar to Baylor than to strain 776. [The alignments shown here differ slightly from our earlier reports (Lednicky *et al*, 1995a, 1997; Stewart *et al*, 1996) in the area of insertions near residue 674. We have arbitrarily chosen to present the alignments that have the smallest effects on the predicted amino acid sequence of T-ag.] The sequences in the variable region of T-ag for the independent viral isolates and the tumor-associated DNAs all differed somewhat from those of both the SV40-776 and Baylor strains. Although many of the sequences were unique, we identified some sequences that were shared by independent sources. One such group was formed by T-ag sequences of VA45-54, Hampil, and Ost5; a second group comprised Pa-57, Rh911, and SVCPC/SVMEN; and a final group contained A2895, SVPML-1, Ep14, Ost10 and EK. In addition, the sequence in CPP15 resembled that of the latter group, lacking only the deletion of residues 700–701.

A close inspection of the variable region sequences (Figure 3a) suggests that SV40-776 at some time in the past may have undergone a deletion of nine bases between nucleotides 2805 and 2806. All the other viral sequences analyzed had insertions at this position. It appears that in this region and at several other positions the most common sequence is not the one exhibited by SV40-776, but the one typified by Baylor (, Figure 3b). The bone tumor-associated sequences appeared more diverse than those found in human brain tumors. Both Ost2 and Ost9 appeared to have a duplication of 9 nucleotides in the region of SV40-776 nucleotide 2805/2806, with Ost2 having a duplication of the region just upstream and Ost9 having a duplication of the region just downstream. Osteosarcoma Ost3 is notable as it had a unique set of changes at the N-terminal portion of the variable region (immediately downstream of nucleotide 3000), making this sequence the most distinctly different of any variations observed to date. [The region just upstream of these changes was sequenced and was identical to that of SV40-776 (Lednicky *et al*, 1997).]

A number of the nucleotide changes listed in Figure 3 would cause predicted amino acid changes in the large T-ag polypeptide (Figure 4). The T-ag C-terminal sequences were compared to those of both SV40-776 (Figure 4) and the Baylor strain (Figure 4b). Both schematics present isolates of monkey kidney origin in the left-hand column, human brain isolates and brain tumor-associated sequences in the center column, and human osteosarcoma-associated sequences in the right-hand column. Identical sequences are lined up horizontally in boxes highlighted with bold lines. The comparison to the reference strain 776 emphasizes the extent of diversity among the different sequences analyzed. The comparison to the Baylor strain reduces the extent of apparent diversity, but serves to highlight that amino acid variation exists in this region of T-ag, regardless of the origin of the sequence. It does not appear that any T-ag C-terminal amino acid patterns were specifically associated with a given species of origin (monkey or human), or with a particular tissue type from which the sequences were recovered (kidney, brain, or bone). In fact, an identical T-ag sequence was found in all three tissue types (e.g., A2895, SVPML-1, Ep14, EK, and Ost10) (Figure 4). Other examples of identical T-ag C-terminal sequences from different sources included (i) VA45-54/Hampil (monkey kidney) and Ost5 (human bone), and (ii) Pa-57/Rh911 (monkey kidney) and SVCPC/SVMEN (human brain). These similarities do not indicate that the infecting viruses were necessarily the same, however, as representatives within a given T-ag variable domain group may have different regulatory regions (Figure 2). The data summarized in Figure 4 also reveal that the diversity within the T-ag variable domain appears to cluster into two subdomains, one area encompassing residues 623–640 and the second, residues from 674 to 702.

Discussion

This report summarizes an extensive sequence analysis of SV40 viral isolates from monkeys and humans and a comparison to SV40 DNAs found associated with human brain and bone tumors. Until now, the reference virus SV40-776 was the only isolate to be completely sequenced. We report here the completed sequences of human isolates SVCPC and SVMEN and of monkey isolates Baylor and VA45-54. In addition, the sequence of another human isolate, SVPML-1, is very nearly complete. The viral sequences were highly conserved, but differences were noted among the isolates. We have retained the numbering system based on SV40-776 to simplify comparisons with the large body of SV40 literature.

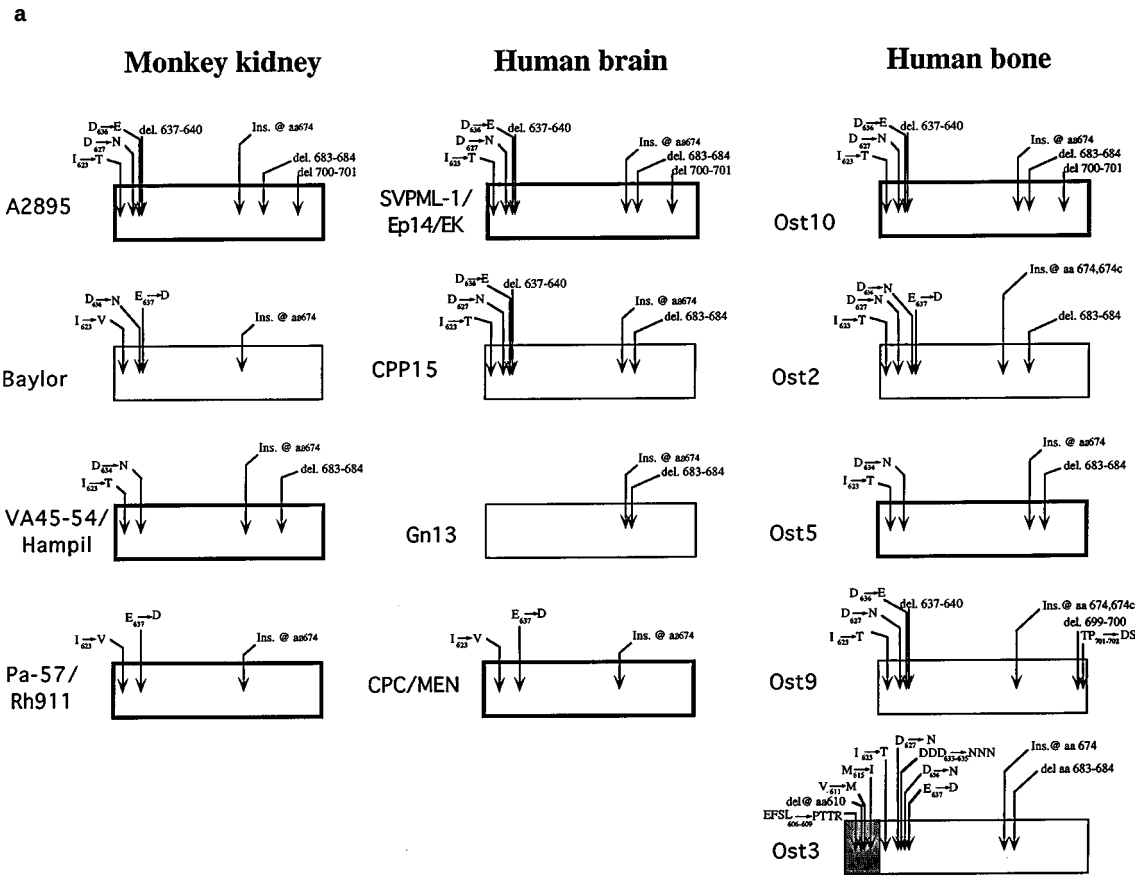
The newly sequenced isolates all differed somewhat from SV40-776, but relatively few nucleotide changes were found in the structural proteins, only a very few of which would cause a change in the predicted amino acid sequence. All of the viruses analyzed (SVCPC/SVMEN, SVPML-1, Baylor, and VA45-54) contained a change at nucleotide 1756, relative to SV40-776, which would result in a change in amino acid 86 of VP1. Although VP1 is the major capsid protein, and presumably the target of SV40 neutralizing antibodies, little information is available concerning the epitopes of SV40 VP1. However, amino acid 86 occurs in the loop between the B and C β sheets predicted in the three-dimensional structure of VP1 (Liddington *et al*,

1991). From studies of JC virus, it has been speculated that the BC loop is an antigenic region of polyomavirus VP1 and that changes in this loop may result in epitope changes across virus strains (Chang *et al*, 1996). Although different serotypes of SV40 have not been reported, these data suggest that the possibility may need to be considered. If the variation at amino acid 86 were to make an antigenic difference, that would influence the design and interpretation of serology studies aimed at determining the frequency and distribution of SV40 antibodies in human populations.

One unexpected result was the absolute conservation of the nucleotide sequence of the agnoprotein coding region. Agnoprotein has been reported to be a maturation, scaffolding protein (Barkan and Mertz, 1981; Jay *et al*, 1981; Cole, 1996) which is not absolutely required for virus replication in tissue culture. However, the strict conservation of the nucleotide sequence observed here raises the possibility that there may be some functional role for the DNA in this region during natural infections.

The majority of the genomic differences observed among SV40 isolates occurred in the previously identified large T-ag variable region. It can be seen at the nucleotide level (Figure 3) and, more readily at the amino acid level (Figure 4), that the differences in the variable region were not evenly distributed, but actually grouped into two subdomains. The significance of this subdomain grouping is unclear at this time. Large T-ag is a multifunctional protein with many well-defined domains. Part of the most C-terminal variable subdomain partially overlaps the host range/adenovirus helper domain (Pipas, 1985; Tornow *et al*, 1985; Fanning, 1992 and references therein; Stewart *et al*, 1996). It has been suggested that this region of large T-ag plays an undefined role in addition of VP1 during virion assembly (Spence and Pipas, 1994). There is also the possibility that this domain of T-ag may facilitate the growth of virus in inappropriate tissues in patients, perhaps leading eventually to disease development. The 'helper domain' of T-ag enables the replication of human adenoviruses in monkey kidney cells, but it is not known if this domain might affect the replication of other viruses in specific human tissues. It will be informative to determine if differences in the large T-ag C-terminal variable region affect particle morphogenesis or tissue tropism.

One goal of this study was to determine if any host-specific features defined different virus isolates. We hypothesized that such a feature, if it existed, would most likely lie in the large T-ag



variable domain or, perhaps, in the regulatory region. However, we did not detect any sequences that could be considered human-specific or monkey-specific. In fact, there were several examples of a particular large T-ag variable region being shared by both monkey viruses and human-associated sequences (e.g., Figure 4, VA45-54, Hampil and Ost5). We also could detect no tissue-specific differences among viral sequences. All of the monkey isolates came originally from kidney tissue (Table 2). There were large T-ag variable region sequences shared by a virus from this tissue (A2895) with sequences from human brain (SVPML-1, Ep14) and from a human osteosarcoma (Ost10). That T-ag C-terminal sequence was the only one recovered from multiple human tumors. Whether it is, in fact, more common among human disease or just fortuitously over-represented in our small sample size is unknown. All the regulatory regions analyzed in the human tumor samples were determined to be archetypal in structure, but there were no significant differences from monkey isolates. Therefore, no host-specific or tissue-specific markers could be identified. SV40 sequences have also been found associated with human mesotheliomas (Carbone *et al*, 1994; Cristaudo *et al*, 1995; Pepper *et al*, 1996); it will be interesting to determine if the SV40 DNAs from this additional tissue type display T-ag variable region sequences similar to those discussed here. Recent SV40 isolates from monkeys have been reported (Ilyinskii *et al*, 1992). The T-ag variable region sequences analyzed in that study were similar to those discussed here, but none were identical. SV40 has traditionally been considered to be a monkey virus, because of the circumstances of the early isolations (Sweet and Hilleman, 1960; Eddy *et al*, 1962). However, there is serological evidence that the virus may have been circulating in the human population for some time (Shah *et al*, 1971, 1972; Zimmermann *et al*, 1983; Geissler *et al*,

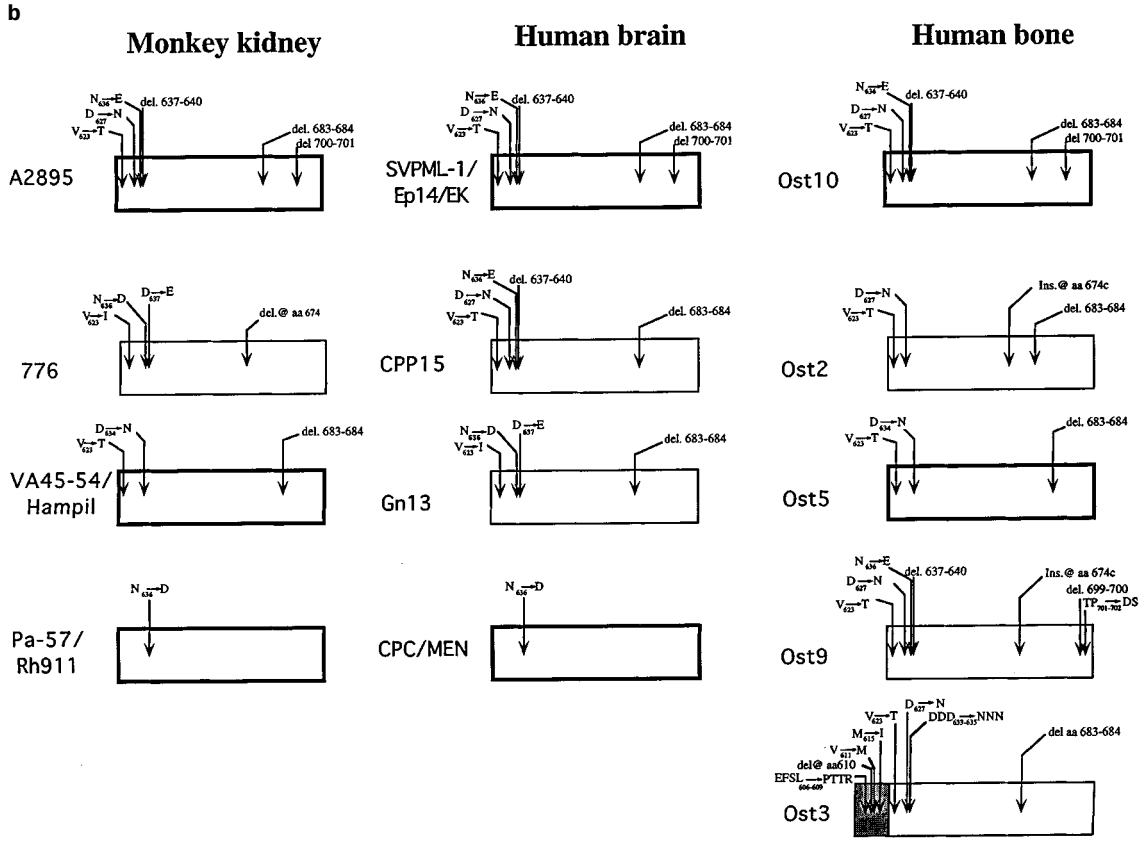


Figure 4 Schematic comparison of the T-ag C-terminal variable domain amino acid sequences of SV40 isolates and tumor-associated sequences to (a) SV40-776 or (b) Baylor strain. The rectangular boxes represent the T-ag C-terminal region from amino acid 622 to 708. Laboratory strains isolated from monkey kidney cells are shown in the left-hand column; human brain isolates and brain-tumor-associated sequences are in the center column, and human osteosarcoma-associated sequences are in the right-hand column. Viral strains and DNA sequences are identified to the left of each box. Sources for the viral sequences are given in Table 2. The nucleotide sequences are shown in Figure 3. Identical T-ag amino acid sequences in DNAs from different tissues are lined up horizontally and the boxes encased in bold lines. The numbering is according to the system for SV40-776; insertions in the region of amino acid 674 are designated with 674 plus a lower-case letter. Arrows indicate the position and type of amino acid changes. The shaded area of the Ost3 box indicates the region immediately upstream of T-ag residue 622.

Table 2 SV40 sequences analyzed

Virus or sequence designation	*Infectious virus isolate*	*Origin*				*Reference or database accession no.*
		Year	*Host species*	*Tissue*	*Sample*	
SV40-776	+	1960	Monkey (green or cynomolgus)	Kidney	Adenovirus type 1 seed stock	Sweet and Hilleman, 1960; Buckler and Salzman, 1986
Baylor	+	1961	Monkey	Kidney	Type 2 Sabin poliovaccine prepared in 1956	Melnick and Stinebaugh, 1962; Stinebaugh and Melnick, 1962; Lednicky *et al*, 1995b; Stewart *et al*, 1996; Lednicky and Butel, 1997
VA45-54	+	1960	Monkey (green)	Kidney	Cell culture, uninoculated	Sweet and Hilleman, 1960; Girardi *et al*, 1962; Stewart *et al*, 1996; Lednicky and Butel, 1997
Hampil	+	≈1960	Monkey (rhesus)	Kidney	Cell culture, uninoculated	Hampil, 1973; Tevethia and Butel, 1973/74
Rh911	+		Monkey (rhesus)	Kidney	Cell culture, uninoculated	Girardi, 1965
A2895	+		Monkey (rhesus)	Kidney	Tumor from hamster injected with kidney cell culture	Eddy *et al*, 1961, 1962; Ilyinskii *et al*, 1992
Pa-57	+		Monkey (patas)	Kidney	Cell culture, uninoculated	Hsiung and Gaylord, 1961; M99362
SVPML-1 (EK)*	+	1970	Human	Brain	Cell culture, PML brain	Weiner *et al*, 1972; Martin, 1989; Ilyinskii *et al*, 1992; Stewart *et al*, 1996
EK*	+	1970	Human	Brain	Cell culture, PML brain	Weiner *et al*, 1972; M99364
SVCPC	+	1995	Human	Brain	Choroid plexus carcinoma	Lednicky *et al*, 1995a; Stewart *et al*, 1996
SVMEN (SV303)	+	1984	Human	Brain	Cloned directly from meningioma	Krieg and Scherer, 1984; Martin, 1989; Stewart *et al*, 1996
CPT6	0		Human	Brain	Choroid plexus tumor	Lednicky *et al*, 1995a
Gn13	0		Human	Nervous system	Ganglioneuroma	Lednicky *et al*, 1995a
Ep14	0		Human	Brain	Ependymoma	Lednicky *et al*, 1995a
CPP15	0		Human	Brain	Choroid plexus papilloma	Lednicky *et al*, 1995a
Ost2	0		Human	Bone	Osteosarcoma	Lednicky *et al*, 1997
Ost3	0		Human	Bone	Osteosarcoma	Lednicky *et al*, 1997
Ost5	0		Human	Bone	Osteosarcoma	Lednicky *et al*, 1997
Ost9	0		Human	Bone	Osteosarcoma	Lednicky *et al*, 1997
Ost10	0		Human	Bone	Osteosarcoma	Lednicky *et al*, 1997

*We sequenced plasmid pSVPML-1 (EK); the partial sequence obtained from GenBank is listed as a separate entry ('EK') to indicate an independent sequence analysis of the virus isolate.

1985). That serology, together with the fact that no host-specific markers have been detected at the sequence level, suggests that SV40 should be thought of as a primate virus rather than solely as a monkey virus.

Natural isolates of SV40 from both monkeys and humans were found to usually have a single 72 bp element in the regulatory region of the viral genome, in contrast to laboratory-adapted strains which typically have duplications in that region. The duplication demonstrably improves growth of the virus in tissue culture (Lednicky *et al*, 1995b). Animal studies will be required to analyze the growth or survival advantage in intact hosts of virus containing a single 72 bp element. We speculate that more inefficient replication by viral archetypes may promote establishment and maintenance of persistent infections by failing to elicit strong immune responses by the infected host.

We conclude from these data that, although the nucleotide sequence of SV40 is remarkably conserved, different strains of the virus do exist. These strains can be distinguished on the basis of the sequences of the regulatory region and the large T-ag variable domain. These regions are not merely mutational hotspots, as cloned archetypal viruses followed through at least 14 serial passages in monkey kidney tissue culture cells failed to acquire mutations in either site (Lednicky and Butel, 1997). Further, we analyzed low-passage stocks of two laboratory strains of SV40 and showed that the stocks contained a mixture of viruses having archetypal or duplicated regulatory regions (Lednicky and Butel, 1997). Once the laboratory (duplicated enhancer) version emerged during tissue culture passage, no other changes ensued in the structure of the regulatory region. There were no changes in the sequence of a given viral T-ag carboxy terminus over many years time. Therefore, we suggest that differences in these locations reflect the existence of viral strains and can serve to distinguish virus isolates.

It appears that isolates with the same regulatory region may have different T-ag C-terminal regions (e.g., CPP15 *versus* Ost9) and *vice versa* (e.g., Pa-57/Rh911 *versus* SVCPC/SVMEN). This raises the question of how such variety might occur. One possibility is convergent evolution. Perhaps a particular large T-ag variable region allows a virus strain to prosper in a particular niche in the natural host. Different viruses might then evolve the same variable region to exploit that environment. The second possibility is that strains containing different regulatory regions and T-ag variable regions may undergo genetic recombination, resulting in a mix-and-match set of genomic sequences. Which of these possibilities may describe the situation for SV40 is a subject for further study.

The source of the virus DNA detected in the human tumors is not known. Although millions of people were exposed to SV40 in unknowingly contaminated virus vaccines (Shah and Nathanson, 1976), many of the tumor samples analyzed came from patients too young to have received a contaminated vaccine. This suggests that SV40 is circulating in the human population, perhaps similar to the known human polyomaviruses JCV and BKV (Frisque, 1994; Shah, 1996). It is possible the presence of SV40 in humans predated the use of contaminated vaccines, but the widespread exposures associated with the poliovaccines may have distributed the virus more broadly. The diversity of SV40 strains detected in the human samples argues against a single source of infection.

The significance of the SV40 DNA found in association with human tumors remains to be established. It is possible that the virus is present merely as a passenger in the tumor tissue, exerting no biological effects on the cells or that some other cofactor must be present and functional in order for the virus to affect cell phenotypes. However, the known oncogenic potential of the virus makes that association cause for concern. Additional studies are warranted to determine the medical significance of the relationship of SV40 DNA to human tumors.

Materials and methods

Sequence analyses

The viral isolates and human tumor-associated viral DNA sequences analyzed in this study are listed in Table 2. The origin of each isolate or sequence, along with any published sequence information, is referenced. Regulatory region and early region sequences have been reported for the Baylor strain (Lednicky *et al*, 1995a,b; Stewart *et al*, 1996), SVCPC (Lednicky *et al*, 1995a; Stewart *et al*, 1996), SVMEN (Lednicky *et al*, 1995a; Stewart *et al*, 1996), SVPML-1 (Martin, 1989; Lednicky *et al*, 1995a; Stewart *et al*, 1996), and EK (Ilyinskii *et al*, 1992). The late regions of these viral isolates were sequenced from viral genomes cloned into either pBR322 or pUC19. Sequencing was performed on an ABI automatic sequencer in the Department of Human and Molecular Genetics, Baylor College of Medicine. The early region and regulatory region sequences of isolate VA45-54 have also been reported (Stewart *et al*, 1996; Lednicky and Butel, 1997); the remaining sequence for VA45-54 was contributed by MJ Tevethia. Also reported are the regulatory region and large T-ag variable region sequences associated with human tumors (Lednicky *et al*, 1995a, 1997) and for strain A2895 (Ilyinskii *et al*, 1992). T-ag variable region sequences for SVPML-1 (EK) and Pa-57 were retrieved from GenBank; the T-ag variable region of SVPML-1 was verified using manual sequencing (Lednicky *et al*, 1995a). The variable region sequences of Hampil and Rh911 isolates were determined by cloning virus from low-passage stocks with subsequent sequencing of several clones; the details of this cloning and sequencing will be reported elsewhere.

Computer programs

Computer programs for DNA sequence analysis, alignment, and translation were from the Wisconsin Package, version 8.0, September 1994 (Genetics Computer Group, Madison, Wisc.) and by Sequencher™ software from the Gene Codes Corp. (Ann Arbor, Mich.). The Sequencher program aligns fragments using a search-and-compare algorithm subject to parameters set by the user. Fragments crossing the plasmid cloning site were aligned manually. All sequence discrepancies were resolved by inspecting the electropherograms generated for each fragment. The full-length contigs were then compared to the sequence of SV40-776 using Sequencher.

Acknowledgements

The work reported in this study was supported in part by National Research Service Award CA09197 from the National Cancer Institute. We thank MJ Tevethia for providing sequence information for strain VA45-54.

References

Barkan A, Mertz JE (1981). DNA sequence analysis of simian virus 40 mutants with deletions mapping in the leader region of the late mRNAs: mutants with deletions similar in size and position exhibit varied phenotypes. *J Virol* **37:** 730–737.

Bergsagel DJ, Finegold MJ, Butel JS, Kupsky WJ, Garcea RL (1992). DNA sequences similar to those of simian virus 40 in ependymomas and choroid plexus tumors of childhood. *New Engl J Med* **326:** 988–993.

Buckler CE, Salzman NP (1986). Annotated nucleotide sequence and restriction site lists for selected papovavirus strains. In: *The Papovaviridae*. Volume 1: *The Polyomaviruses*.Salzman NP (ed). Plenum Press: New York, pp 379–387.

Butel JS (1994). Simian virus 40. In: *Encyclopedia of Virology*. Webster RG, Granoff A (eds).Academic Press: San Diego, vol 3, pp 1322–1329.

Carbone M, Pass HI, Rizzo P, Marinetti MR, Di Muzio M, Mew DJY, Levine AS, Procopio A (1994). Simian virus 40-like DNA sequences in human pleural mesothelioma.*Oncogene* **9:** 1781–1790.

Carbone M, Rizzo P, Procopio A, Giuliano M, Pass HI, Gebhardt MC, Mangham C, Hansen M, Malkin DF, Bushart G, Pompetti F, Picci P, Levine AS, Bergsagel JD, Garcea RL (1996). SV40-like sequences in human bone tumors.*Oncogene* **13:** 527–535.

Chang D, Liou ZM, Ou WC, Wang KZ, Wang M, Fung CY, Tsai RT (1996). Production of the antigen and the antibody of the JC virus major capsid protein VP1.*J Virol Meth* **59:** 177–187.

Cole CN (1996). Polyomavirinae: the viruses and their replication. In: *Fields Virology*, 3rd ed. Fields BN *et al* (eds). Lippincott-Raven: Philadelphia, pp 1997–2025.

Cristaudo A, Vivaldi A, Sensales G, Guglielmi G, Cianca E, Elisei R, Ottenga F (1995). Molecular biology studies on mesothelioma tumor samples: preliminary data on H-Ras, p21 and SV40.*J Environ Pathol Toxicol Oncol* **14:** 29–34.

Eddy BE, Borman GS, Berkeley WH, Young RD (1961). Tumors induced in hamsters by injection of rhesus monkey kidney cell extracts. *Proc Soc Exp Biol Med* **107:** 191–197.

Eddy BE, Borman GS, Grubbs GE, Young RD (1962). Identification of the oncogenic substance in rhesus monkey kidney cell cultures as simian virus 40. *Virology* **17:** 65–75.

Fanning E (1992). Simian virus 40 large T antigen: the puzzle, the pieces, and the emerging picture. *J Virol* **66:** 1289–1293.

Fiers W, Contreras R, Haegeman G, Rogiers R, Van de Voorde A, Van Heuverswyn H, Van Herreweghe J, Volckaert G, Ysebaert M (1978). Complete nucleotide sequence of SV40 DNA.*Nature* **273:** 113–120.

Frisque RJ (1994). JC and BK viruses. In: *Encyclopedia of Virology*. Webster RG, Granoff A (eds).Academic Press: San Diego,vol 2, pp 752–757.

Geissler E (1990). SV40 and human brain tumors. *Prog Med Virol* **37:** 211–222.

Geissler E, Konzer P, Scherneck S, Zimmermann W (1985). Sera collected before introduction of contaminated polio vaccine contain antibodies against SV40. *Acta Virol* **29:** 420–423.

Girardi AJ (1965). Prevention of SV40 virus oncogenesis in hamsters, I. Tumor resistance induced by human cells transformed by SV40. *Proc Natl Acad Sci USA* **54:** 445–451.

Girardi AJ, Sweet BH, Slotnick VB, Hilleman MR (1962). Development of tumors in hamsters inoculated in the neonatal period with vacuolating virus, SV40. *Proc Soc Exp Biol Med* **109:** 649–660.

Hampil B (1973). Personal communication.

Hsiung GD, Gaylord WH Jr (1961). The vacuolating virus of monkeys. I. Isolation, growth characteristics, and inclusion body formation. *J Exp Med* **114:** 975–985.

Ilyinskii PO, Daniel MD, Horvath CJ, Desrosiers RC (1992). Genetic analysis of simian virus 40 from brains and kidneys of macaque monkeys. *J Virol* **66:** 6353–6360.

Jay G, Nomura S, Anderson CW, Khoury G (1981). Identification of the SV40 agnogene product: a DNA binding protein. *Nature* **291:** 346–349.

Krieg P, Amtmann E, Jonas D, Fischer H, Zang K, Sauer G (1981). Episomal simian virus 40 genomes in human brain tumors. *Proc Natl Acad Sci USA* **78:** 6446–6450.

Krieg P, Scherer G (1984). Cloning of SV40 genomes from human brain tumors. *Virology* **138:** 336–340.

Lednicky JA, Butel JS (1997). Tissue culture adaptation of natural isolates of simian virus 40: changes occur in viral regulatory region but not in carboxy-terminal domain of large T-antigen. *J Gen Virol* **78:** 1697–1705.

Lednicky JA, Garcea RL, Bergsagel DJ, Butel JS (1995a). Natural simian virus 40 strains are present in human choroid plexus and ependymoma tumors. *Virology* **212:** 710–717.

Lednicky JA, Stewart AR, Jenkins JJ III, Finegold MJ, Butel JS (1997). SV40 DNA in human osteosarcomas shows sequence variation among T-antigen genes. *Int J Cancer*, **72:** 791–800.

Lednicky JA, Wong C, Butel JS (1995b). Artificial modification of the viral regulatory region improves tissue culture growth of SV40 strain 776. *Virus Res* **35:** 143–153.

Liddington RC, Yan Y, Moulai J, Sahli R, Benjamin TL, Harrison SC (1991). Structure of simian virus 40 at 3.8 Å resolution. *Nature* **354:** 278–284.

Martin JD (1989). Regulatory sequences of SV40 variants isolated from patients with progressive multifocal leukoencephalopathy. *Virus Res* **14:** 85–94.

Martin JD, Li P (1991). Comparison of regulatory sequences and enhancer activities of SV40 variants isolated from patients with neurological diseases. *Virus Res* **19:** 163–172.

Martini F, Iaccheri L, Lazzarin L, Carinci P, Corallini A, Gerosa M, Iuzzolino P, Barbanti-Brodano G, Tognon M (1996). SV40 region and large T antigen in human brain tumors, peripheral blood cells, and sperm fluids from healthy individuals. *Cancer Res* **56:** 4820–4825.

Meinke W, Goldstein DA, Smith RA (1979). Simian virus 40-related DNA sequences in a human brain tumor. *Neurology* **29:** 1590–1594.

Melnick JL, Stinebaugh S (1962). Excretion of vacuolating SV-40 virus (papova virus group) after ingestion as a contaminant of oral poliovaccine. *Proc Soc Exp Biol Med* **109:** 965–968.

Pepper C, Jasani B, Navabi H, Wynford-Thomas D, Gibbs AR (1996). Simian virus 40 large T antigen (SV40LTAg) primer specific DNA amplification in human pleural mesothelioma tissue. *Thorax* **51:** 1074–1076.

Pipas JM (1985). Mutations near the carboxyl terminus of the simian virus 40 large tumor antigen alter viral host range. *J Virol* **54:** 569–575.

Reddy VB, Thimmappaya B, Dhar R, Subramanian KN, Zain BS, Pan J, Ghosh PK, Celma ML, Weissman SM (1978). The genome of simian virus 40. *Science* **200:** 494–502.

Salzman NP, Natarajan V, Selzer GB (1986). Transcription of SV40 and polyoma virus and its regulation. In: *The Papovaviridae*. Vol. 1: *The Polyomaviruses*. Salzman NP (ed). Plenum Press: New York, pp 27–98.

Scherneck S, Rudolph M, Geissler E, Vogel F, Lübbe L, Wählte H, Nisch G, Weickmann F, Zimmermann W (1979). Isolation of a SV40-like papovavirus from a human glioblastoma. *Int J Cancer* **24:** 523–531.

Shah KV (1996). Polyomaviruses. In: *Fields Virology*, 3rd ed. Fields BN *et al* (eds). Lippincott-Raven: Philadelphia, pp 2027–2043.

Shah KV, McCrumb FR Jr, Daniel RW, Ozer HL (1972). Serologic evidence for a simian-virus-40-like infection of man. *J Natl Cancer Inst* **48:** 557–561.

Shah KV, Nathanson N (1976). Human exposure to SV40: review and comment. *Am J Epidemiol* **103:** 1–12.

Shah KV, Ozer HL, Pond HS, Palma LD, Murphy GP (1971). SV40 neutralizing antibodies in sera of US residents without history of polio immunization. *Nature* **231:** 448–449.

Soriano F, Shelburne CE, Gökcen M (1974). Simian virus 40 in a human cancer. *Nature* **249:** 421–424.

Spence SL, Pipas JM (1994). Simian virus 40 large T antigen host range domain functions in virion assembly. *J Virol* **68:** 4227–4240.

Stewart AR, Lednicky JA, Benzick US, Tevethia MJ, Butel JS (1996). Identification of a variable region at the carboxy terminus of SV40 large T-antigen. *Virology* **221:** 355–361.

Stinebaugh S, Melnick JL (1962). Plaque formation by vacuolating virus. *Virology* **16:** 348–349.

Sweet BH, Hilleman MR (1960). The vacuolating virus, S.V.40. *Proc Soc Exp Biol Med* **105:** 420–427.

Tevethia SS, Butel JS (1973/74). Induction of common transplantation antigen by various isolates of papovavirus SV40 and by virus rescued from transformed cells. *Intervirology* **2:** 200–205.

Tornow J, Polvino-Bodnar M, Santangelo G, Cole CN (1985). Two separable functional domains of simian virus 40 large T antigen: Carboxyl-terminal region of simian virus 40 large T antigen is required for efficient capsid protein synthesis. *J Virol* **53:** 415–424.

Weiner LP, Herndon RM, Narayan O, Johnson RT, Shah K, Rubinstein LJ, Preziosi TJ, Conley FK (1972). Isolation of virus related to SV40 from patients with progressive multifocal leukoencephalopathy. *N Engl J Med* **286:** 385–390.

Woloschak M, Yu A, Post KD (1995). Detection of polyomaviral DNA sequences in normal and adenomatous human pituitary tissues using the polymerase chain reaction. *Cancer* **76:** 490–496.

Zimmermann W, Scherneck S, Geissler E (1983). Quantitative determination of papovavirus IgG antibodies in sera from cancer patients, labworkers and several groups of control persons by enzyme-linked immunosorbent assay (ELISA). *Zbl Bakt Hyg I Abt Orig A* **254:** 187–196.