

Novel transcripts of human cytomegalovirus clinical strain found by cDNA library screening

Y.P. Ma, Q. Ruan, Y.H. Ji, N. Wang, M.L. Li, Y. Qi, R. He, Z.R. Sun and G.W. Ren

Virus Laboratory, The Affiliated Shengjing Hospital, China Medical University, Shenyang, Liaoning, P.R. China

Corresponding author: Q. Ruan
E-mail: ruanq@sj-hospital.org

Genet. Mol. Res. 10 (2): 566-575 (2011)
Received September 10, 2010
Accepted December 6, 2010
Published April 5, 2011
DOI 10.4238/vol10-2gmr1059

ABSTRACT. Human cytomegalovirus (HCMV) is a double-stranded DNA virus with the largest genome (~235 kb) of the known human herpes viruses. The coding potential and transcript structures of most HCMV predicted genes have not been identified. New or unknown genes could exist in clinical strains. The SMART (switching mechanism at 5' end of RNA template of reverse transcriptase) technique was used to construct a full-length cDNA library of an HCMV clinical strain in the late expression phase. Randomly selected clones were sequenced. The sequenced expressed sequence tags were used to identify the expression and transcript structures of some predicted and unpredicted genes of HCMV. The transcripts of the UL99, TRIL5/IRL5, UL73 to UL75, UL4, and UL115 genes, which were previously detected, were obtained with full-length structures from this library. Some novel transcripts, including several transcripts of UL/b' genes and three antisense transcripts of UL83, UL87 and UL31 were found. The novel transcripts that were found, particularly the antisense transcripts of UL83, UL87 and UL31, showed that the transcription of HCMV genes is more complex than previously predicted. Our study highlights the

usefulness of the full-length cDNA library for discovering new genes and transcripts of HCMV.

Key words: Human cytomegalovirus; Full-length cDNA library; Novel transcripts; Antisense transcripts

INTRODUCTION

Human cytomegalovirus (HCMV) is a double-stranded DNA virus with the largest genome (~235 kb) of the known human herpesviruses. The open reading frames (ORFs) of HCMV had been mapped on the genome by computer analysis and sequences compared between different HCMV strains and different CMV genomes (Chee et al., 1990; Rigoutsos et al., 2003; Dolan et al., 2004). To date, most HCMV genes have not been extensively characterized with respect to their expression patterns, although the coding potential and transcript structures of a few predicted genes had been identified by analysis of mRNA splicing, Northern blotting, identification of protein, and localization of the coding ORF (Jones and Muzithras, 1991; Adam et al., 1995; Wing and Huang, 1995; Atalay et al., 2002; Akter et al., 2003; Awasthi et al., 2004; Jenkins et al., 2004; Bego et al., 2005). Different transcripts expressed from a single piece of DNA sequence have been observed. Splicing and unsplicing transcripts were previously identified from some predicted genes. Information regarding the nature of transcription from ORFs of the HCMV genome is of value in describing the complex genome, which has the potential to encode over 200 proteins. Studies on HCMV transcript structures have been greatly hampered by the lack of high-throughput methods.

To identify the coding potential and transcript structures of more HCMV predicted genes, and to find new or unknown genes, the construction of cDNA libraries of HCMV is an efficient method (Scott et al., 2002; Bego et al., 2005). Expressed sequence tag (EST) analysis is an effective approach for novel gene identification, homologous gene comparison, and transcription profiling (Li et al., 2002).

In this study, we constructed a full-length cDNA library from human embryonic lung fibroblast (HELFL) cells at late expression phase of HCMV infection using the switching mechanism at 5' end of RNA template of reverse transcriptase (SMART) technique. Partially sequenced ESTs allowed us to identify the expression and transcript structures of some predicted and unpredicted genes of HCMV.

MATERIAL AND METHODS

Virus isolates and specimens

One clinical strain of HCMV, named H, was isolated from a urine sample of a 5-month-old infant hospitalized in Shenjing Hospital of China Medical University. The H strain was passaged six times in HELFL cells, which were maintained in 1640 medium supplemented with 2% fetal bovine serum, 100 IU/mL penicillin and 100 µg/mL streptomycin, at 37°C and 5% CO₂ in a humidified incubator. HELFL cells were inoculated with the strain at a multiplicity of infection of 3-5, and were collected 72 h post-infection when more than 90% cells showed specific cytopathic effects.

Sequencing of UL/b' region of HCMV H strain DNA

Viral DNA of the H strain was purified as described by Alderete et al. (1999). The UL/b' region of HCMV H strain was sequenced using a shotgun sequencing method as described previously (Zabarovsky et al., 1994).

RNA isolation and mRNA purification

Total RNA was isolated from approximately 10^7 HCMV-infected HELF cells using Trizol agent (Invitrogen Company). The integrity and size of the isolated RNA were determined by formaldehyde agarose gel electrophoresis, and the quantity and purity of the RNA were assessed by UV spectrometry. The total RNA sample was processed according to the protocol of the mRNA purification kit (Oligotex mRNA Kits, QIAGEN). At the end of the procedure, mRNA was dissolved in 200 μ L RNase-free H₂O. The integrity of mRNA was analyzed by 1.1% agarose gel electrophoresis along with an RNA marker.

Construction of full-length cDNA library

First-strand cDNA was synthesized according to the protocol of Creator SMART cDNA Library Construction Kit (Clontech, USA): 1 μ g mRNA sample, SMART IV oligonucleotide and CDS III/3' polymerase chain reaction (PCR) primer: [oligonucleotide: 5'-AAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGG-3'; CDS III/3' PCR primer: 5'-ATTCTAGAGGCCGAGGCCGCCGACATG-d(T)30N-1N-3' (N = A, G, C, or T; N-1 = A, G, or C)] were incubated at 72°C for 2 min. First-strand buffer (5X), DTT, dNTP mix and Power Script reverse transcriptase were incubated at 42°C for 1 h.

Amplification of cDNA was performed by long distance PCR according to instructions of the protocol. To prevent distortion of representation of cDNA molecules within the total cDNA population and to maintain the double-stranded cDNA in the exponential phase of amplification, the optimal number of PCR cycles was determined by sampling an aliquot of the PCR-amplified sample at 18 cycles and increments of three cycles thereafter, up to 35 cycles. Finally, 24 PCR cycles were chosen in this case, in which the cDNA did not reach saturation by comparing the ethidium bromide intensities. A 5- μ L aliquot of amplified cDNA (from a total volume of 100 μ L) was electrophoresed on a 1.1% ethidium bromide-containing agarose gel to visualize the amount and the size distribution of the cDNA.

After digestion with restriction enzyme *Sfi*I, the cDNA was size-fractionated using CHROMA SPIN-400 columns included in the kit. Sixteen fractions were collected in separated tubes, and 3 μ L of each fraction was taken to run on a 1.1% agarose gel alongside 0.1 μ g DNA size marker (DL2000, TaKaRa, Japan) at 100 V for 20 min. The peak fractions were determined by visualizing the intensity of the bands under ultraviolet light. Four tubes (tubes 6-9) of fractions containing cDNAs were pooled into a clean 1.5-mL Eppendorf tube. According to the protocol, sodium acetate (3 M), glycogen (20 mg/mL) and 95% ethanol (-20°C) were added to the tube, and the tube was placed in a -20°C freezer overnight. After centrifugation and careful removal of the supernatant, the pellet was washed using 70% ethanol, vacuum-dried and dissolved in 100 μ L deionized H₂O.

The cDNA was ligated into dephosphorylated pBluescript II SK* vector, which was reconstructed from pBluescript II SK vector by introducing *Sfi*I A and *Sfi*I B sites. Next, 1 μ L of the total 1000 μ L ligation mixture was transferred into the competent cells of *Escherichia coli* DH/5 α by the CaCl₂ transformation method. To screen white plaques with insert fragment, the transformed cells were plated on 15-cm Petri dishes containing LB/ampicillin/IPTG/X-Gal. The colonies were counted and the titer of the unamplified library was calculated as: pfu/mL = (number of plaques) \times 10³. The ratio of white plaques (recombinants) to blue plaques (non-recombinants) was rapidly estimated as the recombination efficiency. The percentage of the recombinants = (number of white plaques) / (number of white plaques + number of blue plaques) \times 100%.

To confirm the size of inserted fragments in the library, 24 white colonies were randomly selected and amplified by PCR with M13 primers. PCR products were checked by running on a 1.1% agarose gel alongside a DNA marker (DL2000, TaKaRa).

EST sequencing and data analysis

Totally, 126 white clones were randomly selected from the cDNA library. The purified plasmid DNAs were sequenced at both 5' and 3' ends on an ABI 3730 Genetic Analyzer (Applied Biosystems) using the T7 promoter primer. Raw sequences were first trimmed to remove vector sequence and low-quality sequences using the "Crossmatch" program. ESTs of length less than 100 bp were discarded. The high-quality sequences were submitted to BLAST search at the GenBank database to compare with all available ESTs and genes to date (<http://www.ncbi.nlm.nih.gov/blast>). Sequence assembly of HCMV ESTs was performed with the CAP3 program online (<http://pbil.univ-lyon1.fr/cap3.php>).

Sequence analysis

EST clusters were generated using the online CAP3 program. ORF prediction was performed by the EditSequence program of DNASTar package.

RESULTS

Sequence of UL/b' region of HCMV H strain DNA

A sequence of 18,132 nucleotides of the UL/b' region of HCMV H strain has been assigned GenBank accession No. GQ981646.

Analysis of total RNA and mRNA

The total RNA appeared as a long smear with clear bands of 28S and 18S (data not shown). The OD₂₆₀/OD₂₈₀ ratio of the total RNA was 1.82, and the concentration was 4.62 μ g/ μ L. Therefore, the conclusion could be drawn that the total RNA obtained from the HELF cells infected with the HCMV clinical strain did not degrade apparently and that the purity was high enough for further use. The total amount of the purified mRNA was 9.84 μ g, OD₂₆₀ = 0.571, OD₂₆₀/OD₂₈₀ = 2.028.

Analysis of long distance PCR product and fractionated cDNA

The concentration of double-stranded cDNA was 0.049 $\mu\text{g}/\mu\text{L}$. The size of the amplified cDNA molecules extends up to 3 kb (Figure 1), which indicated that the long distance PCR was successful. The smear of the fractionated cDNA in tubes 1-16 mostly distributed longer than 0.5 kb (data not shown).

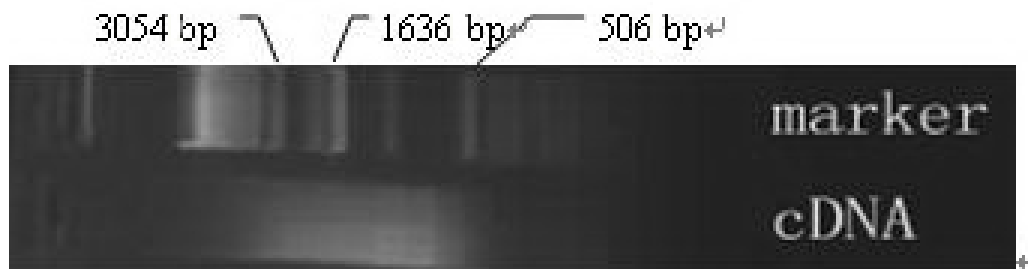


Figure 1. Generation of PCR-amplified cDNA molecules from HCMV-infected HELF cells. A 5- μL aliquot of PCR product was run on an ethidium bromide-containing agarose gel.

Titer of the unamplified library and percentage of the recombinants

A total of 1120 plaques were observed on the library dish including 1062 white plaques and 58 blue plaques. The titer of the unamplified libraries was 1.12×10^6 pfu/mL. The ratio of recombination was 94.8% according to the blue and white screening.

Identification of the cDNA inserts of the recombinants

Amplification of the 24 white clones, which were picked out randomly using M13 primers, showed that the average length of the inserts within the recombinants was 1.2 kb (Figure 2).

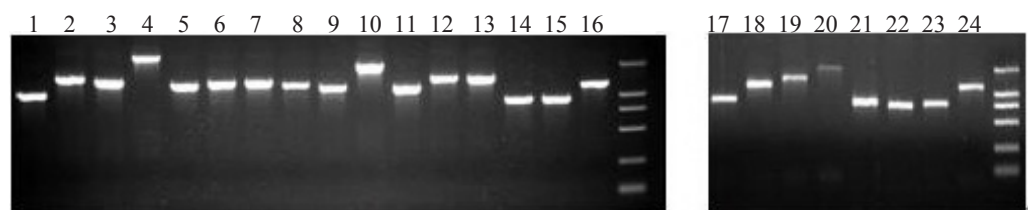


Figure 2. Insert size of cDNA clones. Twenty-four randomly chosen cDNA clones were amplified using M13 primers. The average size of PCR products of insertion fragments is 1.2 kb, determined from the gel. The marker shows 2000, 1000, 750, 500, 250, and 100 bp, respectively.

Characterization of the full-length HCMV cDNA library

After removal of the vector sequences and low-quality sequences, 110 effective sequences from the 126 cDNA clones were obtained. The results of BLAST search

showed that 44 ESTs (40.0% of 110 effective ESTs) came from HCMV, which had high sequence identity with the ESTs and genes of HCMV. And the rest showed highly homology with human ESTs and genes. Sequence assembly of the 44 HCMV ESTs with the CAP3 program online showed that 16 unigenes (6 contigs and 10 singlets) of HCMV were generated with the default options. The largest contig containing 19 ESTs (43.2% of the 44 ESTs) had 98% sequence identity with UL99 gene, and the second largest one containing 5 ESTs showed 97% sequence identity with the sequence of TRL5/IRL5 genes, while the third one containing 4 ESTs was a spliced transcript in the UL73 to UL75 intergenic region. Detailed results are shown in Table 1.

Table 1. Sequence analysis of the 44 expressed sequence tags (ESTs) from the HCMV cDNA library.

Unigene	Number of ESTs	Start position	polyA signal position	3' cleavage site	5' end confirmed before	GenBank No. of sequence
UL99 [#]	3	143907	145135	145161	143922 (Lahijani et al., 1991)	
	16	143911				
TRL5/IRL5 [#]	5	4465	*	*	4595 (Peter et al., 1987)	
UL73~UL75 [#]	3	105658	107724	107764	105683 (Scalzo et al., 2009)	
	1	105659	107724	107764		
US18 [#]	1	206542	206234	206215	207230 (Guo and Huang, 1993)	
UL4 [#]	1	13243	14748	14767	13246 (Degnin et al., 1993)	
UL65 [#]	1	96249	97875	97906	95889 (Scott et al., 2002)	
UL115 [#]	1	164589	163683	163657	164593 (Leatham et al., 1991)	
UL16~UL17 [#]	1	22389	23627	23646		
UL132	1	15543	16141	16169		
UL146~UL132	1	12955	16141	16163		
UL138	1	6316	7160	7178		
UL140~UL141	1	8292	9928	9954		
	1	8718	9928	9954		
UL144	1	11400	12821	12848		
	1	11635	12821	12833		
UL87ANTI [#]	2	130263	129503	129490		GU327645
						GU327647
UL83ANTI [#]	1	120342	121127	121150		GU327648
UL31ANTI [#]	1	37836	37281	37262		GU327646

The positions of the sequences in the table with '[#]' are given with respect to the arrangement of the prototype genome of HCMV strain AD169 (X17403.1) and the others to H strain (GQ981646). ANTI = sequence in the antisense orientation; (*) = the exact positions of the 3' ends were not sequenced successfully.

The transcripts congruent with known genes

Seven unigenes, including 31 ESTs, were homologous to the genes of UL99, TRL5/IRL5, UL73 to UL75, US18, UL4, UL65, and UL115, whose transcript structures have been characterized before. The 5' ends of the UL99 transcripts in this study contained 15 or 11 additional nucleotides upstream of the previously mapped transcription start site (Lahijani et al., 1991). The spliced transcripts in the UL73 to UL75 intergenic region were identical to the results of Scalzo et al. (2009), besides 25 or 26 additional nucleotides at their 5' end. The sequence of another cDNA clone was mostly congruent with the 1.2-kb transcript of UL115 detected by Leatham et al. (1991) (Table 1).

The transcripts of TRL5/IRL5, US18, UL4, and UL65 genes had the 5' ends downstream of the start sites published before (Peter et al., 1987; Degnin et al., 1993; Guo and Huang, 1993; Scott et al., 2002), so the transcripts may have become truncated during preparation.

UL147 ORF and ended at nucleotide 132 of the UL132 ORF, comprising residues from 13606 to 15383. The second intron comprised residues from 15396 to 15428, which was the same as that of the predicted UL132 ORF from nucleotides 149 to 179. Another transcript obtained was mostly congruent with the sequence of the predicted UL138 gene. The 5' end was located at -26 nucleotides upstream of the UL138 ORF and a polyA signal at 319~326 nucleotides downstream of the UL138 ORF. No potential CAAT box or TATA box could be found upstream of the 5' end of the transcript.

DISCUSSION

The ORFs of HCMV had been mapped on the genome by using computer analysis and sequence comparison among different HCMV strains and different CMV genomes (Chee et al., 1990; Rigoutsos et al., 2003; Dolan et al., 2004). The transcription pattern of the HCMV genome is far from being completely defined. Consequently, accurate transcript mapping is an important task to identify all protein coding regions and to understand aspects of gene expression. However, studies on HCMV transcript structures have been greatly hampered by the lack of high-throughput methods.

The SMART system was first used to generate cDNA libraries of human preimplantation embryos (Adjaye et al., 1997, 1999). In this study, a full-length cDNA library of an HCMV clinical strain at late expression phase was created with the SMART method. Despite unavoidable bias, the high-quality and fidelity of sequenced clones from this library enabled us to obtain an inherent molecular phenotype of HCMV RNAs, and could give us rudimentary insights into the gene expression of HCMV. In our study, results of BLAST search showed that 44 sequences from the 110 effective ESTs were congruent with the sequence of the HCMV genome. Although this library is a mixture of transcripts from either HCMV or HELF cells, 40% of the HCMV ESTs in the library make it possible to screen transcripts of HCMV genes.

Based on the 44 HCMV specific sequences, transcripts of 7 known and 6 predicted gene regions, and 3 antisense strands of predicted genes were demonstrated to be transcribed in our study. Among the 44 HCMV specific sequences, 19 were congruent with the UL99 ORF. These sequences were consistent with the 1.6-kb transcript, which had been characterized as a true late RNA by Adam et al. (1995), in whose research a family of eight 3'-coterminal transcripts from ORFs UL93 through UL99 have been analyzed. However, protein of the 1.6-kb transcript was not identified (Martinez et al., 1989; Lahijani et al., 1991). Another sequence in the UL73 to UL75 intergenic region was a spliced transcript containing the splice donor/acceptor sequences and the polyA signal consensus with the UL73 transcript identified in the study of Scalzo et al. (2009). However, the 5' ends of the two transcripts of UL99 and UL73 obtained from our library contained additional nucleotides upstream of the previously mapped transcription start sites. It was emphasized that the precise sequence of the 5' end is, apparently, not crucial for the initiation sites of each transcript.

The HCMV UL/b' region had been found in the Toledo strain and other low-passaged HCMV clinical isolates, which may be essential for viral infection *in vivo*. Seven ESTs from the library were homologous to the sequences of predicted genes in this region. Of the 7 ESTs, one 1.5-kb sequence comprised the total UL146 ORF, a part of the UL147 ORF and the UL132 ORF. This new spliced transcript had not been found by Lurain et al. (2006). In their results, only one 3.7-kb transcript contains UL146 ORF, which was unspliced spanning

UL146 through UL132 (Lurain et al., 2006). Combined with the results of Lurain et al. (2006), a family of 3'-coterminally transcribed transcripts containing ORFs of UL146 through UL132 may exist. In addition, the transcripts of UL138 and UL140-UL141 were identified by sequences of 3 cDNA clones. The 3' ends of the two transcripts were confirmed by the polyA tails, while their 5' end should be identified by RACE or RNase protection assays in the future.

It has been known since the early 1980s that the HCMV genome is abundantly transcribed at late stages of infection. Numerous RNAs are detectable by Northern blot analysis, which are not protein-coding or do not have any relation to a "gene". The functions of these RNAs are poorly defined and some may have a regulatory function. Plenty of antisense transcripts were identified by Zhang et al. (2007). In their study, 55% of the clones contained sequences in an anti-orientation. Two antisense transcripts overlapping the UL87 gene were obtained by screening an HCMV cDNA library with late infection in the study of Zhang et al. (2007). In our study, the library is restricted to polyadenylated messenger RNAs because of the method used to create it. From the library, three novel transcripts were identified as homologous to the antisense strand of the HCMV UL31, UL83 and UL87 predicted genes. An additional 18 adenosine residues in their transcripts, which were not coded by the genome, confirmed that they were indeed polyadenylated RNAs. Different initiation and termination sites and alternative splicing manner of UL87 antisense transcripts were found compared with the present study. Nevertheless, the antisense transcripts of UL31 and UL83 were not previously reported in any study. The three novel antisense transcripts hinted that there should be more antisense transcripts expressed in the HCMV genome.

The primary disadvantage of random sequencing of a cDNA library is that each detected sequence must be individually cloned and sequenced, which can be significantly time-consuming and expensive for a large number of sequences. Nevertheless, EST analysis is not only one of the more effective methods for gene discoveries, gene expression profiling and functional genome studies, but also one of the more efficient ways for identifying different genes (Abernathy et al., 2007; Wang et al., 2007; Malde and Jonassen, 2008; Zhou et al., 2008). The generation of an EST resource for HCMV allows a cDNA-based design of other methods for thoroughly understanding HCMV gene expression. It must be emphasized that the current study represents only the partial sequence analysis of 110 clones taken from an HCMV late-phase cDNA library. The kinetics of gene expression could not be completely elucidated by sequencing the cDNA clones.

ACKNOWLEDGMENTS

Research supported by the National Natural Science Foundation of China (#30672248, #30801254, #30770109, #30700916, and #30901625).

REFERENCES

- Abernathy JW, Xu P, Li P, Xu DH, et al. (2007). Generation and analysis of expressed sequence tags from the ciliate protozoan parasite *Ichthyophthirius multifiliis*. *BMC Genomics* 8: 176.
- Adam BL, Jervoy TY, Kohler CP, Wright GL Jr, et al. (1995). The human cytomegalovirus UL98 gene transcription unit overlaps with the pp28 true late gene (UL99) and encodes a 58-kilodalton early protein. *J. Virol.* 69: 5304-5310.
- Adjaye J, Bolton V and Monk M (1999). Developmental expression of specific genes detected in high-quality cDNA libraries from single human preimplantation embryos. *Gene* 237: 373-383.
- Adjaye J, Daniels R, Bolton V and Monk M (1997). cDNA libraries from single human preimplantation embryos.

- Genomics* 46: 337-344.
- Akter P, Cunningham C, McSharry BP, Dolan A, et al. (2003). Two novel spliced genes in human cytomegalovirus. *J. Gen. Virol.* 84: 1117-1122.
- Alderete JP, Jarrahan S and Geballe AP (1999). Translational effects of mutations and polymorphisms in a repressive upstream open reading frame of the human cytomegalovirus UL4 gene. *J. Virol.* 73: 8330-8337.
- Atalay R, Zimmermann A, Wagner M, Borst E, et al. (2002). Identification and expression of human cytomegalovirus transcription units coding for two distinct Fcγ receptor homologs. *J. Virol.* 76: 8596-8608.
- Awasthi S, Isler JA and Alwine JC (2004). Analysis of splice variants of the immediate-early 1 region of human cytomegalovirus. *J. Virol.* 78: 8191-8200.
- Bego M, Maciejewski J, Khaiboullina S, Pari G, et al. (2005). Characterization of an antisense transcript spanning the UL81-82 locus of human cytomegalovirus. *J. Virol.* 79: 11022-11034.
- Chee MS, Bankier AT, Beck S, Bohni R, et al. (1990). Analysis of the protein-coding content of the sequence of human cytomegalovirus strain AD169. *Curr. Top. Microbiol. Immunol.* 154: 125-169.
- Degnin CR, Schleiss MR, Cao J and Geballe AP (1993). Translational inhibition mediated by a short upstream open reading frame in the human cytomegalovirus gpUL4 (gp48) transcript. *J. Virol.* 67: 5514-5521.
- Dolan A, Cunningham C, Hector RD, Hassan-Walker AF, et al. (2004). Genetic content of wild-type human cytomegalovirus. *J. Gen. Virol.* 85: 1301-1312.
- Guo YW and Huang ES (1993). Characterization of a structurally tricistronic gene of human cytomegalovirus composed of U(s)18, U(s)19, and U(s)20. *J. Virol.* 67: 2043-2054.
- Jenkins C, Abendroth A and Slobedman B (2004). A novel viral transcript with homology to human interleukin-10 is expressed during latent human cytomegalovirus infection. *J. Virol.* 78: 1440-1447.
- Jones TR and Muzithras VP (1991). Fine mapping of transcripts expressed from the US6 gene family of human cytomegalovirus strain AD169. *J. Virol.* 65: 2024-2036.
- Lahijani RS, Otteson EW, Adlish JD and St Jeor SC (1991). Characterization of a human cytomegalovirus 1.6-kilobase late mRNA and identification of its putative protein product. *J. Virol.* 65: 373-381.
- Leatham MP, Witte PR and Stinski MF (1991). Alternate promoter selection within a human cytomegalovirus immediate-early and early transcription unit (UL119-115) defines true late transcripts containing open reading frames for putative viral glycoproteins. *J. Virol.* 65: 6144-6153.
- Li N, Zhao ZH, Liu ZL, Zhao XB, et al. (2002). Analysis of expressed sequence tags from porcine liver organ. *Sci. Agric. Sinica* 35: 1525-1528.
- Lurain NS, Fox AM, Lichy HM, Bhorade SM, et al. (2006). Analysis of the human cytomegalovirus genomic region from UL146 through UL147A reveals sequence hypervariability, genotypic stability, and overlapping transcripts. *J. Virol.* 3: 4.
- Malde K and Jonassen I (2008). Repeats and EST analysis for new organisms. *BMC Genomics* 9: 23.
- Martinez J, Lahijani RS and St Jeor SC (1989). Analysis of a region of the human cytomegalovirus (AD169) genome coding for a 25-kilodalton virion protein. *J. Virol.* 63: 233-241.
- Peter J Greenaway and Gavin WG (1987). Wilkinson nucleotide sequence of the most abundant transcribed early gene of human cytomegalovirus strain AD169. *Virus Res.* 7: 17-31.
- Rigoutsos I, Novotny J, Huynh T, Chin-Bow ST, et al. (2003). *In silico* pattern-based analysis of the human cytomegalovirus genome. *J. Virol.* 77: 4326-4344.
- Scalzo AA, Forbes CA, Smith LM and Loh LC (2009). Transcriptional analysis of human cytomegalovirus and rat cytomegalovirus homologues of the M73/M73.5 spliced gene family. *Arch. Virol.* 154: 65-75.
- Scott GM, Barrell BG, Oram J and Rawlinson WD (2002). Characterisation of transcripts from the human cytomegalovirus genes TRL7, UL20a, UL36, UL65, UL94, US3 and US34. *Virus Genes* 24: 39-48.
- Wang M, Guerrero FD, Pertea G and Nene VM (2007). Global comparative analysis of ESTs from the southern cattle tick, *Rhipicephalus (Boophilus) microplus*. *BMC Genomics* 8: 368.
- Wing BA and Huang ES (1995). Analysis and mapping of a family of 3'-coterminally transcribed transcripts containing coding sequences for human cytomegalovirus open reading frames UL93 through UL99. *J. Virol.* 69: 1521-1531.
- Zabarovsky ER, Kashuba VI, Pettersson B, Petrov N, et al. (1994). Shot-gun sequencing strategy for long-range genome mapping: a pilot study. *Genomics* 21: 495-500.
- Zhang G, Raghavan B, Kotur M, Cheatham J, et al. (2007). Antisense transcription in the human cytomegalovirus transcriptome. *J. Virol.* 81: 11267-11281.
- Zhou RN, Shi R, Jiang SM, Yin WB, et al. (2008). Rapid EST isolation from chromosome 1R of rye. *BMC Plant Biol.* 8: 28.