Prevalence of human cytomegalovirus in colorectal cancer and viral gene expression profiles

Zhengzai Cai
Department of Gastroenterology, Second Affiliated Hospital, Wenzhou Medical University, Wenzhou, China

Lele Ye
Wenzhou Medical University

Yuesheng Zhu
Department of Gastroenterology, Second Affiliated Hospital, Wenzhou Medical University, Wenzhou, China

Ying Lin
Department of Gastroenterology, Second Affiliated Hospital, Wenzhou Medical University, Wenzhou, China

Gangqiang Guo
Department of Microbiology and Immunology, Institute of Molecular Virology and Immunology, Institute of Tropical Medicine, Shool of Basic Medical Sciences, Wenzhou Medical University, Wenzhou, China

Kejing Yan
Department of Microbiology and Immunology, Institute of Molecular Virology and Immunology, Institute of Tropical Medicine, Shool of Basic Medical Sciences, Wenzhou Medical University, Wenzhou, China

Yuqing Wu
Second Clinical College, Wenzhou Medical University, Wenzhou, China

Xiaoxiao Yu
Second Clinical College, Wenzhou Medical University, Wenzhou, China

Ning Ding
Department of Microbiology and Immunology, Institute of Molecular Virology and Immunology, Institute of Tropical Medicine, Shool of Basic Medical Sciences, Wenzhou Medical University, Wenzhou, China

Feng Lin
Department of General Surgery, Taizhou First People's Hospital, Taizhou, China

Shanli Zhu
Department of Microbiology and Immunology, Institute of Molecular Virology and Immunology, Institute of Tropical Medicine, Shool of Basic Medical Sciences, Wenzhou Medical University, Wenzhou, China

Xiangyang Xue (✉ wzxy001@163.com)
Department of Microbiology and Immunology, Institute of Molecular Virology and Immunology, Institute of Tropical Medicine, Shool of Basic Medical Sciences, Wenzhou Medical University, Wenzhou, China

https://orcid.org/0000-0001-8527-2037
Research

**Keywords:** Human cytomegalovirus, Colorectal cancer, Clinical feature, Gene expression profile

**DOI:** [https://doi.org/10.21203/rs.3.rs-28630/v1](https://doi.org/10.21203/rs.3.rs-28630/v1)

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](https://creativecommons.org/licenses/by/4.0/).
Abstract

Background Human cytomegalovirus (HCMV) infection plays a crucial role in the development and progression of cancer. However, the effect of HCMV on colorectal cancer (CRC) remains controversial. This study was performed to explore the pathogenesis of HCMV in CRC.

Methods HCMV DNA was detected in 74 CRC and paired normal samples by PCR. HCMV IEA protein expression was confirmed in 71 CRC biopsies by immunohistochemistry. HCMV gene expression profiles (GEPs) were further analyzed in 5 CRC tissues by transcriptome sequencing. The associations of HCMV infection with clinical features and prognosis were also evaluated.

Results The prevalence rates of HCMV in CRC tissues were 29.73% and 23.17% at the DNA and protein levels respectively, which was significantly higher than those in normal tissues (0%). Transcriptome sequencing to evaluate the GEPs revealed 119 HCMV genes in CRC tissues. The high reads of transcriptions were RNA2.7, RNA4.9, RNA5.0, RL5A, UL82, UL83, and UL70, which correlate with gene expression or regulation. Survival analysis showed that patients with CRC patients and pIEA(++) had longer overall survival (OS) than those with pIEA(+)s at the protein level. However, there was no correlation between pIEA expression and clinical features.

Conclusions HCMV, a common virus found in CRC tissues, is related to the development and progression of CRC. GEP analysis revealed genes correlated with lytic infection. Additionally, genes functioning in gene expression or regulation showed high expression in CRC. We found that CRC patients with HCMV lytic infection have a better prognosis than those with non-HCMV infection. Here, we revealed features of the pathogenic mechanism and provide insight that may be useful for targeted treatment of CRC.

Background

Colorectal cancer (CRC) is one of the most common cancers worldwide and has a high mortality rate. With increasing studies of the etiology and pathogenesis of cancer, molecular-based targeted therapy has enabled specific targeted treatment. However, few target therapies was used in CRC for that the etiology and mechanism of the occurrence and development of CRC remain unclear.

Numerous viruses have been widely recognized as carcinogens [1], such as human papillomavirus which causes cervical cancer and hepatitis B and C viruses which cause liver cancer. Additionally, many studies have suggested that pathogens participate in the occurrence and development of malignant tumors in the digestive tract, such as human papillomavirus-esophageal cancer/CRC/anal cancer, Epstein-Barr virus-esophageal cancer [2], and JC/human cytomegalovirus (HCMV)-CRC [3]. HCMV is regarded as an oncovirus that may enhance the malignancy of cancer cells or tumor-associated cells, a paradigm named as oncomodulation. HCMV has been shown to be associated with malignant glioma, leukemia, gastric cancer, and prostate cancer, among others.
Recently, HCMV has been frequently detected in CRC, and increasing evidence suggests an association between HCMV and CRC [4]. HCMV exerts oncomodulatory effects in the pathogenesis of CRC, as its genes have multiple functions. HCMV can impact infected cells, surrounding tissues, and/or immune reactions [5], encode a homologue of human interleukin-10 (LAvIL-10), control host immunity [6], and alter the tumor microenvironment [7, 8]. Chen et al. showed that in patients with CRC aged < 65 years, those who were HCMV-positive had a more favorable disease-free survival (DFS) rate than HCMV(-) patients; however, HCMV infection was associated with a shorter DFS in elderly patients with CRC [9, 10]. Studies have shown that pUS28 can enhance tumor inflammation by inducing the production of IL-6, RANTES, MCP-1, and fraktaline and activate invasion and metastasis. Polymorphisms in the UL144 gene are related to the clinical outcome of CRC, and HCMV may play an immunomodulatory role in the tumor microenvironment of CRC [11]. pUL123 and pUL122 (pIE1/IE2, pIEA) are expressed in proliferative HCMV infection and may promote virus replication and interfere with host immune responses against tumor cells [12]. However, Knosel et al. showed that HCMV is not associated with the progression and metastasis of CRC and Akintola-Ogunremi et al. failed to detect CMV DNA and protein in 24 CRC samples. Taken together, whether HCMV is closely related to the behavior and prognosis CRC and the underlying mechanism driving these changes remain unclear. Moreover, the viral gene profiles in CRC tissues have not been clearly determined.

Few studies have examined the relationship between HCMV and CRC. Here, we determined the prevalence of HCMV infection at the DNA and protein levels (Totally 865 samples). The gene expression profiles (GEPs) of HCMV in CRC tissues were detected by transcriptome sequencing. Moreover, the correlations of HCMV with the clinical features and prognosis of CRC were analyzed. These results may provide a foundation for new concepts regarding CRC pathogenesis and novel strategies for targeted treatment.

**Methods**

**Study population and specimens**

Seventy-four patients with CRC at the Second Affiliated Hospital of Wenzhou Medical University (Wenzhou, China) were enrolled in the study between February 2017 and December 2017. A total of 148 (74 x 2) paired CRC fresh-frozen and adjacent normal tissue specimens (at least 5 cm from the reception margin) were obtained during surgical resection. Both tissue statuses were confirmed by histopathologic diagnosis performed by two pathologists. The study was approved by the institutional review board of the Second Affiliated Hospital of Wenzhou Medical University. Written informed consent was obtained from each patient. Small pieces (each approximately 500 mg) of tumoral and paired adjacent normal tissues were resected, mixed with 500 µL RNA-Later, stored at 4°C for 1 h, and stored at -20°C for 24 h; the samples were stored over the long-term at -80°C. Moreover, 4 tissue microarrays (TMAs) with 717 biopsies was obtained from Professor Chang (Changhai Hospital, The Navy Military Medical University, Shanghai, China). Of these tissues, those from 669 patients had undergone curative surgery and the other 48 were from paracancerous tissues. The basic information of patients with CRC is shown in Supplementary Table 1.
Polymerase chain reaction (PCR) detection

DNA was extracted from 100 mg of these tissues by QIAamp DNA mini kit (DP304, Qiagen, Hilden, Germany) according to the manufacturer's instructions. The concentration and purity of the extracted DNA was quantified using NanoDrop 1000 spectrophotometer (Thermo Fisher Scientific, MA, USA), and then stored at -20°C until use.

A previous study showed that the UL47, UL56, and UL77 genes of HCMV can be detected simultaneously in five different tumors [13]. Therefore, conservative regions of these three genes were selected to detect HCMV infection in this study to minimize the differences in detection caused by gene mutation. PCR was carried out using specific primers for the UL47, UL55, and UL77 genes of HCMV [13] (Table 1). MRC-5 cells (Medical Research Council cell strain 5, MRC-5, ATCC, Manassas, VA, USA) transfected with HCMV were used as a positive control, and the negative control consisted of sterile double-distilled water. PCR amplification was performed in a T100TM Thermal Cycler (Bio-Rad, Hercules, CA, USA). The PCR samples contained 200 ng DNA template, 1× Taq MasterMix (Tiangen Biotech Co., Ltd., Beijing, China), and 0.3 μL of each specific forward and reverse primers. The total volume was adjusted to 25 μL with double-distilled water. After initial denaturation at 95°C for 5 min, 35 cycles of DNA amplification were performed (95°C for 30 s, annealing for 58°C 30 s, 72°C for 30 s), followed by terminal extension at 72°C for 10 min. Finally, 5 μL of PCR products were electrophoresed on 2% agarose gels and stained with ethidium bromide. Samples for which a band was detected in the correct position on the agarose gel and showing the correct sequence were regarded as HCMV(+).

RNA-seq and GEP analysis

Five HCMV(+) CRC tissues detected by PCR were selected for RNA-seq. GEPs were pre-processed with Cutadapter and FastQC to remove jointed reads and low-quality reads. Tophat (v.2.0.0) software was used to compare the sequences with the human and HCMV genomes. The GEPs of HCMV were determined using Integrated Genomics Viewer and Partek® Genomics Suite™ (version 6.5 beta, Partek, Inc., St. Louis, MO, USA) (Figure 1). Gene expression was calculated as follows:

\[
\text{FPKM} = \frac{\text{Total exon fragments}}{\text{Mapped reads (million)} \times \text{exon length (kb)}}
\]

Miri Shnayder [14] considered more than two HCMV reads in the sample as positive for viral gene expression. Another study considered at least 10 reads as viral-positive infection [15]. Considering that in most cases, only a few sequences were identical to those of specific viruses because of the large differences between the human and viral genomes, we determined that at least one viral fragment (read) in a sample was considered as positive for HCMV infection.
Immunohistochemistry (IHC) detection of pIEA

Four TMAs containing 717 biopsy samples were deparaffinized in xylene and rehydrated through a graded alcohol series. Endogenous peroxidase activity was blocked with a 0.3% solution of hydrogen peroxide and methanol. Subsequently, anti-pIEA antigens (1:50 ZM-0078, ZSGB-BIO, Beijing, China; 1:200, sc69834, Santa Cruz Biotechnology, Dallas, TX USA) with 10 mM citrate buffer (pH 6.0) were retrieved at 98°C for 30 min with 1 mM EDTA buffer (pH 7.5). TMAs were then cooled and placed in PBS solution. After blocking in 20% normal goat serum for 30 min in a humidifier chamber, the TMAs were blotted and covered with antibodies and incubated for 2 h at room temperature (22°C). The arrays were then washed with PBS and incubated for 30 min with EnVision™+Dual Link System-HRP (Dako, Carpinteria, CA, USA). After rinsing 3 times with PBS for 3 min each time, the slides were incubated with DAB reagent (Dako) for 3–5 min and evaluated under a light microscope. The TMAs were then rinsed, counter-stained with hematoxylin, and observed under a Leica microscope (DM4000b; Wetzlar, Germany).

The IHC results were independently assessed by eight researchers including two pathologists blinded to the clinical data. pIEA staining in either cells of tissues was considered as pIEA-positive expression. Scores for pIEA expression were evaluated from the extent of staining as follows: 0 (0–1%), 1 (2–24%), 2 (25–50%), and 3 (51–100%). The number of pIEA(+) cells and total adenocarcinoma cells in each block was accurately counted to determine the staining extent. Grade 0 was regarded as negative, Grades 1 and 2 were identified as low expression (pIEA(+)), and Grade 3 was identified as high expression (pIEA(++)).

Clinical features and survival analysis

Patients with CRC with intact IHC data were included in survival analysis. Continuous clinicopathological data such as patients’ age were classified as dichotomous variables. Our primary outcome of interest was disease-free survival (DFS) and overall survival (OS). DFS was defined as the number of months from the date of undergoing surgery to the date of first relapse. Patients who experienced second primary tumors of other histotypes were counted as censored in DFS analysis. OS was measured in months from the date of undergoing surgery to the date that the patient died. Kaplan–Meier’s analysis with log-rank test and univariate Cox regression analysis was performed to determine the contribution of different levels of pIEA expression to survival.

Statistical analysis

Continuous variables were analyzed using Student’s t-test or non-parametric U test. Categorical variables were analyzed using Chi square test or Fisher’s exact test. Statistical analysis and graphing were performed with SPSS 22.0 software (SPSS, Inc., Chicago, IL, USA), GraphPad Prism 7.0 (GraphPad, Inc.,
La Jolla, CA, USA), and R 3.5.1 (http://www.r-project.org/). A P value <0.05 was considered as statistically significant.

Results

HCMV prevalence is high in CRC tissues

Previous studies showed that the prevalence of HCMV is higher in CRC tissues than in normal intestinal tissues. We evaluated HCMV infection at the DNA and protein levels to confirm its prevalence in CRC tissues. First, we detected the UL47, UL56, and UL77 genes of HCMV in CRC and adjacent tissues by PCR, as these three genes can be detected simultaneously in five different tumors [13]. We defined detection of at least one of these genes as HCMV infection. Of the 74 paired samples, 29.73% patients with CRC were positive for HCMV infection, whereas HCMV was not detected in normal tissues (Figure 2A). Additionally, we detected pIEA of HCMV in 669 CRC tissues and 48 adjacent normal tissues by IHC (Supplementary Figure 1). pIEA was expressed in the cytoplasm and was mainly detected in glandular epithelial cells near the adenoma lumen of the CRC (left panel, Figure 2B). As observed in PCR, the prevalence of pIEA was higher in CRC (23.17%) than in normal tissues (0%, right panel, Figure 2B). These results demonstrate that HCMV was present in patients with CRC and suggested that HCMV is correlated with the occurrence and development of CRC.

HCMV gene expression profile in CRC

Studies have shown that HCMV genes have multiple functions, including oncomodulatory effects. However, no studies have defined the HCMV GEPs in CRC tissues. Therefore, in this study, five HCMV-DNA(+) CRC tissues were randomly selected for transcriptome-seq. However, one samples did not align with any HCMV code sequencing region. The other 4 CRC samples were compared to 119 HCMV genes (Supplementary Table 2); the highest amount of HCMV transcript accumulation is shown in Table 2, of which the RNA 2.7 [16] and RL5A genes were present in all samples. In addition, the RNAs 4.9 [17], RNA 5.0, UL70 [18], UL83 [19], UL150A [20], US8 [18], and US15 were detected in 3 samples; IRS1, RL8A, RL9A, RNA 1.2, TRS1, UL117, UL123, UL135, UL16, UL27, UL31, UL32, UL38, UL40 [20], UL41A, UL44, UL46, UL48A, UL52, UL54, UL69, UL71, UL72, UL82, UL84, UL85, UL87, UL95, UL98, US12, US22, and US9 were detected in 2 samples. Other genes were detected in only one sample (Figure 3). The GEPs showed that RL5A and RNA 2.7 had the largest number of reads in CRC, followed by RNA 4.9, RNA 5.0, UL82, UL83, and UL70. Notably, multiple gene associated with lytic infection [21], such as UL29 [22], UL54 [23, 24], UL83 [19], UL122 [25], and UL123 [25] etc. were detected in patients with CRC. This indicates that HCMV present a lytic infection state in patients with CRC.

HCMV is correlated with the occurrence and development of CRC
A previous study reported that pIEA was closely correlated to HCMV lytic infection. The major immediate early genes (MIE), UL123 and UL122 (IE1/IE2, IEA), play a critical role in subsequent viral gene expression and viral replication efficiency [26]. To further define the infected state of HCMV in patients with CRC, pIEA protein was detected by IHC. The associations of HCMV with survival and the clinical features of patients with CRC were further analyzed. Analysis of clinical characteristics showed that patients with CRC with pIEA(++) expression had longer OS than those patients with pIEA(+) expression (HR = 0.214, 95% CI: 0.047–0.969, P = 0.045, log rank P = 0.028, Supplementary Table 3), whereas no significant association was found with DFS (log rank P = 0.551, Figure 4A). However, no significant difference in survival was observed between pIEA(+) and pIEA(-) patients with CRC. Moreover, we found no significant correlation between pIEA expression and clinical information in 669 patients with CRC (Table 3).

Studies have suggested that HCMV infection affects the efficacy of postoperative radiotherapy and chemotherapy in CRC. Therefore, we also explore the influence of HCMV infection on the efficacy of postoperative radiotherapy and chemotherapy in patients with CRC. We divided the 669 patients into two groups according to chemotherapy administration to analyze the effect of HCMV on CRC. In patients with CRC who had been administered chemotherapy, survival analysis showed that patients with pIEA(++) had a longer OS than those with pIEA(+) (HR\textsubscript{yes} = 0.120 (0.015–0.942), log rank P\textsubscript{yes} = 0.016, Figure 4B), with no difference observed in DFS. No difference was observed among patients with CRC who had not been administered chemotherapy (Figure 4B).

**Discussion**

Chen et al. detected HCMV DNA by PCR in 69 (42.3%) samples resected before formalin fixation (HCMV UL55, UL73, and UL144 genes), whereas only 14 (8.6%) samples from adjacent non-neoplastic tissue showed a positive result [27]. In another study, the HCMV UL73 gene transcript was detected in 42.2% (n = 83) of CRC samples [28]. Dimberg et al. showed that HCMV DNA was significantly higher in cancerous tissue than in paired normal tissue in Swedish (39.8%, n = 119) and Vietnamese (21.9%, n = 83) patients with CRC according to qRT-PCR (the artus CMV TM PCR kit (Qiagen)) [29]. Harkins et al. detected IE1-72 immunoreactivity in 12 (80%) of 15 adenocarcinomas but not in most of the normal colonic epithelium in areas adjacent to the tumor within the same pathological section, nor was IE1-72 detected in tumor-free surgical biopsy specimens of the colon from seven of these patients [30]. Similarly, in this study, we found that the prevalence of HCMV was significantly higher in CRC than in normal colorectal tissues. We evaluated HCMV at the DNA and protein levels for the first time in a large number of samples, supporting the reliability of our results. However, different results were observed for the three genes evaluated in this study (UL47, UL56, and UL77). This may be because of differences in primer specificity and amplification efficiency of PCR for the different genes. Additionally, the number of copies of the HCMV genes may have been too low to be detected [31] because of the differences in the disease states between patients. These factors may explain the different positive rates for different HCMV genes in CRC.

HCMV latent infection is a key part of viral persistence and primary infection or reactivation. HCMV genes have diverse functions, which are closely correlated with the infection state. Previous studies revealed the
gene profiles after HCMV infection by RNA-seq in different cell types. Rossetto et al. found the expression of HCMV genes was correlated with the time after infection in CD14(+) cells: primary infection, latent state, reactivation [32]. Almost all HCMV open reading frames were expressed at 5 days post-infection, only a subset of viral-encoded RNAs but immediate early IE2 and IE1 were detected at 18 days post-infection, and expression of the entire HCMV viral genome was observed, which is mostly consistent with lytic virus replication after reactivation. Guo et al. defined the HCMV gene profiles in peripheral blood mononuclear cells from patients with SLE by RNA-seq [33]. Additionally, Tang et al. detected the HCMV sequence in CRC [13]. Shnayder et al. showed that the infective state is governed mainly by quantitative changes, with a limited number of qualitative changes, in HCMV gene expression. However, the HCMV gene expression profile has not been explored in CRC tissue, which was evaluated by next-generation sequencing in this study. A total of 119 HCMV genes were detected in CRC tissues; 18 HCMV viral genes were detected in patients with SLE [33], among which 17 were detected except for UL112, which was subsequently found to function in immunomodulation [34]. We found that these HCMV genes which were correlated with DNA replication, gene expression regulation, and immunity showed high reads in CRC tissues. Of these genes, multiple lytic infection-associated genes were detected. As lytic-associated genes, UL122 and UL123 (IE1/IE2) play a critical role in viral gene expression and viral replication [26]. pUL83 (pp65) can modulate the expression of other HCMV genes and inhibit NK cell lysis [35]. pIE can promote the expression of early and late genes. The pUL34 interaction with pIE2 and pUL44 contribute to viral gene expression and DNA replication. Additionally, UL136, US29, and US33 are involved in vDNA replication. Similar to previous studies, RNA2.7 showed high levels of transcript accumulation after infection. Genes encoding putative membrane proteins or uncharacterized proteins such as UL1, UL2, UL8, UL59, UL90, UL120, UL127, UL134, and UL148b were not detected [36]. Moreover, in this study, one sample showed HCMV infection by PCR but the HCMV gene was not detected by RNA-seq. This may be because the genes were not transcribed, or the reads were too low to be detected. Further, 115 HCMV genes were detected in one sample, possibly because of the viral titer, disease stage, or other reasons. Unfortunately, the characteristics of the HCMV GEPs in CRC remain unclear because of the small number of samples evaluated. Thus, the effect of HCMV genes on CRC and the infective status of HCMV in CRC require further analysis of larger samples sizes.

Studies have shown HCMV is related to the development and progression of CRC. Harkins et al. reported that pIEA detected by IHC in tumor tissues was associated with progression in the colon (n = 29) [30]. Analysis of the relationship between HCMV and CRC showed unclear results in different populations by PCR. HCMV is associated with a shorter DFS in non-elderly patients with CRC (n = 89), whereas the opposite results were observed in elderly patients (n = 95), which didn't found in our study. Due to the diverse results with small scale of sample size in previous studies, we analyzed the impact of pIEA on CRC in a large sample size. We found pIEA is mainly expressed in the cytoplasm in CRC cells [10], which contrasts the results observed in other cancers. Interestingly, for the first time, we found that patients with CRC with pIEA(++) had improved OS. However, OS can be affected by many factors. There was no difference in DFS in patients with CRC, contrasting the results of previous studies. Thus, pIEA(++) may play a role in prolonging the life of patients with CRC recurrence and promoting the recovery of these
patients through other processes. In our study, pIEA(+) was not associated with survival in patients with CRC who had not been administered chemotherapy, whereas pIEA(++) was positively correlated with OS in patients with CRC who had undergone chemotherapy. We suggest that the expression of pIEA can increase the response of CRC tumor cells to chemotherapy and promote them to increase the tumor immune response. As we known, patients with CRC who do not undergo chemotherapy generally show earlier cancer status and have lighter lesions, suggesting that the beneficial effects of pIEA(++) on OS may mainly play a serious role in cancer lesions. The protection mechanism of pIEA remains unclear; however, our results provide insight into the mechanism of HCMV-CRC which will be further examined in our future research.

**Conclusion**

This is the first study to detect HCMV infection in CRC in a large sample size at the DNA and protein levels. The prevalence of HCMV was higher in CRC tissues than para-carcinoma tissues. HCMV lytic infection-associated genes were found in CRC by RNA-seq. Additionally, patients with CRC with pIEA(++) had a longer OS but not DFS. HCMV may affect CRC via its gene expression, which can influence chemotherapy. The GEPs of HCMV in CRC remains unclear because of the small sample size used in this study, and thus our results should be confirmed in more samples. In conclusion, we revealed features of the pathogenic mechanism and provide insight that may be useful for targeted treatment of CRC.

**Declarations**

The authors declare that they have no competing interests.

**Funding**

This study was supported by grants from the Key R&D Program of Zhejiang Province (Grant No.: 2020C03029), National Natural Science Foundation of China (Grant No.: 81472308, 81672707), and Projects of Wenzhou Municipal Science and Technology Bureau (Grant No.:Y20150054, Y20150057).

**Acknowledgments**

We thank Professor Wenjun Chang assistance with the TMA samples and data; Drs. Zhuohui Gan and Dr. Shengqin Wang for processing RNA-seq data preliminary; Dr. Zhiguang Zhao and Dr. Qunjia Huang help to assess the IHC results.
Authors’ contribution

Z.Z.C, L.L.Y, Y.S.Z: performed the experiments, analyzed and interpreted the data. L.L.Y, X.Y.X: drafted the manuscript. K.J.Y, N.D, Y.L, Y.Q.W, G.Q.G and X.X.Y: performed the experiments and statistical analysis. F.L and X.Y.X: acquired the data and material support. : revised the manuscript. F.L, S.L.Z, X.Y.X: made contribution to the conception and design, analyzed and interpreted the data, supervised the study, provided the project funding, revised the manuscript and finally approved the version of the manuscript for publication. All authors read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Ethics approval and consent to participate

All procedures complied with the ethical standards of the relevant local and national committees on human experimentation. Informed consent or acceptable substitute was obtained from all patients before study inclusion.

Availability of data and materials

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation, to any qualified researcher.

Consent for publication

Not applicable.

References

1. Martin D, Gutkind JS. Human tumor-associated viruses and new insights into the molecular mechanisms of cancer. Oncogene. 2008;27(Suppl 2):31–42.

2. Polz-Gruszka D, Morshed K, Stec A, Polz-Dacewicz M. Prevalence of Human papillomavirus (HPV) and Epstein-Barr virus (EBV) in oral and oropharyngeal squamous cell carcinoma in south-eastern Poland. Infect Agent Cancer. 2015;10:37.
3. Chen H, Chen XZ, Waterboer T, Castro FA, Brenner H. Viral infections and colorectal cancer: a systematic review of epidemiological studies. Int J Cancer. 2015;137:12–24.

4. Chen HP, Chan YJ. The oncomodulatory role of human cytomegalovirus in colorectal cancer: implications for clinical trials. Front Oncol. 2014;4:314.

5. Yoo YD, Chiou CJ, Choi KS, Yi Y, Michelson S, Kim S, Hayward GS, Kim SJ. The IE2 regulatory protein of human cytomegalovirus induces expression of the human transforming growth factor beta1 gene through an Egr-1 binding site. J Virol. 1996;70:7062–70.

6. Slobedman B, Barry PA, Spencer JV, Avdic S, Abendroth A. Virus-encoded homologs of cellular interleukin-10 and their control of host immune function. J Virol. 2009;83:9618–29.

7. Dumortier J, Streblow DN, Moses AV, Jacobs JM, Kreklywich CN, Camp D, Smith RD, Orloff SL, Nelson JA. Human cytomegalovirus secretome contains factors that induce angiogenesis and wound healing. J Virol. 2008;82:6524–35.

8. Mason GM, Jackson S, Okecha G, Poole E, Sissons JG, Sinclair J, Wills MR. Human cytomegalovirus latency-associated proteins elicit immune-suppressive IL-10 producing CD4(+) T cells. PLoS Pathog. 2013;9:e1003635.

9. Chen HP, Jiang JK, Chen CY, Yang CY, Chen YC, Lin CH, Chou TY, Cho WL, Chan YJ. Identification of human cytomegalovirus in tumour tissues of colorectal cancer and its association with the outcome of non-elderly patients. J Gen Virol. 2016;97:2411–20.

10. Chen HP, Jiang JK, Lai PY, Chen CY, Chou TY, Chen YC, Chan CH, Lin SF, Yang CY, Chen CY, et al. Tumoral presence of human cytomegalovirus is associated with shorter disease-free survival in elderly patients with colorectal cancer and higher levels of intratumoral interleukin-17. Clin Microbiol Infect. 2014;20:664–71.

11. HP C, JK J, CH C, WH T, CY Y,YC C, TY C. CH L, virology CYJTJog: Genetic polymorphisms of the human cytomegalovirus UL144 gene in colorectal cancer and its association with clinical outcome. 2015, 96:3613–3623.

12. Michaelis M, Doerr HW, Cinatl J. The story of human cytomegalovirus and cancer: increasing evidence and open questions. Neoplasia. 2009;11:1–9.

13. Tang KW, Alaei-Mahabadi B, Samuelsson T, Lindh M, Larsson E. The landscape of viral expression and host gene fusion and adaptation in human cancer. Nat Commun. 2013;4:2513.

14. Shnayder M, Nachshon A, Krishna B, Poole E, Boshkov A, Binyamin A, Maza I, Sinclair J, Schwartz M, Stern-Ginossar N: Defining the Transcriptional Landscape during Cytomegalovirus Latency with Single-Cell RNA Sequencing. mBio 2018, 9.

15. Kazemian M, Ren M, Lin JX, Liao W, Spolksi R, Leonard WJ. Possible Human Papillomavirus 38 Contamination of Endometrial Cancer RNA Sequencing Samples in The Cancer Genome Atlas Database. J Virol. 2015;89:8967–73.

16. Forte E, Swaminathan S, Schroeder MW, Kim JY, Terhune SS, Hummel M: Tumor Necrosis Factor Alpha Induces Reactivation of Human Cytomegalovirus Independently of Myeloid Cell Differentiation following Posttranscriptional Establishment of Latency. MBio 2018, 9.
17. Kumar A, Tripathy MK, Pasquereau S, Al Moussawi F, Abbas W, Coquard L, Khan KA, Russo L, Algros MP, Valmary-Degano S, et al. The Human Cytomegalovirus Strain DB Activates Oncogenic Pathways in Mammary Epithelial Cells. EBioMedicine. 2018;30:167–83.

18. Dunn C, Chalupny NJ, Sutherland CL, Dosch S, Sivakumar PV, Johnson DC, Cosman D. Human cytomegalovirus glycoprotein UL16 causes intracellular sequestration of NKG2D ligands, protecting against natural killer cell cytotoxicity. J Exp Med. 2003;197:1427–39.

19. Cristea IM, Moorman NJ, Terhune SS, Cuevas CD, O'Keefe ES, Rout MP, Chait BT, Shenk T. Human cytomegalovirus pUL83 stimulates activity of the viral immediate-early promoter through its interaction with the cellular IFI16 protein. J Virol. 2010;84:7803–14.

20. Goodrum FD, Jordan CT, High K, Shenk T. Human cytomegalovirus gene expression during infection of primary hematopoietic progenitor cells: a model for latency. Proc Natl Acad Sci U S A. 2002;99:16255–60.

21. Van Damme E, Van Loock M. Functional annotation of human cytomegalovirus gene products: an update. Front Microbiol. 2014;5:218.

22. Mitchell DP, Savaryn JP, Moorman NJ, Shenk T, Terhune SS. Human cytomegalovirus UL28 and UL29 open reading frames encode a spliced mRNA and stimulate accumulation of immediate-early RNAs. J Virol. 2009;83:10187–97.

23. Dunn W, Chou C, Li H, Hai R, Patterson D, Stolc V, Zhu H, Liu F. Functional profiling of a human cytomegalovirus genome. Proc Natl Acad Sci U S A. 2003;100:14223–8.

24. UniProt Consortium T. UniProt: the universal protein knowledgebase. Nucleic Acids Res. 2018;46:2699.

25. Sinclair J, Sissons P. Latency and reactivation of human cytomegalovirus. J Gen Virol. 2006;87:1763–79.

26. Isomura H, Stinski MF. Coordination of late gene transcription of human cytomegalovirus with viral DNA synthesis: recombinant viruses as potential therapeutic vaccine candidates. Expert Opin Ther Targets. 2013;17:157–66.

27. Chen HP, Jiang JK, Chen CY, Chou TY, Chen YC, Chang YT, Lin SF, Chan CH, Yang CY, Lin CH, et al. Human cytomegalovirus preferentially infects the neoplastic epithelium of colorectal cancer: a quantitative and histological analysis. J Clin Virol. 2012;54:240–4.

28. Chen HP, Jiang JK, Lai PY, Teo WH, Yang CY, Chou TY, Lin CH, Chan YJ. Serological and viraemic status of human cytomegalovirus infection in patients with colorectal cancer is not correlated with viral replication and transcription in tumours. J Gen Virol. 2016;97:152–9.

29. Dimberg J, Hong TT, Skarstedt M, Lofgren S, Zar N, Matussek A. Detection of cytomegalovirus DNA in colorectal tissue from Swedish and Vietnamese patients with colorectal cancer. Anticancer Res. 2013;33:4947–50.

30. Harkins L, Volk AL, Samanta M, Mikolaenko I, Britt WJ, Bland KL, Cobbs CS. Specific localisation of human cytomegalovirus nucleic acids and proteins in human colorectal cancer. Lancet. 2002;360:1557–63.
31. Sandford GR, Schumacher U, Ettinger J, Brune W, Hayward GS, Burns WH, Voigt S. Deletion of the rat cytomegalovirus immediate-early 1 gene results in a virus capable of establishing latency, but with lower levels of acute virus replication and latency that compromise reactivation efficiency. J Gen Virol. 2010;91:616–21.

32. Rossetto CC, Tarrant-Elorza M, Pari GS. Cis and trans acting factors involved in human cytomegalovirus experimental and natural latent infection of CD14 (+) monocytes and CD34 (+) cells. PLoS Pathog. 2013;9:e1003366.

33. Guo G, Ye S, Xie S, Ye L, Lin C, Yang M, Shi X, Wang F, Li B, Li M, et al. The cytomegalovirus protein US31 induces inflammation through mono-macrophages in systemic lupus erythematosus by promoting NF-kappaB2 activation. Cell Death Dis. 2018;9:104.

34. Romania P, Cifaldi L, Pignoloni B, Starc N, D’Alicandro V, Melaiu O, Li Pira G, Giorda E, Carrozzo R, Bergvall M, et al. Identification of a Genetic Variation in ERAP1 Aminopeptidase that Prevents Human Cytomegalovirus miR-UL112-5p-Mediated Immunoevasion. Cell Rep. 2017;20:846–53.

35. Patel M, Vlahava V-M, Forbes SK, Fielding CA, Stanton RJ, Wang ECY. HCMV-Encoded NK Modulators: Lessons From and Genetic Variation. Frontiers in immunology. 2018;9:2214.

36. Cheng S, Caviness K, Buehler J, Smithey M, Nikolich-Zugich J, Goodrum F. Transcriptome-wide characterization of human cytomegalovirus in natural infection and experimental latency. Proc Natl Acad Sci U S A. 2017;114:E10586–95.

Tables

Due to technical limitations, tables are only available as a download in the supplemental files section.

Figures
Figure 1

Flow chart of GEPs analysis.
Figure 2

Detection of HCMV infection in CRC. (A) Gene detection of HCMV by PCR and its positive rates. The single bands on agarose gels at the expected sizes of 212, 200, and 215 bp demonstrated amplification of the UL47, UL56, and UL77 genes, respectively (Left panel). Of the 74 CRC samples, 22 were determined as HCMV infection with a positive rate of 29.73%, and 0% in paired normal tissues (right panel). (B) IHC revealed pIEA expression in CRC tissues and normal colorectal tissues. The brown color indicates positive staining results. pIEA was mainly expressed in glandular epithelial cells and all pIEA was detected in the cytoplasm. The middle image was enlarged by 100x under a microscope, and the left (tumor) and right (normal tissues) images were enlarged by 400x with pIEA expression and non-expression, respectively. The right image shows the prevalence of HCMV infection, with the number on the bar chart representing the positive rate of HCMV infection in groups (%) and N representing the total number of samples in this group.
Figure 3

HCMV gene expression profiles in CRC samples. Five CRC tissues with HCMV infection by PCR were selected for RNA-seq. NC_006273.2 is the genome sequence of the Merlin strain of HCMV. Four CRC samples were compared with the Merlin strain coding sequence. The bottom annotation is the Merlin strain. Each peak represents the abundance of the region.
Survival analysis. (A) Association between different levels of pIEA (+) HCMV infection and survival of patients with CRC. Patients with CRC in the pIEA (++) group showed longer OS than those in the pIEA(+) group (HR = 0.214, 95% CI: 0.047 to 0.969, log rank P = 0.028), but no association with DFS (log rank P = 0.551); no difference in prognosis was observed between pIEA(+) and pIEA(++) patients with CRC (P > 0.05). Log rank P values analyzed by Kaplan Meier (K-M) survival analysis. (B) We divided the patients with CRC according to whether they had been administered chemotherapy and found no difference among different pIEA expression in patients with CRC who had not undergone chemotherapy. In patients with CRC who had undergone chemotherapy, the pIEA(++) group showed longer OS than the pIEA(+) group.
group (HR = 0.120, 95% CI: 0.015–0.942, log rank P = 0.016); no association with DFS was detected (log rank P = 0.614). Different lines represent different expression levels of pIEA HCMV infection. Log-rank P values were from Kaplan-Meier analysis by log-rank test.

**Supplementary Files**

This is a list of supplementary files associated with this preprint. Click to download.

- [20200508Table.xls](#)
- [S1.jpg](#)