

短肽编码基因的研究进展

田原, 杨金娥

■背景资料

随着转录组的深入研究及高分辨率质谱技术的发展, 在模式动物以及人类中已发现了一批由mRNA或非编码RNA(non coding RNA, ncRNA)上的小开放阅读框(small open reading frame, sORF)翻译生成的短肽(short peptides, SEPs). SEPs具有生物学功能, 并且在种间保守。

田原, 杨金娥, 中山大学生命科学学院 基因工程教育部重点实验室 广东省广州市 510275

作者贡献分布: 本文由田原与杨金娥完成; 杨金娥审校。

通讯作者: 杨金娥, 副教授, 510275, 广东省广州市海珠区新港西路135号, 中山大学生命科学学院, 基因工程教育部重点实验室. lssyje@mail.sysu.edu.cn
电话: 021-84115532

收稿日期: 2015-04-28
修回日期: 2015-06-15
接受日期: 2015-06-19
在线出版日期: 2015-11-08

Emerging landscape of short open reading frame-encoded peptides

Yuan Tian, Jin-E Yang

Yuan Tian, Jin-E Yang, School of Life Sciences, Sun Yat-sen University; Key Laboratory of Gene Engineering of the Ministry of Education, Guangzhou 510275, Guangdong Province, China

Correspondence to: Jin-E Yang, Associate Professor, School of Life Sciences, Sun Yat-sen University; Key Laboratory of Gene Engineering of the Ministry of Education, 135 Xingang West Road, Haizhu District, Guangzhou 510275, Guangdong Province, China. lssyje@mail.sysu.edu.cn

Received: 2015-04-28
Revised: 2015-06-15
Accepted: 2015-06-19
Published online: 2015-11-08

Abstract

Short open reading frames (sORFs) are a common feature of genomes of human and other species, but their coding potential remains unknown. Innovations in proteomics and high-

throughput analyses of translation start sites have resulted in the identification of hundreds of putative coding sORFs, and some of them have been verified to be able to translated into short peptides (<100 amino acids). Moreover, recent findings reveal their diverse functions in various biological processes including development and differentiation. This review discusses the translation, identification and biological function of short peptides.

© 2015 Baishideng Publishing Group Inc. All rights reserved.

Key Words: Short peptides; Short open reading frames; Non-coding RNA

Tian Y, Yang JE. Emerging landscape of short open reading frame-encoded peptides. Shijie Huaren Xiaohua Zazhi 2015; 23(31): 4954-4960 URL: <http://www.wjgnet.com/1009-3079/23/4954.asp> DOI: <http://dx.doi.org/10.11569/wcjd.v23.i31.4954>

摘要

人类和其他生物的基因组中存在小开放阅读框(small open reading frame, sORF), 但过去对于这些sORFs是否具有编码能力还无从得知。随着蛋白质组学和高通量翻译起始位点分析方法的发展, 现已鉴定了上百种sORFs, 其中一部分被证实具有翻译能力, 可产生长度在100个氨基酸残基的短肽(short peptides, SEPs)。近年来的研究表明短肽在发育、分化等生物学过程中具有重要功能。本文系统阐述SEPs的翻译过程、鉴定方法和功能。

© 2015年版权归百世登出版集团有限公司所有。

■同行评议者

周南进, 研究员, 江西省医学科学研究所

关键词: 短肽; 小开放阅读框; 非编码RNA

核心提示: 短肽(short peptides)是由短的小开放阅读框(small open reading frame)翻译生成的肽链, 长度多在100个氨基酸以下. 从模式动物到人类, 均检测到具有生物学功能的短肽, 提示他们可能在人类生命活动及疾病的发生发展中扮演重要角色.

田原, 杨金娥. 短肽编码基因的研究进展. 世界华人消化杂志 2015; 23(31): 4954-4960 URL: <http://www.wjgnet.com/1009-3079/23/4954.asp> DOI: <http://dx.doi.org/10.11569/wjcd.v23.i31.4954>

0 引言

人类基因组计划揭示人的基因组共有30亿碱基, 但其中蛋白编码基因只有2.1万个左右, 仅占整个基因组的1.5%^[1], 提示基因组存在大量非蛋白编码区. 随后由美国发起的DNA元件百科全书(encyclopedia of DNA elements, ENCODE)计划旨在找到人类基因组中的所有功能性元件, 结果发现整个人类基因组中大约有76%的DNA片段可以被转录^[2], 在对这些转录产物进行分析发现部分转录产物其中含短的开放阅读框(short open reading frames, sORFs), 提示他们可能编码短的肽链^[3]. 近年来的研究结果显示, 从细菌、酵母等低等的模式动物到人类均检测到sORF可表达短肽(short peptides, SEPs)^[4-6], 并具有生物学功能^[7-9].

1 SEPs简介

SEPs是指由mRNA或非编码RNA(non coding RNA, ncRNA)上的sORF翻译生成的短肽, 长度多在100个氨基酸以下^[10]. 目前发现最短的SEP仅有6个氨基酸残基^[11], 位于S-腺苷甲硫氨酸脱羧酶(AdoMetDC)mRNA的5'端, 通过滞留mRNA上的核糖体进而抑制下游ORF的翻译, 降低AdoMetDC的表达水平, 从而调控动植物的多胺合成^[12].

1.1 SEPs的产生 sORF在基因组上的位置多种多样, 可以位于蛋白编码基因mRNA的5'端非翻译区(5' untranslated region, 5'UTR)^[13-15]、3'UTR区^[16], 或与已知ORF重叠或部分重叠^[17], 或位于非蛋白编码的RNA上^[11]. 位于mRNA 5'UTR区的sORFs也被称为上游ORF(upstream ORF, uORF)^[17]. 目前有两种模

型解释uORFs翻译产生SEPs的过程^[11]: 一种称为重启模型(re-initiation), 核糖体40S小亚基从mRNA的5'甲基鸟苷帽子结构开始扫描, 如果5'UTR区存在uORF, 40S小亚基结合到uORF的起始密码子上, 并进一步与60S大亚基结合, 启动uORF的翻译, 生成SEPs. 当uORF翻译结束后, 大亚基与小亚基解离, 小亚基仍与mRNA结合并继续向3'端扫描, 当遇到下一个起始密码子, 再次启动翻译. 另一种称为渗漏扫描模型(leaky scanning), 当核糖体40S小亚基从5'甲基鸟苷帽子结构开始扫描时, 一部分可以在uORF的起始密码子上与60S大亚基结合启动uORF的翻译, 一部分则忽略uORF的起始密码子, 继续扫描到下游ORF的起始密码子处, 与60S大亚基结合启动ORF的翻译.

虽然以上两种模型可以解释uORFs翻译SEPs的过程, 但无法解释位于mRNA的3'UTR区或与已知ORF重叠或部分重叠的sORFs是如何翻译出SEPs的. 有一种假设认为他们来自其他缩短形式的转录本^[17]. 研究人员通过质谱在白血病细胞系K562中鉴定出两个SEPs, DEDD2-SEP和H2AFX-SEP, 但当他们将全长的mRNA克隆到表达载体后, 却发现并不能正常表达SEPs, 而当克隆5'端缺失部分序列的转录本时, 可以正常表达. 提示DEDD2-SEP和H2AFX-SEP是由其他剪接形式的mRNA翻译生成的, 这些转录本由于在5'端缺失一部分序列, 位于编码序列(coding sequence, CDS)区的sORF就变成了靠近5'端的第一个ORF, 核糖体从5'端开始扫描这种缩短形式的mRNA就能翻译出SEP.

1.2 SEPs使用非经典起始密码子 通过核糖体图谱分析(ribosome profiling)发现, 小鼠胚胎干细胞中的sORFs中约45%的使用经典起始密码子AUG, 其余60%则使用非经典起始密码子, 主要为与AUG相似的起始密码子, 如CUG、GUG等, 占有sORFs的40%, 其余sORFs则使用非AUG起始密码子^[18]. 这些非经典起始密码子与经典起始密码子ATG同源, 常位于Kozak序列附近. 如在K562细胞中鉴定的FRAT2-SEP的起始密码子为非经典的ACG. 而当将ACG突变成ATG后, FRAT2-SEP的表达水平提高近3倍; 提示携带ACG起始密码子的sORFs的翻译效率低于含经典密码子的sORFs^[17]. 也有另

■ 研究前沿

有关SEPs的报道逐年增多, 人们对SEPs的关注度也不断升高. 进一步鉴定新的SEPs, 特别是低丰度的SEPs, 并对已发现的SEPs进行功能解析是未来SEPs研究的重点.

■ 相关报道

本文从产生、保守性、翻译特点、筛选方法、作用机制及生物学功能等多个方面介绍了sORF及其编码短肽的研究进展。国外Slavoff等创建了短肽组学与RNA-seq相结合的方法,首次系统鉴定了人类细胞中的sORF编码短肽。

一种理论认为, sORFs的非经典起始密码子来自于转录后mRNA的RNA编辑^[19], 即在RNA编辑酶的作用下, 将转录产物的起始密码子AUG中的尿嘧啶(U)转换成胞嘧啶(C), 使经典起始密码子转变为同源的非经典密码子, 从而调控sORFs的翻译效率。但目前尚无直接的实验证据揭示sORFs的起始密码子可发生RNA编辑。未来的研究可以通过基因敲除参与RNA编辑的酶后, 来分析其对sORFs的调控。

值得一提的是, 当sORFs与其邻近的ORFs共转录时, 二者形成了多顺反子, 可同时表达两种蛋白。研究人员在蛋白编码基因FRAT2的ORF及其上游的uORF(表达短肽FRAT2-SEP)的C端分别加上c-myc及Flag标签, 将含两种标签的FRAT2 cDNA克隆到表达载体, 可以同时检测到含两种标签的融合蛋白表达, 提示FRAT2 mRNA为双顺反子^[17]。

1.3 SEPs保守性 SEPs在不同物种中具有一定的保守性。Slavoff等^[17]在K562细胞株中鉴定了90个SEPs, 并在29种哺乳动物中分析他们的保守性, 发现SEPs的保守性高于内含子, 低于已知的蛋白编码基因, 且有功能的SEPs保守性更强。如最早在果蝇中发现的SEPs-Tal peptides是调控果蝇胚胎发育重要分子, 现已在其他昆虫中陆续发现了其同源物^[20]。此外, SEPs保守性与其长度有关, Aspden等^[21]根据SEPs的大小将他们分为两类, 一类是较长的SEPs, 占SEPs的83%, 约为80个氨基酸残基, 与已知蛋白具有相似的翻译过程和保守性, 并倾向于含有跨膜结构域。另一类是较短的SEPs, 在SEPs中所占比例小, 约为20个氨基酸残基, 大多数由mRNA的5'UTR区产生, 保守性较差, 翻译效率也较低。

不难想象含有跨膜结构域的SEPs在细胞内有特定的亚细胞定位。如在HEK293T、MEF和COS7等3种细胞中均发现, DEDD2-SEP定位于线粒体膜上^[7]; 而由ncRNA-pncr003:2L翻译出的两种SEPs-sarcolamban A和sarcolamban B, 定位于肌细胞的肌浆网上, 调节Ca²⁺运输和肌肉收缩, 并且在脊椎动物和非脊椎动物中都具有较高的保守性^[22]。

2 筛选鉴定SEPs的方法

目前生物信息学方法预测sORFs的研究还处在初级阶段, 多局限于在已知基因的ORF

附近寻找潜在的sORFs。常用的预测软件有sORFinder^[23]、HALTORF^[24]、uPEPPERoni^[25]等。此外, 通过分析基因序列特征, 如经典起始密码子、翻译终止位点、剪切位点、启动子序列、多聚腺苷酸信号、密码子偏好性、核苷酸组成等也可用于区分有无翻译能力的sORFs^[17]。由于保守性强的sORFs常意味着其有更重要的生物学功能, 因此检测sORFs的保守性也可以作为sORFs是否具有翻译产生功能性SEP的参考。

核糖体图谱分析(ribosome profiling)是一种以高通量DNA测序为基础的技术, 可以在全转录组范围内发现sORFs。其主要原理是用嘌呤霉素(puromycin)处理细胞, 使核糖体从mRNA上解离, 导致肽链延伸终止, 并结合放线菌酮处理, 使翻译起始核糖体停留于翻译起始位点, 再用RNA酶将未与核糖体结合的RNA酶解, 分离得到未被酶解的RNA后进行深度测序, 得到具有翻译潜能的mRNA^[26]。这种技术的优势在于无需根据基因组设计探针, 可以发现新的sORFs。并且嘌呤霉素不会干扰靠近翻译起始位点处延伸复合物的组装, 能较好检测到非经典起始密码子^[18]。但由于起始翻译除了需要核糖体外, 还需要其他起始因子参与, 而且许多核糖体结合到mRNA后并不起始翻译, 因此存在假阳性, 需结合生物信息学等方法进一步甄别^[17]。

另一种较为简单的SEPs鉴定方法是蛋白标签免疫印迹(epitope-tagged detection)^[27]。该方法将表达蛋白标签的序列重组到目的基因的终止密码子上游, 当目的基因表达蛋白时, 在羧基端就会带有蛋白标签, 之后运用免疫印记(Western blot)即可鉴定目的基因是否编码蛋白^[28]。该方法还可以研究uORF与下游ORF的共表达。虽然这种方法可直观地看到SEPs表达情况, 但难以高通量筛选和鉴定表达SEPs的sORFs, 适用于研究少量的sORFs及其表达水平和功能调控。

蛋白质谱也可用于鉴定sORFs翻译的SEPs^[29]。理论上只要细胞或组织内的蛋白丰度达到一定的临界值, 就可以通过电泳分离结合质谱鉴定, 在蛋白质数据库中与已知或预测的寡肽比对, 并进一步推算出对应的核苷酸序列。常用的质谱鉴定方法主要是液质联用串联质谱。串联质谱技术可减少或消除样品

基质中无关物的干扰, 提供丰富的结构信息. Skarszewski等^[25]通过该方法在K562细胞中分别鉴定了多个新的SEPs. 但由于细胞内SEPs的分子量小, 且丰度较低, 在分离和富集步骤中容易损失, 很难得到有效的质谱数据, 质谱鉴定SEPs的技术还需进一步优化和改进.

3 SEPs的作用机制

根据SEPs的定位可以将其分为两类. 第一类是由带有内质网信号肽的初级翻译产物加工而来的短肽. 信号肽引导多肽链定位到粗面内质网和高尔基体, 经加工、修饰后分泌到细胞外, 通过与细胞表面的分子受体相互作用, 介导信号的传导. 第二类SEPs由sORFs翻译而来, 不带有信号肽, 翻译完成后直接在胞质内发挥功能. 已有的研究表明, SEPs可以通过顺式作用的方式调控邻近ORF的翻译效率, 从而影响蛋白的合成^[30,31]、折叠^[32,33]和定位^[34,35]. 如前所述的FRAT2-SEP, 他通过还未阐明的机制抑制下游ORF的翻译^[17]. 类似的SEPs还包括人类SAMDC基因中的uORF, 他表达的SEPs长6个氨基酸残基, 可与多胺结合, 调控下游ORF的翻译. 也有文章报道SEPs通过反式作用的方式抑制其他基因的转录或翻译表达, 如在人类中发现的ASS1-SEP, 长44个氨基酸残基, 能反式调控ASS1基因的表达; 还有与人的EPHX1相关的两个SEPs, 分别长17和26个氨基酸残基, 他们可通过与翻译因子相互作用来抑制EPHX1的翻译.

由于已鉴定的人类SEPs中, 有相当一部分是由位于5'UTR的sORFs翻译而来^[36], 因此目前对uORF生成的SEPs研究得较为深入^[13]. 如哺乳动物β2肾上腺素受体mRNA上存在一个uORF, 能翻译出长19个氨基酸残基的短肽, 名为BUP. 将BUP的起始密码子AUG突变为CTT后, 发现β2肾上腺素受体在COS-7细胞中的表达水平比野生型提高1.9倍. 同时, 在体外用人工合成的BUP处理细胞后也能达到抑制效果^[37]. 研究人员还在编码谷氨酰胺酶亚基蛋白的mRNA上发现一个含有25个密码子的uORF, 该uORF能通过精氨酸途径抑制下游ORF的翻译^[38].

4 SEPs的生物学功能

目前人们对SEPs功能的了解多来自酵母、果

蝇等模式动物中研究的结果. 已有的研究表明, SEPs可以调控昆虫形态发育^[22]、植物的形态发生^[12]、脊椎动物的骨骼肌收缩等^[12]. 模式生物中对短肽簇tarsal-less (tal) peptides的研究较为深入^[39,40]. 他最早由Inagaki等^[41]在果蝇基因组中发现. Tal peptides的编码基因最初被认为是ncRNA, 随后Galindo等^[39]发现该基因上含有10个sORFs, 除sORF5不翻译, 其他9个sORFs可以翻译产生含11-32个氨基酸残基的短肽, 并且sORF1-4具有较高保守性, 可以独立翻译, 而sORF6-10可以被翻译但不具有生物活性. 研究人员发现Tal peptides能够促进Ovo/Svb蛋白的羧基端转录抑制域的水解, 使Ovo/Svb从转录抑制因子变为转录激活因子, 从而激活下游靶基因如miniature和shavenoid的表达, 促进果蝇形成表皮毛^[9]. 也有证据^[42]表明Tal peptides可以通过细胞间隙或穿过磷脂双层膜等非细胞自分泌途径影响相邻的细胞形态. 另一个例子是sarcolamban A和sarcolamban B, 由Magny等^[22]首次在果蝇心肌细胞内发现, 长度为28和29个氨基酸残基, 他们通过调节肌浆网内的Ca²⁺释放影响心肌细胞的收缩.

在植物中表达的小肽在植物形态发生过程中扮演重要角色^[7], 如在拟南芥幼苗根部高水平表达的POLARIS(PLS)基因的mRNA上有一个sORF, 可翻译出一个含36个氨基酸残基的短肽, 在PLS突变株中过表达这种短肽可以回复拟南芥的短根表型, 但机制尚未阐明. 其他调控植物形态发育的小肽还有Rotundifolia(RT4)、Brick1(Brk1)、Enod40^[16]等.

目前在人和鼠中对小肽的研究也已起步, 并已鉴定了一些具有生物学功能的小肽. 如Anderson等^[43]发现了一个长46个氨基酸的小肽MLN, 他可与内质网上的钙泵SERCA特异性结合, 抑制SERCA将Ca²⁺摄入内质网中, 从而调节骨骼肌收缩. 最近, Lee等^[8]研究发现, 线粒体12S rRNA编码的长16个氨基酸的短肽MOTS-c(mitochondrial open reading frame of the 12S rRNA-c)可通过Folate-AICAR-AMPK通路调节胰岛素敏感性, 维持代谢稳态. 此外, 关于SEPs与疾病的研究也有报道^[44-46], 如在1例患有阿尔茨海默病患者的cDNA文库中筛选到一个含有24个氨基酸残基的SEP, 命名

应用要点

SEPs的发现无疑是蛋白水平上的新突破, 他使人们认识到, 细胞内存在的小分子寡肽并不仅仅是“冗余蛋白”, 他是具有生物学意义的功能性分子, 可能在人类生命活动及疾病的发生发展中扮演重要角色.

同行评价

本文介绍了sORF及其编码SEPs的研究进展, 从其产生, 保守性, 翻译特点, 筛选方法, 作用机制及生物学功能等多个方面进行探讨, 基本反映了目前SEPs国内外的研究现状, 有一定的参考价值。

为Humanin。研究^[47]表明, Humanin能够逆转由多种家族性阿尔茨海默病相关基因突变所导致的神经细胞凋亡, 提示他具有治疗阿尔茨海默病的潜力。

5 结论

随着分子生物学和遗传学研究的进展, 生物高新技术的不断开发与应用, 科学家们发现越来越多的新型小分子也参与了细胞生命活动的调控, 如microRNA, lncRNA等。SEPs的发现无疑是蛋白水平上的新突破, 他使人们认识到, 细胞内存在的小分子寡肽并不仅仅是“冗余蛋白”, 而具有生物学意义的功能性分子, 他们可能在人类生命活动及疾病的发生发展中扮演重要角色。

目前关于SEPs的报道在逐年增多, 人们对SEPs的关注度也不断升高。但对SEPs的研究主要还停留在筛选和鉴定方面。随着质谱技术和RNA-seq的逐步改进, 使发现细胞内低丰度的SEPs成为可能。采用生物信息学^[6]和Ribosome profiling^[18]相结合的方法, 已在细菌^[48]、病毒^[49]、植物^[50,51]、酵母^[5]、昆虫^[39]及人类^[47,52]中已陆续发现了一些新的SEPs, 然而这些新SEPs的生物学功能还远未阐明。因此未来SEPs的研究一方面需要建立和优化发现SEPs的策略和技术, 另一方面需大力开展深入的功能研究, 探究其在SEPs生命活动作用。

6 参考文献

- Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, Tanzer A, Lagarde J, Lin W, Schlesinger F, Xue C, Marinov GK, Khatun J, Williams BA, Zaleski C, Rozowsky J, Röder M, Kokocinski F, Abdelhamid RF, Alioto T, Antoshechkin I, Baer MT, Bar NS, Batut P, Bell K, Bell I, Chakraborty S, Chen X, Chrast J, Curado J, Derrien T, Drenkow J, Dumais E, Dumais J, Dutttagupta R, Falconnet E, Fastuca M, Fejes-Toth K, Ferreira P, Foissac S, Fullwood MJ, Gao H, Gonzalez D, Gordon A, Gunawardena H, Howald C, Jha S, Johnson R, Kapranov P, King B, Kingswood C, Luo OJ, Park E, Persaud K, Preall JB, Ribeca P, Risk B, Robyr D, Sammeth M, Schaffer L, See LH, Shahab A, Skancke J, Suzuki AM, Takahashi H, Tilgner H, Trout D, Walters N, Wang H, Wrobel J, Yu Y, Ruan X, Hayashizaki Y, Harrow J, Gerstein M, Hubbard T, Reymond A, Antonarakis SE, Hannon G, Giddings MC, Ruan Y, Wold B, Carninci P, Guigó R, Gingeras TR. Landscape of transcription in human cells. *Nature* 2012; 489: 101-108 [PMID: 22955620 DOI: 10.1038/nature11233]
- Pennisi E. Genomics. ENCODE project writes eulogy for junk DNA. *Science* 2012; 337: 1159, 1161 [PMID: 22955811 DOI: 10.1126/science.337.6099.1159]
- Chu Q, Ma J, Saghatelian A. Identification and characterization of sORF-encoded polypeptides. *Crit Rev Biochem Mol Biol* 2015; 50: 134-141 [PMID: 25857697 DOI: 10.3109/10409238.2015.1016215]
- Hanada K, Zhang X, Borevitz JO, Li WH, Shiu SH. A large number of novel coding small open reading frames in the intergenic regions of the Arabidopsis thaliana genome are transcribed and/or under purifying selection. *Genome Res* 2007; 17: 632-640 [PMID: 17395691]
- Kastenmayer JP, Ni L, Chu A, Kitchen LE, Au WC, Yang H, Carter CD, Wheeler D, Davis RW, Boeke JD, Snyder MA, Basrai MA. Functional genomics of genes with small open reading frames (sORFs) in *S. cerevisiae*. *Genome Res* 2006; 16: 365-373 [PMID: 16510898]
- Frith MC, Forrest AR, Nourbakhsh E, Pang KC, Kai C, Kawai J, Carninci P, Hayashizaki Y, Bailey TL, Grimmond SM. The abundance of short proteins in the mammalian proteome. *PLoS Genet* 2006; 2: e52 [PMID: 16683031]
- Hashimoto Y, Kondo T, Kageyama Y. Lilliputians get into the limelight: novel class of small peptide genes in morphogenesis. *Dev Growth Differ* 2008; 50 Suppl 1: S269-S276 [PMID: 18459982 DOI: 10.1111/j.1440-169X.2008.00994.x]
- Lee C, Zeng J, Drew BG, Sallam T, Martin-Montalvo A, Wan J, Kim SJ, Mehta H, Hevener AL, de Cabo R, Cohen P. The mitochondrial-derived peptide MOTS-c promotes metabolic homeostasis and reduces obesity and insulin resistance. *Cell Metab* 2015; 21: 443-454 [PMID: 25738459 DOI: 10.1016/j.cmet.2015.02.009]
- Kondo T, Plaza S, Zanet J, Benrabah E, Valenti P, Hashimoto Y, Kobayashi S, Payre F, Kageyama Y. Small peptides switch the transcriptional activity of Shavenbaby during Drosophila embryogenesis. *Science* 2010; 329: 336-339 [PMID: 20647469 DOI: 10.1126/science.1188158]
- Ingolia NT. Ribosome profiling: new views of translation, from single codons to genome scale. *Nat Rev Genet* 2014; 15: 205-213 [PMID: 24468696 DOI: 10.1038/nrg3645]
- Andrews SJ, Rothnagel JA. Emerging evidence for functional peptides encoded by short open reading frames. *Nat Rev Genet* 2014; 15: 193-204 [PMID: 24514441 DOI: 10.1038/nrg3520]
- Law GL, Raney A, Heusner C, Morris DR. Polyamine regulation of ribosome pausing at the upstream open reading frame of S-adenosylmethionine decarboxylase. *J Biol Chem* 2001; 276: 38036-38043 [PMID: 11489903]
- Calvo SE, Pagliarini DJ, Mootha VK. Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. *Proc Natl Acad Sci U S A* 2009; 106: 7507-7512 [PMID: 19372376 DOI: 10.1073/pnas.0810916106]
- Abastado JP, Miller PF, Hinnebusch AG. A quantitative model for translational control of the GCN4 gene of *Saccharomyces cerevisiae*. *New Biol*

- 1991; 3: 511-524 [PMID: 1883814]
- 15 Kozak M. Bifunctional messenger RNAs in eukaryotes. *Cell* 1986; 47: 481-483 [PMID: 3779834]
- 16 Sousa C, Johansson C, Charon C, Manyani H, Sautter C, Kondorosí A, Crespi M. Translational and structural requirements of the early nodulin gene *enod40*, a short-open reading frame-containing RNA, for elicitation of a cell-specific growth response in the alfalfa root cortex. *Mol Cell Biol* 2001; 21: 354-366 [PMID: 11113209]
- 17 Slavoff SA, Mitchell AJ, Schwaid AG, Cabili MN, Ma J, Levin JZ, Karger AD, Budnik BA, Rinn JL, Saghatelian A. Peptidomic discovery of short open reading frame-encoded peptides in human cells. *Nat Chem Biol* 2013; 9: 59-64 [PMID: 23160002 DOI: 10.1038/nchembio.1120]
- 18 Ingolia NT, Lareau LF, Weissman JS. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* 2011; 147: 789-802 [PMID: 22056041 DOI: 10.1016/j.cell.2011.10.002]
- 19 Wedekind JE, Dance GS, Sowden MP, Smith HC. Messenger RNA editing in mammals: new members of the APOBEC family seeking roles in the family business. *Trends Genet* 2003; 19: 207-216 [PMID: 12683974]
- 20 Savard J, Marques-Souza H, Aranda M, Tautz D. A segmentation gene in *tribolium* produces a polycistronic mRNA that codes for multiple conserved peptides. *Cell* 2006; 126: 559-569 [PMID: 16901788]
- 21 Aspden JL, Eyre-Walker YC, Phillips RJ, Amin U, Mumtaz MA, Brocard M, Couso JP. Extensive translation of small Open Reading Frames revealed by Poly-Ribo-Seq. *Elife* 2014; 3: e03528 [PMID: 25144939 DOI: 10.7554/eLife.03528]
- 22 Magny EG, Pueyo JI, Pearl FM, Cespedes MA, Niven JE, Bishop SA, Couso JP. Conserved regulation of cardiac calcium uptake by peptides encoded in small open reading frames. *Science* 2013; 341: 1116-1120 [PMID: 23970561 DOI: 10.1126/science.1238802]
- 23 Hanada K, Akiyama K, Sakurai T, Toyoda T, Shinozaki K, Shiu SH. sORF finder: a program package to identify small open reading frames with high coding potential. *Bioinformatics* 2010; 26: 399-400 [PMID: 20008477 DOI: 10.1093/bioinformatics/btp688]
- 24 Vanderperre B, Lucier JF, Roucou X. HAltORF: a database of predicted out-of-frame alternative open reading frames in human. *Database (Oxford)* 2012; 2012: bas025 [PMID: 22613085 DOI: 10.1093/database/bas025]
- 25 Skarszewski A, Stanton-Cook M, Huber T, Al Mansoori S, Smith R, Beatson SA, Rothnagel JA. uPEPPERoni: an online tool for upstream open reading frame location and analysis of transcript conservation. *BMC Bioinformatics* 2014; 15: 36 [PMID: 24484385 DOI: 10.1186/1471-2105-15-36]
- 26 Fritsch C, Herrmann A, Nothnagel M, Szafranski K, Huse K, Schumann F, Schreiber S, Platzer M, Krawczak M, Hampe J, Brosch M. Genome-wide search for novel human uORFs and N-terminal protein extensions using ribosomal footprinting. *Genome Res* 2012; 22: 2208-2218 [PMID: 22879431 DOI: 10.1101/gr.139568.112]
- 27 Hemm MR, Paul BJ, Schneider TD, Storz G, Rudd KE. Small membrane proteins found by comparative genomics and ribosome binding site models. *Mol Microbiol* 2008; 70: 1487-1501 [PMID: 19121005 DOI: 10.1111/j.1365-2958.2008.06495.x]
- 28 Babu M, Butland G, Pogoutse O, Li J, Greenblatt JF, Emili A. Sequential peptide affinity purification system for the systematic isolation and identification of protein complexes from *Escherichia coli*. *Methods Mol Biol* 2009; 564: 373-400 [PMID: 19544035 DOI: 10.1007/978-1-60761-157-8_22]
- 29 Oyama M, Kozuka-Hata H, Suzuki Y, Semba K, Yamamoto T, Sugano S. Diversity of translation start sites may define increased complexity of the human short ORFeome. *Mol Cell Proteomics* 2007; 6: 1000-1006 [PMID: 17317662]
- 30 Darnell JC, Van Driesche SJ, Zhang C, Hung KY, Mele A, Fraser CE, Stone EF, Chen C, Fak JJ, Chi SW, Licatalosi DD, Richter JD, Darnell RB. FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. *Cell* 2011; 146: 247-261 [PMID: 21784246 DOI: 10.1016/j.cell.2011.06.013]
- 31 Morris DR, Geballe AP. Upstream open reading frames as regulators of mRNA translation. *Mol Cell Biol* 2000; 20: 8635-8642 [PMID: 11073965]
- 32 Kimchi-Sarfaty C, Oh JM, Kim IW, Sauna ZE, Calcagno AM, Ambudkar SV, Gottesman MM. A "silent" polymorphism in the MDR1 gene changes substrate specificity. *Science* 2007; 315: 525-528 [PMID: 17185560]
- 33 Zhang G, Hubalewska M, Ignatova Z. Transient ribosomal attenuation coordinates protein synthesis and co-translational folding. *Nat Struct Mol Biol* 2009; 16: 274-280 [PMID: 19198590 DOI: 10.1038/nsmb.1554]
- 34 Mariappan M, Li X, Stefanovic S, Sharma A, Mateja A, Keenan RJ, Hegde RS. A ribosome-associating factor chaperones tail-anchored membrane proteins. *Nature* 2010; 466: 1120-1124 [PMID: 20676083 DOI: 10.1038/nature09296]
- 35 Yanagitani K, Kimata Y, Kadokura H, Kohno K. Translational pausing ensures membrane targeting and cytoplasmic splicing of XBP1u mRNA. *Science* 2011; 331: 586-589 [PMID: 21233347 DOI: 10.1126/science.1197142]
- 36 Ma J, Ward CC, Jungreis I, Slavoff SA, Schwaid AG, Neveu J, Budnik BA, Kellis M, Saghatelian A. Discovery of human sORF-encoded polypeptides (SEPs) in cell lines and tissue. *J Proteome Res* 2014; 13: 1757-1765 [PMID: 24490786 DOI: 10.1021/pr401280w]
- 37 Parola AL, Kobilka BK. The peptide product of a 5' leader cistron in the beta 2 adrenergic receptor mRNA inhibits receptor synthesis. *J Biol Chem* 1994; 269: 4497-4505 [PMID: 8308019]
- 38 Werner M, Feller A, Messenguy F, Piérard A. The leader peptide of yeast gene CPA1 is essential for the translational repression of its expression. *Cell* 1987; 49: 805-813 [PMID: 3555844]
- 39 Galindo MI, Pueyo JI, Fouix S, Bishop SA, Couso JP. Peptides encoded by short ORFs control development and define a new eukaryotic gene

- family. *PLoS Biol* 2007; 5: e106 [PMID: 17439302]
- 40 Kondo T, Hashimoto Y, Kato K, Inagaki S, Hayashi S, Kageyama Y. Small peptide regulators of actin-based cell morphogenesis encoded by a polycistronic mRNA. *Nat Cell Biol* 2007; 9: 660-665 [PMID: 17486114]
- 41 Inagaki S, Numata K, Kondo T, Tomita M, Yasuda K, Kanai A, Kageyama Y. Identification and expression analysis of putative mRNA-like non-coding RNA in *Drosophila*. *Genes Cells* 2005; 10: 1163-1173 [PMID: 16324153]
- 42 Joliet A, Prochiantz A. Transduction peptides: from technology to physiology. *Nat Cell Biol* 2004; 6: 189-196 [PMID: 15039791 DOI: 10.1038/ncb0304-189]
- 43 Anderson DM, Anderson KM, Chang CL, Makarewich CA, Nelson BR, McAnally JR, Kasaragod P, Shelton JM, Liou J, Bassel-Duby R, Olson EN. A micropeptide encoded by a putative long noncoding RNA regulates muscle performance. *Cell* 2015; 160: 595-606 [PMID: 25640239 DOI: 10.1016/j.cell.2015.01.009]
- 44 Wiestner A, Schlemper RJ, van der Maas AP, Skoda RC. An activating splice donor mutation in the thrombopoietin gene causes hereditary thrombocythaemia. *Nat Genet* 1998; 18: 49-52 [PMID: 9425899]
- 45 Liu L, Dilworth D, Gao L, Monzon J, Summers A, Lassam N, Hogg D. Mutation of the CDKN2A 5' UTR creates an aberrant initiation codon and predisposes to melanoma. *Nat Genet* 1999; 21: 128-132 [PMID: 9916806]
- 46 Wen Y, Liu Y, Xu Y, Zhao Y, Hua R, Wang K, Sun M, Li Y, Yang S, Zhang XJ, Kruse R, Cichon S, Betz RC, Nöthen MM, van Steensel MA, van Geel M, Steijlen PM, Hohl D, Huber M, Dunnill GS, Kennedy C, Messenger A, Munro CS, Terrinoni A, Hovnanian A, Bodemer C, de Prost Y, Paller AS, Irvine AD, Sinclair R, Green J, Shang D, Liu Q, Luo Y, Jiang L, Chen HD, Lo WH, McLean WH, He CD, Zhang X. Loss-of-function mutations of an inhibitory upstream ORF in the human hairless transcript cause Marie Unna hereditary hypotrichosis. *Nat Genet* 2009; 41: 228-233 [PMID: 19122663 DOI: 10.1038/ng.276]
- 47 Hashimoto Y, Niikura T, Tajima H, Yasukawa T, Sudo H, Ito Y, Kita Y, Kawasumi M, Kouyama K, Doyu M, Sobue G, Koide T, Tsuji S, Lang J, Kurokawa K, Nishimoto I. A rescue factor abolishing neuronal cell death by a wide spectrum of familial Alzheimer's disease genes and Abeta. *Proc Natl Acad Sci U S A* 2001; 98: 6336-6341 [PMID: 11371646]
- 48 Wadler CS, Vanderpool CK. A dual function for a bacterial small RNA: SgrS performs base pairing-dependent regulation and encodes a functional polypeptide. *Proc Natl Acad Sci U S A* 2007; 104: 20454-20459 [PMID: 18042713]
- 49 Jay G, Nomura S, Anderson CW, Khoury G. Identification of the SV40 agnogene product: a DNA binding protein. *Nature* 1981; 291: 346-349 [PMID: 6262654]
- 50 Casson SA, Chilley PM, Topping JF, Evans IM, Souter MA, Lindsey K. The POLARIS gene of *Arabidopsis* encodes a predicted peptide required for correct root growth and leaf vascular patterning. *Plant Cell* 2002; 14: 1705-1721 [PMID: 12172017]
- 51 Rohrig H, Schmidt J, Miklashevichs E, Schell J, John M. Soybean ENOD40 encodes two peptides that bind to sucrose synthase. *Proc Natl Acad Sci U S A* 2002; 99: 1915-1920 [PMID: 11842184]
- 52 Gleason CA, Liu QL, Williamson VM. Silencing a candidate nematode effector gene corresponding to the tomato resistance gene Mi-1 leads to acquisition of virulence. *Mol Plant Microbe Interact* 2008; 21: 576-585 [PMID: 18393617 DOI: 10.1094/MPMI-21-5-0576]

编辑: 郭鹏 电编: 都珍珍

