mains of TRF1 and TRF2, these structural variations emphasize that the TRFH domain is a versatile framework for interactions with different proteins.

The crystal structure of the $TRF2_{TRFH}$-$Apollo_{TBM}$ complex is corroborated by mutagenesis. Mutations of the conserved hydrophobic residues of Apollo (F504, L506, and P508) or TRF2 (*F120*) completely abolished the interaction both in vitro and in vivo (Fig. 4, F and G). We further assayed the cellular localization of wild-type and mutant Apollo by expressing hemagglutinin (HA)–tagged proteins in human telomerase reverse transcriptase (hTERT)–immortalized human BJ fibroblasts. Although wild-type Apollo showed the expected telomere localization, the L506E/P508A double mutant was distributed throughout the nucleoplasm with no obvious accumulation at telomeres (Fig. 4H). This result confirms the structural information and indicates that the binding of Apollo to the TRFH domain of TRF2 is required for the telomeric localization of Apollo.

We next asked whether other shelterin-associated proteins might contain the F/Y-X-L-X-P motif suggestive of an interaction with the TRFH domain of TRF1 or TRF2. We identified this motif in PinX1, originally identified as a TRF1-interacting protein in a yeast two-hybrid screen (*6*). An 11-residue fragment of PinX1 (R287-D-F-T-L-K-P-K-K-R-R297), referred to as $PinX1_{TBM}$, closely resembles $TIN2_{TBM}$ (fig. S12A), suggesting that it may bind to $TRF1_{TRFH}$ in the same fashion as does $TIN2_{TBM}$. ITC data confirmed the $TRF1_{TRFH}$-$PinX1_{TBM}$ interaction, whereas no measurable interaction was observed between $TRF2_{TRFH}$ and $PinX1_{TBM}$ (fig. S12B). Mutagenesis studies showed that PinX1-L291 and TRF1-*F142* are critical for the interaction, whereas PinX1-P293 is not (fig. S12C). These results are consistent with those of the $TRF1_{TRFH}$-$TIN2_{TBM}$ interaction (Fig. 2D) and indicate that PinX1, like TIN2, binds the TRFH domain of TRF1 but not TRF2. Protein sequence database searches showed many instances of telomere-associated proteins containing the F/Y-X-L-X-P motif (fig. S13). Future studies are needed to address whether this motif mediates the TRF1/TRF2 binding of these telomere-associated proteins in vivo.

Our results indicate that binding to the TRFH docking site involves the sequence F/Y-X-L-X-P in shelterin-associated proteins, which contacts the same molecular recognition surface of the TRFH domains of TRF1 and TRF2 with distinct specificities. Because TRF1 and TRF2 play different roles in telomere length homeostasis and telomere protection (*1*), we propose that the TRFH domains of TRF1 and TRF2 function as telomeric protein docking sites that recruit different shelterin-associated factors with distinct functions to the chromosome ends.

References and Notes

1. T. de Lange, *Genes Dev.* **19**, 2100 (2005).
2. M. van Overbeek, T. de Lange, *Curr. Biol.* **16**, 1295 (2006).
3. X. D. Zhu, B. Kuster, M. Mann, J. H. Petrini, T. de Lange, *Nat. Genet.* **25**, 347 (2000).
4. X. D. Zhu *et al.*, *Mol. Cell* **12**, 1489 (2003).
5. S. Smith, I. Giriat, A. Schmitt, T. de Lange, *Science* **282**, 1484 (1998).
6. X. Z. Zhou, K. P. Lu, *Cell* **107**, 347 (2001).
7. T. Nishikawa *et al.*, *Structure* **9**, 1237 (2001).
8. R. Court, L. Chapman, L. Fairall, D. Rhodes, *EMBO Rep.* **6**, 39 (2005).
9. B. Li, S. Oestreich, T. de Lange, *Cell* **101**, 471 (2000).
10. A. Bianchi, S. Smith, L. Chong, P. Elias, T. de Lange, *EMBO J.* **16**, 1785 (1997).
11. L. Fairall, L. Chapman, H. Moss, T. de Lange, D. Rhodes, *Mol. Cell* **8**, 351 (2001).
12. S. H. Kim, P. Kaminker, J. Campisi, *Nat. Genet.* **23**, 405 (1999).
13. C. Lenain *et al.*, *Curr. Biol.* **16**, 1303 (2006).
14. T. H. Lee, K. Perrem, J. W. Harper, K. P. Lu, X. Z. Zhou, *J. Biol. Chem.* **281**, 759 (2006).
15. S. H. Kim *et al.*, *J. Biol. Chem.* **279**, 43799 (2004).
16. J. Z. Ye *et al.*, *J. Biol. Chem.* **279**, 47264 (2004).
17. Materials and methods are available as supporting material on *Science* Online.
18. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.
19. P. Fotiadou, O. Henegariu, J. B. Sweasy, *Cancer Res.* **64**, 3830 (2004).
20. Coordinates and structure factor amplitudes have been deposited in the Protein Data Bank with access numbers 3BQO ($TRF1_{TRFH}$-$TIN2_{TBM}$), 3BU8 ($TRF2_{TRFH}$-$TIN2_{TBM}$), and 3BUA ($TRF2_{TRFH}$-$Apollo_{TBM}$).

We thank F. Wang and K. Wan for assistance. Work was supported by an NIH grant (to T.de L.) and an American Cancer Society Research Scholar grant and a Sidney Kimmel Scholar award (to M.L.). Use of Life Sciences Collaborative Access Team Sector 21 was supported by the Michigan Economic Development Corporation and the Michigan Technology Tri-Corridor (grant 085P1000817). Use of the Advanced Photon Source was supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under contract no. DE-AC02-06CH11357.

Supporting Online Material

www.sciencemag.org/cgi/content/full/1151804/DC1
Materials and Methods
SOM Text
Figs. S1 to S14
Table S1
References

16 October 2007; accepted 7 January 2008
Published online 17 January 2008;
10.1126/science.1151804
Include this information when citing this paper.

Clonal Integration of a Polyomavirus in Human Merkel Cell Carcinoma

Huichen Feng, Masahiro Shuda, Yuan Chang,* Patrick S. Moore*

Merkel cell carcinoma (MCC) is a rare but aggressive human skin cancer that typically affects elderly and immunosuppressed individuals, a feature suggestive of an infectious origin. We studied MCC samples by digital transcriptome subtraction and detected a fusion transcript between a previously undescribed virus T antigen and a human receptor tyrosine phosphatase. Further investigation led to identification and sequence analysis of the 5387–base-pair genome of a previously unknown polyomavirus that we call Merkel cell polyomavirus (MCV or MCPyV). MCV sequences were detected in 8 of 10 (80%) MCC tumors but only 5 of 59 (8%) control tissues from various body sites and 4 of 25 (16%) control skin tissues. In six of eight MCV-positive MCCs, viral DNA was integrated within the tumor genome in a clonal pattern, suggesting that MCV infection and integration preceded clonal expansion of the tumor cells. Thus, MCV may be a contributing factor in the pathogenesis of MCC.

Polyomaviruses have been suspected as potential etiologic agents in human cancer since the discovery of murine polyoma virus (MuPyV) by Gross in 1953 (*1*). However, although polyomavirus infections can produce tumors in animal models, there is no conclusive evidence that they play a role in human cancers (*2*). These small double-stranded DNA viruses [~5200 base pairs (bp)] encode a variably spliced oncoprotein, the tumor (T) antigen (*3*, *4*), and are divided into three genetically distinct groups: (i) avian polyomaviruses, (ii) mammalian viruses related to MuPyV, and (iii) mammalian polyomaviruses related to simian virus 40 (SV40) (*5*). All four known human polyomaviruses [BK virus (BKV), JCV, KIV, and WUV (*6*, *7*)] belong to the SV40 subgroup. In animals, integration of polyomavirus DNA into the host genome often precedes tumor formation (*8*).

Merkel cell carcinoma (MCC) is a neuroectodermal tumor arising from mechanoreceptor Merkel cells (Fig. 1A). MCC is rare, but its incidence has tripled over the past 2 decades in the United States to 1500 cases per year (*9*). It is one of the most aggressive forms of skin cancer; about 50% of advanced MCC patients

Molecular Virology Program, University of Pittsburgh Cancer Institute, University of Pittsburgh, 5117 Centre Avenue, Suite 1.8, Pittsburgh, PA 15213, USA.

*These authors contributed equally to this work. To whom correspondence should be addressed. E-mail: yc70@pitt.edu (Y.C.); psm9@pitt.edu (P.S.M.)

live 9 months or less. Gene expression profiling studies indicate that MCC may comprise two or more clinically similar diseases with distinct etiologies (*10*). Like Kaposi's sarcoma (KS), MCC occurs more frequently than expected among immunosuppressed transplant and AIDS patients (*11*). These similarities to KS, an immune-related tumor caused by KS-associated herpesvirus (*12*), raise the possibility that MCC may also have an infectious origin.

To search for viral sequences in MCC, we used digital transcriptome subtraction (DTS), a methodology we developed that can identify foreign transcripts by using human high-throughput cDNA sequencing data (*13*). We generated two cDNA libraries from a total of four anonymized MCC tumors. One library was prepared with the use of mRNA from a single tumor (MCC347), and the other was prepared with mRNA pooled from three tumors (MCC337, 343, and 346) to increase the likelihood of detecting rare viral sequences (table S1).

From these two libraries, we respectively pyrosequenced 216,599 and 179,135 cDNA sequences (~150 to 200 bp). These 395,734 cDNA sequences were trimmed with LUCY stringency equivalent to PHRED scores of 20 or higher (*14*). Copolymers of adenine or thymidine [poly(A) and poly (T), respectively], dust (low-complexity), human repeat, and primer adaptor sequences were then removed, leaving 382,747 sequences to form a high-fidelity (HiFi) data set. Of these, 380,352 (99.4%) aligned to human RefSeq RNA, mitochondrial, assembled chromosomes, or immunoglobulin sequences in National Center for Biotechnology Information (NCBI) databases. Of the remaining 2395 HiFi candidate sequences, one transcript (DTS1) from MCC347 cDNA aligned with high homology to African green monkey (AGM) lymphotropic polyomavirus (LPyV) and to human BK polyomavirus T antigen sequences. A second DTS transcript (DTS2) had no homology to deposited polyomavirus sequences but was subsequently identified by aligning HiFi candidates to the full-length viral genome (see below). These two sequences define a previously unknown human polyomavirus that we call Merkel cell polyomavirus (MCV or MCPyV) because of its close association with MCC.

Rapid amplification of cDNA ends (3′-RACE) extended DTS1 to three different cDNAs (Fig. 1B): One transcript terminated at a poly(A) site in the T antigen sequence, and two cDNAs read through this weak poly(A) site to form different length fusions with intron 1 of the human receptor tyrosine phosphatase type G gene (*PTPRG*) (GenBank:18860897) on chromosome 3p14.2. Viral integration at this site was confirmed by sequencing DNA polymerase chain reaction (PCR) products with the use of a viral primer and a human *PTPRG* primer. The same three RACE products were independently cloned from MCC348, a lymph node metastasis from the MCC347 primary tumor, indicating that this tumor was seeded from a single tumor cell already positive for the T antigen–PTPRG fusion transcript.

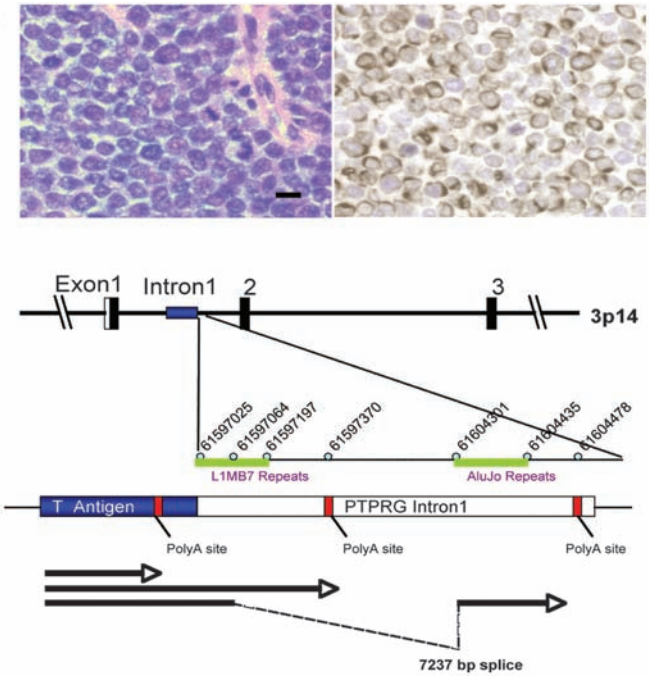

Fig. 1. (**A**) MCC is an aggressive skin cancer derived from Merkel mechanoreceptor cells that expresses neuroendocrine and perinuclear cytokeratin 20 markers, distinguishing it from other small round cell tumors (MCC349, left, hematoxylin and eosin; right, cytokeratin 20 staining, 40×. Scale bar represents 10 μm). (**B**) Discovery of Merkel cell polyomavirus transcripts in (MCC). 3′-RACE mapping of an MCC fusion transcript between the MCV T antigen and human *PTPRG*. A cDNA corresponding to a polyomavirus-like T antigen transcript was found by DTS analysis of MCC. This T antigen cDNA was extended by 3′-RACE to map three mRNA sequences (arrows), one of which terminates at a viral polyadenylation site and two of which extend into flanking human sequence and terminate in intron 1 of the human *PTPRG* gene on chromosome 3p14, indicative of viral DNA integration into the tumor cell genome. The two viral-human chimeric transcripts were generated by read-through of a weak polyadenylation signal in the viral T antigen gene. Identical RACE products were also sequenced from a lymph node metastasis of this primary tumor.

By viral genome walking, we sequenced the complete closed circular genome of MCV (5387 bp, prototype) from tumor MCC350. A second genome, MCV339 (5201 bp), was then sequenced by using MCV-specific primers. The sequences of MCV350 and MCV339 have GenBank accession numbers EU375803 and EU375804, respectively. Both viruses encode sequences with high homology to polyomavirus T antigen, VP1, VP2/3, and replication origin sequences (Fig. 2A). MCV has an early gene expression region [196 to 3080 nucleotides (nt)] containing the T antigen locus with large T and small T open reading frames and a late gene region containing VP1 and VP2/3 open reading frames between 3156 and 5118 nt. The T antigen locus has features conserved with other polyomavirus T antigens, including cr1, DnaJ, pRB1-binding Leu-X-Cys-X-Glu (LXCXE) motif, origin-binding, and helicase/adenosine triphosphatase (ATPase) domains. Mutations in the C terminus of MCV350 and 339 large T open reading frames are predicted to truncate large T protein but are unlikely to affect small T antigen protein expression. The replication origin is highly conserved with that of other polyomaviruses and includes features such as a poly(T) tract and conserved T antigen binding boxes (fig. S1). MCV has highest homology to viruses belonging to the MuPyV subgroup and is most closely related to AGM LPyV (Fig. 2B) (*15*). It is more distantly related to known human polyomaviruses and SV40. The principal differences between MCV350 and MCV339 are a 191-bp (1994 to 2184 nt) deletion in the MCV339 T antigen gene and a 5-bp (5216 to 5220 nt) insertion in the MCV339 late promoter. Excluding these sites, only 41 (0.8%) nucleotides differ between MCV350 and 339.

To investigate the association between MCV infection and MCC, we compared tumors from 10 MCC patients to two tissue control groups. The first control group was composed of unselected tissues from various body sites (including nine skin samples) from 59 patients without MCC (table S2). These samples were taken consecutively on a single surgical day and tested for MCV positivity with two PCR primer sets in the T antigen locus (LT1 and LT3) and one in the VP1 gene (VP1). These primers do not amplify cloned human BKV or JCV genomic DNA or SV40 genome from COS-7 cells. A second control group composed of skin and skin tumor samples from 25 immunocompetent and immunosuppressed patients without MCC were tested with LT1 and VP1 primers (table S2). Samples were randomized and tested in a blinded fashion. Southern blotting of PCR products was performed to increase sensitivity (fig. S2).

Of the 10 MCC tumors from different patients, 8 (80%) were positive for MCV sequences by PCR (Table 1 and table S1). Seven tumors showed robust amplification, and one tumor was positive only after PCR–Southern hybridization. MCC348 (metastasis from MCC347) and

MCC338 (infiltrating tumor from MCC339) were also positive. Two tumors, MCC343 and 346, remained negative after testing with 13 PCR primer pairs spanning the MCV genome. None of the 59 control tissues, including nine skin samples, was positive by PCR alone, but five gastrointestinal tract tissues tested weakly positive after PCR–Southern hybridization (8%, $P < 0.0001$, table S2). Viral T antigen sequences were recovered from three of these samples, confirming low copy number infection. Similarly, only 4 of 25 (16%, $P = 0.0007$, table S2) additional skin and non-MCC skin tumor samples from immunocompetent and immunosuppressed patients tested positive for MCV sequences (Table 2 and table S2).

To determine whether MCV DNA was integrated into the tumor genome, we examined MCC samples by direct Southern blotting without PCR amplification. When MCV DNA in MCC tumor is digested by single-cutter restriction endonucleases, such as EcoRI or BamHI, and probed with viral sequence, four possible patterns are predicted to occur: (i) if the viral DNA exists as freely replicating circular episomes, then a ~5.4 kilobase (kb) band will be present (integrated-concatenated virus will also generate a ~5.4 kb band); (ii) if MCV DNA integrates polyclonally, as might occur during secondary infection of the tumor if MCV is a passenger virus, then diffuse hybridization from different band sizes is expected; (iii) if MCV DNA integrates at one or a few chromosomal sites, then the tumors will have identical or near-identical non-5.4-kb banding patterns; or (iv) if MCV DNA integrates at different chromosomal sites before clonal expansion of the tumor cells, then distinct bands of different sizes will be present (monoclonal viral integration).

Eight of 11 MCC DNA samples (including MCC348 metastasis from MCC347) digested with either BamHI or EcoRI showed robust MCV hybridization, and these corresponded to the same tumors positive by PCR analysis with multiple primers (Fig. 3A and fig. S3). Monoclonal viral integration (pattern iv) was evident with one or both enzymes in six tumors: MCC339, 345, 347, 348, 349, and 352 (solid arrowheads). EcoRI digestion of MCC339, for example, produced two distinct 7.5- and 12.2-kb bands that would arise only if MCV is integrated at a single site in the majority of tumor cells. MCC344 and 350 bands have episomal or integrated-concatemeric bands (open arrowhead, pattern i). MCC352 has a monoclonal integration pattern (solid arrowheads, pattern iv) on BamHI digestion as well as an intense 5.4-kb band (open arrowhead), consistent with an integrated concatemer. All three tumors negative by PCR with ethidium bromide staining (MCC337, 343, and 346) were also negative by direct Southern blotting.

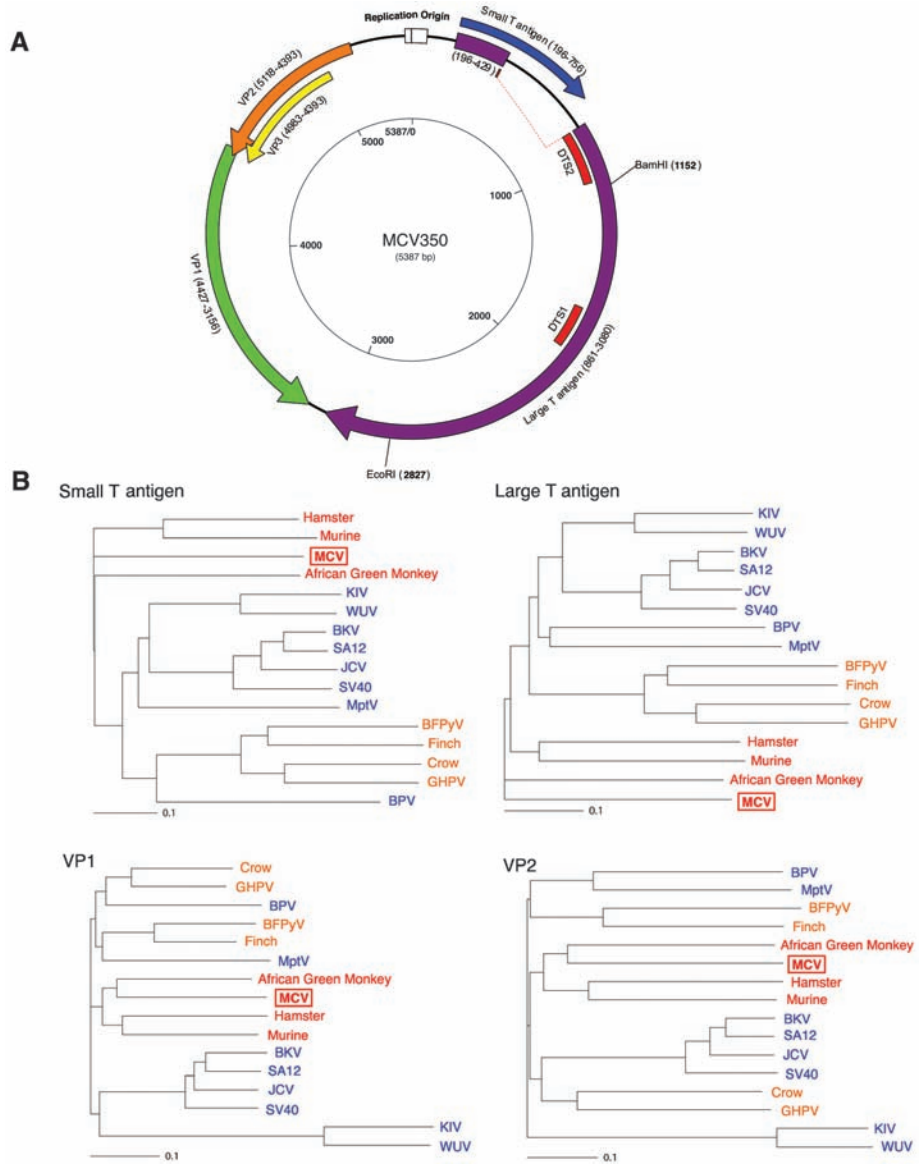

Fig. 2. (**A**) Schematic of MCV genome. Genome walking was used to clone the full MCV genome from tumor MCC350. The genome encodes typical features of a polyomavirus, including large T (purple) and small T (blue) open reading frames. Also shown are predicted VP1 (green) and overlapping VP2 (orange) and VP3 (yellow) genes. DTS1 and DTS2 (red) represent cDNA fragments originally identified by DTS screening. The former was used to identify MCV, and the latter is a spliced transcript with no homology to known polyomavirus sequences. (**B**) Neighbor-joining trees for putative MCV large T, small T, VP1, and VP2 proteins. The four known human polyomaviruses (BKV, JCV, KIV, and WUV) cluster together in the SV40 subgroup (blue), whereas MCV is most closely related to MuPyV subgroup viruses (red). Both subgroups are distinct from the avian polyomavirus subgroup (orange). Scale bars indicate an evolutionary distance of 0.1 amino acid substitutions per position in the sequence.

Table 1. PCR for MCV DNA in MCC tissues. A plus symbol indicates that the sample was strongly positive by ethidium bromide staining only with one or more primers. A minus symbol indicates that the tissue was negative for all primers. Entries with both plus and minus symbols indicate that the sample was negative by ethidium bromide staining but positive after Southern hybridization of PCR products.

MCC cases (n = 10)		
Patient	**Tissue ID**	**MCV positivity**
1	MCC337	–/+
2†	MCC338	+
2	MCC339	+
3	MCC343	–
4	MCC344	+
5	MCC345	+
6	MCC346	–
7	MCC347	+
7‡	MCC348	+
8	MCC349	+
9	MCC350	+
10	MCC352	+
	Total (%)	8/10 (80)

†MCC338 was from an infiltrating tumor in skin tissue adjacent to MCC339 tumor. ‡MCC348 taken from a metastatic lymph node from MCC347.

The Southern blot banding patterns (Fig. 3A) were identical for MCC347 and its metastasis, MCC348, in line with 3′-RACE results (Fig. 1B) and confirming that MCC348 arose as a metastatic clone of MCC347. Because the genomic integration site (the *PTPRG* locus on chromosome 3p14) is mapped for these tumors, we performed Southern blotting with flanking human sequence probes to examine cellular monoclonal integration. NheI-SacI digestion of MCC347 and 348 is predicted to generate a 3.1-kb fragment from the wild-type allele and a 3.9-kb fragment from the allele containing the integrated MCV DNA. Hybridization with a flanking human *PTPRG* sequence probe revealed that the 3.9-kb allele was present in MCC347 and 348 DNA but not in control tissue DNA (Fig. 3B). As predicted, the same fragment hybridized to a MCV T antigen sequence probe, consistent with both cellular and viral monoclonality in this tumor. These results provide evidence that MCV infection and genome integration occurred in this tumor before clonal expansion of tumor cells. MCV in MCC may have some parallels to high-risk human papillomavirus (HPV), which causes cervical cancer mainly after viral episome disruption and integration into the cervical epithelial cell genome (*16*).

If MCV plays a causal role in tumorigenesis, it could conceivably do so by several mechanisms, including T antigen expression, insertional mutagenesis, or both. Our DTS results show tumor expression of MCV T antigen, which has conserved DnaJ (*4*), pocket protein-binding LXCXE (*17*), and pp2A-binding (*18*, *19*) domains previously shown to play roles in polyomavirus-induced cell transformation. Mutational disruption of the *PTPRG* gene, which is suspected to be a tumor suppressor (*20*), could also play a role in MCC, although our Southern blot data suggest that MCV integration occurs at various genomic sites in different MCC tumors.

Our study validates the utility of DTS for the discovery of cryptic human viruses, but it has also revealed some limitations of the approach. Of the four tumors we sampled, only one (MCC347) was infected at high copy number. MCV transcripts in this tumor were present at 10 transcripts per million or about 5 transcripts per tumor cell. In future searches for other directly transforming tumor viruses (*21*), DTS should be used on multiple highly uniform samples sequenced to a depth of 200,000 transcripts or greater. Because DTS is quantitative, it is less likely to be useful in its current form for discovery of low-abundance viruses in autoimmune disorders or other chronic infectious diseases. Discovery of MCV by DTS nonetheless shows that DTS and related approaches (*22*) are promising methods to identify previously unknown human tumor viruses.

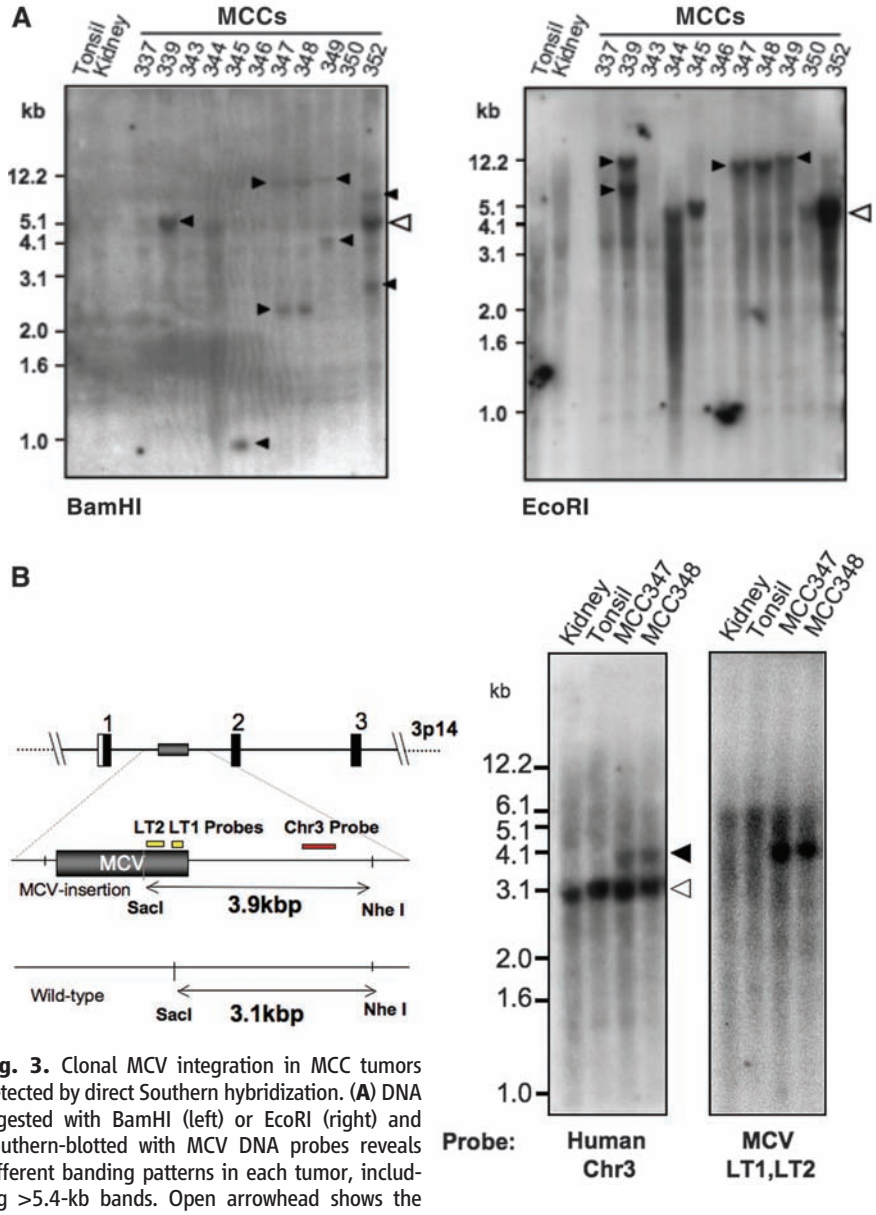

Fig. 3. Clonal MCV integration in MCC tumors detected by direct Southern hybridization. (**A**) DNA digested with BamHI (left) or EcoRI (right) and Southern-blotted with MCV DNA probes reveals different banding patterns in each tumor, including >5.4-kb bands. Open arrowhead shows the expected position for MCV episomal or concatenated-integrated genome (5.4 kb) with corresponding bands present in tumors MCC344 and 350. Tumors MCC339, 345, 347, 348, and 349 have different band sizes and doublet bands (solid arrowheads), consistent with genomic monoclonal integration. MCC352 has a prominent 5.4-kb band as well as higher and lower molecular weight monoclonal integration bands (BamHI), consistent with an integrated concatemer. Tumors MCC337, 343, and 346 have no MCV DNA detected by Southern blotting [bands at 1.5 kb (kidney) and 1.2 kb (MCC346) are artifacts]. (**B**) Viral and cellular monoclonality in MCC347 and 348. Tumor MCC347 and its metastasis MCC348 were digested with SacI and NheI and Southern-blotted with unique human flanking sequence probe [Chr3 (red), left] or viral probes [LT1 and LT2 (yellow), right]. The wild-type human allele is present in all samples at 3.1 kb (left). The MCC tumors, however, have an additional 3.9-kb allelic band formed by MCV DNA insertion into chromosome 3p14. Hybridization with probes for MCV T antigen sequence (yellow, right) generates an identical band.

References and Notes

1. L. Gross, *Proc. Soc. Exp. Biol. Med.* **83**, 414 (1953).
2. D. L. Poulin, J. A. DeCaprio, *J. Clin. Oncol.* **24**, 4356 (2006).
3. S. M. Dilworth, *Nat. Rev. Cancer* **2**, 951 (2002).
4. J. M. Pipas, *J. Virol.* **66**, 3979 (1992).
5. K. A. Crandall, M. Perez-Losada, R. G. Christensen, D. A. McClellan, R. P. Viscidi, *Adv. Exp. Med. Biol.* **577**, 46 (2006).
6. T. Allander *et al.*, *J. Virol.* **81**, 4130 (2007).
7. A. M. Gaynor *et al.*, *PLoS Pathog.* **3**, e64 (2007).
8. D. Hollanderova, H. Raslova, D. Blangy, J. Forstova, M. Berebbi, *Int. J. Oncol.* **23**, 333 (2003).
9. B. Lemos, P. Nghiem, *J. Investig. Dermatol.* **127**, 2100 (2007).
10. M. Van Gele *et al.*, *Oncogene* **23**, 2732 (2004).
11. E. A. Engels, M. Frisch, J. J. Goedert, R. J. Biggar, R. W. Miller, *Lancet* **359**, 497 (2002).
12. Y. Chang *et al.*, *Science* **266**, 1865 (1994).
13. H. Feng *et al.*, *J. Virol.* **81**, 11332 (2007).
14. H. H. Chou, M. H. Holmes, *Bioinformatics* **17**, 1093 (2001).
15. M. Pawlita, A. Clad, H. zur Hausen, *Virology* **143**, 196 (1985).
16. M. Durst, L. Gissmann, H. Ikenberg, H. zur Hausen, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 3812 (1983).
17. J. A. DeCaprio *et al.*, *Cell* **54**, 275 (1988).
18. D. C. Pallas *et al.*, *Cell* **60**, 167 (1990).
19. G. Walter, R. Ruediger, C. Slaughter, M. Mumby, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 2521 (1990).
20. D. M. Pitterle, E. M. Jolicoeur, G. Bepler, *In Vivo (Athens)* **12**, 643 (1998).

Table 2. PCR for MCV DNA in comparison control tissues (n = 84). For detailed description of tissues and tissue sites, see table S2. MCV positivities marked with plus and minus symbols together are as in Table 1. For the various body site tissues, there were 59 samples; for the skin and skin tumor tissues, the sample size was 25 (table S2).

	MCV positivity
Various body site tissues	
Total MCV negative (%)	54/59 (92)
Total MCV positive (%)	5/59 (8)
Appendix control 1	–/+
Appendix control 2	–/+
Gall bladder	–/+
Bowel	–/+
Hemorrhoid	–/+
Skin and skin tumor tissues	
Total MCV negative (%)	21/25 (84)
Total MCV positive (%)	4/25 (16)
Skin	–/+
KS skin tumor 1	–/+
KS skin tumor 2	–/+
KS skin tumor 3	–/+

21. J. Parsonnet, in *Microbes and Malignancy*, J. Parsonnet, Ed. (Oxford Univ. Press, New York, 1999), pp. 3–18.
22. Y. Xu *et al.*, *Genomics* **81**, 329 (2003).
23. We thank the National Cancer Institute–supported Cooperative Human Tissue Network for tissues used in this study, M. Aquafondata for tissue staining, P. S. Schnable for sharing cDNA data sets used in DTS pilot testing, O. Gjoerup and R. D. Wood for helpful comments, and J. Zawinul for help with the manuscript. Supported in part by funds from NIH R33CA120726 and the Pennsylvania Department of Health. The Pennsylvania Department of Health specifically disclaims responsibility for any analyses, interpretations, or conclusions.

Supporting Online Material
www.sciencemag.org/cgi/content/full/1152586/DC1
Materials and Methods
Figs. S1 to S3
Tables S1 to S5
References

5 November 2007; accepted 8 January 2008
Published online 17 January 2008;
10.1126/science.1152586
Include this information when citing this paper.

Worldwide Human Relationships Inferred from Genome-Wide Patterns of Variation

Jun Z. Li,[1,2]*† Devin M. Absher,[1,2]* Hua Tang,[1] Audrey M. Southwick,[1,2] Amanda M. Casto,[1] Sohini Ramachandran,[4] Howard M. Cann,[5] Gregory S. Barsh,[1,3] Marcus Feldman,[4]‡ Luigi L. Cavalli-Sforza,[1]‡ Richard M. Myers[1,2]‡

Human genetic diversity is shaped by both demographic and biological factors and has fundamental implications for understanding the genetic basis of diseases. We studied 938 unrelated individuals from 51 populations of the Human Genome Diversity Panel at 650,000 common single-nucleotide polymorphism loci. Individual ancestry and population substructure were detectable with very high resolution. The relationship between haplotype heterozygosity and geography was consistent with the hypothesis of a serial founder effect with a single origin in sub-Saharan Africa. In addition, we observed a pattern of ancestral allele frequency distributions that reflects variation in population dynamics among geographic regions. This data set allows the most comprehensive characterization to date of human genetic variation.

In the past 30 years, the ability to study DNA sequence variation has dramatically increased our knowledge of the relationships among and history of human populations. Analyses of mitochondrial, Y chromosomal, and autosomal markers have revealed geographical structuring of human populations at the continental level (*1*–*3*) and suggest that a small group of individuals migrated out of eastern Africa and their descendants subsequently expanded into most of today's populations (*3*–*6*). Despite this progress, these studies were limited to a small fraction of the genome, to limited populations, or both, and yield an incomplete picture of the relative importance of mutation, recombination, migration, demography, selection, and random drift (*7*–*10*). To substantially increase the genomic and population coverage of past studies (e.g., the HapMap Project), we have examined more than 650,000 single-nucleotide polymorphisms (SNPs) in samples from the Human Genome Diversity Panel (HGDP-CEPH), which represents 1064 fully consenting individuals from 51 populations from sub-Saharan Africa, North Africa, Europe, the Middle East, South/Central Asia, East Asia, Oceania, and the Americas (*11*). This data set is freely available (*12*) and allows a detailed characterization of worldwide genetic variation.

We first studied genetic ancestry of each individual without using his/her population identity. This analysis considers each person's genome as having originated from K ancestral but unobserved populations whose contributions are described by K coefficients that sum to 1 for each individual. To increase computational efficiency, we developed new software, *frappe*, that implements a maximum likelihood method (*13*) to analyze all 642,690 autosomal SNPs in 938 unrelated and successfully genotyped HGDP-CEPH individuals (*14*). Figure 1A shows the results for $K = 7$; those for $K = 2$ through 6 are in fig. S1. At $K = 5$, the 938 individuals segregate into five continental groups, similar to those re-

[1]Department of Genetics, Stanford University School of Medicine, Stanford, CA 94305–5120, USA. [2]Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305–5120, USA. [3]Department of Pediatrics, Stanford University School of Medicine, Stanford, CA 94305–5120, USA. [4]Department of Biological Sciences, Stanford University, Stanford, CA 94305–5120, USA. [5]Foundation Jean Dausset-Centre d'Etude du Polymorphisme Humain (CEPH), 75010 Paris, France.

*These authors contributed equally to this work.
†Present address: Department of Human Genetics, University of Michigan, 5789A MS II, Ann Arbor, MI 48109–5618, USA.
‡To whom correspondence should be addressed. E-mail: marc@charles.stanford.edu (M.F.); cavalli@stanford.edu (L.L.C.S.); myers@shgc.stanford.edu (R.M.M.)

Supporting Online Material (MS#1152586)

Materials and methods

Human tissue samples

Human Merkel cell carcinoma tissues were obtained from the Cooperative Human Tissue Network as frozen excess biopsy samples (Table S1). All MCC tumors except MCC352 were reconfirmed in our laboratory by H&E and cytokeratin 20 immunostaining; of these all except MCC350 were positive for cytokeratin 20. MCC350 represents metastasis to lymph node. Due to sampling issues, we were unable to identify MCC tumor cells in this portion of tissue taken for our examination. We relied on the original pathology report as evidence for MCC. Four cases (MCC347, MCC337, MCC343, and MCC346) from 4 men ranging in age from 38 to 84 years were used for DTS analysis. Control samples were collected from excess surgical tissues as a consecutive series of anonymized pathology collections from a single operating day or collected from ongoing studies of Kaposi's sarcoma (Table S2). All tissues were tested under University of Pittsburgh IRB exemption status for anonymous excess pathology tissues not required for patient diagnosis.

Generation of cDNA library for pyrosequencing

Total RNA was extracted from MCC tissues using RNeasy Midi Kit (Qiagen, Alameda, CA) and treated with DNase I (Ambion, Austin, TX) to remove genomic DNA. Integrity of tissue RNAs was analyzed by an Agilent 2100 bioanalyzer (Quantum Analytics, Foster City, CA) using the RNA 6000 Nano Reagent Kit. mRNA was purified with Dynabeads® mRNA Purification Kit (Invitrogen). Double strand cDNA was synthesized with oligo(dT) primer using the SuperScript™ Double-stranded cDNA Synthesis Kit (Invitrogen). Five micrograms of MCC cDNA was used for pyrosequencing after confirming cDNA quality on an Agilent bioanalyzer (Quantum Analytics) at 454 Life Sciences (Roche). The cDNA sample was fractionated (300-500 bp) and blunted for ligation with two different adaptors. These two adaptors provide unique priming sequences for both amplification and sequencing, and form the basis of the single-strand template library for pyrosequencing (*S1*). Sequencing was performed on two cDNA libraries: one library from a single case (MCC347) and another library of three pooled cases (MCC337, 343 and 346).

Digital Transcriptome Subtraction

The sequence data was first trimmed using LUCY (*S2*) with stringency similar to Phred scores of 20 or higher (-error 0.01 0.01), and long reads over 50 bp (-m 50). Only high quality sequences obtained after Lucy trimming were used for further subtraction with SeqClean (http://compbio.dfci.harvard.edu/tgi/software/). First, poly(A/T), dust (low-complexity), human repeat (http://www.girinst.org) and adaptor sequences were removed to obtain a high fidelity (HiFi) dataset. These HiFi sequences were then aligned against human databases, including human RefSeq RNA (ftp://ftp.ncbi.nih.gov/refseq/H_sapiens/mRNA_Prot), mitochondrial and assembled chromosomes (ftp://ftp.ncbi.nih.gov/refseq/H_sapiens/H_sapiens), and human immunoglobulin variable sequences (ftp://ftp.ncbi.nih.gov/blast/db/FASTA) with a minimum hit length of 30 bp (-x 95). The remaining candidate sequences were then aligned to online GenBank nonredundant (NR) using BLASTX program in the netblast package (ftp://ftp.ncbi.nih.gov/blast/executables/). DTS analysis was performed using stand-alone executables on a Mac Pro (Apple, Cupertino CA).

RACE analysis on MCV transcripts

Both rapid amplification of 5'- and 3'- cDNA ends (RACE) were performed on MCC347 and MCC348 with GeneRacer Kit (Invitrogen) according to the manufacturer's instructions. Primers used for RACE are listed in Table S3. Primer M1L and M3 were used in 5'-RACE. Primer M2L and M4

were used in 3'-RACE. The PCR fragments were isolated from agarose gels, extracted with QIAEX II Gel Extraction Kit (Qiagen), and ligated into pCR 2.1 vector (Invitrogen) for DNA sequencing.

Consensus PCR for VP1

Consensus PCR on the polyomavirus VP1 region was performed as previously described (*S3*). Genomic DNAs from MCC339, MCC344, MCC347, and MCC350 were PCR amplified using Platinum Taq DNA polymerase (Invitrogen) with VP1 consensus primers VP1-1 as in Table S3. Cycling conditions for the first PCR were 5 min at 95 °C, followed by 45 cycles each of 94 °C for 30 sec, 46 °C for 1 min and 72 °C for 1 min, and final elongation at 72 °C for 10 min. Nested PCR was performed with consensus VP1-2 primers (Table S3) using 4 μl of the first PCR product as template in a similar reaction at 95 °C for 5 min, 45 cycles of 94 °C for 30 sec, 56 °C for 30 min and 72 °C for 30 sec, and 72 °C for 10 min. PCR fragments were recovered from agarose gel, cloned in pCR2.1 cloning vector (Invitrogen) and sequenced. Based on the sequencing results, specific primers (VP1-iF and VP1-iR) were designed in the MCV350 VP1 region.

MCV genome sequencing

Primers for genome sequencing are listed in Table S4. The viral genome was bi-directionally sequenced with >3 fold bidirectional coverage. First, successive outward PCR was performed from the 3' end of the T antigen sequence to a conserved VP1 site with primers M6 and VP1-iR, and from the 5' end of the T antigen sequences to a conserved VP1 site with primers M5 and VP1-iF. Walking primers (W1-W10) were then designed to sequence the long PCR products. This sequence data was finally used to design 13 PCR primer sets (contig1-contig13) that encircle the genome. These PCR products were used for confirmatory second and third sequencing rounds. All PCR reactions were performed with High Fidelity Platinum Taq DNA polymerase (Invitrogen).

MCV detection by PCR-Southern blotting

Genomic DNA was extracted by standard phenol-chloroform method and DNA quality was confirmed by β-actin PCR. One hundred nanograms of genomic DNA was amplified using Taq DNA polymerase (Invitrogen) in a final volume of 50 μl. Cycling conditions were 3 min at 94 °C, followed by 31 cycles each of 94 °C for 45 sec, 58 °C for 30 sec and 72 °C for 45 sec, and final elongation of 15 min at 72 °C. PCR primers are listed in Table S5. PCR of the T antigen locus was performed with LT1 and LT3 primer sets (internal Southern probes were generated with M1-M2 and LT5, respectively) and for the VP1 gene with VP1 primer set (internal probe generated with VP1.3). Absence of MCV genome in MCC343 and 346 was confirmed using genomic 13 PCR primer sets contig1-contig13 (Table S4). Southern blotting for PCR products was performed as described below for genomic Southern blotting, using 15 μl of the 50μl PCR reaction product as the starting DNA and electrophoresis on 1% agarose gels.

To avoid potential contamination of template DNA, PCR mixtures were prepared in an isolated room and template DNA was prepared in an UV-irradiated clean hood. Recombinant DNA harboring MCV DNA sequence was not amplified at the same time as tissue samples to avoid cross contamination between PCR samples. Negative PCR controls contained all components except DNA template. All samples including control samples were randomized and blinded to the scientist performing PCR-Southern throughout testing. Fisher exact 2-tailed tests were used to compare positivity rates between groups.

Genomic Southern blotting for MCV and *PTPRG*

Genomic Southern probes were generated by PCR using primers listed in Table S5. Genomic Southern blotting for virus monoclonality was performed on fifteen micrograms of control or tumor DNA, digested overnight with 60 units of EcoRI or BamHI (New England Biolabs, Ipswich MA).

Genomic Southern blotting for cellular monoclonality was performed with 10 micrograms of control or tumor DNA digested with 100 units NheI and 100 units SacI (New England Biolabs) overnight. DNA digestions were separated on 0.7% agarose gels at 80 volts with ethidium bromide staining to confirm complete digestion. Digested genomic DNA was then transferred onto a nitrocellulose membrane (Amersham) with 10x SSC (*S4*). Membranes were hybridized with probe overnight at 42 °C and rinsed in 0.2x SSC with 0.5% SDS at 60 °C for MCV probes, and 72 °C for the Chr 3 probe in the *PTPRG* gene. Probes were labeled with [α^{32}P] dCTP (~$3x10^7$ dpm/ml) using the Readiprime II Random Prime Labelling System (Amersham). MCV DNA probes (LT1, LT2, P1, P3, P6, P9, and P12) covering 3.2 kb of non-overlapping MCV350 genome (Fig. S3) were combined for Southern blot detection of MCV genome.

Supporting figures

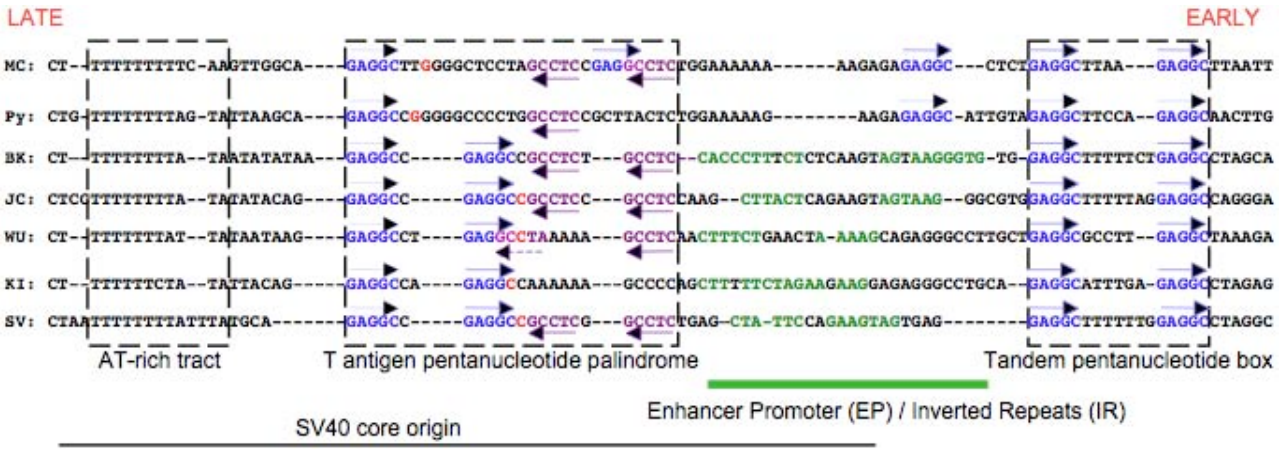

Figure S1. MCV replication origin sequence is highly conserved with other polyomaviruses. The MCV350 replication origin (5360-5387 and 1-69 nt) has seven conserved pentameric T antigen-binding sites (forming a pentanucleotide palindrome and a tandem pentanucleotide box), a homopolymeric T tract and semiconserved inverted repeats found in other polyomaviruses. (MC: Merkel Cell polyomavirus, Py: Mouse polyomavirus, BK: human BK polyomavirus, JC: human JC polyomavirus, WU: human WU polyomavirus, KI: human KI polyomavirus, SV: Simian virus 40)

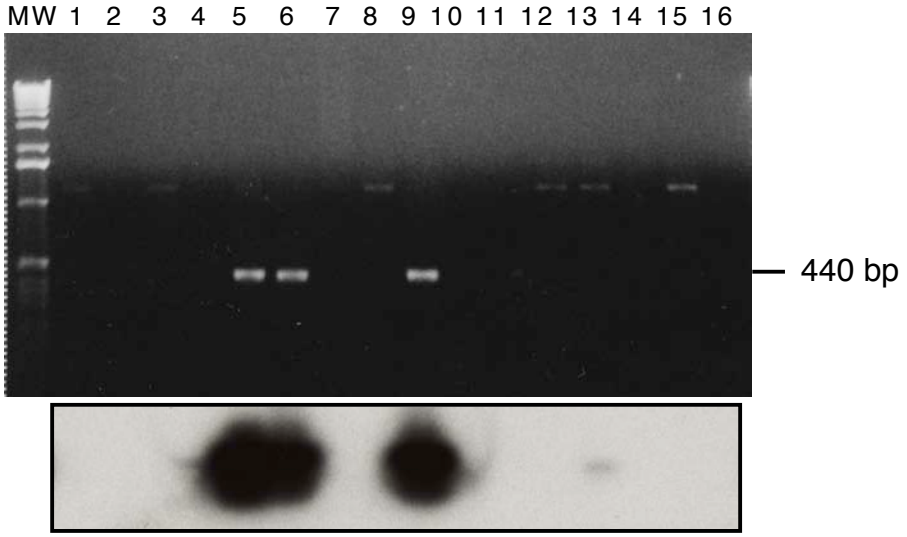

Figure S2. Representative results of PCR-Southern blotting for detection of MCV in MCC and control tissues. Samples were tested and scored in a randomized and blinded fashion. Top panel, agarose gel (ethidium bromide) of amplification products using LT1 primers (Table S5). Bottom panel, hybridization of LT1 PCR products to the [α^{32}P]-dCTP-labeled M1-M2 internal probe after transfer of DNA to nitrocellulose membrane. MCC tissue DNA in lanes 1 (MCC346), 5 (MCC348), 6 (MCC344), 9 (MCC339), and 15 (MCC343); DNA-negative control (H_2O) in lanes 2, 10 and 11; and surgical control tissue DNA in lanes 3, 4, 7, 8, 12, 13, 14 and 16. Lanes 5, 6 and 9 (MCC348, 344 and 339 respectively) show robust signal after both ethidium bromide staining and Southern blotting. The weak signal in lane 13 is from a control gall bladder tissue DNA that is positive only after Southern blotting of the PCR product. MCC346 (lane 1) and 343 (lane 15) are negative (see Text).

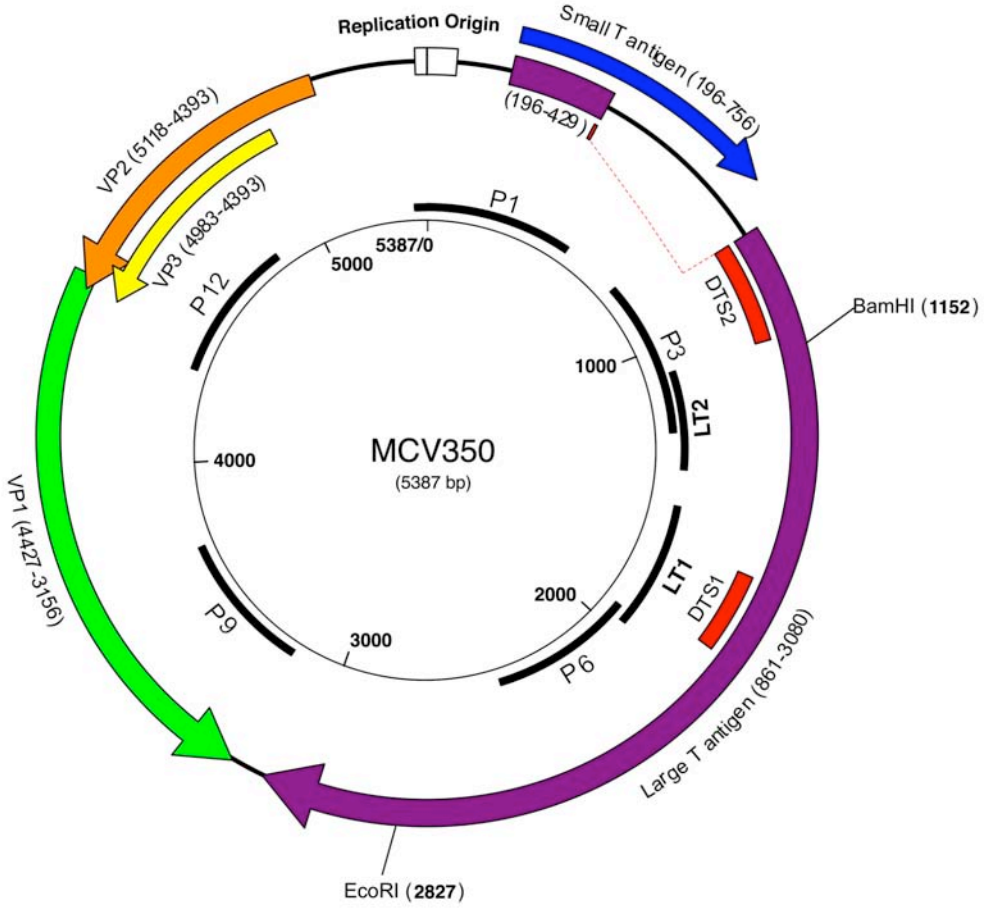

Figure S3. Positions of the probes used for genomic Southern blotting to detect MCV. Unless indicated otherwise, all seven probes were used together for MCV detection.

Supporting tables

Table S1. Clinicopathological and PCR data for MCC.

Patient	Tissue ID	Age	Sex	Race	Cytokeratin 20
1	MCC337	84	Male	White	+
2	MCC338*	79	Male	White	+
	MCC339				+
3	MCC343	79	Male	White	+
4	MCC344	57	Male	White	+
5	MCC345	77	Male	Black	+
6	MCC346	38	Male	Unknown	+
7	MCC347	56	Male	White	+
	MCC348†				+
8	MCC349	58	Female	White	+
9	MCC350	58	Male	White	-
10	MCC352	58	Male	White	ND‡

PCR Results for MCC Cases (n, 10)

Patient	Tissue ID	LT1	LT3	VP1	Summary§
1	MCC337	-/+	-/-	-/-	-/+
2	MCC338†	+/+	+/+	+/+	+/+
	MCC339	+/+	+/+	+/+	+/+
3	MCC343	-/-	-/-	-/-	-/-
4	MCC344	+/+	+/+	+/+	+/+
5	MCC345	-/-	+/+	-/-	+/+
6	MCC346	-/-	-/-	-/-	-/-
7	MCC347	+/+	+/+	-/-	+/+
	MCC348‡	+/+	+/+	-/-	+/+
8	MCC349	+/+	+/+	-/+	+/+
9	MCC350	+/+	+/+	+/+	+/+
10	MCC352	ND	+/+	+/+	+/+
No. of Positives (%)		6/9 (67)	7/10 (70)	5/10 (50)	8/10 (80)

* MCC338: Infiltrating tumor in skin tissue adjacent to MCC339 tumor.
† MCC348: Metastatic MCC to lymph node from MCC347.
‡ ND: Not Determined.
§ +/+ : Strongly positive by ethidium bromide staining and Southern hybridization of PCR products with one or more primers.
-/- : Negative for both ethidium bromide staining and Southern hybridization on all primers.
-/+ : Negative by ethidium bromide staining but positive after Southern hybridization of PCR products.

Table S2. Clinicopathological and PCR data for control tissues from various body site and skin tissues.

Tissue Type	Total No. (No. MCV positive)*
Various Body Site Controls	
Colon	5
Small Bowel	3 (1)
Hemorrhoid	1 (1)
Gall Bladder	7 (1)
Appendix	9 (2)
Mouth	1
Vein	2
Heart	1
Kidney	1
Skin	9
Hernia	2
Hematolymphoid tissues	
Lymph node	1
Tonsil	5
B cell CLL	1
Myeloid hyperplasia	1
Posttransplant lymphoma	1
HIV+ large cell lymphoma	1
Miscellaneous tissues	
Lipoma	1
Fibrous tissue	2
Fistula track	1
Meningioma	1
Breast cancer	1
Lung cancer	1
Prostate	1
Skin Tissue Controls	
Normal skin (1 HIV+)	6 (1)†
Kaposi's sarcoma (4 HIV+)	15 (3)†
Malignant Melanoma	1
Inflammatory Skin	3

PCR Results for Comparison Control Tissues (n, 84)

Positive Tissues Only	LT1	LT3	VP1	Summary‡
Various Body Sites (n, 59)				
Appendix	-/+	-/+	-/+	-/+
Appendix	-/-	-/+	-/+	-/+
Gall Bladder	-/+	-/-	-/-	-/+
Bowel	-/-	-/+	-/+	-/+
Hemorrhoid	-/-	-/-	-/+	-/+
Skin or Skin Tumors (n, 25)				
Skin	-/+	ND	-/+	-/+
KS skin tumor	-/+	ND	-/+	-/+
KS skin tumor	-/+	ND	-/-	-/+
KS skin tumor	-/-	ND	-/+	-/+

* Each tissue sample from a single patient.

† None of the samples from HIV+ patients were positive for MCV.

‡ -/+ : Negative by ethidium bromide staining but positive after Southern hybridization of PCR products on one or more primers.

Table S3. Primers used for the MCV cloning.

Name	Nucleotide position*	Purpose	Sequence
M1L	1894-1864	5'-RACE	TTCTCTTGCAGTAATTTGTAAGGGGACTTAC
M3	1848-1827	5'-RACE	TTTCAGGCATCTTATTCACTCC
M2L	1707-1734	3'-RACE	AGCAGGCATGCCTGTGAATTAGGATGTA
M4	1784-1805	3'-RACE	TTTTTGCTCTACCTTCTGCACT
VP1-1F		VP1 Consensus PCR	CCAGACCCAACTARRAATGARAA
VP1-1R		VP1 Consensus PCR	AACAAGAGACACAAATNTTTCCNCC
VP1-2F		VP1 Consensus PCR	ATGAAAATGGGGTTGGCCCNCTNTGYAARG
VP1-2R		VP1 Consensus PCR	CCCTCATAAACCCGAACYTCYTCHACYTG
M6	1827-1848	Genome Cloning	GGAGTGAATAAGATGCCTGAAA
VP1-iR	3480-3461	Genome Cloning	ATGGGTGAAAAACCCCTACC
M5	1796-1770	Genome Cloning	GGTAGAGCAAAAATTCTTAATAGCAGA
VP1-iF	3508-3527	Genome Cloning	CTAGGCAACCCATGAAGAGC

*Nucleotide position is based on MCV350 genome.

Table S4. Primers used for genome sequencing.

Name	Nucleotide position*	Purpose	Sequence
W1	411-4130	Primer walking	ACTCTTGCCACACTGTAAGC
W2	1290-1272	Primer walking	CAGGGGAGGAAAGTGATTC
W3	4268-4288	Primer walking	GGGTAATGCTATCTTCTCCAG
W4	946-929	Primer walking	TATTCGTATGCCTTCCCG
W5	4293-4316	Primer walking	CACAGATAATACTTCCACTCCTCC
W7	5260-5278	Primer walking	TTATCAGTCAAACTCCGCC
W8	5294-5312	Primer walking	TCAATGCCAGAAACCCTGC
W9	166-148	Primer walking	AACAGCAGAGGAGCAAATG
W10	96-78	Primer walking	TCTGCCCTTAGATACTGCC
contig1f	5344-5363	overlapping contigs	TTGGCTGCCTAGGTGACTTT
contig1r	518-499	overlapping contigs	CCAGGACCTCTGCAAAATCT
contig2f	354-373	overlapping contigs	GGAATTGAACACCCTTTGGA
contig2r	879-860	overlapping contigs	ATATAGGGGCCTCGTCAACC
contig3f	730-749	overlapping contigs	TGCTTACTGCATCTGCACCT
contig3r	1287-1268	overlapping contigs	GGGAGGAAAGTGATTCATCG
contig4f	1132-1151	overlapping contigs	AGGAACCCACCTCATCCTCT
contig4r	1641-1619	overlapping contigs	AAATGGCAAAACAACTTACTGTT
contig5f	1538-1561	overlapping contigs	AAACAACAGAGAAACTCCTGTTCC
contig5r	2088-2069	overlapping contigs	GAGCCTTGTGAGGTTTGAGG
contig6f	1934-1953	overlapping contigs	AGAGGCCAGCTGTAATTGGA
contig6r	2437-2418	overlapping contigs	GCAGCAAAGCTTGTTTTTCC
contig7f	2328-2349	overlapping contigs	TTTGAAAAGAAGCTGCAGAAAA
contig7r	2885-2866	overlapping contigs	TGTATCAGGCAAGCACCAAA
contig8f	2763-2783	overlapping contigs	CACTTTTTCCCAAAGGCAAAT
contig8r	3282-3263	overlapping contigs	TTACCCAAAGCCCTCTGTTG
contig9f	3187-3206	overlapping contigs	GAGGCCTTTTGAGGTCCTTT
contig9r	3687-3667	overlapping contigs	TCAGACAGGCTCTCAGACTCC
contig10f	3599-3618	overlapping contigs	ATAGAGGGCCCACTCCATTC
contig10r	4107-4088	overlapping contigs	TCTGCCAATGCTAAATGAGG
contig11f	3949-3969	overlapping contigs	CCTGACACAGGAATACCAGCA
contig11r	4504-4485	overlapping contigs	GCAAACTCCAGATTGGCTTC
contig12f	4329-4349	overlapping contigs	TTTTGGAACTGAGGCAACATT
contig12r	4829-4810	overlapping contigs	TAACTGTGGGGGTGAGGTTG
contig13f	4765-4784	overlapping contigs	TACCCACGAAACATCCCTGT
contig13r	5386-5367	overlapping contigs	AGCCTCTGCCAACTTGAAAA

*Nucleotide position is based on MCV350 genome.

Table S5. PCR Primers and Probes used for MCV detection.

Name	Nucleotide position*	Sense	Antisense
		Primers for MCV PCR	
LT1	1514-1953	TACAAGCACTCCACCAAAGC	TCCAATTACAGCTGGCCTCT
LT3	571-879	TTGTCTCGCCAGCATTGTAG	ATATAGGGGCCTCGTCAACC
VP1	4137-3786	TTTGCCAGCTTACAGTGTGG	TGGATCTAGGCCCTGATTTTT
		PCR Primers for Southern hybridization Probes	
M1-M2	1711- 1889	GGCATGCCTGTGAATTAGGA	TTGCAGTAATTTGTAAGGGGACT
LT5	253-855	GCTCCTAATTGTTATGGCAACA	TGGGAAAGTACACAAAATCTGTCA
VP1.3	4107-3599	TCTGCCAATGCTAAATGAGG	ATAGAGGGCCCACTCCATTC
P1	5344-518	TTGGCTGCCTAGGTGACTTT	CCAGGACCTCTGCAAAATCT
P3	730-1287	TGCTTACTGCATCTGCACCT	GGGAGGAAAGTGATTCATCG
P6	1934-2437	AGAGGCCAGCTGTAATTGGA	GCAGCAAAGCTTGTTTTTCC
P9	3187-3687	GAGGCCTTTTGAGGTCCTTT	TCAGACAGGCTCTCAGACTCC
P12	4329-4829	TTTTGGAACTGAGGCAACATT	TAACTGTGGGGGTGAGGTTG
LT2	1054-1428	CTGGGTATGGGTCCTTCTCA	TGGTGAAGGAGGAGGATCTG
Chr.3	61563308 – 61563830†	TTTCAGACGGAAGCGAAGTT	ACCACGATTTGGAAAACAGC

*Nucleotide position is based on MCV350 genome.
† Nucleotide position is based on NT_022517.17.

Supporting references

S1. M. Margulies *et al.*, *Nature* **437**, 376 (2005).
S2. H. H. Chou, M. H. Holmes, *Bioinformatics* **17**, 1093 (2001).
S3. R. Johne, D. Enderlein, H. Nieper, H. Muller, *J Virol* **79**, 3883 (2005).
S4. E. Cesarman, P.S. Moore, P.H. Rao, G. Inghirami, D. M. Knowles, Y. Chang, *Blood* **86**, 2708 (1995).