



**WIE-biotech**

立效生物

# (多) 核糖体图谱

基础+技能版

方案组  
2022/8/8

---

## 目录

核糖体图谱介绍.....	2
核糖体图谱主要的未预见的发现.....	4
核糖体图谱技术.....	6
多聚核糖体图谱技术.....	20

## 核糖体图谱介绍

蛋白质在生命活动中执行着各种功能，所以它们在细胞内是处于持续严格的控制之下的。根据分子生物学的中心法则，整个蛋白质组的产生包括四个主要的调控步骤：**RNA 合成**（包括表观遗传和转录调控）、**RNA 降解**、**蛋白质合成**（即翻译调控）和**蛋白质降解**。随着蛋白质组学技术的发展，研究表明翻译调节占有所有调节幅度的一半以上，超过了所有其他调节的总和，因此翻译调控可以说是生物体中最重要的调控步骤。与全基因组的“基因组学”、所有转录本的“转录组学”、所有蛋白质的“蛋白质组学”对应，在“翻译组学”中人们研究与翻译有关的所有元素。**核糖体图谱是非常了不起的创新，它揭示了太多 mRNA 翻译以前非被发现的信息。**RNA 协会应该对 **Jonathan Weissman 和 Nick Ingolia** 表示感谢，他们不仅想出了这个点子，还对方法学进行了仔细的优化。这个方法的起源可以追溯至 50 年前多核糖体的发现，当用核酸内切酶处理会将其转化成单个保护有 mRNA 片段的核糖体。**Joan Steitz 和 Marilyn Kozak 首先利用这种特性来鉴定和测序原核生物噬菌体 RNA 和真核生物呼肠病毒 RNA 的翻译起始位点。**Sandra Wolin 和 Peter Walter 对这一方法基序延伸，揭示了核糖体在催乳素起始密码子和终止密码子上的停滞现象和 SRP 与新生肽链作用的节点。因为测序方法在那个时代的限制，这些实验需要使用无细胞翻译系统，最好只翻译单个 mRNA（或者是至多四个呼肠 mRNA）。

现代深度测序技术和相关的扩增短 mRNA 片段获得 cDNA 文库的方法，使在完整细胞中获得核糖体在 mRNA 上的结合位置得到了可能。我们认为这个方法最大的创新之一是使用 **E. Coli RNase I 生成受到保护的 mRNA 片段**，这种酶相比之前使用的核酸酶（RNase A、RNase T1 和微球菌核酸酶）几乎没有序列特异性。RNase I 酶切序列的 5' 末端，**核糖体在 ORF 上的位置可以单碱基精确度地获得**，图谱显示出明显的**三核苷酸周期性**，这允许了翻译的阅读框的分配以及将翻译的核糖体产生的图谱与其他原因产生的图谱区分开。

在大多数分析中，受到保护的 mRNA 片段的长度分布展现出**单个对称的峰**，中位数是 28-29nt（酿酒酵母）或者 30-31nt（哺乳动物细胞），这可能是因为哺乳动物 60S 核糖体亚基更大。但是，~16nt，~21nt 和 40-65nt 的受保护片段也被 Nick Guydosh 和 Rachel Green 等人获得。16nt 片段来源于 80S 核糖体停留在截断（部分消化）mRNA 物理 3' 端；40-65nt 片段被认为是当前面的核糖体停滞时堆在一起的 80S 核糖体产生

---

的。~21nt 片段更加有趣，似乎是来源于翻译中的 80S 核糖体的另外一种构象：这种构象可以被茴香霉素稳定，而放线菌酮可以稳定~30nt 的片段(Pat Brown 课题组)。

使用冷冻电镜或者 X 射线晶体学来比较这两种结构一定非常有趣。

为了得到各 mRNA 的核糖体足迹强度，我们需要知道各 mRNA 本身的丰度，这可以通过以下方式达成：从细胞提取物中分离 poly A+RNA，将其片段化成~30nt，然后将这些片段进行深度测序。这可以获得各 mRNA 的平均核糖体密度，但是因为这仅仅是平均的，不能反映 mRNA 在多核糖体梯度显示的双峰分布，其中一些完全被抑制或者未翻译。

## 核糖体图谱主要的未预见的发现

一、可能是最意外的发现,在小鼠 ES 细胞,有着大量翻译的短上游 ORF(uORF),有着类似的起始密码子(CUG>GUG>UUG>ACG>其他),这些 uORF 是 AUG 起始的 uORF 的四倍。毫无疑问这些 uORF 被真实地翻译了,但是细胞收获的步骤可能会扩大这些相似位点的频率(见下面)。GCN4 mRNA 的前例提示这些 uORF 的功能可能是调控编码蛋白的 ORF 的翻译,而不是产生具有生物学功能的肽链,但仍然有待去看这是不是对所有的 uORFs 都成立。

二、对于很多编码蛋白的 ORF,注释的起始密码子被发现不是唯一的框内起始位点,在某些情况甚至不是主要的起始位点。因此相比之前认为的可能有更多的 mRNA 编码 N 末端拓展或截断的亚型。

三、在果蝇胚胎和 S2 细胞(同时也包括酿酒酵母和人类纤维母细胞),大量的低丰度的终止逃逸的例子,导致 C 末端拓展的蛋白亚型的产生,在部分例子中包含进化上保守的氨基酸序列。

四、核糖体足迹的三核苷酸周期,与三尖杉碱的联合使用,使发生在两个相对长的相交的阅读框的翻译的许多例子被发现。另外, Caia Duncan 和 Juan Mata 在裂殖酵母中发现 19 例编码蛋白质的 mRNA 从相同基因组区域的两条链中转录。

五、核糖体图谱发现酿酒酵母、果蝇 S2 细胞和脊椎动物中多种盖帽和聚腺苷酸化的 mRNA,这些在之前被分类为“长非编码 RNA”(因为缺乏超过 100 个密码子的 ORF)实际上是被翻译的:它们有短的靠近 5'端的 ORF,产生特征性的三核苷酸周期性的足迹(除了当翻译发生在两个相交的阅读框的时候),它们受到无义介导的降解,这一过程是依赖于翻译的。在某些情况,多肽产物通过质谱或者同位素标记被探测到,其中一些短 ORF 在进化上是保守的,即使是在斑马鱼胚胎和人类细胞中,提示这些编码的多肽可能有生物学功能。

六、Weissman 课题组发明了绝妙的方法,在体内使用生物素标记靠近 ER 或者线粒体的酵母核糖体 60S 亚基,因此允许在建立图谱的时候选择特定的核糖体亚群。对于多种分泌的蛋白质,图谱展示出核糖体在 ER 吸引 Sec61/Sec63 或者 Ssh1 转座子所需要的最短的新生肽链。线粒体研究揭示内膜蛋白无一例外的是共转录地插入,但是其他蛋白的共转录插入受到放线菌酮添加的强力促进,因为可以给核糖体更多时间来靠近线粒体。

七、使用三尖杉碱处理小鼠 ES 细胞数次可以获得衍生速率，大概是 330 个密码子每分钟，该速率在不同 mRNA 之间的差异很小。

八、核糖体停滞和暂停的以及有时与核糖体堆叠偶联的例子，在 ES 细胞和酿酒酵母的连续脯氨酸密码子处被发现，也在组氨酸饥饿时在组氨酸密码子处被发现。

九、Nick Guydosh 和 Rachel Green 发现 10%的酿酒酵母 mRNA 在三种阅读框中都有 3UTR 核糖体足迹，在 *dom34Δ* 株中显著提升。它们被认为是由通常的 ABCE1 (Rli1) 催化的终止后核糖体释放的失败（或去组装）导致的，结果是 80S 核糖体在 3UTR 上迁移而不翻译。在野生株中它们很可能通过与 Dom34 和 Hbs1 蛋白结合而被解救（例如回收）。

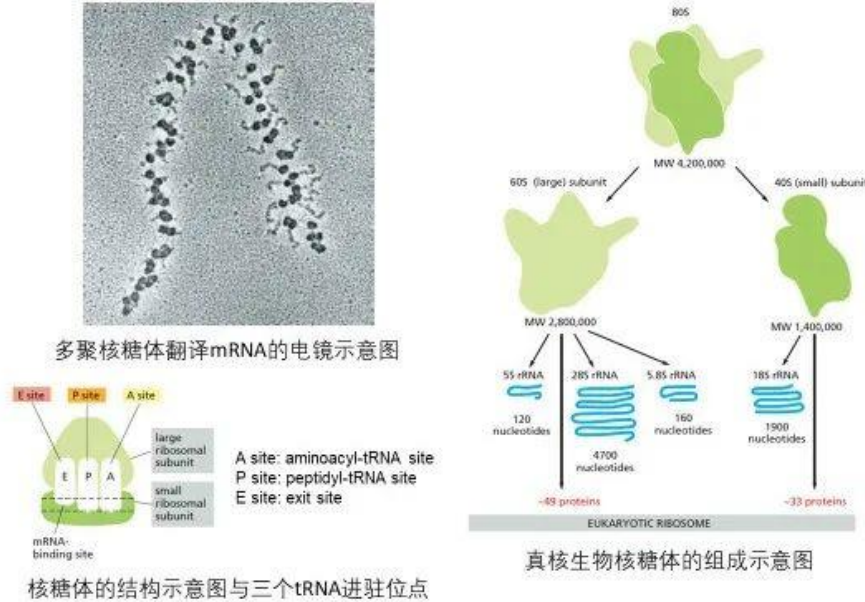
我们一直困扰的一个问题，是细胞收获和裂解过程在何种程度上扰乱稳定状态的核糖体分布。除非材料可以迅速冷却至 $\sim 0^{\circ}\text{C}$ ，多核糖体逃离可能发生，使起始位点、短的 ORF 和长 ORF 的 5 端的核糖体减少。为了避免这种情况，很多操作人员会在 CHX 中培养细胞 5 分钟，之后冷却。这可以避免逃离，但是会使起始位点的足迹比稳定状态时多得多，因为 CHX 不抑制扫描或起始，所以额外的 80S 起始复合物可以结合至抑制剂开始发挥作用时有空的起始位点的 mRNA。尽管这一问题先前被部分作者发现，没有人定量地描述这一现象。对于可以捕获所有扫描的核糖体的高效的起始位点，我们推测 100%的空位的占据会在 CHX 抑制开始的 10 秒内完成。（这一估计基于以下事实：90%的翻译合成两种球蛋白[平均 143 个氨基酸，平均多核糖体数目为 5]的网织红细胞在  $37^{\circ}\text{C}$  时核糖体需要花 $\sim 30$  秒合成球蛋白链，各 mRNA 的翻译起始必须平均 6 秒开始一次）。如果起始位点发生了变异，使其效率降低，只捕获 10%的扫描核糖体，所有空位点的占据也能在 2 分钟以内完成。所以使用 CHX 孵化 3-5 分钟会导致起始位点相同的强度，而本身的效率却相差 10 倍以上。此外，80S、核糖体停留在 5 端附近的位点将会阻止任何扫描核糖体到达下游位点，这可能导致在一般背景的 uORF 起始位点观察到的足迹信号比高效率的下游起始位点的信号要大。

同样的问题会出现，如果细胞使用三尖杉碱孵化 3--5 分钟，它既不抑制起始也不抑制延伸（或者多核糖体逃脱），但是阻止翻译从起始状态到延伸状态，所以它会在所有的起始位点留下明显的 80S 核糖体足迹，但是，像 CHX，它会完全的隐藏它们在稳定条件下的起始速率。

# 核糖体图谱技术

## 一、实验原理

Ribo-seq的实验原理——了解核糖体结构与翻译机器

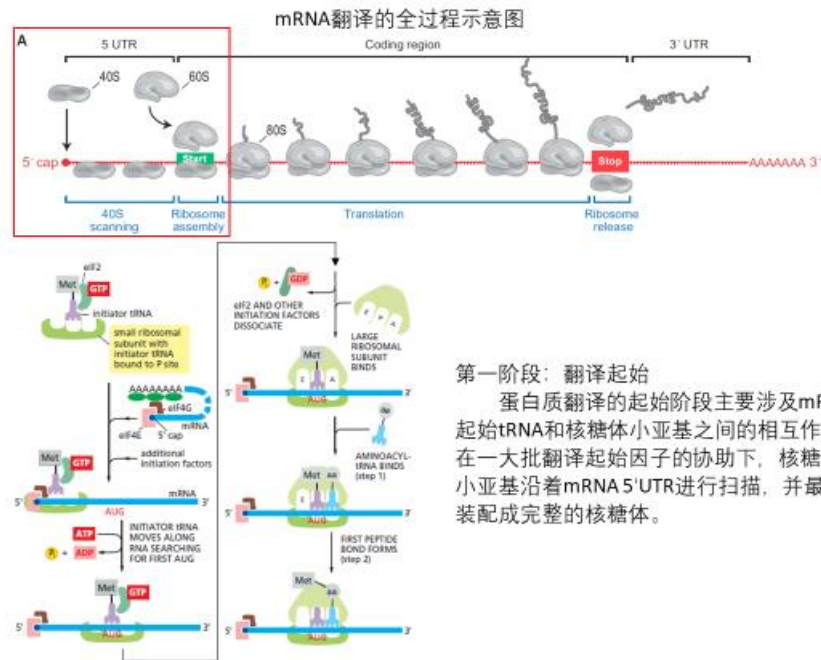


——公众号：RNA 研究团队

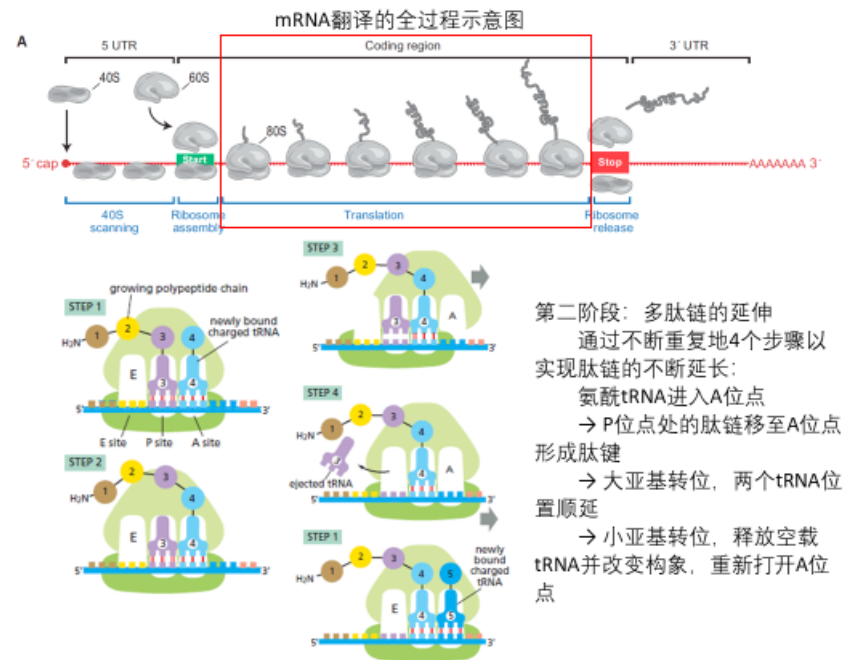
要建立一个能捕获并监控正在翻译的 RNA 的技术，我们自然会思考什么东西是与它关联最紧密的，而我们能通过那个东西对正在翻译的和没有翻译的 RNA 进行区分，那么答案显而易见就是核糖体。所以我们来聊一聊核糖体的构造以及翻译机器的运作，让大家对此有个大概的了解。对于核糖体的结构，比较值得我们注意的是，一个完整的来自真核生物的核糖体包含大小亚基两个部分，而大小亚基又是由若干大小不一的 rRNA 和蛋白质所构成的。核糖体大小亚基的结合形成了三个 tRNA 的进驻位点，分别是 A、P、E 位，对应结合氨酰 tRNA、肽酰 tRNA 和空载 tRNA，这个结构不论是在翻译过程中还是在后面应用到 Ribo-seq 技术中都是一个核心的概念。



## Ribo-seq的实验原理——了解核糖体结构与翻译机器

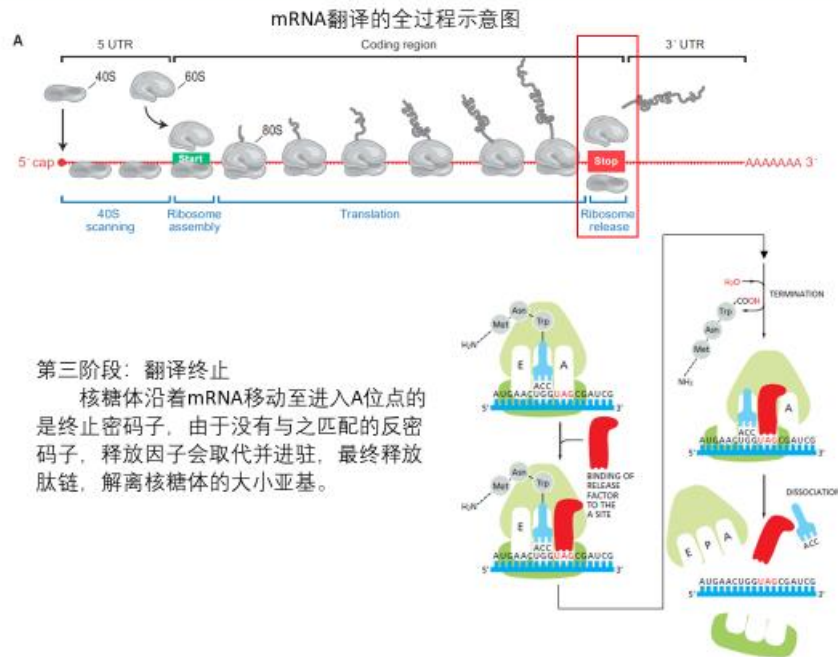


## Ribo-seq的实验原理——了解核糖体结构与翻译机器





## Ribo-seq的实验原理——了解核糖体结构与翻译机器

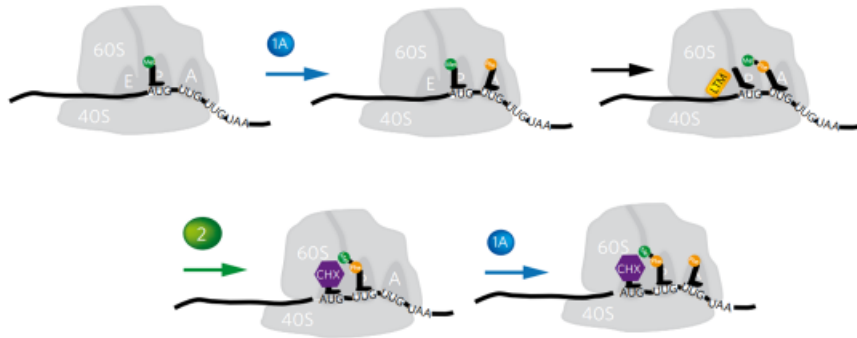


接下来我们回顾一下细胞内 mRNA 的翻译过程，接触过分子生物学的同学对这张图应该并不陌生，所以我们这里就简单地概述一下翻译机器是怎么运作的，过程中所涉及的一系列辅助因子暂时都不进行拓展。mRNA 的翻译过程主要分为翻译起始、肽链延伸和翻译终止三个阶段。翻译起始阶段主要涉及 mRNA、起始 tRNA 和核糖体小亚基之间的相互作用。在一大批翻译起始因子的协助下，核糖体小亚基沿着 mRNA 5' UTR 进行扫描，并最终装配成完整的核糖体。肽链延伸阶段主要通过不断循环以下 4 个步骤以实现蛋白质的合成：氨酰 tRNA 进入 A 位点 → P 位点处的肽链转移至 A 位点并形成肽键 → 核糖体大亚基转位，两个 tRNA 位置顺延 → 核糖体小亚基转位，释放空载 tRNA 并改变构象，重新打开 A 位点。当核糖体沿着 mRNA 移动至进入 A 位点的是终止密码子时，由于没有与之匹配的反密码子，释放因子会取代并进驻，最终释放肽链，解离核糖体的大小亚基，这就是翻译的终止阶段。

## Ribo-seq的实验原理——了解核糖体结构与翻译机器

用于研究翻译行为的两种常用药物

——放线菌酮 (Cycloheximide, CHX) 和拉克霉素 (Lactimidomycin, LTM)



CHX和LTM抑制肽链延伸的机制示意图

CHX和LTM同为抑制肿瘤增殖的戊二酰亚胺类化合物，主要通过竞争性结合60S核糖体大亚基上的E位点，阻止空载tRNA的离开从而实现肽链延伸的阻断。

CHX在蛋白翻译的任何时候都能结合至核糖体上，而LTM只能在翻译起始时进行结合和阻断，在蛋白翻译时无法进入核糖体。研究表明，这两种不同的作用机制，可能是由于大小和结构的不同所致。

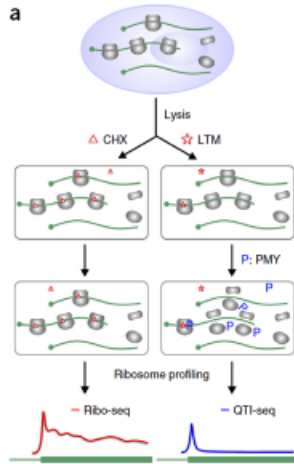
在了解了核糖体的结构和翻译机器的运作后，我们最后来谈一下用于研究翻译行为的两种常用化合物——放线菌酮（CHX）和拉克霉素（LTM）。这两种化合物都是可以用来抑制肿瘤增殖的戊二酰亚胺类化合物，它们主要通过竞争性结合核糖体大亚基上的E位点，阻止空载 tRNA 的离开从而实现肽链延伸的阻断。有趣的是，放线菌酮在蛋白质翻译的任何时候都能结合到核糖体上，而拉克霉素只能在翻译起始时进行结合和阻断而在翻译过程中无法进入核糖体。研究表明，这两种不同的作用机制可能是由于化合物的大小和结构不同所致。拉克霉素的三维结构及其附加的内酯环可能会阻碍它与E位点的结合，使之成为一种“慢速”的抑制剂，也就是只有在没有竞争对手的情况下才能有效结合，即在翻译的起始阶段。

## 二、Ribo-seq 的实验流程

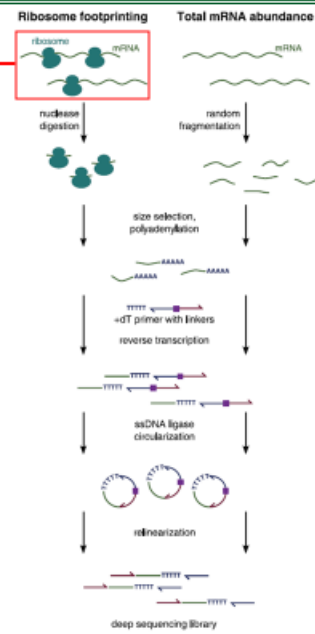
## Ribo-seq的实验流程

### 第一步：核糖体的固定

对培养好的细胞进行短暂的放线菌酮处理，使翻译中的核糖体固定在RNA上。随后对细胞进行充分的裂解。



图示Ribo-seq（使用放线菌酮）和QTI-seq（使用拉克霉素和嘌呤霉素）处理步骤及最终输出结果的差异。



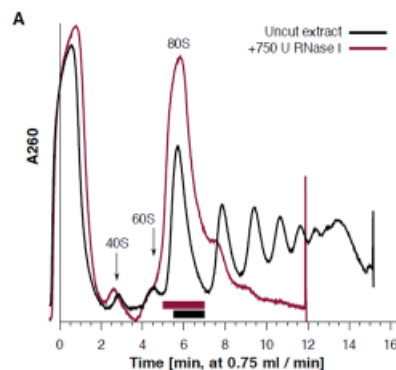
Gao X, Wan J, Liu B, et al. Nature Methods, 2015, 12(2):147.  
Ingolia, N. T., Ghaemmaghami, S., Newman, J. R., et al. Science, 2009, 324:218–223.

如上图中右边这幅示意图所示，Ribo-seq 和 total RNA 测序的流程被摆在了一起。为了能对数据进行更好的分析，这两种测序往往是协同进行的，而今天我们主要把重点放在 Ribo-seq 上。对于 Ribo-seq 技术我们首先考虑的就是要对自然翻译状态下的核糖体进行捕获，所以第一步就是**对核糖体进行固定**，阻滞它在 RNA 上的移动。常规的做法就是使用放线菌酮对培养好的细胞进行短暂的孵育处理，然后对细胞进行裂解。左边的示意图是 Gao 等人基于 Ribo-seq 的流程所发展出来的 QTI-seq，它将放线菌酮处理替换为上述提到的拉克霉素和嘌呤霉素处理，此时只有结合在翻译起始位点处的核糖体可以被阻滞，因而该方法能够特异性地对参与翻译起始的核糖体进行定位和定量。

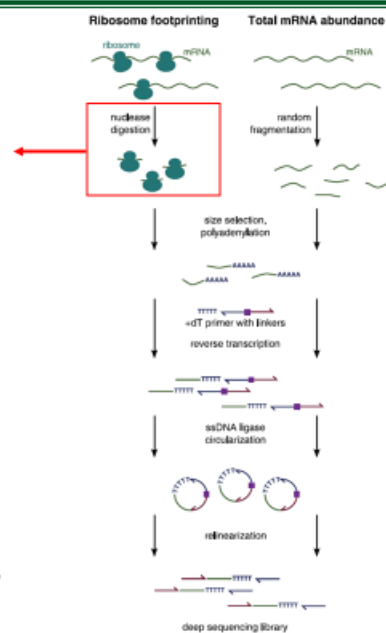
## Ribo-seq的实验流程

### 第二步：核酸酶生成印迹和核糖体分离

在细胞裂解液中加入RNase I消化，使RNA充分片段化，留下被单个核糖体包裹保护的印迹片段。



图示对核酸酶消化的裂解物进行蔗糖密度梯度分析。



Gao X, Wan J, Liu B, et al. Nature Methods, 2015, 12(2):147.  
Ingolia, N. T., Ghaemmaghami, S., Newman, J. R., et al. Science, 2009, 324:218–223.

## Ribo-seq的实验流程

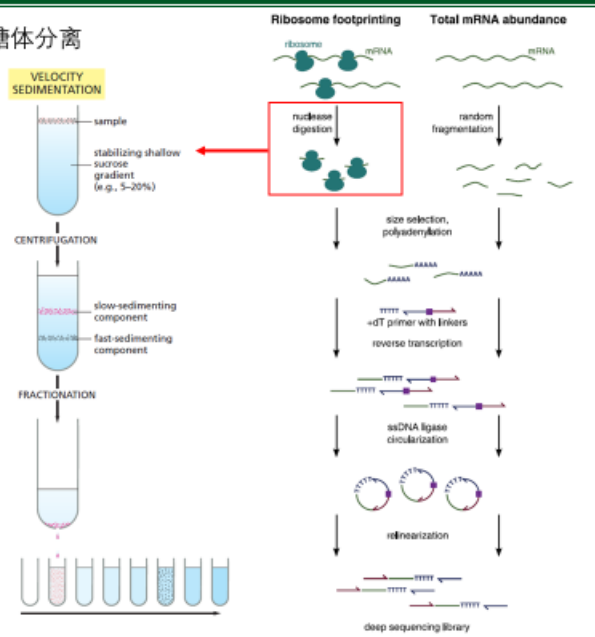
### 第二步：核酸酶生成印迹和核糖体分离

为了将裂解液中的单核糖体与其他细胞组分（细胞器、未被翻译的RNA、解离的核糖体大小亚基等）进行区分，常常采取密度梯度离心的策略。目前也有商业化的基于分子筛原理富集核糖体的试剂盒。

对分离后的单核糖体使用常规Trizol法提取印迹RNA。

#### 密度梯度离心：

含有多种组分的样品在惰性密度梯度介质（蔗糖）中，通过一定的离心力，按不同的沉降速率或密度分布分配到梯度中某些特定位置上，形成不同区带的分离方法，实现单一组分的获取。



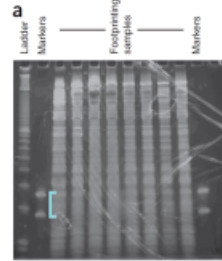
Gao X, Wan J, Liu B, et al. Nature Methods, 2015, 12(2):147.  
Ingolia, N. T., Ghaemmaghami, S., Newman, J. R., et al. Science, 2009, 324:218-223.

经过了上述处理后我们获得了受核糖体保护的完整 RNA，但是研究对象是正在翻译的 RNA，因此那些没有被核糖体覆盖的部分其实是我们不需要的，要想办法进行去除。这里为了获得被单核糖体所包裹的 RNA 片段，采用了核酸内切酶 RNase I 进行消化。右图所示是通过密度梯度离心法分离不同细胞组分后所进行的测定，用吸光度来表征含量，可以发现裂解液在酶切处理前后可以获得单核糖体的量会显著增加。此时的裂解液虽然经过片段化，但依旧是包含众多杂质的混合物。为了将裂解液中的单核糖体与其他细胞组分（细胞器、未被翻译的 RNA、解离的核糖体大小亚基等）进行区分，往往采取密度梯度离心的策略。密度梯度离心的主要原理是，含有多种组分的样品在惰性密度梯度介质（蔗糖）中，通过一定的离心力，会按不同的沉降速率或密度分布分配到梯度中某些特定位置上，形成不同区带，从而实现单一组分的获取。当然目前对于分离核糖体也有更加便利的方法，就是通过商业化的基于分子筛原理的用于富集核糖体的试剂盒。随后就是对分离后的单核糖体使用常规的 Trizol 法来提取印迹 RNA。

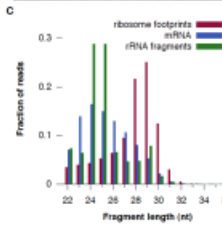
## Ribo-seq的实验流程

### 第三步：印迹片段筛选和纯化（关键）

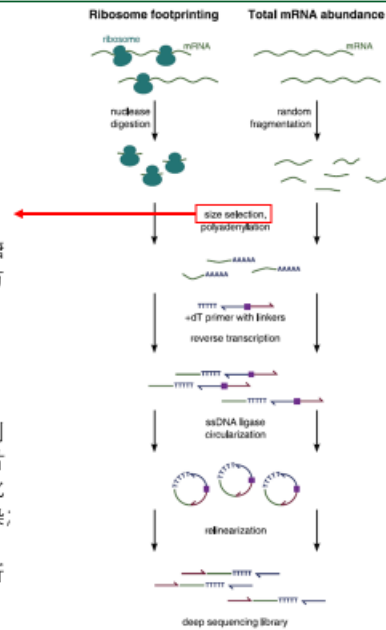
对从核糖体中回收的印迹片段进行凝胶电泳（变性电泳）分离，割取26-30nt大小的RNA凝胶，回收并纯化，用于后续建库测序。



根据片段大小对核糖体印迹进行筛选。蓝色方框指示的是割胶的区域（26-30nt）。



图示核糖体印迹序列的长度分布。其中rRNA片段是由样品中核酸酶消化核糖体所产生的rRNA污染；mRNA片段是由实验最开始进行的随机片段化的所产生。



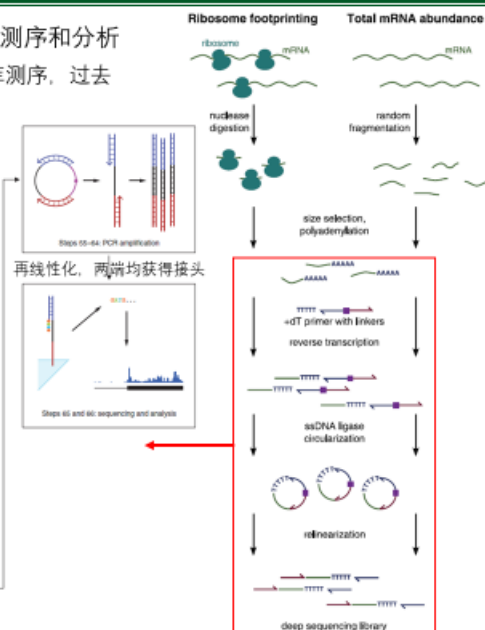
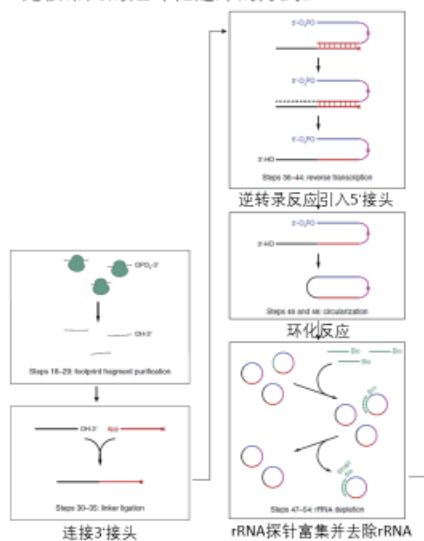
Gao X, Wan J, Liu B, et al. Nature Methods, 2015, 12(2):147.  
Ingolia, N. T., Ghaemmghami, S., Newman, J. R., et al. Science, 2009, 324:218–223.

第三步流程是印迹片段的筛选和纯化，也是 Ribo-seq 里面最为关键的步骤。即使进行了上述的分离步骤，也没有办法保证获得的 RNA 样品里百分之百都是印迹片段。所以，我们要再次基于片段大小对印迹 RNA 进行凝胶电泳（变性）分离，割取 26-30nt 大小的 RNA 凝胶，回收纯化。选择这个范围大小的片段进行割取，主要是考虑到单个核糖体覆盖 RNA 的序列长度。另外从下面这张序列长度分布图我们也可以看出为什么选取 26-30nt 的位置，在这个位置中样品受到 rRNA 片段和 mRNA 片段的干扰也是最小的。

## Ribo-seq的实验流程

### 第四步：对核糖体印迹片段进行建库、测序和分析

对回收所得的核糖体印迹片段进行建库测序。过去比较常用的是环化建库的方式。

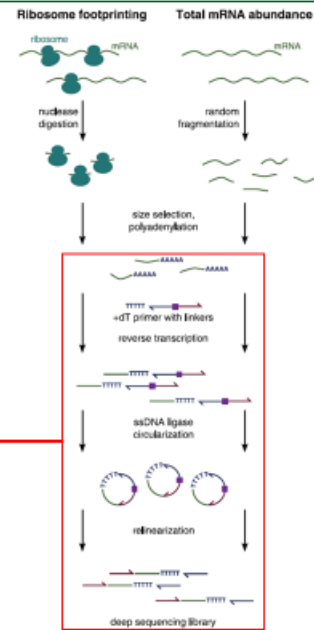


Ingolia, N. T., Brar, G. A., Rouskin, S., et al. Nature Protocols, 2012, 7(8):1534.  
Ingolia, N. T., Ghaemmghami, S., Newman, J. R., et al. Science, 2009, 324:218–223.

## Ribo-seq的实验流程

### 第四步：对核糖体印迹片段进行建库、测序和分析

由于核糖体印迹片段较小，也可采取流程更为简单的小RNA建库方法。同时，上述的去除rRNA步骤也可以提前至印迹片段筛选之前进行，并采用去除效率更高的商业化试剂盒。

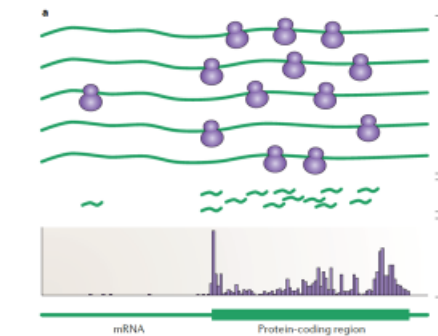


Ingolia, N. T., Brar, G. A., Rouskin, S., et al. Nature Protocols, 2012, 7(8):1534.  
Ingolia, N. T., Ghaemmaghami, S., Newman, J. R., et al. Science, 2009, 324:218–223.

Ribo-seq 流程的最后就是对获得的核糖体印迹片段进行建库、测序和分析。在这里就暂时不对建库的过程做太多说明。过去 Ingolia 等人比较常用的是环化建库的方式，如左图所示，主要包括连接 3' 接头、逆转录反应引入 5' 接头、环化、探针富集并去除 rRNA、再线性化等步骤。当然，由于核糖体印迹片段较小，我们也可以采取流程更为简单的小 RNA 建库方法。同时，上述的去除 rRNA 步骤也可以适当提前至印迹片段筛选之前进行，并采用去除效率更高的商业化试剂盒。

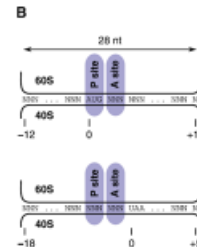
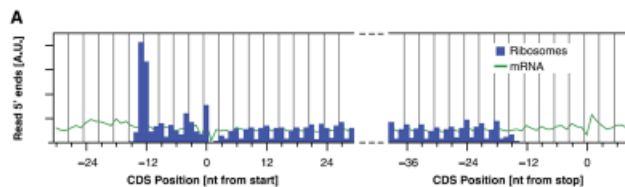
## Ribo-seq的结果呈现

### 结果——精确到密码子分辨率的核糖体密度分布情况



Ribo-seq数据可以精确地反映转录本上每一个位置核糖体的密度分布情况。因此判断编码区域或ORF的范围。

在分析过程中值得注意的是，由于核糖体包裹的RNA范围不仅限于三个密码子的长度，所以在翻译起始和终止位点附近的核糖体印迹分布会向外延伸偏移。



Ingolia, N. T., Nature Reviews Genetics, 2014, 15(3):205.  
Ingolia, N. T., Ghaemmaghami, S., Newman, J. R., et al. Science, 2009, 324:218–223.

对 Ribo-seq 结果的呈现，最为直观的就是可以精确地反映转录本上每一个位置核糖体的密度分布情况，从而判断编码区域或 ORF 的范围。同时，因为核糖体印迹片段的读长分布



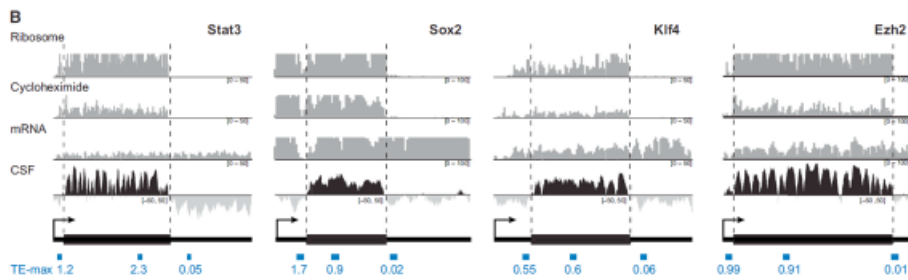
会在密码子的三个碱基上呈现一个高-低-低的趋势，所以这个技术甚至可以将分辨率精确到单密码子上。另外在分析过程中值得注意的是，由于核糖体保护的 RNA 范围不仅限于三个密码子的长度，所以在翻译起始和终止位点附近的核糖体印迹分布会适当向外延伸偏移。

### Ribo-seq的结果呈现

#### 结果二——翻译效率 (Translational efficiency, TE)

$$TE = \frac{\text{the number of reads observed among ribosome-associated RNA}}{\text{the number of reads in total poly-A}^+ \text{ RNA}}$$

简单而言，翻译效率反映了测序数据中RNA的“使用率”——在总RNA中有多少RNA结合了核糖体或是被翻译。



图示4种mRNA在不同区域的核糖体密度分布和翻译效率

- “Ribosome”表示不经任何处理下核糖体关联的读长覆盖情况；
- “Cycloheximide”表示在放线菌酮处理下核糖体关联的读长覆盖情况；
- “mRNA”表示polyA富集的总RNA的读长覆盖情况；
- “CSF”分值用于表征目标序列在何种程度上符合蛋白质编码区域所期望的进化模式。简单来说就是保守性。

Guttman, M., Russell, P., Ingolia, N. T., et al. Cell, 2013, 154(1):240-51.

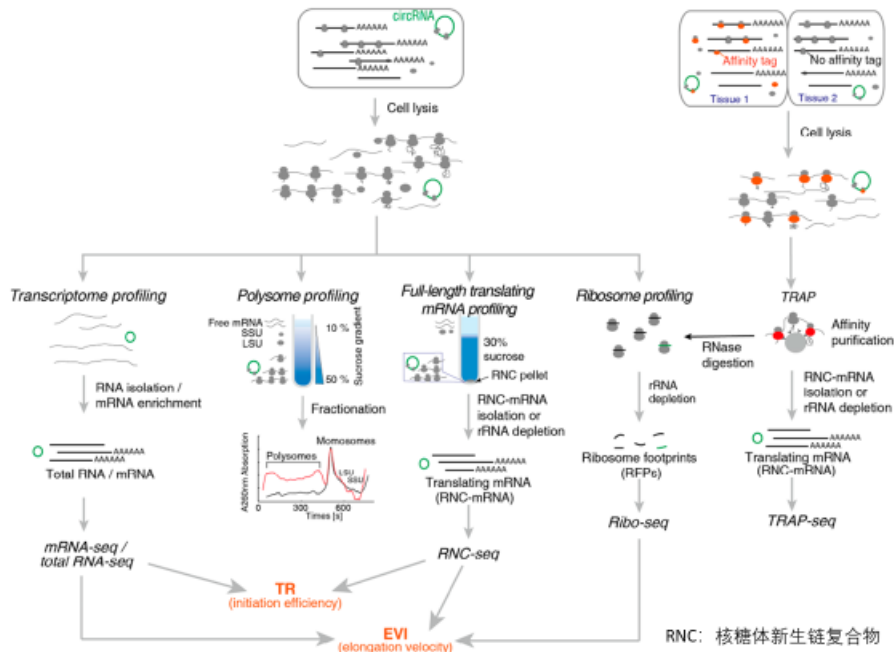
若是谈到对 Ribo-seq 结果的间接呈现，就不得不提翻译效率的计算了，这也是前面提到的为什么 Ribo-seq 和 RNA-seq 常常会同时进行的原因。简单来说，翻译效率的计算公式，就是反映了测序数据中 RNA 的“使用率”——在总 RNA 中有多少 RNA 结合了核糖体或是被翻译。通过翻译效率的计算，可以在某种程度上判断一段序列是否具有编码的能力，但这并不是绝对的，还需要后续实验的佐证。上图展示了 4 种 mRNA 在不同区域的核糖体密度分布情况和翻译效率。“Ribosome”是指不经任何处理下核糖体关联的读长覆盖情况；“Cycloheximide”是指在放线菌酮处理下核糖体关联的读长覆盖情况；“mRNA”就是常规的 RNA-seq 结果；“CSF”则起到一个判断序列保守性的作用，众所周知编码序列的保守程度是要远高于非编码序列的，所以这里把它纳入作为一个考量因素。结果显示，像 CDS、5’ UTR 这类翻译活跃的区域所计算出来的翻译效率值会显著高于 3’ UTR。

### 三、其他应用于研究正在翻译的 RNA 的方法

第三部分在这里我想要再聊一下其他用于研究正在翻译的 RNA (核糖体新生链复合物中的 mRNA, RNC-mRNA) 的方法。由于 RNA 是蛋白质合成的蓝图，所以研究正在翻译的 RNA 就成为了翻译组学的首要任务。目前发展了几种比较经典的方法用来分析正在翻译的 RNA 的不同特征，这些方法的基本原理如图所示。



## 其他用于研究正在翻译的RNA的方法



首先是多聚核糖体图谱技术 (Polysome profiling)，这个技术可以说是核糖体图谱技术的一个前身，主要依靠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和 mRNA 的复合物。当一条 mRNA 上结合的核糖体数目越多时，它在离心时的沉降速率就会越快，因此结合有不同数量核糖体的 mRNA 就可以通过离心在溶液中分开，使用 Northern 杂交、微阵列或 RT-PCR 等方法对分离出来的各组分中的 mRNA 进行分析，就可以反映转录本翻译的分布。这个技术通常用于检测翻译中的大变化，其主要局限在于难以对所有的 RNC-mRNA 进行深入分析，这是由于蔗糖梯度的体积大，每个组分中 RNC-mRNA 的浓度都很低，回收的总量一般只够 RT-PCR 定量，除非使用大量的起始材料，否则很难进行高通量分析。

然后是正在翻译 mRNA 的全长测序技术 (RNC-seq)。该技术能有效地解决上述多聚核糖体图谱技术出现的问题，主要是将细胞裂解物负载至单一浓度的蔗糖缓冲液中，通过冷冻超速离心分离所有的细胞组分，并使 RNC 沉积在底部。而 RNC-mRNA 则可从沉积的 RNC 中直接抽提回收，有效避免了高浓度蔗糖的干扰。RNC-seq 可以更有效地揭示正在翻译 mRNA 全长信息，包括丰度和类型等。但是由于核糖体与 mRNA 是非共价结合的，所以 RNC 非常脆弱，在细胞裂解后容易发生解离或降解。因此该技术的难点也在于如何分离完整的 RNC，处理不当就得不到全长的 RNC-mRNA，从而造成定量的失准。

最后我们来看一下正在翻译的核糖体亲和纯化测序技术 (TRAP-seq)。该技术主要通过组织特异性的启动子表达核糖体大亚基的 L25 蛋白，并在 C 端连接亲和标签 (如多聚组氨酸、绿色荧光蛋白等)，最后利用抗体将含有标签的核糖体抓取下来，从而获得结合在此类核糖体上的 RNC-mRNA。该技术由于无需使用超速离心，所以分离所得的核糖体不会被与之共沉淀的其他细胞组分污染，而且具有在复杂组织里从特定细胞类型中分离 RNC-mRNA 的优势。但是这个技术的缺点也很明显，如果对象是细胞，那么就必须要稳定地表达这些被标记的核糖体蛋白；如果对象是动植物，构建稳定的转基因品系也是必须的。与此同时，这些被标记的核糖体蛋白的过量生产也有可能改变它们的结构和性质，所以该系统不再处于正常的生理条件之下。

## 其他用于研究正在翻译的RNA的方法

	Ribo-Seq	Polysome Profiling	RNC-Seq	TRAP-Seq
可获得的RNC-mRNA	被核糖体保护的片段	全长	全长	全长
可选用的高通量检测方法	下一代测序	微矩阵/下一代测序	微矩阵/下一代测序	微矩阵/下一代测序
通量需求	高	低	低	低
读长需求	22–35nt	任意	任意	任意
统计单个mRNA上的核糖体数量	X	√	X	X
获取核糖体的位置、密度分布	√	X	X	X
鉴定ORF和uORF	√	X	X	X
组织特异性	X	X	X	√
在正常生理条件下	√	√	√	X
实验繁琐程度	复杂	简单	简单	复杂

## 其他用于研究正在翻译的RNA的方法

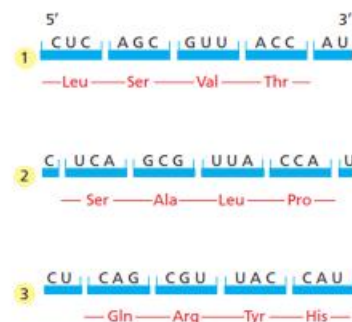
### Ribo-seq的局限性

- 实验处理引入的数据失真  
核糖体图谱技术的关键是迅速抑制翻译，以捕获核糖体的瞬时状态。这样的处理可能会引入干扰。
- 无法精确表示蛋白质的合成速率  
核糖体图谱技术本质上只是检测了RNA上的平均核糖体密度，对RNA翻译速率的测量要在完成翻译后才能真正的实现。
- 印迹大小片段的污染  
核糖体图谱技术极容易受到印迹大小的非编码RNA或rRNA的污染，它们会给最终的数据带来干扰。
- 样本数目和测序通量  
目前，核糖体图谱技术所需要的样本输入量高于RNA-seq，对测序通量的要求也比较苛刻，无法用于单细胞研究。

这里列出了刚才提到的 4 种用于研究正在翻译 RNA 的方法在各方面的对比，可以看出来 Ribo-seq 相比其他技术虽然流程更加复杂，对读长和通量的要求也更加苛刻，但是却能够获得更多的信息。当然一个技术肯定不是百分之百完美的，在这里也借机总结一下它的局限性和不足之处：首先，由于 Ribo-seq 是基于迅速抑制翻译的原理来捕获核糖体的瞬时状态，所以相应的实验处理可能会引入干扰导致数据的失真；再者，Ribo-seq 本质上只是对 RNA 的平均核糖体密度进行检测，其实无法精确地测量蛋白质的合成速率；其次 Ribo-seq 的实验过程中一定要慎防印迹大小的非编码 RNA 或 rRNA 片段的污染，它们会为后续分析带来极大

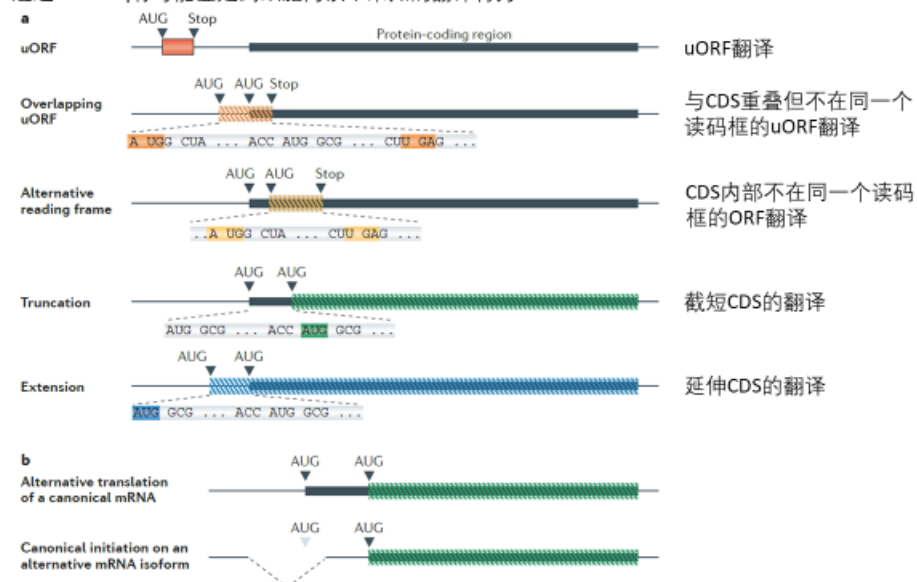
## Ribo-seq的应用

- Extensive upstream initiation at non-AUG codons in yeast<sup>3</sup>
- Extensive upstream initiation at non-AUG codons in mammalian cells and translation at the 5' end of some long interspersed non-coding RNAs (lincRNAs)<sup>4,21,22</sup>
- Many upstream open reading frames (uORFs) that start at AUG and non-AUG codons in yeast, the translation of which is induced by meiosis<sup>19</sup>
- Diverse short reading frames in human cytomegalovirus<sup>29</sup>



Ingolia, N. T., *Nature Reviews Genetics*, 2014, 15(3):205.

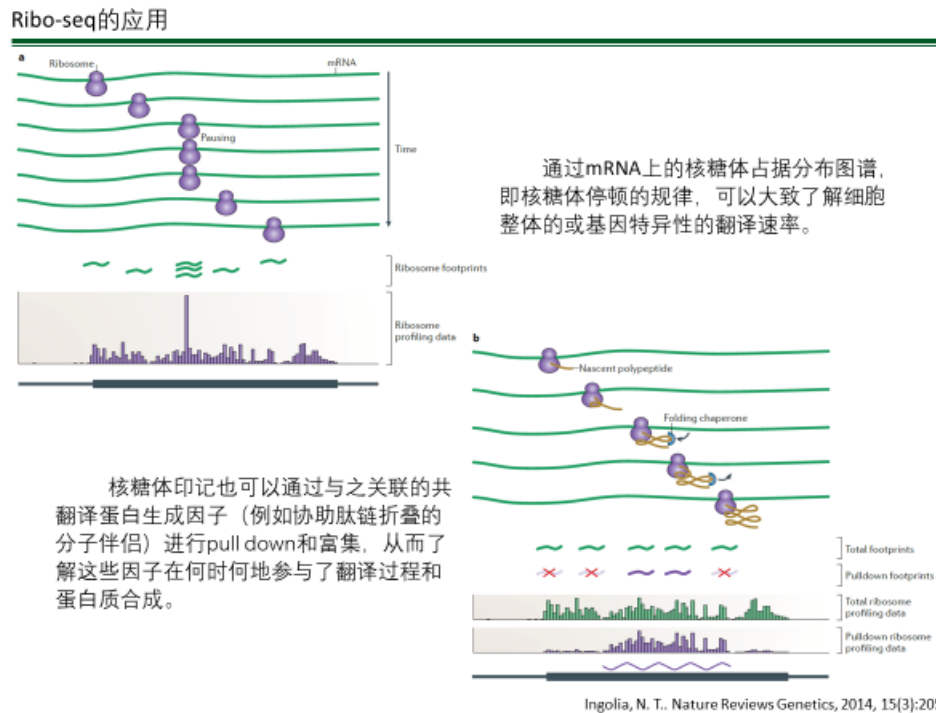
通过Ribo-seq有可能鉴定到细胞内以下未知的翻译行为



Ingolia, N. T., Nature Reviews Genetics, 2014, 15(3):205.

在了解了 Ribo-seq 的实验原理和基本流程之后，最后一部分我们来谈一下 Ribo-seq 的应用，或是说研究者们通过 Ribo-seq 的数据能挖掘出什么新的发现。首先 Ribo-seq 技术最主要的应用就是为那些原来没被注释过的可以翻译的序列进行注释。这里再引入一个概念，就是上文频繁提到的 ORF 可变读码框，所谓的可变其实是在 RNA 上指随着密码子相对碱基

位置的偏移可以产生 3 种翻译阅读框，因此原则上同一个 RNA 序列可以指定三个完全不同的氨基酸序列，但现实中往往只有一个可以被真正编码蛋白。Ribo-seq 技术的出现极大拓展了人们对 ORF 翻译潜能的认识，并有可能鉴定到以下细胞内未知的翻译行为，包括 uORF 翻译、与 CDS 不在同一个读码框的 ORF 翻译、CDS 的截短延伸以及图中还未提及的曾被注释为非编码 RNA 的翻译行为。

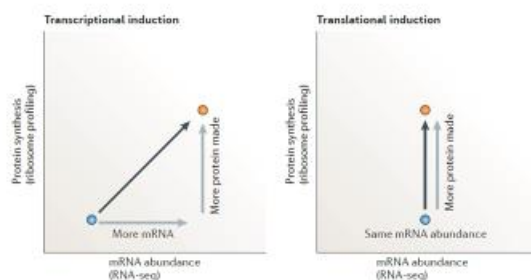


其次，Ribo-seq 技术还可以用于研究蛋白质合成的机制。譬如左上的示意图通过 mRNA 上的核糖体占据分布图谱，即核糖体停顿的规律，我们可以大致地了解细胞整体的或某个特定基因的翻译速率；又如右下的示意图，核糖体印记也可以通过与之关联的共翻译蛋白生成因子（像协助肽链折叠的分子伴侣）进行富集，从而了解这些因子在何时何地参与了翻译过程和蛋白质合成。

## Ribo-seq的应用

### 应用三——研究基因表达在翻译层面上的调控

- MicroRNAs (miRNAs) affect both mRNA abundance and translation of target genes<sup>91</sup>
- Translational repression of ribosome biogenesis and increased translation of membrane proteins as embryonic stem cells differentiate into embryoid bodies<sup>14</sup>
- miRNAs first repress translation and then promote mRNA turnover<sup>92</sup>
- Active site mammalian target of rapamycin (mTOR) inhibitors translationally repress ribosome biogenesis<sup>57</sup>
- Just-in-time translational induction in yeast meiosis and sporulation; translational control by transcriptional changes in mRNA 5' leaders<sup>19</sup>
- In *C. elegans*, miRNAs simultaneously decrease mRNA abundance and translation<sup>93</sup>
- Only uORFs that start at AUG codons repress translation<sup>9</sup>
- Out-of-frame internal initiation induces nonsense-mediated mRNA decay<sup>18</sup>



通过比对RNA-seq和Ribo-seq的数据可以区分基因表达调控的不同模式：转录诱导调控导致相应的RNA丰度和蛋白质合成同时增加；翻译诱导调控只会引起蛋白质合成的增加。

Ingolia, N. T. Nature Reviews Genetics, 2014, 15(3):205.

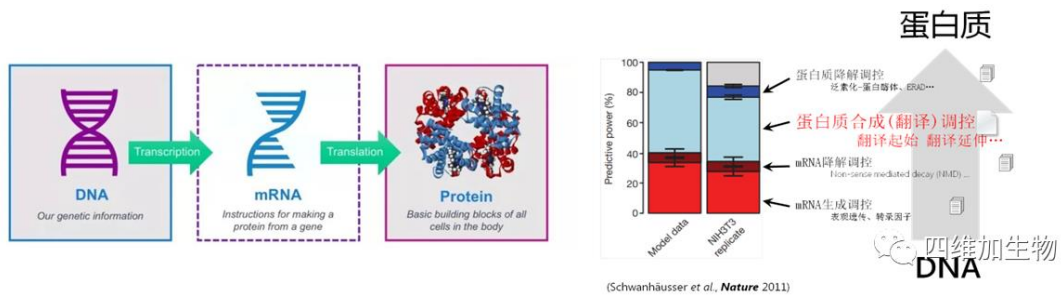
最后，Ribo-seq 技术还可以用于研究基因表达在翻译层面上的调控。下面这张示意图就是通过比对 RNA-seq 和 Ribo-seq 的数据以区分基因表达调控的不同模式：若基因受到转录诱导调控，必然会导致其相应的 RNA 丰度和蛋白质合成同时增加；而若基因只受到翻译诱导调控，那么就只会引起蛋白质合成的增加。



## 多聚核糖体图谱技术

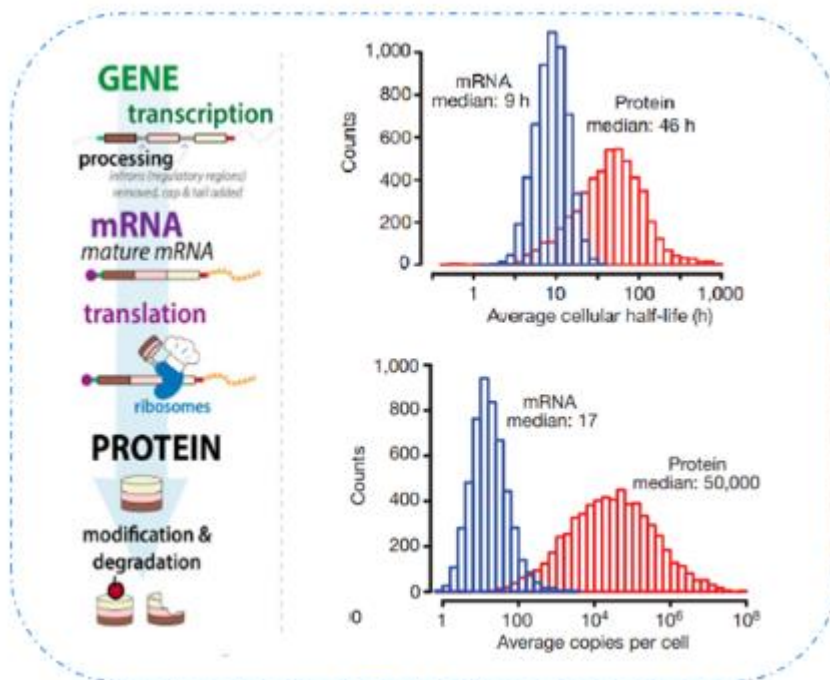
### 一、翻译调控是细胞内最重要的调控方式

- 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、翻译调控、蛋白质降解调控是个主要调控阶段；
- 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；
- 翻译组学的研究对生物工程、合成生物学等应用科学具有直接的指导意义。



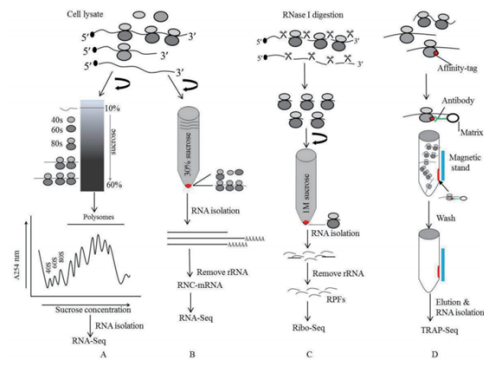
### 二、翻译组研究更具有理论和实践意义

- 蛋白质是几乎所有生理过程的实际执行者，是生物体功能的重要体现者；
- 转录组学结果和蛋白组结果二者间相关性较低，匹配度较差，无法对基因和蛋白的功能及调控途径进行最为有效的机制分析；
- 翻译组是蛋白组的最佳体现者，是连接转录组和蛋白组的桥梁，研究翻译组有助于更好地阐释从 mRNA 到蛋白质之间翻译水平调控机制研究。



### 三、Polysome profiling 是最经典的翻译组检测技术

## Polysome profiling是最经典的翻译组检测技术



A: 依次为多聚核糖体图谱技术; B: 翻译谱分析技术; C: 核糖体图谱技术; D: 核糖体电泳和电化技术

图1 翻译组研究技术流程

研究翻译组的主要技术有:

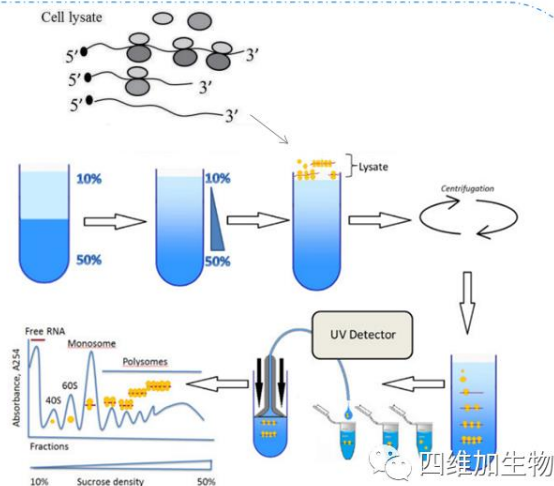
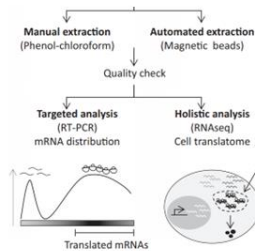
技术比较	Polysome profiling	RNC-RNS-seq	Ribosome profiling	TRAP
同一mRNA上核糖体数量	✓	×	×	×
获取翻译状态、核糖体位置信息	×	×	✓	×
获取非编码区、翻译调控信息	✓	✓	×	✓

Polysome profiling是研究翻译组的经典技术,

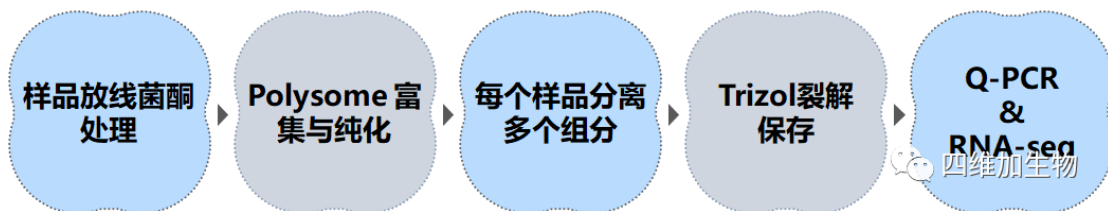
是评价翻译效率的“黄金标准”

## Polysome profiling的技术原理

- 该技术利用蔗糖密度梯度离心的方法分离多聚核糖体。
- 将蔗糖溶液中的RNA进行分离纯化,判断待测基因在细胞内的翻译状况。

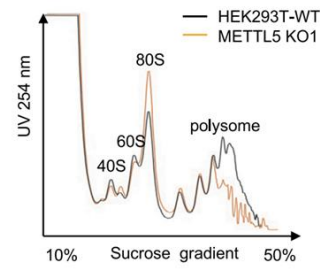
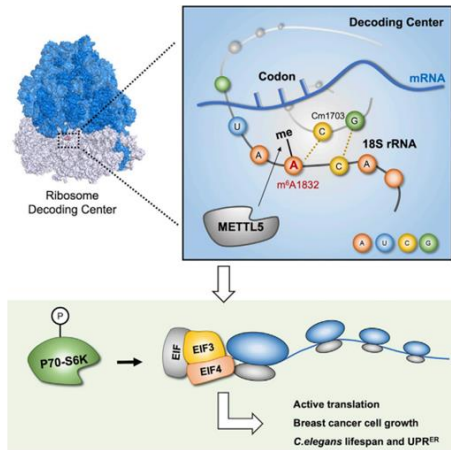


## Polysome profiling的技术路线



## 四、Polysome profiling 的应用研究

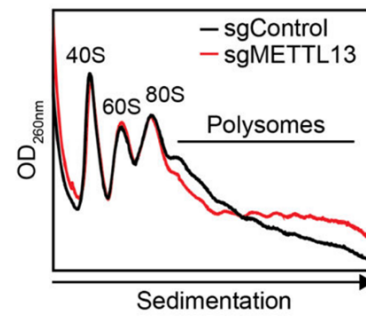
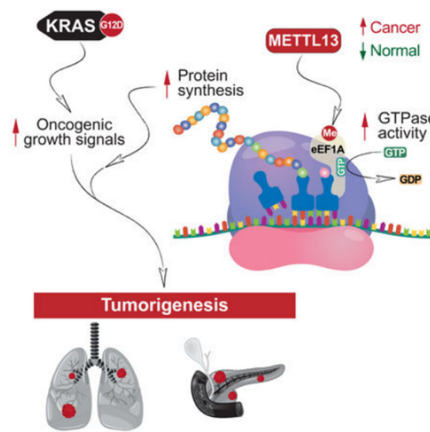




Polysome profiling证实METTL5缺失导致细胞多核糖体部分的显著减少，80S单核糖体的积累，整体翻译水平下降。

核糖体18S m<sup>6</sup>A甲基转移酶METTL5促进翻译起始和乳腺癌细胞生长

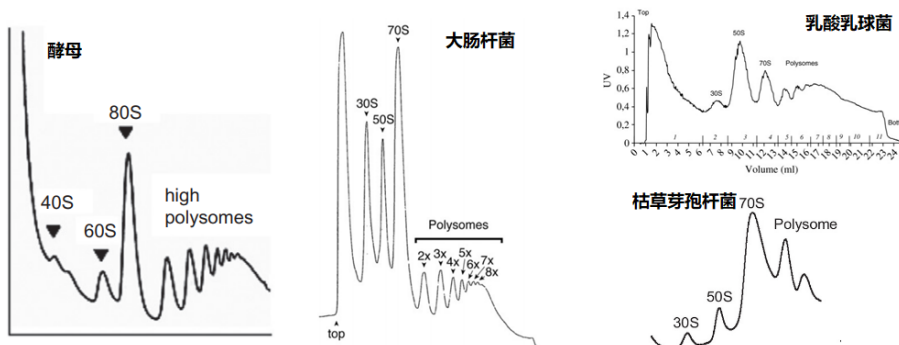
Cell Rep. 2020 Dec 22;33(12):108544.



使用polysome profiling的方法，检测在METTL13敲除和WT细胞中polysome分布是否差异显著。结果发现，在敲除细胞中，heavy polysome/light polysome的比值升高，提示METTL13敲除确实影响了蛋白翻译过程。

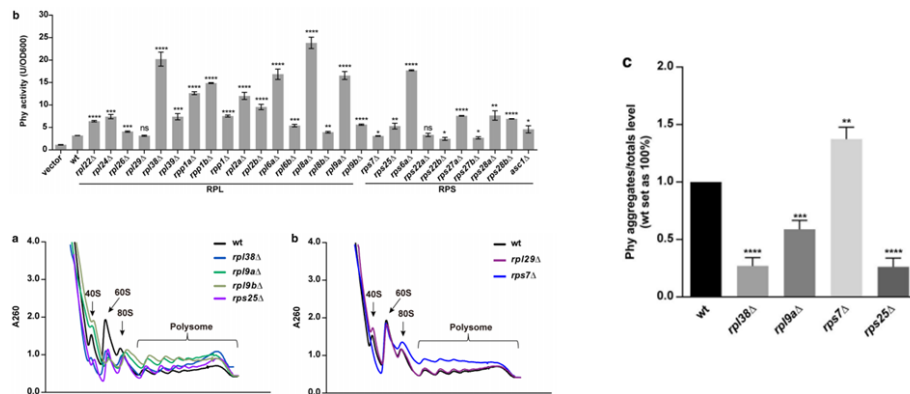
mettl13介导的eEF1A55二甲基化促进细胞蛋白质合成

Cell. 2019 Jan 24;176(3):491-504.e21.



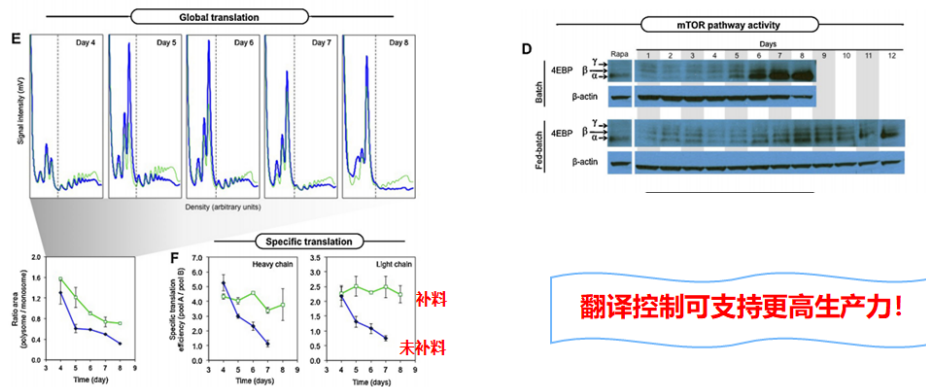
多聚核糖体技术被广泛用来研究动物、植物及酵母细胞的 mRNA 的翻译状况，为生物改造提供了可能

## 翻译组调控显奇功：核糖体蛋白“裁员”提高外源蛋白表达的折叠效率



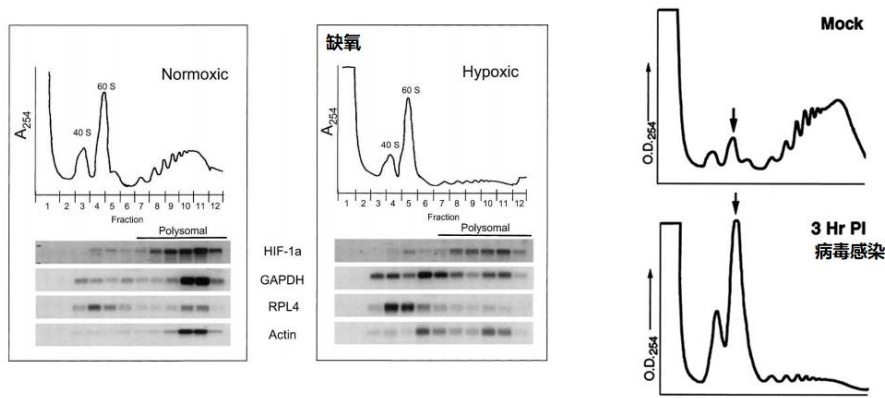
- 毕赤酵母中多个核糖体蛋白敲除可使外源蛋白的表达活性成倍升高；
- “增强”菌株中有更低的 60s 核糖体亚基和更高比例的多核糖体，翻译延伸速度较慢；
- “增强”菌株中的新生肽链折叠状态更好，更能耐受蛋白酶的攻击。

## 补料分批培养可保持CHO 细胞更高的翻译活性，提高重组蛋白产量



翻译控制可支持更高生产力！

- 营养补充后多核糖体含量增加，多核糖体向单体的大部分转变发生在第 5 天；
- 与分批培养相比，补料后的翻译活性提高 2 倍，第 5 天差异最大。



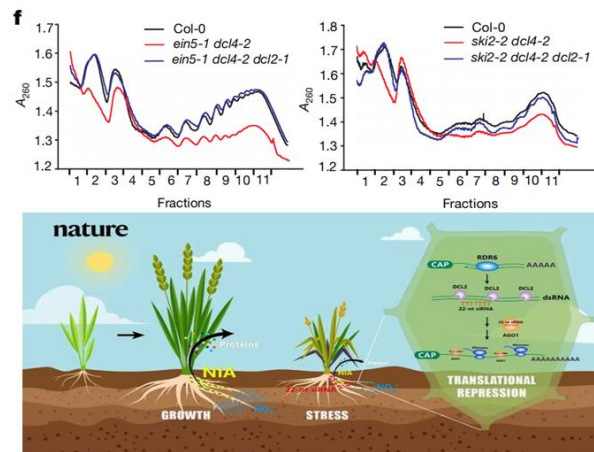
RNA. 2007 Jul;13(7):1116-31.

Proc Natl Acad Sci U S A. 1999 Nov 9;96(23):13118-23.

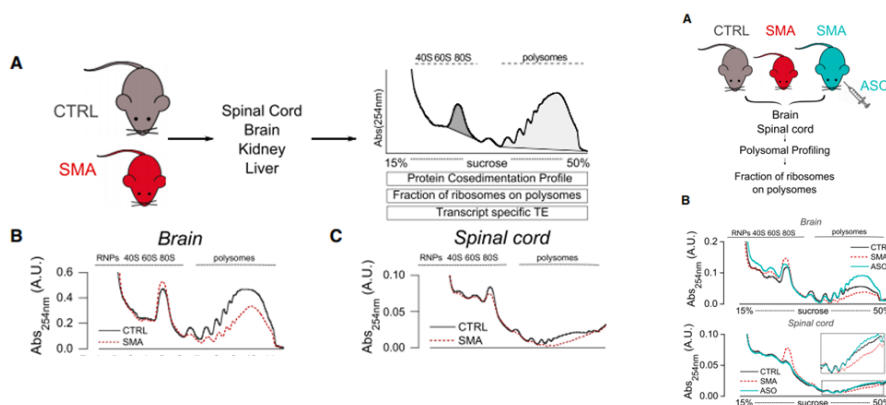
多聚核糖体技术被广泛用来研究生物体在逆境胁迫下的翻译应答

### 植物22-nt siRNA介导翻译抑制与胁迫适应

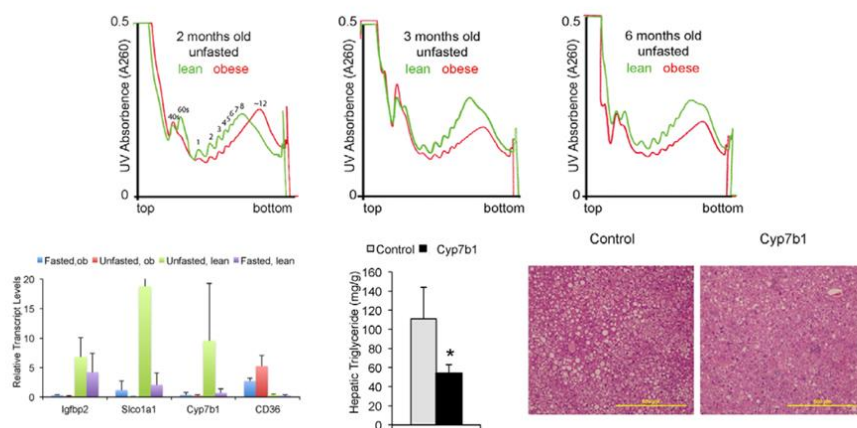
对22nt siRNA积累的植株进行Polysome profiling分析, 结果显示多聚核糖体数量明显减少。这说明22-nt siRNA主要在翻译水平上抑制基因的表达



植物在逆境条件下, 可能通过调控 22-nt siRNA 的水平, 控制细胞中的蛋白翻译, 抑制生长, 减少能量消耗, 同时激活自身胁迫翻译, 保证自身生存。



通过对 SMA 小鼠进行转录组和翻译组 (Polysome profiling) 的比较研究, 在疾病早期, 一系列 RNA 的翻译已发生了明显的变化, 揭示了 SMN 蛋白不足导致机体翻译水平的改变是造成 SMA 的主要原因。



- 肥胖小鼠肝脏的 ER 表现出蛋白质合成普遍减少，对多核糖体结合转录物的综合分析揭示了肥胖翻译组中蛋白质合成机制、线粒体成分和胆汁酸代谢的广泛下调；
- 组织多聚体图谱可以探索在营养摄入和疾病病理反应中动态调节的内稳态反应的新机制。

文献阅读建议：Deficiency of ribosomal proteins reshapes the transcriptional and translational landscape in human cells