RESEARCH ARTICLE SUMMARY

COMPARATIVE GENETICS

Systematic discovery of cap-independent translation sequences in human and viral genomes

Shira Weingarten-Gabbay, Shani Elias-Kirma, Ronit Nir, Alexey A. Gritsenko, Noam Stern-Ginossar, Zohar Yakhini, Adina Weinberger, Eran Segal*

INTRODUCTION: The recruitment of the ribosome to a specific mRNA is a critical step in the production of proteins in cells. In addition to a general recognition of the "cap" structure at the beginning of eukaryotic mRNAs, ribosomes can also initiate translation from a regulatory RNA element termed internal ribosome entry site (IRES) in a cap-independent manner. IRESs are essential for the synthesis of many human and viral proteins and take part in a variety of biological functions, such as viral infections, the response of cells to stress, and organismal development. Despite their importance, we lack systematic methods for discovering and characterizing IRESs, and thus, little is known about their position in the human and viral genomes and the mechanisms by which they recruit the ribosome.

RATIONALE: Our method enables accurate measurement of thousands of fully designed sequences for cap-independent translation activity. By using a synthetic oligonucleotide library, we can determine the exact composition of the sequences tested and can profile sequences from hundreds of different viruses, as well as the human genome, in a single experiment. In addition, synthetic design enables the construction of oligos in which we carefully and systematically mutate native IRESs and measure the effect of these mutations on expression. This reverse-genetics approach enables the characterization of the regulatory elements that recruit the ribosome and provide specificity in translation.

**RESULTS:** We uncover thousands of human and viral sequences with cap-independent translation activity, which provide a 50-fold increase in the number of sequences known to date. Unbiased screening of cap-independent activity across human transcripts demonstrates enrichment of regulatory elements in the untranslated region in the beginning of transcripts (5′UTR). However, we also find enrichment in the untranslated region located downstream of the coding sequence (3′UTR), which suggests a mechanism by which ribosomes are recruited to the 3′UTR to enhance the translation of an upstream sequence. A genome-wide profiling of positive-strand RNA viruses ([+]ssRNA) reveals the existence of translational elements along their coding regions. This finding suggests that [+]ssRNA viruses can translate only part of their genome, in addition to the synthesis and cleavage of a premature polyprotein. Our analysis reveals two classes of functional elements that drive cap-independent translation: (i) highly structured elements and (ii) unstructured elements that act through a short sequence motif. We show that many 5′UTRs can attract the ribosome by Watson-Crick base pairing with the 18S ribosomal RNA, a structural RNA component of the small ribosomal subunit (40S). In addition, we systematically investigate the functional regions of the 18S rRNA involved in these interactions that enhance cap-independent translation.

CONCLUSION: These results reveal the wide existence of cap-independent translation sequences in both humans and viruses. They provide insights on the landscape of translational regulation and uncover the regulatory elements underlying cap-independent translation activity. ■
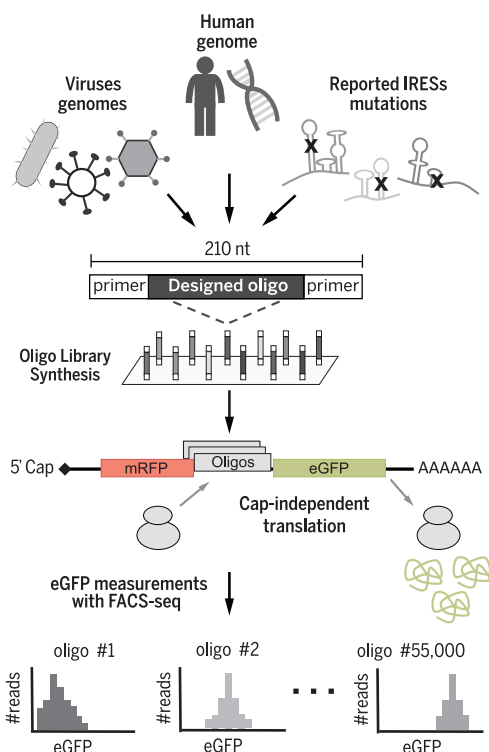

High-throughput bicistronic assay provides insights on translational regulation in humans and viruses. (**A**) A library of thousands of designed oligonucleotides as synthesized and cloned into a bicistronic reporter. Measurements of enhanced green fluorescent protein (eGFP) production, representing cap-independent translation activity, were performed with fluorescence-activated cell sorting and deep sequencing (FACS-seq). (**B**) The landscape of cap-independent translation sequences in humans and viruses and the identified cis-regulatory elements driving their activity.
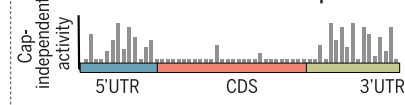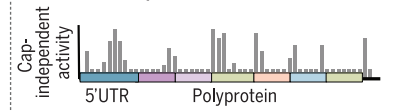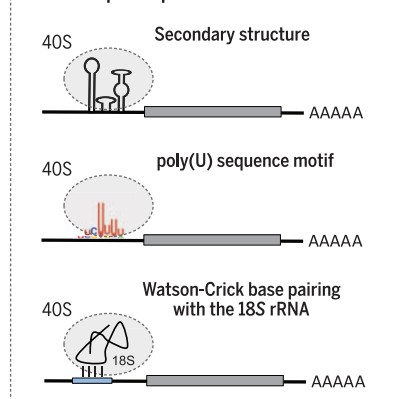
COMPARATIVE GENETICS

Systematic discovery of cap-independent translation sequences in human and viral genomes

Shira Weingarten-Gabbay,[1,2] Shani Elias-Kirma,[1,2] Ronit Nir,[1,2] Alexey A. Gritsenko,[3,4,5] Noam Stern-Ginossar,[6] Zohar Yakhini,[7,8] Adina Weinberger,[1,2] Eran Segal[1,2]*

To investigate gene specificity at the level of translation in both the human genome and viruses, we devised a high-throughput bicistronic assay to quantify cap-independent translation. We uncovered thousands of novel cap-independent translation sequences, and we provide insights on the landscape of translational regulation in both humans and viruses. We find extensive translational elements in the 3′ untranslated region of human transcripts and the polyprotein region of uncapped RNA viruses. Through the characterization of regulatory elements underlying cap-independent translation activity, we identify potential mechanisms of secondary structure, short sequence motif, and base pairing with the 18S ribosomal RNA (rRNA). Furthermore, we systematically map the 18S rRNA regions for which reverse complementarity enhances translation. Thus, we make available insights into the mechanisms of translational control in humans and viruses.

Translation of mRNA is a fundamental process subjected to extensive levels of regulation. Despite the importance of translation control in regulating gene expression, we have lacked high-throughput methods to investigate mRNA translation. Recent technological advances have allowed us to identify and quantify the production of proteins in cells with ribosome profiling (1). However, a systematic characterization of the functional cis-regulatory elements that govern this process is still missing.

Translation initiation in eukaryotes generally involves the recognition of the m7GpppX cap structure at the 5′ end of the transcript (2). However, ribosomes can also initiate translation from a cis-regulatory element in the mRNA termed the internal ribosome entry site (IRES). Ever since its initial discovery in picornaviruses a few decades ago (3, 4), numerous studies have demonstrated that IRESs are crucial for proper regulation of viral and human genes. Many positive-strand RNA viruses or [+]ssRNA viruses, which make up more than one-third of known virus genera (5), are naturally uncapped and rely heavily on IRES-dependent translation for expressing their genome. IRESs are thus essential

for viral infections and the resulting pathologies, and they serve as specific targets for antiviral therapeutic drugs (6). Emerging reports demonstrate novel mechanisms by which IRESs also take part in a variety of biological functions in mammalians even under conditions in which cap-dependent translation is intact, such as the translation of two different proteins from a single bicistronic transcript (7) and accurate gene expression during organismal development (8).

IRESs can act by various mechanisms to recruit the 40S subunit of the ribosome, including the formation of RNA structures, interaction with IRES trans-acting factors (ITAFs), and Watson-Crick base pairing with the 18S ribosomal RNA (rRNA) (9–11). However, experimental limitations have prevented a systematic identification and characterization of these elements (12), and despite much research, only a few dozen cellular and viral IRESs have been discovered to date.

A recent study used an in vitro selection strategy to survey the entire human genome for RNA sequences that enhance cap-independent translation and assayed thousands of native genomic fragments (13). However, although increasing the throughput of measurements is necessary, it is not sufficient to understand the function of these elements. To decipher the cis-regulatory elements driving IRES activity, we need to systematically manipulate many native sequences and to measure their expression. Furthermore, because many [+]ssRNA viruses express their entire proteome using IRES-dependent translation, the identification of these sequences is essential for understanding viral gene regulation. However, current approaches use native genomic fragments as the input sequences and are not applicable to

many viruses, some of which cannot currently be cultured in the laboratory. Thus, the landscape of translational regulation in the vast majority of viruses remains unknown.

To address these fundamental questions, we devised a high-throughput bicistronic assay to quantify cap-independent translation activity of thousands of native and synthetic sequences from human and viral genomes. We provide insights regarding translational regulation in humans and viruses and systematically characterize the cis-regulatory elements involved.

Accurate measurements of 55,000 designed sequences

We designed a library of 55,000 oligonucleotides to screen for novel cap-independent translation sequences in human and hundreds of viral genomes, and to decipher the cis-regulatory elements driving IRES activity (Fig. 1A) [(14) and tables S1, S3 to S5, S7, and S8]. To accurately measure the expression of each of these oligos, we devised a high-throughput bicistronic reporter assay using fluorescence-activated cell sorting (FACS) and high-throughput DNA sequencing (15) (Fig. 1B) (14).

To gauge the accuracy of our measurements, we designed 15 replicates for previously reported IRESs with unique barcodes. We found agreement between oligos with different barcodes for independent synthesis, cloning, sorting, and sequencing (fig. S1). Our measurements also were reproducible between biological replicates [correlation coefficient (R) = 0.90] (Fig. 1C). To evaluate the accuracy of our assay in comparison with each oligo's individual measurement, we isolated 25 clones from the library pool and measured the expression of each isolated clone using flow cytometry, which also agreed between individual measurements and those extracted from the pooled sequence measurements $(R = 0.96)$ (Fig. 1D). Finally, we compared our approach with the traditional luciferase reporter system by cloning 13 oligos from the library into bicistronic luciferase plasmids. Testing their activity with a dual luciferase assay, we found a high correlation between these measurements and the computed expression scores of the library $(R = 0.89)$ (fig. S2).

To identify whether our results were due to cryptic promoter activity, we also performed high-throughput promoter measurements (Fig. 1B). Oligos for which >20% of the reads were obtained in the enhanced green fluorescent protein–positive (eGFP+) population were considered active promoters and were removed from all downstream analyses. In addition, we devised a high-throughput assay to identify cryptic splicing events by quantifying the reduction in the levels of intact bicistronic transcripts in cells. For each oligo from the eGFP+ population, we computed the ratio between deep-sequencing reads obtained from cDNA and genomic DNA (gDNA) samples (fig. S3, A and B). Oligos for which we detected prominent reduction, indicating cryptic splice sites, were removed from all downstream analyses (14). Notably, our measurements successfully captured

[1]Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot 76100, Israel. [2]Department of Molecular Cell Biology, Weizmann Institute of Science, Rehovot 76100, Israel. [3]The Delft Bioinformatics Laboratory, Department of Intelligent Systems, Delft University of Technology, Delft, Netherlands. [4]Platform Green Synthetic Biology, Delft, Netherlands. [5]Kluyver Centre for Genomics of Industrial Fermentation, Delft, Netherlands. [6]Department of Molecular Genetics, Weizmann Institute of Science, Rehovot 76100, Israel. [7]Department of Computer Science, Technion, Haifa, Israel. [8]Agilent Laboratories, Tel-Aviv, Israel.
*Corresponding author. E-mail: eran.segal@weizmann.ac.il

the previously identified X-linked inhibitor of apoptosis protein (XIAP) and the eukaryotic initiation factor eIF4G1 IRESs that show active cryptic splice sites in some bicistronic plasmids (*16*) (fig. S3B). A lack of correlation between eGFP expression and the computed splicing score (R = −0.07) (fig. S3C) supports the idea that cryptic splicing events are not the predominate signal driving the expression of eGFP
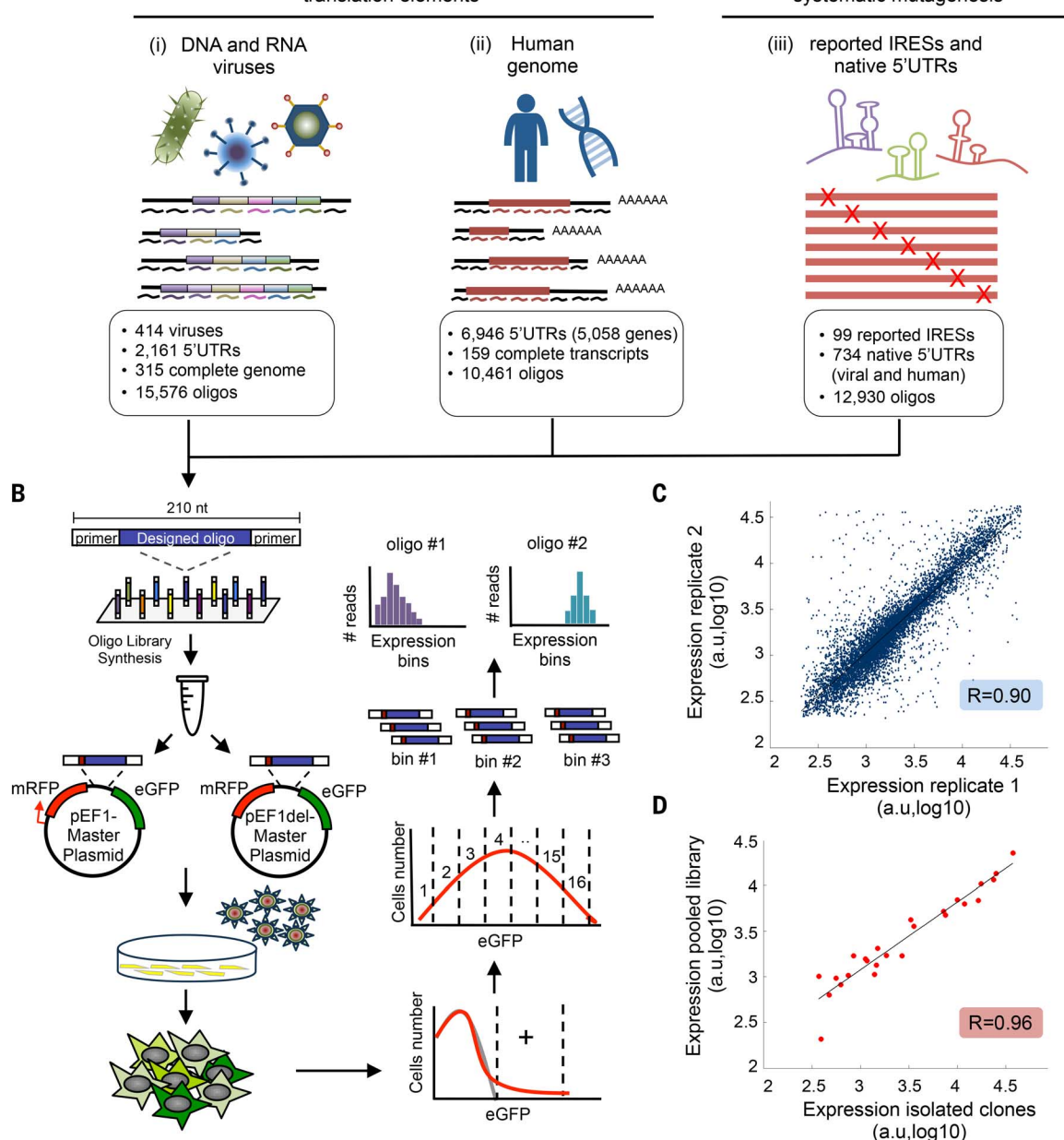

Fig. 1. Synthetic library design and measurements. (**A**) Design of the tested synthetic oligonucleotides: (i) viral 5′UTRs and fragments of complete viral genomes, (ii) human 5′UTRs and fragments of complete transcripts, and (iii) systematic mutagenesis for reported IRESs (*17*) and native 5′UTRs. (**B**) Schematic representation of a high-throughput bicistronic reporter assay: 55,000 designed ssDNA oligos 210 nt in length were synthesized by using oligonucleotide library synthesis technology (*15, 52, 53*). For cap-independent translation measurements, we cloned the library into a lentiviral bicistronic plasmid between mRFP and eGFP reporters and infected H1299 cells resulting in integration of a single oligo into each cell. We then sorted the resulting pool of cells into 16 bins on the basis of eGFP expression with FACS. Next, we used deep sequencing to compute an expression score for each designed oligo on the basis of the distribution of sequence reads across expression bins. For promoter activity measurements, we cloned the library into a plasmid that lacks intrinsic promoter and sequenced eGFP+ population. (**C**) Reproducibility of expression measurements: A comparison between two biological replicates of independent sorting of the library into 16 expression bins. a.u., Arbitrary units. (**D**) Accuracy of expression measurements: 25 clones, each expressing a single oligo, were measured for each clone individually by flow cytometry. A comparison between these isolated measurements and those calculated from the pooled expression measurements is shown.

in the library. To test this directly, we performed quantitative real-time polymerase chain reaction (qRT-PCR) with three sets of primers for different regions on the mRFP cistron. Although a clear reduction was obtained for cells expressing the XIAP IRES, we detected no differences between the empty vector and the eGFP⁺ population, which provides additional evidence that most of the positive oligos that we identified do not contain a cryptic splice site (fig. S4).

Identification of human and viral 5′UTRs with cap-independent translation activity

To confirm that our assay can identify sequences with cap-independent translation activity, we included all the reported IRESs (*17*) in our library design. In cases of IRESs longer than the ssDNA oligo length, we designed multiple oligos spanning the entire sequence of the investigated IRES [table S1, (*14*)]. Our assay successfully captures the activity of 43 of the 119 cellular and viral IRESs reported, including prominent IRESs such as c-Myc, p53, and Apaf-1 (fig. S5). Moreover, these measurements reveal the location of the functional regulatory elements within some long IRESs, such as the encephalomyocarditis virus (EMCV) IRES (fig. S6A). However, our library, which is limited to ~200–nucleotide (nt)–length sequences, cannot detect some long complex IRESs such as the hepatitis C virus (HCV) IRES (fig. S6B).

To identify novel human 5′ untranslated regions (5′UTRs) with cap-independent translation activity, we tested 6946 native 5′UTRs from the human genome. We focused on genes that remained associated with polysomes in eight different conditions, in which cap-dependent translation was suppressed (table S2) (*18–25*); genes that contain short complementary sequences to the 18*S* rRNA that function as short IRES elements (*26–30*); and genes with alternative isoforms that differ in their translation start site. Previous reports estimated that 5 to 10% of cellular mRNAs recruit the ribosome through cap-independent mechanisms (*31*). However, a systematic screening of 5′UTRs for cap-independent activity has not been carried out to date, and thus, only a handful of 5′UTRs and genes that harbor cap-independent activity are currently known. Our assay revealed 583 genes with positive expression encompassing various biological processes, such as translation (eIF2B4, RPL9, RPL41); transcription (SOX5, GATA1, HMGA1); signal transduction (PI4KB, IGF1, BID) and others (Fig. 2A and table S3). Examining the activity of randomly selected 5′UTRs (*14*), we determined that ~10% of human 5′UTRs harbor cap-independent sequences (fig. S7). Note that gene ontology analysis revealed that no specific biological process, cellular component, or molecular function was enriched, which suggests that cap-independent translation is a global mechanism shared among genes with various functions.

To systematically screen for cap-independent translation sequences in viruses, we tested 2161 native 5′UTRs of all the annotated open reading frames (ORFs) in 414 RNA and DNA viruses. We selected viruses from families for which IRES ele-

ments were described, such as the picornaviridae and dicistroviridae, as well as viruses that cause human pathologies such as human papilloma viruses (HPV), herpesviruses, and human immunodeficiency virus (HIV). Our assay identified 471 novel 5′UTRs with positive cap-independent translation activity in a wide span of viruses, including human cytomegalovirus (HCMV), Kaposi's sarcoma–associated herpesvirus (KSHV), adenovirus, simian virus 40 (SV40), HPV, and many others (Fig. 2B and table S4). The observed activity for this large collection of heterogeneous genomes included in the design and representing viruses from different groups {double-stranded DNA (dsDNA), ssDNA, dsRNA, [+]ssRNA, [–]ssRNA, and retroviruses} suggests that cap-independent translation is used by a variety of viruses beyond the genomes tested here.

Comparison between human and viral 5′UTRs reveals that the fraction of cap-independent sequences is higher in viruses and that they are more active than human 5′UTRs in general (Wilcoxon rank-sum test, $P < 10^{-6}$) (Fig. 2C). Sequence analysis of active 5′UTRs identified specific differences. Viral 5′UTRs have lower GC content and higher minimal free energy (MFE) in comparison with their human counterparts (*t* test, $P < 10^{-55}$ and $P < 10^{-37}$, respectively) (Fig. 2C). To test if these features are associated with expression levels, we compared the GC content and MFE for all active and inactive 5′UTRs from both human and viral origin. Indeed, active 5′UTRs have lower GC content and higher MFE (*t* test, $P < 10^{-49}$ and $P < 10^{-51}$ respectively) (Fig. 2D).

Systematic mutagenesis reveals two functional classes of IRESs

Reverse-genetics approaches using mutation scanning have been successfully used to uncover the cis-regulatory elements driving IRES activity (*32, 33*). However, current techniques have limited ability to construct and measure mutated sequences for a large number of IRESs in many positions. As our method enables the construction of a large number of fully design sequences, we performed a systematic mutagenesis for 99 reported IRESs and 734 viral and human 5′UTRs (Fig. 3A). To evaluate our ability to detect cis-regulatory elements using scanning mutagenesis, we examined the ODC1 IRES with known cis-regulatory elements (*32*). Notably, our assay captures the two elements that stimulate ODC1 IRES activity (Fig. 3B) and demonstrates that systematic mutagenesis can decipher IRES regulatory elements.

Examining all sequences with positive activity (*n* = 100) reveals two different mutagenesis profiles: (i) IRESs for which expression is reduced only when a specific position is mutated and (ii) IRESs for which mutation in most positions greatly reduces expression (Fig. 3, C and D). We termed these two classes "local" and "global" sensitivity, respectively. These two classes may represent differences in the underlying mechanism for IRES activity. IRESs can either act through a short sequence motif, such as ITAF binding sites, in which only mutations in a specific motif reduce activity (local sensitivity). Or

IRES activity can involve the formation of a secondary structure, in which mutations at various positions can disrupt the overall structure and result in reduced activity (global sensitivity). Computing the MFE for the two classes separately reveals that globally sensitive IRESs have more structured sequences (i.e., significantly lower MFE values) compared with local sensitivity IRESs (*t* test, $P < 10^{-5}$) (Fig. 3E). Note that in the case of local-sensitivity, most of the inactivating mutations reside within ~60 nt upstream of the AUG triplet (Fig. 3D) which suggests that proximity of the functional element to the start codon is essential for the activity of viral and cellular IRESs.

We examined the high-throughput measurements that we had performed for splicing and promoter activities and found no significant differences between the two clusters (*t* test, $P > 0.2$ and Wilcoxon rank-sum test, $P > 0.3$, respectively) (fig. S8, A and B). Moreover, mRNA measurements of the mRFP and eGFP cistrons in cells expressing two individual IRESs from the local cluster (DAP5 and ELG1) confirm that eGFP expression is driven by intact bicistronic mRNA (fig. S8C).

Searching for enriched short sequence motifs for positive IRESs in all native sequences (*14*) reveals an enriched poly(U) motif [feature motif model hypergeometric *P* value (*34*), $P < 10^{-160}$] and that this motif is significantly enriched in positions for which mutations caused reduction in expression in our scanning mutagenesis assay (Fisher's exact test, $P < 10^{-3}$) (fig. S9A). This enrichment cannot solely be explained by low GC content because we do not find enrichment for poly(A) in active sequences (Fisher's exact test, $P > 0.5$) (fig. S9, A and B). Next, we divided the native sequences into "structured" and "unstructured" groups according to the computed MFE and searched each group for enriched motifs for positive IRESs. Although the poly(U) motif was enriched in positive unstructured IRESs, no specific motif was found for positive structured IRESs (Fig. 3F, compare *P* values), adding to the evidence for differences in the underlying mechanism of these two classes.

Mapping 18*S* rRNA regions that enhance cap-independent translation

Base pairing of mRNA and the 18*S* rRNA is the underlying mechanism by which some short cellular and viral IRESs recruit the 40*S* subunit of the ribosome (*11, 26–30*) (Fig. 4A). Thus, searching for complementary sequences to the 18*S* rRNA can be used to predict IRES activity, yet little is known about the 18*S* rRNA regions involved in this process. To systematically map these regions, we designed 171 oligos with sequences complementary to human 18*S* rRNA encompassing its entire 1869 nt with high-resolution and 164-nt overlap between fragments (Fig. 4A and table S5). For each position on the 18*S* rRNA, we computed the averaged expression of all the oligos containing the corresponding complementary sequence. This analysis uncovers one distinct region (nt 812 to 1233) for which

complementary sequences have cap-independent translation activity (Fig. 4B), referred to here as the "active region."

To test if this region is accessible for interactions with a translated mRNA, we compared our functional measurements with the reported 18S rRNA positions that contact mRNA in the eukaryotic initiation complex (*35*). It is striking that the two helices that interact with the mRNA upstream of the start codon, h23 and h26, are contained within the active region (Fig. 4B). These helices enhance the translation of IGF1R and HCV IRESs, respectively, by Watson-Crick base pairing (*36–38*). In addition, three short IRESs complementary to the 18S rRNA: tobacco etch virus (TEV) (*28*), poliovirus type 2 (*29*), and Gtx (*26*), map to the active region, which demonstrates that our systematic mapping successfully captures reported positions of the 18S rRNA involved in cap-independent translation. Notably, our assay reveals additional positions for

which complementary sequences present cap-independent translational activity, which suggests the existence of additional short IRES elements. Computing the enrichment of all the possible substrings of length *k* that are contained in a string, or *k*-mers, derived from the active region in oligos with positive cap-independent translation from the entire library reveals 134 novel significant elements [Wilcoxon rank-sum test, *P* < 0.01, false discovery rate (FDR) corrected]. Note that the fraction of enriched *k*-mers derived from the 18S rRNA–active region is higher than the inactive regions (two-proportion *z* test, *P* < 10⁻⁷) (Fig. 4C and table S6), which provides additional evidence that this region is involved in cap-independent translation.

To investigate the contribution of 18S rRNA complementarity to the translation of viral and human genes, we examined the presence of short complementary elements in functional regulatory sequences identified by scanning mutagenesis.

We found that the "UACUCCC" and "UUCCUUU" elements, which were originally discovered in the TEV and poliovirus type 2 IRESs, positively regulate the translation of additional viral and human sequences (Fig. 4D and figs. S10 and S11A). To directly test the effect of the UACUCCC element on expression, we designed native sequences that contain two UACUCCC sites and oligos in which we mutated each site separately and the two sites together (fig. S11B). Notably, the effect on expression of the same element varies between transcripts and even between different sites within a single transcript, which suggests that additional parameters such as the relative distance to the AUG and the surrounding sequence contribute to its activity. Finally, to gauge the functionality of the novel short elements uncovered here, we examined four *k*-mers from different positions of the 18S rRNA active region and found that they reside within regulatory sequences of viral and cellular cap-independent sequences (Fig. 4E).
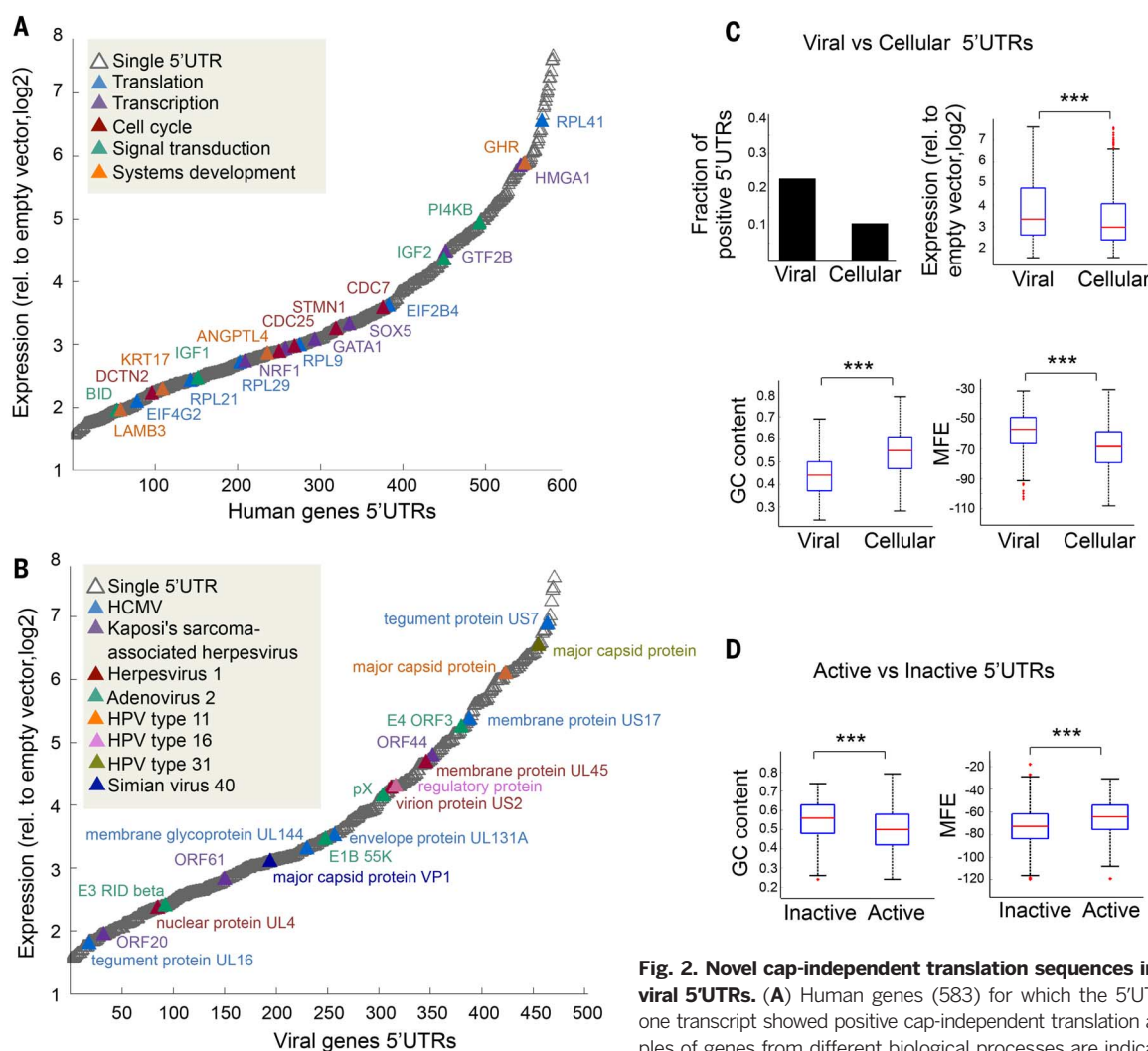

Fig. 2. Novel cap-independent translation sequences in human and viral 5′UTRs. (**A**) Human genes (583) for which the 5′UTR of at least one transcript showed positive cap-independent translation activity. Examples of genes from different biological processes are indicated. (**B**) Viral genes (471) from 414 different viruses for which 5′UTR exhibited positive cap-independent translation activity. Examples of genes from different viruses are indicated. (**C**) Comparison between human and viral 5′UTRs of (i) the fraction of positive 5′UTRs from all 5′UTRs; (ii) expression levels of positive 5′UTRs (*P* < 10⁻⁶, Wilcoxon rank-sum test); (iii) GC content (*P* < 10⁻⁵⁵, *t* test); and (iv) MFE (*P* < 10⁻³⁷, *t* test). (**D**) Comparison of GC content (*P* < 10⁻⁴⁹, *t* test) and MFE (*P* < 10⁻⁵¹, *t* test) for all active and inactive 5′UTRs (human and viral).

Human 3'UTRs contain cap-independent translation sequences

Although some reports indicate the presence of cellular IRESs in coding sequences and 3'UTRs (7, 13, 39–42), the vast majority of the investigated IRESs reside within the 5'UTR region. However, we cannot tell whether IRESs are truly enriched in this region because most of the studies focus on 5'UTRs, under the assumption that translation elements are located upstream to the coding sequence. For this reason, little is known about cap-independent translation activity in non-5'UTR regions.

We performed an unbiased screen of cap-independent translation elements across the entire length of 159 complete human transcripts. For each transcript, we designed nonoverlapping oligos encompassing its entire sequence (Fig. 5A and table S7). We focused on genes for which IRES elements were described or genes associated with translating polysomes under conditions of cap-dependent inhibition in at least two independent studies (table S2) (18–25). To quan-

tify the abundance of active elements per region, we examined activity across the 5'UTR, the coding sequence, and the 3'UTR for all 159 transcripts when centered on either the start or stop codons. As expected, a significant enrichment in activity is obtained for the 5'UTR region compared with the coding sequence with 43 positive oligos of the 191 tested (two-proportion z test, $P < 10^{-5}$) (Fig. 5B). Note that we also find a significant enrichment of cap-independent translation elements in the 3'UTR region as compared with the coding sequence with 381 positive oligos of the 1296 tested (two-proportion z test, $P < 10^{-33}$) (Fig. 5B).

We then examined promoter and splicing activities. As expected, promoter activity is significantly enriched in the 5'UTR region, which may stem from residual core promoters elements, and is depleted from the coding sequence and 3'UTR regions (two-proportion z test, $P < 10^{-12}$) (Fig. 5C). Splicing activity is significantly reduced in the 3'UTR region compared with the 5'UTR and the coding sequence (Wilcoxon rank-sum test, $P <$

10^{-5}) (Fig. 5D). These results are in line with a nonsense-mediated decay (NMD) mechanism that selectively degrades mRNAs harboring premature termination codons that are upstream of the last splicing junction, which results in a stop codon located in the last exon (43). In addition, we confirmed the presence of intact bicistronic transcripts for a few oligos from the 3' UTR region, using qRT-PCR measurements (fig. S12).

These controls demonstrate that the activity measured in the 3'UTR mostly represents cap-independent translation. Notably, cap-independent sequences in the 5'UTR and the 3'UTR regions are not mutually exclusive, and genes present cap-independent activity in the 5'UTR [DNA cross-link repair 1A (DCLRE1A)], the 3'UTR [dystrophin-associated glycoprotein 1 (DAG1)] and activating transcription factor 6 (ATF6)], and in both regions [fibroblast growth factor 1 (FGF1)] (Fig. 5E). Although the group of transcripts selected may have an overall higher fraction of cap-independent translation elements because of our selection
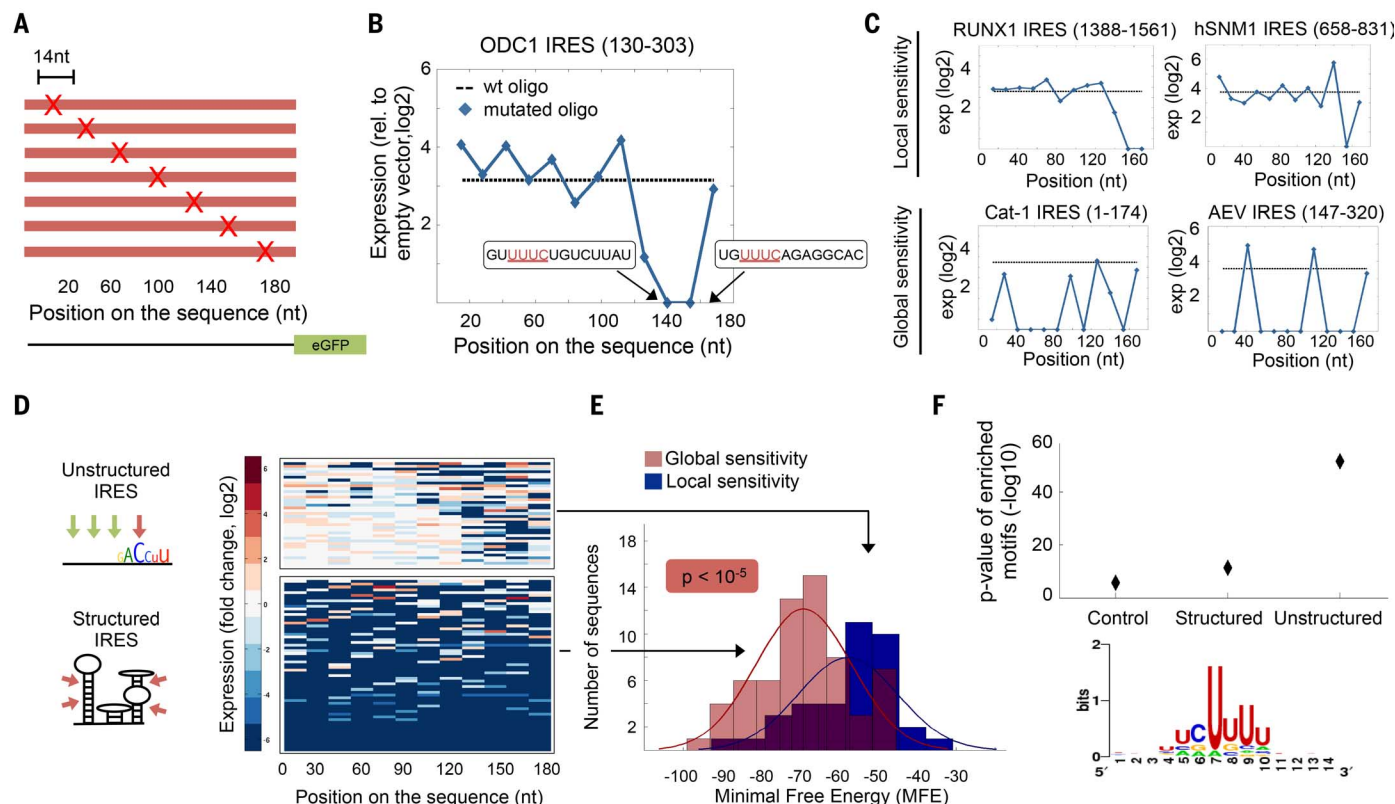
Fig. 3. Systematic scanning mutagenesis for reported IRESs and native 5'UTRs. (**A**) Illustration of systematic scanning mutagenesis. Each oligo contains a 14-nt window in which all nucleotides were mutated. (**B**) Each blue diamond represents a designed mutated oligo in the library. The original sequences of two windows, for which mutations cause reduction in expression, are shown. Highlighted in red are the two UUUC motifs that stimulate IRES-dependent translation of the ODC1 IRES (32). (**C**) Examples of scanning mutagenesis profiles of four reported IRESs showing local and global sensitivity. (**D**) Heat map of scanning mutagenesis profiles for 100 native sequences tested. Each row represents a different IRES and each column, a different position within the mutated IRES. IRESs were clustered into local and global

sensitivity (k-means clustering, $k = 2$). (**E**) MFE was calculated for wild-type IRESs in each cluster separately. A significant shift toward lower MFE values, representing more structural sequences, is obtained for global sensitivity cluster ($P < 10^{-5}$, t test). (**F**) Native sequences were divided into "structured" and "unstructured" groups according to their calculated MFE value (<−80.2 and >−48.10, respectively). Enriched sequence motifs that discriminate between active and inactive IRESs were scanned in each group separately with feature motif model (34) compared with a control, random group of sequences, regardless of activity levels and MFE. The feature motif model hypergeometric P values of enriched motifs for each group and the top-hit motif logo that is enriched in the unstructured group are shown.

criteria, on the basis of our experimental design, we expect no bias in selecting for translation elements in the 3′UTR region for this group of transcripts. This suggests that the observed enrichment in the 3′UTR may apply to other human genes as well.

To validate our findings, we cloned three positive oligos from the 3′UTR region into a monocistronic reporter plasmid downstream of a hairpin structure that attenuates cap-dependent translation. We obtained positive activity for all the three oligos tested, which suggests that they can attract the ribosome directly (fig. S13A). In-
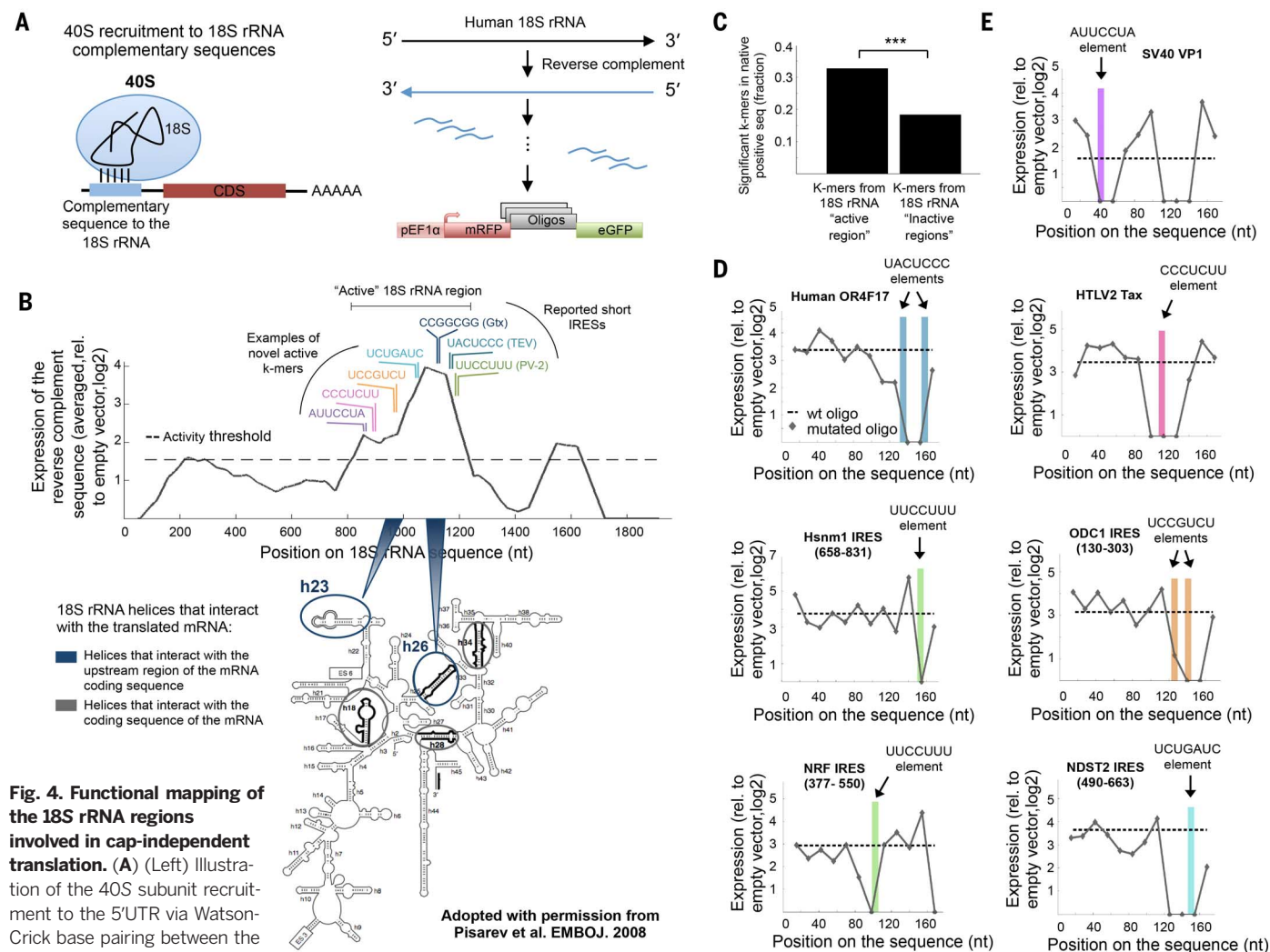
troducing various deletions to one of these sequences led to reduction in expression in both bicistronic and monocistronic reporter plasmids (fig. S13, B and C), which supported our findings that there are functional cap-independent sequences in the 3′UTR region.

[+]ssRNA viruses contain cap-independent sequences in the polyprotein region

More than one-third of known virus genera are [+]ssRNA viruses (44), including HCV, poliovirus, and foot-and-mouth disease virus (FMDV). As

[+]ssRNA viruses replicate in the cytoplasm, most of them are naturally uncapped and, therefore, rely heavily on cap-independent mechanisms for gene expression (45). The production of viral proteins includes the synthesis of a single polyprotein precursor from an IRES element in the 5′UTR followed by protease cleavage that gives rise to the mature proteins (45) (Fig. 6A).

In addition to the canonical mechanism, which ensures equimolar amounts of the viral proteins, we hypothesized that [+]ssRNA viruses may also directly attract the ribosome to enhance the translation of a specific protein. To test this hypothesis,

Fig. 4. Functional mapping of the 18S rRNA regions involved in cap-independent translation. (**A**) (Left) Illustration of the 40S subunit recruitment to the 5′UTR via Watson-Crick base pairing between the 18S rRNA and a reverse-complement sequence. (Right) Design of high-resolution mapping of 18S rRNA regions with reverse-complement fragments that can enhance cap-independent translation. Each fragment represents a single oligo in the library. (**B**) (Top) Expression measurements of 18S rRNA reverse-complement oligos. The positions of the reported short IRES elements of TEV, poliovirus type 2, and Gtx are indicated. (Bottom) Regions on the 18S rRNA that contact the translated mRNA in the eukaryotic initiation complex (35). Helices that contact the mRNA upstream and downstream of the start codon are marked in different colors, and the corresponding positions of the helices that contact upstream regions are denoted on the functional map. (**C**) Comparison between the enrichment of k-mers derived from the 18S rRNA active and inactive regions in oligos with cap-independent translation activity. The fraction of significant k-mers (P < 0.01, Wilcoxon rank-sum test; FDR controlled with

Benjamini Hochberg procedure) from all tested k-mers in each group is shown. Higher fraction of significant k-mers is obtained for the active region (P < 10⁻⁷, two-proportion z test). (**D**) Examples of human sequences that use the UACUCCC (TEV) or UUCCUUU (poliovirus type 2) short IRES elements. Scanning mutagenesis profiles are shown. Each gray dot represents a designed mutated oligo. The positions of the elements (up to 1 mismatch) within the native sequences are indicated. (**E**) Examples of viral and cellular sequences that contain novel k-mers with complementarity to the 18S rRNA active region. We selected four significant k-mers from the analysis in (C) that span various positions on the active region. Examples of scanning mutagenesis profiles of four viral and cellular sequences are shown. The positions of the k-mers (up to 1 mismatch) within the native sequences are indicated.

we profiled cap-independent translation elements across the complete genomes of 131 [+]ssRNA viruses coding a single polyprotein. For each genome, we designed nonoverlapping oligos encompassing its entire length (Fig. 6A and table S8). Notably, in contrast to human transcripts, where the coding sequence is relatively depleted from cap-independent translation elements, polyproteins of uncapped [+]ssRNA viruses (picornaviridae and hepaciviruses) have no significant differences in activity at their 5′UTR and polyprotein regions (two-proportion z test, $P > 0.1$) (Fig. 6, B and C). Note that elevated expression levels in the polyprotein region are specific for uncapped viruses and are not obtained for flaviviruses, [+]ssRNA viruses that modify their genomic RNA with a cap structure synthesized by a viral enzyme (two-proportion z test, $P < 10^{-16}$) (Fig. 6, B and D). Capped flaviviruses have activity levels at their polyprotein similar to those found in coding sequences of human genes ($P > 0.05$) (Fig. 6, B and D).

Examining our measurements for promoter and splice site activities across the viral genomes reveals that the uncapped [+]ssRNA viruses do not contain cryptic promoters or splice sites in the polyprotein region (fig. S14, A and B). Note

that the 5′UTRs of uncapped [+]ssRNA viruses are depleted from promoters and splicing elements in comparison with human 5′UTRs as expected for sequences that were not evolved in the cell nucleus, which demonstrates our ability to accurately detect regulatory elements using high-throughput promoter and splicing activity measurements.

To validate our finding, we cloned two positive sequences from the polyprotein region of the simian sapelovirus 1 and the Ljungan virus to a bicistronic luciferase plasmid. Examining their activity in cells revealed higher expression than the empty vector and the reported Bcl-2 IRES (fig. S15). We tested if the translation elements colocalize with the annotated mature proteins (Fig. 6E) (14) and found no significant colocalization for the tested genomes, which may be due to differences in translation mechanisms for different viruses, selection of different start position when the mature protein is cleaved from the polyprotein precursor or directly translated from the viral genome, or unknown viral ORFs.

Discussion

We present a systematic high-throughput study for the identification and characterization of cis-

regulatory elements that recruit the ribosome in a cap-independent manner and thereby expand by ~50-fold the set of all such sequences known to date. Our unbiased profiling revealed the wide existence of cap-independent translation elements in 3′UTRs of human transcripts and the polyprotein region of [+]ssRNA viruses.

Our results suggest a mechanism by which uncapped [+]ssRNA viruses translate only part of their genome. We speculate that these viruses, which evolved efficient sequence elements for cap-independent translation in the 5′UTR region, can exploit similar elements to facilitate the translation of individual proteins. The conditions under which viruses use these elements and which translated proteins result require further study. Small molecules that interfere with the secondary structure of viral IRESs can serve as specific antiviral agents (6). In that sense, our identification of ~900 IRES elements in RNA viruses increases the number of potential targets for future drug development.

Note that some plant viruses use cap-independent translation elements in their 3′UTR, in a mechanism that involves RNA circularization to place the scanning ribosome near the 5′ end (46–48). Although by a different mechanism, eukaryotic
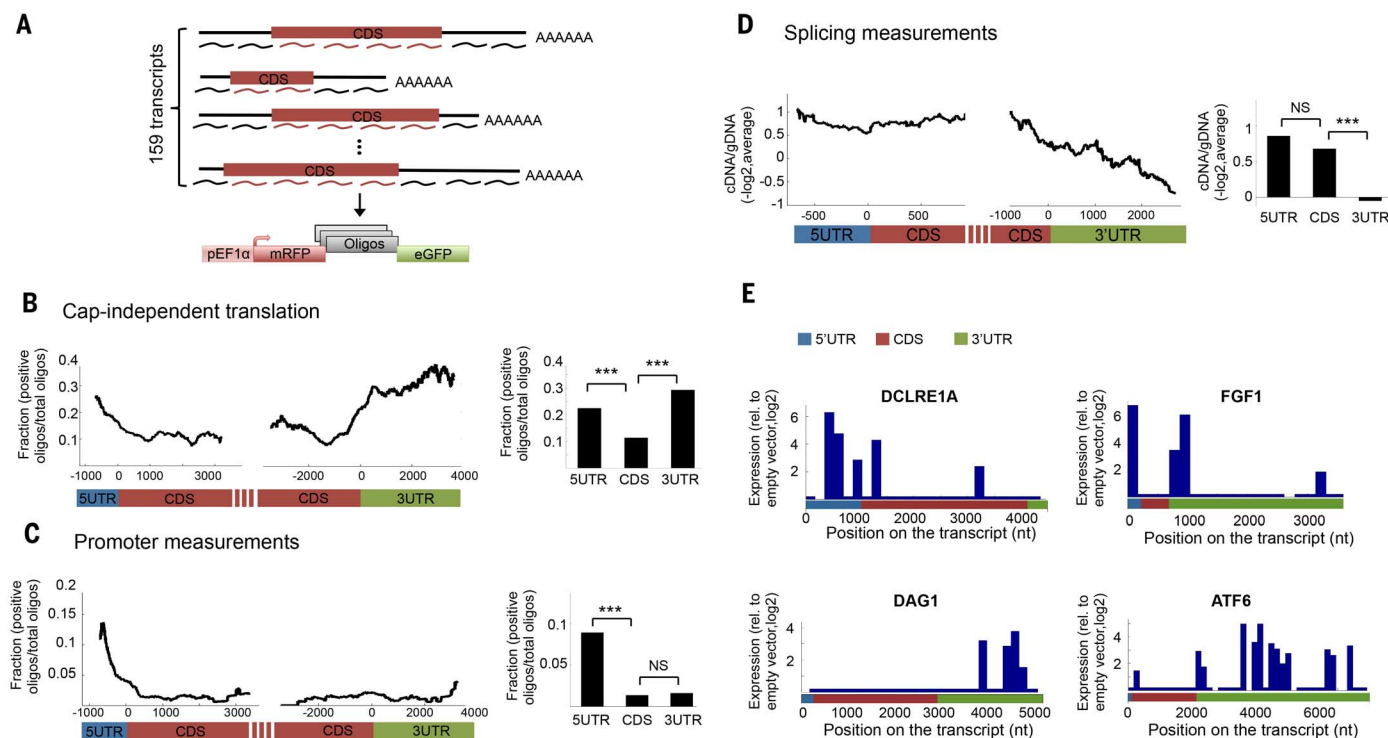

Fig. 5. Unbiased screen for cap-independent translation elements across human transcripts. (**A**) Design of unbiased screen for cap-independent translation elements in 159 human transcripts; each fragment represents a single oligo in the library. (**B**) Cap-independent activity across the 5′UTR, the coding sequence, and the 3′UTR. Transcripts were aligned by their translation start or stop sites and the fraction of positive oligos was computed for each position (y axis). Bar chart denotes the fraction of positive oligos per region. A significant enrichment can be seen for the 5′UTR and the 3′UTR regions in comparison with the coding sequence ($P < 10^{-5}$ and $P < 10^{-33}$, respectively; two-proportion z test). (**C**) Promoter activity, as obtained

from high-throughput promoter assay, across the 5′UTR, the coding sequence, and the 3′UTR. A significant enrichment can be seen for the 5′UTR, but not the 3′UTR, region in comparison with the coding sequence ($P < 10^{-12}$ and $P > 0.5$, respectively; two-proportion z test). (**D**) Splicing activity, calculated as −log of cDNA/gDNA reads number, across the 5′UTR, the coding sequence, and the 3′UTR. Similar splicing activity is obtained for 5′UTR and the coding sequence regions, whereas a significant reduction is obtained for the 3′UTR region ($P > 0.6$ and $P < 10^{-5}$, respectively; Wilcoxon rank-sum test). (**E**) Examples of cap-independent translation activity across four different transcripts.

mRNAs are also circularized in cells through interaction between eIF4G and the poly(A)-binding protein. In addition, tethering the ribosome to the 3′UTR via artificial MS2 coat protein–binding sites or the EMCV IRES can enhance the translation of an upstream reporter gene (49). In light of our findings that cap-independent translation elements frequently occur in 3′UTRs and emerging evidence of high ribosome abun-

dance in the 3′UTR region of eukaryotic transcripts (50, 51), it will be interesting to investigate if a similar mechanism to that described in plant viruses also exists for human transcripts.

As in the case of transcriptional control, in which different elements are combined to generate a variety of expression patterns, the cis-regulatory elements that underlie cap-independent translation act through diverse mechanisms

and spread a broad range of expression. Secondary structure, GC content, complementarity to the 18S rRNA, and stretches of "UUUUU" that we report here are just part of the sequence features governing activity levels. Our large-scale functional activity assay results in thousands of newly discovered cap-independent regulatory sequences in human genes and viruses, bringing us closer toward a mechanistic and
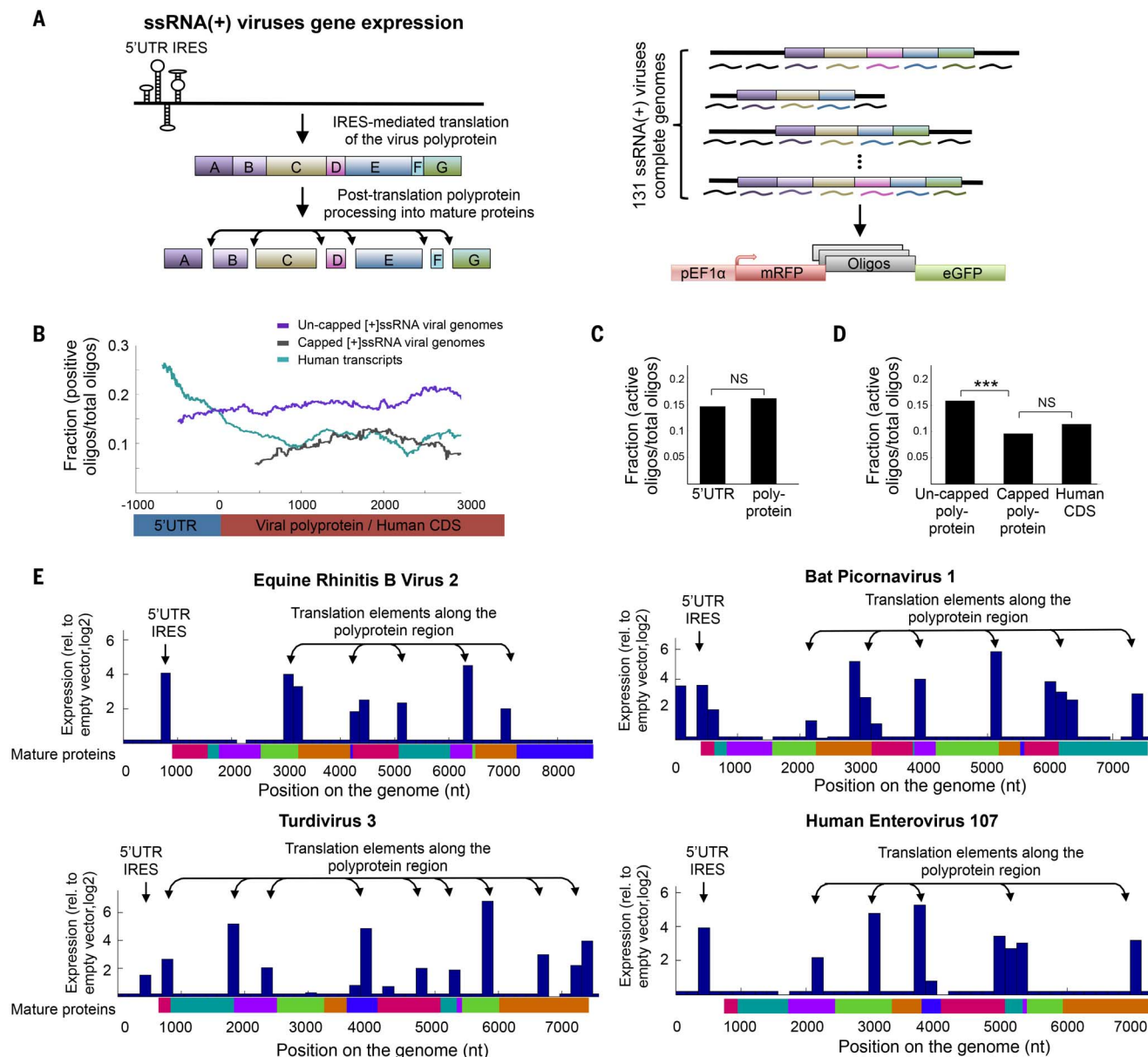
Fig. 6. Genome-wide profiling of cap-independent translation elements in [+]ssRNA viruses. (**A**) (Left) Illustration of canonical gene expression in positive-sense ssRNA genomes. An IRES element in the 5′UTR of the genomic RNA directs the ribosome to synthesize a single polyprotein precursor. Next, the polyprotein is cleaved by proteases to give rise to both the structural and nonstructural viral mature proteins, which are therefore produced in equimolar amounts (45). (Right) Design of genome-wide screen for cap-independent translation elements in 131 [+]ssRNA viral genomes. (**B**) Cap-independent activity across the 5′UTR and the polyprotein regions. A similar profile of human transcripts is shown for comparison (from Fig. 5B). (**C**) The fraction of positive oligos of uncapped viral genomes per region. (**D**) The fraction of positive oligos of uncapped [+]ssRNA viruses, capped [+]ssRNA viruses, and human coding-sequence regions. A significant enrichment can be seen for uncapped viral polyprotein as compared with capped [+]ssRNA viruses ($P < 10^{-16}$, two-proportion z test). (**E**) Examples of cap-independent translation activity across the genome of four different uncapped [+]ssRNA viruses. Mature proteins track: annotations of the mature protein positions from the NCBI database.

quantitative understanding of this important mode of regulation.

Materials and methods
Synthetic library design
Reported IRESs

The full sequences of all the reported cellular and viral IRESs were downloaded from the IRESite database (*17*). In cases of IRESs longer than 174 nt, their sequence was dissected into fragments of 174 nt with overlap of 124 nt between oligos (i.e., 50-nt distance between the start positions of two sequential oligos).

5′UTRs of human genes

We composed a list of 5058 genes from the human genome according to the following criteria: (i) We collected all the genes that remained associated with polysomes in eight different microarray studies, in which cap-dependent translation was suppressed (*18–25*) (table S2). (ii) Genes for which we located complementary sequences to the 18*S* rRNA that were shown to act as functional IRES elements (*26–30*) in their 5′UTR. (iii) Genes with alternative isoforms that differ in their translation start site. In addition, as a random group of genes, we included all the 783 genes from the library of annotated reporter cell clones (LARC) from Uri Alon's lab (*54*). In this library, yellow fluorescent protein (YFP) was fused to endogenous proteins using random integration of retroviruses carrying YFP reporter. Therefore, these genes were not subjected to any selection and should not be specifically enriched for cap-independent translation elements. The full transcripts were downloaded from National Center for Biotechnology Information (NCBI), NIH, and the 174-nt sequence upstream of the annotated start codon for each transcript was extracted.

5′UTRs of viral genes

We composed a list of 414 viruses from the NCBI database. We made sure to represent both RNA and DNA viruses and selected viruses according to two criteria: (i) Families of viruses that were reported to use IRES elements for gene expression: Picornaviridae, Dicistroviridae, Flaviviridae, and Retroviridae. (ii) Viruses that cause clinical pathologies such as herpesvirus, HPV, and HIV. The genomes of the viruses were downloaded from NCBI, and the 174-nt sequence upstream of all the annotated genes was extracted.

Scanning mutagenesis to decipher the cis-regulatory elements

We composed a list of wild-type sequences for scanning mutagenesis including (i) reported IRESs from the IRESite database (mammalians and viruses of humans, vertebrates, and invertebrates). In cases of IRESs longer than 174 base pairs (bp), IRESs were divided into fragments of 174 bp with 86-bp overlap, and the last three fragments were selected; (ii) 5′UTRs of genes that were found to be associated with translating polysomes under conditions of cap-dependent inhibition in at least two independent studies (table S2) (*18–25*); (iii) 5′UTRs of genes that

contain a reported short IRES element with complementarity to the 18*S* rRNA (*26–30*); and (iv) 5′UTRs of all genes from viral genomes for which at least one IRES element was reported. For each wild-type sequence 12 nonoverlapping mutated oligos were designed. Each designed oligo contains a 12- to 14-nt window for which all the nucleotides were mutated.

Sequences complementary to 18S rRNA

The sequence of the human 18*S* rRNA (1869 nt, NR_003286) was downloaded from NCBI. Its reverse-complement sequence was partitioned into 171 fragments of 174 nt with 164-nt overlap between fragments (i.e., 10-nt distance between the start positions of two sequential oligos).

Screening human transcripts

We selected 159 human transcripts according to the following criteria: (i) genes for which IRES elements were described in the literature, and (ii) genes that were found to be associated with translating polysomes under conditions of cap-dependent inhibition in at least two independent studies (table S2) (*18–25*). The full transcript sequences were downloaded from NCBI and non-overlapping 174-nt-long fragments were extracted.

Screening viral genomes

We included the full-length genomes of all the 315 RNA viruses from the list that we composed for the viral 5′UTRs screen. The full genome sequences of the RNA viruses were downloaded from NCBI and nonoverlapping 174-nt-long fragments were extracted.

Experimental procedures
Cell culture

Human embryonic kidney cells 293T (HEK 293T) were cultured in Dulbecco's modified Eagle's medium (Gibco) supplemented with 10% fetal bovine serum (FBS) [Biological Industries, Beit-Haemek, Israel (BI)] and 1% penicillin and streptomycin (P.S., BI). H1299 human lung carcinoma cells were cultured in RPMI 1640 medium (Gibco), supplemented with 10% FBS and 1% P.S. Cells were kept at 37°C in a humidified atmosphere containing 5% CO_2 and were frozen in complete media with 7% dimethyl sulfoxide (DMSO) (Sigma). Trypsin-EDTA solution C (BI) was used to detach cells from culture dishes.

Plasmids

pEF1_XIAP, pEF1_EMCV, pEF1_HCV, and pEF1-del_XIAP plasmids were kindly provided by N. Rahm and A. Telenti (The Institute of Microbiology of the University Hospital Center, Lausanne, Switzerland) (*55*). pMDL, pVSV-G, and pRSV-Rev helper plasmid for lentiviruses packaging were kindly provided by M. Selitrennik and S. Lev (Weizmann Institute of Science, Israel).

Synthetic library production and amplification

The production and amplification steps were adopted from a protocol that was previously described for yeast promoters (*15*). We used Agilent oligo library synthesis technology to produce a pool of 55,000 different fully designed

single-stranded 210-oligomers (Agilent Technologies, Santa Clara, CA). Each designed oligo contains common priming sites and restriction sites for Asc I and Rsr II at the ends, leaving 174 for the variable region. The library was synthesized using Agilent's on-array synthesis technology (*52, 53*) and then provided to us as an oligo pool in a single tube (10 pmol). The pool of oligos was dissolved in 200 μl Tris-ethylenediaminetetraacetic acid (Tris-EDTA). We divided 11 ng of the library (1:50 dilution) into 32 tubes and amplified each tube using PCR. Each PCR reaction contained 24 μl of water with 0.346 ng DNA, 10 μl of 5× Herculase II reaction buffer, 10 μl of 2.5 mM deoxynucleotide triphosphate (dNTPs) each, 2.5 μl of 20 μM forward (Fw) primer, 2.5 μl of 20 μM reverse (Rv) primer, and 1 μl Herculase II fusion DNA polymerase (Agilent Technologies). The parameters for PCR were 95°C for 1 min, 14 cycles of 95°C for 20 s, and 68°C for 1 min, each, and finally one cycle of 68°C for 4 min. The oligonucleotides were amplified using constant primers in the length of 48 nt, which have 18-nt complementary sequence to the single-stranded 210-mers and a tail of 30 nt to allow recognition of products that were not properly cut by restriction enzymes in the next step. Primers sequences follow, underline represents the 18-nt complementary sequence to the ssOligos: TCAGTCGCCGCTGCCAGCTCTCGCACTCT-<u>TCTCGGCGCGCCAGTCCT</u> (Fw primer), TTCTT-CCGCCGCTCCGCCGTCGCGTTTCTC<u>TGCGTCC-GGTCCGAGTCG</u> (Rv primer). The PCR products from all 32 tubes were joined and concentrated using Amicon Ultra, 0.5 ml 30K centrifugal filters (Merck Millipore) for DNA purification and concentration. The concentrated DNA was then purified using a PCR mini-elute purification kit (Qiagen) according to the manufacturer's instructions.

Construction of reporter master plasmids

Two lentiviral plasmids: pEF1_XIAP and pEF1-del_XIAP (without promoter) (*55*) were used as a vector backbone to create the recipient plasmids for the library. All plasmids except the XIAP IRES region were amplified by a PCR reaction with primers that add unique restriction sites for Asc I and Rsr II (for synthetic library cloning), as well as a Bam HI site on both primers to allow plasmid self-ligation. The primers used for this reaction were as follows: GGTggatccGGGTTGGGTGCG-GACCGatggtgagcaagggcgaggag (Fw primer), CAAggatccCAACACACCCGGCGCGCGCCctagtttaaacgtct-agagccac (Rv primer). Each PCR reaction contained 30.5 μl water with 20 ng of recipient plasmid, 10 μl of 5× Phusion polymerase reaction buffer, 1 μl of 10 mM dNTPs mix, 2.5 μl of 10 μM 5′ primer, 2.5 μl of 10 μM 3′ primer, 1 μl Phusion polymerase (Thermo Fisher Scientific), and 2.5 μl of DMSO (Thermo Fisher Scientific). The parameters for PCR were 95°C for 30 s, 30 cycles of 95°C for 30 s, 60°C for 1 min, and 72°C for 5 min, each, and finally one cycle of 72°C for 7 min. The amplified vectors were separated from unspecific fragments by electrophoresis on a 1% agarose gel stained with GelStar (Cambrex Bio Science Rockland), cut from the gel, and purified using

a gel extraction kit (Qiagen). Purified vectors were cut with Bam HI [New England Biolabs (NEB)] for 1 hour at 37°C. To digest the original plasmids templates Dpn I (NEB) was added to the reaction. Products were cleaned with PCR purification kit (Qiagen), and 50 ng of the amplified vectors were self-ligated for 20 min at room temperature using Quick DNA Ligase enzyme (NEB). Next, ligated plasmids were transformed into *Escherichia coli* (self-made Hit-DH5α cells) by using heat shock, positive colonies were grown in 2YT medium, and the plasmids were purified using a plasmid mini-kit (RBC BioScience).

Synthetic library cloning into reporter plasmids

The amplified synthetic library was cloned into the two master plasmids described above to create the pEF1-mRFP-oligos-eGFP library for cap-independent translation measurements and pEF1del-mRFP-oligos-eGFP library for promoter activity measurements. Library cloning into the master plasmids was adopted from a protocol that was previously described for yeast promoters (*15*). Purified library DNA (720 ng total) was cut with the unique restriction enzymes Asc I and Rsr II (Fermentas FastDigest) for 2 hours at 37°C in four 40-μl reactions containing 4 μl fast digest (FD) buffer, 1 μl of Asc I enzyme, 2.5 μl of Rsr II enzyme, 0.8 μl of dithiothreitol (DTT), and 18 μl of DNA, followed by a heat inactivation step of 20 min at 65°C. Digested DNA was separated from smaller fragments and uncut PCR products by electrophoresis on a 2.5% agarose gel stained with GelStar (Cambrex Bio Science Rockland). Fragments the size of 210 bp were cut from the gel and eluted using electroelution Midi GeBAflex tubes (GeBA, Kfar Hanagid, Israel). Eluted DNA was precipitated by using standard Na acetate–isopropanolol protocol. The master plasmids were cut with Asc I and Rsr II (Fermentas FastDigest) for 2.5 hours at 37°C in a reaction mixture containing 9 μl of FD buffer, 3 μl of each enzyme, 3 μl of alkaline phosphatase (Fermentas), and 4.5 μg of the plasmid in a total volume of 90 μl, followed by a heat inactivation step of 20 min at 65°C. Digested DNA was purified using a PCR purification kit (Qiagen). The digested plasmids and DNA library were ligated for 0.5 hours at room temperature in two 10-μl reactions, each containing 150 ng plasmid and the library in a molar ratio of 1:1, 1 μl CloneDirect 10× ligation buffer, and 1 μl CloneSmart DNA ligase (Lucigen Corporation), followed by a heat inactivation step of 15 min at 70°C. Ligated DNA (14 μl) was transformed into a tube of *E. coli* 10G electrocompetent cells (Lucigen) divided into seven aliquots (25 μl each), which were then plated on 28 Luria broth (LB) agar (200 mg/ml amp) 15-cm plates. To ensure that all 55,000 oligos are represented, we collected ~1.3 ×10^6 colonies (2 ×10^6 in the pEF1del plasmid) 16 hours after transformation, by scraping the plates into LB medium. Library pooled plasmids were purified using a NucleoBond Xtra maxi kit (Macherey Nagel). To ensure that the collected plasmids represent a ligation of single inserts, we performed colony PCR on 96 random colonies. The volume of each PCR reaction was 30 μl; each reaction contained a random colony picked from an LB plate, 3 μl of 10× DreamTaq (Thermo Fisher Scientific) buffer, 3 μl 2 mM dNTPs mix, 1.2 μl of 10 μM 5′ primer, 1.2 μl of 10 μM 3′ primer, 0.3 μl DreamTaq polymerase. The parameters for PCR were 95°C for 5 min, 30 cycles of 95°C for 30 s, 68°C for 30 s, and 72°C for 40 s, each, and finally one cycle of 72°C for 5 min. The primers used for this reaction were CCACAACGAGGACTACACCA (Fw primer) and GT-AGGTCAGGGTGGTCACGA (Rv primer). Of the 96 colonies tested for each of the pEF1-mRFP-oligos-eGFP and pEF1del-mRFP-oligos-eGFP libraries, only one and two colonies had multiple inserts, respectively.

Lentiviruses production and infections

HEK 293T cells were used for lentivirus production. Twenty-five 6-cm plates were coated with poly-lysin 0.001% (Sigma), incubated for 1 hour, and washed three times with phosphate-buffered saline (PBS). Cells (5 × 10^5) were plated on each of the 25 plates, 24 hours before transfection. Cells were cotransfected with three helper plasmids (pMDL, pVSV-G, and pRSV-Rev) and one of the library plasmids (pEF1-mRFP-oligos-eGFP or pEF1del-mRFP-oligos-eGFP). Each transfection included 100 μl of Dulbecco's minimum essential medium (DMEM) with no serum or antibiotics, 18 μl of FuGENE 6 transfection reagent (Promega), 2.7 μg of library plasmids, 1.7 μg of pMDL, 0.9 μg of pVSV-G, and 0.7 μg pf pRSV-Rev. Transfection was performed according to the manufacturer's instructions. After 24 hours, medium was replaced with fresh DMEM, and after an additional 24 hours, ~90 ml of virus-containing media was collected, filtered with 0.45-μm filters (mercury), divided into five 50 ml tubes and stored in –80°C. To determine the titer of the produced lentiviruses, 5 × 10^5 H1299 cells were plated on 10-cm plates 24 hours before infection. We thawed one frozen tube of viruses and performed serial dilutions of 1:1, 1:5, 1:100, 1:500, and 1:1000 in DMEM. Diluted virus-containing media (3.5-ml samples) were added to 1.5 ml RPMI media in each plate (total volume of 5 ml) in addition to 5 μl of Polybrene (AL-118, Sigma). After 24 hours cells were washed three times with PBS, and fresh RPMI complete medium was added. After an additional 24 hours, cells from each plate were harvested and plated on a 15-cm plate. We analyzed the percentages of mRFP$^+$ cells (representing viral integrations) for each dilution using flow cytometry and determined the exact viral titer in the original sample. We computed the amount of viruses needed for multiplicity of infection (MOI) of 0.1 and repeated the infection protocol with the calculated viruses volume for 25 10-cm plates with 5 × 10^5 H1299 cells in each plate. A total of 1.25 × 10^6 cells were infected so that each designed sequence in the library (*n* = 55,000) was independently integrated into ~23 individual cells on average.

Sorting libraries by FACS

H1299 cells from both libraries were grown for 5 days after infection in RPMI medium and split 1:3 into 15-cm plates 1 day before sorting. On the day of sorting, cells were trypsinized, centrifuged, resuspended in sterile PBS, and filtered using cell-strainer capped tubes [Becton Dickinson (BD) Falcon]. Sorting was performed with BD FACSAria II SORP (special-order research product) at low sample flow rate and a sorting speed of ~8000 cells/s. pEF1-mRFP-oligos-eGFP library sorting: To sort cells that integrated the reporter construct successfully (10% of the infected population), we determined a gate according to mRFP fluorescence so that only mRFP-expressing cells were sorted. Library sorting was conducted in two steps. In the first step, we sorted eGFP$^+$ population, which represent sequences with cap-independent translation activity. To set the eGFP gate above background level, H1299 cells infected with an empty vector (pEF1-mRFP-eGFP) were used. Thirty 15-cm plates of library-infected cells were grown for sorting, and a total of 10^5 cells were collected (5% of library). To obtain high-resolution measurements for the sequences with positive expression, we performed a second step in which eGFP$^+$ population was grown for an additional week and resorted into 16 bins according to eGFP levels. We collected a total of 4.4 × 10^6 cells for all 16 bins. In addition, we sorted mRFP$^+$ cells, representing the entire library of 55,000 designed oligos to determine the representation of each oligo in the original library. pEF1del-mRFP-oligos-eGFP sorting: To obtain sequences with positive promoter activity, we sorted eGFP$^+$ population. To ensure a measure comparable with the pEF1-mRFP-oligos-eGFP library, we used the exact same gate for eGFP and ran the same number of cells in the FACS as we did when sorting the eGFP$^+$ population for the pEF1-mRFP-oligos-eGFP library. Twenty 15 cm plates of infected cells were grown for sorting and a total of 60,000 cells (2.5% of library) were collected.

Isolated clones measurements

Three isolated clones from each of the 16 expression bins were grown from single cells that were sorted into a 96-well plate. After 28 days, cells were analyzed in Flow Cytometry for eGFP expression and genomic DNA (gDNA) was purified. DreamTaq DNA polymerase (Thermo Fisher Scientific) was used to amplify the library from 200 ng gDNA, with same conditions and primers as in the library colony PCR. The PCR product was Sanger sequenced from the PCR Fw primer. Colonies for which two or more eGFP peaks were obtained (representing two or more single cells that were sorted into the same well) were removed from the analysis.

Preparing samples for sequencing

In order to maintain the complexity of the library amplified from gDNA, PCR reactions were carried out on a gDNA amount calculated to contain an average of 200 copies of each oligo included in the sample. The amounts of cells used for each gDNA purification were 3.5 × 10^7 cells (170 μg gDNA) of total pEF1-mRFP-oligos-eGFP population, 7 × 10^6 cells (10 μg gDNA) for each expression

bin, 3×10^8 cells (1500 µg gDNA) of total pEF1del-mRFP-oligos-eGFP population, and 6×10^7 cells (20 µg) of eGFP⁺ population. gDNA was purified using DNeasy blood and tissue kit (Qiagen) or blood and cell culture DNA maxi kit (Qiagen). For each population, a two-step nested PCR was performed in multiple tubes (to include the required amount of gDNA), each containing 100 µl (in both steps). In the first step, each reaction contained 10 µg gDNA, 50 µl of Kapa Hifi ready mix X2 (KAPA Biosystems), 5 µl of 10 µM 5′ primer, and 5 µl of 10 µM 3′ primer. The parameters for the first PCR were 95°C for 5 min, 18 cycles of 94°C for 30s, 65°C for 30s, and 72°C for 40s, each, and finally one cycle of 72°C for 5 min. The primers used for this reaction were CCA-CAACGAGGACTACACCA (Fw primer) and GTA-GGTCAGGGTGGTCACGA (Rv primer). In the second PCR step, each reaction contained 5 µl of the first PCR product (uncleaned), 50 µl of Kapa Hifi ready mix X2 (KAPA Biosystems), 5 µl 10 of µM 5′ primer, and 5 µl of 10 µM 3′ primer. The PCR program was similar to the first step, using 24 cycles. Specific primers corresponding to the constant region of the plasmid were used. The 5′ primer also had a unique upstream 5-nt barcode sequence (underlined) (5′-XXXXX TAGGG-CGCGCCAGTCCT-3′), and three different barcodes were used for each bin. The 3′ primer was common to all bins (5′-NNNNN CTCACCATCGGT-CCGAGTCG-3′, where 'N's represent random nucleotides). The concentration of the PCR samples was measured using a monochromator (Tecan i-control), and the samples were mixed in ratios corresponding to their ratio in the population. The library was separated from unspecific fragments by electrophoresis on a 2% agarose gel stained by EtBr, cut from the gel, and cleaned in 2 steps: gel extraction kit (Qiagen) and SPRI beads (Agencourt AMPure XP). The sample was assessed for size and purity at the Tapestation, using high sensitivity D1K screenTape (Agilent Technologies, Santa Clara, California). 100 ng library DNA were used for library preparation for NGS; specific Illumina adaptors were added, and DNA was amplified using 8 amplification cycles, protocol adopted from Blecher-Gonen *et al.* (*56*). The sample was reanalyzed using Tapestation.

RNA purification for splicing measurements

RNA was purified from 2×10^6 eGFP⁺ library cells, using the Nucleospin RNA II kit (Macherey-Nagel). cDNA was prepared from 1 µg RNA using Verso cDNA synthesis kit (Thermo Fisher Scientific) and random hexamers. PCR on the cDNA as template was carried out using Kapa Hifi ready mix X2 (KAPA Biosystems) with the library constant primers. To enable direct comparison with the gDNA sample, we used exactly the same PCR reaction setup and program.

Quantitative real-time PCR (qRT-PCR)

For each sample, RNA was purified from $\sim 2 \times 10^6$ cells using Nucleospin RNA II kit (Macherey-Nagel). cDNA was prepared from 1 µg RNA using Verso cDNA synthesis kit (Thermo Fisher Scien-

tific) and random hexamers, and diluted 1:8 for real-time PCR. Real-time PCR was carried out using KAPA SYBR FAST qPCR Kit (Kapa Biosystems) in a StepOnePlus machine (Applied Biosystems). Expression levels were normalized to the glyceraldehyde phosphate dehydrogenase (GAPDH) housekeeping gene. Primers used were GAGTTCATGCGCTTCAAGGTGC (mRFP Fw1), TGGAGCCGTACTGGAACTGAGG (mRFP Rv1) GCAGGACGGCGAGTTCATCTAC (mRFP Fw2), TCCTTCAGCTTCAGCCTCATCTTG (mRFP Rv2), CTGAAGGGCGAGATCAAGATGAG (mRFP Fw3), CCTCGTTGTGGGAGGTGATGTC (mRFP Rv3), GCCACAAGTTCAGCGTGTCC (eGFP Fw), GTAGC-GGCTGAAGCACTGCAC (eGFP Rv), GTCGGAGT-CAACGGATTTGG (GAPDH Fw), AAAAGCAGCCC-TGGTGACC (GAPDH Rv).

Construction of luciferase reporter plasmids

Bicistronic luciferase plasmids: Sequences of 13 selected oligos and the linker sequence from the empty vector were inserted instead of the HCV IRES in the bicistronic luciferase plasmid pRL-HCV_IRES-pFL, by using a restriction free (RF) cloning method (*57*).

Monocistronic luciferase plasmids: The tested sequences were amplified from bicistronic luci-fearse plasmids together with the Firefly coding sequence and the terminator, and cloned into the monocistronic luciferase plasmid php-Bcl2_IRES-FL downstream of a hairpin structure. The amplified sequences were cloned instead of the Bcl-2 IRES and the respective downstream sequences using the restriction enzymes Not I and Psp XI (NEB).

Deletion of sequences in different luciferase plasmids: Deletions in the length of 70 bp were made using the Transfer PCR (TPCR) method (*58*). Briefly, a primer flanking the deletion and a downstream primer were used to create a mega-primer that, in turn, is used for entire plasmid synthesis in the same reaction. Deletions were created on bicistronic and monocistronic plasmids using the same approach. Primers used for Luciferase assay construct cloning are shown in table S1.

Luciferase reporter assay

Dual luciferase assay was performed as described before (*59*). Briefly, 2.5×10^4 H1299 cells were plated per well in a 24-well plate, 24 hours before transfection. Cells were transfected with 30 ng bicistronic plasmid by using Lipofectamine 2000 transfection reagent, and medium was exchanged after 6 hours. Renilla and Firefly luciferase activities were measured 48 hours posttransfection by using the Dual-Luciferase Reporter Assay System (Promega, Madison, WI), according to the manufacturer's protocol, with a Veritas Lumino-meter (Promega). Monocistronic luciferase assay: 1.5×10^5 H1299 cells were plated per well in a 6-well plate, 24 hours before transfection. Cells were transfected with 180 ng monocistronic plasmid by using Lipofectamine 2000 transfection reagent, and medium was exchanged after 6 hours. Firefly luciferase activity was measured 48 hours posttransfection using the Luciferase Assay Sys-

tem (Promega), according to the manufacturer's protocol. RNA was extracted from cell lysate by using TRI reagent (Sigma-Aldrich), and Firefly mRNA levels were quantified by using qRT–PCR. Expression levels were normalized to the GAPDH housekeeping gene. The primers used for the reaction follow: TCGGTAAAGTTGTTC-CATTTTTTGAAG (Fluc Fw primer) and GGATT-GTTTACATAACCGGACATAATCATAG (Fluc Rv primer), GTCGGAGTCAACGGATTTGG (GAPDH Fw), AAAAGCAGCCCTGGTGACC (GAPDH Rv).

Computational analyses
Mapping deep-sequencing reads

To determine the identity of the oligo after sequencing, we made sure that the first 25 nt of the variable region were unique and distinguishable from other designed oligos in the library of 3 nt or more. In the cases where the first 25 nt of the designed oligo were similar to other oligos (less than 3 nt difference), a unique 8-mer barcode sequence was designed instead of the first 8 nt of the variable region. DNA was sequenced on HiSeq-2000 or NextSeq-500 sequencer. For cap-independent translation measurements, we obtained ~26 million reads for the full pEF1-mRFP-oligos-eGFP library (before sorting) so that 88% of the designed oligos (48,658 of 55,000) had coverage of ≥10 reads. We obtained ~4 million reads for all the 16 expression bins in biological replicate number 1 and ~7 million reads in biological replicate number 2. For promoter activity measurements, we obtained ~31 million reads of the full pEF1del-mRFP-oligos-eGFP library (before sorting), so that 92% of the designed oligos (50,559 of 55,000) had sequencing coverage of ≥10 reads. About 1 million reads were obtained for the eGFP⁺ population. For splicing measurements, we obtained ~1 million reads for each of the gDNA and cDNA samples. As reference sequence for mapping, we constructed *in silico* an "artificial library chromosome" by concatenating all the sequences of the 55,000 designed oligos with spacers of 50 N's. Single-end HiSeq or NextSeq reads in the length of 50 or 75 nt, respectively, were trimmed to 45 nt containing the common priming site and the unique 25 nt of the oligo's variable region. Trimmed reads were mapped to the artificial library chromosome using Novoalign aligner, and the number of reads for each designed oligo was counted in each sample.

Computing expression scores for cap-independent translation

Deep-sequencing reads from each expression bin and the full pEF1-mRFP-oligos-eGFP library were mapped to the unique 25 nt of the designed 55,000 oligos. For oligos that had ≥2 reads in at least two adjacent bins (representing independent PCR and sequencing), we computed the mean expression as the weighted average of eGFP expression bins, where the weight of each bin is the fraction of the oligo reads number in this bin of its total reads in all 16 bins. For oligos that were not detected in two adjacent bins and had >100 reads in the full

pEF1-mRFP-oligos-eGFP library (e.g., oligos that were represented in the library but had negative eGFP expression), we assigned a score of background eGFP expression as determined by the empty vector (pEF1-mRFP-eGFP). For oligos that had fewer than 100 reads in the full pEF1-mRFP-oligos-eGFP library, we assigned a NaN value.

Computing expression scores for promoter activity

For each of the 55,000 oligos, we computed the fraction of reads number in the eGFP$^+$ population out of the total number of reads in the full pEF1del-mRFP-oligos-eGFP library. Because the overall eGFP$^+$ cells were 2.5% of the population when sorting the library in the FACS, we set the activity threshold to the 97.5th percentile of the computed fractions of all oligos, which is 0.2. Oligos with computed fraction of ≥0.2 were considered as positive promoters and, therefore, removed from cap-independent translation analysis.

Assessment of splicing activity in the library

For each oligo, we computed the log2 ratio between the number of reads in the cDNA sample and the number of reads in the gDNA sample. To compute a threshold for oligos with splicing activity, we fitted a normal distribution based on the right side of the histogram (fig. S3B) and extracted the mean and standard deviation (SD). We set a threshold 1.5 SD from the mean of this histogram (fig. S3B). Oligos with computed log ratio of ≤−2.5 were considered as having splicing activity and, therefore, were removed from cap-independent translation analysis.

Testing the enrichment of k-mers with complementarity to the 18S rRNA in native cap-independnet sequences

Short k-mers 7 nt in length were extracted from the active region (nt 812 to 1233) of the 18S rRNA and the inactive regions (i.e, position on the 18S rRNA for which expression is lower than the activity threshold). For each k-mer, we compared the expression measurements of oligos that contain its sequence (up to one mismatch) and oligos that do not contain its sequences from all the native oligos in the library (n = 23,623) by Wilcoxon rank-sum test.

Statistical analyses

To assess the difference between two groups of values that are distributed normally (e.g., GC content, MFE, cDNA/gDNA ratio), we used Student's t test. In the case of fluorescence-based expression measurements of cap-independent translation and promoter activity for which a detection boundary exists and the data thus do not distribute normally, we performed a nonparameteric Wilcoxon rank-sum test. To show that the fraction of sequences with cap-independent translation or promoter activity is higher in one group compared with the other, we used a two-proportion z test (e.g., cap-independent activity of 3′UTRs versus the CDS of human transcripts and cap-independent activity of the polyprotein

region of uncapped versus capped [+]ssRNA viruses). To test for enrichment of a motif within a group of sequences [e.g., the enrichment of poly(U) element within regulatory sequences], we used Fisher's exact test. To test for colocalization between cap-independent translation sequences and the annotated mature proteins of [+]ssRNA viruses, we compared the distances between the start positions of mature proteins and the nearest translation elements to the distances obtained when the translation elements were randomly shuffled across the virus genome. When testing multiple hypotheses (e.g., multiple k-mers with complementary sequences to the 18S rRNA and multiple uncapped [+]ssRNA genomes), P values were corrected using the Benjamini-Hochberg procedure.

REFERENCES AND NOTES

1. N. T. Ingolia, S. Ghaemmaghami, J. R. Newman, J. S. Weissman, Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* 324, 218–223 (2009). pmid: 19213877

2. N. Sonenberg, A. G. Hinnebusch, Regulation of translation initiation in eukaryotes: Mechanisms and biological targets. *Cell* 136, 731–745 (2009). doi: 10.1016/j.cell.2009.01.042; pmid: 19239892

3. S. K. Jang 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). pmid: 2839690

4. J. Pelletier, N. Sonenberg, Internal initiation of translation of eukaryotic mRNA directed by a sequence derived from poliovirus RNA. *Nature* 334, 320–325 (1988). doi: 10.1038/334320a0; pmid: 2839775

5. J. A. den Boon, A. Diaz, P. Ahlquist, Cytoplasmic viral replication complexes. *Cell Host Microbe* 8, 77–85 (2010). doi: 10.1016/j.chom.2010.06.010; pmid: 20638644

6. G. Lozano et al., Local RNA flexibility perturbation of the IRES element induced by a novel ligand inhibits viral RNA translation. *RNA Biol.* 12, 555–568 (2015). doi: 10.1080/15476286.2015.1025190; pmid: 25775053

7. X. Du et al., Second cistron in CACNA1A gene encodes a transcription factor mediating cerebellar development and SCA6. *Cell* 154, 118–133 (2013). pmid: 23827678

8. S. Xue et al., RNA regulons in Hox 5′ UTRs confer ribosome specificity to gene regulation. *Nature* 517, 33–38 (2015). doi: 10.1038/nature14010; pmid: 25409156

9. A. B. Sachs, P. Sarnow, M. W. Hentze, Starting at the beginning, middle, and end: Translation initiation in eukaryotes. *Cell* 89, 831–838 (1997). doi: 10.1016/S0092-8674(00)80268-8; pmid: 9200601

10. L. Balvay, R. Soto Rifo, E. P. Ricci, D. Decimo, T. Ohlmann, Structural and functional diversity of viral IRESes. *Biochim. Biophys. Acta* 1789, 542–557 (2009). doi: 10.1016/j.bbagrm.2009.07.005; pmid: 19632368

11. J. Dresios, S. A. Chappell, W. Zhou, V. P. Mauro, An mRNA-rRNA base-pairing mechanism for translation initiation in eukaryotes. *Nat. Struct. Mol. Biol.* 13, 30–34 (2006). doi: 10.1038/nsmb1031; pmid: 16341227

12. S. R. Thompson, So you want to know if your message has an IRES? *Wiley Interdiscip. Rev. RNA* 3, 697–705 (2012). doi: 10.1002/wrna.1129; pmid: 22733589

13. B. P. Wellensiek et al., Genome-wide profiling of human cap-independent translation-enhancing elements. *Nat. Methods* 10, 747–750 (2013). doi: 10.1038/nmeth.2522; pmid: 23770754

14. Materials and methods are available as supplementary material on Science Online.

15. E. Sharon et al., Inferring gene regulatory logic from high-throughput measurements of thousands of systematically designed promoters. *Nat. Biotechnol.* 30, 521–530 (2012). doi: 10.1038/nbt.2205; pmid: 22609971

16. B. T. Baranick 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). doi: 10.1073/pnas.0710650105; pmid: 18326627

17. M. Mokrejs et al., IRESite—a tool for the examination of viral and cellular internal ribosome entry sites. *Nucleic Acids Res.*

38 (Database), D131–D136 (2010). doi: 10.1093/nar/gkp981; pmid: 19917642

18. J. D. Blais et al., Activating transcription factor 4 is translationally regulated by hypoxic stress. *Mol. Cell. Biol.* 24, 7469–7482 (2004). doi: 10.1128/MCB.24.17.7469-7482.2004; pmid: 15314157

19. M. Bushell et al., Polypyrimidine tract binding protein regulates IRES-mediated gene expression during apoptosis. *Mol. Cell* 23, 401–412 (2006). doi: 10.1016/j.molcel.2006.06.012; pmid: 16885029

20. R. Genolet, T. Araud, L. Maillard, P. Jaquier-Gubler, J. Curran, An approach to analyse the specific impact of rapamycin on mRNA-ribosome association. *BMC Med. Genomics* 1, 33 (2008). doi: 10.1186/1755-8794-1-33; pmid: 18673536

21. G. Johannes, M. S. Carter, M. B. Eisen, P. O. Brown, P. Sarnow, Identification of eukaryotic mRNAs that are translated at reduced cap binding complex eIF4F concentrations using a cDNA microarray. *Proc. Natl. Acad. Sci. U.S.A.* 96, 13118–13123 (1999). doi: 10.1073/pnas.96.23.13118; pmid: 10557283

22. T. Kawai, J. Fan, K. Mazan-Mamczarz, M. Gorospe, Global mRNA stabilization preferentially linked to translational repression during the endoplasmic reticulum stress response. *Mol. Cell. Biol.* 24, 6773–6787 (2004). doi: 10.1128/MCB.24.15.6773-6787.2004; pmid: 15254244

23. I. R. Powley et al., Translational reprogramming following UVB irradiation is mediated by DNA-PKcs and allows selective recruitment to the polysomes of mRNAs encoding DNA repair enzymes. *Genes Dev.* 23, 1207–1220 (2009). doi: 10.1101/gad.516509; pmid: 19451221

24. X. Qin, P. Sarnow, Preferential translation of internal ribosome entry site-containing mRNAs during the mitotic cycle in mammalian cells. *J. Biol. Chem.* 279, 13721–13728 (2004). doi: 10.1074/jbc.M312854200; pmid: 14739278

25. J. D. Thomas, G. J. Johannes, Identification of mRNAs that continue to associate with polysomes during hypoxia. *RNA* 13, 1116–1131 (2007). doi: 10.1261/rna.534807; pmid: 17488873

26. S. A. Chappell, G. M. Edelman, V. P. Mauro, A 9-nt segment of a cellular mRNA can function as an internal ribosome entry site (IRES) and when present in linked multiple copies greatly enhances IRES activity. *Proc. Natl. Acad. Sci. U.S.A.* 97, 1536–1541 (2000). doi: 10.1073/pnas.97.4.1536; pmid: 10677496

27. G. C. Owens, S. A. Chappell, V. P. Mauro, G. M. Edelman, Identification of two short internal ribosome entry sites selected from libraries of random oligonucleotides. *Proc. Natl. Acad. Sci. U.S.A.* 98, 1471–1476 (2001). doi: 10.1073/pnas.98.4.1471; pmid: 11171975

28. V. Zeenko, D. R. Gallie, Cap-independent translation of tobacco etch virus is conferred by an RNA pseudoknot in the 5′-leader. *J. Biol. Chem.* 280, 26813–26824 (2005). doi: 10.1074/jbc.M503576200; pmid: 15911616

29. R. Nicholson, J. Pelletier, S. Y. Le, N. Sonenberg, Structural and functional analysis of the ribosome landing pad of poliovirus type 2: In vivo translation studies. *J. Virol.* 65, 5886–5894 (1991). pmid: 1656077

30. S. A. Chappell, V. P. Mauro, The internal ribosome entry site (IRES) contained within the RNA-binding motif protein 3 (Rbm3) mRNA is composed of functionally distinct elements. *J. Biol. Chem.* 278, 33793–33800 (2003). doi: 10.1074/jbc.M303495200; pmid: 12824175

31. J. W. Hershey, N. Sonenberg, M. B. Mathews, Principles of translational control: An overview. *Cold Spring Harb. Perspect. Biol.* 4, a011528 (2012). doi: 10.1101/cshperspect.a011528; pmid: 23209153

32. S. Pyronnet, L. Pradayrol, N. Sonenberg, A cell cycle-dependent internal ribosome entry site. *Mol. Cell* 5, 607–616 (2000). doi: 10.1016/S1097-2765(00)80240-3; pmid: 10882097

33. S. D. Baird, S. M. Lewis, M. Turcotte, M. Holcik, A search for structurally similar cellular internal ribosome entry sites. *Nucleic Acids Res.* 35, 4664–4677 (2007). doi: 10.1093/nar/gkm483; pmid: 17591613

34. E. Sharon, S. Lubliner, E. Segal, A feature-based approach to modeling protein-DNA interactions. *PLOS Comput. Biol.* 4, e1000154 (2008). doi: 10.1371/journal.pcbi.1000154; pmid: 18725950

35. A. V. Pisarev, V. G. Kolupaeva, M. M. Yusupov, C. U. Hellen, T. V. Pestova, Ribosomal position and contacts of mRNA in eukaryotic translation initiation complexes. *EMBO J.* 27, 1609–1621 (2008). doi: 10.1038/emboj.2008.90; pmid: 18464793

36. D. Matsuda, V. P. Mauro, 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). doi: 10.1073/pnas.1413472111; pmid: 25313046

37. A. A. Malygin, O. A. Kossinova, I. N. Shatsky, G. G. Karpova, 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). doi: 10.1093/nar/gkt632; pmid: 23873958

38. Z. Meng, N. L. Jackson, O. D. Shcherbakov, H. Choi, S. W. Blume, The human IGF1R IRES likely operates through a Shine-Dalgarno-like interaction with the G961 loop (E-site) of the 18S rRNA and is kinetically modulated by a naturally polymorphic polyU loop. *J. Cell. Biochem.* **110**, 531–544 (2010). pmid: 20432247

39. S. Cornelis *et al.*, Identification and characterization of a novel cell cycle-regulated internal ribosome entry site. *Mol. Cell* **5**, 597–605 (2000). doi: 10.1016/S1097-2765(00)80239-7; pmid: 10882096

40. M. M. Candeias *et al.*, Expression of p53 and p53/47 are controlled by alternative mechanisms of messenger RNA translation initiation. *Oncogene* **25**, 6936–6947 (2006). doi: 10.1038/sj.onc.1209996; pmid: 16983332

41. C. H. Herbreteau *et al.*, HIV-2 genomic RNA contains a novel type of IRES located downstream of its initiation codon. *Nat. Struct. Mol. Biol.* **12**, 1001–1007 (2005). doi: 10.1038/nsmb1011; pmid: 16244661

42. S. Bradley, S. Narayanan, M. Rosbash, NAT1/DAP5/p97 and atypical translational control in the *Drosophila* Circadian Oscillator. *Genetics* **192**, 943–957 (2012). doi: 10.1534/genetics.112.143248; pmid: 22904033

43. Y. F. Chang, J. S. Imam, M. F. Wilkinson, The nonsense-mediated decay RNA surveillance pathway. *Annu. Rev. Biochem.* **76**, 51–74 (2007). doi: 10.1146/annurev.biochem.76.050106.093909; pmid: 17352659

44. A. Jakubiec, I. Jupin, Regulation of positive-strand RNA virus replication: The emerging role of phosphorylation. *Virus Res.* **129**, 73–79 (2007). doi: 10.1016/j.virusres.2007.07.012; pmid: 17764774

45. L. A. Ball, Replication strategies of RNA viruses, in *Fields Virology*, volume 1, Bernard N. Fields, Peter M. Howley, Diane E. Griffin, Robert A. Lamb, Malcolm A. Martin, Bernard Roizman, Stephen E. Straus, David M. Knipe, Eds. (Lippincott Williams & Wilkins, Philadelphia, ed. 4, 2001), chap. 5.

46. K. Treder, E. L. Pettit Kneller, E. M. Allen, Z. Wang, K. S. Browning, W. A. Miller, The 3′ cap-independent translation element of Barley yellow dwarf virus binds eIF4F via the eIF4G subunit to initiate translation. *RNA* **14**, 134–147 (2008). doi: 10.1261/rna.777308; pmid: 18025255

47. B. M. Gazo, P. Murphy, J. R. Gatchel, K. S. Browning, A novel interaction of cap-binding protein complexes eukaryotic initiation factor (eIF)4F and eIF(iso)4F with a region in the 3′-untranslated region of satellite tobacco necrosis virus. *J. Biol. Chem.* **279**, 13584–13592 (2004). doi: 10.1074/jbc.M311361200; pmid: 14729906

48. I. N. Shatsky, S. E. Dmitriev, I. M. Terenin, D. E. Andreev, Cap- and IRES-independent scanning mechanism of translation initiation as an alternative to the concept of cellular IRESs. *Mol. Cells* **30**, 285–293 (2010). doi: 10.1007/s10059-010-0149-1; pmid: 21052925

49. K. Y. Paek *et al.*, Translation initiation mediated by RNA looping. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 1041–1046 (2015). doi: 10.1073/pnas.1416883112; pmid: 25583496

50. T. P. Miettinen, M. Björklund, Modified ribosome profiling reveals high abundance of ribosome protected mRNA fragments derived from 3′ untranslated regions. *Nucleic Acids Res.* **43**, 1019–1034 (2015). doi: 10.1093/nar/gku1310; pmid: 25550424

51. N. R. Guydosh, R. Green, Dom34 rescues ribosomes in 3′ untranslated regions. *Cell* **156**, 950–962 (2014). doi: 10.1016/j.cell.2014.02.006; pmid: 24581494

52. M. A. Cleary *et al.*, Production of complex nucleic acid libraries using highly parallel in situ oligonucleotide synthesis. *Nat. Methods* **1**, 241–248 (2004). doi: 10.1038/nmeth724; pmid: 15782200

53. E. M. LeProust *et al.*, Synthesis of high-quality libraries of long (150mer) oligonucleotides by a novel depurination controlled process. *Nucleic Acids Res.* **38**, 2522–2540 (2010). doi: 10.1093/nar/gkq163; pmid: 20308161

54. A. Sigal *et al.*, Dynamic proteomics in individual human cells uncovers widespread cell-cycle dependence of nuclear proteins. *Nat. Methods* **3**, 525–531 (2006). doi: 10.1038/nmeth892; pmid: 16791210

55. K. Kazadi *et al.*, Genomic determinants of the efficiency of internal ribosomal entry sites of viral and cellular origin. *Nucleic Acids Res.* **36**, 6918–6925 (2008). doi: 10.1093/nar/gkn812; pmid: 18978018

56. R. Blecher-Gonen *et al.*, High-throughput chromatin immunoprecipitation for genome-wide mapping of in vivo protein-DNA interactions and epigenomic states. *Nat. Protoc.* **8**, 539–554 (2013). doi: 10.1038/nprot.2013.023; pmid: 23429716

57. T. Unger, Y. Jacobovitch, A. Dantes, R. Bernheim, Y. Peleg, Applications of the Restriction Free (RF) cloning procedure for molecular manipulations and protein expression. *J. Struct. Biol.* **172**, 34–44 (2010). doi: 10.1016/j.jsb.2010.06.016; pmid: 20600952

58. A. Erijman, A. Dantes, R. Bernheim, J. M. Shifman, Y. Peleg, Transfer-PCR (TPCR): A highway for DNA cloning and protein engineering. *J. Struct. Biol.* **175**, 171–177 (2011). doi: 10.1016/j.jsb.2011.04.005; pmid: 21515384

59. S. Weingarten-Gabbay *et al.*, The translation initiation factor DAP5 promotes IRES-driven translation of p53 mRNA. *Oncogene* **33**, 611–618 (2014). doi: 10.1038/onc.2012.626; pmid: 23318444

SUPPLEMENTARY MATERIALS

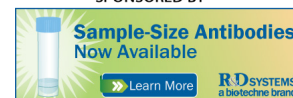www.sciencemag.org/content/351/6270/aad4939/suppl/DC1
Materials and Methods
Figs. S1 to S15
Tables S2 and S9
Reference (*60*)
Data Tables S1 and S3 to S8

20 September 2015; accepted 11 November 2015
10.1126/science.aad4939

Science

AAAS

Systematic discovery of cap-independent translation sequences in human and viral genomes
Shira Weingarten-Gabbay *et al.*
Science **351**, (2016);
DOI: 10.1126/science.aad4939

This copy is for your personal, non-commercial use only.