Identification of Cryptic Putative IRESs Initiating the Translation of Nonstructural Proteins Encoded by the HRV16 Genome

Bingtian Shi

Chinese Center for Disease Control and Prevention

Qinqin Song

Chinese Center for Disease Control and Prevention

Xiaonuan Luo

Shenzhen University

Juan Song

Chinese Center for Disease Control and Prevention

Dong Xia

Chinese Center for Disease Control and Prevention

Zhiqiang Xia

Chinese Center for Disease Control and Prevention

Mi Liu

Chinese Center for Disease Control and Prevention

Wenjun Wang

Chinese Center for Disease Control and Prevention

Ruifang Wang

Chinese Center for Disease Control and Prevention

Haijun Du

Chinese Center for Disease Control and Prevention

Jun Han (✉ hanjun_sci@163.com)

National Institute for Viral Disease Control and Prevention <https://orcid.org/0000-0002-0803-4403>

Research Article

Keywords: Cryptic Putative, Nonstructural Proteins Encoded, ribosome

Posted Date: March 23rd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-348410/v1>

Abstract

Cap-dependent initiation of translation is a canonical mechanism adopted by eukaryotic cells. Internal ribosome entry site (IRES)-dependent translation is a mechanism distinct from 5' cap-dependent translation. IRES elements are located mainly in the 5'-untranslated regions (UTRs) of viral and eukaryotic mRNAs. In addition, IRESs are found in the coding regions of some viral and eukaryotic genomes and initiate the translation of some functional truncated isoforms. Here, via IRES-initiated expression of proteins, bicistronic vectors and ribosome profiling of the human rhinovirus 16 (HRV16), we found that the coding region of the nonstructural proteins P2 and P3 contained 5 putative IRES elements. These 5 putative IRESs were located within nucleotides 4286-4585, 5002-5126, 6245-6394, 6619-6718 and 6629-6778 and initiated green fluorescent protein (GFP) expression *in vitro*. This alternative mechanism might be effective and economical for eliminating the time and raw material required to synthesize the full-length polyprotein.

1. Introduction

The canonical eukaryotic translation mechanism is 5' cap-dependent (m⁷GpppN) [1]. However, picornaviruses have adopted an alternative IRES-dependent translation mechanism to initiate polyprotein translation [2-6]. Studies show that the mRNA of multiple viruses and a minority (<10%) of the mRNA in eukaryotic cells contain IRES elements and that classical internal ribosome entry sites (IRESs) are located in the 5'-untranslated region (UTR). According to the secondary structure of their host RNA, IRESs can be divided into four types. Type I IRESs are found in *Enterovirus* and *Rhinovirus* genomes [7,8]. Type II IRESs are contained in *Cardiovirus* and *Aphthovirus* genomes [9]. HCV and HCV-like IRESs are found in some members of the *Flaviviridae* and *Picornaviridae* families, such as hepatitis C Virus [10], classical swine fever virus [11,12], porcine teschovirus-1 [13], and simian virus 2 [14]. Intergenic region (IGR) IRESs are originally found in the cricket paralysis virus genome and exist widely in members of the *Discistoviridae* family [15]. In contrast, IRESs in eukaryotic cells are difficult to classify into different types because of their different structures. In addition to being located in the 5'-UTRs of eukaryotic cellular genes [16-18], some IRESs are also found in the coding region of eukaryotic cellular genes, such as those encoding the 14-3-3 and prion proteins [19,16]. In addition, our previous study showed that multiple putative IRESs are located in the coding region of the Coxsackievirus B type 3 (CVB3) genome (unpublished). IRESs in the coding regions of viral genomes might initiate the translation of specific proteins during a particular virus life cycle. The *Picornaviridae* family is one of the largest known virus families and includes many important human and animal pathogens. Picornaviruses are nonenveloped RNA viruses possessing a single-stranded RNA (positive-sense (+)) genome (7-8 kb) composed of a 5'-NTR, an open reading frame (ORF), a 3'-NTR and poly(A) tail. The ORF is translated into a large polyprotein, which is proteolytically cleaved by viral proteases (2A, 3C) to release 4 structural proteins (VP4, VP2, VP3, and VP1) and 7 nonstructural proteins (2A^{pro}, 2B, 2C, 3A, 3B, 3C^{pro}, 3D^{pol} and, in some genera L) [20,21].

Human rhinovirus 16 (HRV16) belongs to the *Rhinovirus* genus in the family *Picornaviridae*. The HRV 16 genome is a single-stranded positive-sense RNA genome with a length of approximately 7.5 kb [21].

According to the viral genome structure and classical virus replication mechanism, nonstructural proteins associated with viral replication, such as 2A^{pro}, 2B, 2C, 3A, 3B, 3C^{pro} and 3D^{pol}, are synthesized after the structural proteins. However, nonstructural proteins are the central players in the RNA replication and transcription machinery during the life cycle of RNA viruses. 3D^{pol}, an RNA-dependent RNA polymerase (RdRp), is indispensable for both replication and transcription of the viral genome. 3D^{pol} uses VPg (3B) as a primer to initiate the replication process. Both 2A^{pro} and 3C^{pro} are cleaved to form viral functional components; the molecular weight of the large protein is approximately 240 kDa. Synthesis of such a large protein is certain to affect the efficiency of viral replication. Therefore, we believe that there may be a more effective mechanism by which viruses can synthesize proteins from the genome. Considering this possibility in combination with results from previous studies, we believe that IRESs are contained within the coding region of the viral genome in addition to the 5' noncoding region.

Therefore, HRV16, was used as a model, system to search for putative IRES elements in the viral genome. To support the hypothesis, a complete experimental scheme was designed to search for putative IRESs, and then verify their function. We found 5 putative IRESs with length varying from 100 bp to 300 bp in the coding region of the nonstructural proteins P2 and P3 in the viral genome. Thus, HRV16 utilizes putative IRESs within the coding region to initiate the translation of nonstructural proteins.

2. Material And Methods

2.1. Cells and viruses

Baby hamster kidney (BHK-21) cells were cultured with minimum essential medium Eagle (Sigma-Aldrich) containing 10% fetal bovine serum (FBS) (Gibco), penicillin-streptomycin (PS) (Hyclone) and glutamine (Hyclone). Human embryonic kidney (HEK 293T) cells and H1-HeLa cells were cultured with Dulbecco's modified Eagle's medium (DMEM) (Gibco) containing 10% FBS, PS and glutamine. All three cell lines were cultured in a humidified incubator with 5% CO₂ at 37°C. Notably, H1-HeLa cells were cultured at 33°C after infection with HRV 16.

2.2. Plasmids and transfections

The nucleotide sequences of corrected ORFs from the ATG initiation codon of P2 or P3 to the 3' terminal of P2 or P3 were inserted upstream of the 5' end the green fluorescent protein (GFP) gene in the pEGFP-N1 vector (Invitrogen) to generate the fusion expression plasmids pP2(2969-4861), pP2(3632-4861), pP2(3926-4861), pP2(4256-4861), pP2(4586-4861), pP3(4586-7084), pP3(4874-7084), pP3(5177-7084), pP3(5672-7084), pP3(5993-7084), pP3(6164-7084), pP3(6395-7084) and pP3(6596-7084).

A series of bicistronic luciferase vectors were constructed on the basis of p-IRES.CHECK, which contained Renilla luciferase (R-Luc) and Firefly luciferase (F-Luc) genes interspersed with a hairpin structure (AGCTGGGCCGGGCGCGGCCGCGCCCGGCCAGGTACCTGGCCGGGCGCGGCCGCGCCCGGCCA). The 300-nucleotide sequences located immediately upstream of the ORFs in P2 and P3 were inserted between the

hairpin structure and the F-Luc gene to construct pP2-IRES-(3956-4255), pP2-IRES-(4181-4480), pP2-IRES-(4286-4585), pP3-IRES-(4877-5176), pP3-IRES-(5372-5671), pP3-IRES-(5693-5992), pP3-IRES-(6095-6394), pP3-IRES-(6419-6718) and pP3-IRES-(6629-6928). The 300-nucleotide sequences above were gradually truncated to 150 bp and 100 bp isoforms. These truncated isoforms were inserted between the hairpin structure and the F-Luc gene to generate pP2-IRES-(4286-4435), pP2-IRES-(4360-4509), pP2-IRES-(4436-4585), pP3-IRES-(4877-5026), pP3-IRES-(4952-5101), pP3-IRES-(5027-5176), pP3-IRES-(4952-5051), pP3-IRES-(4977-5076), pP3-IRES-(5002-5101), pP3-IRES-(5027-5126), pP3-IRES-(5052-5151), pP3-IRES-(5077-5176), pP3-IRES-(5002-5126), pP3-IRES-(6095-6244), pP3-IRES-(6169-6318), pP3-IRES-(6245-6394), pP3-IRES-(6245-6344), pP3-IRES-(6270-6369), pP3-IRES-(6295-6394), pP3-IRES-(6419-6568), pP3-IRES-(6494-6643), pP3-IRES-(6569-6718), pP3-IRES-(6569-6668), pP3-IRES-(6594-6693), pP3-IRES-(6619-6718), pP3-IRES-(6629-6778), pP3-IRES-(6704-6853), pP3-IRES-(6779-6928), pP3-IRES-(6629-6728), pP3-IRES-(6654-6753) and pP3-IRES-(6679-6778). The encephalomyocarditis virus (EMCV) IRES sequence was cloned between the hairpin structure and the F-Luc gene in p-IRES.CHECK to generate the positive control plasmid p-P-IRES.CHECK. p-IRES.CHECK was used as the negative control plasmid.

The F-Luc gene in p-IRES.CHECK was replaced with the nucleotide sequence of the EMCV IRES and the GFP gene to generate the vector pEMCV-GFP. The nucleotide sequence of the EMCV IRES in pEMCV-GFP was replaced by the nucleotide (nt) 4286-nt 4585, nt 5002-nt 5126, nt 6245-nt 6394, nt 6619-nt 6718 and nt 6629-nt 6778 to generate the vectors pP2-IRES(4286-4585)-GFP, pP3-IRES(5002-5126)-GFP, pP3-IRES(6245-6394)-GFP, pP3-IRES(6619-6718)-GFP and pP3-IRES(6629-6778)-GFP, respectively. The nucleic acid sequences of putative IRESs were shuffled randomly to form scrambled putative IRES sequence. The nucleotide sequence of the EMCV IRES in pEMCV-GFP was replaced by random scrambled nucleotides sequences based on the EMCV IRES and 5 putative IRESs to generate the corresponding negative control vectors, designated pEMCV-GFP(s), pP2-IRES(S4286-4585)-GFP, pP3-IRES(S5002-5126)-GFP, pP3-IRES(S6245-6394)-GFP, pP3-IRES(S6619-6718)-GFP and pP3-IRES(S6629-6778)-GFP. The nucleic acid sequences of R-Luc gene, hairpin structure and F-Luc gene in p-IRES.CHECK were replaced by the GFP gene to generate the vector pIRES-GFP.

When the cell density reached 70-80%, all plasmids were transiently transfected into cells with Lipofectamine 3000 transfection reagent (Invitrogen) according to manufacturer's instructions.

2.3. Western blotting

Cells were collected with cell scrapers and were then centrifuged at $1000 \times g$ for 3 min. Cell pellets were washed with phosphate buffer saline (PBS) and then lysed with cell lysis buffer (Beyotime) on ice for 40 min. Cell supernatants were harvested by centrifugation at 4°C and $15000 \times g$ for 10 min. Protein samples were separated by SDS-PAGE and were then transferred to nitrocellulose (NC) membranes (GE). NC membranes were blocked with 5% skim milk in PBS at room temperature for 2 h and were then incubated with anti-GFP (Proteintech), anti-GRP 78 (Abcam) and anti- β -actin (Abcam) monoclonal antibodies overnight on a shaker at 4°C . Next, NC membranes were incubated with a horseradish peroxidase-conjugated anti-mouse IgG secondary antibody (Abcam) at room temperature for 1 h. Specific

protein bands on the NC membranes were detected with enhanced chemiluminescence (ECL) detection kit (PerkinElmer).

2.4. Luciferase assay

BHK-21 cells were cultured in 96-well plates to 70-80% confluence and were then transfected with bicistronic luciferase plasmids. Luciferase activity was measured with a Dual-Luciferase Reporter Gene Assay Kit (Beyotime) according to the manufacturer's instructions at 24 h post-transfection. The ratio of F-Luc activity to R-Luc activity (F-Luc/R-Luc) showed the translation initiation ability of the putative IRES.

2.5. RNA preparation and quantitative reverse transcription-PCR (qRT-PCR)

Bicistronic reporter plasmids containing the final truncated putative positive IRES sequences were transfected into BHK-21 cells. Total RNA was extracted with TRIzol reagent (Sigma-Aldrich) 24 h post transfection and reverse transcribed to cDNA with a reverse transcription kit (Takara). Two pairs of specific primers targeting the R-Luc and F-Luc genes were designed with Oligo software (supplementary materials 2). qRT-PCR was performed with SYBR *Premix Ex Taq* II (Takara) using the primers described above.

2.6. The inhibitory of rupintrivir (AG7088) on HRV 16 3C^{pro} activity

BHK-21 cells were cultured in 6-well plates to 70-80% confluence and were then transfected with pP3(4586-7084). The BHK-21 cells were treated with 1 μ M, 5 μ M and 10 μ M AG7088 (MCE), an irreversible inhibitor of 3C protease of HRV, at 24 h post-transfection, respectively. Western blotting analysis was performed to detect the full-length proteins expressed by pP3(4586-7084) at 48 h post-transfection.

2.7. Ribosome profiling preparation

H1-HeLa cells were infected with HRV 16 at a multiplicity of infection (m.o.i.) of 0.01 in serum-free DMEM at 33°C for 1 h. H1-HeLa cells were harvested at 0, 21 and 24 h.p.i. (hours post infection). The detailed procedures were as follows: first, at different time points post infection, H1-HeLa cells were incubated with ice-cold DMEM containing cycloheximide (CHX, final concentration of 100 μ g/ml) (Amresco) for 1 min. The cell flask was placed on ice after aspiration of the medium. Second, cells were washed with ice-cold PBS containing cycloheximide (final concentration of 100 μ g/ml), and the flask was then placed on ice after aspiration of PBS. Cells were lysed with 800 μ l of ice-cold polysome lysis buffer (17.56 mM Tris-HCl, pH 7.4; 131.7 mM NaCl; 4.39 mM MgCl₂; 1.88 mM DTT; 1% Triton X-100; 0.01 U/ml DNase I (Thermo Fisher); 100 μ g/ml cycloheximide). After a 10 min incubation on ice with periodic inversion, the cell lysate supernatant was centrifuged at 4°C and 15000 $\times$ g for 10 min. The OD₂₆₀ of the diluted cell lysate supernatant was measured with a Nano Drop 2000 spectrophotometer (Thermo Fisher). The ribosome-protected segments were subjected by next-generation sequencing (Novogene).

2.8. Endoplasmic reticulum stress (ERS) induced by thapsigargin (TG)

ERS increased the expression of glucose-regulated protein 78 (GRP 78) has been reported in previous studies[22,23]. BHK-21 cells were treated with 0.25 μ M TG (Sigma-Aldrich) for 12 h. The expression of GRP 78 was detected.

BHK-21 cells were cultured in 96-well plates to 70-80% confluence and were then transfected with the bicistronic luciferase plasmids. The BHK-21 cells were treated with or without 0.25 μ M TG at 12 h post-transfection. Luciferase activity was detected at 24 h post-transfection.

2.9. RNA secondary structure prediction

The secondary structures of putative IRES were predicted by RNA structure prediction website "RNAfold web server" (<http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi>) [24].

3. Results

3.1. Identification of putative IRESs within coding regions in the HRV16 genome initiating the translation of nonstructural proteins According to previous studies [19,16], the target genome was inserted into pEGFP-N1. After transfection, GFP-fused proteins translated via de novo synthesis and dependent on an IRES-mediated mechanism were identified by Western blotting with an anti-GFP antibody. If the same putative IRESs appeared as two or more inserted fragments, bands of the same sizes would be detected in the adjacent lanes on the Western blot. To investigate the presences of putative IRES in the nonstructural protein coding region of the HRV16 genome, the sequence of the ORFs from each start codon (AUG) to the C-terminal codon of P2 or P3 were cloned into the vector pEGFP-N1 separately. If two start codons were very close to each other, only the longer sequence with two start codons was selected for cloning into pEGFP-N1. Thus, 5 sequences in the P2 region and 8 sequences in the P3 region were individually cloned into the vector pEGFP-N1; the resulting plasmids were designated pP2(2969-4861), pP2(3632-4861), pP2(3926-4861), pP2(4256-4861), pP2(4586-4861), pP3(4586-7084), pP3(4874-7084), pP3(5177-7084), pP3(5672-7084), pP3(5993-7084), pP3(6164-7084), pP3(6395-7084) and pP3(6596-7084). The 5 sequences in the P2 coding region were nt 2969-nt 4861, nt 3632-nt 4861, nt 3926-nt 4861, nt 4256-nt 4861, nt 4586-nt 4861 and 8 sequences in the P3 coding region were nt 4586-nt 7084, nt 4874-nt 7084, nt 5177-nt 7084, nt 5672-nt 7084, nt 5993-nt 7084, nt 6164-nt 7084, nt 6395-nt 7084, and nt 6596-nt 7084. Because the full-length 5' termini of P2 and P3 did not contain an AUG, the first AUG in the P2 sequence was selected in the VP1 coding region, and the first AUG in the P3 sequence was selected in the 2C coding region (Fig. 1a). The molecular weights of the GFP-fusion proteins corresponding to the abovementioned sequences were approximately 98 kDa, 75 kDa, 64 kDa, 51 kDa, 39 kDa, 121 kDa, 110 kDa, 99 kDa, 81 kDa, 69 kDa, 63 kDa, 54 kDa and 47 kDa. After pP2(2969-4861), pP2(3632-4861), pP2(3926-4861), pP2(4256-4861), and pP2(4586-4861) were transfected into BHK-21 cells, multiple protein bands appeared in each lane after Western blotting with an anti-GFP antibody. Specifically, 3 bands, 3 bands, 5 bands, 7 bands and 2 bands were found in the lanes of samples transfected with pP2(2969-4861), pP2(3632-4861), pP2(3926-4861), pP2(4256-4861), and pP2(4586-4861), respectively (Fig. 1 b). Bands corresponding to the same molecular weights namely, approximately of 55 kDa, 40 kDa

and 38 kDa detected in many adjacent lanes (Figure 1b). After removal of GFP-tag, the nucleotide sequences corresponding to the remaining part of the three proteins were nt 4256-nt 4861, nt 4481-nt 4861 and nt 4586- nt 4861, individually. 300-bp sequences upstream of the above nucleotide sequences were nt 3956- nt 4255, nt 4181- nt 4480 and nt 4286- nt 4585, individually. We concluded that the putative IRESs in the P2 region were located within the nt 3956-nt 4255, nt 4181-nt 4480 and nt 4286- nt 4585 sequences. Multiple protein bands were also detected after transfection of pP3(4586-7084), pP3(4874-7084), pP3(5177-7084), pP3(5672-7084), pP3(5993-7084), pP3(6164-7084), pP3(6395-7084) and pP3(6596-7084) (Fig. 1c). Specifically, 7, 7, 8, 7, 6, 7, 5 and 4 bands were found in the lanes of samples transfected with pP3(4586-7084), pP3(4874-7084), pP3(5177-7084), pP3(5672-7084), pP3(5993-7084), pP3(6164-7084), pP3(6395-7084) and pP3(6596-7084), respectively. Bands corresponding to the same 6 molecular weights namely, approximately 100 kDa, 75 kDa, 70 kDa, 58 kDa, 38 kDa and 32 kDa appeared in many adjacent lanes (Fig. 1c). After removal of GFP-tag, the nucleotide sequences corresponding to the remaining part of the six proteins were nt 5177-nt 7084, nt 5672-nt 7084, nt 5993-nt 7084, nt 6395-nt 7084, nt 6719-nt 7084 and nt 6929-nt 7084, individually. 300-bp sequences upstream of the above nucleotide sequences were nt 4877-nt 5176, nt 5372-nt 5671, nt 5693-nt 5992, nt 6095-nt 6394, nt 6419-nt 6718 and nt 6629-nt 6928, individually. We concluded that the putative IRESs in the P3 region were located within the nt 4877-nt 5176, nt 5372-nt 5671, nt 5693-nt 5992, nt 6095-nt 6394, nt 6419-nt 6718 and nt 6629-nt 6928 sequences. In summary, we concluded that 9 putative IRESs might be located in the nonstructural protein coding region of the HRV16 genome.

3.2. Confirmation of putative IRES elements within the nonstructural proteins coding region of the HRV16 genome through bicistronic vectors

The above results indicate that 9 putative IRESs are located upstream of 9 potential IRES-dependent proteins sequences. To search for putative IRESs sequences, according to a previous study [16,25,26], a dual-luciferase reporter plasmid (p-IRES.CHECK) with a hairpin structure ($\Delta G_{cal} = -74.4 \text{ kcal mol}^{-1}$) between the R-Luc and F-Luc sequences was constructed (Fig. 2a and 2b, left panel) based on psiCHECKTM-2. Then, 300-bp sequences upstream of the above protein sequences were inserted individually downstream of the hairpins. The 9 putative IRES candidates, namely, nt 3956-nt 4255, nt 4181-nt 4480, nt 4286-nt 4585, nt 4877-nt 5176, nt 5372-nt 5671, nt 5693-nt 5992, nt 6095-nt 6394, nt 6419-nt 6718, and nt 6629-nt 6928, were inserted upstream of F-Luc in the vector to generate pP2-IRES-(3956-4255), pP2-IRES-(4181-4480), pP2-IRES-(4286-4585), pP3-IRES-(4877-5176), pP3-IRES-(5372-5671), pP3-IRES-(5693-5992), pP3-IRES-(6095-6394), pP3-IRES-(6419-6718) and pP3-IRES-(6629-6928), respectively. The EMCV IRES sequence was inserted into p-IRES.CHECK to generate p-P-IRES.CHECK, which was used as the positive control (Fig. 2a and 2b). We established a criterion of an F-Luc/R-Luc ratio of more than 3-fold greater than that of the negative control after transfection of a plasmid containing a putative IRES sequence (Fig. 2a and 2b, right panels). The above plasmids were transfected into BHK-21 cells for 24 h and the F-Luc/R-Luc ratio of each reporter vector was calculated relative to that of the negative control. The ratios were 1.67, 1.68, 8.17, 24.52, 0.21, 1.21, 45.47, 190.49 and 10.64 after transfection of pP2-IRES-(3956-4255), pP2-IRES-(4181-4480), pP2-IRES-(4286-4585), pP3-IRES-(4877-5176), pP3-IRES-(5372-5671), pP3-IRES-(5693-5992), pP3-IRES-(6095-6394), pP3-IRES-(6419-6718) and pP3-IRES-(6629-6928), respectively. pP2-IRES-(4286-4585), pP3-IRES-(4877-5176), pP3-IRES-(6095-6394), pP3-IRES-(6419-6718) and pP3-IRES-(6629-6928) increased the F-Luc/R-Luc ratio to more than 3-fold greater than that in the negative

control BHK-21 cells. Thus, we concluded that putative IRES elements are located within the nt 4286-nt 4585 sequence in the P2 protein coding region and within the nt 4877-nt 5176, nt 6095-nt 6394, nt 6419-nt 6718 and nt 6629-nt 6928 sequences in the P3 protein coding region.

3.3. Location of the putative IRES initiating the translation of the P2 nonstructural protein

To precisely map the locations of these 5 putative IRESs in the P2 or P3 region of the genome, 15 truncated fragments with 150-bp deletions and 75-bp successive overlaps were inserted into above the bicistronic vectors to generate pP2-IRES-(4286-4435), pP2-IRES-(4360-4509), pP2-IRES-(4436-4585), pP3-IRES-(4877-5026), pP3-IRES-(4952-5101), pP3-IRES-(5027-5176), pP3-IRES-(6095-6244), pP3-IRES-(6169-6318), pP3-IRES-(6245-6394), pP3-IRES-(6419-6568), pP3-IRES-(6494-6643), pP3-IRES-(6569-6718), pP3-IRES-(6629-6778), pP3-IRES-(6704-6853) and pP3-IRES-(6779-6928) (Fig. 3a-e, left panels). To map the putative IRES within the nt 4286 nt-nt 4585 sequences, pP2-IRES-(4286-4435), pP2-IRES-(4360-4509), and pP2-IRES-(4436-4585), as well as pP2-IRES-(4286-4585) and p-P-IRES.CHECK, were transfected into BHK-21 cells, and the F-Luc/R-Luc ratios were calculated. The F-Luc/R-Luc ratios in cells transfected with pP2-IRES-(4286-4435), pP2-IRES-(4360-4509), pP2-IRES-(4436-4585), pP2-IRES-(4286-4585) and p-P-IRES.CHECK were 10.9, 3.1, 8.88, 19.22 and 59.26, respectively (Fig. 3a, right panel). The F-Luc/R-Luc ratio in cells transfected with pP2-IRES-(4286-4435) or pP2-IRES-(4436-4585) was lower than that of cells transfected with the positive control (p-P-IRES.CHECK) and approximately half of that in cells transfected with pP2-IRES-(4286-4585). Therefore, we concluded that the putative IRES is located between nt 4286 to nt 4585.

3.4. Location of the putative IRES initiating the translation of the P3 nonstructural protein

Similarly, to map the putative IRES within the nt 4877-nt 5176 sequence, pP3-IRES-(4877-5026), pP3-IRES-(4952-5101), and pP3-IRES-(5027-5176), as well as pP3-IRES-(4877-5176) and p-P-IRES.CHECK, were transfected, and the results showed that the F-Luc/R-Luc ratios in cells transfected with these plasmids were 3.66, 206.92, 58.32, 11.53 and 39.3, respectively (Fig. 3b, right panel). The F-Luc/R-Luc ratio in cells transfected with pP3-IRES-(4952-5101) and pP3-IRES-(5027-5176) exceeded that in cells transfected with pP2-IRES-(4877-5176) or the positive control vector. To further map the putative IRES location within the nt 4952-nt 5101 and nt 5027-nt 5176 sequences, 6 truncated fragments with 100-bp deletions were inserted into above the bicistronic vectors. The 6 resulting vectors were transfected, and the F-Luc/R-Luc ratios in cells transfected with pP3-IRES-(4952-5051), pP3-IRES-(4977-5076), pP3-IRES-(5002-5101), pP3-IRES-(4952-5101), pP3-IRES-(5027-5126), pP3-IRES-(5052-5151), pP3-IRES-(5077-5176), pP3-IRES-(5027-5176), pP3-IRES-(4877-5176) and p-P-IRES.CHECK were 2.27, 1.05, 100.05, 92.15, 125.81, 4.22, 2.88, 48.79, 6.42 and 41.97, respectively. The F-Luc/R-Luc ratio in cells transfected with pP3-IRES-(5002-5101) or pP3-IRES-(5027-5126) exceeded that in cells transfected with pP3-IRES-(4952-5101), pP3-IRES-(5027-5176) and the positive control vector, respectively (Fig. 4a, right panel). Because of successive overlap of the major nucleotide sequences 5002-5101 and 5027-5126, the whole sequence encompassing nt 5002-nt 5126 was inserted in p-IRES.CHECK to generate pP3-IRES-(5002-5126). After transfection, the F-Luc/R-Luc ratios in cells transfected with pP3-IRES-(5002-5101), pP3-IRES-(5027-5126), pP3-IRES-(5002-5126) and p-P-IRES.CHECK were 219.9, 211.29, 212.94 and 40.18, respectively (Fig. 4b). The putative IRES activity levels of pP3-IRES-(5002-5101), pP3-IRES-(5027-5126) and pP3-IRES-(5002-5126) were approximate. Therefore, we concluded that the sequence between nt 5002 and nt 5126 was one of the putative IRESs in the P3 region. To map the putative IRES within the nt 6095-nt 6394 sequence, pP3-IRES-(6095-6244), pP3-IRES-(6169-6318), pP3-

IRES-(6245-6394), pP3-IRES-(6095-6394) and p-P-IRES.CHECK were transfected into BHK-21 cells, and the F-Luc/R-Luc ratios were calculated. The F-Luc/R-Luc ratios in cells transfected with pP3-IRES-(6095-6244), pP3-IRES-(6169-6318), pP3-IRES-(6245-6394), pP3-IRES-(6095-6394) and p-P-IRES.CHECK were 1.41, 14.52, 32.15, 41.07 and 37.04, respectively. The F-Luc/R-Luc ratios in cells transfected with pP3-IRES-(6245-6394) were more than 0.7-fold that in cells transfected with pP3-IRES-(6095-6394) (Fig. 3c, right panel). To further map the putative IRES within the nt 6245-nt 6394, 3 truncated segments (100 bp) were inserted into the above bicistronic vectors to generate pP3-IRES-(6245-6344), pP3-IRES-(6270-6369) and pP3-IRES-(6295-6394). After transfection, the F-Luc/R-Luc ratios in cells transfected with pP3-IRES-(6245-6344), pP3-IRES-(6270-6369), pP3-IRES-(6295-6394), pP3-IRES-(6245-6394), pP3-IRES-(6095-6394) and p-P-IRES.CHECK were 6.94, 1.79, 4.35, 19.31, 24.24 and 36.7, respectively (Fig. 4c, right panel). The F-Luc/R-Luc ratio in cells transfected with pP3-IRES-(6245-6344) or pP3-IRES-(6295-6394) was less than 0.7-fold that in cells transfected with pP3-IRES-(6245-6394). Therefore, we concluded that the putative IRES is located within a 150-nucleotide region spanning nt 6245-nt 6394. To map the putative IRES within the nt 6419-nt 6718 sequence, pP3-IRES-(6419-6568), pP3-IRES-(6494-6643), pP3-IRES-(6569-6718), pP3-IRES-(6419-6718) and p-P-IRES.CHECK were transfected, the F-Luc/R-Luc ratios in cells transfected with pP3-IRES-(6419-6568), pP3-IRES-(6494-6643), pP3-IRES-(6569-6718), pP3-IRES-(6419-6718) and p-P-IRES.CHECK were 1.5, 1.67, 41.03, 23.3 and 40.2, respectively (Fig. 3d, right panel). The F-Luc/R-Luc ratio in cells transfected with pP3-IRES-(6569-6718) exceeded that in cells transfected with F-Luc/R-Luc pP3-IRES-(6419-6718). To further map the putative IRES within the nt 6569-nt 6718 sequence, 3 truncated segments (100 bp) were inserted into the above bicistronic vectors to generate pP3-IRES-(6569-6668), pP3-IRES-(6594-6693), and pP3-IRES-(6619-6718). After transfection, the F-Luc/R-Luc ratios in cells transfected with pP3-IRES-(6569-6668), pP3-IRES-(6594-6693), pP3-IRES-(6619-6718), pP3-IRES-(6569-6718), pP3-IRES-(6419-6718), and p-P-IRES.CHECK were 1.46, 0.89, 57.44, 35.74, 20.38 and 38.05, respectively. The F-Luc/R-Luc ratio in cells transfected with pP3-IRES-(6619-6718) exceeded that in cells transfected with pP3-IRES-(6569-6718) or pP3-IRES-(6419-6718) (Fig. 4d, right panel). We thus concluded that the putative IRES is located within a 100-nucleotide region spanning nt 6619 to 6718. To map the putative IRES within the nt 6629-nt 6928 sequence, pP3-IRES-(6629-6778), pP3-IRES-(6704-6853), and pP3-IRES-(6779-6928) were transfected. The F-Luc/R-Luc ratios in cells transfected with pP3-IRES-(6629-6778), pP3-IRES-(6704-6853), pP3-IRES-(6779-6928), pP3-IRES-(6629-6928) and p-P-IRES.CHECK were 11.01, 1.65, 7.73, 13.61 and 35.56, respectively (Fig. 3e, right panel). The F-Luc/R-Luc ratio in cells transfected with pP3-IRES-(6629-6778) was more than 0.7-fold that in cells transfected with pP3-IRES-(6629-6928). To further map the putative IRES within the nt 6629-6778 sequence, 3 truncated segments (100 bp) were inserted into the above bicistronic vectors to generate pP3-IRES-(6629-6728), pP3-IRES-(6654-6753), and pP3-IRES-(6679-6778). After transfection, the F-Luc/R-Luc ratio in cells transfected with pP3-IRES-(6629-6728), pP3-IRES-(6654-6753), pP3-IRES-(6679-6778), pP3-IRES-(6629-6778), pP3-IRES-(6629-6928) and p-P-IRES.CHECK were 6.3, 0.34, 2.93, 12.87, 15.41 and 59.71, respectively (Fig. 4e, right panel). The F-Luc/R-Luc ratio in cells transfected with of pP3-IRES-(6629-6728) was less than 0.7-fold that in cells transfected with pP3-IRES-(6629-6778). Therefore, we concluded that the putative IRES is located within a 150-nucleotide region spanning nucleotides 6629 to 6778. In summary, we found that 5 putative IRESs were located at nt 4286-nt 4585 in the 2C region, and at nt 5002-5126, nt 6245-nt 6394, nt

6619-nt 6718 and nt 6629-nt 6778 in the P3 region. The positions of the 5 putative IRESs in the HRV16 genome are shown (Fig. 5). 3.5. Verification of the function of the putative IRESs To verify the function of these putative IRESs in initiating protein expression in vitro, p-IRES.CHECK-GFP was constructed by inserting the hairpin structure downstream of R-Luc and replacing the F-Luc gene with the GFP gene based on p-IRES.CHECK (Fig. 6a). The abovementioned putative IRESs were inserted between the hairpin and GFP genes to generate vectors pP2-IRES(4286-4585)-GFP, pP3-IRES(5002-5126)-GFP, pP3-IRES(6245-6394)-GFP, pP3-IRES(6619-6718)-GFP and pP3-IRES(6629-6778)-GFP. The EMCV IRES sequence was inserted to generate pEMCV-GFP as the positive control vector. The scrambled putative IRES sequence shuffled randomly was inserted between the hairpin structure and the GFP gene to generate the negative control plasmid. After transfection into 293T cells for 24 h, GFP proteins in each group were detected through Western blotting with an anti-GFP antibody. Except for the nt 6245-nt 6394 and scrambled sequences, all sequences, including nt 4286-nt 4585, nt 5002-nt 5126, nt 6619-nt 6718 and nt 6629-nt 6778, effectively initiated GFP expression (Fig. 6). 3.6. Ribosome profiling of the HRV16 genome Cycloheximide stalls ribosomes on mRNA by blocking translation. To understand the ribosome profile of the HRV16 genome and further confirm the putative IRES sequences, H1-HeLa cells were treated with cycloheximide 21 and 24 h.p.i. The ribosome-protected segments were subjected to next generation sequencing. RiboSeq reads mapped to the HRV16 genome were counted (Supplemental materials 3). The ribosome profiling results showed that successive nucleotide sequences with more than 150 reads covered the positions of the 5 putative IRESs described above (Table 1). The locations were basically consistent with the genomic locations (Fig. 7). This consistency supported additional evidence for the biological function of the putative IRESs verified above by a bioinformatic approach.

4. Discussion

Translation initiation is a key step in protein synthesis in living cells [27,28]. The 5'-UTR of most eukaryotic mRNAs contains a cap structure that interacts with ribosomes to initiate translation [28]. The cap-dependent translation initiation mechanism is the canonical mechanism and is used by eukaryotes and most viruses. IRES elements were initially found in the 5'-UTR of the EMCV genome [29]. Viruses utilize an IRES-dependent translation mechanism to synthesize viral protein [30] and shut down the 5' cap-dependent translation initiation mechanism of host cells in the endoplasmic reticulum. Approximately 10% of eukaryotes mRNAs contain IRES elements, which are related to cell growth, maturation, apoptosis, stress, and cycle regulation [31,32,30]. IRESs are RNA regions with certain structures that can recruit eukaryotic ribosomes and then initiate translation under certain conditions in which cap-dependent mechanisms are suppressed, such as DNA damage [33,34] and heat shock [35]. In general, most known IRESs are located in the 5'-UTRs of mRNAs. However, a few IRESs are located in the coding regions of viral or eukaryotic genomes. In murine hepatitis virus (MHV), an IRES is located in mRNA 5; its 280 nucleotides span ORF 5a and ORF 5b, the 3' border comprises the initiation codon of ORF 5b[36].

GFP has generally been used as a tag in protein expression applications. Green fluorescence can be observed through fluorescence microscopy to indicate the expression of the fusion protein. In addition, a

low molecule weight protein fused with GFP has an increased molecular weight and can be detected with an anti-GFP antibody, obviating the need for a monoclonal antibody against a specific protein. As previous study showed, if one gene contains several putative IRESs, the fusion protein whose translation is initiated by the putative IRES can be detected in the same lane by Western blotting with an anti-GFP antibody. In this research, the Western blots showed that 9 bands corresponding to truncated proteins with the same molecular weight and located in the same position in multiple lanes, indicating that these proteins might be expressed through an IRES-dependent mechanism.

The indicated molecular weight of the protein expressed by pP2(2969-4861) was approximate to the predicted molecular weight (S Fig. 1). The insignificant differences in size might be difficult to distinguish by SDS-PAGE separation. The molecular weight of 3ABCD-GFP (111.8 kDa) was less than that of the predicted full-length fusion protein (121.8 kDa) (Fig. 1c). We concluded that this discrepancy was due to cleavage by the viral protease 3C^{pro}, because the protein expressed by pP3(4586-7084) contained a 3C^{pro} cleavage site at the junction of proteins 2C and 3A. Upon treatment of 5 μ M or 10 μ M rupintrivir, the predicted full-length fusion protein expressed by pP3(4586-7084) was detected. In addition, the inhibitory on 3C^{pro} activity was related to concentrations of rupintrivir (S Fig. 2). The region encoding the 3ABCD protein spanned nucleotides 4862-7084 in the HRV16 genome, and the predicted full-length fusion protein expressed from pP3(4874-7084) did not contain the 3C^{pro} cleavage site at the junction of proteins 2C and 3A. Additionally, the 25-kDa bands observed in every lane might be correspond to free GFP cleaved from fusion protein (Fig. 1b and 1c)[37].

In previous studies, the nucleotide sequence lengths of IRESs in coding regions were generally found to be less than 300 bp[36,16]. To identify the authenticity of putative candidate IRES elements, the putative IRES sequences with a length of 300 bp were cloned into a bicistronic expression vector containing the R-Luc and F-Luc genes. In addition, the classical EMCV IRES was inserted into a bicistronic expression vector to generate a positive control vector, similar to the method used in other research [38]. A hairpin structure ($\Delta G_{cal} = -74.4 \text{ kcal mol}^{-1}$) was inserted between R-Luc and F-Luc genes to guarantee that F-Luc gene was translated without interference of R-Luc gene. As a barrier, the stable hairpin structure prevented ribosomes from reading through the stop codon of the R-Luc ORF but did not affect the expression of downstream genes [39,40,16]. The R-Luc gene in the bicistronic expression vector was translated in a cap-dependent manner. In contrast, F-Luc gene expression was dependent on the inserted nucleotide sequence. Putative IRES activity was represented by the F-Luc/R-Luc ratio relative to that in the negative control cells. According to our previous research (unpublished), the criterion for a true putative IRES was an F-Luc/R-Luc ratio at least 3-fold greater than that in the negative control cells. The results showed that 5 putative IRES elements were located in the nonstructural protein coding region of the HRV16 genome. Deletion analysis was conducted to map the ranges of the 5' and 3' boundaries of these putative IRESs. If the putative IRES activity of the truncated isoform was greater than 0.7-fold that of the intact nucleotide sequence, the putative IRES element was deemed to be located in the truncated region.

According to this criterion, the putative IRES activity of nucleotide sequence 4286-4585 was dependent on nt 4286-nt 4435 and nt 4436-nt 4585 (Fig. 3a). In mapping the putative IRES within the nt 6095-nt 6394 region, the putative IRES activity of nucleotide sequences 6245-6394 was found to be greater than 0.7-fold that of the full-length nucleotide sequence 6095-6394 (Fig. 3c). Further deletion mutation of nucleotide sequence 6245-6394 led to a significant reduction in putative IRES activity (Fig. 4c). Similar experimental results were found in mapping putative IRESs within the nt 6629-nt 6928 region (Fig. 3e and Fig. 4e). The putative IRES activity of the truncated region was greater than 0.7-fold that of the full-length sequences, showing that the truncated region was essential for putative IRES activity [32].

After truncation, the putative IRES activity levels of nucleotide sequences 4952-5101 and 5027-5176 were much higher than that of nucleotide sequence 4877-5176 (Fig. 3b). The putative IRES activity increased appropriately after further truncation (Fig. 4a). Similarly, in mapping the putative IRES within nt 6419-nt 6718, putative IRES activity was found to be slightly increased when the nucleotide sequences were truncated stepwise (Fig. 3d and 4d). This phenomenon might have arisen from alleviation of steric hindrance on ribosome binding to the IRES or IRES trans-acting factors (ITAFs) after truncation. The putative IRES activity was almost equal in nucleotide sequences 5002-5101 and 5027-5126. In addition, the main segment of these two regions overlapped. Therefore, the putative IRES was located roughly within the 125-nucleotide region spanning nucleotides 5002-5126. Ribosome profiling showed that the ribosome positions in the genome “snapshot” were in the abovementioned putative IRES regions [41].

According to our previous study [16], to exclude the possibility of alternative splicing in the constructed vectors, copies of two cistrons were detected by RT-qPCR after transfection. The mRNA copy numbers of F-Luc and R-Luc were not significantly different, indicating that no major aberrant splicing isoforms were contained in the constructed vectors (S Fig. 3 and S Table 1).

The TG used as an inducer of ERS has been reported in previous studies[23,42]. The expression of GRP 78 was increased in BHK-21 cells treated with TG (S Fig 4 a). Stress could lead the upregulation of IRES-mediated translation[43,44]. The ratios of F-Luc/R-Luc in TG-treated BHK-21 cells increased significantly versus that of BHK-21 cells without treatment of TG (S Fig 4 b). The result supported the idea that ERS enhanced the IRES-mediated translation and further illustrated the presence of putative IRESs in the coding region of HRV 16 genome.

Our results indicated that the putative IRES can effectively initiate the expression of GFP *in vitro* (Fig. 6). The stable secondary structure was essential for putative IRES activity[45]. Obvious stem loops were found in the secondary structures of putative IRESs. The possibility of the shuffled sequence forming a stable secondary structure was much less than that of the putative IRES (supplementary materials 4). The potential secondary structure in the shuffled sequence might lead to insignificant cap-independent translation.

In general, IRES elements are located within the 5'-UTRs of cellular or viral mRNAs upstream of the AUG initiation codon and initiate the translation of full-length proteins [46]. However, several reports have shown that IRESs, within the coding regions of viral genomes, drive protein translation. The RNA genome

of HIV-1 contains two IRESs. The first one is located in the 5'-UTR and mediates viral replication during the G₂/M phase of the cell cycle [26]. The other one is located within the coding region of *gag* and initiates the expression of a novel Gag isoform through an AUG initiation codon in the *gag* coding region. This Gag isoform may participate in HIV-1 replication *in vitro* [47]. In addition, the HIV-2 virion contains three isoforms of Gag (57 kDa, 50 kDa and 44 kDa) and two N-terminally truncated shorter isoforms are translated by an IRES element located in the coding region of Gag. As integrated Gag polyproteins participate in viral capsid assembly and genome packaging, the two shorter Gag isoforms might be synthesized independently and perform other functions of Gag [40]. In the canonical mechanism, a single integrated polyprotein is synthesized through the IRES in the 5'-UTR of viral genomic RNA and is then cleaved by viral proteases 2A and 3C to generate structural and nonstructural proteins [21]. The polyprotein synthesis needed to meet a possible requirement for certain viral proteins during the infection cycle of HRV16 might cost time and waste energy.

Considering our results, we concluded that the translation of proteins with various molecular weights is mediated through different putative IRESs in the HRV16 genome. The translational efficiency of viral nonstructural proteins is inversely related to the order of the corresponding protein coding region in the genome. A segment of 3D (52.3 kDa) dependent on the putative IRES within nt 6619-nt 6718 or nt 6629-nt 6778 was translated in the highest amount, and whether this protein is conducive to viral replication needs further study. In addition, the translation of a protein containing a segment of 3B and the complete 3CD protein was initiated by the putative IRES located between nucleotides 5002 and 5126, and the translation of a protein containing a segment of the 2C protein was initiated by a putative IRES located between nucleotides 4286 and 4585. These nonstructural proteins efficiently promote virus replication.

Declarations

Author contributions

Jun Han and Bingtian Shi designed and conceived the experiments; Bingtian Shi performed the experiments, analyzed data and wrote the original draft; Qinqin Song, Xiaonuan Luo, Juan Song, Dong Xia, Zhiqiang Xia, Mi Liu, Wenjun Wang, Ruifang Wang and Haijun Du supervised data; Jun Han reviewed and edited the article.

Funding

This work was supported by China Mega-Project for Infectious Disease [2018ZX10102001, 2018ZX10711001, 2018ZX10734401 and 2018ZX10734404],

and SKLID Development Grant (2011SKLID104)

Conflict of interest

The authors declare that there are no conflicts of interest.

References

1. Merrick WC (2004) Cap-dependent and cap-independent translation in eukaryotic systems. *Gene* 332:1-11. doi:10.1016/j.gene.2004.02.051
2. Lindeberg J, Ebendal T (1999) Use of an internal ribosome entry site for bicistronic expression of Cre recombinase or rtTA transactivator. *Nucleic acids research* 27 (6):1552-1554. doi:10.1093/nar/27.6.1552
3. Ohlmann T, Jackson R (1999) The properties of chimeric picornavirus IRESes show that discrimination between internal translation initiation sites is influenced by the identity of the IRES and not just the context of the AUG codon. *RNA (New York, NY)* 5 (6):764-778. doi:10.1017/s1355838299982158
4. Pelletier J, Sonenberg N (1988) Internal initiation of translation of eukaryotic mRNA directed by a sequence derived from poliovirus RNA. *Nature* 334 (6180):320-325. doi:10.1038/334320a0
5. ROJAS-EISENRING IA, CAJERO-JUAREZ M, DEL ANGEL RM (1995) Cell proteins bind to a linear polypyrimidine-rich sequence within the 5'-untranslated region of rhinovirus 14 RNA. *69* (11):6819-6824
6. Yang D, Wilson J, Anderson D, Bohunek L, Cordeiro C, Kandolf R, McManus B (1997) In vitro mutational and inhibitory analysis of the cis-acting translational elements within the 5' untranslated region of coxsackievirus B3: potential targets for antiviral action of antisense oligomers. *Virology* 228 (1):63-73. doi:10.1006/viro.1996.8366
7. Sweeney TR, Abaeva IS, Pestova TV, Hellen CU (2014) The mechanism of translation initiation on Type 1 picornavirus IRESs. *The EMBO journal* 33 (1):76-92. doi:10.1002/embj.201386124
8. Martínez-Salas E, Francisco-Velilla R, Fernandez-Chamorro J, Lozano G, Diaz-Toledano R (2015) Picornavirus IRES elements: RNA structure and host protein interactions. *Virus research* 206:62-73. doi:10.1016/j.virusres.2015.01.012
9. Lee KM, Chen CJ, Shih SR (2017) Regulation Mechanisms of Viral IRES-Driven Translation. *Trends Microbiol* 25 (7):546-561. doi:10.1016/j.tim.2017.01.010
10. Khawaja A, Vopalensky V, Pospisek M (2015) Understanding the potential of hepatitis C virus internal ribosome entry site domains to modulate translation initiation via their structure and function. *Wiley Interdiscip Rev RNA* 6 (2):211-224. doi:10.1002/wrna.1268
11. Zhu Z, Wang Y, Yu J, Wan L, Chen J, Xiao M (2010) Classical swine fever virus NS3 is an IRES-binding protein and increases IRES-dependent translation. *Virus research* 153 (1):106-112. doi:10.1016/j.virusres.2010.07.013
12. Xiao M, Wang Y, Zhu Z, Yu J, Wan L, Chen J (2009) Influence of NS5A protein of classical swine fever virus (CSFV) on CSFV internal ribosome entry site-dependent translation. *J Gen Virol* 90 (Pt 12):2923-2928. doi:10.1099/vir.0.014472-0
13. Chard LS, Kaku Y, Jones B, Nayak A, Belsham GJ (2006) Functional analyses of RNA structures shared between the internal ribosome entry sites of hepatitis C virus and the picornavirus porcine

- teschovirus 1 Talfan. *J Virol* 80 (3):1271-1279. doi:10.1128/JVI.80.3.1271-1279.2006
14. Chard LS, Bordeleau ME, Pelletier J, Tanaka J, Belsham GJ (2006) Hepatitis C virus-related internal ribosome entry sites are found in multiple genera of the family Picornaviridae. *J Gen Virol* 87 (Pt 4):927-936. doi:10.1099/vir.0.81546-0
 15. Pfingsten J, Costantino D, Kieft J (2007) Conservation and diversity among the three-dimensional folds of the Dicistroviridae intergenic region IRESes. *Journal of molecular biology* 370 (5):856-869. doi:10.1016/j.jmb.2007.04.076
 16. Luo XN, Song QQ, Yu J, Song J, Wang XL, Xia D, Sun P, Han J (2017) Identification of the internal ribosome entry sites (IRES) of prion protein gene. *The international journal of biochemistry & cell biology* 93:46-51. doi:10.1016/j.biocel.2017.10.014
 17. Graber TE, Baird SD, Kao PN, Mathews MB, Holcik M (2010) NF45 functions as an IRES trans-acting factor that is required for translation of cIAP1 during the unfolded protein response. *Cell death and differentiation* 17 (4):719-729. doi:10.1038/cdd.2009.164
 18. Qin X, Sarnow P (2004) Preferential translation of internal ribosome entry site-containing mRNAs during the mitotic cycle in mammalian cells. *J Biol Chem* 279 (14):13721-13728. doi:10.1074/jbc.M312854200
 19. Song QQ, Luo XN, Chi MM, Song J, Xia D, Han J (2020) Identification of Internal Ribosomal Entry Site inside Open Reading Frame of 14-3-3beta Gene. *Biomedical and environmental sciences : BES* 33 (4):273-276. doi:10.3967/bes2020.037
 20. Song J, Chi M, Luo X, Song Q, Xia D, Shi B, Han J (2019) Non-Structural Protein 2B of Human Rhinovirus 16 Activates Both PERK and ATF6 Rather Than IRE1 to Trigger ER Stress. *Viruses* 11 (2):133. doi:10.3390/v11020133
 21. Lin JY, Chen TC, Weng KF, Chang SC, Chen LL, Shih SR (2009) Viral and host proteins involved in picornavirus life cycle. *Journal of biomedical science* 16:103. doi:10.1186/1423-0127-16-103
 22. Fu Y, Cai J, Xi M, He Y, Zhao Y, Zheng Y, Zhang Y, Xi J, He Y (2020) Neuroprotection Effect of Astragaloside IV from 2-DG-Induced Endoplasmic Reticulum Stress. *Oxidative medicine and cellular longevity* 2020:9782062. doi:10.1155/2020/9782062
 23. Fang DL, Wan Y, Shen W, Cao J, Sun ZX, Yu HH, Zhang Q, Cheng WH, Chen J, Ning B (2013) Endoplasmic reticulum stress leads to lipid accumulation through upregulation of SREBP-1c in normal hepatic and hepatoma cells. *Molecular and cellular biochemistry* 381 (1-2):127-137. doi:10.1007/s11010-013-1694-7
 24. Gruber AR, Lorenz R, Bernhart SH, Neubock R, Hofacker IL (2008) The Vienna RNA websuite. *Nucleic Acids Res* 36 (Web Server issue):W70-74. doi:10.1093/nar/gkn188
 25. Ricci EP, Herbreteau CH, Decimo D, Schaupp A, Datta SA, Rein A, Darlix JL, Ohlmann T (2008) In vitro expression of the HIV-2 genomic RNA is controlled by three distinct internal ribosome entry segments that are regulated by the HIV protease and the Gag polyprotein. *Rna* 14 (7):1443-1455. doi:10.1261/rna.813608

26. Brasey A, Lopez-Lastra M, Ohlmann T, Beerens N, Berkhout B, Darlix JL, Sonenberg N (2003) The leader of human immunodeficiency virus type 1 genomic RNA harbors an internal ribosome entry segment that is active during the G2/M phase of the cell cycle. *J Virol* 77 (7):3939-3949. doi:10.1128/jvi.77.7.3939-3949.2003
27. Jackson RJ, Hellen CU, Pestova TV (2010) The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat Rev Mol Cell Biol* 11 (2):113-127. doi:10.1038/nrm2838
28. Sonenberg N, Hinnebusch AG (2009) Regulation of translation initiation in eukaryotes: mechanisms and biological targets. *Cell* 136 (4):731-745. doi:10.1016/j.cell.2009.01.042
29. Jang S, Kräusslich H, Nicklin M, Duke G, Palmenberg A, Wimmer E (1988) A segment of the 5' nontranslated region of encephalomyocarditis virus RNA directs internal entry of ribosomes during in vitro translation. *Journal of virology* 62 (8):2636-2643. doi:10.1128/jvi.62.8.2636-2643.1988
30. Hanson PJ, Zhang HM, Hemida MG, Ye X, Qiu Y, Yang D (2012) IRES-Dependent Translational Control during Virus-Induced Endoplasmic Reticulum Stress and Apoptosis. *Front Microbiol* 3:92. doi:10.3389/fmicb.2012.00092
31. Komar AA, Hatzoglou M (2014) Cellular IRES-mediated translation. *Cell Cycle* 10 (2):229-240. doi:10.4161/cc.10.2.14472
32. Weingarten-Gabbay S, Elias-Kirma S, Nir R, Gritsenko AA, Stern-Ginossar N, Yakhini Z, Weinberger A, Segal E (2016) Comparative genetics. Systematic discovery of cap-independent translation sequences in human and viral genomes. *Science* 351 (6270). doi:10.1126/science.aad4939
33. Sharathchandra A, Katoch A, Das S (2014) IRES mediated translational regulation of p53 isoforms. *Wiley Interdiscip Rev RNA* 5 (1):131-139. doi:10.1002/wrna.1202
34. Ray PS, Grover R, Das S (2006) Two internal ribosome entry sites mediate the translation of p53 isoforms. *EMBO reports* 7 (4):404-410. doi:10.1038/sj.embor.7400623
35. Rubtsova MP, Sizova DV, Dmitriev SE, Ivanov DS, Prassolov VS, Shatsky IN (2003) Distinctive properties of the 5'-untranslated region of human hsp70 mRNA. *J Biol Chem* 278 (25):22350-22356. doi:10.1074/jbc.M303213200
36. Jendrach M, Thiel V, Siddell S (1999) Characterization of an internal ribosome entry site within mRNA 5 of murine hepatitis virus. *Archives of virology* 144 (5):921-933. doi:10.1007/s007050050556
37. Wachalska M, Graul M, Praest P, Luteijn RD, Babnis AW, Wiertz E, Bienkowska-Szewczyk K, Lipinska AD (2019) Fluorescent TAP as a Platform for Virus-Induced Degradation of the Antigenic Peptide Transporter. *Cells* 8 (12):1590. doi:10.3390/cells8121590
38. Wein N, Vulin A, Falzarano MS, Szigyarto CA, Maiti B, Findlay A, Heller KN, Uhlen M, Bakthavachalu B, Messina S, Vita G, Passarelli C, Brioschi S, Bovolenta M, Neri M, Gualandi F, Wilton SD, Rodino-Klapac LR, Yang L, Dunn DM, Schoenberg DR, Weiss RB, Howard MT, Ferlini A, Flanigan KM (2014) Translation from a DMD exon 5 IRES results in a functional dystrophin isoform that attenuates dystrophinopathy in humans and mice. *Nat Med* 20 (9):992-1000. doi:10.1038/nm.3628
39. Abaeva IS, Pestova TV, Hellen CUT (2016) Attachment of ribosomal complexes and retrograde scanning during initiation on the Halastavi árva virus IRES. *Nucleic Acids Research* 44 (5):2362-

2377. doi:10.1093/nar/gkw016

40. Herbreteau CH, Weill L, Decimo D, Prevot D, Darlix JL, Sargueil B, Ohlmann T (2005) HIV-2 genomic RNA contains a novel type of IRES located downstream of its initiation codon. *Nature structural & molecular biology* 12 (11):1001-1007. doi:10.1038/nsmb1011
41. Szavits-Nossan J, Ciandrini L (2020) Inferring efficiency of translation initiation and elongation from ribosome profiling. *Nucleic Acids Res* 48 (17):9478-9490. doi:10.1093/nar/gkaa678
42. Sheng W, Wang G, Tang J, Shi X, Cao R, Sun J, Lin YH, Jia C, Chen C, Zhou J, Dong M (2020) Calreticulin promotes EMT in pancreatic cancer via mediating Ca(2+) dependent acute and chronic endoplasmic reticulum stress. *Journal of experimental & clinical cancer research : CR* 39 (1):209. doi:10.1186/s13046-020-01702-y
43. Sharathchandra A, Lal R, Khan D, Das S (2012) Annexin A2 and PSF proteins interact with p53 IRES and regulate translation of p53 mRNA. *RNA Biol* 9 (12):1429-1439. doi:10.4161/rna.22707
44. Shi Y, Yang Y, Hoang B, Bardeleben C, Holmes B, Gera J, Lichtenstein A (2016) Therapeutic potential of targeting IRES-dependent c-myc translation in multiple myeloma cells during ER stress. *Oncogene* 35 (8):1015-1024. doi:10.1038/onc.2015.156
45. Wang G, Guo X, Silveyra P, Kimball SR, Floros J (2009) Cap-independent translation of human SP-A 5'-UTR variants: a double-loop structure and cis-element contribution. *American journal of physiology Lung cellular and molecular physiology* 296 (4):L635-647. doi:10.1152/ajplung.90508.2008
46. Stoneley. M, Willis. AE (2004) Cellular internal ribosome entry segments: structures, trans-acting factors and regulation of gene expression. *Oncogene* 23 (18):3200-3207. doi:10.1038/sj.onc.1207551
47. Buck CB, Shen X, Egan MA, Pierson TC, Walker CM, Siliciano RF (2001) The human immunodeficiency virus type 1 gag gene encodes an internal ribosome entry site. *J Virol* 75 (1):181-191. doi:10.1128/JVI.75.1.181-191.2001

Tables

Table 1 The positions of successive nucleotide sequences with more than 150 reads basically covered those of the 5 validated IRESs identified above

Location of IRESs in the HRV16 genome	Reads peaks at 21 h.p.i.	Reads peaks at 24 h.p.i.
nt 4286-nt 4585	nt 4362-nt 4412	nt 4362t-nt 4490
	nt 4417-nt 4447	nt 4508-nt 4591
	nt 4457-nt 4490	
	nt 4549-nt 4581	
nt 5002-nt 5126	nt 4969-nt 5022	nt 4907-nt 5022
		nt 5079-nt 5098
nt 6245-nt 6394	nt 6224-nt 6256	nt 6181-nt 6259
	nt 6300-nt 6352	nt 6300-nt 6353
nt 6619-nt 6718	nt 6677-nt 6711	nt 6672-nt 6718
nt 6629-nt 6778	nt 6750-nt 6776	nt 6672-nt 6718
		nt 6747-nt 6834

Figures

Figure 1

Screening of putative IRESs in nonstructural protein coding regions of the HRV16 genome. (a) Schematic diagram of the positions of an ORF (nt 2969-nt 4861) and an ORF (nt 4586-nt 7084) in the intact P2 gene (nt 3185-nt 4861) and P3 gene (nt 4862-nt 7084), respectively, in the HRV16 genome. (b) Screening of IRESs in the P2 coding region. Schematic diagram of inserted sequences fused with the GFP gene sequence. The positions of the inserted sequences in the HRV16 genome are shown in a schematic

diagram (left panel). After pP2(2969-4861), pP2(3632-4861), pP2(3926-4861), pP2(4256-4861) and pP2(4586-4861) were transfected individually into BHK-21 cells, fusion proteins were detected by Western blotting with an anti-GFP antibody. Potential IRES-dependent proteins are indicated by the red arrows (right panel). (c) Fusion protein expression from ORFs located in the P3 coding region. Schematic diagram of inserted sequences fused with the GFP gene sequence. The positions of the inserted ORFs in the HRV16 genome are shown in a schematic diagram (left panel). After pP3(4586-7084), pP3(4874-7084), pP3(5177-7084), pP3(5672-7084), pP3(5993-7084), pP3(6164-7084), pP3(6395-7084) and pP3(6596-7084) were transfected individually into BHK-21 cells, fusion proteins were detected by Western blotting with an anti-GFP antibody. Potential IRES-dependent proteins are indicated by the red arrows and might be IRES-dependent (right panel)

Figure 2

Identification of putative IRESs in nonstructural proteins coding regions. A hairpin structure was inserted between two luciferase cistrons as a barrier of ribosome sliding. (a) Schematic diagram of the dual-luciferase reporter vectors p-IRES.CHECK, pP2-IRES-(3956-4255), pP2-IRES-(4181-4480), pP2-IRES-(4286-4585) and p-P-IRES.CHECK used for IRES identification in the P2 coding region (left panel). The characters (N, 1, 2, 3 and P) on the Y-axis represent these respective vectors (right panel). (b) Schematic

diagram of the dual-luciferase reporter vectors p-IRES.CHECK, pP3-IRES-(4877-5176), pP3-IRES-(5372-5671), pP3-IRES-(5693-5992), pP3-IRES-(6095-6394), pP3-IRES-(6419-6718), pP3-IRES-(6629-6928) and pP-IRES.CHECK used for IRES identification in the P3 coding region (left panel). The characters (N, 1, 2, 3, 4, 5, 6 and P) on the Y-axis represent these respective vectors (right panel). BHK-21 cells were transfected with the plasmids described above. The activity of F-Luc and R-Luc was measured 24 h post transfection. The F-Luc/R-Luc ratio relative to that of the negative control (F-Luc/R-Luc/N) was calculated. Experiments were repeated independently three times. *P, positive. N, negative

Figure 3

Determination of putative IRES boundaries by stepwise 150-bp depletion. (a) Mapping the putative IRES within nt 4286-nt 4585. Schematic diagrams of truncated isoforms (150 bp) with successive overlap of 75 bp (left panel). Schematic diagrams of the dual-luciferase reporter vectors p-IRES.CHECK, pP2-IRES-(4286-4435), pP2-IRES-(4360-4509), pP2-IRES-(4436-4585), pP2-IRES-(4286-4585) and p-P-IRES.CHECK (middle panel), which are represented by N, 1, 2, 3, 4 and P, respectively, on the Y-axis (right panel). (b) Mapping the putative IRES within nt 4877-nt 5176. Schematic diagrams of truncated isoforms (150 bp) with successive overlap of 75 bp (left panel). Schematic diagrams of the dual-luciferase reporter vectors p-IRES.CHECK, pP3-IRES-(4877-5026), pP3-IRES-(4952-5101), pP3-IRES-(5027-5176), pP3-IRES-(4877-5176) and p-P-IRES.CHECK (middle panel), which are represented by N, 1, 2, 3, 4 and P, respectively, on the Y-axis (right panel). (c) Mapping the putative IRES within nt 6095-nt 6394. Schematic diagrams of truncated isoforms (150 bp) with successive overlap of 75 bp (left panel). Schematic diagrams of the dual-luciferase reporter vectors p-IRES.CHECK, pP3-IRES-(6095-6244), pP3-IRES-(6169-6318), pP3-IRES-(6245-6394), pP3-IRES-(6095-6394) and p-P-IRES.CHECK (middle panel), which are represented by N, 1, 2, 3, 4 and P, respectively, on the Y-axis (right panel). (d) Mapping the putative IRES within nt 6419-nt 6718. Schematic diagrams of truncated isoforms (150 bp) with successive overlap of 75 bp (left panel). Schematic diagrams of the dual-luciferase reporter vectors p-IRES.CHECK, pP3-IRES-(6419-6568), pP3-IRES-(6494-6643), pP3-IRES-(6569-6718), pP3-IRES-(6419-6718) and p-P-IRES.CHECK (middle panel), which are represented by N, 1, 2, 3, 4 and P, respectively, on the Y-axis (right panel). (e) Mapping the putative IRES within nt 6629-nt 6928. Schematic diagrams of truncated isoforms (150 bp) with successive overlap of 75 bp (left panel). Schematic diagrams of the dual-luciferase reporter vectors p-IRES.CHECK, pP3-IRES-(6629-6778), pP3-IRES-(6704-6853), pP3-IRES-(6779-6928), pP3-IRES-(6629-6928) and p-P-IRES.CHECK (middle panel), which are represented by N, 1, 2, 3, 4 and P, respectively, on the Y-axis (right panel). BHK-21 cells were transfected with the plasmids described above. The activity of F-Luc and R-Luc was measured 24 h post transfection. The F-Luc/R-Luc ratio relative to that in the negative control cells (F-Luc/R-Luc/N) was calculated. Experiments were repeated independently three times. *P, positive. N, negative

Figure 4

Further determination of putative IRES boundaries in truncated isoforms (150 bp). (a) Mapping putative IRESs within the nt 4952-nt 5101 and nt 5027-nt 5176. Schematic diagrams of truncated isoforms (100 bp) with successive overlap of 75 bp (left panel). Schematic diagrams of the dual-luciferase reporter vectors p-IRES.CHECK, p3-IRES-(4952-5051), p3-IRES-(4977-5076), p3-IRES-(5002-5101), p3-IRES-(4952-5101), p3-IRES-(5027-5126), p3-IRES-(5052-5151), p3-IRES-(5077-5176), p3-IRES-(5027-5176),

pP3-IRES-(4877-5176) and p-P-IRES.CHECK (middle panel), which are represented by N, 1, 2, 3, 4, 5, 6, 7, 8, 9 and P, respectively, on the Y-axis (right panel). (b) Verification of putative IRES activity in nucleotide sequence 5002-5126. Schematic diagrams of the relative positions of nucleotide sequence 5002-5101, 5027-5126 and 5002-5126 (left panel). Schematic diagrams of the dual-luciferase reporter vectors p-IRES.CHECK, pP3-IRES-(5002-5101), pP3-IRES-(5027-5126), pP3-IRES-(5002-5126) and p-P-IRES.CHECK (middle panel), which are represented by N, 1, 2, 3 and P, respectively, on the Y-axis (right panel). (c) Mapping the putative IRES within nt 6245-nt 6394. Schematic diagrams of the truncated isoforms (100 bp) with successive overlap of 75 bp (left panel). Schematic diagrams of dual-luciferase reporter vectors p-IRES.CHECK, pP3-IRES-(6245-6344), pP3-IRES-(6270-6369), pP3-IRES-(6295-6394), pP3-IRES-(6245-6394), pP3-IRES-(6095-6394) and p-P-IRES.CHECK (middle panel), which are represented by N, 1, 2, 3, 4, 5 and P, respectively, on the in Y-axis (right panel). (d) Mapping the putative IRES within nt 6569-nt 6718. Schematic diagrams of truncated isoforms (100 bp) with successive overlap of 75 bp (left panel). Schematic diagrams of the dual-luciferase reporter vectors p-IRES.CHECK, pP3-IRES-(6569-6668), pP3-IRES-(6594-6693), pP3-IRES-(6619-6718), pP3-IRES-(6569-6718), pP3-IRES-(6419-6718) and p-P-IRES.CHECK (middle panel), which are represented by N, 1, 2, 3, 4, 5 and P, respectively, on the in Y-axis (right panel). (e) Mapping the putative IRES within nt 6629-nt 6928. Schematic diagrams of truncated isoforms (100 bp) with successive overlap of 75 bp (left panel). Schematic diagrams of the dual-luciferase reporter vectors p-IRES.CHECK, pP3-IRES-(6629-6728), pP3-IRES-(6654-6753), pP3-IRES-(6679-6778), pP3-IRES-(6629-6778), pP3-IRES-(6629-6928) and p-P-IRES.CHECK (middle panel), which are represented by N, 1, 2, 3, 4, 5 and P, respectively, on the Y-axis (right panel). BHK-21 cells were transfected with the plasmids described above. The activity of F-Luc and R-Luc was measured 24 h post transfection. The F-Luc/R-Luc ratio relative to that in the negative control cells (F-Luc/R-Luc/N) was calculated. Experiments were repeated independently three times. *P, positive. N, negative

Figure 5

Schematic diagrams of the distribution of the 5 putative IRESs in the HRV16 genome. The putative IRES spanning nt 4286-4585 was located in the 2C region. The IRES spanning nt 5002-5126 crossed the 3A and 3B regions. The putative IRESs spanning nt 6245-6394, 6619-6718 and 6629-6778, were located in the 3D region

Figure 6

Confirmation of the function of the putative IRESs in initiating protein translation. (a) Schematic diagrams of a series of vectors containing nucleotide sequences of the IRESs (top) or random scrambled sequences of the corresponding IRESs (bottom). On the basis of p-IRES.CHECK, the F-Luc gene was replaced with the GFP gene. (b-d) Western blotting for β-actin and GFP proteins whose translation was initiated by putative IRESs in nucleotide sequences 4286-4585, 5002-5126, 6619-6718, 6629-6778 and

6245-6394. The corresponding scrambled IRES sequences are shown as 4286 nt-4585 nt (s), 5002 nt-5126 nt (s), 6619 nt-6718 nt (s), 6629 nt-6778 nt (s), 6245 nt-6394 nt (s) and EMCV IRES (s). The EMCV IRES was used as the positive control

Figure 7

Mapping the ribosome density and distribution on HRV 16 mRNA by ribosome profiling. H1-HeLa cells were infected with HRV16 (m.o.i. =0.01) and treated with cycloheximide 21, 24 h.p.i. The Y-axis shows the density (reads) of ribosomes binding to HRV16 mRNA. The X-axis shows the relative positions in the HRV16 genome. H1-HeLa cells not infected with HRV16 were used as negative control cells

Supplementary Files

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

- [SFiguresandSTable1.doc](#)
- [supplementarymaterials.rar](#)