



蛋白质动力学揭示基因表达噪声来源

廖 浩 (报告)

张晓燕 (翻译)

主要内容

- 1.背景介绍
- 2.基因表达模型
- 3.模型验证
- 4.总结

1. 背景介绍

■ 基因表达噪声

相同的基因在表达上出现细胞间的差异，即基因表达出现一定的随机性

■ 影响

通过研究发现基因表达噪音深刻影响着生物功能，因此在基因回路中精确找到噪声来源具有重要意义

1.背景介绍

■ 噪声来源

■ 外噪声

主要产生于细胞间的差异，如RNA聚合酶或者核糖体等细胞原件的差异

■ 内噪声

启动子重建、转录、翻译、mRNA和蛋白种类退化相关的生化反应

1. 背景介绍

- 虽然已有几种方法可以有效分离外来噪音和内部噪音，但是能够区分不同内部噪音的方法还非常不足。
- 目前通过实验验证内噪声主要来源有：
 1. 由概率合成产生的mRNA泊松波动和独立mRNA转录子的衰减（mRNA增减波动）
 2. 不同转录状态间的启动子转换（启动子波动）

寻找并发展出一种方法，在假设蛋白质水平是系统唯一的可观察状态
的条件下，来确定mRNA增减波动与启动子波动对基因表达内部噪音
的影响

1.背景介绍

- 在一定水平，mRNA增减波动与启动子波动两者能够产生相似的稳态蛋白质噪音水平，故利用细胞的蛋白丰度的稳态分布识别内部表达噪音来源是不够的。

- 单分子FISH

在原位杂交（单分子FISH）中加入单分子荧光的方式去计算单细胞内独立mRNA转录子的数量

但是它要求独立mRNA分子足够长，能够在上面结合至少20个荧光标签

2. 基因表达模型

$$CV^2 = \text{mRNA birth/death fluctuations} + \text{promoter fluctuations}$$

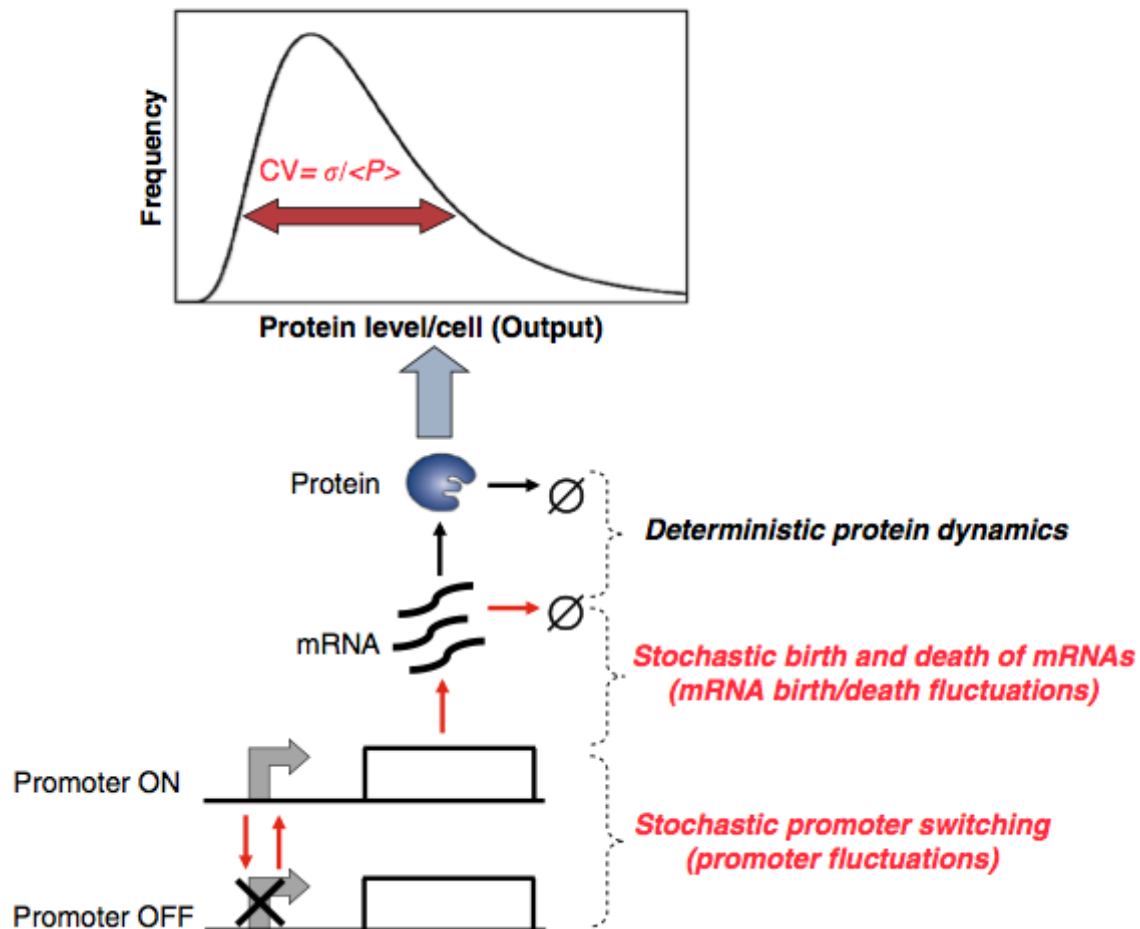


图1 含两种不同噪音来源的基因表达的随机混合模型的原理图

2. 基因表达模型

- 每一个转录事件都带来一次几率B这么多的mRNA分子的爆发，其平均值用 $\langle B \rangle$ 表示，代表一个预期值
- 如果 $m(t)$ 表示mRNA在时间 t 时的数量， $P_j(t)$ 表示在时间 t 时mRNA的数量为 j 的几率，则

$$\begin{aligned}\frac{dP_0(t)}{dt} &= \gamma_m P_1(t) - k_m P_0(t) \\ \frac{dP_j(t)}{dt} &= \sum_{i=1}^j k_m P_{j-i}(t) \alpha_i + \gamma_m (j+1) P_{j+1}(t) \\ &\quad - (\gamma_m j + k_m) P_j(t), \quad j \in \{1, 2, \dots\}\end{aligned}\quad (1)$$

k_m 代表转录活动的频率， γ_m 代表mRNA衰减速率

2. 基因表达模型

- mRNA数量的波动传播到下游影响蛋白质水平，蛋白质水平被假定为系统可观测的唯一状态，忽略由蛋白分子随机增减而产生的泊松噪音。那么，蛋白质动力学公式可确定为：

$$\frac{dp(t)}{dt} = k_p m(t) - \gamma_p p(t) \quad (2)$$

k_p 表示mRNA转录速率， γ_p 表示蛋白质降解速率， $p(t)$ 表示在时间 t 时的蛋白数量。那么，公式(1)和(2)构成了随机混合基因表达模型，其中当蛋白水平不断变化且可确定时，mRNA随时间的变化是离散和随机的。

2. 基因表达模型

- 蛋白表达噪音可通过变异系数的平方值被量化，定义为 $CV^2 = \sigma^2 / \langle p \rangle^2$ ，其中 σ^2 表示蛋白质水平的变化， $\langle p \rangle$ 表示蛋白丰度的平均值

$$\overline{CV^2} = \frac{\eta_m k_p}{(\gamma_m + \gamma_p) \langle p \rangle} = \underbrace{\frac{k_p}{(\gamma_m + \gamma_p) \langle p \rangle}}_{\text{mRNA birth/death fluctuations}} + \underbrace{\frac{(\eta_m - 1) k_p}{(\gamma_m + \gamma_p) \langle p \rangle}}_{\text{Promoter fluctuations}} \quad (3)$$

其中

$$\langle p \rangle = \frac{\langle B \rangle k_p k_m}{\gamma_p \gamma_m} \quad (4)$$

$$\eta_m = \frac{\langle B^2 \rangle + \langle B \rangle}{2 \langle B \rangle} \quad (5)$$

表示mRNA种类数量的稳态法诺因子（即噪音强度）

2. 基因表达模型

- 公式(3)和(4)显示，增大转录活动的频率 k_m 能够增加稳态平均蛋白丰度，但是降低稳态噪音水平 $CV^2 \times \langle p \rangle$ 保持固定。
- 因此 η_m 正比于公式(3)中比例因子，可以通过变换 CV^2 和 $\langle p \rangle$ 的比例来确定
- 而且，比例因子同样依赖于mRNA的翻译速率 k_p 以及蛋白质和mRNA的半衰期，因此，这个方法需要对这些参数的先验知识，此外，量化蛋白质丰度需要测量绝对分子数量而不是荧光强度
- 故仅仅靠蛋白噪音大小的稳态测量是不足以确定 η_m 值的，也不能够区分内部噪音的不同来源

2. 基因表达模型

- 小分子药物可特异性地阻碍转录和翻译。在应答一个完整的对转录和翻译的阻碍活动，平均蛋白水平 $\langle p(t) \rangle$ 以如下公式在衰减：

$$\frac{\langle p(t) \rangle}{\langle p(0) \rangle} = \exp(-\gamma_m t) \text{ or } \frac{\langle p(t) \rangle}{\langle p(0) \rangle} = \frac{\gamma_p \exp(-\gamma_m t) - \gamma_m \exp(-\gamma_p t)}{\gamma_p - \gamma_m}, \quad (6)$$

t代表从添加药物起始的时长。因此，mRNA和蛋白的半衰期能够通过测量 $\langle p(t) \rangle$ 来量化

通过研究了蛋白质水平（变异系数的平方）在应答这些小分子药物的变化时细胞与细胞之间怎样发生变异，这些变化为未知的系统参数提供了信息，例如 η_m

2. 基因表达模型

- 当基因表达出现转录阻碍，设当时间 $t=0$ 时转录活动的频率 $k_m=0$ ，假设系统在阻碍前处于稳态，蛋白噪音水平（通过变异系数的平方 $CV^2(t)$ 测得）的变化如下列公式：

$$\frac{CV^2(t)}{CV^2} = f(\gamma_m, \gamma_p, \eta_m, t) \quad (7)$$

CV^2 是公式(3)给定的初始蛋白噪音水平，

$f(\gamma_m, \gamma_p, \eta_m, t)$ 是关于 t 的一个单调递增函数

- 随着转录阻碍，mRNA水平以指数形式降低，导致细胞间变异的增多，mRNA噪音的增大会传播到编码的蛋白质

2. 基因表达模型

- 表达噪音主要来自启动子波动时，mRNA的数目足够多，mRNA就会发生降解。因此，当转录活动被阻碍之后，mRNA的降解大致可确定为发生在单一细胞内，这类细胞的蛋白质水平降低并产生一个缓慢增大的 $CV^2(t)$ 。
- 当表达噪音来自mRNA增减波动时，转录阻碍会引起细胞的mRNA随机降解，这类细胞会发生蛋白噪音水平快速增加。

2. 基因表达模型

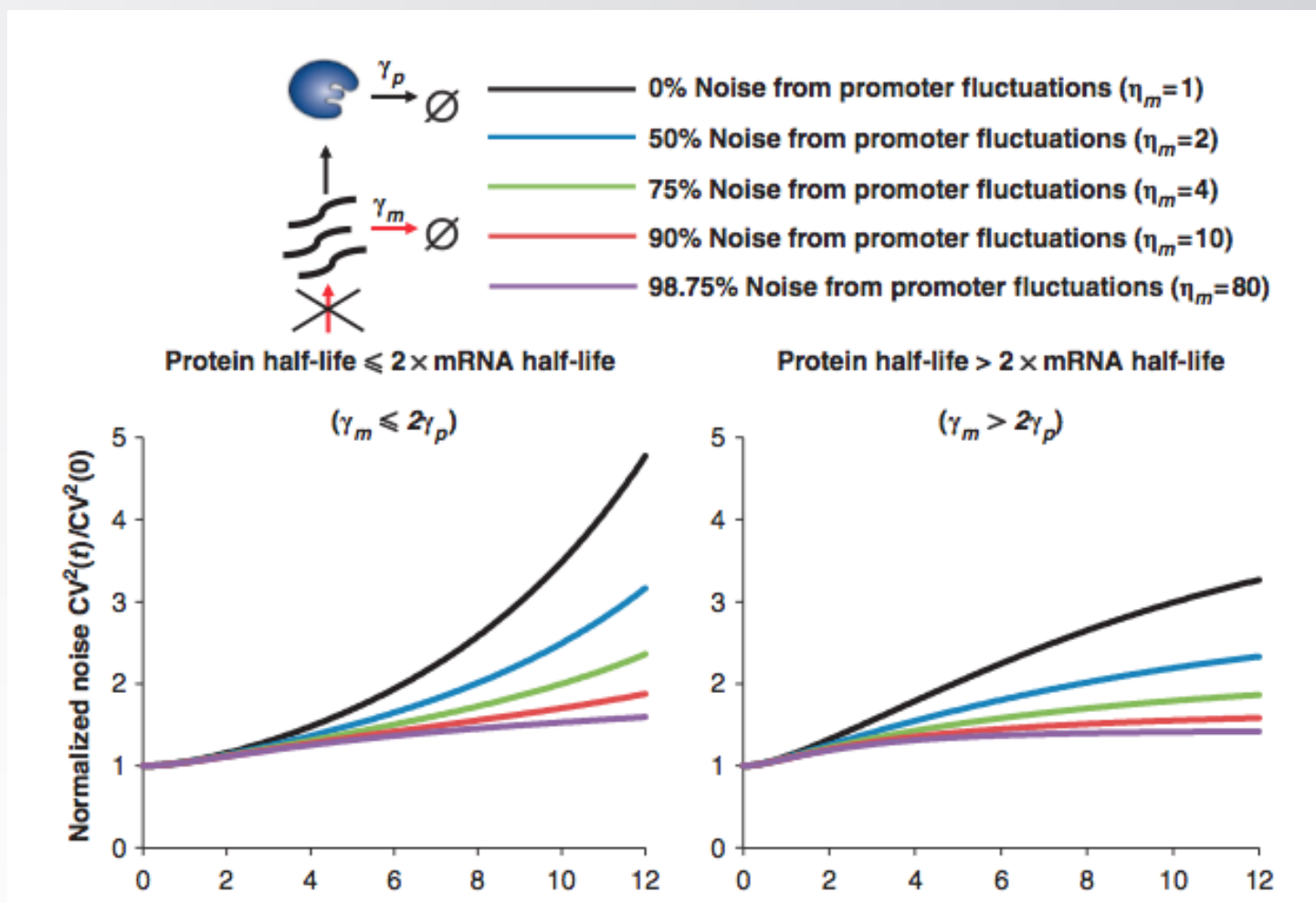


图2 转录阻碍后的噪音变化

2. 基因表达模型

- 公式(7)显示，转录阻碍后蛋白噪音水平的上升取决于 η_m ， Y_m 以及 Y_p 。
- 假设蛋白和mRNA的降解速率为已知或已由蛋白衰减动力学（公式(6)）得到确定数值， η_m 的大小能够直接由 $CV^2(t)$ 决定。
- 如果绝对蛋白分子数目的数据是有效的，那么转录发生的频率 k_m 和mRNA翻译速率 k_p 可以从稳态蛋白平均值和噪音水平（公式(3)和(4)）得到。
- 因此，结合 $p(t)$ 随机时刻（公式(6)和(7)）的瞬态数据和其稳态的数值（公式(3)和(4)），便可推断出基因表达模型的所有参数。

2. 基因表达模型

- 这个模型的一个限制是它不适用于mRNA半衰期过短于蛋白质半衰期的情况。当mRNA半衰期很短时，一旦转录发生抑制mRNA转录子就会很快衰减到0。没有可利用的mRNA，蛋白质会降解，并不再继续提供mRNA动力学的任何数据信息
- 报告系统需要用在mRNA和蛋白降解反应是一级反应的情况，存在已知的可比速率。此外，小分子药物在用于扰乱转录/翻译时是不应改变衰减速率的
- 当 η_m 的数值较大时，转录抑制后的蛋白表达变异性的变化对 η_m 不敏感，因此，通过这个方法我们可能得不到对 η_m 数值大小的精确估计

3.模型验证

■ HIV-1 LTR基因表达噪音

HIV-1 LTR启动子的毒性蛋白的随机性表达在复制激活与整合后潜伏期（病毒休眠期，类似噬菌体的溶原现象）之间密切影响着病毒命运决策。这是一个很重要的识别基因表达内部噪音来源的系统

报告系统

流式细胞仪测量失稳GFP（d2GFP）载体，显示半衰期为2.5-h，量化测定d2GFP mRNA半衰期为3-h，所以对d2GFP来说 $\gamma_m \approx \gamma_p$

3.模型验证

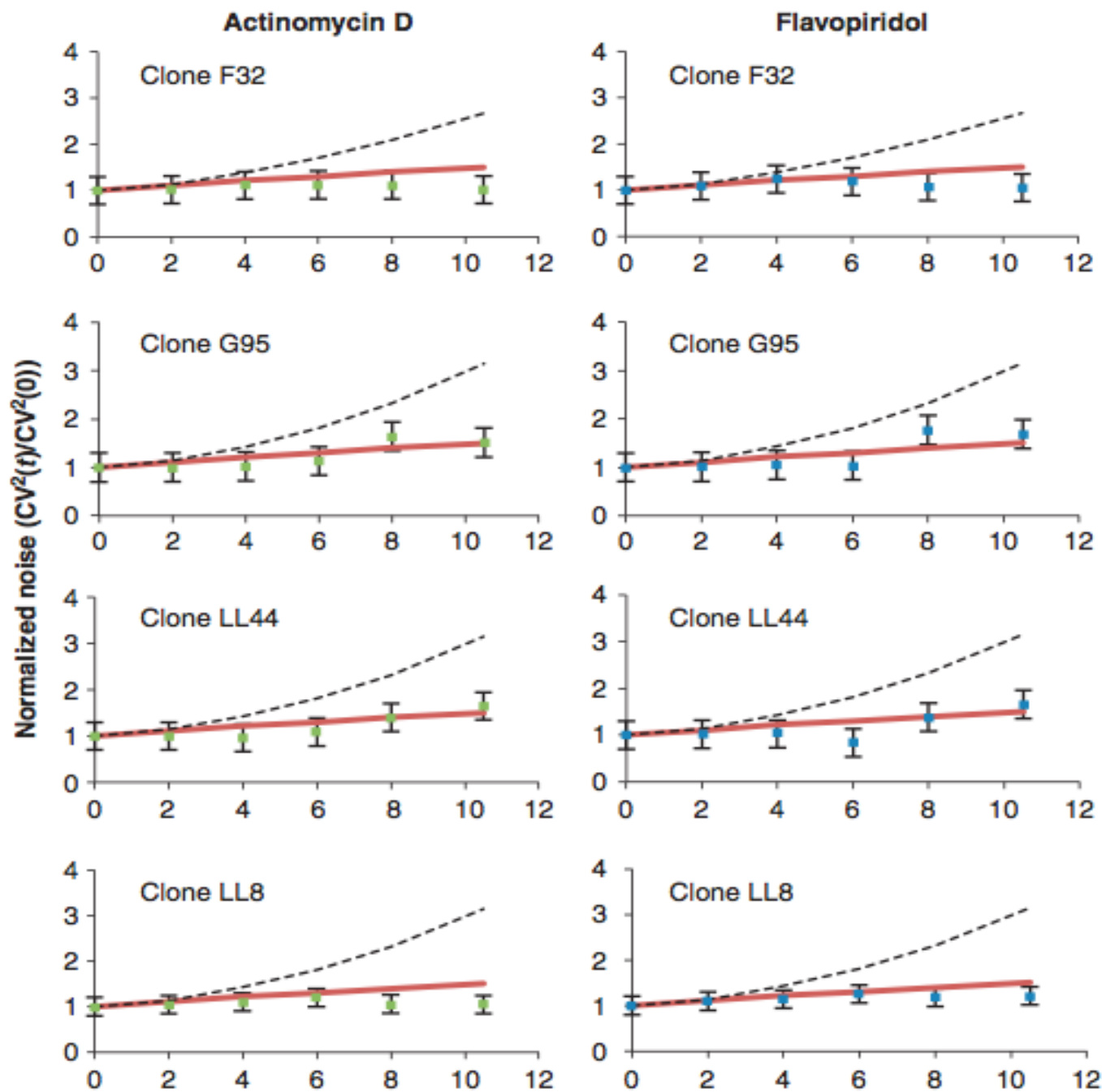
- 为了量化HIV-1 LTR表达噪音，我们利用前人的国际标准化组织克隆种群文库，每个种群携带一个编码LTR-d2GFP的最小载体的单一拷贝，在人类基因上的一个独特位点处整合。通过流式细胞仪的数据能够找到内部噪音以及排除对噪音测量的外界影响。
- 系统中每个细胞包含一个的LTR与d2GFP的整合拷贝和一个LTR与mCherry的整合拷贝，即以双色荧光报告系统应用于过滤外来噪音。

3.模型验证

- 在过滤了外来噪音之后，国际标准化组织克隆种群会显示出一个稳态噪音水平与平均蛋白丰度的比值，其倒数与mRNA增减波动和启动子波动是一致的，这恰恰说明仅用稳态蛋白噪音水平区分内部噪音的不同来源是不充分的
- 为区分mRNA增减波动和启动子波动，我们分析了发生扰乱后HIV-1 LTR表达噪音的动力学变化

3.模型验证

- 放线菌素D，反正快并且通过阻滞RNA聚合酶II（RNAP II）的延伸能够有效抑制转录活动
- 另一种是Flavopiridol，它通过抑制P-TEFb和RNAP II的相互作用来抑制转录活动
- 给任何isoclonal种群引入任意转录抑制剂所导致的d2GFP荧光强度水平的降低，都与前文中2.5-h d2GFP蛋白的半衰期和3-h d2GFP mRNA的半衰期是一致，说明报告衰减动力学没有受到小分子药物的影响。
- 为应答上文中两种转录抑制剂，做出荧光强度水平的细胞间变异随着时间呈增长趋势



3.模型验证

- 转录被抑制后，基因表达噪音的瞬态增长的观测结果与由于毒性启动子造成的mRNA转录爆发后所引起的HIV-1 LTR启动子波动为主要噪音来源的模型所观测到的结果是一致的（红色曲线， $\eta_m=15$ ）。
- 与之相比，对一个以mRNA增减波动为基因表达噪音主要来源的模型（黑色虚线， $\eta_m=1$ ）来讲，所有克隆的瞬态增长 $CV^2(t)$ 观测结果都过高。

3.模型验证

- 数据表明，在4个不同的整个位点上，启动子波动在HIV-1 LTR启动子基因表达噪音中占主要来源，HIV-1 LTR启动子的mRNA法诺因子（ η_m ）至少为10
- 通过mRNA smFISH法直接测量mRNA种群状态也显示HIV-1 LTR的转录爆发水平很高（ $\eta_m \approx 80$ ）进一步验证了实验结果

4.总结

- 一个用来区分mRNA增减波动和启动子波动的简单的方法，即通过蛋白表达噪音动力学系统来观测细胞对扰乱活动的应答。这种扰乱方式与其他方法互补，例如mRNA smFISH法。
- 优点是只要求蛋白水平的数据，并且能在mRNA高表达水平时应用
- 缺点是系统应答扰乱时表达噪音的变化对外来噪音比较敏感，因此外来噪音必须通过适当控制细胞或利用双色报告系统被排除

4,总结

- 在转录抑制后测定表达噪音的瞬态变化，结果显示由HIV-1 LTR引起的mRNA转录爆发是病毒基因表达的一个重要噪音来源，这些转录爆发现象成为HIV-1毒性蛋白表达水平引起细胞高变异性的一个重要管理机制。
- 蛋白水平的波动给我们带来很重要的信息，而监测这些波动的统计特性可以成为探索遗传回路管理机制的一个重要工具

Thank you