



# Genome-wide modeling of transcription kinetics reveals patterns of RNA production delays

全基因组的转录动力学建模揭示RNA产物延迟的模型

吴浩



一、背景简介

二、实验流程

三、结论

四、讨论



# 一、背景简介

基因转录是严格调控的动力学过程，而转录延迟对基因表达动力学有重要影响。研究转录延迟对进一步研究如何调控细胞快速应答是必需的。

研究进展：

- ①肿瘤坏死因子TNF- $\alpha$ 诱导基因的相关研究，发现与剪接相关的RNA产物延迟。
- ②转录延迟可降低负反馈环或促进及时转录间的振荡，作出最佳诱导。
- ③结合ChIP-Seq和GRO-Seq的数据研究转录延伸动力学。

本文建立转录的时间动力学模型，使用全基因组的高通量时间过程数据，测量转录活性和mRNA浓度。运用常微分方程与非参数统计建模，解释转录延迟这个生物学过程。

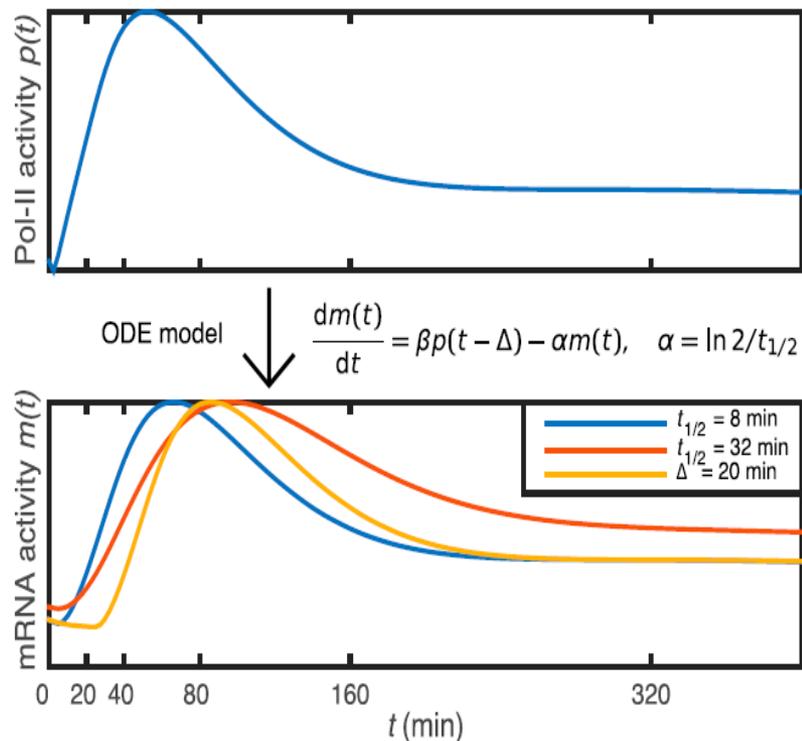
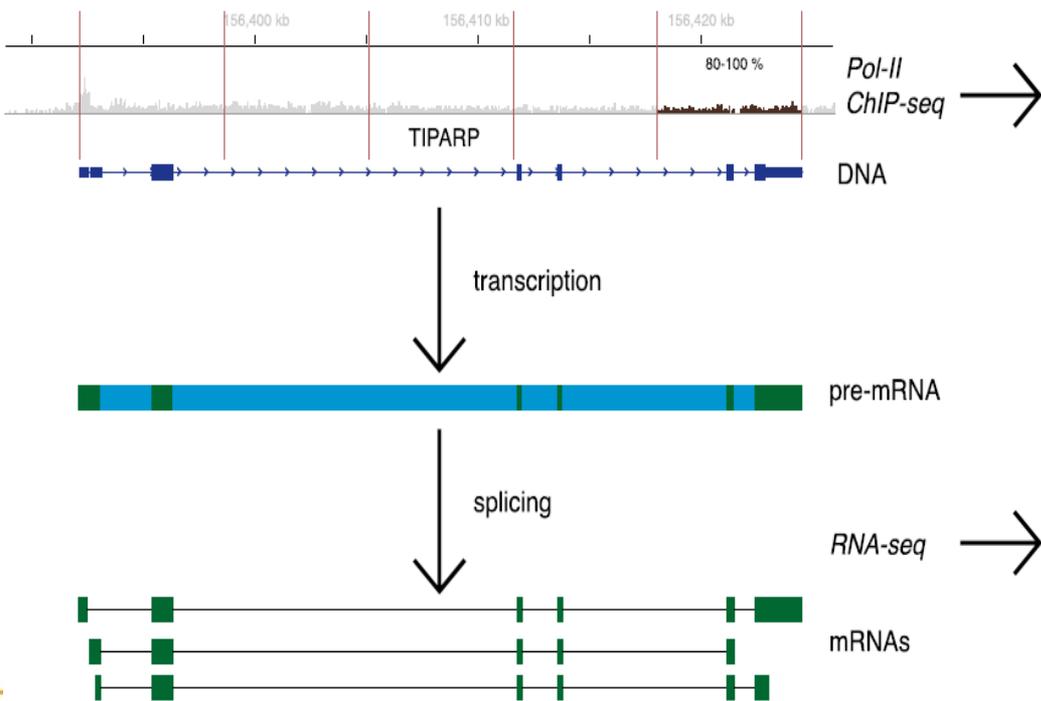
## 二、实验流程

转录延迟：从转录终止到mRNA产生的过程。  
 (RNA聚合酶从pre-mRNA上脱离到加工成成熟的mRNA)

MCF-7乳腺癌细胞系  
 (无E2培养3d)

雌二醇 (E2) 刺激

雌激素受体  
 (ER)





## ODE模型

$$\frac{dm(t)}{dt} = \beta p(t - \Delta) - \alpha m(t)$$

$m(t)$ : 成熟mRNA的浓度  
 $p(t)$ : 基因3末端转录速率  
 $\Delta$ : 转录延迟时间 (min)  
 $\alpha$ : mRNA的降解速率  
 $\beta$ : 与 $p(t)$ 大小有关  
 $t$ : 收集数据的10个时间点

RNA聚合酶 II 的ChIP-Seq时间过程数据（酶位于基因3末端最后20%的区域）定量转录激活  $p(t)$ 。

mRNA的RNA-Seq时间过程数据（map到已注释的转录本）定量成熟mRNA  $m(t)$ 。

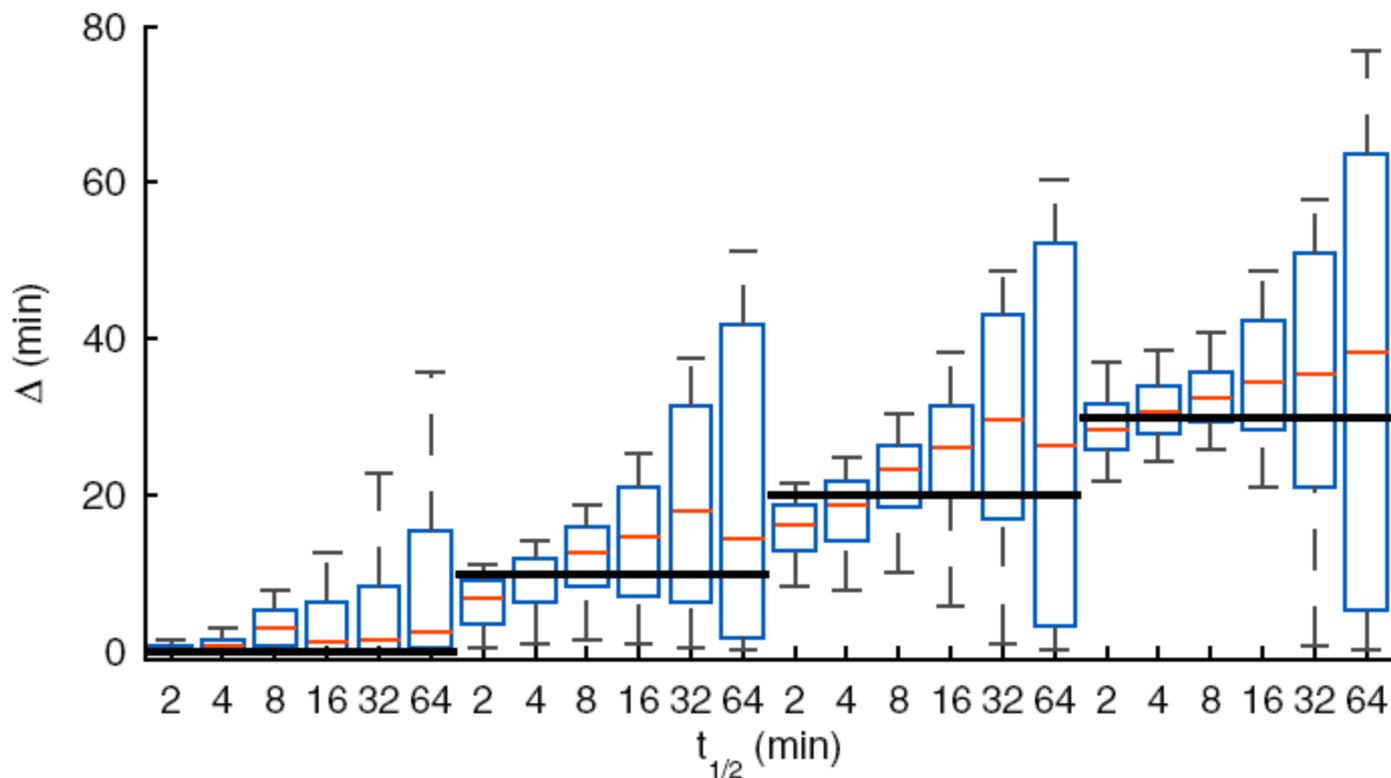
运用Gaussian process（高斯过程）、Bayesian（贝叶斯定理）、MCMC（马尔柯夫链蒙特卡洛理论）数据拟合，后验分布计算出参数。



### 三、结论

## 1. 模拟数据（验证可靠性）

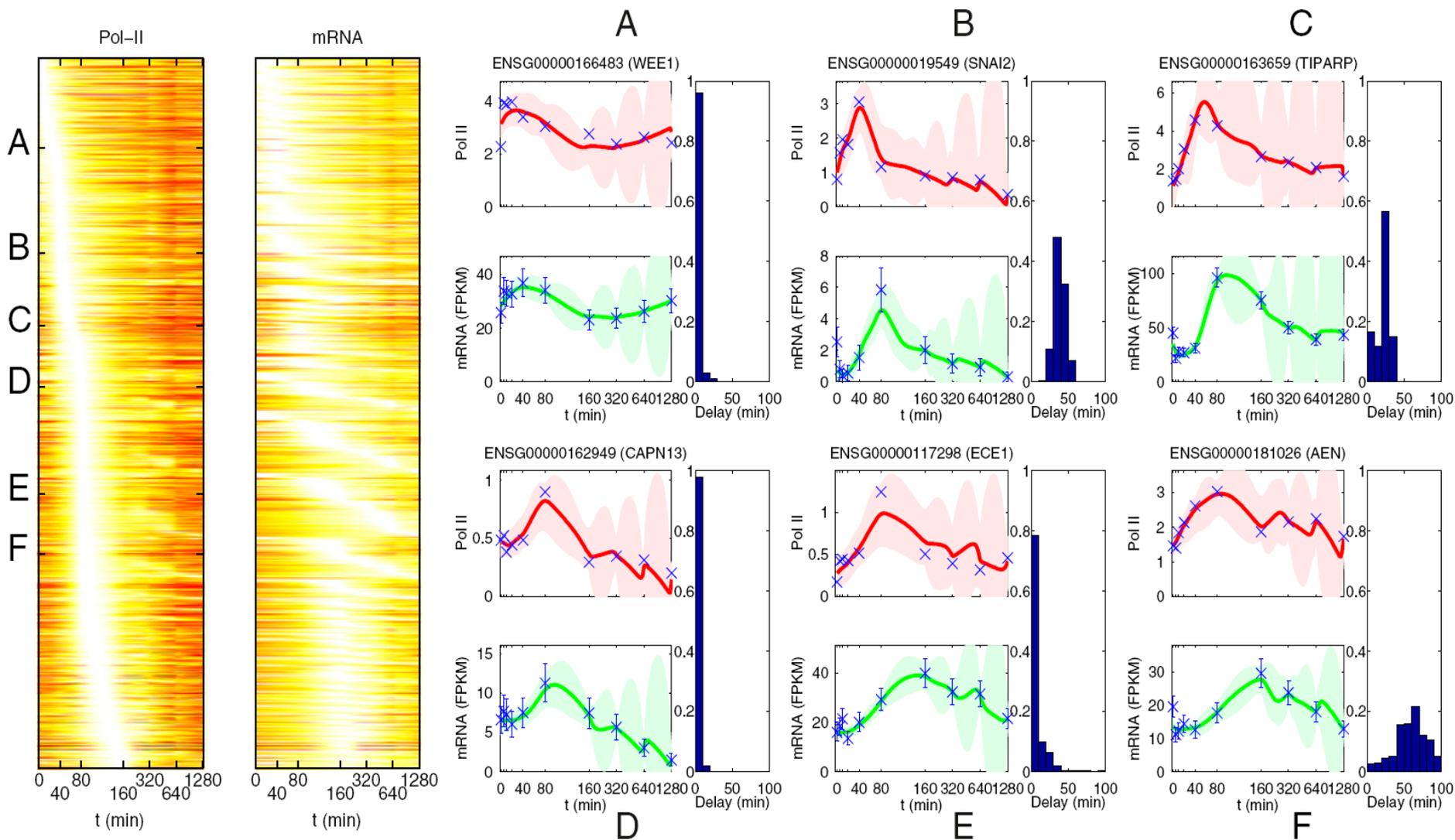
选用TIPARP基因的pol- II 数据检验 $p(t)$



参数后验分布箱线图：真实值在**25%**后验分布以上，可准确推断 $\Delta$ 在可信区间。

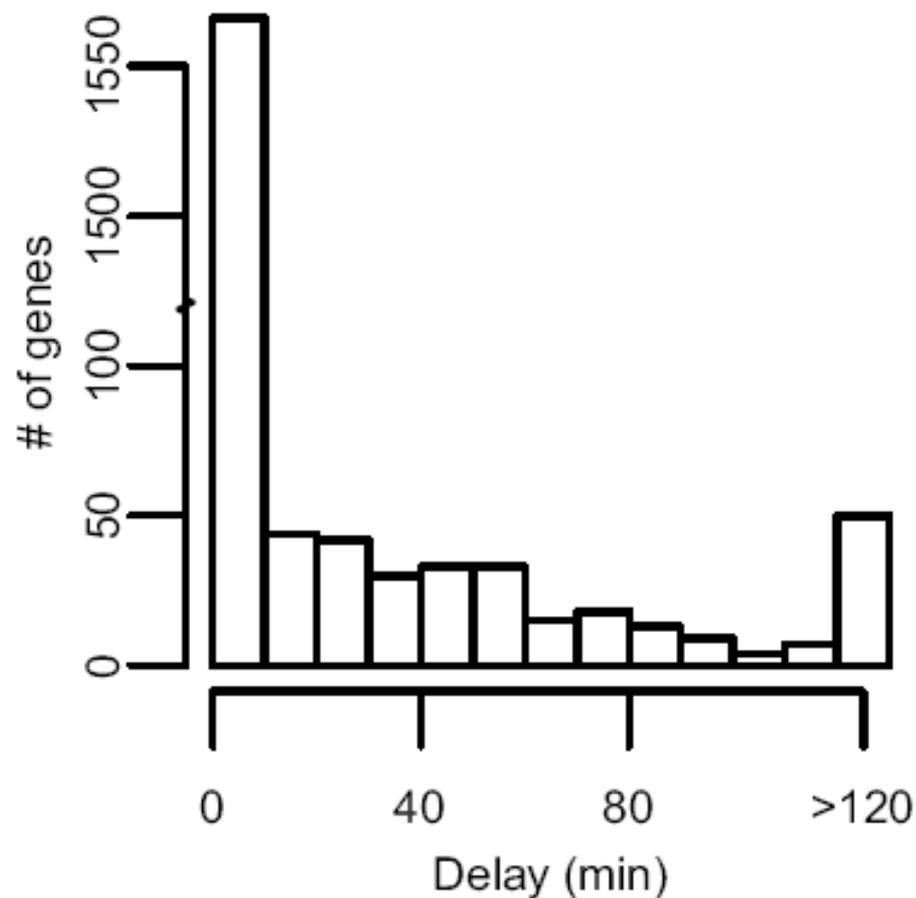


## 2. ER信号



通过数据拟合模型，筛选到**1864**个基因显示有相当大转录延迟，其中**50**个基因由于延迟的后验中值  $> 120\text{min}$  被认为不可靠舍去。

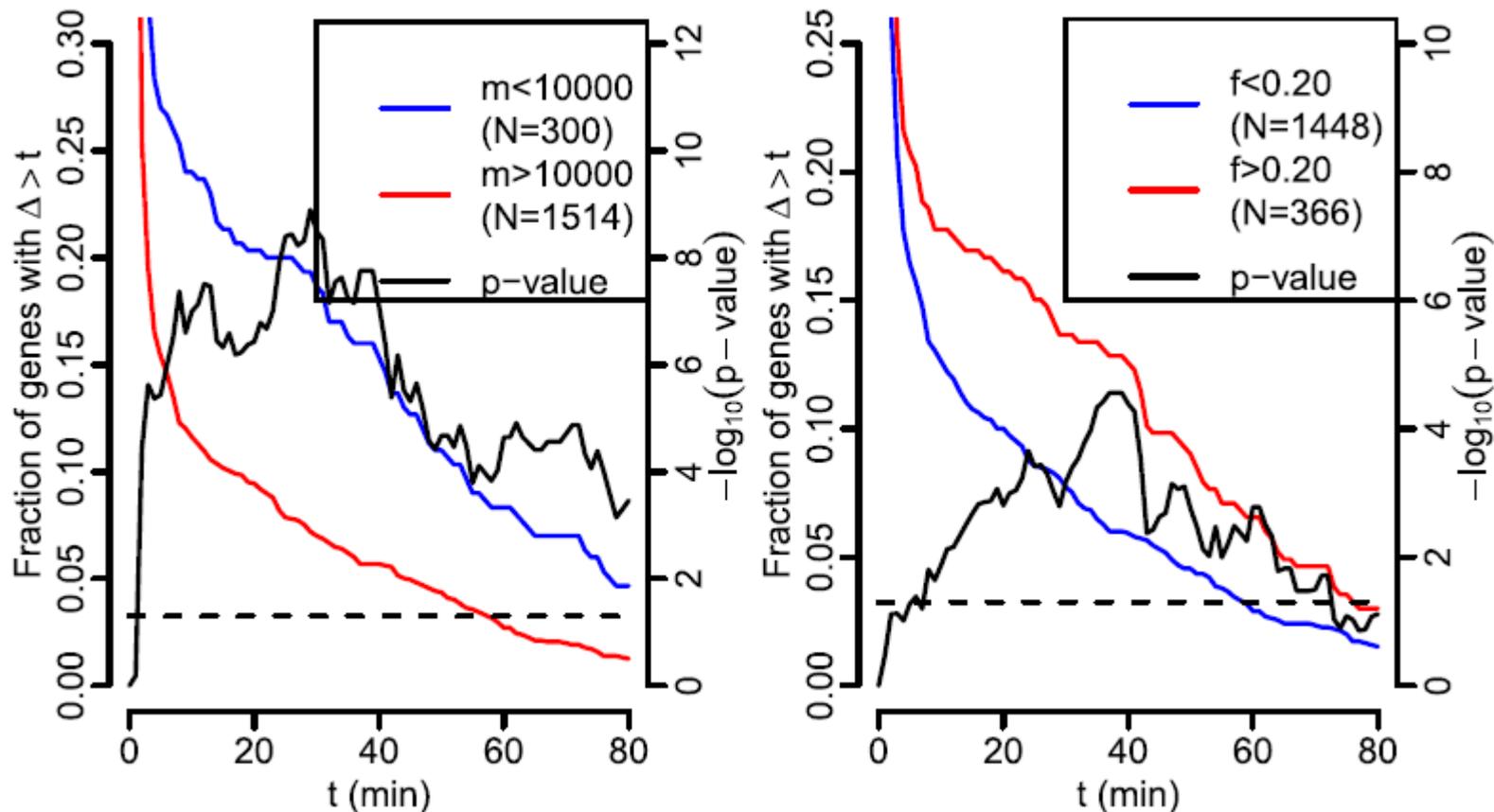
最终**1814**个基因中，**204**个基因（**11%**）的延迟后验中值  $> 20\text{min}$ ，属于长延迟基因。而早期诱导基因的延迟估计值更加精确。





### 3.长延迟基因的基因特征

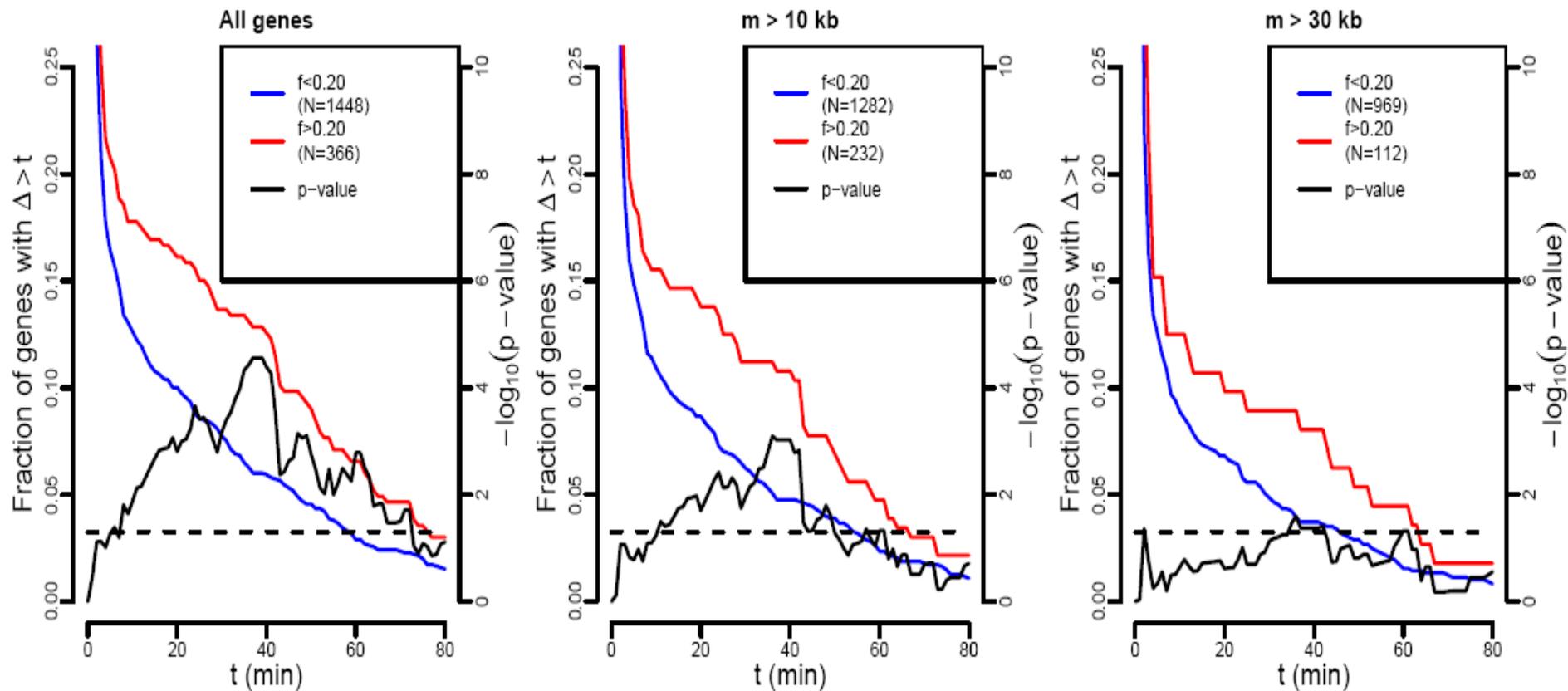
之前研究表明延迟与剪接过程有关。

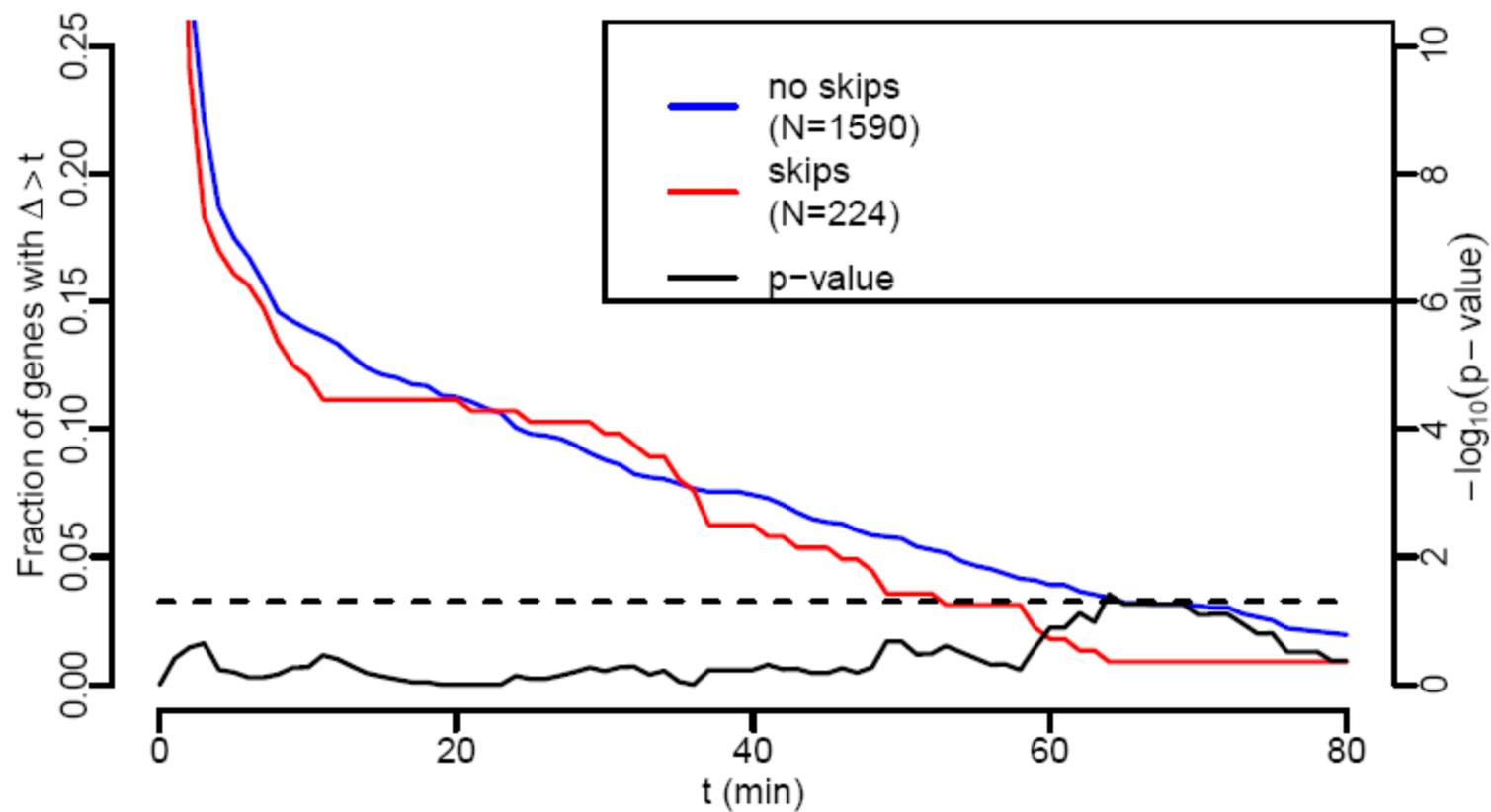


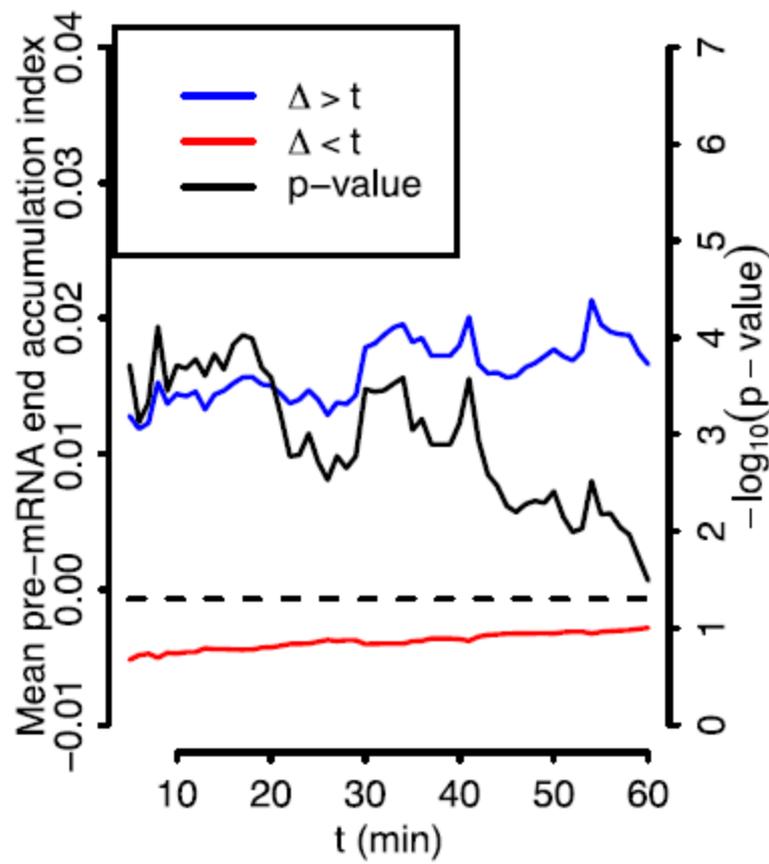
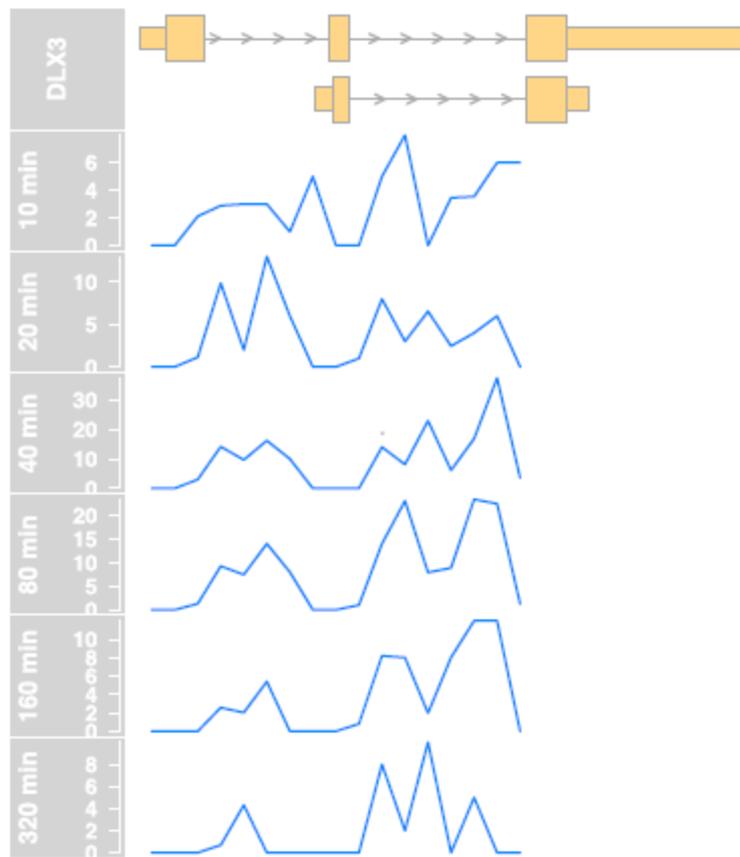
左：具有短的pre-mRNA的基因有长延迟。

右：最后一个内含子的长度占转录本比例大的基因具有长延迟。（弱）

区分：pre-mRNA长度、最后一个内含子长度分别对延迟的影响。











- (1) 用一句话概括此研究的主要结论或创新点。
- (2) 此研究对你有何启发。
- (3) 此研究还存在哪些问题可以改进。







