New Microbiologica, 43, 2, 58-63, 2020, ISSN 1121-7138

FULL PAPER

Characterization of Human Cytomegalovirus UL16 and UL17 Transcripts

Zhongyang Liu[1], Yanping Ma[1], Ying Lu[2], Liying Han[3], Shuang Gao[4], Bo Zheng[5], Ying Qi[1], Yujing Huang[1], Qiang Ruan[1*]

[1]*Virus Laboratory, Affiliated Shengjing Hospital, China Medical University, Shenyang, Liaoning, 110004, China;*
[2] *Department of Pathogen Biology, College of Basic Medical Sciences, Liaoning Medical University, Jinzhou, Liaoning, 121000, China;*
[3] *Children's Hospital of Nanjing Medical University, Nanjin g, Jiangsu, 210008, China;*
[4]*The Clinical Laboratory, Shenyang Women's and Children's Hospital, Shenyang, Liaoning, 110014, China;*
[5]*The Pediatric Department, Shenyang Women's and Children's Hospital, Shenyang, Liaoning, 110014, China*

SUMMARY

Up to now, the UL16-17 region of *human cytomegalovirus* (HCMV) has not been well characterized at the level of mRNA and protein, especially for the Han strain, the first clinical HCMV strain in China. In previous studies, three transcripts were detected from the UL16-17 region by northern blot analysis for Merlin strain. Transcriptions of UL16 and UL17 were also studied by 5' rapid amplification of cDNA ends (5'RACE) and deep sequencing for AD169 and Towne strains, respectively. However, details of 3' end of UL16 and UL17 transcripts have never been confirmed by 3'RACE. The expressing phage of the UL16-17 region needs further research by northern blot, too. In the present study, cDNA library screening, northern blot and RACE were used to identify the transcription characteristics of the UL16-17 region. Mainly, 3 clusters of transcripts with the same 3' end were found to be expressed from the UL16-17 region in both Han and AD169 strains. The lengths of the core transcripts among the 3 clusters were 1,254nt, 718nt and 468nt, respectively. The corresponding 5' ends are at nt23119, nt23655, nt23905 in the HCMV Han genome. The consistent 3' end is located at nt24372 in the Han genome. The 1,254nt and 468nt transcripts are transcribed in early and late phases, and the 718nt transcript is transcribed only in the late phase.

Received November 7, 2019 *Accepted* March 29, 2020

INTRODUCTION

HCMV infects 60%-90% the human population (Zuhair *et al.*, 2019). HCMV could cause severe congenital abnormalities and fatal opportunistic infections in neonates and immunosuppressed patients (Griffiths *et al.*, 2015). HCMV possess 230–235 kb of double-stranded genome DNA, which is predicted to contain 751 open reading frames (ORFs) (Stern-Ginossar *et al.*, 2012). Low passaged clinical isolates have 19 additional ORFs, which were absent from the laboratory strain AD169 genome (Cha *et al.*, 1996). 248 transcriptional start sites, 116 transcriptional termination sites and 80 splicing junctions were reported in the Towne strain (Balázs *et al.*, 2017). Many of these sites are not verified by RACE.

Point mutations and deletion mutations between different strains arise with the spread and passage of HCMV. The mutation could change the virulence, fertility and resistance to the immune system of HCMV (Wilkinson *et al.*, 2015). There is doubt regarding the HCMV pathogenicity study with the laboratory strain. Han is the first clinical isolated strain in China constructed into bacterium artificial chromosome (Zhao *et al.*, 2016). It could be used as a platform for the study of China's specific HCMV vaccine and the epidemiology of HCMV in China. The transcription study of the Han strain is far from complete.

In the UL16-17 region, two ORFs with 693nt and 315nt for UL16 and UL17, respectively, were first predicted by sequence analysis in the AD169 strain (Chee *et al.*, 1990). An additional translatable ORF with "CTG" as initiation codon within UL16 ORF was observed by ribosome profiling (Stern-Ginossar *et al.*, 2012). By northern blot analysis, a 1.4-kb transcript of UL16 was initially reported to express from early (E) phase during AD169 strain infection (Kaye *et al.*, 1992). Three transcripts including the 1.4kb transcript were detected with another probe at 72 hours post infection (hpi) (Stern-Ginossar *et al.*, 2012). Three kinds of transcription start sites were discovered by 5'RACE in AD169 strain (Akter, 2002). Deep sequencing revealed there are 7 kinds of transcripts referring to the UL16-17 region, including a UL14L-UL15A-UL16-UL17 long transcript and a UL20AS-UL19AS1 antisense strand transcript (Balázs *et al.*, 2017). Another deep sequencing study revealed that UL17 has more transcription products than UL16 at 72hpi (Gatherer *et al.*, 2011). Knowledge of mRNA structure and expression phase could provide instructive information for a function study of genes. However, no exact mRNA structure and expression phase of the UL16-17 region have been reported so far. In the present study, 3 clusters of transcripts with a coterminal site were identified from the UL16-17 region by cDNA library

Key words:
Cytomegalovirus, UL16·UL17, Transcript, Open reading frames.

Corresponding author:
Qiang Ruan
E-mail: ruanqiang2000@aliyun.com

screening, northern blot and RACE. Accurate structures of the 3 clusters of transcripts were obtained from both a clinical isolate and the AD169 strain.

MATERIALS AND METHODS

Cells and Viruses

Human embryonic lung fibroblast (HELF) was acquired from Wuhan Institute of Virology. The HELF was cultured in minimal essential medium (MEM) supplemented with 10% fetal bovine serum (Biological Industries, Israel) and 100 unit/ml penicillin and 0.1 mg/ml streptomycin. The HCMV Han strain was isolated from a urine sample of a congenitally HCMV-infected infant at the Affiliated Shengjing Hospital of China Medical University, Shenyang, China. The AD169 strain was a gift from Wuhan Institute of Virology. The HCMV strains were inoculated into HELF cells maintained in MEM supplemented with 2% fetal bovine serum and penicillin-streptomycin in an incubator at 37˚C, 5% CO2. When 100% of cells showed cytopathic effects, cells and supernatant were frozen and thawed to release the virions. The mixture was centrifuged at 10,000g, 4˚C for 30min to remove the cell debris. Virus was titrated by typical soft agar plaque assay (Yamanishi *et al.*, 1978).

RNA preparation

For immediate early (IE) and early (E) RNAs preparation, cycloheximide (CHX) (Sigma, USA) and phosphonoacetic acid (PAA) (Sigma, USA) were added to the HELF culture medium with a final concentration of 100μg/ml, respectively. Two hours later, the cells were inoculated with HCMV strains at an MOI of 3. The CHX treated cells were harvested in Trizol reagent (Invitrogen, USA) at 12hpi for preparation of IE RNAs. The PAA treated cells were harvested at 24hpi for preparation of E RNAs. For preparation of late (L) RNAs, the cells without any drug treatment were infected at an MOI of 3 and harvested at 72hpi. Uninfected HELF cells were harvested for preparation of mock RNAs. Total RNAs were extracted using a standard phenol-chloroform method (Craig & Raskas, 1974). The integrity of the isolated RNAs was analyzed by formaldehyde agarose gel electrophoresis. The concentration of the RNAs was estimated by optical density value detection (Ma *et al.*, 2011).

Screening of HCMV cDNA library

UL16-17 specific transcripts were screened with PCR from a full-length cDNA library of HCMV Han strain previously constructed in pBlue-script SK vector (Ma *et al.*, 2011). 8600 clones in 8600 tubes make up the junior grade library. Mixtures per 100 tubes from the junior grade library make up the intermediate grade library. Mixtures per 10 tubes from the intermediate grade library make up the superior grade library. There are 9 tubes of mixture in the superior grade library. A pair of UL17 gene specific primers (*Table 1*) were designed and used to screen UL16-17 positive clones from the superior grade library to the junior grade library. The PCR reaction conditions were: 98°C for 5 min, 30 cycles of 95°C for 20 sec, 55°C for 20 sec and 72°C for 40 sec, followed by final elongation at 72°C for 3 min. Inserts of the positive clones were sequenced using vector-targeting primers M13F and M13R on an ABI PRISM 3730 DNA analyzer (Applied Biosystems, CA).

Table 1 - *Gene specific primers used in the study.*

Primers	*Sequence*	*Position of the 5'end*[&]
cDNA library screening		
UL17 F	5' -ACTGTCGGTTGGAAATGTTG -3'	nt23985
UL17 R	5'-TGTTGTTTACTCCTCCTCTT -3'	nt24262
Northern Blot		
UL17P-F	5'-GTCACTGTCGGTTGGAAAT -3'	nt23982
UL17P-R*	5'-GGACATGATCGGCGTTAT-3'	nt24342
UL16P-F	5'-TTGCTTATCTGCCTCTTCCT -3'	nt23192
UL16P-R*	5'-CGCAGCGTAAACAGTCATAG-3'	nt23571
RACE		
3'-U	5'-GATGTGTATTTCTATCTGCTAC -3'	nt23315
3'-M	5'-GAACTGTCCGCGACCTTGGCG -3'	nt23682
3'-D	5'-ATGGATCACGCGCTCTTCACAC -3'	nt23942
5'-O	5'-CTTCGGGAGACCGTGCCGGATC -3'	nt24218
5'-I	5'-CGACATCAGTGGCTCTGAGTC -3'	nt24245

Note: [&]The positions of the 5' ends of the primers were referred to the Han strain. *To synthesize the northern blot probes, the promoter sequence of T7 RNA polymerase (5'-TAATACGACTCACTATAGGG-3') was added to the 5' ends of the reverse primers. The characters of F, R, P, U, M, D, O and I were abbreviated from words of forward, reverse, probe, upstream, midstream, downstream, outer and inner, respectively.

Northern blot

Northern blot analysis was performed using DIG Northern Starter Kit (Roche, 12039672910, DEU). Two UL16-17 specific single strand RNA probes, which target the 3' end of UL17 coding region and the 5'end of UL16 coding region respectively, were prepared according to the manufacturer's introductions (*Figure 1*). The RNA probe was labeled in an in-vitro transcription reaction with digoxigenin-11-UTP using a labeling mixture and an optimized transcription buffer. The primers for producing the probes are listed in *Table 1*. The efficiencies of the probes were detected by grade dilution following the protocol. 10μg total RNA for each sample was separated on a 1% formaldehyde denaturing agarose gel along with the DIG-labeled RNA molecular weight marker I (Roche, DEU) under 50v for 4~6hrs. The separated RNAs were transferred onto a positively charged nylon film (Millipore, USA) by capillary transfer, and then baked at 80˚C for 2 hrs. The membrane was pre-hybridized in the DIG EasyHyb pre-hybrid solution at 65°C for 30 min. The membrane was then hybridized to the denatured UL16 or UL17 specific probes overnight at 65°C in hybridization bags with hybridization oven. After washing twice in 2×SSC, 0.1% SDS, twice in 0.1 ×SSC, 0.1% SDS, and once in washing buffer, the membrane was blocked with the blocking solution at room temperature for 30 min followed by incubation with the anti-digoxigenin antibody conjugated to alkaline phosphatase for 30 min at room temperature. The membrane was incubated in the detection buffer for 5 min and the blots were detected with the CDP-Star using the ChemiDoc™ XRS System (Bio-Rad, USA).

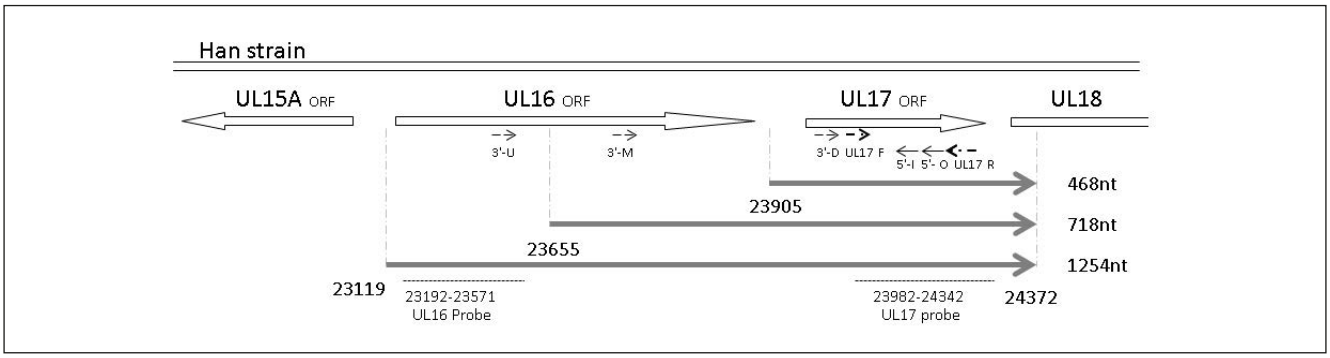

Figure 1 - *Results of cDNA library screening and relative positions of primers, probes used in the study. The small arrows below the open reading frame (ORF) indicate the primers used in cDNA library screening, 3'RACE, and 5'RACE. The large gray arrows indicate the 3 kinds of main transcripts gained in the cDNA library screening. The dashes indicate the target positions of the two probes in northern blot.*

Rapid amplification of cDNA 3' end

Rapid amplification of cDNA 3' end (3'RACE) was used to detect the 3' termini of the UL16 and UL17 transcripts by 3'-Full race core set with PrimeScript™ RTase (TAKARA, Japan). After treatment with TURBO DNA-free™ Kit (Ambion, USA) to remove possible DNA contamination, L RNAs from HCMV Han or AD169 infected HELF cells were reverse transcribed using 3'RACE adaptor at 42°C for 60 min and at 70°C for 15 min according to the manufacturer's instruction. The cDNA was amplified using UL16-17 specific primers (*Table 1*) coupled with 3'RACE Outer primer provided in the kit. Amplification was performed with initial denaturation at 94°C for 3 min followed by 30 cycles at 94°C for 30 sec, 55°C for 30 sec, and 72°C for 2 min, then a terminal extension at 72°C for 10 min. After separation with 1.5% agarose gel, the PCR products were recovered with Wizard® SV Gel and PCR Clean-Up System (Promega, USA) and ligated to pCR2.1 vector (Invitrogen, USA). The ligated products were chemically transformed into DH5α competent cells (TIANGEN, China). The selected clones were further detected by PCR with M13F/R primers, and positive clones were selected for sequencing with the M13 primers on an ABI PRISM 3730 DNA analyzer (Applied Biosystems, USA).

Rapid amplification of cDNA 5' end

Rapid amplification of cDNA 5' end (5'RACE) was used to detect the 5' termini of the UL16-17 transcripts. Total L RNAs were prepared as described in 3' RACE and used for detection with 5'-Full Race Kit (TAKARA, Japan). Alkaline Phosphatase was used to delete the nicked 5' phosphate group of the degraded transcript in the RNA preparations. The unbroken 5'cap of the transcript was removed by treatment with Tobacco Acid Pyrophosphatase (TAP), which leaves only one phosphate. Then, the 5' RACE adaptor was linked to the left phosphate group. cDNAs were generated by reverse transcription with reverse Transcriptase (RTase) and random 9 mers. The 5' ends of UL16-17 cDNA sequences were amplified by nest PCR using 5'-O and 5'-I primers (*Table 1*) coupled with 5'-Full Race outer/inner primers provided in the kit, respectively. The PCR cycles were as follows: an initial denaturation step at 94°C for 3 min, 20 cycles at 94°C for 30 sec, 55°C for 30 sec, 72°C for 2 min, and then terminal extension at 72°C for 10 min in the first-round PCR. The same amplification condition was used in the second-round PCR reaction, except for 25 cycles. Reactions without TAP and RTase, which exclude interference from the 5'-terminal phosphate of rRNA, tRNA and cracked mRNA, or from DNA contamination, were performed as two control reactions. The PCR products were separated by 1.5% agarose gel electrophoresis, purified and ligated into pCR2.1 vector. The ligated products were chemically transformed into DH5α competent cells. Positive clones in PCR identification were selected for sequencing with the M13 primers on the ABI PRISM 3730 DNA analyzer (Applied Biosystems, USA).

BLAST search and sequence analysis

Classic nucleotide to nucleotide BLAST was performed on the NCBI website. The nucleotide positions shown in this study correspond to the sequence of the HCMV Han strain and AD169 strain, respectively. DNA alignment was done by MegAlign using Clustal W algorithms. The Editseq program of the DNA star package was used to search for possible open reading frames (ORF) in the identified transcripts. The Eukaryotic promoter database in the Swiss Institute of Bioinformatics (CHE) was used as a standard to search the TATA-box, CCAAT-box, and GC-box existing in the UL16-17 region. The Signalscan program on the BIMAS website was used to search the transcription factor binding site.

Accession number

The sequences used in the analysis are accessible at NCBI-GenBank under accession numbers KJ426589.1, FJ527563.1, LT907985.1 and NC_006273.2 for Han, AD169, Towne and Merlin strains, respectively.

RESULTS

Transcripts from UL16-17 gene region screened from cDNA library

Three kinds of UL16-17 specific transcripts with approximate lengths of 468nt, 718nt and 1,254nt were identified from the cDNA library (*Figure 1*). The ~468nt transcript, which derived from 4 cDNA clones, contains UL17 ORF only. The ~718nt transcript, which derived from 2 clones, contains ORFL49W.iORF1 (Stern-Ginossar *et al.*, 2012) and UL17 ORF, while the ~1,254nt transcript, which derived from 22 clones, contains all three reported ORFs. All three transcripts were unspliced, with the same 3' end but different 5' ends. The 3' terminus is at nt24372 of the

HCMV Han genome, which is downstream of a polyadenylation signal AATAAA at nt24354-24358 after UL17 ORF.

Transcripts from UL16-17 gene region identified by Northern blot

To analyze potential transcripts from the UL16-17 region during HCMV infection, IE, E and L phase RNAs from Han or AD169 infected HELFs, and the RNAs from mock-infected cells were detected by northern blot with UL16 or UL17 specific single strand RNA probe (*Figure 2*). Using the UL17 specific probe, mainly 3 different transcripts of ~500nt, ~750nt and ~1,300nt were detected in the L RNAs, and two transcripts of ~500nt and ~1,300nt were detected in the E RNAs from both Han and AD169 infected HELF cells. Using the UL16 specific probe, one transcript of about 1,300nt in length was identified in E and L RNAs from Han and AD169 infected cells. No UL16 or UL17 specific band was observed in IE and mock RNA preparations. The lengths of UL16 and UL17 specific transcripts detected by northern blot are in accord with the results of the cDNA library screening. In some repetitive detection, a ~3000nt transcript was discovered in AD169 and Han strain with UL16 and UL17 specific probe at L RNAs, respectively.

3' and 5' ends of transcripts from UL16-17 gene region obtained by RACE

L RNAs of Han and AD169 infected cells were used to further identify the 3' and 5' termini of the UL16-17 transcripts (*Figure 3*). 27 and 32 positive clones of Han and AD169 strains were selected for sequencing in 3'RACE, respectively. 24 and 21 positive clones of Han and AD169 strains were selected for sequencing in 5'RACE, respectively. Consistent with the cDNA library screening, 3'RACE showed that all the 3'ends of the transcripts located at nt24372 of Han strain. The transcription start sites are mainly located at nt23119, nt23655 and nt23905 for Han strain. All three sites were identified from at least 4 clones. Beside the dom-

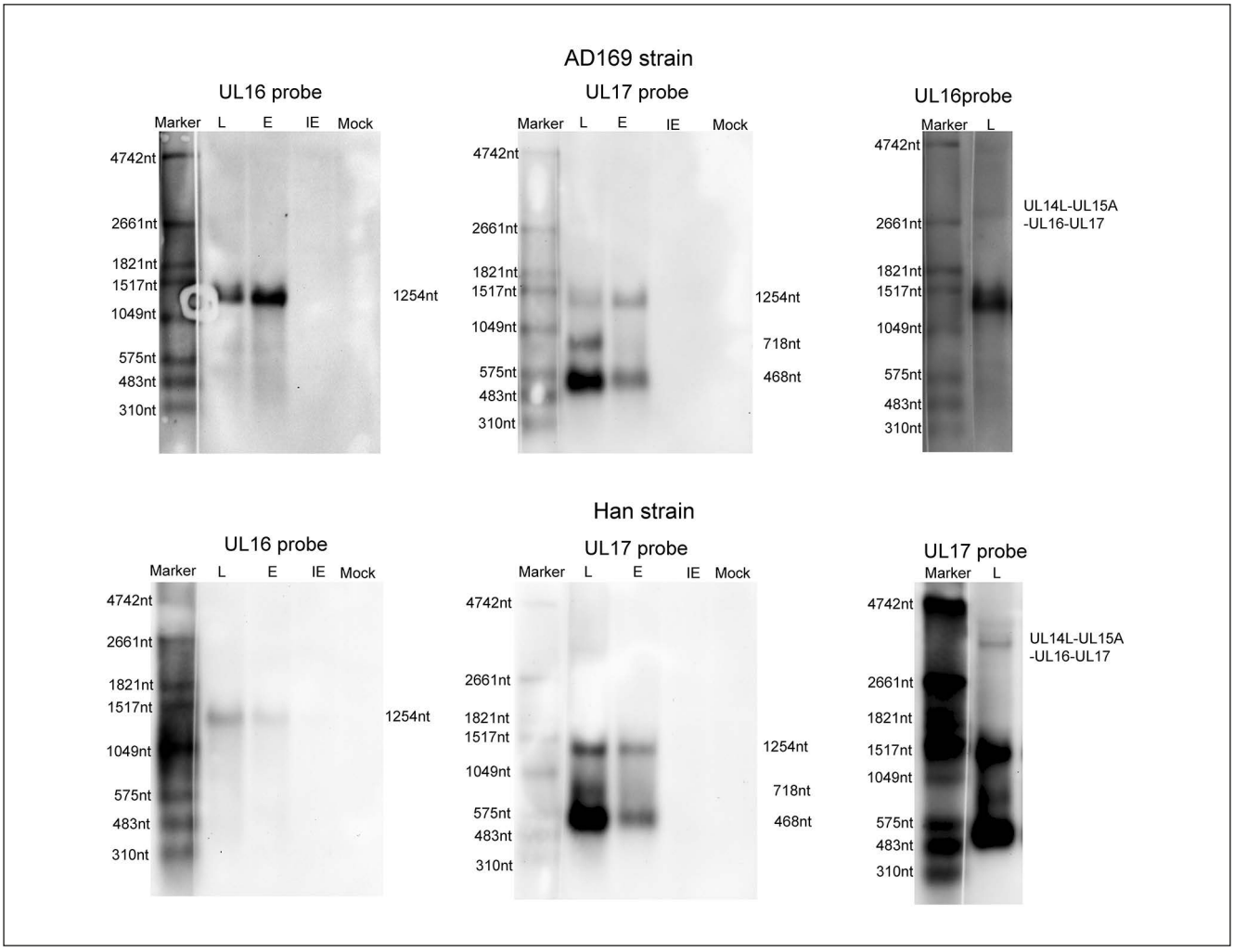

Figure 2 - *Northern blot analysis of UL16 and UL17 transcripts. Immediate early (IE), early (E) and late (L) phase RNAs from Han or AD169 infected HELF cells or RNA from mock infected cells were detected by northern blot. Using UL17 specific probe, which is located at the 3'end of UL17 gene, mainly three transcripts about 1,300nt, 750nt and 500nt were detected in late RNA preparations, two transcripts about 1,300nt and 500nt were detected in the early RNA preparation from cells infected with either Han or AD169 strain. Using UL16 specific probe, which is located at 5' end of UL16 open reading frame, one transcript of 1,300nt was detected in early and late RNAs from cells infected with either Han or AD169 strain. A ~3000nt transcript was detected in late RNA preparations for AD169 and Han strains with UL16 or UL17 specific probe, respectively.*

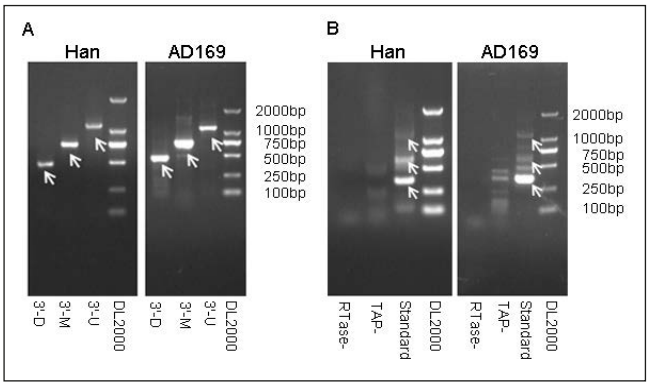

Figure 3 - *RACE results for UL16 and UL17 transcripts. A: 3'RACE results for UL16 and UL17 transcripts of Han and AD169 strain. The cDNAs of late phase RNAs from either Han or AD169 infected cells were amplified using three UL16-17 specific primers (Table 1) and the outer primer provided in the kit. Sequencing results of the indicated products showed that all the transcripts have the same transcription terminal site in both Han and AD169 strains downstream of the UL17 ORF. B: 5'RACE results for UL16 and UL17 transcripts of Han and AD169 strain. The late phase RNAs were reverse transcribed using random 9-mers from Han and AD169 infected cells, respectively. The cDNA was nest amplified using primers 5' -O and 5' -I together with the primers provided in the kit. The PCR products were detected by electrophoresis. The fragments in the bands indicated by white arrows were subsequently recovered, cloned, and sequenced. Three different transcription start sites at UL16-17 region were obtained. 'TAP-' and 'RTase-' represent the RNA samples treated without tobacco acid pyrophosphatase and reverse transcriptase, respectively.*

inant transcription start sites, there are other sites located at nt22917, nt22947, nt22957, nt23007, nt23117, nt23422, nt23591 and nt23652. For AD169 strain**,** the main transcription start sites are nt22418, nt22954 and nt23204, respectively. Besides the dominant transcription start sites, there are other sites located at nt22279, nt22285, nt22391, nt22416, nt22890 and nt22951. The 3' and main 5' ends of UL16-17 transcripts from the AD169 strain are the same as those from the Han strain. Finally, the lengths of 3 main transcripts are 1,254nt, 718nt and 468nt without consideration of the polyA tail.

Analysis of cis element in UL16-17 gene region and coding potential of UL16-17 transcripts

A conventional TATA box with core sequence of TATAAGAA locates at nt23874-23881 or nt23173-23180 aligned to the Han strain and AD169 strain, respectively. This TATA box is the promoter of the 468nt transcript. No canonical TATA box was found upstream of the coding sequences of the 1254nt and 718nt transcripts. However, some potential binding sites of transcription factors, such as AP-1, CTF and SP1, were found within 75bp upstream of the transcription start site of the two transcripts. A conventional polyadenylation signal of AATAAA locates at nt24354-24359 or nt 23653-23658 aligned to the Han strain and AD169 strain, respectively. The 1,254nt transcript spans UL16, ORFL49W.iORF1 and UL17 ORFs, the 718nt transcript initiates within the UL16 ORF and spans both ORFL49W.iORF1 and UL17 ORFs, while the 418nt transcript contains only UL17 ORF. UL17 ORF of the Han strain is 15nt longer than the AD169, Merlin and Towne strains due to a missense mutation at the stop codon.

DISCUSSION

The previous study described the UL16 transcription phase (Kaye *et al.*, 1992), transcription start sites (Akter, 2002), length of 3 transcripts from UL16-17 region, 3 ORFs (Stern-Ginossar *et al.*, 2012), 7 transcripts isoforms (Balázs *et al.*, 2017) by northern blot, 5'RACE, ribosome profiling and deep sequencing. Now, this research supports these results and supplements new information with cDNA library screening, infection phase northern blot and RACE with Han strain.

In this study, transcripts from the UL16 and UL17 regions could be divided into 3 clusters. The core transcript in the 3 clusters are 1,254nt, 718nt and 468nt. The clusters that 1,254nt and 468nt exist in begin to express at an early phase of infection, while the cluster that 718nt exists in begin to express at a late phase of infection. The transcription phase of UL16 is the same as the result reported (Kaye *et al.*, 1992). The transcription phase of UL17 has been confirmed by northern blot study for the first time. The ~3000nt band in northern blot is considered to be the UL14L-UL15A-UL16-UL17 transcript, which is 2782nt in Towne strain as reported. The band in the AD169 strain is consistent with 2782nt, but the band in the Han strain seems a little bigger. There is no mutation at the promoter sequence of UL14L-UL15A-UL16-UL17 between the Towne, AD169 and Han strains. The DNA length of the Han strain is even 30bp shorter than the other two strains at the UL14L-UL15A-UL16-UL17 gene region, so there could be another upstream promoter for the UL14L-UL15A-UL16-UL17 transcript in the Han strain.

From the RACE results, different 5' terminals around the 3 main transcripts are confirmed. Analyzing with the northern blot result, clusters 1254nt and 718nt have many different transcription start sites and relatively low mRNA level, cluster 468nt has a conservative transcription start site and a relatively high mRNA level. This may be due to the presence or absence of a typical promoter. If there is a typical "TATA" box, the transcription initiation site is accurate and single. If not, there is a cluster of initiation sites, which may lead to different protein products. Some UL16 transcripts exceed 1365nt in both the Han and AD169 strain. This could lead to a new UL16 ORF, which is 36bp longer than the one previously reported without frameshift mutation.

In 3'RACE sequencing, some 718nt and 1254nt transcripts have splice junctions. These junction sites are irregular, don't follow the "GT-AG" rule (Ohno *et al.*, 2018) and have no repeatability. In addition, these sites can't be confirmed by the 5'RACE and northern blot. The 5'RACE sequencing results showed no splice junction. The difference may derive from the reverse transcription method. The primers are oligo dT and random 9 mer for 3'RACE and 5'RACE, respectively. The 3'RACE sequencing result of the 468nt transcript showed no splice junction either. This indicates that the gap occurs as the cDNA extending process in 3'RACE. Therefore, we conclude that there is no splice junction in UL16-17 transcripts.

Due to a missense mutation of the stop codon, UL17 ORF in the Han strain is 15bp longer than in the AD169, Mer-

lin and Towne strains. Whether this extra 15bp leads to some biological change needs further study. There is no large deletion mutation or frame shift mutation between different strains in the UL16-17 region, which indicates that UL16-17 is a relatively conservative gene region. The different length of UL17 ORF could influence the immune reaction or pathogenicity of the Han strain. This indicates that the native strain is beneficial for vaccine development and for the HCMV epidemiology study.

In the HCMV genome, some genes can be transcribed into multiple transcripts with different lengths of 5´-untranslated region (5´UTR) through different transcription start sites, such as UL4, UL44 and UL57 (Kiehl *et al.*, 2003; Chang *et al.*, 1989; Leach *et al.*, 1989). 5´UTR could affect the transportation and reaction to the translation factor of mRNA (Moore *et al.*, 2018). In the present study, UL17 ORF has 3 transcripts with different 5´UTRs expressed in different infection phases, which could be regulated by different transacting factors. Different promoters of UL17 could play specific roles in different infection phases in various cell environments, in order to adapt to the host.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (81672028 and 81371788) and the Outstanding Scientific Fund of Shengjing Hospital.

References

Akter P. (2002) Transcript mapping in human cytomegalovirus strain AD169. PhD thesis, University of Glasgow.

Balázs Z., Tombácz D., Szücs A., Csabai Z., Megyeri K., et al. (2017). Long-read sequencing of human cytomegalovirus transcriptome reveals RNA isoforms carrying distinct coding potentials. *Sci Rep.* **7**, 15989.

Cha T.A., Tom E., Kemble G.W., Duke G.M., Mocarski E.S., et al. (1996). Human cytomegalovirus clinical isolates carry at least 19 genes not found in laboratory strains. *J Virol.* **70**, 78-83.

Chang C.P., Malone C.L., Stinski M.F. (1989). A human cytomegalovirus early gene has three inducible promoters that are regulated differentially at various times after infection. *J Virol.* **63**, 281-290.

Chee M.S., Bankier A.T., Beck S., Bohni R., Brown C.M., et al. (1990). Analysis of the protein-coding content of the sequence of human cytomegalovirus strain AD169. *Curr Top Microbiol Immunol.* **154**, 125-169.

Craig E.A., Raskas H.J. (1974). Effect of cycloheximide on RNA metabolism early in productive infection with adenovirus 2. *J Virol.* **14**, 26-32.

Gatherer D., Seirafian S., Cunningham C., Holton M., Dargan D.J., et al. (2011). High-resolution human cytomegalovirus transcriptome. *Proc Natl Acad Sci USA.* **108**, 19755-19760.

Griffiths P., Baraniak I., Reeves M. (2015). The pathogenesis of human cytomegalovirus. *J Pathol.* **235**, 288-297.

Kaye J., Browne H., Stoffel M., Minson T. (1992). The UL16 gene of human cytomegalovirus encodes a glycoprotein that is dispensable for growth in vitro. *J Virol.* **66**, 6609-6615.

Kiehl A., Huang L., Franchi D., Anders D.G. (2003). Multiple 5' ends of human cytomegalovirus UL57 transcripts identify a complex, cycloheximide-resistant promoter region that activates oriLyt. *Virology.* **314**, 410-422.

Leach F.S., Mocarski E.S., et al. (1989). Regulation of cytomegalovirus late-gene expression: differential use of three start sites in the transcriptional activation of ICP36 gene expression. *J Virol.* **63**, 1783-1791.

Ma Y., Wang N., Li M., Gao S., Wang L., et al. (2011). An antisense transcript in the human cytomegalovirus UL87 gene region. *Virol J.* **8**, 515.

Ma Y.P., Ruan Q., Ji Y.H., Wang N., Li M.L., et al. (2011). Novel transcripts of human cytomegalovirus clinical strain found by cDNA library screening. *Genet Mol Res.* **10**, 566-575.

Moore K.S., von Lindern M. (2018). RNA Binding Proteins and Regulation of mRNA Translation in Erythropoiesis.*Front Physiol.* **9**, 910.

Ohno K., Takeda J.I., Masuda A. (2018). Rules and tools to predict the splicing effects of exonic and intronic mutations. *Wiley Interdiscip Rev RNA.* **9**.

Stern-Ginossar N., Weisburd B., Michalski A., Le V.T., Hein M.Y., et al. (2012). Decoding human cytomegalovirus. *Science.* **338**, 1088-1093.

Wilkinson G.W., Davison A.J., Tomasec P., Fielding C.A., Aicheler R., et al. (2015). Human cytomegalovirus: taking the strain. *Med Microbiol Immunol.* **204**, 273-284.

Yamanishi K., Fogel M., Rapp F. (1978). Effect of caffeine on the replication of nonirradiated and ultraviolet-irradiated cytomegalovirus. *Intervirology.* **10**, 241-253.

Zhao F., Shen Z.Z., Liu Z.Y., Zeng W.B., Cheng S., et al. (2016). Identification and BAC construction of Han, the first characterized HCMV clinical strain in China. *J Med Virol.* **88**, 859-870.

Zuhair M., Smit G.S.A., Wallis G., Jabbar F., Smith C., et al. (2019). Estimation of the worldwide seroprevalence of cytomegalovirus: A systematic review and meta-analysis. *Rev Med Virol.* **29**, e2034.