Supporting Online Material for

Clonal Integration of a Polyomavirus in Human Merkel Cell Carcinoma

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

*To whom correspondence should be addressed. E-mail: yc70@pitt.edu (Y.C.), psm9@pitt.edu (P.S.M.)

Published 17 January 2008 on *Science* Express
DOI: 10.1126/science.1152586

This PDF file includes:

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

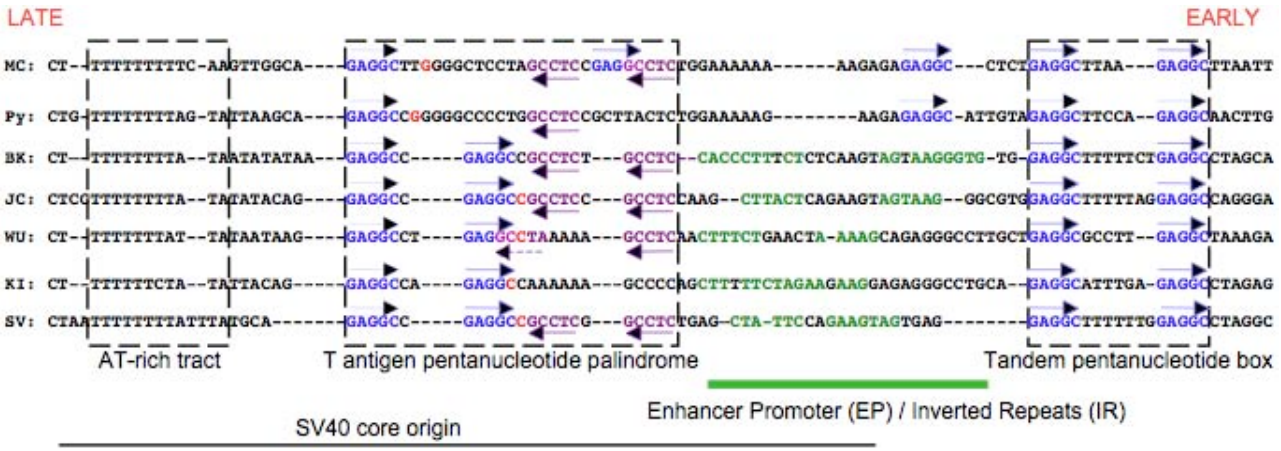

Figure S1. MCV replication origin sequence is highly conserved with other polyomaviruses. The MCV350 replication origin (5360-5387 and 1-69 nt) has seven conserved pentameric T antigen-binding sites (forming a pentanucleotide palindrome and a tandem pentanucleotide box), a homopolymeric T tract and semiconserved 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)

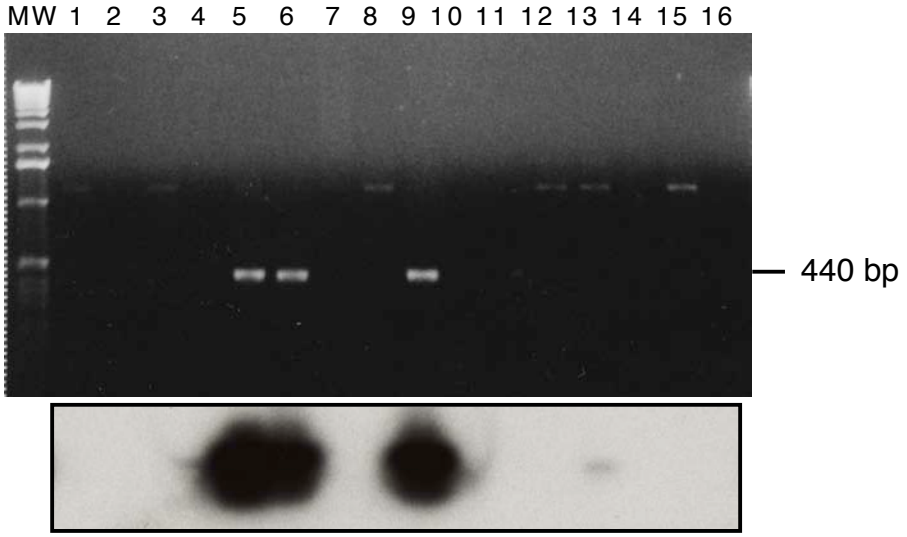

Figure S2. Representative results of PCR-Southern blotting for detection of MCV in MCC and control tissues. Samples were tested and scored in a randomized and blinded fashion. Top panel, agarose gel (ethidium bromide) of amplification products using LT1 primers (Table S5). Bottom panel, hybridization 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).

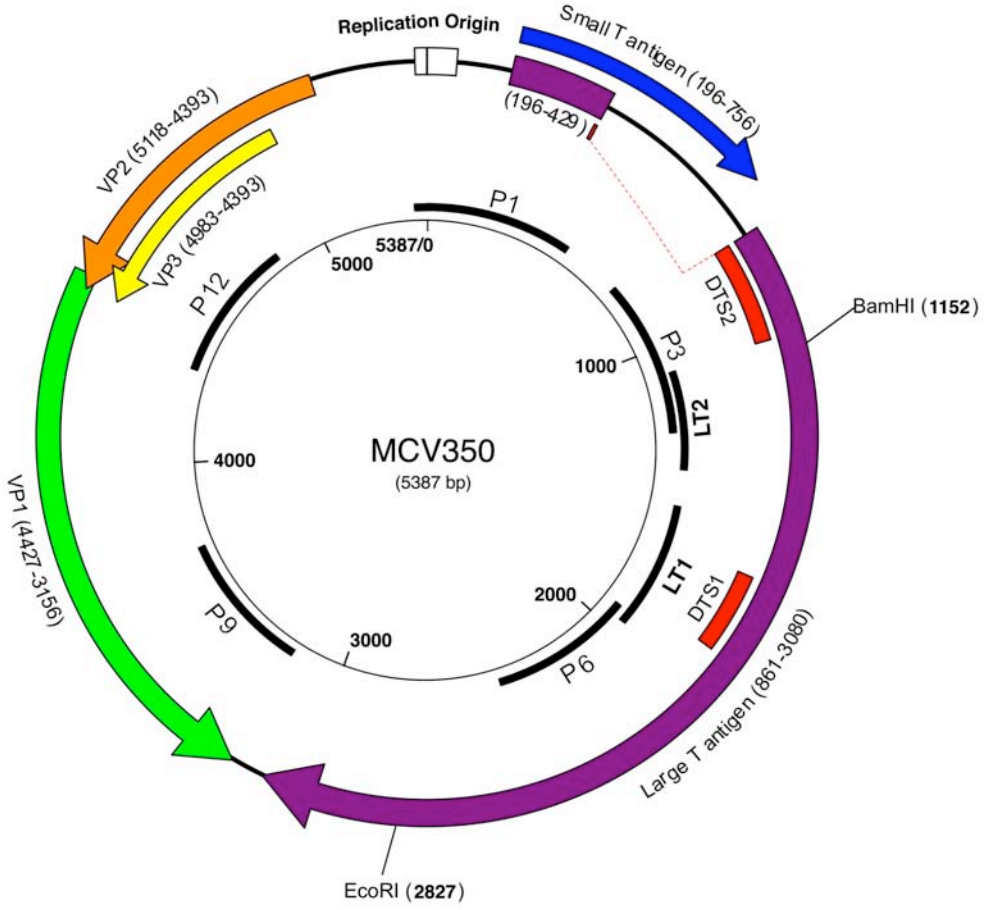

Figure S3. Positions of the probes used for genomic Southern blotting to detect MCV. Unless indicated otherwise, all seven probes were used together for MCV detection.

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).