

The background of the slide features a series of dynamic, overlapping brushstrokes in various shades of green, ranging from light lime to dark forest green. These strokes are primarily oriented diagonally, creating a sense of movement and energy. The strokes are most concentrated on the left side of the slide and become sparser towards the right.

用流式细胞仪定量分析哺乳动物翻译起始位点

主讲人：占传松（植科）李舟（生科）

核糖体起始复合物对AUG的识别过程

Eukaryotic initiation uses several complexes

43S complex
eIF2, eIF3
Met-tRNA_i



Cap-binding complex + mRNA
eIF4A, B, E, G



43S complex binds to 5' end of mRNA



48S complex forms at initiation codon
eIF2, EIF3
eIF1, 1A
eIF4A, B, F

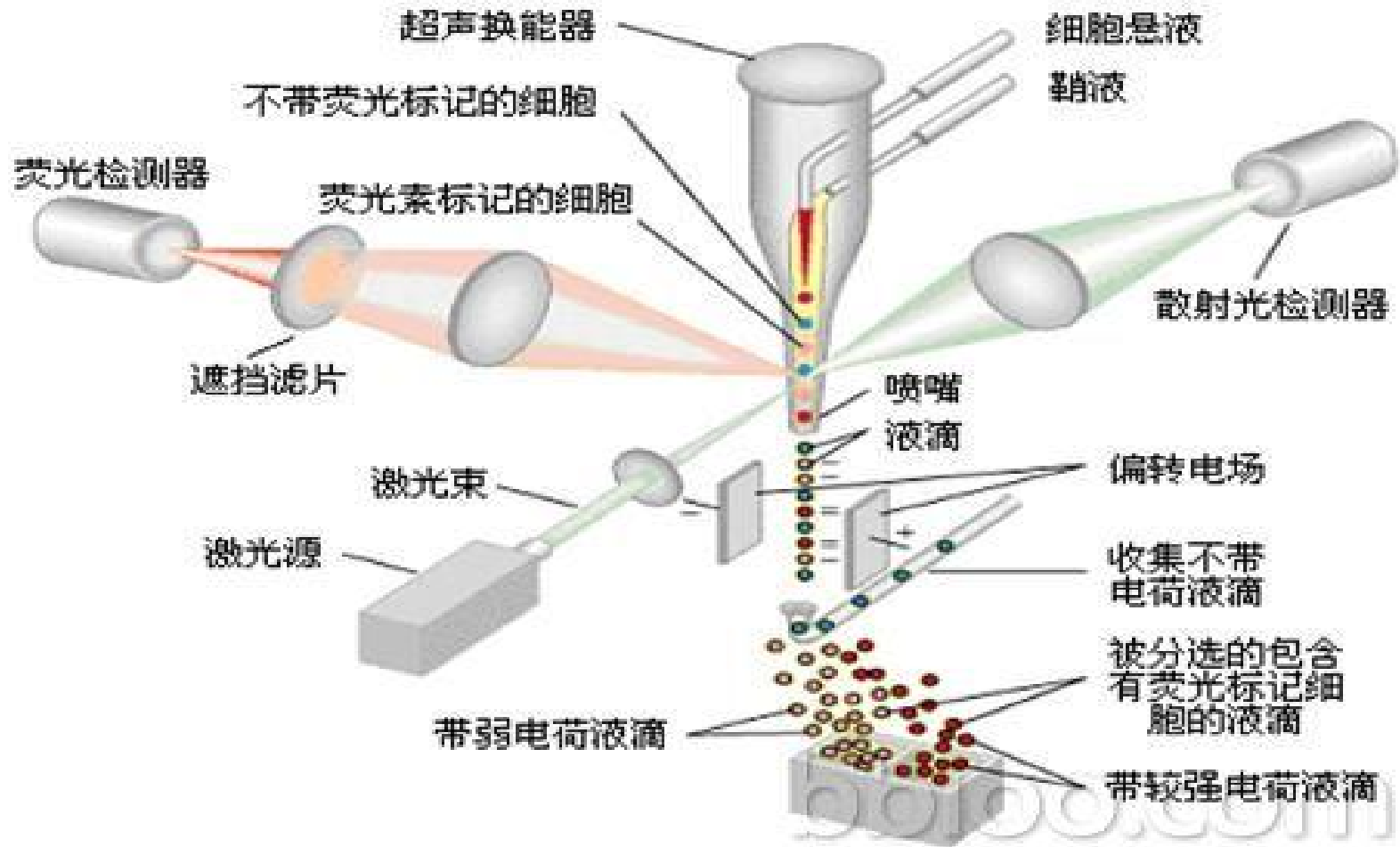


实验方法：FACS-seq

1. 一种结合了流式细胞分选术和高通量DNA测序的新方法。

高通量DNA测序技术视频简介。

流式细胞分选术原理图



报告基因 (**reporter gene**)

1. 报告基因 (reporter gene) 是一个其表达产物非常容易被鉴定的基因。把它的编码序列和基因表达调节序列相融合形成嵌合基因，或与其它目的基因相融合，在调控序列控制下进行表达，从而利用它的表达产物来标定目的基因的表达调控，筛选得到转化体。



实验目标：

1. 检测每个可能的含有AUG的翻译起始序列的翻译水平并建模。
2. 确定哪些其他实验变量影响了我们的翻译起始水平测量的程度。
3. 进行的人类TIS效率的定量全基因组分析。
4. 运用TIS序列空间的全面分析，使我们能够搜索候选肿瘤突变驱动。
5. 运用‘漏’核糖体扫描模型分析蛋白质组多样性。

位置权重矩阵 (PWM)

1. 单分子PWM:

$$E = k \cdot \exp\left(\sum_{i=-6}^{+5} C_{b,i}\right)$$

where k was a proportionality constant and $C_{b,i}$ was the coefficient for base $b = \{U, C, A, \text{ or } G\}$ at position $i = \{-6, -5, -4, -3, -2, -1, +4, \text{ or } +5\}$. The coefficients were normalized such

其中k是比例常数

$C_{b,i}$ 是碱基 $b = \{U, C, A \text{ 或 } G\}$ 在 $i = \{-6, -5, -4, -3, -2, -1, +4 \text{ 或 } +5\}$ 的位置系数

双分子位置权重矩阵

1. 双分子位置权重矩阵：

$$E = k \cdot \exp \left(\sum_{i=-6}^{+5} C_{b,i} + \sum_{i=-6}^{+5} C_{b_1,i_1,b_2,i_2} \right)$$

where $C_{b,i}$ matched the mononucleotide PWM coefficients, and C_{b_1,i_1,b_2,i_2} was the interaction coefficient for base b_1 at position i_1 and base b_2 at position i_2 (subject to $i_1 \neq i_2$). The coefficients were

$C_{b,i}$ 相当于单核苷酸的PWM系数， C_{b_1,i_1,b_2,i_2} 是碱基 b_1 在 i_1 和碱基 b_2 在 i_2 位置上的相互作用系数

目标一的实现方法：

1. 运用基因报告系统，该系统的运用减少了外在噪音从而提高了检测的准确度。
2. 报告文库的建立。对于文库的建立有一个要求：文库要足够大 $65,536$ (4^8)，要能代表99%以上的可能性。(NNNNNNAUGNN)
3. 运用FACS-seq法对新建文库进行分离筛选和相关分析。
4. 建立位置单核苷酸权重矩阵模型 (position weightmatrix)。

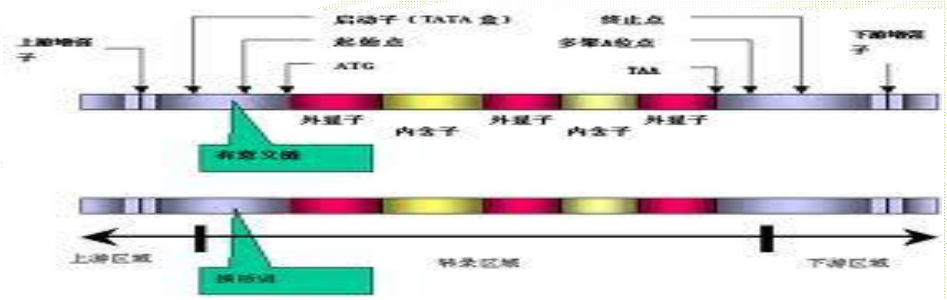
目标二的实现方法：

1. 在多种细胞系中观察翻译起始再现的水平是否相同或相似. 结果发现尽管其它因素能够影响翻译起始效率，但我们的实验表明，我们的报告系统是足以评估在不同实验条件的翻译起始效率相对水平。
2. 通过计算报告文库中的与每个TIS序列相连接的65, 536个mRNA序列折叠能量来判断mRNA的二级结构是否会对TIS报告序列的表达水平有影响。（结果是无明显影响）

目标三的实验方法：

1. 构建双核苷酸位置权重矩阵模型（二核苷酸的PWM）。对所有编码人类蛋白TISs分析起始效率。
2. 比较蛋白质编码序列向上游开放阅读框（uORF）TISs的起始效率。

开放阅读框



1. 开放阅读框（英语：Open reading frame；缩写：ORF；其他译名：开放阅读框架、开放式阅读框架，开放读架等）是生物个体的基因组中，可能是蛋白质编码序列的部分。基因中的ORF包含并位于开始编码与终止编码之间。由于一段DNA或RNA序列有多种不同读取方式，因此可能同时存在许多不同的开放阅读框架。开放阅读框包含一段可以编码蛋白的碱基序列，不能被终止子打断。

目标四的实验方法：

1. 在体细胞突变的癌症目录（COSMIC）中筛选除-3或+4的其他位置上的体细胞突变TIS。通过使用可公开获得的基因表达数据，基因拷贝数数据和文献检索来评估候选突变。最后挑选出7个TIS突变子，并通过常规的流式细胞术分析，进而观察我们测量的TIS表达有效性值和我们预测的有效性值是否一致。

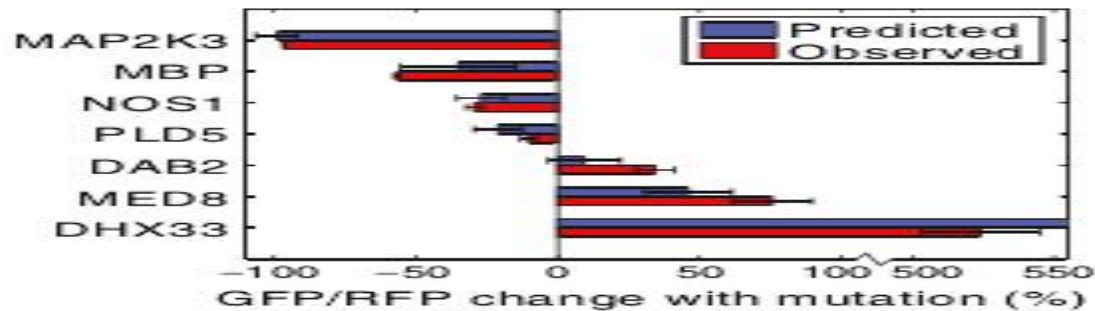


Figure 7. Effect of mutations on TIS reporter.

TIS sequences from Table 1 (both with and without mutations) were inserted into the TIS reporter. The TIS sequences were analyzed individually by conventional flow cytometry, and the change in TIS efficiency caused by the TIS mutation was calculated. The observed changes in TIS reporter efficiencies were compared to the dinucleotide PWM predictions. Error bars represent 95% CI ($n = 3$). The predicted change in TIS efficiency for DHX33 was 798% (off-scale). Source data are available online for this figure.

搜索候选肿瘤突变驱动



目标五的实验方法：

1. 要识别N端截短蛋白异构体，我们开发了将我们的TIS效率值来预测截断ORF的相对平移的定量漏扫描模型。

$$X = \frac{(100 - kE_1)}{100E_2}$$

（参数E1和E2是由第一和第二TIS 序列来分别确定的相对效率值，常数k涉及到相对效率值起始的绝对可能性）K=0.86

从我们的模型预测表明，‘漏’核糖体扫描大致增加了蛋白质组多样性。尖杉酯碱排除干扰。

此篇文章之我见

1. 创新点：这篇文章最大的亮点在于运用了一种结合了流式细胞分选术和高通量DNA测序的新方法，使得对所有可能的哺乳动物的TIS序列将进行系统分析成为可能。
2. 启发点：现代生物科技的发展不再是单一、片面，而需要从整体和系统的角度去把握。
3. 改进点：由于数据的噪音，我们目前还不能肯定，任何单一的价值可能恰恰代表起始效率。另外流式细胞仪和下一代DNA测序仪都十分昂贵。因此想从事相关研究必需要有较高的科研平台。

Thank you !

谢谢观看！