DOI:10.12113/202408007

内部核糖体进入位点(IRES)计算识别方法的研究进展

陈哲伟,张 萌,宋晓峰*,赵 健*
(南京航空航天大学 自动化学院,南京 211106)

摘 要:内部核糖体进入位点(Internal ribosome entry site,IRES)是一种RNA顺式作用元件,可以不依赖帽子结构启动翻译过程。其最早在病毒RNA中被报道,随后在真核生物的mRNA中也被发现。近来,IRES元件还被证实存在于环状RNA中,并介导其翻译。IRES元件介导的非帽依赖翻译起始机制在多个关键细胞生命活动过程中发挥着重要作用,并与多种人类复杂疾病的发生发展密切相关,引起越来越多研究者的关注。由于IRES元件的实验鉴定方法费时费力,因此,更为快速便捷的计算识别方法在前期实验数据基础上陆续被开发出来。本文对IRES元件进行了简要概述,总结了目前已知的IRES元件计算识别方法,并且建立了独立测试集对其预测性能进行了深入比较。

关键词:内部核糖体进入位点;非帽依赖;翻译调控;机器学习;深度学习
中图分类号:R318.04 **文献标志码**:A **文章编号**:1672-5565(2025)04-249-12

Review of computational methods for predicting internal ribosome entry site (IRES)

CHEN Zhewei, ZHANG Meng, SONG Xiaofeng*, ZHAO Jian*
(*College of Automation Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing* 211106, *China*)

Abstract: Internal ribosome entry site (IRES) is a cis-regulatory element that can initiate translation in cap-independent manner. It was first discovered in viral RNA sequences, and then cellular mRNAs were found to contain IRESs. Recently, IRES elements have also been reported in circRNAs and mediate their translation. Cap-independent translation mediated by IRES plays important roles in many cellular processes and is involved in many human diseases, receiving more and more attentions. Since identifying IRES elements by experimental method is time-consuming and laborious, faster and more convenient computational methods have been developed based on experimentally validated data. In this paper, we briefly introduce IRES elements, summarize existing IRES predicting methods, and compare their prediction performance on an independent testing dataset.

Keywords: Internal ribosome entry site; Cap-independent translation; Translation regulation; Machine learning; Deep learning

在真核生物中,mRNA的翻译起始通常依赖其5'端的N7-甲基鸟苷(N7-methylguanosine, m7G)帽子结构,被称为帽依赖翻译起始机制。但在某些特殊情况下,比如细胞应激[1]和病毒感染[2],帽依赖

收稿日期:2024-08-14;**修回日期**:2024-09-14;**网络首发日期**:2024-10-17.
网络首发地址:https://link.cnki.net/urlid/23.1513.Q.20241016.1715.002
基金项目:国家自然科学基金面上项目(No. 62273175);江苏省重点研发计划项目(No. BE2022843);中央高校基本科研业务费资助(No. 1003-XAA23018).
*通信作者:宋晓峰,男,教授,研究方向:生物信息学.E-mail: xfsong@nuaa.edu.cn;
赵健,男,副研究员,研究方向:模式识别与生物信息学. E-mail: zhaojian@nuaa.edu.cn.
引用格式:陈哲伟,张 萌,宋晓峰,等. 内部核糖体进入位点(IRES)计算识别方法的研究进展[J].生物信息学,2025,23(4):249-260.
CHEN Zhewei, ZHANG Meng, SONG Xiaofeng, et al. Review of computational methods for predicting internal ribosome entry site (IRES) [J].Chinese Journal of Bioinformatics,2025,23(4):249-260.

翻译起始机制会被抑制,部分 mRNA 转而通过 IRES 元件介导的非帽依赖的翻译起始机制编码蛋白。IRES 元件可绕过5'端帽子结构,直接招募核糖体进入 mRNA 内部,启动蛋白质翻译过程。此前的研究表明 IRES 元件在病毒感染、细胞应激反应和癌症的发生发展过程中起着关键作用[3-7]。据估计,大约 10% 的细胞 mRNA 可能通过 IRES 来启动翻译[7]。尽管目前已鉴定出了一些 IRES 元件,但仍有大量 IRES 元件尚未被发现。由于通过实验鉴定 IRES 元件费时费力,用于 IRES 元件的计算识别方法开始受到越来越多的关注。目前已有多种计算识别方法被提出,其原理和适用对象各不相同,因此本文对这些方法进行综述,并对其分类性能进行了测试和比较。

1 IRES 概述

IRES 元件最早被发现于脑心肌炎病毒和脊髓灰质炎病毒中,Pelletier J 等[8]在 1988 年通过构建双顺反子系统验证了脑心肌炎病毒的 RNA 序列中存在内部核糖体进入位点,Jang S K 等[9]也发现脊髓灰质炎病毒的5'非翻译区(Untranslated region, UTR)存在序列能和核糖体结合。随后,Macejak D G 和 Sarnor P[10]在 1991 年发现细胞 mRNA 中也存在 IRES 元件,他们发现在被病毒感染的宿主细胞中,当帽依赖翻译起始被抑制时,编码免疫球蛋白重链结合蛋白(Immunoglobulin heavy chain binding protein, BIP)的 mRNA 仍能进行翻译。经实验排除了核糖体通读的可能,验证了 BIP mRNA 的5'端可以直接招募核糖体并启动翻译。随着越来越多的 IRES 元件被实验验证,人们发现 IRES 起始非帽依赖的翻译需要特殊的 IRES 反式作用因子(IRES transacting factor, ITAF)的帮助[11-14]。ITAF 通常是核蛋白或者是在细胞核和细胞质之间穿梭的蛋白质,但其作用机制尚不完全清楚。目前的研究认为 ITAF 可以作为 RNA 的分子伴侣,通过改变或稳定 IRES 的二级结构,允许其进一步与蛋白质或 40S 核糖体亚基结合。或者,ITAF 作为接头蛋白质,形成其他蛋白质或 40S 核糖体亚基结合的锚点[15]。

IRES 元件分为病毒 IRES 和细胞 IRES。病毒 IRES 通常具有更复杂的结构特征,这些结构在起始翻译时与核糖体的结合密切相关。根据其结构以及对翻译起始因子(Initiation factors, IFs)和 IRES 反式作用因子的需求,可以将病毒 IRES 元件大致分为四类[16](图 1(a))。Ⅰ类病毒 IRES 代表为蟋蟀麻痹病毒(Cricket paralysis virus, CrPV),可直接招募 40S 小核糖体亚基与 RNA 结合,无需任何翻译起始因子和 IRES 反式作用因子。Ⅱ类病毒 IRES 代表为丙型肝炎病毒(Hepatitis C virus, HCV),也可直接招募 40S 小核糖体亚基与 RNA 结合,但其通常需要翻译起始因子协助。Ⅲ与Ⅳ类病毒 IRES 不能直接结合 40S 小核糖体亚基,需要一些典型的翻译起始因子、氨基酰-tRNA 和 IRES 反式作用因子的帮助。Ⅲ类与Ⅳ类代表病毒分别为脑心肌炎病毒(Encephalomyocarditis virus, EMCV)和脊髓灰质炎病毒(Poliovirus, PV)。它们的主要区别在于Ⅲ类病毒 IRES 招募 40S 核糖体小亚基直接在结合位点启动翻译,而Ⅳ类病毒 IRES 则需要 40S 核糖体小亚基对下游进行扫描并寻找起始密码子来启动翻译。相比于病毒 IRES 元件,细胞 IRES 元件的结构特征更少,且序列的保守性更小。根据核糖体的招募方式,可粗略地将细胞 IRES 元件分为两类[17](图 1(b)):Ⅰ类细胞 IRES 通过与 ITAF 结合来招募核糖体;Ⅱ类细胞 IRES 通过与 18S 核糖体 RNA 序列互补配对来招募核糖体。总的来说,病毒 IRES 拥有相似的二级结构与翻译起始机制,而细胞 IRES 的序列及结构特征则更加多样化。

IRES 不仅能启动 mRNA 的翻译,Chen 等[18]的研究证实 IRES 还可以启动环状 RNA 的翻译。随着研究的深入,已有相当多的环状 RNA 被发现具有编码功能,其翻译产生的蛋白质参与多种生物过程和人类复杂疾病的调控[19]。Legnini 等[20]发现通过 IRES 翻译产生一个含有 250 个氨基酸的蛋白质,该蛋白质可以调节肌细胞的增殖。Pamudurti 等[21]发现果蝇基因组中的部分环状 RNA 与核糖体结合,其中环状 RNA circMbl 的 UTR 区域通过 IRES 启动非帽依赖翻译。Yang 等[22]通过实验鉴定出 Circ-FBXW7 的 UTR 全长片段可以招募核糖体起始翻译,其截短片段的 IRES 活性则大大减弱。Circ-FBXW7 翻译出的蛋白 FBXW7-185aa 在癌症细胞中表达上调,能够抑制细胞增殖并减缓细胞周期加速过程。Li 等[23]则发现环状 RNA 的开放阅读框上游存在 IRES 元件,其编码产物 EIF6-224aa 直接作用于乳腺癌致癌基因 MYH9。已有研究表明,IRES 介导的非帽依赖翻译起始机制与血管再生、细胞凋亡等生物过程密切相关,并在癌症等人类复杂疾病中扮演着重要角色[24]。

IRES 元件的识别对于非帽依赖翻译起始调控机制及其功能研究至关重要,其可作为新的治疗靶点或用于开发新型 RNA 药物。目前最常用于验证 IRES 活性的实验方法是双顺反子报告系统,该系统将待测的包含 IRES 的片段放置于两个开放阅读框

之间,第一个开放阅读框由5'端帽子结构招募核糖体启动翻译,第二个开放阅读框则通过两者间的待测片段招募核糖体启动翻译,由此通过比较两个开放阅读框的蛋白质表达量即可定量待测序列的IRES活性。然而,这一过程会受到待测序列启动子活性或可变剪接事件的干扰,进而出现假阳性结果,因此仍需设计额外的对照实验作进一步验证。综上,通过实验方法鉴定IRES元件耗时耗力,迫切需要开发精确高效的计算方法来节约成本,从而促进其作用机制和生物学功能的深入挖掘。本文总结了现存的IRES元件计算识别方法,并对这些方法进行了性能比较(表1)。

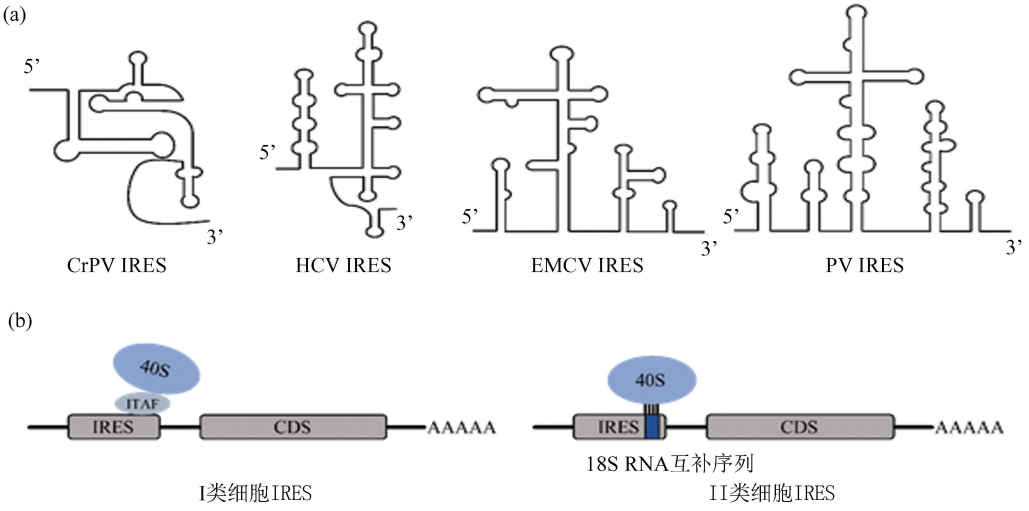

图 1　病毒 IRES 结构和细胞 IRES 起始翻译的机制

Fig.1　Structure of viral IRES and mechanism of translation initiation driven by cellular IRES

注:(a)四类病毒 IRES 结构;(b)细胞 IRES 起始翻译机制.

表 1　IRES 预测工具比较

Table 1　Comparison of IRES predicting tools

工具	发表年份	预测对象	方法或模型	数据集来源	是否可用	网址
IRSS[25]	2009	病毒 IRES	计算结构相似性得分	IRESite,UTRdb,Rfam	否	http://140.135.61.9/ires
VIPS[26]	2013	病毒 IRES	计算结构相似性得分,假结结构判断	IRESite,UTRbd,Rfam	否	http://140.135.61.250/vips
IRESPred[27]	2016	病毒与细胞 IRES	支持向量机	IRESite,GenBank	否	http://bioinfo.net.in/IRESPred
IRESfinder[28]	2018	细胞 IRES	逻辑回归	Weingarten-Gabbay et al.[29]	是	https://github.com/xiaofengsong/IRESfinder
IRESpy[30]	2019	病毒与细胞 IRES	XGBoost	Weingarten-Gabbay et al.[29] Komar et al.[31] IRESite	是	https://irespy.shinyapps.io/IRESpy
DeepCIP[32]	2023	环状 RNA IRES	S-LSTM,GCN	Chen et al.[33]	是	https://github.org/zjupgx/DeepCIP

2　IRES 元件计算识别方法

2.1　基于结构相似性识别方法

2.1.1　IRSS

2009 年,Wu 等[25]最早开发了用于识别病毒基因序列中 IRES 元件的在线搜索系统 IRSS。该搜索系统主要包含两个步骤:第一步,输入的序列通过滑动窗口划分成多个片段,利用 RNALfold 程序预测 RNA 序列的二级结构;第二步,利用 RNA Align 程序将预测的 RNA 二级结构与 Rfam 数据库中已知的 IRES 结构做比对,然后计算结构相似性得分 R。R 被定义为比对匹配得分(Alignment match score, ALEN)和距离得分(Distance score, DIST)的比值,

即 R = ALEN/DIST。其中比对匹配得分代表二级结构之间的匹配程度,距离得分代表结构比对时匹配上的长度。对于 R 阈值的选取,IRSS 将丙型肝炎病毒(Hepatitis C virus, HCV)和瘟疫病毒(Poliovirus, PV)的5'UTR序列作为正样本,来自 UTRdb 数据库的其他5'UTR序列作为负样本。通过计算在不同大小窗口下相似性得分 R 值的分布,利用线性判别分析(Linear discriminant analysis, LDA)方法确定阈值。

为评估 IRSS 的预测准确率,选取 UTRdb 数据库中丙型肝炎病毒和瘟疫病毒的 IRES 作为测试数据。结果显示,IRSS 在滑动窗口 L = 250 时预测能力最强,在两种病毒的 IRES 数据上分别取得了66.7%和72.3%的预测准确率。IRSS 提供了网页服务,输入序列要求 FASTA 格式或纯文本格式,窗口大小 L 和 R 的阈值是可调参数。默认的窗口大小 L 为 250,默认的 R 阈值为 1.4。输出内容为文本格式,包括 RNA 序列、二级结构和预测的最小自由能。IRSS 的预测对象受限于病毒序列,比对过程所需时间较长,不适用于处理大量数据。目前,其网页服务已无法使用。

2.1.2 VIPS

2013 年,Hong 等[26]在 IRSS 的基础上开发了新的病毒 IRES 预测系统 VIPS。该系统增加了假结结构预测模块,以进一步增强对病毒 IRES 元件的预测能力。VIPS 从 NCBI GenBank 和 Rfam 数据库中选取了病毒 IRES 作为结构比对的模板,这些 IRES 可以根据其二级结构分为四组。其中,Ⅰ组包括 16 条蟋蟀麻痹病毒 IRES;Ⅱ组包括 17 条脑心肌炎病毒 IRES;Ⅲ组包括3 096条丙型肝炎病毒 IRES;Ⅳ组包括 213 条脊髓灰质炎病毒 IRES。上述 IRES 元件被选作正样本,相应病毒序列中 IRES 元件外的编码区则被选作负样本。对于每组样本,采用线性判别分析方法确定结构相似性得分 R 的阈值以区分正负样本。同时,假结结构也是识别 IRES 元件的重要因素。VIPS 在结构相似性的基础上,进一步通过是否含有假结结构来更好地预测病毒 IRES 元件。

基于结构相似性得分 R,VIPS 在四组训练集上的分类准确率分别为 92.3%、90.3%、82.4% 和 77.6%;通过进一步判别是否含有假结结构,预测准确率有了较为显著的提升,分别达到 98.5%、90.8%、82.4%和80.4%。用户可以使用 VIPS 提供的网页服务预测病毒 IRES,输入序列为纯文本格式,限制在 5 000 个核苷酸以内。用户还可以调整 R 的阈值,选择是否使用假结结构预测模块,Ⅰ、Ⅱ、Ⅲ和Ⅳ组 IRES 的 R 默认阈值分别为 1.61、1.98、1.87和1.58。其输出内容包括 RNA 序列、二级结构、结构相似性得分、假结预测结果和最小自由能。考虑到 VIPS 仅在训练集上对其分类性能进行了评估,并未经独立测试集检验,因此该预测模型可能存在过拟合。目前,其网页服务已经无法访问。

2.2 基于机器学习模型识别方法

2.2.1 IRESPred

2016 年,Kolekar 等[27]基于支持向量机,开发了适用于病毒和细胞 IRES 的预测方法 IRESPred。正样本来自于 IRESite 数据库中收录的 IRES 元件,负样本则来自于管家基因(Housekeeping gene)的 5'UTR序列、真核生物编码序列(Coding sequence, CDS)和病毒 CDS 序列。在去除正样本中的冗余序列并删除合成序列后,最终构建的数据集包含 189 条正样本和 189 条负样本,其中正样本包含了 58 条人类序列、58 条其它真核生物序列和 73 条病毒序列,负样本则包含了 97 条管家基因序列、46 条细胞 CDS 序列和 46 条病毒 CDS 序列。IRESPred 总共从序列中提取了 35 个特征,其中 8 个特征是 UTR 相关序列特征,如序列长度、上游密码子 AUG 个数和二级结构等,另外 27 个特征为 UTR 与 27 种小亚基核糖体蛋白(Small subunit ribosomal proteins, SSRPs)的结合概率。

数据集被划分成训练集和测试集,比例为 1∶1。作者在训练集上利用十折交叉验证对支持向量机模型的参数进行寻优。经寻优后,IRESPred 在测试集上取得了 75.5%的准确率和 75.3%的敏感性,并且对实验验证的细胞 IRES 元件达到了 73.2%的预测准确率,对病毒 IRES 元件达到了 91.8%的预测准确率。用户可以使用 IRESPred 提供的网页服务预测 IRES,输入序列要求为 FASTA 格式,长度在 15~7 500 个碱基之间,每次输入不能超过 10 条核苷酸序列。IRESPred 使用 CDS 区序列作为负样本,而正样本中的 IRES 元件则位于 UTR 区,考虑到 UTR 与 CDS 二者在序列及结构特征上的固有差别,负样本的选择弱化了模型的泛化能力。此外,IRESPred 的网页服务限制输入序列的数目,因而不适用于批量预测 IRES。IRESPred 网页服务已失效,目前无法访问。

2.2.2 IRESfinder

2018 年,本课题组[28]基于逻辑回归算法,开发了一款用于预测真核生物 IRES 的方法 IRESfinder。该方法首次使用 Weingarten-Gabbay 等[29]通过高通量实验鉴定的 IRES 元件构建数据集。Weingarten-Gabbay 等[29]通过高通量的双顺反子实验鉴定出数千个具有 IRES 活性的序列以及大量不具有 IRES 活性的序列,并且测定了序列的剪切活性和启动子活性以排除干扰。通过对剪切活性和启动子活性的筛

选,首先从人类序列中去除可能的假阳性样本。接着,选取 IRES 活性大于 600 的序列作为正样本,与阴性对照组活性相同的序列作为负样本。最终,构建了包含 401 条正样本和 401 条负样本的标准数据集,从正负样本中各挑选出 40 条序列作为测试集。此外,为了验证模型的泛化能力,还额外构建了两个独立测试集,其正样本源自文献中收集的经实验验证的 IRES 元件,负样本从先前的数据集中随机挑选。其中,第一个独立测试集只选取了 IRES 的核心序列,第二个独立测试集选取了包含 IRES 的完整 5'UTR序列。

为解决 k-mer 特征在 k 值较大时导致的矩阵稀疏问题,对其进行了改进,提出了 framed k-mer 特征。研究选取 $k=\{1,2,3,4,5\}$,从序列中提取了 2 500 个framed k-mer 特征。十折交叉验证结果显示出模型平均 AUC 达到0.782。由于特征维度远大于训练集的样本数(361),作者对特征进行筛选以避免模型过拟合。首先,移除在 90%以上样本中值为 0 的特征;随后,利用单个特征训练模型,移除 AUC 得分低于 0.6 的特征;最后,采用序列前向选择(Sequence forward selection, SFS)算法挑选出了排名靠前的 19 个特征。利用选定的特征训练的模型平均 AUC 值提高到0.825。进一步分析发现,IRES 元件更倾向于包含 T、TNNTT、TNNC、TTNA、TCTNT 和 AAT 短序列,这一结果表明 T(或 U)核苷酸可能对其招募核糖体起着关键作用。在三个独立测试集上,IRESfinder 分别取得了 65%、73.1%和 69.2%的准确率。用户可以使用 IRESfinder 提供的可执行程序预测 IRES,其输入为 FASTA 格式文件,输出为待测序列的概率值及判定结果。另外,IRESfinder 提供了可选模式,可以通过滑动窗口寻找 IRES 核心区域。考虑到 IRESfinder 采用的逻辑回归算法是线性模型,不能捕捉到复杂的非线性关系,其泛化能力会受到一定程度的影响。

2.2.3　IRESpy

2019 年,Wang 等[30]基于极限梯度提升(Extreme gradient boosting, XGBoost)算法,开发了一款快速且可靠的 IRES 在线预测工具 IRESpy。其训练数据同样来源于 Weingarten-Gabbay 等[29]的研究,IRESpy 选取了活性超过 600 的序列作为正样本,剩余序列作为负样本。构建的训练集包含了 2 129 个 IRES,18 743 个非 IRES,其比例约为 1∶9。对于独立测试集的构建,作者从 IRESite 数据库选取 64 条细胞 IRES 和 52 条病毒 IRES 作为正样本,51 条管家基因序列作为负样本。IRESpy 提取了序列特征、结构特征和混合特征。其中序列特征为 k-mer,包含了局部 k-mer 和全局 k-mer 特征;结构特征为 Q_{MFE} 特征,其被定义为序列与随机序列的最小自由能比值,可以用于描述二级结构的复杂度;混合特征为三联体特征,结合了一级序列特征和三联体的结构特征,包含了序列信息和结构信息。

IRESpy 采用十折交叉验证来对模型参数进行寻优,寻优后在测试集上的 AUC 值达到 0.775。IRESpy 的预测概率阈值设置为 0.1,据此在独立测试集上取得了 77.8%的准确率和 79.6%的敏感性。作者还进一步分析了预测的 IRES 概率值与实验测定的 IRES 活性值之间的关系,发现高活性(>600)序列的预测值显著高于低活性序列。最后,作者利用可解释性算法 SHAP,发现 U 富集序列,如"U"、"UU"、"UUU"、"UUUU"、"CU"和"UGU",与较高的 IRES 预测概率相关。IRESpy 为用户提供网页服务,输入为 FASTA 格式的文件,用户可以自由设定预测阈值(默认为 0.1),输出结果包含预测概率值及判定结果。

2.3　基于深度学习模型识别方法 DeepCIP

2023 年,Zhou 等[31]开发了首个环状 RNA 中的 IRES 元件预测工具 DeepCIP。该工具基于多模态深度学习算法,集成了序列信息与结构信息。其数据集取自 Chen 等[32]通过高通量实验鉴定的环状 RNA 中的 IRES 元件,包含 4 531 个正样本和 9 616 个负样本。由于数据集的正负样本数量不平衡,据此 DeepCIP 分别构建了三个正负样本平衡的训练集,由此训练出的模型经软投票方式集成。由于缺少独立测试集,DeepCIP 将数据集中与环状 RNA 数据库比对上的序列作为测试集正样本,随机挑选了相同数量的负样本。DeepCIP 分别构建了序列模型和结构模型:序列模型对 RNA 序列采用了独热(One-hot)、核苷酸化学性质(Nucleotide chemical properties, NCP)、二核苷酸理化性质(Dinucleotide physicochemical properties, DPCP)三种编码方式,采用句子状态长短时记忆网络(Sentence-state long short term memory, S-LSTM)模型提取特征;结构模型的输入为 RNA 序列的二级结构信息,采用图卷积神经网络(Graph convolutional network, GCN)提取结构特征。DeepCIP 在融合上述两个模型的输出后,给出 IRES 预测结果。

经评估,DeepCIP 在测试集上取得了 0.727 的 AUC,并且在文献收集的 21 条环状 RNA IRES 中成功预测出 15 个。该研究还分析了 IRES 预测概率与其实验测定活性之间的相关性,发现活性较高的 IRES 往往具有较高的预测概率值。作者利用 DeepCIP 对未被证实含有 IRES 元件但具备翻译潜

能的环状 RNA 进行了预测,得到 340 个可能包含 IRES 元件的环状 RNA,并且发现这些 IRES 常常存在于环状 RNA 的反向剪接位点附近。最后,作者利用深度学习可解释性算法,积分梯度(Integrated gradients, IG),挖掘序列中对预测值贡献度较高的基序(Motif),发现这些 motif 大多是 AU 富集的序列,此结果与之前研究发现一致[34]。DeepCIP 为用户提供了可运行的程序,输入为 FASTA 格式的文件,输出为 CSV 格式的文件。其提供了多个输入参数,用户可以设置概率阈值,并可切换模式,选择滑动窗口大小和步长后将输入序列分成多个子序列进行预测。

3 IRES 识别方法的性能比较

3.1 数据集

目前可用的 IRES 元件识别工具有 IRESfinder, IRESpy 和 DeepCIP。分别比较了这些工具对于线性 RNA 和环状 RNA 中 IRES 元件的识别能力。为了公正客观地评价上述工具的预测性能,需要建立一个与上述算法的训练数据集没有重合序列的独立测试集。具体地,从文献中收集了 45 条经双顺反子报告系统验证的 IRES 作为正样本[35-64],包含了细胞 IRES 和病毒 IRES,选取相同数量的管家基因 5'UTR序列作为负样本,构建了线性 IRES 测试集。同时,也构建了环状 RNA IRES 独立测试集,由于缺乏负样本,只从文献中收集了经双顺反子报告系统验证的 27 条环状 RNA IRES 元件[20-23, 65-85],长度从 42~478 不等,记为环状 IRES 测试集。

3.2 评价指标

为了全面评估模型的性能,选取了多个评价指标,包括准确率(Accuracy, ACC)、敏感性(Sensitivity, SN)、特异性(Specificity, SP)、马修斯相关系数(Matthew's correlation coefficient, MCC)、受试者工作特征曲线下面积(Area under receiver operating characteristic, AUROC)和 F1 分数(F1-score)。这些指标的计算公式如下:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$SN = \frac{TP}{TP + FN} \tag{2}$$

$$SP = \frac{TN}{TN + FP} \tag{3}$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{4}$$

$$F1 - \text{score} = \frac{2 \times TP}{2 \times TP + FP + FN} \tag{5}$$

其中,TP、TN、FP 和 FN 分别表示真阳性,真阴性,假阳性和假阴性。

3.3 实验参数设置

对于 IRES 的在线预测服务,IRESfinder 和 DeepCIP 提供了可运行的程序,IRESpy 提供了网页服务。分别利用三种算法对测试集序列进行预测,预测时均使用程序提供的默认参数。另外,由于缺少新的 IRES 预测工具,本文构建了一个基于深度学习的 IRES 预测模型。该模型由卷积神经网络(Convolutional neural networks, CNN),门控循环单元(Gate recurrent unit, GRU)和自注意力机制(Self-attention)组成,称为 CNN_GRU_ATT 模型。

模型的输入序列经过独热编码后转变为 L×4 的矩阵,其中 L 表示序列的长度。输入矩阵通过卷积神经网络提取特征,卷积结果进入门控循环单元来捕捉序列上下文关系,其输出通过自注意力机制层来提取更重要的特征,最终经过全连接层和 sigmoid 激活函数得到预测结果。CNN_GRU_ATT 模型的具体超参数设置见表 2。为了测试模型对线性 IRES 和环状 RNA IRES 的预测能力,分别利用 IRESpy 和 DeepCIP 的训练数据训练 CNN_GRU_ATT 模型。此外,IRESfinder、DeepCIP 和 CNN_GRU_ATT 的预测阈值被设置为 0.5,而 IRESpy 的预测阈值被设置成 0.1。

3.4 线性 IRES 预测

在线性 RNA 的 IRES 测试集上,IRESfinder, IRESpy, DeepCIP 和 CNN_GRU_ATT 的性能比较结果如图 2 和表 3 所示。结果显示出,总体上 IRESfinder 在该测试集上预测性能最优,分别取得了 53.3% 的敏感性、86.7% 的特异性、70% 的准确率、0.424的马修斯相关系数、0.64 的 F1 得分和 0.765 的 AUROC 得分,仅特异性和 AUROC 值略低于 CNN_GRU_ATT 模型。IRESpy 与 CNN_GRU_ATT 的预测性能次之,DeepCIP 在此测试集上的预测性能最差,准确率和 AUROC 等指标远低于其他方法。IRESfinder 是基于人类 IRES 数据训练的算法,但是对测试集中的病毒 IRES 也具有预测能力,表明人类 IRES 与病毒 IRES 可能存在共同的序列特征。虽然 IRESpy 的预测准确率与 CNN_GRU_ATT 模型相同,但其 AUROC 得分略低,表明 CNN_GRU_ATT 模型的分类能力更强。

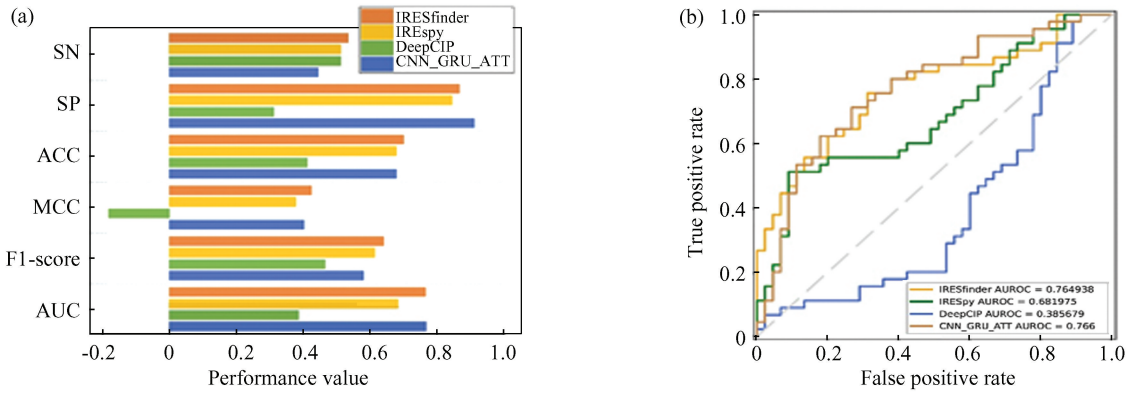

图 2　不同方法对线性 IRES 预测性能比较

Fig.2　Prediction performance comparison of different methods for linear IRESs

注:(a)不同模型在多个评价指标上比较结果;(b)不同模型的 AUROC 曲线比较结果.

表 2　CNN-GRU-ATT 模型的超参数设置

Table 2　Hyperparameter setting of the CNN_GRU_ATT model

超参数	设定值
卷积核大小	3
第一层卷积通道数	64
第二层卷积通道数	16
GRU 隐藏层大小	16
注意力机制维度	16
Dropout 层失活概率	0.2
全连接层维度	32
学习率	2×10^{-3}

DeepCIP 工具主要被设计用于预测环状 RNA 中的 IRES 元件,其在线性测试集的预测性能最差,这说明环状 RNA IRES 与线性 IRES 存在显著的差异。此外,IRESfinder,IRESpy 和 CNN_GRU_ATT 对 IRES 预测的敏感性较低,而特异性较高,说明它们对真实 IRES 的识别能力较差,对非 IRES 的识别能力强。然而,在实际应用中,往往更注重对 IRES 的预测能力。因此,后续研究需要进一步改善预测工具的敏感性。

表 3　线性 IRES 的预测结果

Table 3　Prediction result for linear IRESs

工具	敏感性(%)	特异性(%)	准确率(%)	马修斯相关系数	F1 分数	受试者工作特征曲线下面积
IRESfinder	53.33	86.67	**70.00**	**0.424 3**	**0.640 0**	0.764 9
IRESpy	51.11	84.44	67.78	0.377 1	0.613 3	0.682 0
DeepCIP	51.11	31.11	41.11	−0.181 4	0.464 6	0.385 7
CNN_GRU_ATT	44.44	**91.11**	67.78	0.402 0	0.579 7	**0.766 9**

3.5　环状 IRES 预测

在环状 RNA IRES 测试集上,IRESfinder、IRESpy、DeepCIP 和 CNN-GRU-ATT 的预测结果如表 4 所示。每条 IRES 都标注了宿主环状 RNA 名称和序列长度信息,大于阈值的预测得分用粗体突出显示。IRESfinder,IRESpy,DeepCIP 和 CNN_GRU_ATT 分别正确预测出 12、9、17 和 15 条 IRES。在我们构建的环状 IRES 测试集上,DeepCIP 对环状 RNA IRES 的预测能力最强,CNN_GRU_ATT 和 IRESfinder 次之,IRESpy 的预测能力最差。IRESfinder 和 IRESpy 两款基于线性 IRES 数据训练的模型对环状 RNA IRES 的预测能力较差,这印证了环状 RNA IRES 与线性 IRES 存在差异。我们构建的 CNN_GRU_ATT 模型只基于序列特征对 IRES 进行预测,而 DeepCIP 结合了序列特征与结构特征。DeepCIP 对环状 RNA IRES 的预测能力更强,表明环状 RNA IRES 有其独特的结构特征。总的来说,对环状 RNA IRES 的预测能力还有待提高,需要开发预测准确率更高的方法。

4　总　结

蛋白质是生命活动的主要承担者,其翻译起始机制可分为两类:帽依赖翻译起始和 IRES 介导的非帽依赖翻译起始。IRES 作为一种顺式作用元件,可在帽依赖翻译被抑制或不可用的情况下,起始非帽

依赖翻译。现有研究表明,非帽依赖翻译起始在细胞应激反应、病毒蛋白翻译等生命过程中发挥着至关重要的作用,且与包括癌症在内的多种人类复杂疾病密切相关。尽管目前已经鉴定出许多 IRES 元件,但是仍有大量的 IRES 元件未被识别。鉴于 IRES 元件的实验鉴定方法费时费力,为了节约时间及成本,深入探索 IRES 元件的作用机制及功能,亟需开发相关计算识别方法。

表 4 环状 RNA IRES 的预测结果

Table 4 Prediction results for circRNA IRESs

环状 RNA	长度	IRESfinder 得分	IRESpy 得分	DeepCIP 得分	CNN-GRU-ATT 得分
Circ-E-Cad	**313**	**0.363 2**	**0.038 5**	**0.605 1**	**0.593 0**
CircEIF6	148	**0.534 5**	0.085 8	**0.659 3**	**0.696 4**
CircSMO	148	**0.615 4**	**0.136 1**	**0.707 8**	**0.550 4**
CircPLCE1	334	**0.850 2**	0.052 6	0.041 2	0.340 0
CircEZH2	156	0.479 7	0.025 0	0.356 2	0.215 1
CircATG4B	139	0.220 5	0.039 2	**0.719 7**	**0.591 5**
Circ-FBXW7	62	0.441 7	0.069 4	**0.539 9**	**0.611 9**
Circ-SHPRH	127	**0.689 1**	0.046 9	0.378 2	0.498 4
CircZNF609	211	0.374 3	0.019 2	**0.542 7**	**0.674 3**
Circβ-catenin	107	**0.682 4**	**0.206 6**	0.423 9	0.450 2
CircPINTexon2	478	**0.698 0**	**0.291 9**	0.200 7	**0.647 7**
Circ-AKT3	216	**0.528 9**	0.026 8	0.256 8	0.215 9
CircFNDC3B	77	0.351 2	0.019 7	**0.540 8**	0.455 1
CircMAPK1	54	**0.787 7**	**0.111 2**	**0.514 0**	0.266 6
Circ-HER2	149	0.473 9	0.082 6	**0.708 7**	**0.562 2**
CircSEMA4B	130	0.374 6	0.074 1	**0.692 2**	**0.740 6**
CircZKSCAN1	151	0.283 8	**0.111 5**	**0.628 7**	**0.640 6**
CircZKSCAN1	151	0.329 1	**0.107 0**	**0.606 1**	**0.727 9**
cGGNBP2	139	**0.557 9**	**0.177 9**	**0.591 3**	**0.683 9**
ecircCUX1	138	0.127 8	0.041 3	**0.660 4**	**0.582 9**
ecircCUX1	107	0.185 7	0.037 5	**0.647 3**	0.413 6
vcircRNA_000048	42	0.498 6	**0.109 6**	**0.570 1**	**0.559 8**
CircPDHK1	149	**0.641 9**	0.076 8	0.308 8	0.447 6
CirCTRIM1	141	**0.693 6**	**0.130 5**	**0.605 9**	**0.711 9**
CirCYThdC2	174	**0.819 3**	0.063 9	0.375 3	0.309 4
Circ-HGF	47	0.086 4	0.037 3	0.498 9	0.338 2
CircRALGPS2	245	0.478 9	0.049 0	0.218 6	0.155 2

目前可用的 IRES 元件计算识别方法包括识别细胞 IRES 的 IRESfinder、识别细胞和病毒 IRES 的 IRESpy 以及识别环状 RNA IRES 的 DeepCIP。本文建立了一个基于深度学习模型的 IRES 预测方法,并建立独立测试集比较了这些方法的预测性能。然而,上述方法对于新数据的预测能力均不太理想,仍有待进一步提升。其中,较为关键的原因在于,上述方法所用的实验验证的 IRES 元件长度仅为 174 个核苷酸,而真实 IRES 元件的长度不定。此外,受限于训练样本长度,其主要采用序列特征,然而研究表明 IRES 元件活性与其结构密切相关。最后,目前开发的算法大多数基于机器学习模型,而深度学习模型相较于机器学习模型有着更高的精度,将来需要开发更多基于深度学习模型的 IRES 元件计算识别方法。

参考文献(References)

[1]PRIGGS K A, BUSHELL M, WILLIS A E. Translational regulation of gene expression during conditions of cell stress [J]. Molecular Cell, 2010, 40(2): 228-237. DOI:10.1016/j.molcel.2010.09.028.

[2]KNELLER E L, RAKOTONDRAFARA A M, MILLER W A. Cap-independent translation of plant viral RNAs [J]. Virus Research, 2006, 119(1): 63-75. DOI:10.1016/j.virusres.2005.10.010.

[3]BELTRAN M, PUIG I, PEÑA C,et al. A natural antisense transcript regulates Zeb2/Sip1 gene expression during Snail1-induced epithelial-mesenchymal transition [J]. Genes & Development, 2008, 22(6): 756-769. DOI:10.1101/gad.455708.

[4]GUNASEELAN S, WONG K Z, MIN N, et al. Prunin suppresses viral IRES activity and is a potential candidate for treating enterovirus A71 infection[J]. Science Translational Medicine, 2019, 11(516): eaar5759. DOI: 10.1126/scitranslmed.aar5759.

[5]WANG Jinlong, ZHANG Xing, GREENE G H, et al. PABP/purine-rich motif as an initiation module for cap-independent translation in pattern-triggered immunity [J]. Cell, 2022, 185(17): 3186-3200. e17. DOI: 10.1016/j.cell.2022.06.037.

[6]FAN Xinyi,ZHAO Zitong, MA Liying, et al. PTBP1 promotes IRES-mediated translation of cyclin B1 in cancer[J]. Acta Biochimica et Biophysica Sinica, 2022, 54(5): 696-707. DOI:10.3724/abbs.2022046.

[7]STONELEY M, WILLIS A E. Cellular internal ribosome entry segments: Structures, trans-acting factors and regulation of gene expression[J]. Oncogene, 2004, 23(18): 3200-3207. DOI:10.1038/sj.onc.1207551.

[8]PELLETIER J,SONENBERG N. Internal initiation of translation of eukaryotic mRNA directed by a sequence derived from poliovirus RNA[J]. Nature, 1988, 334(6180): 320-325. DOI:10.1038/334320a0.

[9]JANG S K, KRäUSSLICH H G, NICKLIN M J, et al. A segment of the 5' nontranslated region of encephalomyocarditis virus RNA directs internal entry of ribosomes during in vitro translation [J]. Journal of Virology, 1988, 62(8): 2636-2643. DOI:10.1128/jvi.62.8.2636-2643.1988.

[10]MACEJAK D G,SARNOW P. Internal initiation of translation mediated by the 5' leader of a cellular mRNA [J]. Nature, 1991, 353(6339): 90-94. DOI: 10.1038/353090a0.

[11]MEEROVITCH K,SVITKIN Y V, LEE H S, et al. La autoantigen enhances and corrects aberrant translation of poliovirus RNA in reticulocyte lysate[J]. Journal of Virology, 1993, 67(7): 3798-3807. DOI:10.1128/jvi.67.7.3798-3807.1993.

[12]PESTOVA T V,SHATSKY I N, HELLEN C U. Functional dissection of eukaryotic initiation factor 4F: the 4A subunit and the central domain of the 4G subunit are sufficient to mediate internal entry of 43S preinitiation complexes[J]. Molecular and Cellular Biology, 1996, 16(12): 6870-6878. DOI:10.1128/mcb.16.12.6870.

[13]NIEPMANN M, PETERSEN A, MEYER K, et al. Functional involvement of polypyrimidine tract-binding protein in translation initiation complexes with the internal ribosome entry site of foot-and-mouth disease virus[J]. Journal of Virology, 1997, 71(11): 8330-8339. DOI:10.1128/jvi.71.11.8330-8339.1997.

[14]PILIPENKO E V,PESTOVA T V, KOLUPAEVA V G, et al. A cell cycle-dependent protein serves as a template-specific translation initiation factor[J]. Genes & Development, 2000, 14(16): 2028-2045. DOI: 10.1101/gad.14.16.2028.

[15]KING H A, COBBOLD L C, WILLIS A E. The role of IRES trans-acting factors in regulating translation initiation [J]. Biochemical Society Transactions, 2010, 38(6): 1581-1586. DOI:10.1042/bst0381581.

[16]KIEFT J S. Viral IRES RNA structures and ribosome interactions[J]. Trends in Biochemical Sciences, 2008, 33(6): 274-283. DOI:10.1016/j.tibs.2008.04.007.

[17]YANG Yun, WANG Zefeng. IRES-mediated cap-independent translation, a path leading to hidden proteome [J]. Journal of Molecular Cell Biology, 2019, 11(10): 911-919. DOI:10.1093/jmcb/mjz091.

[18]CHEN Changyou, SARNOW P. Initiation of protein synthesis by the eukaryotic translational apparatus on circular RNAs[J]. Science, 1995, 268(5209): 415-417. DOI: 10.1126/science.7536344.

[19]CHENG Jiahui, LI Guangyue, WANG Wenmeng, et al. Circular RNAs with protein-coding ability in oncogenesis [J]. Biochimica et Biophysica Acta Reviews on Cancer, 2023, 1878(4): 188909. DOI: 10.1016/j.bbcan.2023.188909.

[20]LEGNINI I, DI TIMOTEO G, ROSSI F, et al. Circ-znf609 is a circular RNA that can be translated and functions in myogenesis[J]. Molecular Cell, 2017, 66(1): 22-37.e9. DOI:10.1016/j.molcel.2017.02.017.

[21]PAMUDURTI N R, BARTOK O, JENS M, et al. Translation of circRNAs[J]. Molecular Cell, 2017, 66(1): 9-21.e7. DOI:10.1016/j.molcel.2017.02.021.

[22]YANG Yibing, GAO Xinya, ZHANG Maolei, et al. Novel role of FBXW7 circular RNA in repressing glioma tumorigenesis[J]. Journal of the National Cancer Institute, 2018, 110(3): 304-315. DOI:10.1093/jnci/djx166.

[23]LI Yaming, WANG Zekun, SU Peng, et al. circ-EIF6 en-

codes EIF6-224aa to promote TNBC progression via stabilizing MYH9 and activating the Wnt/beta-catenin pathway [J]. Molecular Therapy, 2022, 30(1): 415-430. DOI: 10.1016/j.ymthe.2021.08.026.

[24] KOMAR A A, HATZOGLOU M. Cellular IRES-mediated translation: the war of ITAFs in pathophysiological states [J]. Cell Cycle, 2011, 10(2): 229-240. DOI:10.4161/cc.10.2.14472.

[25] WU T Y, HSIEH C C, HONG Junjie, et al. IRSS: A web-based tool for automatic layout and analysis of IRES secondary structure prediction and searching system in silico[J]. BMC Bioinformatics, 2009, 10: 160. DOI:10.1186/1471-2105-10-160.

[26] HONG Junjie, WU T Y, CHANG T Y, et al. Viral IRES prediction system - a web server for prediction of the IRES secondary structure in silico[J]. PloS One, 2013, 8(11): e79288. DOI:10.1371/journal.pone.0079288.

[27] KOLEKAR P, PATASKAR A, KULKARNI-KALE U, et al. IRESPred: Web server for prediction of cellular and viral internal ribosome entry site (IRES)[J]. Scientific Reports, 2016, 6: 27436. DOI:10.1038/srep27436.

[28] ZHAO Jian, WU Jing, XU Tianyi, et al. IRESfinder: Identifying RNA internal ribosome entry site in eukaryotic cell using framed k-mer features[J]. Journal of Genetics and Genomics, 2018, 45(7): 403-406. DOI:10.1016/j.jgg.2018.07.006.

[29] Weingarten-Gabbay S, ELIAS-KIRMA S, NIR R, et al. Systematic discovery of cap-independent translation sequences in human and viral genomes[J]. Science, 2016, 351(6270):240. DOI:10.1126/science.aad4939.

[30] WANG Junhui, GRIBSKOV M. IRESpy: An XGBoost model for prediction of internal ribosome entry sites[J]. BMC Bioinformatics, 2019, 20: 409. DOI: 10. 1186/s12859-019-2999-7.

[31] KOMAR A A, MAZUMDER B, MERRICK W C. A new framework for understanding IRES-mediated translation[J]. Gene, 2012, 502(2): 75-86. DOI:10.1016/j.gene.2012.04.039.

[32] ZHOU Yuxuan, WU Jingcheng, YAO Shihao, et al. DeepCIP: A multimodal deep learning method for the prediction of internal ribosome entry sites of circRNAs[J]. Computers in Biology and Medicine, 2023, 164: 107288. DOI: 10.1016/j.compbiomed.2023.107288.

[33] CHEN Chunkan, CHENG Ran, DEMETER J, et al. Structured elements drive extensive circular RNA translation[J]. Molecular Cell, 2021, 81(20): 4300-4318.e13. DOI:10.1016/j.molcel.2021.07.042.

[34] FAN Xiaojun, YANG Yun, CHEN Chuyun, et al. Pervasive translation of circular RNAs driven by short IRES-like elements[J]. Nature Communications, 2022, 13: 3751. DOI:10.1038/s41467-022-31327-y.

[35] LURIA N, SMITH E, LACHMAN O, et al. Isolation and characterization of a novel cripavirus, the first Dicistroviridae family member infecting the cotton mealybug Phenacoccus solenopsis[J]. Archives of Virology, 2020, 165(9): 1987-1994. DOI:10.1007/s00705-020-04702-7.

[36] MOREIRA-SOTO A, ARROYO-MURILLO F, SANDER A L, et al. Cross-order host switches of hepatitis C-related viruses illustrated by a novel hepacivirus from sloths[J]. Virus Evolution, 2020, 6(2): veaa033. DOI:10.1093/ve/veaa033.

[37] CANUTI M, WILSON L, BOWES V, et al. A novel calicivirus discovered in trumpeter swans (Cygnus buccinator) expands the richness of known avian caliciviruses[J]. Current Research in Microbial Sciences, 2022, 3: 100169. DOI:10.1016/j.crmicr.2022.100169.

[38] SICULELLA L, GIANNOTTI L, TESTINI M, et al. In steatotic cells, ATP-Citrate Lyase mRNA is efficiently translated through a cap-independent mechanism, contributing to the stimulation of de novo lipogenesis[J]. International Journal of Molecular Sciences, 2020, 21(4):1206. DOI:10.3390/ijms21041206.

[39] COURTILLON C, BRIAND F X, ALLÉE C, et al. Description of the first isolates of guinea fowl corona and picornaviruses obtained from a case of guinea fowl fulminating enteritis[J]. Avian Pathology, 2021, 50(6): 507-521. DOI:10.1080/03079457.2021.1976725.

[40] HEDIL M, NAKASU E Y T, NAGATA T, et al. New features on the genomic organization of a novel dicistrovirus identified from the sweet potato whitefly Bemisia tabaci[J]. Virus Research, 2020, 288: 198112. DOI:10.1016/j.virusres.2020.198112.

[41] KOK V J T, TANG Jiaying, ENG G W L, et al. SFPQ promotes RAS-mutant cancer cell growth by modulating 5'-UTR mediated translational control of CK1α[J]. Nucleic Acids Research Cancer, 2022, 4(3): zcac027. DOI:10.1093/narcan/zcac027.

[42] YU Ruyuan, HU Yameng, ZHANG Shuxia, et al. LncRNA CTBP1-DT-encoded microprotein DDUP sustains DNA damage response signalling to trigger dual DNA repair mechanisms[J]. Nucleic Acids Research, 2022, 50(14): 8060-8079. DOI:10.1093/nar/gkac611.

[43] RITTORE C, MÉCHIN D, SANCHEZ E, et al. TNFR1-d2 carrying the p.(Thr79Met) pathogenic variant is a potential novel actor of TNFα/TNFR1 signalling regulation in the pathophysiology of TRAPS[J]. Scientific Reports, 2021, 11: 4172. DOI:10.1038/s41598-021-83539-9.

[44] PHUNG B, CIESLA M, SANNA A, et al. The x-linked DDX3X RNA helicase dictates translation reprogramming and metastasis in melanoma[J]. Cell Reports, 2019, 27(12): 3573-3586.e7. DOI:10.1016/j.celrep.2019.05.069.

[45] JAUD M, PHILIPPE C, VAN DEN BERGHE L, et al. The

PERK branch of the unfolded protein response promotes DLL4 expression by activating an alternative translation mechanism[J]. Cancers, 2019, 11(2):142. DOI:10.3390/cancers11020142.

[46]LI Hui, CHEN Yuhang, ZHANG Junshi, et al. Identification of the internal ribosome entry sites in the 5'-untranslated region of the c-fos gene[J]. International Journal of Molecular Medicine, 2021, 47(4):56. DOI:10.3892/ijmm.2021.4889.

[47]LEE H C, FU Chuanyang, LIN C Y, et al. Poly(U)-specific endoribonuclease ENDOU promotes translation of human CHOP mRNA by releasing uORF element-mediated inhibition[J]. the EMBO Journal, 2021, 40(11): e104123. DOI:10.15252/embj.2019104123.

[48]SHI Bingtian, SONG Qinqin, LUO Xiaonuan, et al. Identification of cryptic putative IRESs within the ORF encoding the nonstructural proteins of the human rhinovirus 16 genome[J]. Archives of Virology, 2021, 166(12): 3373-3386. DOI:10.1007/s00705-021-05209-5.

[49]SHI Bingtian, SONG Qinqin, LUO Xiaonuan, et al. Identification of an IRES within the coding region of the structural protein of human rhinovirus 16[J]. Journal of Medical Virology, 2022, 94(6): 2653-2661. DOI:10.1002/jmv.27507.

[50]SONG Qinqin, LUO Xiaonuan, SHI Bingtian, et al. Exploration of IRES Elements within the ORF of the coxsackievirus B3 genome[J]. Biomedical and Environmental Sciences, 2022, 35(4): 322-333. DOI:10.3967/bes2022.043.

[51]WANG Hongbo, ZHU Yunjia, HU Lijun, et al. Internal ribosome entry sites mediate cap-independent translation of bmi1 in nasopharyngeal carcinoma[J]. Frontiers in Oncology, 2020, 10: 1678. DOI:10.3389/fonc.2020.01678.

[52]WANG Xinying, VLOK M, FLIBOTTE S, et al. Resurrection of a viral internal ribosome entry site from a 700 year old ancient northwest territories cripavirus[J]. Viruses, 2021, 13(3):493. DOI:10.3390/v13030493.

[53]WARNAKULASURIYARACHCHI D, CERQUOZZI S, CHEUNG H H, et al. Translational induction of the inhibitor of apoptosis protein HIAP2 during endoplasmic reticulum stress attenuates cell death and is mediated via an inducible internal ribosome entry site element[J]. Journal of Biological Chemistry, 2004, 279(17): 17148-17157. DOI:10.1074/jbc.M308737200.

[54]SHI Yijiang, SHARMA A, WU Hong, et al. Cyclin D1 and c-myc internal ribosome entry site (IRES)-dependent translation is regulated by AKT activity and enhanced by rapamycin through a p38 MAPK- and ERK-dependent pathway[J]. Journal of Biological Chemistry, 2005, 280(12): 10964-10973. DOI:10.1074/jbc.M407874200.

[55]MENG Zheng, JACKSON N L, CHOI H, et al. Alterations in RNA-binding activities of IRES-regulatory proteins as a mechanism for physiological variability and pathological dysregulation of IGF-IR translational control in human breast tumor cells[J]. Journal of Cellular Physiology, 2008, 217(1): 172-183. DOI:10.1002/jcp.21486.

[56]COLEMAN J, MISKIMINS W K. Structure and activity of the internal ribosome entry site within the human p27 Kip1 5'-untranslated region[J]. RNA Biology, 2009, 6(1):84-89. DOI:10.4161/rna.6.1.7572.

[57]YANG Lin, GU Lubing, LI Zhuoya, et al. Translation of TRAF1 is regulated by IRES-dependent mechanism and stimulated by vincristine[J]. Nucleic Acids Research, 2010, 38(13): 4503-4513. DOI:10.1093/nar/gkq183.

[58]ZHANG Wei, WANG Xia, XIAO Zhifeng, et al. Mapping of the minimal internal ribosome entry site element in the human embryonic stem cell gene OCT4B mRNA[J]. Biochemical and Biophysical Research Communications, 2010, 394(3): 750-754. DOI:10.1016/j.bbrc.2010.03.064.

[59]GROVER R, SHARATHCHANDRA A, PONNUSWAMY A, et al. Effect of mutations on the p53 IRES RNA structure: Implications for de-regulation of the synthesis of p53 isoforms[J]. RNA Biology, 2011, 8(1): 132-142. DOI:10.4161/rna.8.1.14260.

[60]PETZ M, THEM N, HUBER H, et al. La enhances IRES-mediated translation of laminin B1 during malignant epithelial to mesenchymal transition[J]. Nucleic Acids Research, 2012, 40(1): 290-302. DOI:10.1093/nar/gkr717.

[61]ZHANG Hailong, LIU Tao, YI Sha, et al. Targeting MYCN IRES in MYCN-amplified neuroblastoma with miR-375 inhibits tumor growth and sensitizes tumor cells to radiation[J]. Molecular Oncology, 2015, 9(7):1301-1311. DOI:10.1016/j.molonc.2015.03.005.

[62]GAO Wenqing, LI Qi, ZHU Ruiyu, et al. La Autoantigen induces ribosome binding protein 1 (RRBP1) expression through internal ribosome entry site (IRES)-mediated translation during cellular stress condition[J]. International Journal of Molecular Sciences, 2016, 17(7):1174. DOI:10.3390/ijms17071174.

[63]KUNZE M M, BENZ F, BRAUß T F, et al. sST2 translation is regulated by FGF2 via an hnRNP A1-mediated IRES-dependent mechanism[J]. Biochimica et Biophysica Acta, 2016, 1859(7): 848-859. DOI:10.1016/j.bbagrm.2016.05.005.

[64]LI Qi, GAO Wenqing, DAI Wenyan, et al. ATF2 translation is induced under chemotherapeutic drug-mediated cellular stress via an IRES-dependent mechanism in human hepatic cancer Bel7402 cells[J]. Oncology Letters, 2016, 12(6): 4795-4802. DOI:10.3892/ol.2016.5274.

[65]GAO Xinya, XIA Xin, LI Fanying, et al. Circular RNA-encoded oncogenic E-cadherin variant promotes glioblastoma tumorigenicity through activation of EGFR-STAT3 signalling[J]. Nature Cell Biology, 2021, 23(3): 278-291. DOI:

10.1038/s41556-021-00639-4.

[66] WU Xujia, XIAO Songhua, ZHANG Maolei, et al. A novel protein encoded by circular SMO RNA is essential for Hedgehog signaling activation and glioblastoma tumorigenicity[J]. Genome Biology, 2021, 22: 33. DOI: 10.1186/s13059-020-02250-6.

[67] LIANG Zhenxing, LIU Huashan, XIONG Li, et al. A novel NF-κB regulator encoded by circPLCE1 inhibits colorectal carcinoma progression by promoting RPS3 ubiquitin-dependent degradation[J]. Molecular Cancer, 2021, 20: 103. DOI: 10.1186/s12943-021-01404-9.

[68] ZHONG Jian, YANG Xuesong, CHEN Junju, et al. Circular EZH2-encoded EZH2-92aa mediates immune evasion in glioblastoma via inhibition of surface NKG2D ligands[J]. Nature Communications, 2022, 13: 4795. DOI: 10.1038/s41467-022-32311-2.

[69] PAN Zihao, ZHENG Jun, ZHANG Jiebin, et al. A novel protein encoded by exosomal circatg4b induces oxaliplatin resistance in colorectal cancer by promoting autophagy[J]. Advanced Science, 2022, 9(35): e2204513. DOI: 10.1002/advs.202204513.

[70] ZHANG Maolei, HUANG Nunu, YANG Xuesong, et al. A novel protein encoded by the circular form of the SHPRH gene suppresses glioma tumorigenesis[J]. Oncogene, 2018, 37(13): 1805-1814. DOI: 10.1038/s41388-017-0019-9.

[71] XIA Xin, LI Xixi, LI Fanying, et al. A novel tumor suppressor protein encoded by circular AKT3 RNA inhibits glioblastoma tumorigenicity by competing with active phosphoinositide-dependent Kinase-1[J]. Molecular Cancer, 2019, 18: 131. DOI: 10.1186/s12943-019-1056-5.

[72] LIANG Weicheng, WONG C W, LIANG Puping, et al. Translation of the circular RNA circβ-catenin promotes liver cancer cell growth through activation of the Wnt pathway[J]. Genome Biology, 2019, 20: 84. DOI: 10.1186/s13059-019-1685-4.

[73] ZHANG Maolei, ZHAO Kun, XU Xiaoping, et al. A peptide encoded by circular form of LINC-PINT suppresses oncogenic transcriptional elongation in glioblastoma[J]. Nature Communications, 2018, 9: 4475. DOI: 10.1038/s41467-018-06862-2.

[74] PAN Zihao, CAI Jianye, LIN Jiatong, et al. A novel protein encoded by circFNDC3B inhibits tumor progression and EMT through regulating Snail in colon cancer[J]. Molecular Cancer, 2020, 19: 71. DOI: 10.1186/s12943-020-01179-5.

[75] LI Jie, MA Maoguang, YANG Xuesong, et al. Circular HER2 RNA positive triple negative breast cancer is sensitive to Pertuzumab[J]. Molecular Cancer, 2020, 19: 142. DOI: 10.1186/s12943-020-01259-6.

[76] WANG Xuehui, JIAN Wei, LUO Qifeng, et al. CircSEMA4B inhibits the progression of breast cancer by encoding a novel protein SEMA4B-211aa and regulating AKT phosphorylation[J]. Cell Death & Disease, 2022, 13(9): 794. DOI: 10.1038/s41419-022-05246-1.

[77] SONG Runjie, MA Shuoqian, XU Jiajia, et al. A novel polypeptide encoded by the circular RNA ZKSCAN1 suppresses HCC via degradation of mTOR[J]. Molecular Cancer, 2023, 22: 16. DOI: 10.1186/s12943-023-01719-9.

[78] LI Hui, LAN Tian, LIU Hailing, et al. IL-6-induced cGGNBP2 encodes a protein to promote cell growth and metastasis in intrahepatic cholangiocarcinoma[J]. Hepatology, 2022, 75(6): 1402-1419. DOI: 10.1002/hep.32232.

[79] YANG Feng, HU Anpei, GUO Yanhua, et al. p113 isoform encoded by CUX1 circular RNA drives tumor progression via facilitating ZRF1/BRD4 transactivation[J]. Molecular Cancer, 2021, 20: 123. DOI: 10.1186/s12943-021-01421-8.

[80] ZHANG Yunshan, ZHU Min, ZHANG Xing, et al. Micropeptide vsp21 translated by Reovirus circular RNA 000048 attenuates viral replication[J]. International Journal of Biological Macromolecules, 2022, 209(Pt A): 1179-1187. DOI: 10.1016/j.ijbiomac.2022.04.136.

[81] ZHENG Weiwei, WANG Linchao, GENG Shang, et al. CircYthdc2 generates polypeptides through two translation strategies to facilitate virus escape[J]. Cellular and Molecular Life Sciences, 2024, 81: 91. DOI: 10.1007/s00018-024-05148-9.

[82] HE Haorong, WEI Yuanhang, CHEN Yuqi, et al. High expression circRALGPS2 in atretic follicle induces chicken granulosa cell apoptosis and autophagy via encoding a new protein[J]. Journal of Animal Science and Biotechnology, 2024, 15: 42. DOI: 10.1186/s40104-024-01003-w.

[83] HUANG Bo, REN Junwu, MA Qiang, et al. A novel peptide PDHK1-241aa encoded by circPDHK1 promotes ccRCC progression via interacting with PPP1CA to inhibit AKT dephosphorylation and activate the AKT-mTOR signaling pathway[J]. Molecular Cancer, 2024, 23: 34. DOI: 10.1186/s12943-024-01940-0.

[84] SAUNDERS J T, KUMAR S, BENAVIDES-SERRATO A, et al. Translation of circHGF RNA encodes an HGF protein variant promoting glioblastoma growth through stimulation of c-MET[J]. Journal of Neuro-oncology, 2023, 163: 207-218. DOI: 10.1007/s11060-023-04331-5.

[85] LI Yaming, WANG Zekun, YANG Jingwen, et al. CircTRIM1 encodes TRIM1-269aa to promote chemoresistance and metastasis of TNBC via enhancing CaM-dependent MARCKS translocation and PI3K/AKT/mTOR activation[J]. Molecular Cancer, 2024, 23: 102. DOI: 10.1186/s12943-024-02019-6.