

LETTER TO THE JOURNAL

Discovery of a novel viroid-like circular RNA in colorectal cancer

Viroids, the smallest known infectious agents, were initially discovered in plants [1], and have caused significant agricultural diseases [2–4]. Recently, viroids have been identified in fungi [5–7] and bacteria [8], but none have been identified in animals. To date, no studies have explored the presence of viroids in colorectal cancer (CRC). We employed a reference-free computational method to identify a novel viroid-like circular RNA (circRNA) in CRC patients. Our study suggests that a broader class of viroids may exist in living systems.

We utilized a computer algorithm developed by Qingfa Wu's team [9, 10], which is unique in its homology independence. It employs the splitting longer reads into shorter fragments (SLS) technique and progressively filters overlapping small RNAs 2 (PFOR2), facilitating the detection of novel viroid-like circRNAs through deep RNA sequencing. SLS segments long RNA sequences into 21-nt virtual small RNAs, followed by PFOR2 analysis, which retains only 21-nt virtual internal small RNAs (ISRs) for circRNA assembly [9]. Our goal was to employ reference-free computer algorithms to investigate the existence of viroids in CRC tissues, aiming to gain new insights into the pathogenesis and treatment of CRC.

Through deep sequencing, we compiled whole transcriptome data from 12 clinical pairs of CRC samples (Supplementary Table S1). The data analysis flowchart is depicted in Figure 1A. After excluding low-quality bases and irrelevant reads, clean, high-quality reads were aligned to the human genome (hg38) using Tophat, with matched reads being discarded. The remaining reads were transformed into virtual small RNAs using the SLS program, followed by contig assembly using PFOR2. To enhance computational efficiency, the “step” parameter in the SLS

program was set to 8. During this process, over 10,000 contigs were assembled. These contigs were compared to the human genome (hg38) using BLAST, and highly homologous contigs were removed, leaving only nonhomologous ones. After screening, 5,235 contigs remained, with a GC content of 51.35% and a total length of 1,114,562 bp. The length distribution is illustrated in Figure 1B, with most fragment lengths clustered between 100 bp and 200 bp. To more accurately eliminate contigs derived from the human genome, the 5,235 contigs were individually aligned to the human genome under more stringent conditions, resulting in the identification of 130 contigs (Supplementary Table S2). Throughout the experimental design, we systematically discarded sequencing data homologous to the human genome whenever possible and used non-homologous data as the foundational dataset for PFOR2 program operations.

A total of 130 primer pairs were designed based on these sequences (Supplementary Table S3) and subsequently verified through PCR and Sanger sequencing in CRC tissues, ultimately identifying three contigs. However, only one of the three contigs tested positive in the in situ hybridization analysis using the BaseScope™ Detection Reagent Kit v2-RED, leading to its identification as a novel viroid-like circRNA in CRC patients. We propose naming it CRC-associated viroid (CCAV).

The nucleotide sequence has been submitted to GenBank, with the GenBank accession number OR538373 (<https://www.ncbi.nlm.nih.gov/nuccore/OR538373>). CCAV spans 114 nt and has a GC content of 42.11% (Figure 1C). Based on our data processing, we argue that CCAV is not derived from splicing of the human genome. Comparison with genomes of other species revealed no homology. The secondary structure of CCAV, predicted using RNAfold, shows a circular ring, distinct from the rod-like conformation with extensive base-pairing that is characteristic of all known viroids (Figure 1C).

We utilized the BaseScope™ Detection Reagent Kit v2-RED to examine CCAV levels in fresh CRC pathological sections sourced from the First Affiliated Hospital of

List of abbreviations: CRC, colorectal cancer; circRNA, circular RNA; CCAV, colorectal cancer-associated viroid; DEGs, differentially expressed genes; GO, Gene Ontology; ISRs, internal small RNAs; KEGG, Kyoto Encyclopedia of Genes and Genomes; MFE, minimum free energy; PFOR2, progressively filters overlapping small RNAs 2; SLS, splitting longer reads into shorter fragments.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2024 The Author(s). *Cancer Communications* published by John Wiley & Sons Australia, Ltd. on behalf of Sun Yat-sen University Cancer Center.

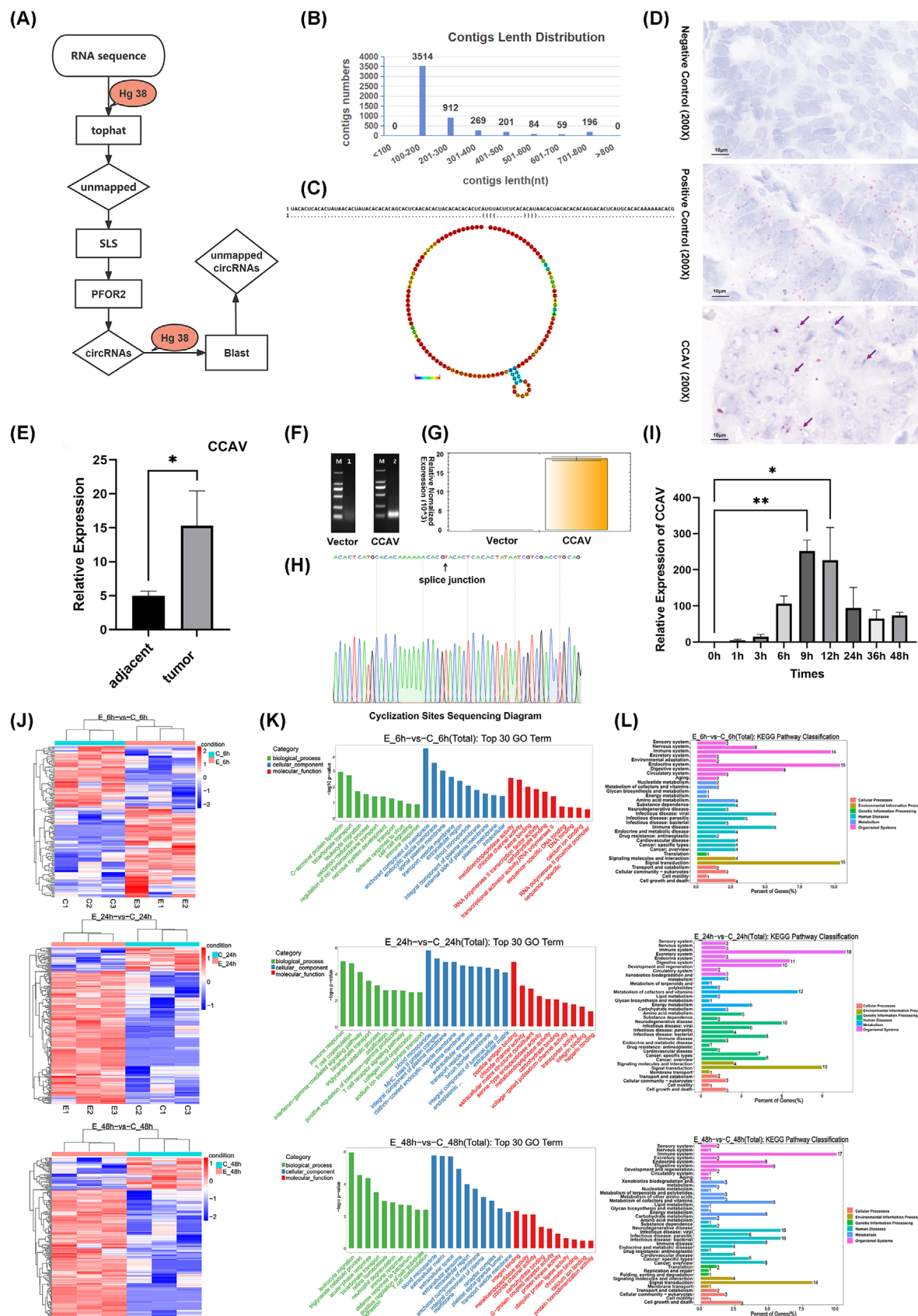


FIGURE 1 Discovery of a novel viroid-like circRNA. (A) Data analysis flowchart. High-quality reads from RNA deep sequencing were aligned to the human genome (hg38) using Tophat, with matched reads discarded. The remaining reads were transformed into virtual small RNAs using the SLS program, followed by contig assembly using PFOR2, resulting in over 10,000 contigs. To enhance computational efficiency, the “step” parameter in the SLS program was set to 8. These contigs were compared with the human genome (hg38) using BLAST,

Kunming Medical University. The BaseScope™ Probe for CCAV was designed for the target region 61-13 bp. The presence of CCAV was confirmed in fresh pathological sections using in situ hybridization analysis with the BaseScope™ Detection Reagent Kit v2-RED (Figure 1D). As depicted in the figure, the expression of CCAV in the tissue was not highly pronounced.

In further validation experiments, CCAV was detected in both tumorous and adjacent tissues of 100 CRC patients with stages I, II, III, and IV (Supplementary Table S4). During this step, linear RNAs were pre-digested by RNase R. Compared to adjacent tissues, CCAV exhibited significantly higher expression in cancer tissues (* $P < 0.05$) (Figure 1E, Supplementary Figure S1).

Numerous studies have confirmed the association between the occurrence of CRC and bacteria. Therefore, we isolated 36 strains of bacteria from fresh CRC tissues and confirmed the presence of CCAV in these 36 bacteria strains through qPCR and Sanger sequencing (Supplementary Table S5). The results demonstrated the presence of CCAV in all 36 strains of bacteria. Further experiments are required to determine whether bacteria act as carriers or hosts of CCAV and to elucidate the nature of their relationship.

The CCAV overexpression vector was constructed to further investigate the role of CCAV in CRC. The vector pCE-RB-Mam-NeoR served as the backbone (Supplemen-

tary Figure S2). To induce circularization in vivo, side flanking repeat sequences (~180 bp 5' flanking and ~200 bp 3' flanking sequence) were added to both sides of the CCAV complete sequence (Supplementary Figure S2). We confirmed that the overexpression vector of CCAV could form loops after being transfected into 293T cells (Figure 1F-H). The CCAV expression level in the overexpression vector-transfected group was significantly higher than in the empty vector control group (Figure 1F-G), and the splice junction sequencing was correct (Figure 1H).

The expression of CCAV was measured at various time points (0h, 1h, 3h, 6h, 9h, 12h, 24h, 36h, and 48h) after transfecting SW480 cells with the CCAV overexpression vector, with peak expression observed at 9h and 12h (* $P < 0.05$, ** $P < 0.01$) (Figure 1I). Based on these results, SW480 cell samples transfected with the CCAV overexpression vector were collected at 6h, 24h, and 48h for deep sequencing, aiming to further understand the changes in SW480 cells after CCAV infection.

In this analysis, whole transcriptome sequencing of 21 samples was completed, yielding a total of 321.41 GB of CleanData. The number of differentially expressed genes (DEGs) was 595, 144, and 235 at 6h, 24h, and 48h, respectively (Figure 1J). Differential mRNA Gene Ontology (GO) analysis (Figure 1K) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis (Figure 1L) were performed using the hypergeometric distribution

and highly homologous contigs were removed, leaving only nonhomologous ones. (B) Length distribution of the 5,235 contigs. The 5,235 sequences were processed using FasParser software, with most fragment lengths clustered between 100 bp and 200 bp. (C) Sequence and minimum free energy (MFE) secondary structure of CCAV. The secondary structure of CCAV was analyzed using the RNAfold web server. The optimal secondary structure, shown in dot-bracket notation, had an MFE of -0.50 kcal/mol. The thermodynamic ensemble's free energy was -2.22 kcal/mol, with the most frequent structure in the ensemble accounting for 6.16% of the total, and an ensemble diversity of 10.58. (D) CCAV BaseScope analysis. The BaseScope™ Detection Reagent Kit v2-RED was used to examine CCAV levels in fresh CRC pathological sections from the First Affiliated Hospital of Kunming Medical University. The BaseScope™ Probe for CCAV targeted the region 61-13 bp. Control probes included the BaseScope™ Negative Control Probe DapB-IZZ and the BaseScope™ Positive Control Probe Human (Hs)-PPIB-IZZ. The presence of CCAV was confirmed in the fresh pathological sections, although its expression was not highly pronounced. As shown by the arrows in the figure. (E) qPCR data. The graph shows the relative expression of CCAV in tumor tissues and adjacent normal tissues from 100 patients: 9 patients in stage I, 51 in stage II, 32 in stage III, and 8 in stage IV. CCAV expression was significantly higher in tumor tissues compared to adjacent tissues (* $P < 0.05$). (F-H) CCAV Overexpression Vector in 293T Cell Line. PCR electrophoresis images comparing empty and overexpressed vectors (F). The vector pCE-RB-Mam-NeoR served as the backbone, with side flanking repeat sequences added to both sides of the CCAV complete sequence to induce circularization in vivo. The overexpression vector of CCAV successfully formed loops after transfection into 293T cells (G, H). The expression level of CCAV in the overexpression vector-transfected group was significantly higher than in the empty vector control group (G), and the splice junction sequencing was confirmed to be correct (H). (I) Expression of CCAV at various time points after transfection of SW480 cells with the CCAV overexpression vector. The expression of CCAV peaked at 9h and 12h (* $P < 0.05$, ** $P < 0.01$). (J-L) Differentially expressed genes. SW480 cell samples transfected with the CCAV overexpression vector were collected at 6h, 24h, and 48h for deep sequencing. Differential transcripts were identified based on a P value ≤ 0.05 and a fold-change ≥ 2 . The number of differentially expressed genes (DEGs) was 595, 144, and 235 at 6h, 24h, and 48h, respectively (J). Differential mRNA Gene Ontology (GO) analysis (K) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis (L) were performed using the hypergeometric distribution test. Whole transcriptome deep sequencing revealed that DEGs generated at different time points after transfection were mainly associated with viral infection, cellular immune diseases, and tumors (L). Abbreviations: SLS, splitting longer reads into shorter fragments; PFOR2, progressively filters overlapping small RNAs 2; CRC, colorectal cancer; CCAV, colorectal cancer-associated viroid; MFE, minimum free energy; PPIB, Peptidylprolyl Isomerase B; DapB, 4-hydroxy-tetrahydronicotinate reductase; qPCR, quantitative real-time PCR; DEGs, differentially expressed genes; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes.

test. Whole transcriptome deep sequencing revealed that DEGs generated at different time points after transfection of SW480 cells with the CCAV overexpression vector were mainly associated with viral infection, cellular immune diseases, and tumors (Figure 1L).

While substantial experimental evidence is still needed to understand its role in tumor development, this discovery represents the identification of a mammalian viroid for the first time. It changes the traditional understanding that viroids exist only in plants and expands the range of viroids. This breakthrough opens new avenues for investigating human diseases and warrants further exploration.

AUTHOR CONTRIBUTIONS

Yunzhang Hu, Changning Liu, Meini Wu, and Ningzhu Hu contributed to the study's conceptualization or design. Wenliang Li and Ning Xu enrolled patients, provided study material, and elaborated the clinical information. Meini Wu, Jianfang Li, Yanhan Li, Jing Sun, and Jiandong Shi carried out the experiments and helped collect and assemble the data. Changning Liu and Jing Li performed and interpreted bioinformatic analysis. Meini Wu wrote the manuscript with input from all authors. All authors were involved in the critical review of the manuscript and approved the final version.

ACKNOWLEDGEMENTS

The authors extend their gratitude to all participants, their families, and collaborating clinicians for their invaluable contribution to this study.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

The research was approved by the Ethics Committee of the First Affiliated Hospital of Kunming Medical University (2017L3). Informed consent was signed and received from each patient.

FUNDING INFORMATION

This study was supported by grants from the National Natural Science Foundation of China (No. 31500748) and the CAMS Innovation Fund for Medical Sciences (No. 2017-I2M-3-022).

DATA AVAILABILITY STATEMENT

All data can be obtained by contacting the corresponding author via email. The SRA accession number for the whole transcriptome sequencing data mentioned in this paper is PRJNA675665 and PRJNA1005762. The complete genome sequences of CCAV in this study were submitted to GenBank under the accession numbers OR538373.

Meini Wu¹
Wenliang Li²
Ningzhu Hu¹
Changning Liu³
Jianfang Li¹
Yanhan Li¹
Ning Xu⁴
Jiandong Shi¹
Jing Sun¹
Jing Li³
Yunzhang Hu¹ 

¹*Institute of Medical Biology, Chinese Academy of Medical Sciences & Peking Union Medical College, Kunming, Yunnan, P. R. China*

²*Department of Colorectal Surgery, Yunnan Cancer Hospital & The Third Affiliated Hospital of Kunming Medical University & Yunnan Cancer Center, Kunming, Yunnan, P. R. China*

³*CAS Key Laboratory of Tropical Plant Resources and Sustainable Use, Yunnan Key Laboratory of Crop Wild Relatives Omics, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Kunming, Yunnan, P. R. China*

⁴*Department of Oncology, First Affiliated Hospital of Kunming Medical University, Kunming, Yunnan, P. R. China*

Correspondence

Yunzhang Hu, Institute of Medical Biology, Chinese Academy of Medical Sciences & Peking Union Medical College, 935 Jiaoling Road, Wuhua district, Kunming 650118, Yunnan, P. R. China.
Email: huyunzhangym@126.com

Meini Wu, Wenliang Li, Ningzhu Hu, and Changning Liu contributed equally to this paper.

ORCID

Yunzhang Hu  <https://orcid.org/0000-0001-7428-5967>

REFERENCES

- Diener TO. Potato spindle tuber "virus". IV. A replicating, low molecular weight RNA. *Virology*. 1971;45(2):411-428.
- Daròs JA, Elena SF, Flores R. Viroids: an Ariadne's thread into the RNA labyrinth. *EMBO Rep*. 2006;7(6):593-598.
- Diener TO. The viroid: biological oddity or evolutionary fossil? *Adv Virus Res*. 2001;57:137-184.
- Mascia T, Gallitelli D. Chapter 51 - Economic Significance of Satellites. In: Hadidi A, Flores R, Randles JW, Palukaitis P, editors. *Viroids and Satellites*. Boston: Academic Press; 2017. p. 555-563.

5. Dong K, Xu C, Kotta-Loizou I, Jiang J, Lv R, Kong L, et al. Novel Viroid-Like RNAs Naturally Infect a Filamentous Fungus. *Adv Sci (Weinh)*. 2023;10(3):e2204308.
6. Forgia M, Navarro B, Daghighi S, Cervera A, Gisel A, Perotto S, et al. Hybrids of RNA viruses and viroid-like elements replicate in fungi. *Nat Commun*. 2023;14(1):2591.
7. Lee BD, Neri U, Roux S, Wolf YI, Camargo AP, Krupovic M, et al. Mining metatranscriptomes reveals a vast world of viroid-like circular RNAs. *Cell*. 2023;186(3):646-661.e4.
8. Zheludev IN, Edgar RC, Lopez-Galiano MJ, de la Peña M, Babaian A, Bhatt AS, et al. Viroid-like colonists of human microbiomes. *bioRxiv*. 2024.
9. Wu Q, Wang Y, Cao M, Pantaleo V, Burgyan J, Li WX, et al. Homology-independent discovery of replicating pathogenic circular RNAs by deep sequencing and a new computational algorithm. *Proc Natl Acad Sci U S A*. 2012;109(10):3938-3943.
10. Zhang Z, Qi S, Tang N, Zhang X, Chen S, Zhu P, et al. Discovery of replicating circular RNAs by RNA-seq and computational algorithms. *PLoS Pathog*. 2014;10(12):e1004553.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.