In format provided by Leppek, K. *et al.* (doi:10.1038/nrm.2017.103)

Supplementary information S2 | **Viral IRESs and systematic IRES discovery**

Viral IRESs are ancient mRNA structures evolved to hijack cellular translation

Viral IRES elements are RNA motifs that can directly recruit the 40S small ribosomal subunit close to the start codon (reviewed in refs.[1–4]). This provides sustained translation of the viral genome by host ribosomes as eukaryotic cells inhibit cap-mediated translation initiation upon viral infection. Since the first discovery of IRES elements in poliovirus and encephalomyocarditis virus (EMCV) in 1988[5,6], many viral IRESs have been well characterized. Thanks to comparative analysis, there is a growing appreciation of the diversity of viral IRES structures that recruit the translation machinery to the 5' UTR. Indeed, viral IRESs vary greatly in their requirements for eIFs. Based on these requirements, viral IRESs have been classified into four types that encompass their co-factor needs, RNA structure, location of the initiation codon relative to the IRES, and their activity in an *in vitro* translation system (reviewed in ref.[7]) (see the figure, part **a**). Only the highly structured and compact Type IV IRESs, found in ssRNA+ Dicistroviridae such as cricket paralysis virus (CrPV), *Plautia stali* intestine virus (PSIV), or taura syndrome virus (TSV), do not require any host initiation factors or initiator methionyl $tRNA_i$, but can directly recruit the 40S[8] (see the figure, part **b**). One extreme example of this recruitment is the molecular mimicry of a tRNA-like structure by the IRESs of the Dicistroviridae, including CrPV and TSV. The pseudoknot I (PKI) of the intergenic region IRES of CrPV domain 3 mimics the tRNA anticodon stem-loop bound to a cognate mRNA codon[7,9]. PKI binds the 40S which is stabilized by elongation factor eEF2, occupies the aminoacyl (A)-site in the decoding center in the 80S very similar to tRNAs, and is translocated in the 80S ribosome[10–13]. Thereby, the IRES RNA undergoes a structural change to a more stretched conformation[14]. While this structural mimicry appears to be limited to the small region of PKI in the decoding center as PKI globally does not look like a tRNA, the overall IRES structure interacts with the tRNA binding sites on both subunits. Surprisingly, this impressive structural feature of the CrPV IRES RNA is not only able to recruit and initiate translation by eukaryotic ribosomes, but is even able to promote translation in bacteria, albeit less efficiently[15]. In its crystal structure, the CrPV IRES occupies the tRNA space on both eukaryotic and prokaryotic ribosomes. This elegantly illustrates how IRESs in RNAs function as structure-driven signals for translation initiation spanning billions of years of evolution. Not only IRESs located in the 5' UTR can mimic a tRNA shape. The turnip crinkle virus (TCV) IRES has a tRNA-like IRES structure, as suggested by in-line probing[16], and resides in the 3' UTR[16]. This IRES acts as a translational enhancer by binding the 60S from this locus which stimulates initiation upon mRNA circularization[17].

Besides tRNA mimicry, base pairing of rRNA with viral IRES elements at the ribosome binding site is another mechanism to directly recruit the translation machinery, akin to the Shine-Dalgarno (SD) ribosomal binding sequence in bacterial and archaeal mRNAs. For example, the Type III hepatitis C virus (HCV) IRES directly base pairs with 3 nt in both helix 26 and expansion segment (ES) 7 of 18S rRNA[18,19], as shown by hydroxyl radical structure probing and structural mutagenesis. The observation that ES7 is a central anchor point for the highly structured HCV IRES on the human 80S ribosome was confirmed by cryo-electron microscopy (cryo-EM) studies[20] (see the figure, part **a**). This direct interaction induces conformational changes in the 40S subunit which positions it accurately at the initiation codon as shown by footprinting[21,22]. Moreover, variations in the RNA structure of HCV-like IRESs itself can dictate their behavior and may regulate their activity upon stress[23]. Importantly, initiation by the HCV IRES starts with 40S binding, but also requires binding to the eIF2 ternary complex and eIF3[21,22] for start codon recognition. In comparison, EMCV IRES requires eIF4A, eIF4B, and eIF4G as co-factors. Together, viral

IRES elements further highlight the notion that the more compact and structured a viral IRES is, the lower its requirement for assisting RBPs and canonical translation initiation factors[8]. Yet, despite the wealth of knowledge about the defined secondary and tertiary RNA structures of viral IRESs, their interactions with the ribosome are less well understood (reviewed in ref.[24]). After decades of work, we only have cryo-EM structures for CrPV[10–13] and HCV[20,25–27] IRESs on the ribosome, which show completely different modes of interaction (see the figure). In addition, besides in-depth analysis of viral IRES RNA structures, there are few structural or sequence motifs that are in common between the different classes of IRESs. This variability limits bioinformatics prediction of IRESs[28–30], especially in the context of cellular genomes, which also harbor IRES elements.

Alternative modes of internal, structure-dependent translation initiation

Alternative internal initiation can also occur in a cap-dependent but scanning-free manner, via ribosome "shunting", or cap-independent translation enhancer (CITE) elements. In ribosome shunting, the 40S is bound at the cap followed by discontinuous scanning, in which large 5' UTR regions are bypassed to reach the initiation codon. Ribosome shunting has been reported for selective translation of several viral RNA genomes such as the adenovirus late mRNA 5' UTR[31,32], in which complex RNA structures of the tripartite leader facilitate shunting by complementarity to the 18S rRNA. Shunting in cellular mRNAs has been proposed for the human heat shock protein 70 (Hsp70) 5' UTR[32] and the structured 5' UTR of beta-secretase 1 (BACE1) mRNA, an enzyme important in Alzheimer's disease[33]. This mechanism is predominant when eIF4F is limited, for example upon infection or during heat stress, but it is not well understood and no common RNA motifs or structures have been associated with it so far. In contrast to shunting, CITEs reside mostly in the 3' UTR (reviewed in ref.[34]). They exclusively recruit eIFs in a cap-independent fashion and deliver them to the 5' end by long-distance base pairing with the 5' UTR, bridging the 3' and 5' UTRs[35–37]. 3'CITEs are found in uncapped plant RNA viruses and in some cellular mRNAs such as human apoptotic protease activating factor 1 (Apaf-1) acting in apoptosis when eIF4E is limited[38]. CITEs require scanning from the uncapped 5' end for initiation. However, many 3'CITEs function equally well when placed in the 5' UTR. The histone H4 mRNA employs a different method for initiation[39]. This mRNA has a very short 5' UTR and two structural elements in its coding region, one of which internally binds eIF4E cap-independently while the other sequesters the cap and thereby positions the ribosome onto the start codon. In fact, cryo-EM data suggest that the histone H4 mRNA directly base pairs with the 18S rRNA to tether the 40S ribosomal subunit[40].

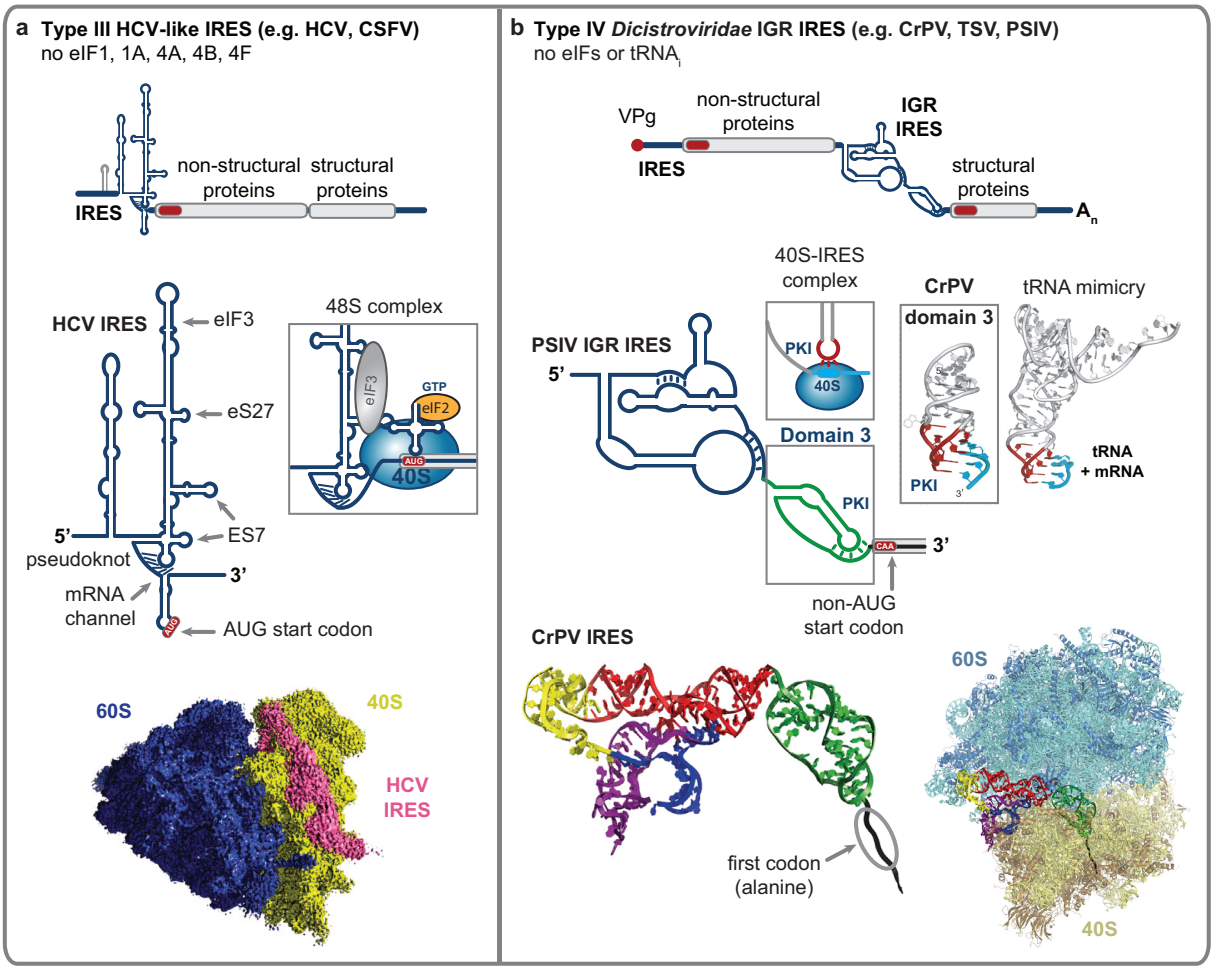
Systematic discovery of physiologically-relevant cellular IRESs

In contrast to stress-linked cellular IRESs, certain IRESs are active during and required for normal embryo development and are efficient in mediating cap-independent translation initiation in a physiological and tissue-specific manner. This finding provides the rationale for systematic approaches aimed at increasing the list of functional IRESs under physiological conditions. Indeed, recent unbiased methods systematically screened the transcriptome for novel IRES activity *in vitro*[41] and in mammalian cells[42]. The *in vivo* screen measured IRES activity of a synthetic library of 174 nt long oligonucleotides in a bicistronic reporter assay, which places an RNA fragment between two reporter ORFs. While translation of the upstream ORF is cap-dependent, internal translation of the second reporter ORF is stimulated by the IRES in the intergenic region. This analysis led to the identification of putative cap-independent translation elements in hundreds of viral and human 5' UTRs, as well as unexpected novel IRESs in 3' UTRs[42]. Many of them appear to recruit the 40S ribosome by direct 18S rRNA base pairing. However,

In format provided by Leppek, K. *et al.* (doi:10.1038/nrm.2017.103)

perhaps because the relatively short fragment length of the library was not compatible with extensive RNA structures, the predominant novel IRESs reported from the screen were short linear poly(U) motifs or small highly structured elements that did not share any obvious common motifs. The importance of their local RNA structure, assisting co-factors or activity in their natural mRNA context remains to be tested. In fact, empirical and systematic discovery of potential novel IRES elements particularly demand meticulous independent validation of their activity in a set of carefully controlled experiments, as reviewed elsewhere[43,44]. Experimental limitations and caveats have indeed largely restricted the discovery and confirmation of new functional IRESs[45–48]. Despite these caveats, the established examples of cellular IRESs represent a means to fine-tune gene expression by internal ribosome recruitment.

Figure | **Viral IRESs are classic examples of RNA structures that mediate cap-independent translation initiation**



IRES elements in viral genomes are categorized into four types according to their features such as requirement for stimulating co-factors such as eIFs. Structurally related viral IRESs use distinct mechanisms to mediate internal ribosomal recruitment to an mRNA. **a** The type III hepatitis C virus (HCV)-like IRESs (~300 nt in length) includes HCV and classical swine fever virus (CSFV) and require a minimal set of eIF3 and the ternary complex (GTP-eIF2-Met-tRNA$_i$) as co-factors to bind and position the 40S ribosome close to the start codon. This step forms the 48S complex. The cryo-EM structure of the

HCV IRES-ribosome complex confirms that the IRES only contacts the 40S subunit, including interactions of rRNA and ribosomal proteins with the IRES as indicated with arrows on the RNA structure model. **b** The type IV IRESs (~200 nt in length) of the Dicistroviridae genomes include the intergenic region (IGR) IRESs of the cricket paralysis virus (CrPV), taura syndrome virus (TSV) and *Plautia stali* intestine virus (PSIV). A genome-linked viral protein (VPg, red circle) is covalently linked to the 5' end instead of the cap which is incompatible with cap-dependent initiation. Non-structural proteins are translated by a 5' IRES whereas structural proteins are translated by the IGR IRES. The PSIV IRES RNA structure includes three pseudoknots (PK). These IRESs require no co-factors or initiator Met-tRNA$_i$ for ribosome recruitment but rather mimic the shape of a tRNA for direct 40S binding and translocation inside the decoding center of the 80S ribosome. This is exemplified for the domain 3 (pseudoknot 1, PKI) of the CrPV IGR IRES that mimics the tRNA-mRNA interaction (red, blue regions) and docks into the P-site of the 40S ribosome (ribbon representation from crystal structures reproduced with permission from ref.[7]). The cryo-EM structure of the CrPV IRES RNA (ribbon representation with PKI in green) and bound to the 80S ribosome was reproduced with permission from ref.[11]. Type I and II viral IRES (not shown) require more co-factors and are functionally related to cap-dependent initiation.
Part **a** is from REF. 20, Macmillan Publishers Limited. Part **b** is reproduced with permission from REF. 7, Elsevier; and REF. 11, Elsevier.

References:

1. Jackson, R. J. Alternative mechanisms of initiating translation of mammalian mRNAs. *Biochem. Soc. Trans.* **33,** 1231–1241 (2005).
2. Sonenberg, N. & Hinnebusch, A. G. Regulation of translation initiation in eukaryotes: mechanisms and biological targets. *Cell* **136,** 731–745 (2009).
3. Jackson, R. J., Hellen, C. U. T. & Pestova, T. V. The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat. Rev. Mol. Cell Biol.* **11,** 113–127 (2010).
4. Plank, T. D. M. & Kieft, J. S. The structures of nonprotein-coding RNAs that drive internal ribosome entry site function. *Wiley Interdiscip. Rev. RNA* **3,** 195–212 (2012).
5. Pelletier, J. & Sonenberg, N. Internal initiation of translation of eukaryotic mRNA directed by a sequence derived from poliovirus RNA. *Nature* **334,** 320–325 (1988).
6. Jang, S. K. *et al.* A segment of the 5' nontranslated region of encephalomyocarditis virus RNA directs internal entry of ribosomes during in vitro translation. *J. Virol.* **62,** 2636–2643 (1988).
7. Kieft, J. S. Viral IRES RNA structures and ribosome interactions. *Trends Biochem. Sci.* **33,** 274–283 (2008).
8. Filbin, M. E. & Kieft, J. S. Toward a structural understanding of IRES RNA function. *Curr. Opin. Struct. Biol.* **19,** 267–276 (2009).
9. Costantino, D. a, Pfingsten, J. S., Rambo, R. P. & Kieft, J. S. tRNA-mRNA mimicry drives translation initiation from a viral IRES. *Nat. Struct. Mol. Biol.* **15,** 57–64 (2008).
10. Schuler, M. *et al.* Structure of the ribosome-bound cricket paralysis virus IRES RNA. *Nat Struct Mol Biol* **13,** 1092–1096 (2006).
11. Fernández, I. S., Bai, X. C., Murshudov, G., Scheres, S. H. W. & Ramakrishnan, V. Initiation of translation by cricket paralysis virus IRES requires its translocation in the ribosome. *Cell* **157,** 823–831 (2014).
12. Koh, C. S., Brilot, A. F., Grigorieff, N. & Korostelev, A. A. Taura syndrome virus IRES initiates translation by binding its tRNA-mRNA-like structural element in the ribosomal decoding center. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 9139–44 (2014).
13. Murray, J. *et al.* Structural characterization of ribosome recruitment and translocation by type IV IRES. *Elife* **5,** 1–24 (2016).
14. Muhs, M. *et al.* Cryo-EM of ribosomal 80s complexes with termination factors reveals the translocated cricket paralysis virus IRES. *Mol. Cell* **57,** 422–433 (2015).
15. Colussi, T. M. *et al.* Initiation of translation in bacteria by a structured eukaryotic IRES RNA. *Nature* **519,** 110–113 (2015).
16. McCormack, J. C. *et al.* Structural domains within the 3' untranslated region of Turnip crinkle virus. *J Virol* **82,** 8706–8720 (2008).
17. Stupina, V. A. *et al.* The 3' proximal translational enhancer of Turnip Crinkle Virus binds to 60S ribosomal subunits. *RNA* **14,** 2379–2393 (2008).
18. Malygin, A. A., Kossinova, O. A., Shatsky, I. N. & Karpova, G. G. HCV IRES interacts with the 18S rRNA to activate

the 40S ribosome for subsequent steps of translation initiation. *Nucleic Acids Res*. **41,** 8706–8714 (2013).
19. Matsuda, D. & Mauro, V. P. Base pairing between hepatitis C virus RNA and 18S rRNA is required for IRES-dependent translation initiation in vivo. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 15385–15389 (2014).
20. Quade, N., Boehringer, D., Leibundgut, M., van den Heuvel, J. & Ban, N. Cryo-EM structure of Hepatitis C virus IRES bound to the human ribosome at 3.9-Å resolution. *Nat. Commun.* **6,** 7646, 1–9 (2015).
21. Pestova, T. V, Shatsky, I. N., Fletcher, S. P., Jackson, R. J. & Hellen, C. U. A prokaryotic-like mode of cytoplasmatic eukaryotic ribosome binding to the initiation codon during internal translation initiation of hepatitis C and classical swine fever virus RNAs. *Genes Dev*. **12,** 67–83 (1998).
22. Kieft, J. S., Zhou, K., Jubin, R. & Doudna, J. A. Mechanism of ribosome recruitment by hepatitis C IRES RNA. *RNA* **7,** 194–206 (2001).
23. Jaafar, Z. A., Oguro, A., Nakamura, Y. & Kieft, J. S. Translation initiation by the hepatitis C virus IRES requires eIF1A and ribosomal complex remodeling. *Elife* **5,** 1–23 (2016).
24. Balvay, L., Rifo, R. S., Ricci, E. P., Decimo, D. & Ohlmann, T. Structural and functional diversity of viral IRESes. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1789,** 542–557 (2009).
25. Spahn, C. M. *et al*. Hepatitis C virus IRES RNA-induced changes in the conformation of the 40s ribosomal subunit. *Science* **291,** 1959–1962 (2001).
26. Boehringer, D., Thermann, R., Ostareck-Lederer, A., Lewis, J. D. & Stark, H. Structure of the hepatitis C virus IRES bound to the human 80S ribosome: Remodeling of the HCV IRES. *Structure* **13,** 1695–1706 (2005).
27. Hashem, Y. *et al*. Hepatitis-C-virus-like internal ribosome entry sites displace eIF3 to gain access to the 40S subunit. *Nature* **503,** 539–543 (2013).
28. Baird, S. D., Turcotte, M., Korneluk, R. G. & Holcik, M. Searching for IRES. *RNA* **12,** 1755–1785 (2006).
29. Baird, S. D., Lewis, S. M., Turcotte, M. & Holcik, M. A search for structurally similar cellular internal ribosome entry sites. *Nucleic Acids Res*. **35,** 4664–4677 (2007).
30. Xia, X. & Holcik, M. Strong eukaryotic IRESs have weak secondary structure. *PLoS One* **4,** 4–6 (2009).
31. Yueh, A. & Schneider, R. J. Selective translation initiation by ribosome jumping in adenovirus- infected and heat-shocked cells. *Genes Dev*. **10,** 1557–1567 (1996).
32. Yueh, A. & Schneider, R. J. Translation by ribosome shunting on adenovirus and hsp70 mRNAs facilitated by complementarity to 18S rRNA. *Genes Dev*. **14,** 414–421 (2000).
33. Koh, D. C., Edelman, G. M. & Mauro, V. P. Physical evidence supporting a ribosomal shunting mechanism of translation initiation for BACE1 mRNA. *Transl. (Austin, Tex.)* **1,** e24400, 1–8 (2013).
34. Shatsky, I. N., Dmitriev, S. E., Andreev, D. E. & Terenin, I. M. Transcriptome-wide studies uncover the diversity of modes of mRNA recruitment to eukaryotic ribosomes. *Crit. Rev. Biochem. Mol. Biol*. **49,** 164–177 (2014).
35. Miller, W. A., Wang, Z. & Treder, K. The amazing diversity of cap-independent translation elements in the 3′-untranslated regions of plant viral RNAs. *Biochem. Soc. Trans*. **35,** 1629–1633 (2007).
36. Mokrejs, M. *et al*. IRESite--a tool for the examination of viral and cellular internal ribosome entry sites. *Nucleic Acids Res*. **38,** D131-6 (2010).
37. Terenin, I. M., Andreev, D. E., Dmitriev, S. E. & Shatsky, I. N. A novel mechanism of eukaryotic translation initiation that is neither m7G-cap-, nor IRES-dependent. *Nucleic Acids Res*. **41,** 1807–1816 (2013).
38. Andreev, D. E., Dmitriev, S. E., Zinovkin, R., Terenin, I. M. & Shatsky, I. N. The 5' untranslated region of Apaf-1 mRNA directs translation under apoptosis conditions via a 5' end-dependent scanning mechanism. *FEBS Lett*. **586,** 4139–4143 (2012).
39. Martin, F. *et al*. Cap-assisted internal initiation of translation of histone H4. *Mol. Cell* **41,** 197–209 (2011).
40. Martin, F. *et al*. Ribosomal 18S rRNA base pairs with mRNA during eukaryotic translation initiation. *Nat. Commun*. **7,** 12622, 1–7 (2016).
41. Wellensiek, B. P. *et al*. Genome-wide profiling of human cap-independent translation-enhancing elements. *Nat. Methods* **10,** 747–750 (2013).
42. Weingarten-Gabbay, S. *et al*. Systematic discovery of cap-independent translation sequences in human and viral genomes. *Science* **351,** 1–24 (2016).
43. Thompson, S. R. So you want to know if your message has an IRES? *Wiley Interdisciplinary Reviews: RNA* **3,** 697–705 (2012).
44. Terenin, I. M., Smirnova, V. V., Andreev, D. E., Dmitriev, S. E. & Shatsky, I. N. A researcher's guide to the galaxy of IRESs. *Cell. Mol. Life Sci*. **74,** 1431–1455 (2017).
45. Van Eden, M. E., Byrd, M. P., Sherrill, K. W. & Lloyd, R. E. Demonstrating internal ribosome entry sites in eukaryotic mRNAs using stringent RNA test procedures. *RNA* **10,** 720–730 (2004).
46. Bert, A. G., Grépin, R., Vadas, M. A., Goodall, G. J. & Gre, R. Assessing IRES activity in the HIF-1a and other cellular 5' UTRs. *RNA* **12,** 1074–1083 (2006).
47. Baranick, B. T. *et al*. Splicing mediates the activity of four putative cellular internal ribosome entry sites. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 4733–4738 (2008).
48. Jackson, R. J. The current status of vertebrate cellular mRNA IRESs. *Cold Spring Harb. Perspect. Biol*. **5,** a011569, 1–20 (2013).