



DATABASE

IRESbase: A Comprehensive Database of Experimentally Validated Internal Ribosome Entry Sites



Jian Zhao^{1,#}, Yan Li^{2,3,#}, Cong Wang^{1,#}, Haotian Zhang^{2,#}, Hao Zhang², Bin Jiang⁴, Xuejiang Guo^{2,*}, Xiaofeng Song^{1,*}

¹ Department of Biomedical Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China

² State Key Laboratory of Reproductive Medicine, Department of Histology and Embryology, Nanjing Medical University, Nanjing 211166, China

³ Center of Pathology and Clinical Laboratory, Sir Run Run Hospital, Nanjing Medical University, Nanjing 211166, China

⁴ College of Automation Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China

Received 6 June 2019; revised 24 March 2020; accepted 25 March 2020

Available online 6 June 2020

Handled by Domenica D'Elia

KEYWORDS

IRES;
Circular RNA;
Long non-coding RNA;
Eukaryotic IRES;
Viral IRES

Abstract Internal ribosome entry sites (**IRESs**) are functional RNA elements that can directly recruit ribosomes to an internal position of the mRNA in a cap-independent manner to initiate translation. Recently, IRES elements have attracted much attention for their critical roles in various processes including translation initiation of a new type of RNA, **circular RNA** (circRNA), with no 5' cap to support classical cap-dependent translation. Thus, an integrative data resource of IRES elements with experimental evidence will be useful for further studies. In this study, we present IRESbase, a comprehensive database of IRESs, by curating the experimentally validated functional minimal IRES elements from literature and annotating their host linear and circular RNAs. The current version of IRESbase contains 1328 IRESs, including 774 **eukaryotic IRESs** and 554 **viral IRESs** from 11 eukaryotic organisms and 198 viruses, respectively. As IRESbase collects only IRES of minimal length with functional evidence, the median length of IRESs in IRESbase is 174 nucleotides. By mapping IRESs to human circRNAs and **long non-coding RNAs** (lncRNAs), 2191 circRNAs and 168 lncRNAs were found to contain at least one entire or partial IRES sequence. IRESbase is available at <http://reprod.njmu.edu.cn/cgi-bin/iressbase/index.php>.

* Corresponding authors.

E-mail: guo_xuejiang@njmu.edu.cn (Guo X), xfsong@nuaa.edu.cn (Song X).

Equal contribution.

Peer review under responsibility of Beijing Institute of Genomics, Chinese Academy of Sciences and Genetics Society of China.

<https://doi.org/10.1016/j.gpb.2020.03.001>

1672-0229 © 2020 The Authors. Published by Elsevier B.V. and Science Press on behalf of Beijing Institute of Genomics, Chinese Academy of Sciences and Genetics Society of China.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Introduction

Translation initiation is a crucial and highly regulated step during protein synthesis in eukaryotes. Generally, eukaryotic mRNAs recruit ribosomes to initiate translation by a canonical cap-dependent scanning mechanism, which requires a methylated cap structure at its 5' end and is mediated by the eukaryotic translation initiation factors (eIFs). Instead of the cap-dependent manner, a subset of eukaryotic mRNAs can initiate an internal translation through specialized sequences, called internal ribosome entry sites (IRESs), which can directly recruit 40S ribosomal subunits to initiate translation.

The IRES elements were first discovered about 30 years ago in picornaviruses [1], followed by the discovery of novel IRES in many other pathogenic viruses, *e.g.*, encephalomyocarditis virus, foot-and-mouth disease virus, and human immunodeficiency virus [2–4]. The IRES elements ensure the efficient viral translation, when canonical eIFs necessary for mRNA recruitment are inhibited in infected cells. Besides the viral mRNAs, IRESs were also found in a subset of eukaryotic mRNAs, *e.g.*, *XIAP*, *TP53*, *EIF4G2*, and *APAF1* [5–8]. The IRESs in eukaryotic mRNAs allow their translation to continue under many physiologically stressful conditions when the cap-dependent translation is suppressed. In addition, for those eukaryotic IRES-containing mRNAs with long and highly structured 5'-UTRs (incompatible with the cap-dependent scanning mechanism), the IRES elements ensure their efficient translation under normal physiological conditions even when cap-dependent translation is fully active [9]. The increasing number of discovered eukaryotic IRESs suggests that the IRES-driven initiation mechanism accounts for a significant proportion of eukaryotic mRNA translation. Although the studies of IRESs have been limited, eukaryotic mRNAs bearing IRESs seem to play important roles in diverse processes, including stress response, cell survival, apoptosis, mitosis, tumorigenesis, and progression [10–14].

Two IRES databases, IRESdb and IRESite, were built in 2002 and 2005, respectively [15,16]. In addition, the Rfam database collected IRESs as a type of *cis*-regulatory RNA element [17]. However, these databases only include IRESs in mRNAs. Recently, the IRES elements have also been found in the circular RNAs (circRNAs) and long non-coding RNAs (lncRNAs) [18]. The circRNAs, *circ-FBXW7*, *circ-ZNF609*, *circ-SHPRH*, *circPINTexon2*, and *circ β -catenin*, were found to be translated into polypeptides, mediated by IRES elements [19–23]. The translation of two small open reading frames (ORFs; 120 nucleotides and 141 nucleotides long) within the lncRNA *meloe* was achieved through an IRES-dependent mechanism [24]. Identification of the translation of circRNAs and lncRNAs driven by IRES elements has attracted more attention recently.

The IRESdb database only has 30 viral IRESs and 50 eukaryotic IRESs, and the IRESite database contains 125 IRESs in total from 43 viruses and 70 eukaryotic mRNAs. The Rfam database builds 32 RNA families with about 11 viral IRESs and 21 eukaryotic IRESs. In this study, by manual curation, we developed a novel non-redundant public database and named it IRESbase. This database contains updated experimentally validated IRESs, including 554 viral IRESs, 691 human IRESs, and 83 IRESs from other eukaryotic organisms.

In order to facilitate development of new IRES-dependent functional regulation studies, IRESbase collected rich informa-

tion about human IRESs. This database includes information regarding the genomic positions, sequence conservation, single nucleotide polymorphisms (SNPs), nucleotide modifications, targeting microRNAs (miRNA), host gene, host transcripts (mRNA, circRNA, and lncRNA), GO terms (biological process, molecular function, and cellular component), KEGG pathway annotations, and validation assay information. In the database, 2191 circRNAs and 168 lncRNAs contain at least one entire or partial IRES element. Among these transcripts, 1012 circRNAs contain an ORF longer than 300 nucleotides, and all the 168 lncRNAs contain at least two small ORFs (ranging from 60 to 300 nucleotides in length). Our IRESbase database allows users to search, browse, download, and submit experimentally validated IRESs.

Data collection and processing

Curation of IRES elements

To construct IRESbase with high quality information, the experimentally validated IRES sequences were manually searched from literature published before Oct. 14, 2019 using PubMed with ten keywords in the titles or abstracts (Figure S1). These ten keywords are “IRES,” “IRESs,” “internal ribosome entry site,” “internal ribosome entry sites,” “internal ribosomal entry site,” “internal ribosomal entry sites,” “internal ribosome entry segment,” “internal ribosomal entry segment,” “internal ribosome entry sequence,” and “internal ribosomal entry sequence”. We manually checked the abstracts and the corresponding full texts of the retrieved literature, to see whether a novel IRES is discovered and validated experimentally. For the literature concerning previously discovered IRESs, we further retrieved the original publications through their citations. For published reviews and databases of IRESs, we checked all the cited references of their collected IRESs, and located the original publications that discovered and validated these IRESs. The sequences of experimentally validated IRESs together with experiment information essential for the validation were extracted. The validation experiment information includes bicistronic backbone, positive and negative controls, second cistron expression, tested cell types, methods to analyze RNA structure, and *in vitro* translation experiments. Instead of collecting all the truncated 5' UTR sequences experimentally validated to contain IRES activity as IRESite did, only the shortest one was preserved in IRESbase to reduce data redundancy (Figure S2). In total, from 236 literature, 1328 experimentally validated IRES sequences, including 554 viral IRESs, 691 human IRESs, and 83 IRESs from other eukaryotic organisms (mouse, fly, yeast, *etc.*), were collected.

Collection of IRES sequences

As the available IRES sequence-related information varied in different literature, we manually extracted the IRES sequence using different methods (Figure S1). These methods can be roughly classified into three categories: (1) extracting IRES sequence directly from the figure of its two-dimensional (2D) structure; (2) extracting IRES sequence indirectly through the information of its host transcript and position; and (3) extracting IRES sequence by using the NCBI BLAST tool with the forward and reverse primers reported in the literature.

Note that if more than one of the overlapping sequences were experimentally validated to contain IRES activity in literature, only the minimal functional sequence in them was selected (Figure S2). In addition, the IRES sequences derived from different literature were compared pair wise, and if the sequence identity was over 90%, only the shortest one was selected.

Annotation of IRESs

We searched and collected the verified 2D structures of IRESs (or IRES host 5' UTRs) from literature. The 3D structures of IRESs were manually searched and collected from the PDB, PDBj, and PDBe databases [25–27]. The sequences in the collected IRES 2D structures were manually extracted from the corresponding figures, and then the IRES-matched regions were calculated by sequence alignment. In addition, for all IRESs, the 2D structure and the minimal free energy (MFE) were also calculated using RNAfold of Vienna RNA package [28].

For human IRESs, genomic position, sequence conservation, SNPs, RNA modification sites, potential host circular RNA, and miRNA–IRES interaction data were provided. As shown in Figure S3, the IRES genomic positions were obtained by mapping IRES sequences to their host transcripts, and then chromosomal positions were extracted from NCBI transcript annotation files (GRCh37.p13 and GRCh38.p10) [29]. Based on the genomic positions, the phyloP scores (hg38 phyloP100way) were extracted from the UCSC database [30]. The NCBI dbSNP database (150 release) was searched for the common SNPs and SNPs within the IRES sequences were flagged [31]. The RMBase v2.0 database was searched for RNA modification sites within the human IRES regions, including N^6 -methyladenosine (m^6A) sites, N^1 -methyladenosine (m^1A) sites, pseudouridine (ψ) sites, and 2'-*O*-methylation (2'-*O*-Me) sites [32]. The potential miRNA–IRES interactions were predicted by miRanda (version: aug2010) with a score threshold of 150 [33].

Annotation of IRES host genes and transcripts

The host genes and transcripts of IRESs were collected from the literature, if available. For IRES sequences without detailed host gene or transcript information in the literature, nucleotide BLAST in NCBI was performed to search for their host genes and transcripts with a sequence identity of at least 95% [34]. For viral IRESs, the closest gene with coding sequence (CDS) downstream of an IRES element was regarded as its host gene. For human IRESs, besides mRNAs, we also aligned IRESs to lncRNAs and circRNAs, and only sequences containing an IRES fragment of at least 30 nucleotides were considered as candidate IRES host transcripts. The cutoff value (30 nucleotides) was defined according to the minimum length of known IRES elements.

For each IRES entry, mRNA and lncRNA information was extracted from the NCBI nucleotide database, and the circRNA information was extracted from circBase, circRNADb, and circAtlas [35–37]. The ORF (< 300 nucleotides) and its corresponding peptide sequence were predicted for each lncRNA, and the ORF (> 300 nucleotides) and its corresponding protein sequence were predicted for each circRNA. The coding potential score of ORFs was calculated by CPAT [38]. The sequence conservation of ORFs in lncRNA was assessed and indicated using

the mean phyloP score [30]. Gene level functional information, *i.e.*, OMIM and Gene Ontology (GO) terms of biological process, molecular function, and cellular component, was extracted from the NCBI Gene database. The pathway information was retrieved from the KEGG PATHWAY database [39].

Database construction

The IRESbase database was constructed by the assembly of all curated minimal IRES sequences and structure information, related transcript and gene information, as well as annotations. IRESbase was configured in a typical XAMPP (X-Windows + Apache + MySQL + PHP + Perl) integrated environment.

Implementation and results

Database content

IRESbase is intended to be a comprehensive database that includes both eukaryotic and viral IRESs with experimental evidence and covers not only the information on IRES sequences and structures, but also the information on IRES host transcripts and genes. The database currently contains a total of 1328 IRES entries, including 774 eukaryotic IRES entries and 554 viral IRES entries. Each eukaryotic IRES entry contains three basic types of records – IRES, gene, and transcript records if available, whereas each viral IRES entry contains the IRES and transcript records. In total, 1328 IRES records, 725 eukaryotic gene records, 3633 eukaryotic transcript records (1274 mRNAs, 168 lncRNAs, and 2191 circRNAs), and 588 viral transcript records are included in IRESbase. The data sources and the structure of IRESbase are shown in Figure 1.

The types of data available varied across different organisms. Human IRES records have the largest number of attributes (32 in total, if available); the IRES records for other eukaryotic organisms and viruses have 20 and 24 attributes, respectively (Figure 1). All IRES records have the following 18 attributes, including IRESbase ID, organism, sequence, length, host gene ID and symbol, validation assay information, 2D structure, MFE, PubMed ID, and reference. The other attributes for human records are host gene synonym, host transcript ID and alignment result, genomic positions (location and strand), sequence conservation, SNP, modification sites, and miRNA–IRES interaction. Host gene synonym, host transcript ID and alignment result are also contained in the IRES records for other eukaryotic organisms. The viral IRES record contains six other attributes including viral genome ID (*e.g.*, NC_006273), genome shape, virus type, genomic position, CDS of regulation, and 3D structure.

As shown in Figure 1, the eukaryotic gene records commonly have nine attributes, which are gene ID, gene symbol, gene synonym, gene note, GO biological process, GO molecular function, GO cell component, KEGG pathway, and IRESbase ID. For human, besides the nine attributes mentioned above, three other attributes (OMIM ID, OMIM phenotype, and HGNC ID) are included. The eukaryotic transcript records for IRES host mRNAs, lncRNAs, and circRNAs contain 12, 13, and 11 attributes, respectively. The 12 attributes of eukaryotic mRNA records are transcript ID, name, version, sequence, length, gene symbol, IRES alignment result, CDS

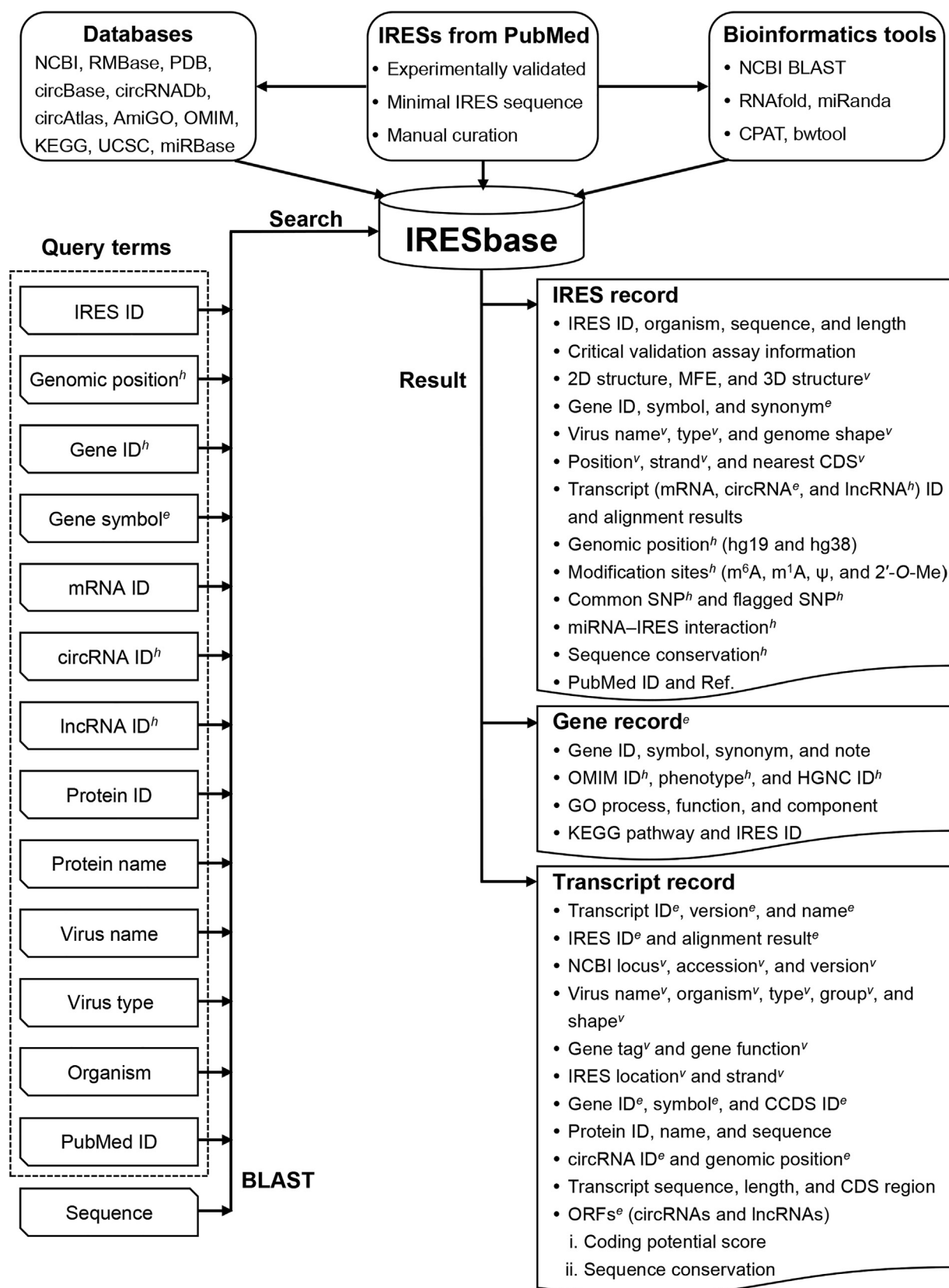


Figure 1 Data sources and the IRESbase structure

For query and record terms, information exclusively available for eukaryotes, humans, and viruses is indicated with “*e*”, “*h*”, and “*v*” in superscript, respectively; while the remaining terms without superscripts are commonly used for different organisms.

region, protein ID, name, sequence, and CCDS ID. For lncRNA records, besides the first seven attributes in the mRNA record, the other six attributes are gene ID, ORF, ORF's coding potential score, ORF's sequence conservation, predicted protein sequence, and length. For circRNA records, the 11 attributes are circRNA ID (circBase, circRNADb, and circAtlas ID), genomic location, strand, gene symbol, best transcript, IRES alignment result, ORF, coding potential score, predicted protein sequence, protein length, and reference. The viral transcript records have 21 attributes, including NCBI locus, accession, version, virus name, organism, type, group, transcript sequence, length, shape, gene ID, name, tag, function, CDS region, protein ID, name, sequence, IRES-base ID, IRES location, and strand.

Database interface

Simple, advanced, and BLAST search

IRESbase was developed in a user-friendly mode, providing a search engine to find the IRESs of interest. The homepage (Figure S4) offers a search box for querying the IRESbase by IRES ID, gene ID, transcript ID, protein ID, PubMed ID, chromosomal location, organism name, gene name, protein name, or a substring of any of the aforementioned keywords. If a keyword (e.g., *FUBP1* or HCMV) is added into the search bar, the results will be shown in a tabular format. By clicking on the IRES ID (e.g., *hsa_ires_00089.1* or *vir_ires_00484.1*), the detailed information is shown in three parts: the IRES information (Figure 2), its host gene information, and its host transcript information (Figure 3). Furthermore, terms in the page are cross-linked to several external databases including NCBI, UCSC, AmiGO, KEGG, HGNC, RMBase, circBase, circRNADb, and circAtlas [30–32,35–37,39–41].

In addition, IRESbase also provides three advanced search options for users, including (1) BLAST search, (2) Advanced search, and (3) Chromosomal location.

- (1) BLAST search: this option was designed to find if there is a similar IRES in the RNA sequence of interest. It provides four IRES sequence datasets (e.g., human, other eukaryotic organisms, virus, and all organisms) for users. Users can input more than one query sequence in FASTA format. The search results are summarized in a table containing job title, IRES ID, score, expect, identity, gaps, and strand (Figure 4). By clicking the 'Job title,' the alignment detail is shown at the nucleotide level.
- (2) Advanced search: in this option, users can use one or more query fields together to retrieve the IRESs of interest. Different query fields are provided for users to search IRESs of different organisms, including virus, human, and other eukaryotic organisms (Figure 1).
- (3) Chromosomal location: this option was designed for users to find if there is a verified IRES located in the circRNAs of interest through the input of a specific chromosomal region in humans (hg19 or hg38).

Data browse, download, and submission

Instead of searching for a specific IRES, all entries of IRESbase were grouped by organism, GO & KEGG, chromosome, and transcript type. Users can browse the IRES group of

interest by clicking the count on the left of the corresponding page. Unlike other existing IRES databases, IRESbase supports users to download all the IRES records together in a simple tab-delimited format or to download each IRES record separately in a standard PageTab (page layout tabulation) format. The PageTab format file only contains the basic IRES information and its validation assay information. In addition, users can also select the IRES attributes of interest to download in an Excel-compatible format.

IRESbase allows and encourages users to submit novel experimentally validated IRESs. For the submission, users should provide the basic IRES information (e.g., IRES sequence, organism, host gene ID, host gene name, and host transcript ID) and validation assay information, including bicistronic backbone, positive and negative controls, second cistron expression, tested cell types, methods to analyze RNA structure, *in vitro* translation experiments, and PubMed ID. In addition, users can also submit a novel verified IRES by emailing a PageTab format file, in which all the essential information should be assembled. The submitted record will be included in IRESbase after manual check by our curators.

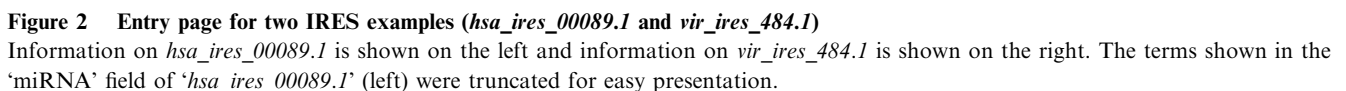
Comparison with existing IRES databases

The current version of IRESbase contains the largest collection of experimentally validated IRESs in 11 eukaryotic organisms and 198 viruses. When compared with the existing IRES databases, IRESbase contains the largest number of eukaryotic IRESs and viral IRESs, including 1328 IRESs, while Rfam, IRESdb, and IRESite only contains about 32, 80, and 135 IRESs, respectively (Figure 5A). In particular, IRESbase is the first database containing IRESs discovered in circRNAs and lncRNAs (Figure 5C). Moreover, IRESbase has minimal data redundancy. Unlike the IRESite database which collects all the truncated overlapping 5' UTR sequences validated to contain IRES activity, IRESbase only stores the shortest functional sequences as the core IRES region. The median length of IRESs in IRESbase is 174 nucleotides, which is much shorter than that in IRESite (442 nucleotides; Figure 5B).

In addition, IRESbase provides richer information and more useful functions than other IRES databases. Besides the experimentally verified 2D structure of 56 IRESs, IRESbase also collected the 3D structure of 12 IRESs from literature. Moreover, IRESbase predicted 2D structure for all of the IRESs with Vienna RNA package [28]. IRESbase also provides more annotations, including genomic positions, sequence conservation, SNPs, nucleotide modifications, and targeting miRNAs for human IRESs. And apart from the IRES host transcripts collected from literature, IRESbase also provides many other potential IRES host transcripts (mRNAs, lncRNAs, and circRNAs) predicted by sequence identity. Compared with existing databases, IRESbase provides more query fields (e.g., genomic location and gene synonym) and search methods (fuzzy, advanced, and BLAST search). Additionally, the entire IRESbase dataset or the matched specific terms can be easily batch downloaded from IRESbase.

Data statistics and analyses

In addition to 5' UTRs, IRESs are also found in the CDS regions [42]. In the 691 human IRESs, 60% IRESs are located



the gene set enrichment analysis was performed for the human IRES host genes using Metascape (www.metascape.org) [43]. As shown in Figure 5E, the most significant term is

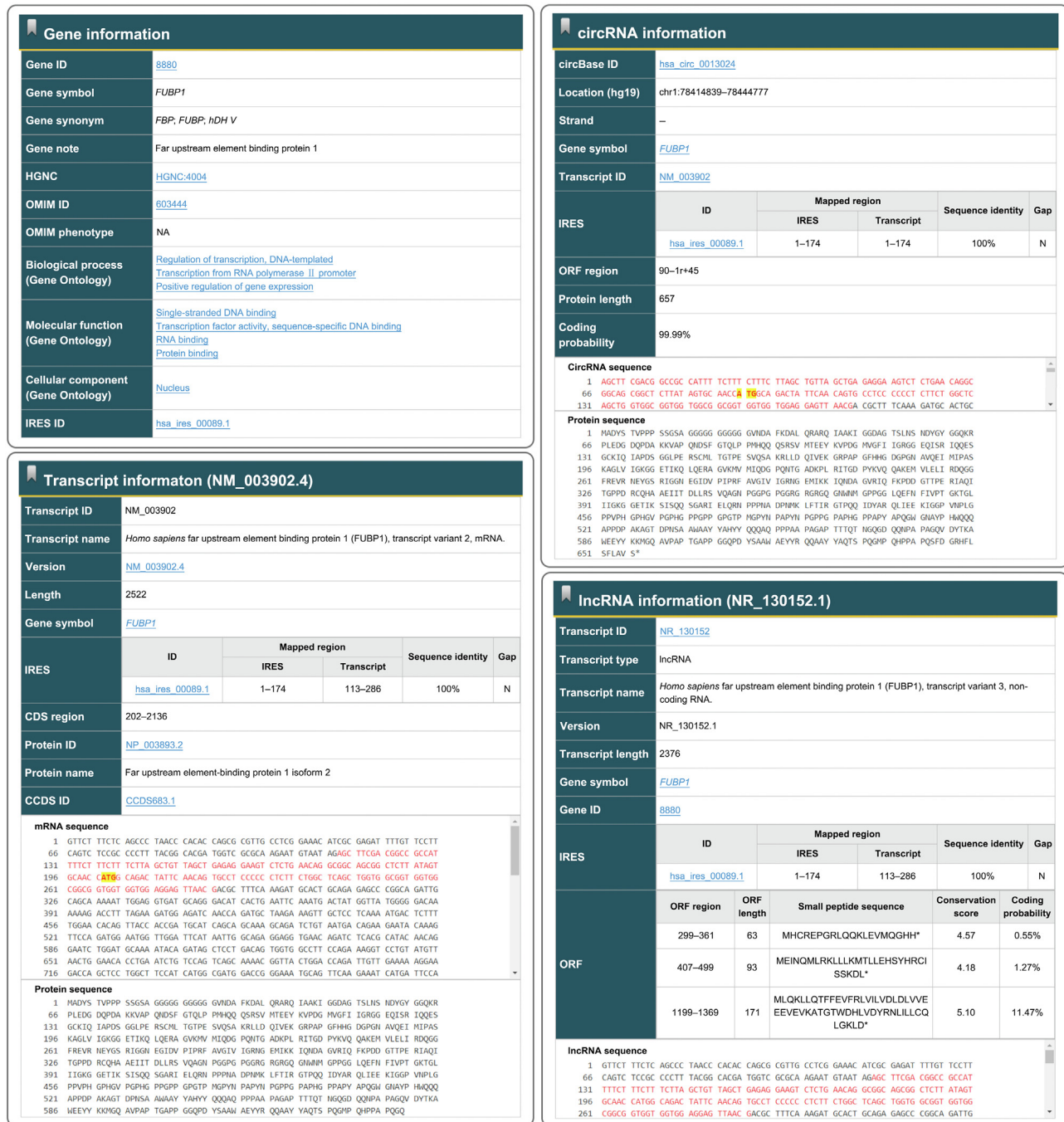


Figure 3 Gene and transcript records for a human IRES entry (*hsa_ires_00089.1*)

The IRES region is marked in red and the boundary codons of the CDS region are highlighted with yellow.

‘hsa05200: Pathways in cancer.’ The enriched terms indicate that human IRESs may play important roles in cancer, cellular stress response, virus infection, and cell signaling.

As early as in 1995, IRESs in mRNAs were proved to be effective in initiating the protein synthesis of artificial circRNAs [44]. In recent years, IRESs in circRNAs are also demonstrated to initiate the translation of linear transcripts (e.g., bicistronic transcripts) [19–23]. These studies indicate that the activity of IRES is not RNA type-specific and IRES can play roles in any kinds of host RNAs. Currently, only nine

circRNAs and one lncRNA have been found to be translated via IRES. Therefore, in order to identify more potential coding circRNAs and lncRNAs, mRNA-derived IRESs were mapped to human circRNAs and lncRNAs by using the pairwise sequence alignment. The result shows that 371 circRNAs and 83 lncRNAs contain full IRES sequences; 1777 circRNAs and 77 lncRNAs contain partial IRES sequence (at least 30 nucleotides). Furthermore, circRNA- and lncRNA-derived IRESs were also mapped to mRNAs. Consequently, eight novel mRNAs are discovered to contain IRESs.

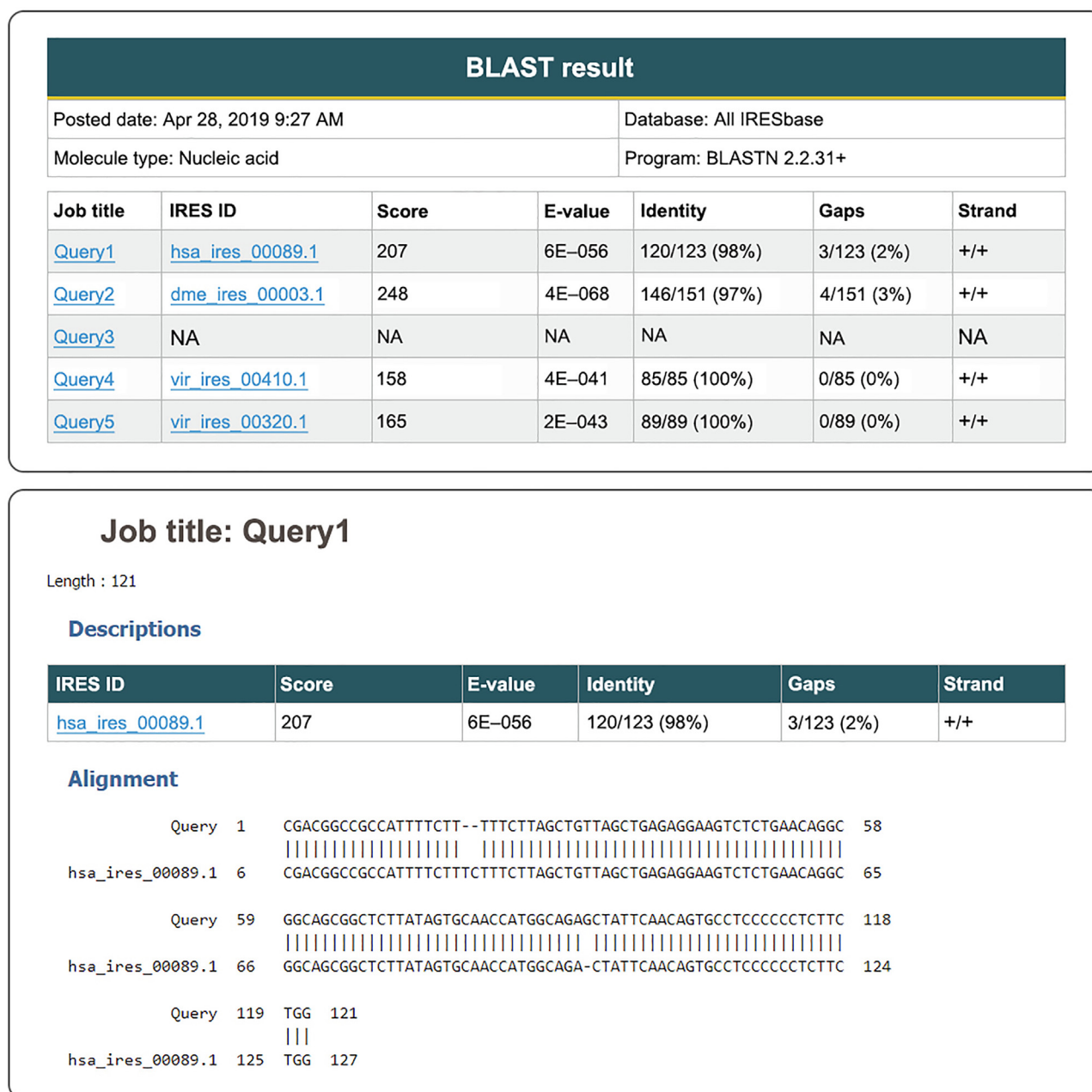


Figure 4 Result page of BLAST search exemplified with *hsa_ires_00089.1*

In order to investigate the potential coding circRNAs and lncRNAs, ORFs were predicted for the circRNAs and lncRNAs containing a full or partial IRES sequence. Of the circRNAs with an entire mRNA-derived IRES element, 261 contain an ORF (> 300 nucleotides); of the circRNAs with a partial mRNA-derived IRES element, 722 contain an ORF (> 300 nucleotides). All of lncRNAs were predicted to contain small ORFs (longer than 60 nucleotides and shorter than 300 nucleotides). And in 61 lncRNAs, the IRESs were found to be located between two small ORFs, which indicates that these lncRNAs may encode multiple small peptides. In addition, the coding potential score was calculated for the circRNAs and lncRNAs, and the results suggest that 779 circRNAs and three lncRNAs are able to encode proteins and small peptides, respectively.

Discussion and perspectives

In recent years, as a functional RNA element which recruits ribosome to initiate translation in a cap-independent manner, IRES has been shown to mediate circRNA and lncRNA translation [18,24,45]. Although only a few peptides encoded by circRNAs have been discovered, these peptides play important roles in various biological processes [19–23]. Besides the presence of ORF, coding ability of circRNAs further depends on the presence of IRES element. However, so far only a few IRESs are available in existing IRES databases, and the translation initiation mechanism via IRES is still elusive, especially for cellular IRESs [46]. Therefore, there is an urgent need for a comprehensive curated dataset of all experimentally validated

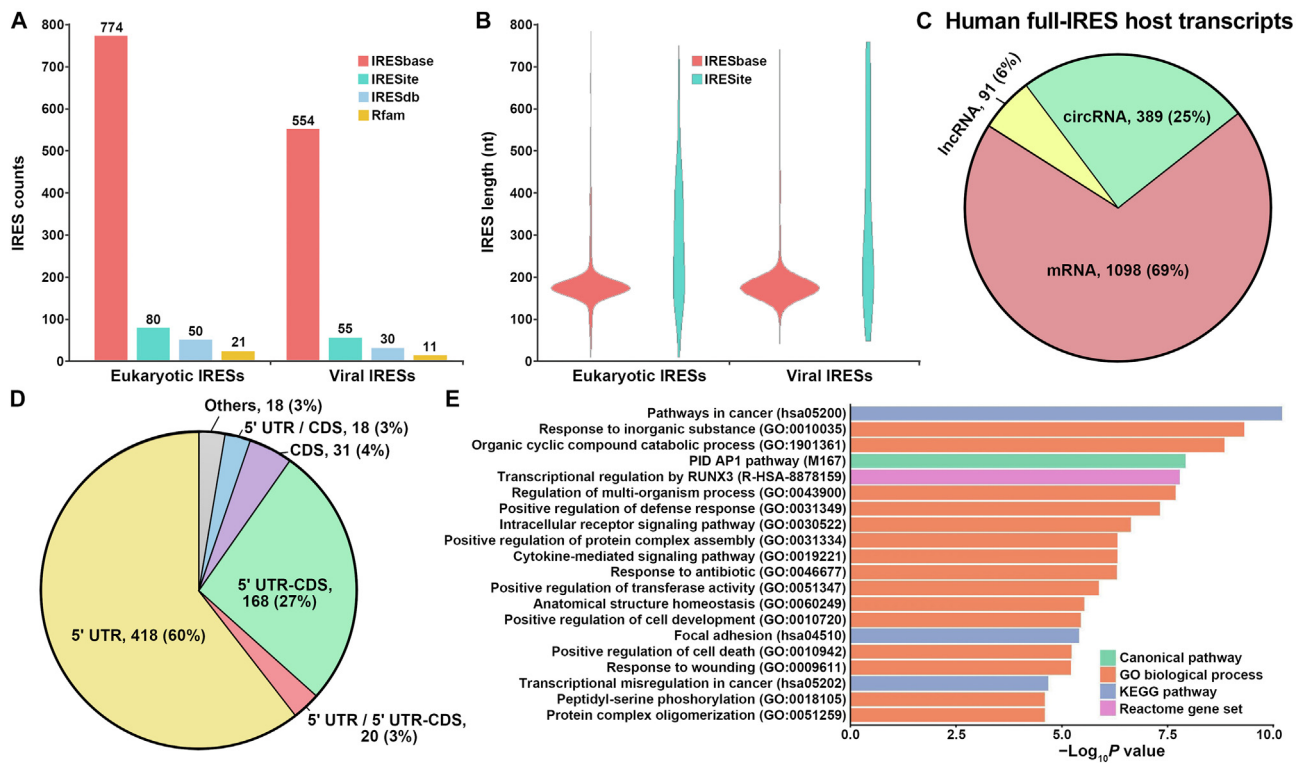


Figure 5 Statistics and analysis of human IRESs

A. Comparison with other IRES databases based on the number of IRESs curated. **B.** Comparison with IRESite based on the IRES length. **C.** The statistics of human full-IRES host transcripts (mRNAs, lncRNAs, and circRNAs). **D.** The statistics of human IRESs location in host mRNAs. **E.** The pathway and biological process enrichment analysis for the human IRES host genes (colored by enriched term).

IRES elements, which could be a rich resource for functional studies and provides clues for protein coding ability of circRNAs and lncRNAs.

In the near future, we will collect and update more IRES related information, including regulatory proteins, critical sites or regions for their activity, and experimentally verified biological function. We also plan to expand the database with experimentally validated sequences without IRES activity, which can not only help prevent repeated validations, but also facilitate the development of bioinformatics tools for the identification of IRES elements. Finally, as an open-access IRES database, we hope researchers can not only use its content but also help us update the database with information on new IRESs and provide us with a feedback.

Availability

IRESbase is freely accessible at <http://reprod.njmu.edu.cn/cgi-bin/iressbase/index.php>.

CRedit author statement

Jian Zhao: Methodology, Data Curation, Formal analysis, Visualization, Writing - Original Draft. **Yan Li:** Investigation, Data Curation. **Cong Wang:** Methodology, Software. **Haotian Zhang:** Methodology, Software. **Hao Zhang:** Investigation. **Bin**

Jiang: Resources. **Xuejiang Guo:** Conceptualization, Supervision, Writing - Review and Editing, Funding acquisition. **Xiaofeng Song:** Conceptualization, Supervision, Writing - Review and Editing, Funding acquisition. All authors read and approved the final manuscript.

Competing interests

The authors have declared no competing interests.

Acknowledgments

The study was supported by the National Natural Science Foundation of China (Grant No. 61973155), the National Key R&D Program of China (Grant No. 2016YFA0503300), the National Natural Science Foundation of China (Grant Nos. 61571223, 81971439, and 81771641), the Program for Distinguished Talents of Six Domains in Jiangsu Province (Grant No. YY-019), the Fok Ying Tung Education Foundation (Grant No. 161037), the Fundamental Research Funds for the Central Universities (Grant No. NP2018109), the National Science Foundation of the Jiangsu Higher Education Institutions (Grant No. 18KJB310006), and the Scientific Research Foundation of Nanjing Medical University (Grant No. 2242019K3DN02), China.

Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.gpb.2020.03.001>.

ORCID

0000-0001-7857-4766 (Zhao, J)
 0000-0003-0420-4222 (Li, Y)
 0000-0003-4669-9702 (Wang, C)
 0000-0002-7181-0962 (Zhang, H)
 0000-0001-7043-5438 (Zhang, H)
 0000-0002-1156-2557 (Jiang, B)
 0000-0002-0475-5705 (Guo, X)
 0000-0001-7445-4302 (Song, X)

References

- [1] Pelletier J, Sonenberg N. Internal initiation of translation of eukaryotic mRNA directed by a sequence derived from poliovirus RNA. *Nature* 1988;334:320–5.
- [2] Jang SK, Kräusslich HG, Nicklin MJ, Duke GM, Palmenberg AC, Wimmer E. A segment of the 5′ nontranslated region of encephalomyocarditis virus RNA directs internal entry of ribosomes during *in vitro* translation. *J Virol* 1988;62:2636–43.
- [3] Kühn R, Luz N, Beck E. Functional characterization of the internal translation initiation site of foot-and-mouth disease virus. *J Virol* 1990;64:4625–31.
- [4] Buck CB, Shen X, Egan MA, Pierson TC, Walker CM, Siliciano RF. The human immunodeficiency virus type 1 *gag* gene encodes an internal ribosome entry site. *J Virol* 2001;75:181–91.
- [5] Holcik M, Korneluk RG. Functional characterization of the X-linked inhibitor of apoptosis (XIAP) internal ribosome entry site element: role of La autoantigen in XIAP translation. *Mol Cell Biol* 2000;20:4648–57.
- [6] Ray PS, Grover R, Das S. Two internal ribosome entry sites mediate the translation of p53 isoforms. *EMBO Rep* 2006;7:404–10.
- [7] Henis-Korenblit S, Strumpf NL, Goldstaub D, Kimchi A. A novel form of DAP5 protein accumulates in apoptotic cells as a result of caspase cleavage and internal ribosome entry site-mediated translation. *Mol Cell Biol* 2000;20:496–506.
- [8] Coldwell MJ, Mitchell SA, Stoneley M, MacFarlane M, Willis AE. Initiation of Apaf-1 translation by internal ribosome entry. *Oncogene* 2000;19:899–905.
- [9] Komar AA, Hatzoglou M. Cellular IRES-mediated translation: the war of ITAFs in pathophysiological states. *Cell Cycle* 2011;10:229–40.
- [10] Thakor N, Holcik M. IRES-mediated translation of cellular messenger RNA operates in eIF2 α -independent manner during stress. *Nucleic Acids Res* 2012;40:541–52.
- [11] Nevins TA, Harder ZM, Korneluk RG, Holcik M. Distinct regulation of internal ribosome entry site-mediated translation following cellular stress is mediated by apoptotic fragments of eIF4G translation initiation factor family members eIF4GI and p97/DAP5/NAT1. *J Biol Chem* 2003;278:3572–9.
- [12] Spriggs KA, Bushell M, Mitchell SA, Willis AE. Internal ribosome entry segment-mediated translation during apoptosis: the role of IRES-*trans*-acting factors. *Cell Death Differ* 2005;12:585–91.
- [13] Pyronnet S, Dostie J, Sonenberg N. Suppression of cap-dependent translation in mitosis. *Genes Dev* 2001;15:2083–93.
- [14] Walters B, Thompson SR. Cap-independent translational control of carcinogenesis. *Front Oncol* 2016;6:128.
- [15] Bonnal S, Boutonnet C, Prado-Lourenço L, Vagner S. IRESdb: the internal ribosome entry site database. *Nucleic Acids Res* 2003;31:427–8.
- [16] Mokrejš M, Vopálenký V, Kolenatý O, Mašek T, Feketová Z, Sekyrová P, et al. IRESite: the database of experimentally verified IRES structures. *Nucleic Acids Res* 2006;34:D125–30.
- [17] Kalvari I, Argasinska J, Quinones-Olvera N, Nawrocki EP, Rivas E, Eddy SR, et al. Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families. *Nucleic Acids Res* 2018;46:D335–42.
- [18] Pamudurti NR, Bartok O, Jens M, Ashwal-Fluss R, Stottmeister C, Ruhe L, et al. Translation of circRNAs. *Mol Cell* 2017;66:9–21.e7.
- [19] Yang Y, Gao X, Zhang M, Yan S, Sun C, Xiao F, et al. Novel role of *FBXW7* circular RNA in repressing glioma tumorigenesis. *J Natl Cancer Inst* 2018;110:304–15.
- [20] Legnini I, Di Timoteo G, Rossi F, Morlando M, Briganti F, Sthandier O, et al. *Circ-ZNF609* is a circular RNA that can be translated and functions in myogenesis. *Mol Cell* 2017;66:22–37.e9.
- [21] Zhang M, Huang N, Yang X, Luo J, Yan S, Xiao F, et al. A novel protein encoded by the circular form of the *SHPRH* gene suppresses glioma tumorigenesis. *Oncogene* 2018;37:1805–14.
- [22] Zhang M, Zhao K, Xu X, Yang Y, Yan S, Wei P, et al. A peptide encoded by circular form of *LINC-PINT* suppresses oncogenic transcriptional elongation in glioblastoma. *Nature Commun* 2018;9:4475.
- [23] Liang WC, Wong CW, Liang PP, Shi M, Cao Y, Rao ST, et al. Translation of the circular RNA *circ β -catenin* promotes liver cancer cell growth through activation of the Wnt pathway. *Genome Biol* 2019;20:84.
- [24] Charpentier M, Croyal M, Carbonnelle D, Fortun A, Florenceau L, Rabu C, et al. IRES-dependent translation of the long non coding RNA *meloe* in melanoma cells produces the most immunogenic MELOE antigens. *Oncotarget* 2016;7:59704–13.
- [25] Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. *Nucleic Acids Res* 2000;28:235–42.
- [26] Kinjo AR, Suzuki H, Yamashita R, Ikegawa Y, Kudou T, Igarashi R, et al. Protein Data Bank Japan (PDBj): maintaining a structural data archive and resource description framework format. *Nucleic Acids Res* 2012;40:D453–60.
- [27] Gutmanas A, Alhroub Y, Battle GM, Berrisford JM, Bochet E, Conroy MJ, et al. PDBE: Protein Data Bank in Europe. *Nucleic Acids Res* 2014;42:D285–91.
- [28] Hofacker IL. Vienna RNA secondary structure server. *Nucleic Acids Res* 2003;31:3429–31.
- [29] O’Leary NA, Wright MW, Brister JR, Ciuffo S, Haddad D, McVeigh R, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res* 2016;44:D733–45.
- [30] Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The human genome browser at UCSC. *Genome Res* 2002;12:996–1006.
- [31] Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, et al. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* 2001;29:308–11.
- [32] Xuan JJ, Sun WJ, Lin PH, Zhou KR, Liu S, Zheng LL, et al. RMBase v2.0: deciphering the map of RNA modifications from epitranscriptome sequencing data. *Nucleic Acids Res* 2018;46:D327–34.
- [33] Betel D, Wilson M, Gabow A, Marks DS, Sander C. The microRNA.org resource: targets and expression. *Nucleic Acids Res* 2008;36:D149–53.
- [34] Johnson M, Zaretskaya I, Raytselis Y, Merezuk Y, McGinnis S, Madden TL. NCBI BLAST: a better web interface. *Nucleic Acids Res* 2008;36:W5–9.
- [35] Glažar P, Papavasiliou P, Rajewsky N. circBase: a database for circular RNAs. *RNA* 2014;20:1666–70.

- [36] Chen X, Han P, Zhou T, Guo X, Song X, Li Y. circRNADb: a comprehensive database for human circular RNAs with protein-coding annotations. *Sci Rep* 2016;6:34985.
- [37] Ji P, Wu W, Chen S, Zheng Y, Zhou L, Zhang J, et al. Expanded expression landscape and prioritization of circular RNAs in mammals. *Cell Rep* 2019;26:3444–60.e5.
- [38] Wang L, Park HJ, Dasari S, Wang S, Kocher JP, Li W. CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model. *Nucleic Acids Res* 2013;41:e74.
- [39] Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res* 1999;27:29–34.
- [40] Carbon S, Ireland A, Mungall CJ, Shu S, Marshall B, Lewis S, et al. AmiGO: online access to ontology and annotation data. *Bioinformatics* 2009;25:288–9.
- [41] Gray KA, Yates B, Seal RL, Wright MW, Bruford EA. Genenames.org: the HGNC resources in 2015. *Nucleic Acids Res* 2015;43:D1079–85.
- [42] Weingarten-Gabbay S, Elias-Kirma S, Nir R, Gritsenko AA, Stern-Ginossar N, Yakhini Z, et al. Systematic discovery of cap-independent translation sequences in human and viral genomes. *Science* 2016;351:aad4939.
- [43] Tripathi S, Pohl MO, Zhou Y, Rodriguez-Frandsen A, Wang G, Stein DA, et al. Meta- and orthogonal integration of influenza “OMICs” data defines a role for UBR4 in virus budding. *Cell Host Microbe* 2015;18:723–35.
- [44] Chen CY, Sarnow P. Initiation of protein synthesis by the eukaryotic translational apparatus on circular RNAs. *Science* 1995;268:415–7.
- [45] Yang Y, Fan XJ, Mao MW, Song XW, Wu P, Zhang Y, et al. Extensive translation of circular RNAs driven by *N*⁶-methyladenosine. *Cell Res* 2017;27:626–41.
- [46] Yamamoto H, Unbehaun A, Spahn CMT. Ribosomal chamber music: toward an understanding of IRES mechanisms. *Trends Biochem Sci* 2017;42:655–68.