张弓，男，博士，暨南大学生命科学与技术学院教授，博士生导师，翻译组学实验室负责人。现为国家优秀青年基金获得者、国家863青年科学家、国家万人计划青年拔尖人才、中国蛋白质组学专业委员会(CNHUPO)理事。主要研究领域为翻译组学、生物信息学与系统生物学。

首次实现高等生物翻译中mRNA全长测序、翻译速率全局测定、tRNA组全定量测序，将中心法则定量化，由此建立翻译组学新学科领域。发现翻译速率调控是决定蛋白质折叠的关键因素之一，对1972年诺贝尔化学奖理论“安芬森原则”作出颠覆性更新，并以此将蛋白质可溶性表达效率提高数千倍。

研发FANSe系列超高精度大规模测序序列比对算法，同类算法中精度最高，有效解决大规模测序的可验证性、可重复性问题；建立了第一个高度实用化的基因测序分析云平台，在多个临床案例中发挥关键作用，受邀作TED演讲。其指导的高中生在Molecular Biosystems杂志上发表研究论文并被作为当期封面文章。课题组测定了中国缅甸远征军遗骸的全基因组序列，被境内外多家媒体专访报道。

翻译组学：方法及应用

赵 晶，张 弓*

(暨南大学生命与健康工程研究院，功能蛋白质广东省普通高校重点实验室，广州 510632)

摘要：所有产生生物学功能的蛋白质都必须由翻译过程来合成，翻译调控是中心法则诸多层次调控中最重要、最显著的调控环节，其对蛋白质组的调控幅度超过其他所有调控层次的总和。因此对翻译进行总体的研究非常必要，涉及生命科学的方方面面。翻译组学是总体上研究细胞内翻译状况的新兴学科，已在蛋白质组、肿瘤研究、微生物抗逆研究等方面做出了全新的贡献，极大地拓展了人们在翻译调控和蛋白质合成方面的视野。本文综述了近年来研究翻译组的主要方法，及其在基础研究与工程应用方面的巨大潜力。

关键词：翻译组；RNC-mRNA；蛋白质组；翻译调控

Translatomics: methods and applications

ZHAO Jing, ZHANG Gong*

(Key Laboratory of Functional Protein Research of Guangdong Higher Education Institutes, Institute of Life and Health Engineering, Jinan University Guangzhou510632, China)

Abstract: All functional proteins are synthesized via translation process. Translational regulation is the most important and significant step in the flow of Central Dogma. Its regulatory amplitude exceeds the sum of all other regulatory steps. Therefore, it is essential to investigate the translation in global scale, as it involves in

收稿日期：2016-06-28

基金项目：国家高技术研究发展计划(863计划)青年科学家专题项目(SS2014AA020504)；国家自然科学基金项目(81322028)；中国博士后科学基金项目(2016M592601)

***通信作者：**E-mail: zhanggong-uni@qq.com

all aspects of life. Translatomics is an emerging discipline for such investigations. It has exhibited its unique and innovative power on proteomics, cancer research and stress response of bacteria, expanding the view for the impact of protein synthesis. This review summarizes the main methods of translatomics and the applications on research and industry.

Key Words: translatome; RNC-mRNA; proteome; translation regulation

蛋白质是几乎所有生命活动的实际执行者。一个细胞内通常存在少则数千多则数万种不同的蛋白质，在精细地调控下共同完成复杂的生理过程。中心法则是现代生物学中最重要最基本的规律之一，阐述了遗传信息如何从DNA经过转录生成RNA，然后翻译生成蛋白质的生物信息转移途径[1]。根据中心法则，从基因组生成整个蛋白质组需要有RNA生成调控(含表观遗传调控和转录调控)、RNA降解调控、蛋白质生成调控(即翻译调控)、蛋白质降解调控这四个主要调控阶段。其中，翻译调控占所有调控比例的一半以上，超过其他所有调控的总和，是细胞内最重要的调控方式[2]。近年来，针对其他三个调控阶段的研究都有大量成果发表，已全面走向组学研究：研究全体基因的有“基因组学”，研究全体转录产物的有“转录组学”，研究全体蛋白质的有“蛋白质组学”。照此逻辑，研究翻译的总体也理应有“翻译组学”。但由于研究方法发展缓慢，这一提法只是近年来才少量地出现[3]。而对于翻译调控这一“调控之王”整体性的研究的缺失，说明现有生命过程的全局性研究还未触及主要矛盾，对中心法则这一生物核心信息流的研究仍存在严重的断层。

对翻译调控的全局研究之所以少见，主要有两个原因：一是参与翻译过程的生物大分子复杂多样，核酸和蛋白质都有参与调控，因而研究翻译需要联合多个学科的知识和方法，对研究者的知识技术背景提出了很高的要求；二是翻译机器十分精密和脆弱，翻译调控系统的变化往往发生在分钟级别，这给各种分析手段的发展带来了极大的困难。近年来，相关组学技术相继被突破，翻译的各部分均能被有效地进行全局研究，形成了“翻译组学”这一新兴学科。可喜的是，这些突破主要是由中国科学家贡献的。

本文将对翻译组学这一新兴学科的研究对象、主要研究方法及其在生物学和应用上的重要意义做一简要介绍。

1 翻译组学研究对象和主要研究技术

广义的翻译组(translatome)指直接参与翻译过程的所有元件，包括但不限于核糖体、正在翻译的mRNA(translating mRNA，又称RNC-mRNA)、tRNA、调控性RNA(如miRNA、lncRNA等)、新生肽链(nascent polypeptide chain)、各种翻译因子等；狭义的翻译组特指正在翻译的mRNA。本文所述翻译组为广义的翻译组。由于参与翻译过程的元件包含RNA和蛋白质两大类，并由复杂而精细的大分子机器调控，兼具静态构成和动态变化的特性。针对不同的组分，翻译组学发展出了多种多样的研究方法。

1.1 获取正在翻译的mRNA的方法

mRNA是蛋白质合成的信息蓝图，研究正在翻译的mRNA就是翻译组研究的首要任务。由于核糖体和RNA的结合为非共价结合，整个大分子复合体构象非常脆弱，在组织细胞破碎后很容易因为物理震动或化学环境变化而解离或被降解，因而研究难度很大。历史上先后出现的技术有多聚核糖体分析技术(polysome profiling)、核糖体亲和纯化技术(translating ribosome affinity purification, TRAP)、核糖体足迹谱技术(ribosome profiling)和翻译谱分析技术(translatome profiling)。

Polysome profiling出现于20世纪60年代，利用核糖体沉降系数较大的特性，用蔗糖密度梯度离心的方法分离多聚核糖体。由于核糖体是细胞里密度最高的生物大分子，一条mRNA上结合的核糖体越多，在蔗糖密度梯度离心时的沉降速率越快，因此结合不同数量核糖体的mRNA通过离心在溶液中会被分开。离心结束后将蔗糖溶液由底部缓慢抽出，即可对结合有不同数量核糖体的组分进行分离，可以直观地反映细胞内翻译的状态。在环境胁迫时，翻译的全局状态常有非常明显的改变，如在高渗压力下单个核糖体的组分明显增加[4-5]，而在氧化压力下单个mRNA上结合的

核糖体数量呈明显增加[5]。如有必要，可以对各组分中的mRNA进行分析[6]。

值得一提的是，人们一度认为活跃翻译的mRNA通常结合着多个核糖体进行翻译，而非活跃的mRNA结合的核糖体很少。然而近年来的研究表明，即便只结合了单个的核糖体，翻译也照样可能是活跃的。例如，在公认翻译非常活跃的HEK293细胞和处于富营养培养基上对数生长的大肠杆菌中，其单核糖体组分(原核为70S，真核为80S)都占绝对主要地位[5,7]；将单核糖体组分分离出来后置入无细胞翻译系统中，其核糖体可以继续翻译过程生成蛋白质[8]。这些都证明翻译的活跃程度与mRNA上核糖体的数量并无关系。

Polysome profiling技术的最大缺点是难以对翻译中mRNA(即RNC-mRNA)进行全面的分析。由于蔗糖密度离心的体积通常较大，造成了其中RNC-mRNA浓度很低；而溶液中高浓度的蔗糖，也使RNC-mRNA回收率很低。该种方法回收所得RNC-mRNA量一般仅够做PCR定量，但很难达到芯片或测序等全谱分析所需样品量，因而用这种方法进行RNC-mRNA的全谱分析有一定局限性[9]。

RNC-mRNA直接分离技术则有效地规避了这个问题。既然翻译活跃程度与核糖体数量无关，那直接把核糖体组分离心下来岂不省事？RNC-mRNA直接分离技术使用30%的单一浓度蔗糖溶液作为缓冲液，通过超速冷冻离心将核糖体组分与游离mRNA以及其他细胞组份分离开来。由于所有的核糖体组分全部被压在管底的沉淀中，弃去上清并保留沉淀抽提RNA，即可得到RNC-RNA。利用高通量测序技术等对获得的RNC-mRNA进行分析，即可得到特定翻译状态下所有正在翻译的全长mRNA信息，这种测序方法又被称为RNC-seq，通过分析可得RNC-mRNA的丰度与类型。这样有效解决了polysome profiling体积大浓度低的问题，排除了高浓度蔗糖的干扰，回收率高；而且这样得到的核糖体-新生肽链复合物仍然保留翻译活性，给予合适缓冲液即可恢复蛋白翻译能力[8]，可兼容后续多种研究方案。2013年本课题组首次报道了高等哺乳动物的全长RNC-seq[10]。RNC-seq的技术难点在于RNC-mRNA复合体的分离，因为RNC-mRNA复合体极为脆弱，非常容易在处理过程中造成核糖体解离或mRNA断裂，处理不当极易得不到全长RNC-mRNA，造成RNC-mRNA定量失准。

2009年由Weissman课题组首次发表的ribosomal profiling技术(又称Ribo-seq)[11]，用低浓度RNase处理核糖体-新生肽链复合物，降解掉没有核糖体覆盖的mRNA片段，再去除核糖体，利用二代测序技术检测被核糖体保护的约22~30 bp的RNA小片段(称为ribosome footprints, RFPs，又称ribosome protected fragments, RPFs，两者等价)，可得到核糖体分布的位置信息，据此可得到每种转录本上核糖体的分布，可推测起始密码子位置(包括非ATG起始)以及uORFs等信息[12]，这是RNC-seq所做不到的。但该技术的弊端在于操作复杂、测序偏好性大、成本高的问题；而且由于影响RFP数量的因素众多，包括进入翻译的mRNA数量、翻译起始效率和翻译延伸速率等，若翻译完全被阻滞在某一mRNA上的某处时，则该mRNA某区域上的RFP将会非常集中，因此得知RFP的数量并不能直接反映翻译的活跃程度。而且由于该方法只能分析有核糖体结合的蛋白编码区域(coding sequence, CDS)，因此也不能获得与翻译调控高度相关的非翻译区段(untranslated region, UTR)的信息。

Ribosome affinity purification(RAP, 或称translating RAP, TRAP)由Gerber课题组2002年报道，通过用组织特异性启动子表达结合核糖体大亚基的L25(Rpl25p)蛋白质，并在C端连接亲和标签(Affinity Tag，例如His，GFP等)，转化细胞，然后使用抗体将含有标签的核糖体与其他组织的核糖体分离，同时可获得结合在这类核糖体上正在翻译的mRNA[13]。该方法可以特异富集复杂样品中难以分离组织的RNC-mRNA进行后续分析而无需考虑来自其他细胞或组织的污染[14]。但TRAP需要对细胞系进行稳转使其表达含标签的核糖体蛋白，针对动植物个体则需要先构建转基因动植物，流程繁琐、耗时冗长，且并不能适用尚未建立稳定转化体系的物种[14-15]。而且过表达核糖体蛋白会干扰正常生理状况，嵌有融合蛋白的核糖体的翻译性能也与正常的核糖体不同，因而不可能准确反映正常生理状况下的翻译状况，因此，这种技术的应用有一定局限性。但在难以分离特定组织细胞的复杂组织中，TRAP技术有其独特优势[16]。

以上技术的特性比较见表1。

表1 翻译组研究技术对比

比较项目	Polysome profiling	RNC-seq	Ribo-seq	TRAP
RNC-mRNA的回收	全长	全长	仅核糖体覆盖片段	全长
翻译中mRNA回收难度	体积大，高糖，难回收	体积小，易回收	/	体积小，易回收
可采用的高通量测定方法	芯片、测序	芯片、测序	测序	芯片、测序
测序通量要求	低(多组分)	低(单组分)	高	低
Reads长度	任意	任意	22~30 nt	任意
检测序列变异	容易	容易	较难	容易
获取UTR、翻译调控区信息	可以	可以	不能	可以
获取核糖体位置、密度、阅读框位置、翻译暂停区域、uORFs、起始密码子检测	不能	不能	能	不能
获取同一mRNA上的核糖体数量	能	不能	不能	不能
获取组织特异性的翻译信息	需先分离组织	需先分离组织	需先分离组织	用组织特异性启动子表达带标签的核糖体蛋白
生理条件	可以	可以	可以	不能
操作简便性	较容易	简便	复杂	很复杂
样品量要求	大量	少量	大量	少量
杂质污染	不容易	不容易	容易	不容易
主要技术挑战	蔗糖密度的线性度，离心条件的控制	不加RNase抑制剂下的降解	酶切条件重复性	转基因模型，抗体下拉，降解

1.2 获取tRNA组信息的方法

tRNA在翻译过程中负责识别mRNA上的密码子并转运相应的氨基酸，是mRNA与蛋白质之间的桥梁，也是翻译组学中非常重要的一个组成部分。tRNA的种类与浓度高度影响着蛋白质合成的种类与翻译速度[5,17-18]。由于tRNA种类繁多(hg19人基因组中有610个tRNA基因，斑马鱼中甚至高达12 292个)[19]、物理化学特征极为相似、碱基修饰严重、二级和三级结构极为稳定等因素，tRNA组的全定量测定一直是个难题，成为细胞中最后一类尚未被完整测序的RNA分子。1996年和1999年，通过二维电泳的方式，Dong等[20]和Kanaya等[21]分别分离并定量测定了大肠杆菌和枯草芽孢杆菌中的几十种tRNA，实现了较为简单的原核生物的tRNA组测定，但对于tRNA种类高达几百甚至上万的真核生物则无能为力。2015年，暨南大学Zhong等[5]利用含有特异CCA尾巴的引物进行tRNA组的特异反转录并进行了二代测序，配合高容错性的算法FANSe2[22]，首次在单碱基分辨率水平实现了tRNA组的全定量测序技术。之后，Zheng等[23]和Cozen等[24]分别独自利用来自细菌的经过工程改造的去甲基化酶AlkB处理tRNA组降低修饰碱基含量，也分别实现了高分辨率的全长tRNA组测序与定量。这些新的测序方法使得定量tRNA测序变得切实可行，使得我们得以全面监控tRNA的种类以及其表达水平。

tRNA的各种修饰也是其完成生物学功能的必要结构，而tRNA也是目前已知的修饰最多的RNA种类，70~90个碱基的长度中可以有多达十余个碱基被修饰，所有的修饰种类加起来超过80种。这些修饰的异常可能会导致翻译的重大失调，甚至导致疾病[25]。利用质谱技术可以对tRNA修饰进行全局性研究：将tRNA酶解成更小的片段后，进行LC-MS/MS质谱分析，与tRNA一级序列相比对，即可知晓其修饰碱基及修饰类型[26]。

1.3 新生肽链折叠状态的测定

由于现阶段蛋白质质谱技术的精度仍然远远落后于测序技术，达不到解析新生肽链序列的能力；

同样因为蛋白质组学手段的局限，现阶段尚无全局性检测新生肽链构象的方法，因此目前只能对单一蛋白质新生肽链构象进行研究。较为通用简便的方法是有限酶切法，即将核糖体-新生肽链复合物(ribosome nascent-chain complex, RNC)分离，然后使用proteinase K等特异性较低的蛋白酶，在低浓度或低温下处理，蛋白构象较为松散的区段被切碎，而折叠稳定的区段则难以被酶切而保留完整的片段。通过western blot或放射自显影的方法即可观察到该蛋白质是否已在核糖体上形成了稳定的结构[8, 18]。

对于RNC中新生肽链的精细结构的解析则困难得多。由于核糖体体积巨大，难以通过结晶衍射的方法获得高分辨率的新生肽链结构；而冰冻蚀刻方法也达不到足够的分辨率；另外，由于翻译延伸速率有一定的随机性[4,8,18]，在一个细胞群体中同时存在着各种不同构象并动态变化的新生肽链，因而难以进行三维重构。核磁共振(NMR)可以测定溶液中生物大分子的动态构象，较为适合解析新生肽链的结构，但必须排除来自核糖体的干扰。2006年，Christodoulou研究组使用NMR解析了Ig2蛋白的新生肽链构象[27]，2016年同一研究组通过在重水中培养细菌并使用^{1}H和^{15}N进行代谢标记，分离纯化出含有目标蛋白质的大量RNC，再使用NMR进行解析，避开了核糖体的信号干扰，成功研究了α-synuclein的新生肽链构象[28]。

主流翻译组学研究技术路线如图1所示：

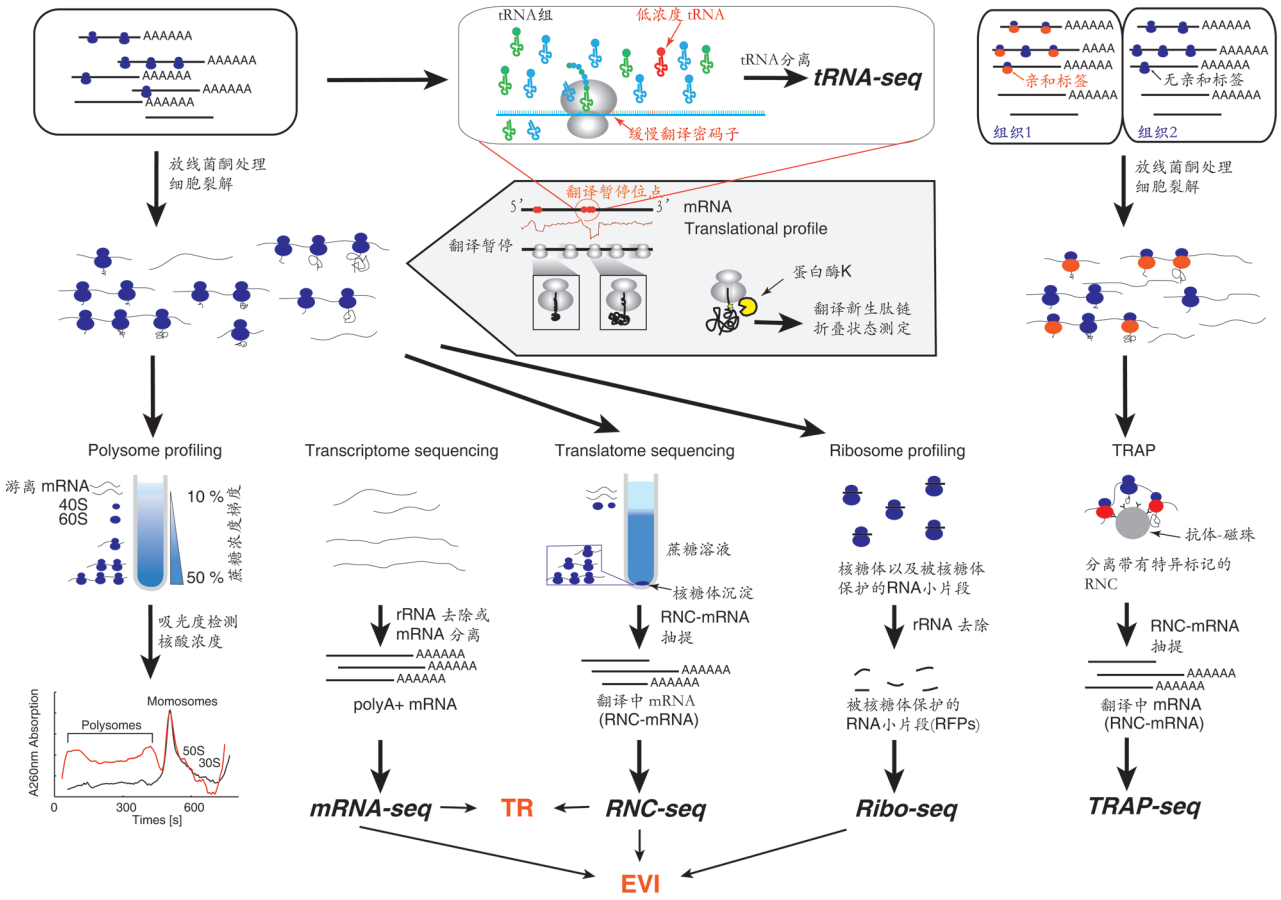

图1　翻译组研究技术区别与联系示意图

2　翻译组学对生物学基础理论的贡献

2.1　定量化中心法则

由于翻译调控在中心法则的信息传递中占据绝对主要地位，但是各个物种中的mRNA丰度和蛋白质丰度相关性R^2普遍较差，仅有0.01~0.5[29]，从mRNA到蛋白质的定量传递关系一直没有得到阐明。2013年，张弓课题组[10]利用翻译组学研究方法首次发现稳态细胞中正在翻译的mRNA相对

含量、mRNA长度和蛋白质相对含量之间的三元线性关系，R^2达到了0.94。这标志着中心法则提出55年后终于推进到定量化的阶段，打通了中心法则中基因信息流从转录本到功能蛋白质定量的关键步骤，同时也实现了用研究核酸的方法研究和定量蛋白质组。由于大规模测序的灵敏度和精度远高于定量蛋白质组学方法，因而在稳态细胞中可用翻译组测序取代蛋白质组定量，大大提高检测精度的同时降低成本。

2.2 翻译暂停决定蛋白质的折叠

细胞中直接翻译出来的多肽链必须折叠成特定的三维构象才能发挥功能，否则不仅失去功能，还会给细胞带来各种负面效应，甚至导致疾病。1972年获得诺贝尔化学奖的“安芬森原则(Anfinsen’s Dogma)”指出：蛋白质的一级结构决定其三维结构构象。根据安芬森原则，一般认为沉默突变不改变氨基酸序列，也就不会影响蛋白质的折叠。然而，同一氨基酸可由不同密码子编码，而在细胞中不同密码子对应的tRNA浓度相差可达一个数量级以上，自然翻译延伸速率也会受到影响。Zhang等[8]基于tRNA浓度提出预测翻译速率的数学模型，并用实验证实蛋白质某些区段翻译速率较慢，是由于该区段密集分布着若干个tRNA浓度较低的密码子，导致核糖体在通过这些区域时翻译缓慢，这些翻译缓慢的区段被称为翻译暂停位点。翻译暂停位点并非随机分布，一般在蛋白质结构域分隔处下游20~70氨基酸处[30]，见图1。用提高相应tRNA浓度或沉默突变方法可以加速翻译暂停位点的翻译速率。但是即使氨基酸序列不变，也有可能导致蛋白质难以正确折叠而丧失功能，甚至导致错误的细胞定位。研究发现较长的蛋白质中翻译暂停位点一般较多，在翻译中使翻译速率和结构域折叠速率相适应，帮助形成正常功能的蛋白质，且该机制完全独立于分子伴侣等折叠机制[8]。随后人们发现许多疾病，包括癌症、神经系统疾病等都可以由沉默突变导致[8]。翻译暂停理论突破了安芬森原则，指出翻译速率也会决定蛋白质的折叠，是生化基础理论的一大革新[32]。

2.3 翻译起始效率和翻译延伸速率的二维调控模式

翻译起始和翻译延伸作为翻译调控的两大主要阶段，其全局测定是翻译调控研究所必须首先面对的问题。

通过平行测定同一组样品的mRNA和RNC-mRNA，即可计算出每个基因的翻译比率(Translation ratio, TR)值。在真核生物中，由于翻译起始是蛋白质合成主要限速步骤，因此TR值可以近似代表翻译起始效率[10]。原核生物中，翻译起始要活跃得多，因此翻译延伸往往成了限速步骤。且原核生物大量存在多顺反子，因此直接测定每个基因的翻译起始效率尚无有效方法。

相比之下，翻译延伸速率的测定则困难得多。2011年，Ingolia等[33]使用ribosome profiling技术对小鼠胚胎干细胞进行了分析，使用药物阻止翻译起始，然后追踪核糖体在mRNA上的分布随时间的改变。由于误差很大，这种方法只能测定大量基因的平均翻译延伸速率，而无法实现单基因水平测定；再者他们使用药物进行翻译起始的抑制，并非生理状况下的测定，因而其结果存在着很明显的人为偏倚。这些技术上的严重缺陷使得他们得出了一个显然有悖事实的结论：不同类型基因的平均延伸速率看上去没有显著差异。

Lian等[17]利用RNA-seq、RNC-seq及Ribo- seq三种技术同时对同一细胞系进行测序，使用超高精度序列比对算法FANSe2对测序数据进行处理，根据三种测序方法得到的数据计算出了4株细胞在生理条件下8 000~10 000个基因每个基因的翻译起始效率(以翻译比率TR表征)和翻译延伸速率(以延伸速率指数EVI表征)。研究结果表明tRNA浓度介导的密码子选择是几乎唯一的调控翻译延伸速率的因素，并鉴定出了具有很强的细胞特异性的翻译延伸缓慢基因所喜用的密码子。研究结果表明几十年来人们一直用来预测翻译延伸速率的各种密码子偏好指数(如codon adaptation index, codon bias index, tRNA adaptation index等)与实测的翻译延伸速率基本没有相关性。该研究还直接用实验数据证实mRNA二级结构的稳定性对翻译延伸速率在全局上没有显著影响，终结了30年来饱受争议的问题[17]。

翻译起始效率和翻译延伸速率有一定的关联，形成二维调控模式(图2)：

(1) 大多数基因具有普通的TR和EVI；

(2) 部分基因具有高的TR，即翻译起始极为活

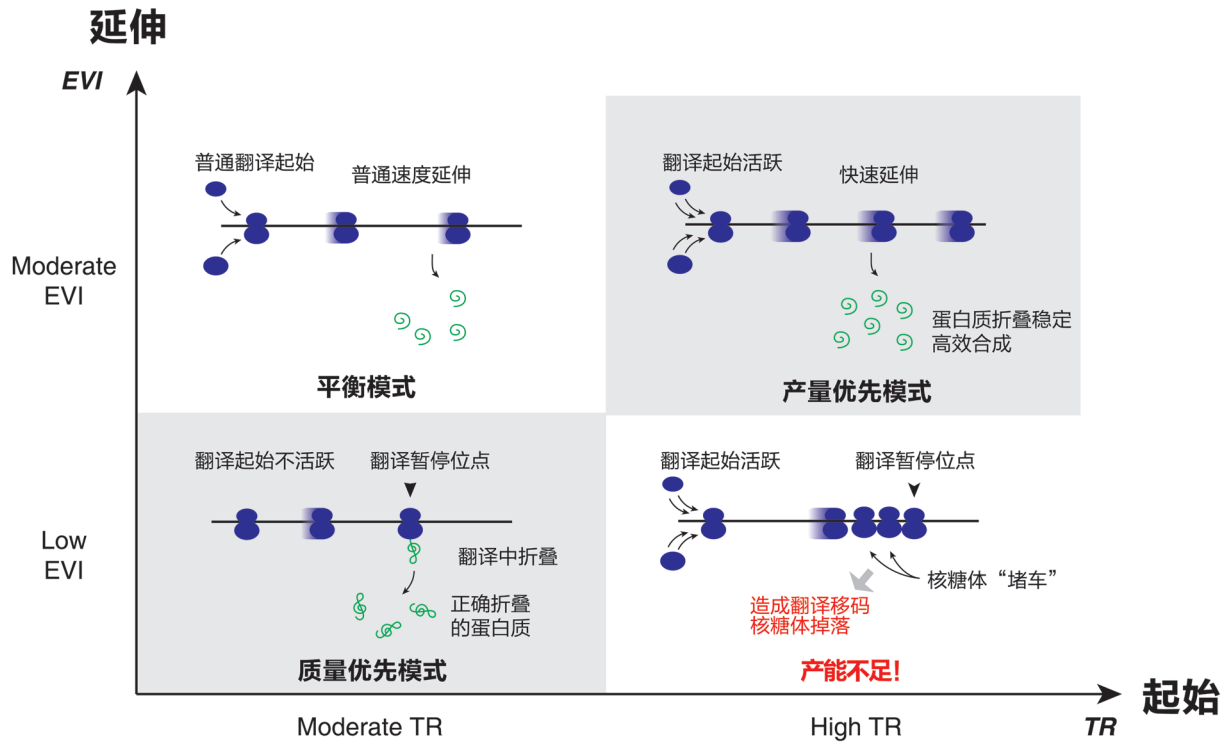

图2　翻译起始效率(TR)和翻译延伸速率(EVI)的二维调控模式

跃，这部分蛋白质的折叠很稳定，无需翻译暂停，因此提高翻译起始效率可大大提高其产量；

(3) 部分基因具有低EVI，即翻译延伸速率缓慢，翻译暂停位点多，这类基因往往编码多结构域蛋白，为保证折叠效率(蛋白质生产质量)而降低翻译延伸速率；

(4) 不存在翻译起始效率很高和翻译延伸速率很低的基因，这是细胞为了避免核糖体“堵车”或核糖体掉落等导致无效的蛋白质翻译。

3　翻译组学的应用

3.1　肿瘤方面应用

癌症是人类生命的重大威胁之一，以其难于治愈、易于复发的特性让人谈之色变。恶性癌组织一般没有清晰的边缘，手术难以根除，而且由于其基因组的不稳定性使得突变频发，使靶向药物失效或对化疗药物产生耐药性。靶向药物总是有限的，因而终将败于不断突变的癌细胞。从这个意义上来说，一切基于基因组变异的癌症治疗方案都是治标不治本的权宜之计。癌细胞基因组上一切的变化最终都必须通过转录并翻译为蛋白质才能最终体现。因此从翻译层次入手，揭示癌细胞的生存与发展的根本规律，才能为癌症的防治提供真正有效的基础。

Wang等[10]利用翻译组测序技术对肺癌细胞和正常细胞进行分析，发现肺癌细胞中全局性mRNA的翻译起始效率(以TR表征)普遍比正常细胞上调，其中TR上调幅度最大的123个基因几乎完整体现了癌细胞的各种表型，精准程度远胜于SILAC定量蛋白质组，说明优先翻译的基因决定着细胞的功能与表型。

这项研究也发现，在正常细胞中，长度越短的基因(对应着小蛋白质)越容易被翻译，而癌细胞中的短基因翻译优先级更加整体上调。Guo等[34]提出理论并用数学模型仿真的方法证实，由于蛋白质合成过程中会出错(占主导地位的是折叠错误)，因此合成短蛋白质的成品率较高，能量利用效率较高；但大型蛋白质的功能较强，需要保证折叠效率而不是产量。因此为了保证蛋白质组(而不是单个蛋白质)整体的功能性，翻译起始效率随基因长度逐渐降低是“性价比”最高的模式。癌细胞翻译速率升高，折叠错误更多，更难生产折叠正确的大型蛋白质，因此提高小分子蛋白质的翻译效率是提高生存优势所必需的。

在癌细胞翻译延伸的整体调控方面，Lian等[17]测定了生理条件下每个基因的翻译起始效率，发现癌细胞中促癌基因的翻译速率显著降低，而下调翻译速率能有效保证这些基因蛋白的正确折叠，从而使这些促癌基因正常行使功能；与此同时，已知的抑癌基因在癌细胞中的翻译速率被主动上调，使其难以正确折叠成有功能的蛋白质，从而不能发挥抑癌作用。这一正一反的翻译延伸调控，使得癌细胞主动、选择性地维持了其恶性表型。

对翻译起始和延伸的整体研究，为癌症细胞表型研究、分子机制研究和新治疗策略提供了全新的研究方向。

3.2 微生物抗逆机制研究

微生物需要快速应对各种环境中的不利因素(例如抗生素)，而抗生素导致的活性氧(ROS)累积所造成的氧化压力是细菌所最经常遇到的不利环境。抗生素导致的氧化自由基在加药后立即产生，而细菌拥有应对氧化压力的专门系统(如SoxRS系统和OxyR系统)，通过转录调控来应对氧化应激，需要20~30 min才能起效[35]。Zhong等[5]发现，在加入过氧化氢之后1~2 min，细菌胞内蛋白质的合成立即减慢到几乎停滞的状态。但是，氧化应激并没有造成细胞内大面积的蛋白质氧化损伤，翻译起始速率丝毫不受影响，而翻译延伸速率则严重减慢。通过tRNA组全定量测序技术、qRT-PCR和凝胶电泳的三重验证，发现几乎所有的tRNA在氧化应激开始时都出现了严重的下调。而人工添加tRNA则可以加速翻译，在氧化应激的条件下使大肠杆菌更快地生长，保护其对抗更高浓度的过氧化氢，甚至耐受环丙沙星引起的ROS。研究表明，在正常状况下，过高的tRNA水平会减少翻译暂停，造成蛋白质折叠失败，导致细菌生长受阻；而在氧化压力下，较高的tRNA水平则能帮助细菌抵抗压力，维持生存能力。传统认为，细菌的调控主要是转录调控，而翻译调控只占很小的比例，该研究不仅发现了细菌快速应对氧化应激的全新机制，显示出tRNA介导的翻译调控是细菌抵抗逆境的重要防线，还为细菌耐药提供了全新的视角。

肺炎链球菌(*S. pneumoniae*)是严重威胁人类健康的主要病原菌之一，持续高效的从宿主方摄取铁离子是肺炎链球菌赖以生存的必要条件。在肺炎链球菌中已被发现的有PiaABC，PiuABC和PitABC三种转铁蛋白复合体，但即便是将这三种蛋白质全部敲除，肺炎链球菌仍然能摄取足够的铁而生存[36]。Yang等[36]通过RNC-seq和iTRAQ蛋白质质谱定量技术，联合搜寻在缺铁胁迫时中高表达的基因，独立发现了多个未知转铁蛋白候选基因，并通过实验验证了其中SPD_1609是一个新的铁结合蛋白(iron-binding protein)，作为传统三种转铁蛋白的备份系统。该研究不仅发现了新的转铁蛋白，还建立了一套行之有效的发现新的功能蛋白的方法策略[36]。

3.3 指导蛋白质组鉴定以及新蛋白的发掘

人类蛋白质组计划(Human Proteome Project, HPP)是继人类基因组计划之后又一大型国际联合项目，力图鉴定所有编码基因所编码的蛋白质以及它们在各组织、细胞、亚细胞的分布情况。质谱技术的广泛应用使得大规模鉴定蛋白质组成为可能，但现有蛋白质组学技术仍然还有诸多的缺陷，不能够完全对蛋白质组进行测定。翻译组测序精度比现有质谱技术高、不受蛋白质理化性质影响，且同样可以提供基因的翻译证据，还能轻易达到极高的覆盖度，间接测定蛋白质序列的改变。Zhong等[36]提出可以用翻译组分析技术辅助蛋白质组鉴定，用以找到已知和未知编码基因的翻译证据。人类蛋白质组计划中国团队合作完成了肝癌细胞系的转录组-翻译组-蛋白质组联合分析，利用高精度测序分析算法FANSe2[22]从每种肝癌细胞系中定量测定了14 000~15 000个正在翻译的基因的表达量，发表了一系列成果[38-41]。由于为人类蛋白质组计划提供了重要数据，翻译组学已成为人类蛋白质组计划的四大核心支柱之一，被人类蛋白质组组织(HUPO)列为2014年人类蛋白质组计划的首要突出贡献[37]。

新蛋白(即由以往认为的非编码RNA所编码的蛋白质)是人类蛋白质组计划中另一个亟需解决的难题，应用翻译组技术在新蛋白寻找上也取得了瞩目的成绩。翻译组测序不依赖原有基因的注释即可给出细胞内蛋白质确定性的翻译证据；不存在抗体“低染”不确定性的问题，因此可以很有效地检测细胞/组织特异性表达蛋白，还可以在RNC-mRNA中发现一些以往被误认为是非编码

RNA的基因。例如在肺癌细胞中，应用RNC-seq发现了1 397个RefSeq数据库中被标注为非编码RNA的基因，而该数据表明这些标注为非编码的基因很有可能被翻译为蛋白质，因此人类基因组可能需要进行系统性的重新注释[37]。以上证据表明翻译组测序方法不但可以作为蛋白质组学的重要补充，还可以独立地作为发现新蛋白质的一种有效手段。在2016年5月的中国蛋白质组学大会上，翻译组证据被列为新蛋白发现的必要证据。

3.4 利用翻译暂停理论优化蛋白折叠效率

蛋白质工程中，人们将外源基因转入细菌或酵母等工业表达体系中进行诱导表达，可大量生产蛋白质，降低成本，例如胰岛素的成功合成。但绝大部分具有药用价值或工业价值的蛋白质在细菌或酵母等表达体系中无法正确折叠，致使目的蛋白表达量低或进入包涵体，活性蛋白质产率很低。传统上人们应用包涵体复性、分子伴侣共表达、密码子优化等手段优化蛋白质折叠，但往往收效甚微，且通用性较差。根据Zhang等[8]提出的翻译暂停理论，优化翻译暂停即可使外源蛋白质正确折叠。这种策略在不改变蛋白质氨基酸序列和诱导表达条件的情况下，可成百上千倍地提高蛋白质的可溶性产量、提高蛋白的生物活性，且可与传统优化方法联用。2014年，Chen等[18]首次报道了利用上述算法模型的翻译暂停理性重设计策略和验证方法，在大肠杆菌中优化了一种重要的抗病毒药物蓝藻抗病毒蛋白N(Cyanovirin-N, CVN)的可溶性表达。在未改变其氨基酸序列，也未改变诱导表达条件的基础上，仅应用翻译暂停理性重设计策略，就使得可溶性目的蛋白产量同比野生型增加2 000倍，且优化后的CVN蛋白抗病毒比活性显著高于野生型。该成果证明了翻译暂停理论重设计法在工业生产上的巨大潜力，表明了基于翻译暂停理论重设计蛋白策略的优越性。

4 总结与展望

翻译调控比转录调控更快更灵敏，使得生物在面对逆境时可以做出快速的应对措施以保证生物体的存活，极大地提高了基因表达调控的灵活性与复杂性，体现出翻译调控对于生物的重要性。翻译组学则让人们能够全局性地研究这一生命调控过程中最重要的一步，打通了基因组到蛋白质组的关键一环，不但具有重大的理论意义，也引致全新的应用。我们相信，翻译组学这一新兴学科将在广泛的研究与应用领域展现其强大的生命力。

参 考 文 献

[1] Shapiro JA. Revisiting the central dogma in the 21st century. Ann N Y Acad Sci, 2009, 1178: 6-28

[2] Schwanhausser B, Busse D, Li N, et al. Global quantification of mammalian gene expression control. Nature, 2011, 473(7347): 337-342

[3] Halbeisen RE, Gerber AP. Stress-dependent coordination of transcriptome and translatome in yeast. PLoS Biol, 2009, 7(5): e1000105

[4] Zhang G, Fedyunin I, Miekley O, et al. Global and local depletion of ternary complex limits translational elongation. Nucleic Acids Research, 2010, 38 (14): 4778-4787

[5] Zhong J, Xiao C, Gu W, et al. Transfer RNAs Mediate the Rapid Adaptation of Escherichia coli to Oxidative Stress. PLoS Genet, 2015, 11(6): e1005302

[6] King HA, Gerber AP. Translatome profiling: methods for genome-scale analysis of mRNA translation. Brief Funct Genomics, 2014, 15(1):22-31

[7] Morello LG, Hesling C, Coltri PP, et al. The NIP7 protein is required for accurate pre-rRNA processing in human cells. Nucleic Acids Res, 2011, 39(2): 648-665

[8] Zhang G, Hubalewska M, Ignatova Z. Transient ribosomal attenuation coordinates protein synthesis and co-translational folding. Nat Struct Mol Biol, 2009, 16(3): 274-280

[9] Ho JJ, Wang M, Audas TE, et al. Systemic Reprogramming of Translation Efficiencies on Oxygen Stimulus. Cell Rep, 2016, 14(6): 1293-1300

[10] Wang T, Cui Y, Jin J, et al. Translating mRNAs strongly correlate to proteins in a multivariate manner and their translation ratios are phenotype specific. Nucleic Acids Res, 2013, 41(9): 4743-4754

[11] Ingolia NT, Ghaemmaghami S, Newman JR, et al. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science, 2009, 324(5924): 218-223

[12] Ingolia NT. Ribosome Footprint Profiling of Translation throughout the Genome. Cell, 2016, 165(1): 22-33

[13] Inada T, Winstall E, Tarun SJ, et al. One-step affinity purification of the yeast ribosome and its associated proteins and mRNAs. RNA, 2002, 8(7): 948-958

[14] Heiman M, Kulicke R, Fenster RJ, et al. Cell type-specific mRNA purification by translating ribosome affinity purification (TRAP). Nat Protoc, 2014, 9(6): 1282-1291

[15] Wang Y, Jiao Y. Translating ribosome affinity purification (TRAP) for cell-specific translation profiling in developing flowers. Methods Mol Biol, 2014, 1110: 323-328

[16] Mustroph A, Juntawong P, Bailey-Serres J. Isolation of plant polysomal mRNA by differential centrifugation and ribosome immunopurification methods. Methods Mol Biol, 2009, 553: 109-126

[17] Lian X, Guo J, Gu W, et al. Genome-Wide and Experimental Resolution of Relative Translation Elongation Speed at Individual Gene Level in Human Cells. PLoS Genet, 2016, 12(2): e1005901

[18] Chen W, Jin J, Gu W, et al. Rational design of translational pausing without altering the amino acid sequence dramatically promotes soluble protein expression: a strategic demonstration. J Biotechnol, 2014, 189: 104-113

[19] Chan PP, Lowe TM. GtRNAdb: a database of transfer RNA genes detected in genomic sequence. Nucleic Acids Res, 2009, 37(Database issue): 93-97

[20] Dong H, Nilsson L, Kurland CG. Co-variation of tRNA abundance and codon usage in Escherichia coli at different growth rates. J Mol Biol, 1996, 260(5): 649-663

[21] Kanaya S, Yamada Y, Kudo Y, et al. Studies of codon usage and tRNA genes of 18 unicellular organisms and quantification of Bacillus subtilis tRNAs: gene expression level and species-specific diversity of codon usage based on multivariate analysis. Gene, 1999, 238(1): 143-155

[22] Xiao CL, Mai ZB, Lian XL, et al. FANSe2: a robust and cost-efficient alignment tool for quantitative next-generation sequencing applications. PLoS One, 2014, 9(4): e94250

[23] Zheng G, Qin Y, Clark WC, et al. Efficient and quantitative high-throughput tRNA sequencing. Nat Methods, 2015, 12(9): 835-837

[24] Cozen AE, Quartley E, Holmes AD, et al. ARM-seq: AlkB-facilitated RNA methylation sequencing reveals a complex landscape of modified tRNA fragments. Nat Methods, 2015, 12(9): 879-884

[25] 周觅, 刘如娟, 王恩多. tRNA转录后修饰的功能及其与人类疾病的关系. 生命科学, 2014, 10: 1032-1037.

[26] Puri P, Wetzel C, Saffert P, et al. Systematic identification of tRNAome and its dynamics in Lactococcus lactis. Mol Microbiol, 2014, 93(5): 944-956

[27] Hsu ST, Fucini P, Cabrita LD, et al. Structure and dynamics of a ribosome-bound nascent chain by NMR spectroscopy. Proc Natl Acad Sci U S A, 2007, 104(42): 16516-16521

[28] Deckert A, Waudby CA, Wlodarski T, et al. Structural characterization of the interaction of alpha-synuclein nascent chains with the ribosomal surface and trigger factor. Proc Natl Acad Sci U S A, 2016, 113(18): 5012-5017

[29] Maier T, Guell M, Serrano L. Correlation of mRNA and protein in complex biological samples. FEBS Lett, 2009, 583(24): 3966-3973

[30] Zhang G, Ignatova Z. Generic algorithm to predict the speed of translational elongation: implications for protein biogenesis. PLoS One, 2009, 4(4): e5036

[31] Sauna ZE, Kimchi-Sarfaty C. Understanding the contribution of synonymous mutations to human disease. Nat Rev Genet, 2011, 12(10): 683-691

[32] Fredrick K, Ibba M. How the sequence of a gene can tune its translation. Cell, 2010, 141(2): 227-229

[33] Ingolia NT, Lareau LF, Weissman JS. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. Cell, 2011, 147(4): 789-802

[34] Guo J, Lian X, Zhong J, et al. Length-dependent translation initiation benefits the functional proteome of human cells. Mol Biosyst, 2015, 11(2): 370-378

[35] Seo SW, Kim D, Szubin R, et al. Genome-wide Reconstruction of OxyR and SoxRS Transcriptional Regulatory Networks under Oxidative Stress in Escherichia coli K-12 MG1655. Cell Rep, 2015, 12(8): 1289-1299

[36] Yang XY, He K, Du G, et al. Integrated Translatomics with Proteomics to Identify Novel Iron-Transporting Proteins in Streptococcus pneumoniae. Front Microbiol, 2016, 7: 78

[37] Zhong J, Cui Y, Guo J, et al. Resolving chromosome-centric human proteome with translating mRNA analysis: a strategic demonstration. J Proteome Res, 2014, 13(1): 50-59

[38] Liu Y, Ying W, Ren Z, et al. Chromosome-8-coded proteome of Chinese Chromosome Proteome Data set (CCPD) 2.0 with partial immunohistochemical verifications. J Proteome Res, 2014, 13(1): 126-136

[39] Wang Q, Wen B, Wang T, et al. Omics evidence: single nucleotide variants transmissions on chromosome 20 in liver cancer cell lines. J Proteome Res, 2014, 13(1): 200-211

[40] Chang C, Li L, Zhang C, et al. Systematic analyses of the transcriptome, translatome, and proteome provide a global view and potential strategy for the C-HPP. J Proteome Res, 2014, 13(1): 38-49

[41] Zhang C, Li N, Zhai L, et al. Systematic analysis of missing proteins provides clues to help define all of the protein-coding genes on human chromosome 1. J Proteome Res, 2014, 13(1): 114-125