Orthogonal control of expression mean and variance by epigenetic features at different genomic loci

Pablo

College of Life Science and Technology

Huazhong Agricultural University

Introduction

Gene **expression mean**: average number of the products of genes
Gene **expression noise**: random variability in the number of the products of genes

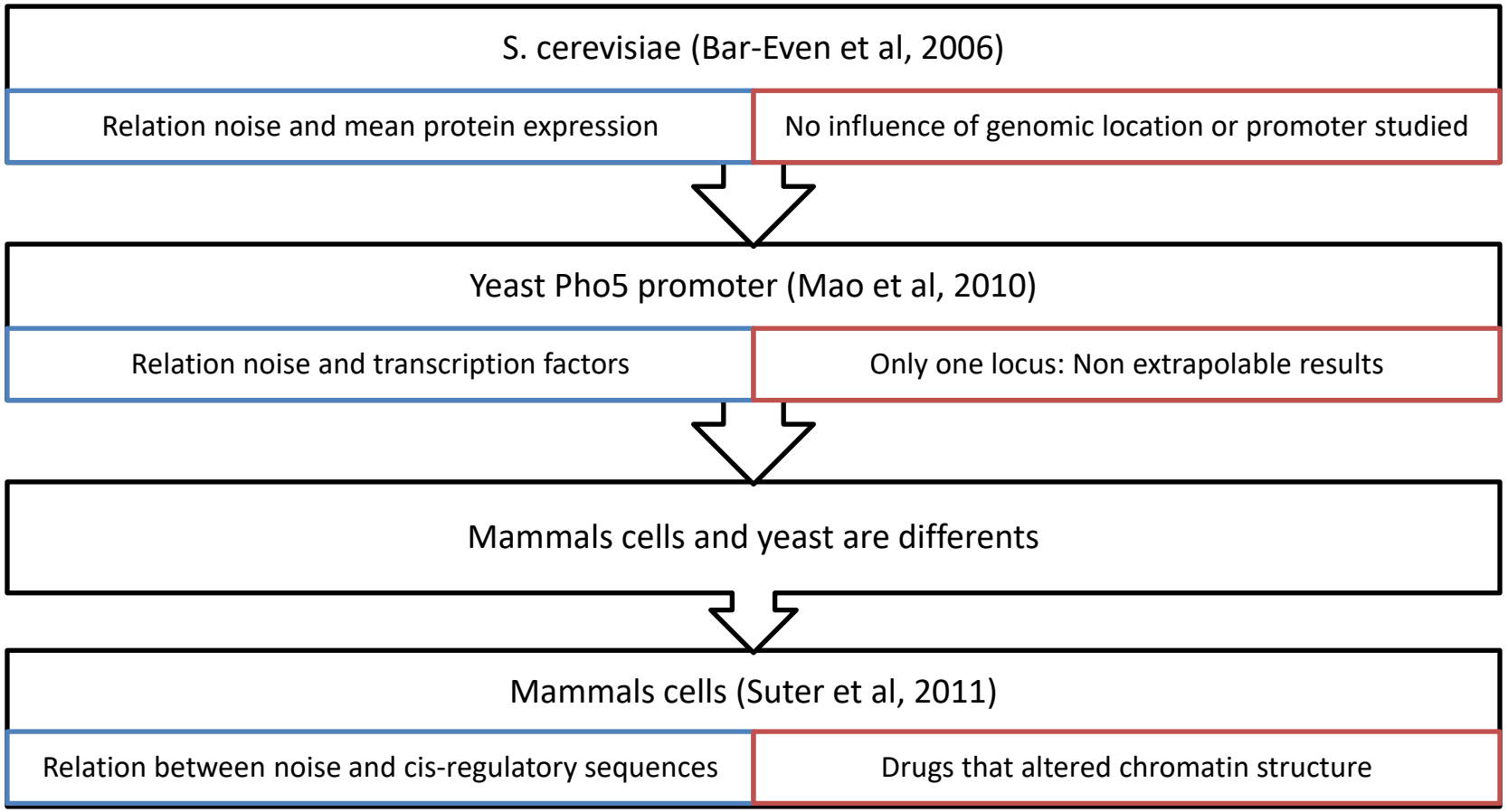
Isogenic cell populations grown under identical conditions exhibit non-genetic heterogeneity

Gene expression **noise** has been shown to be an important source of non-genetic heterogeneity in mammalian cells

Validated by **single-cell analysis techniques** such as flow cytometry, high-throughput microscopy, and recently single-molecule RNA fluorescent in situ hybridization (smFISH)

The mechanistic roles and impact of key **molecular factors** on gene expression noise remain largely undissected

What do we know about transcriptional noise?



How the genomic environment influences gene expression noise regulation in a mammalian system?

Experiment design: No relation between promoter sequence and genomic location

1. Lentiviral-based system:
   - semi-random integration in different genomic locations while maintaining the same promoter architecture
   - lentiviral promoters exhibit many features of eukaryotic promoters (TATA box, cis acting elements and nucleosomes along the promoter)

2. Generate single-cell clones spanning hundreds of integration positions: study the influence of genomic location on mRNA and protein expression noise

3. Measure the chromatin state of promoters integrated into different genomic locations: discover which molecules regulate expression noise

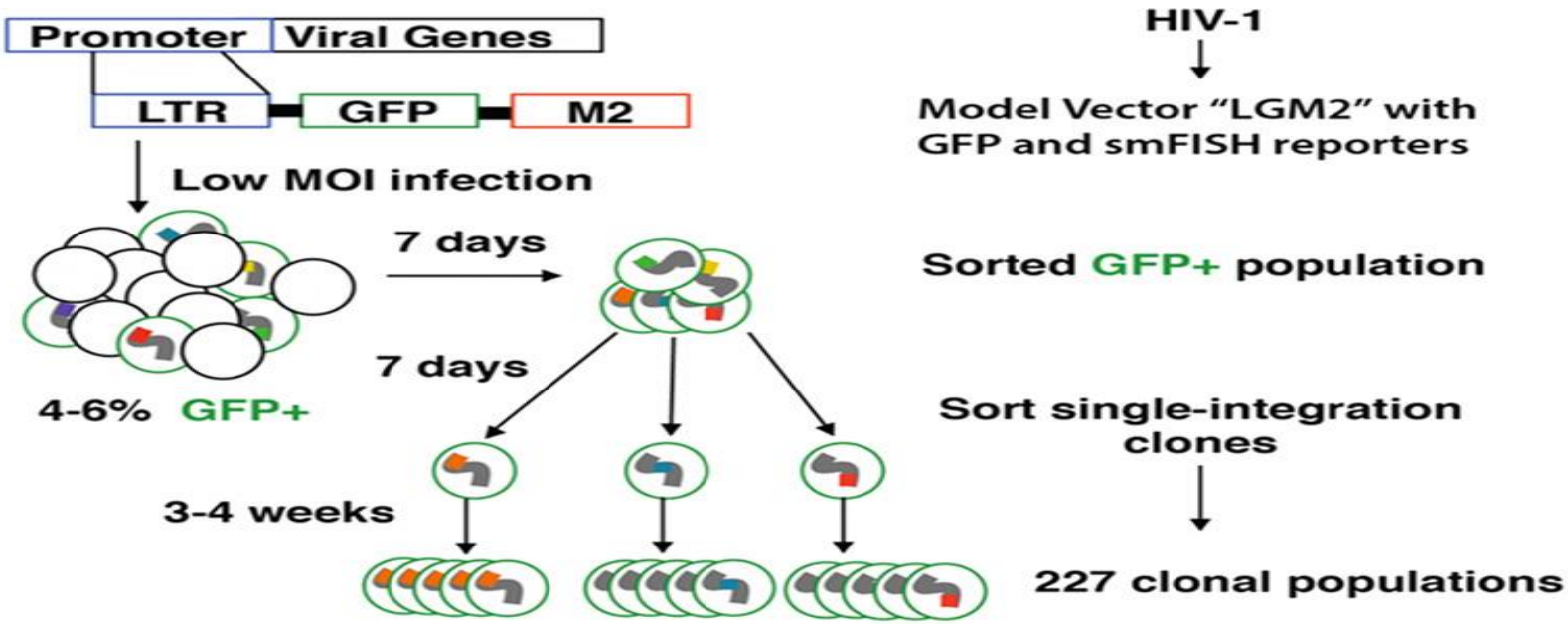
4. Quantifie mRNA levels by smFISH and protein levels by flow cytometry: know if burst size or burst frequency correlates with gene expression mean

   Burst size: number of transcripts generated during a short interval during which a “bursty” promoter is mediating transcription

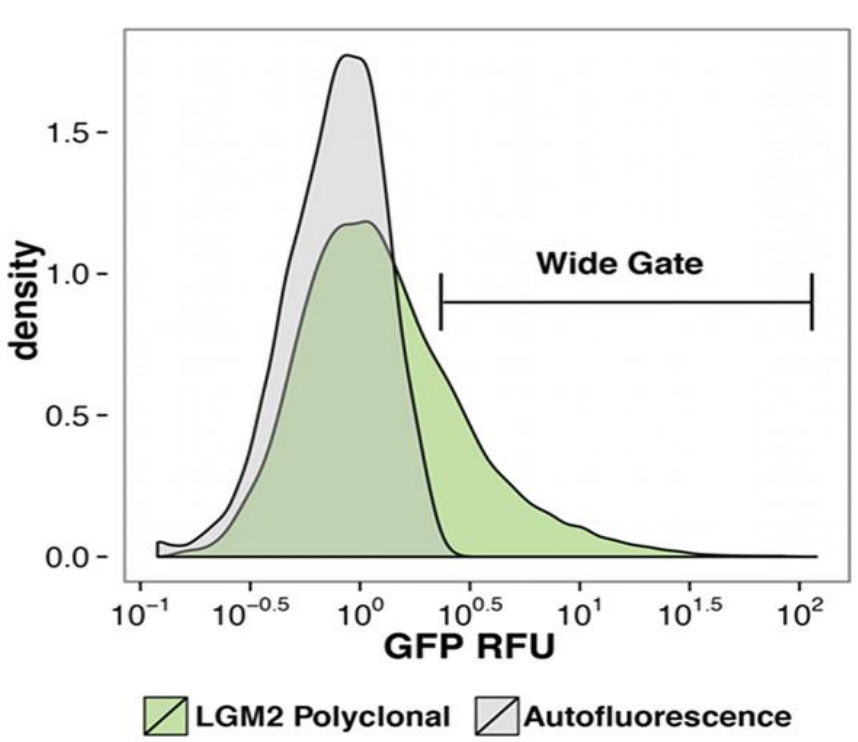
   Burst frequency: the rate of promoter transitions into the productive bursty state

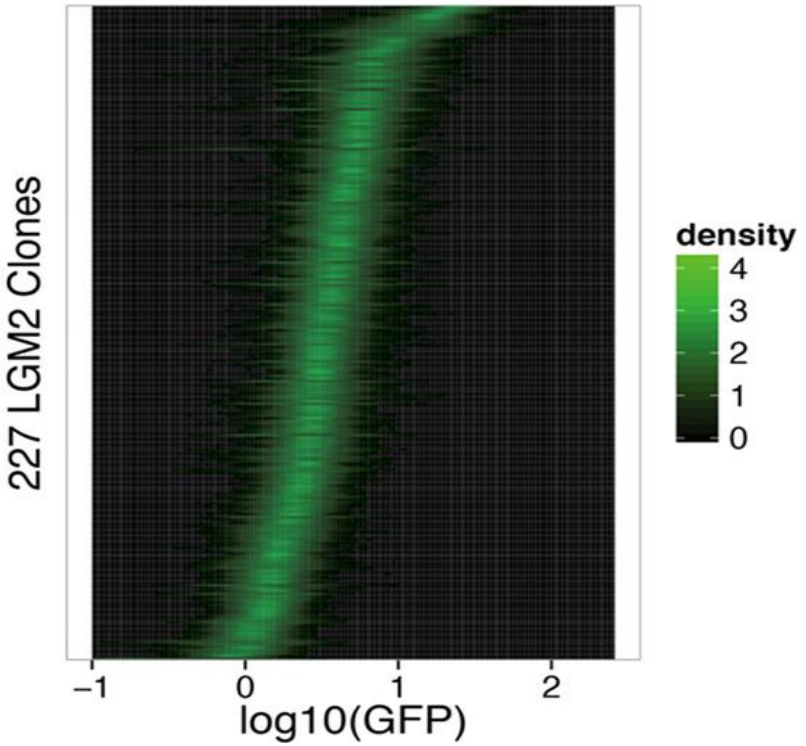
Results

Model vector (LGM2) consisting of the HIV-1 LTR driving dual protein (GFP) and RNA (M2 smFISH) reporters



1. Cells were infected with HIV-1 LGM2 and allowed to reach steady state expression for 7 days
2. GFP-positive cells were sorted and allowed to expand for 7 days
3. From this GFP+ population, single cells were allowed to expand for 3–4 weeks to generate 227 clonal populations



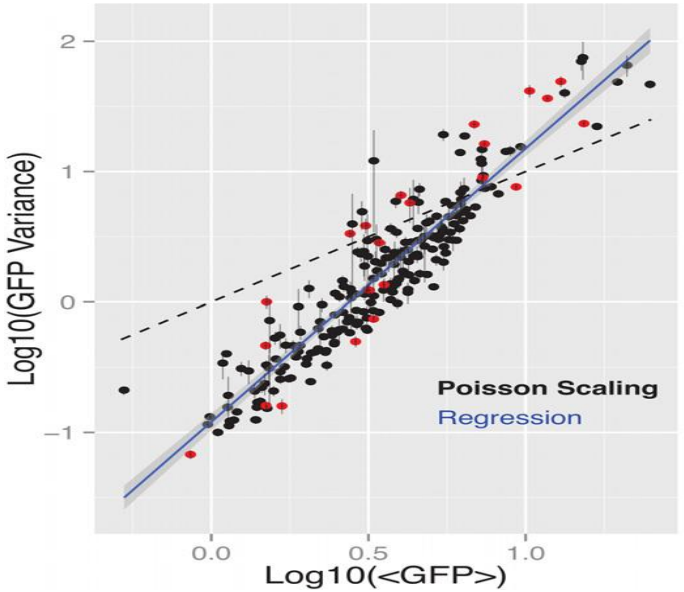


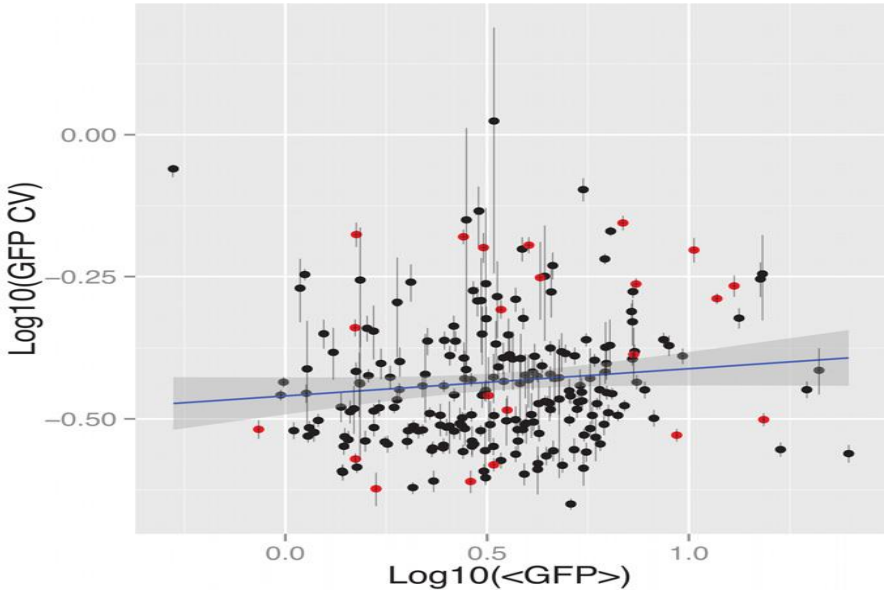
- 227 clones were chosen from the wide gate. Another set of 191 clones were randomly chosen to validate future results (no differences)

- Similar mean but different width (noise: random variability in the number of the products of genes). Evidence of noise independence from mean expression across the set of clones

2 options in relation between mean and variability:

- Promoter always in a productive state: distribution variance would scale linearly with mean, and the coefficient of variation (CV) would decrease (Poisson process)
- Promoter transitions from a Off and On state: RNA is produced in infrequent bursts, and variance would relate to mean through a power-law relathionship (non-Poissonian process)

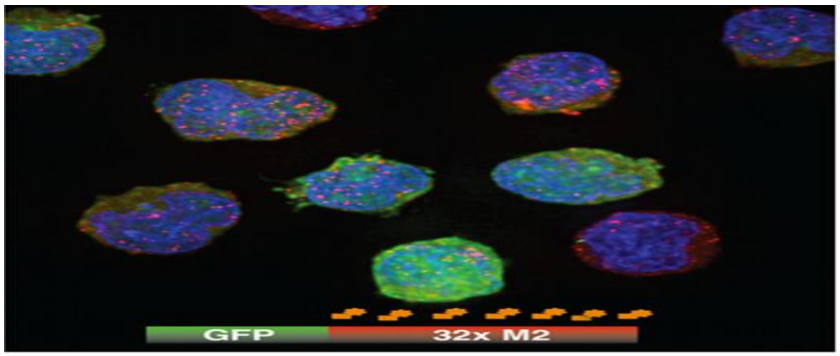


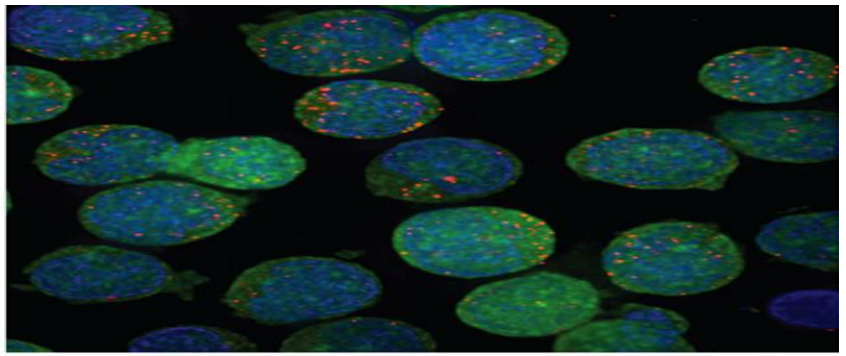


- GFP fluorescence from each clone was measured via flow cytometry
- High relationship between distribution mean and variance is distinct from Poisson scaling and is consistent with **promoter transitions between "Off" and "On" states**
- GFP CV (dimensionless expression noise) and GFP mean suggests <u>independent control of mean expression and expression noise across integration positions</u>

Results consistent with a non-Poissonian bursting transcriptional mechanism

- However, the GFP observations are post-transcriptional processes that may obscure important underlying information

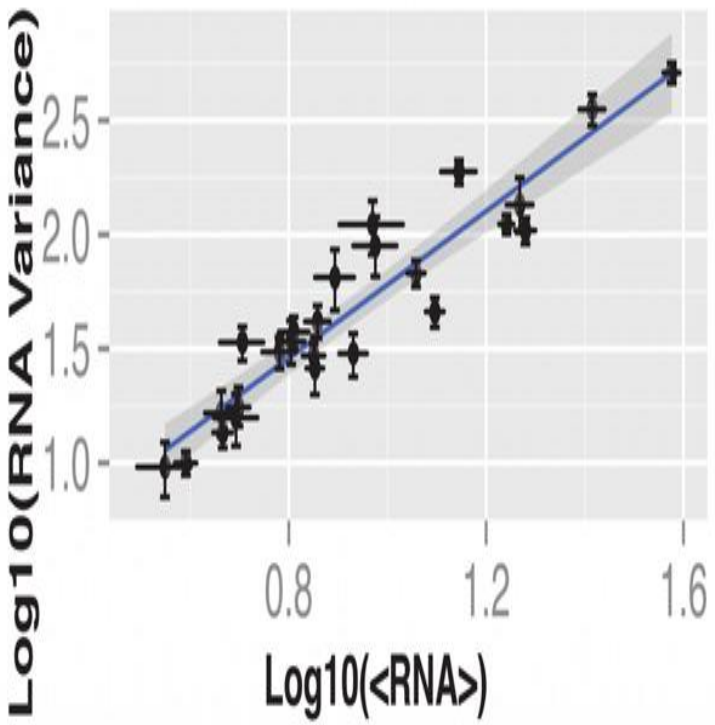


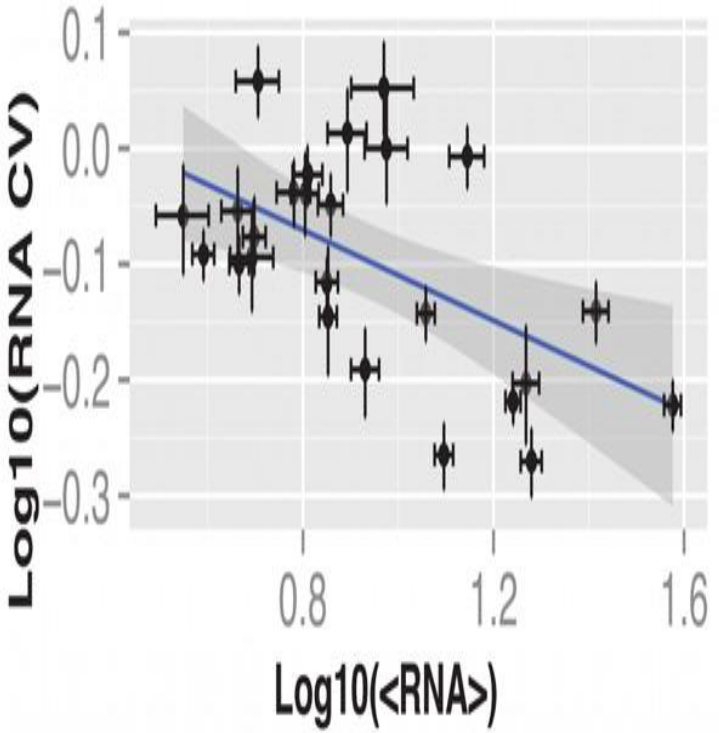


Analysis of 25 clones by smFISH using a probe against the M2 array

Proceeded to quantify exact RNA copy numbers per cell in both high noise and low noise clones

Transcriptional bursting and RNA and GFP expression heterogeneity in high and low noise clones



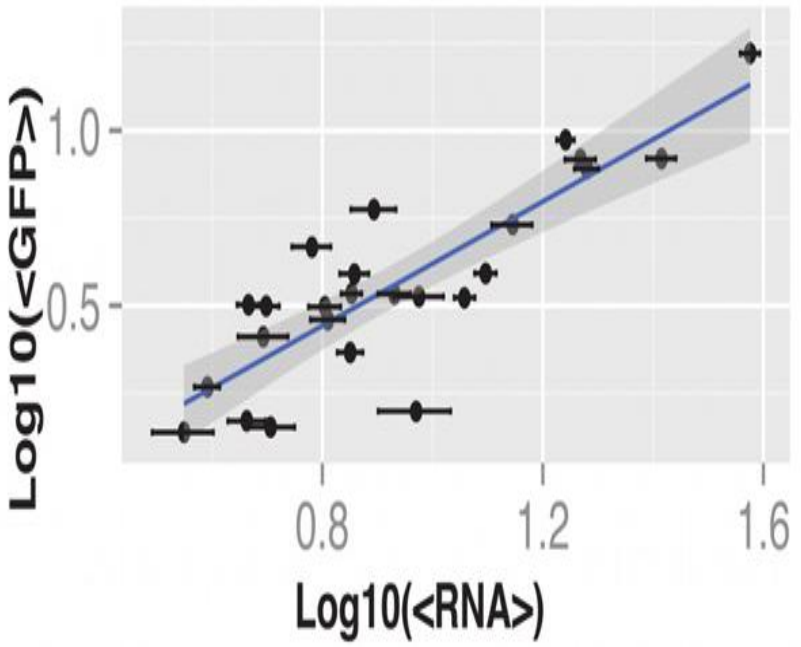


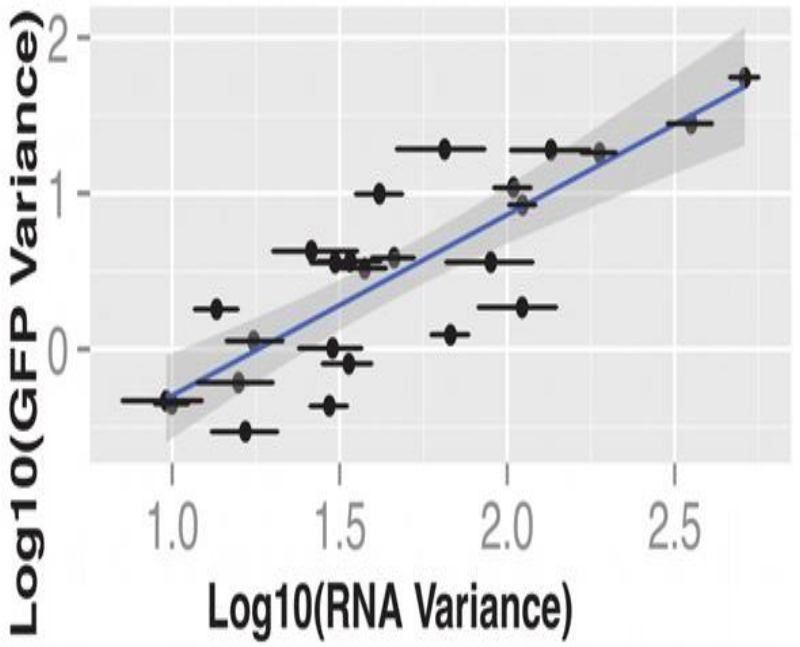
- RNA mean and variance follow a power-law relationship, but the slope of GFP mean vs GFP variance was higher: post-transcriptional steps may augment noise

- RNA CV is not dependent on RNA mean: non Poissonian (same as GFP mean and CV)

Assumed that translation is a constant-rate first-order process that does not vary between clonal populations:

-RNA mean and variance would strongly predict protein mean and variance

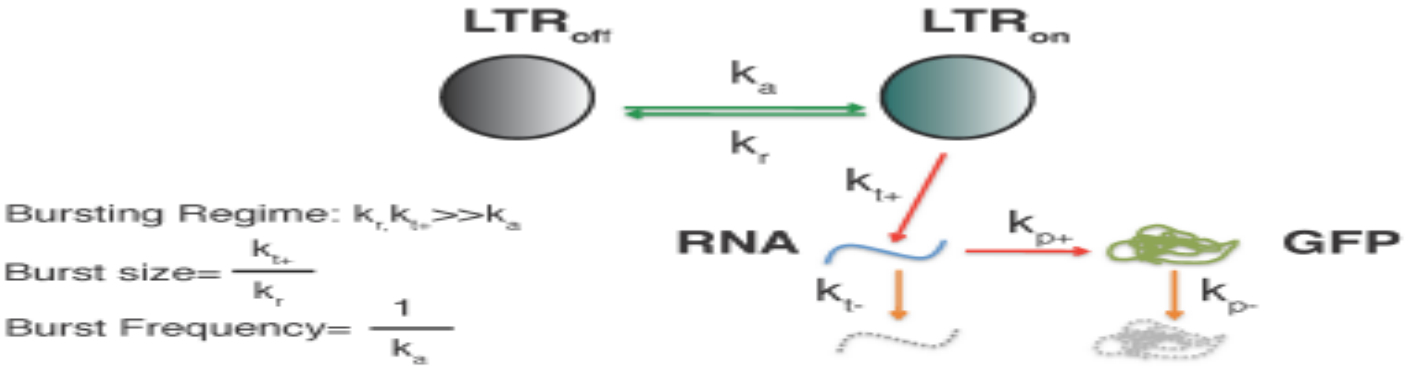




- Variation in RNA mean predicts variations in GFP mean
- RNA variance predominantly explains GFP variance

This suggests that **protein distribution is predominantly determined by RNA distribution**, with little contributions from post-transcriptional processes

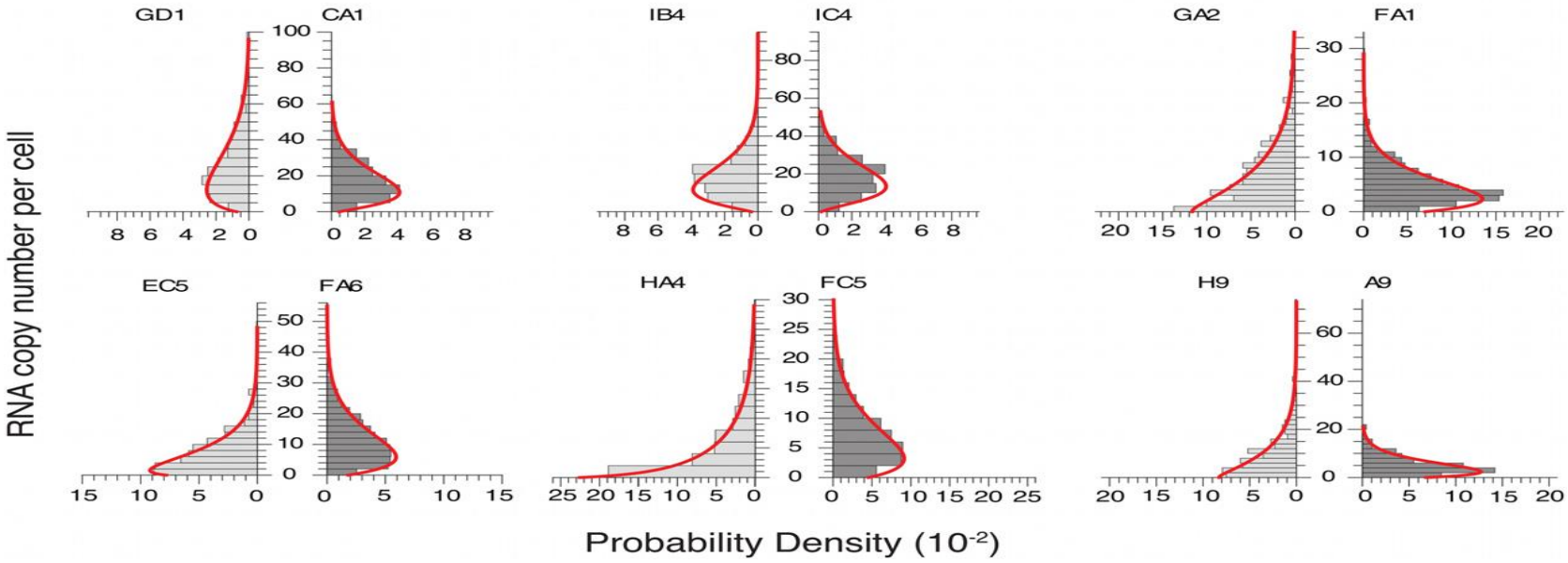
Mean and noise expression are independent: distinct molecular mechanisms could regulate them allowing cells to precisely tune gene expression distributions



$$P(n \ \ RNA) = \frac{\left(\frac{k_{t+}}{k_{t-}}\right)^n \Gamma\left(\frac{k_a}{k_{t-}} + \frac{k_r}{k_{t-}}\right) \Gamma\left(\frac{k_a}{k_{t-}} + n\right) {}_1F_1\left(\frac{k_a}{k_{t-}} + n; \frac{k_a}{k_{t-}} + n + \frac{k_r}{k_{t-}}; -\frac{k_{t+}}{k_{t-}}\right)}{\Gamma(n+1)\,\Gamma\left(\frac{k_a}{k_{t-}}\right)\Gamma\left(\frac{k_a}{k_{t-}} + n + \frac{k_r}{k_{t-}}\right)}$$

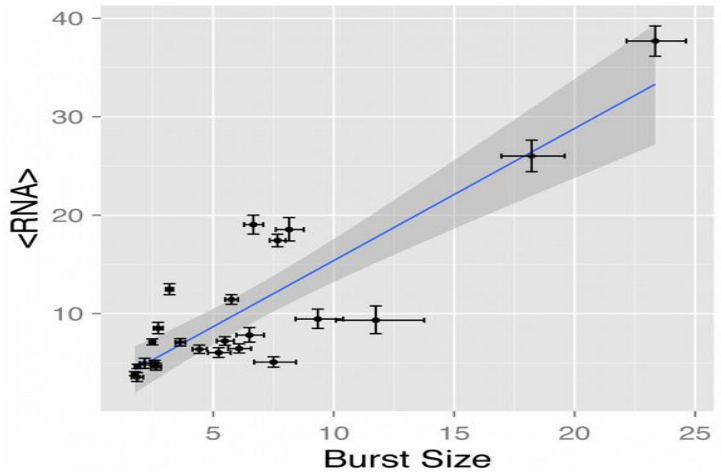
Maximum-likelihood estimation (MLE) of kinetic parameters was performed for the two-state transcription model:

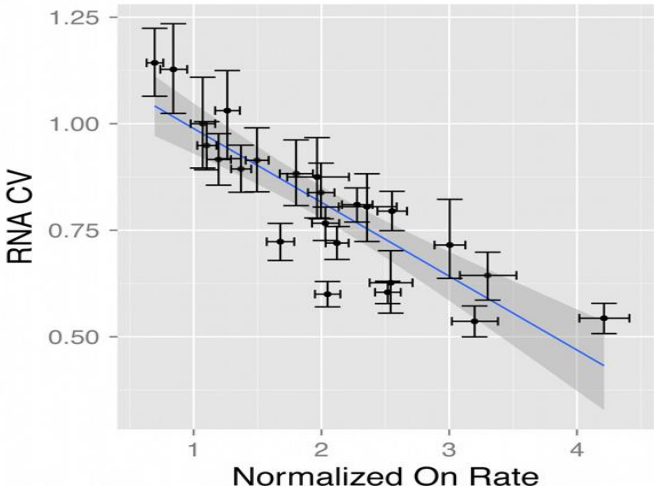
Each clone can be described by the average rate of promoter transitions to the ‘On’ state and the average number of transcripts produced in the ‘On’ state.

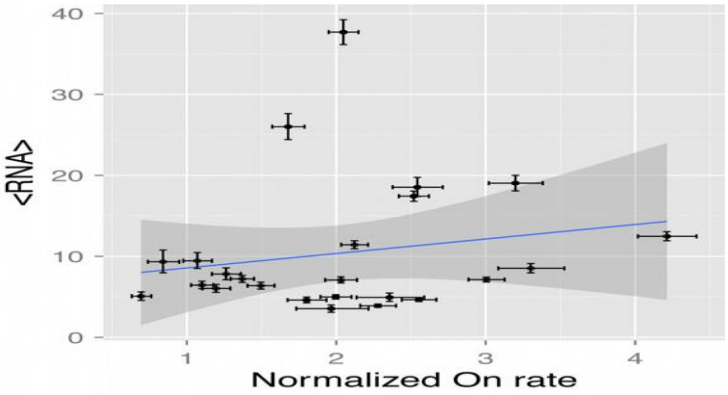


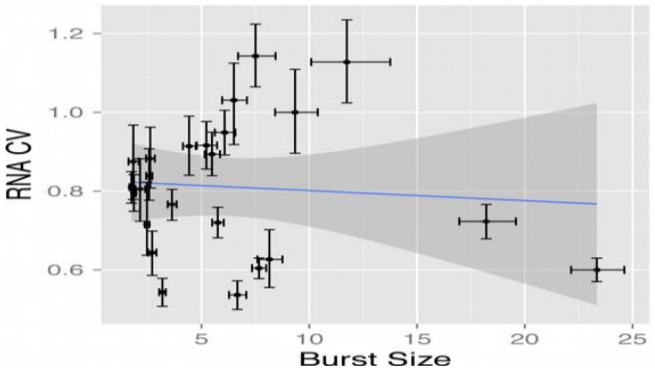
•The two state model can effectively fit both the low and high noise clones mRNA histograms

•**Burst size and frequency vary across genomic locations**



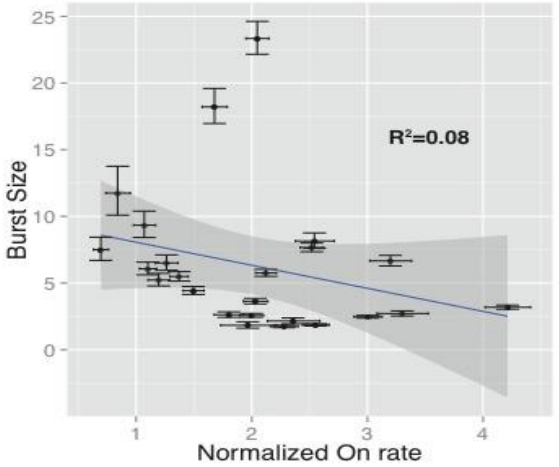






- There is a relation between:
  - Burst size and RNA mean
  - Promoter 'On' rate (burst frequency) and expression noise

- No correlations between:
  - Burst frequency and RNA mean
  - Burst size and RNA noise
  - Burst size and burst frequency



Burst size is related to mean and burst frequency to noise

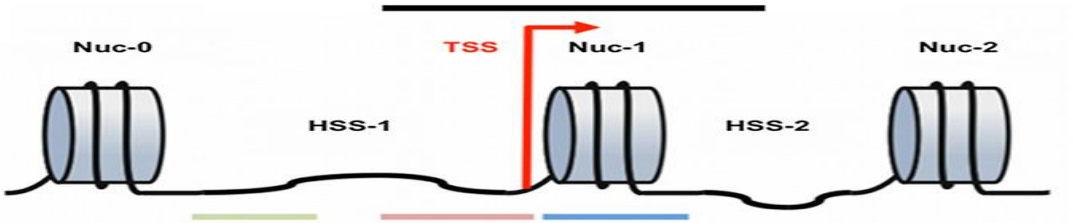
Burst size and frequency vary across genomic locations

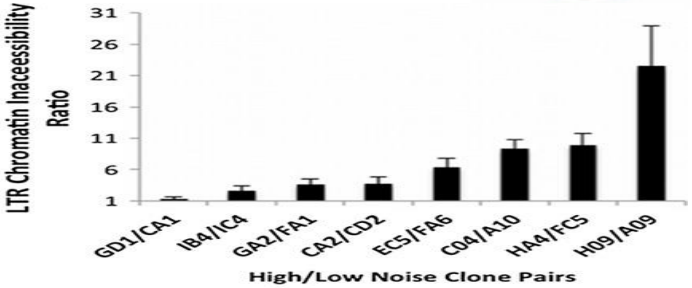
As mean and noise are independent, **burst size and frequency may thus have distinct molecular regulation**

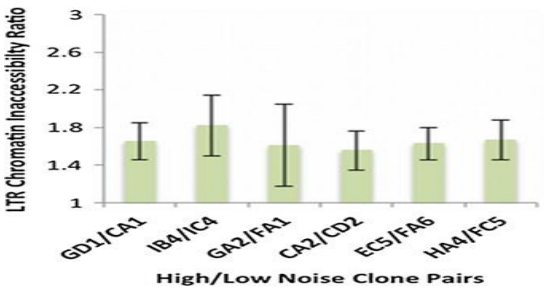
What mechanism may differentially affect burst size and frequency remains to be discovered

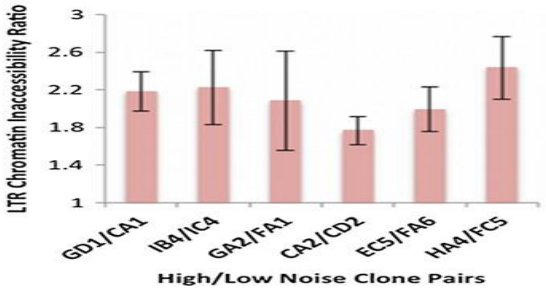
Maybe chromatin or nucleosome dynamics might control the promoter On rate and expression noise.

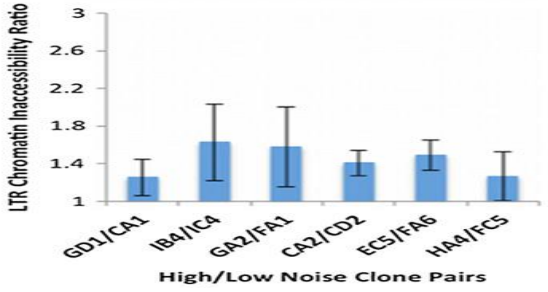
DNAse I sensitivity assays to quantitatively measure the chromatin accessibility of the LTR across different genomic locations



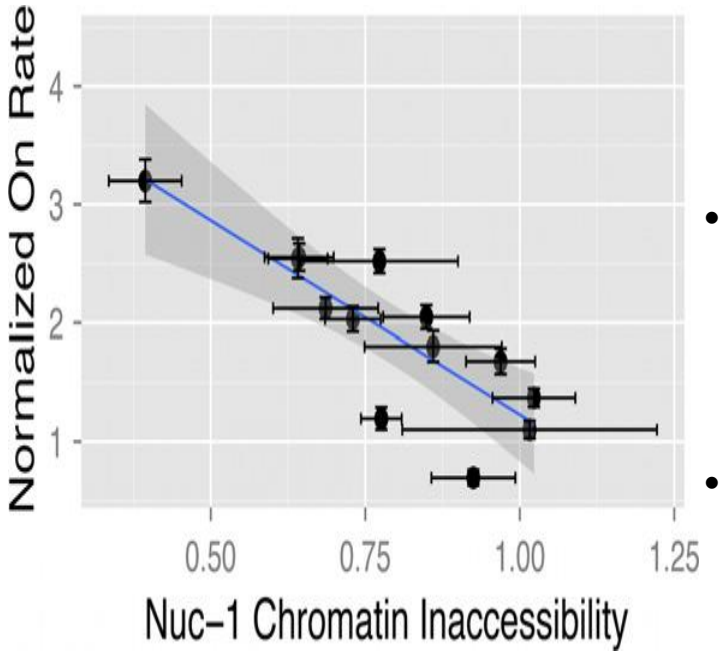








- 6 pairs of clones with similar RNA/protein mean but different noise: the ratio of chromatin inaccessibility between high and low noise clone pairs of the same mean was > 1 in all cases, implying that **clones that are integrated into more closed chromatin display noisier gene expression**
- To gain a more detailed molecular picture of chromatin features regulating gene expression noise, performed DNase I sensitivity analysis of 3 shorter regions along the length of the viral promoter:
  - Nuc-0, a nucleosome 450 nucleotides upstream of the TSS
  - The hypersensitive site (HSS) between Nuc-1 and Nuc-0, a region that contains binding sites for critical transcription factors NF-kB and Sp1
  - Nuc-1, a nucleosome that has previously been shown to be important for LTR-mediated gene expression



- In HSS low noise clones have more open chromatin: HSS has binding site for transcription factors
- Burst size is uncorrelated to the chromatin state of the promoter

The expression noise is nearly inversely correlated with the chromatin density at Nuc-1

Increased chromatin density at the transcription start site results in more infrequent transitions to the On promoter state and greater expression noise

Nucleosome at the transcription start site regulates gene expression noise and burst frequency

Conclussions

- Uncorrelated expression mean and noise suggest primarily orthogonal control across genomic locations
- Analysis by smFISH reflects properties of the full set of clones
- RNA distribution shape is highly related to protein distribution shape
- Systematic fitting of RNA distributions reveals that a two-state model can describe both low and high noise clones
- There is a differential control of expression mean and noise by burst size and rate of promoter ON transitions
- Nucleosome occupancy at the transcription start site regulates gene expression noise and burst frequency

Paper relevance

1. Expression mean and noise are uncorrelated across integration positions
2. Independence between expression mean and noise can be explained by the independent control of gene expression mean by burst size, and noise by promoter On rate
3. Chromatin density at the promoter can explain the promoter activation rate but does not provide an explanation for burst size

Consequences of the results

- The chromatin environment regulating gene expression noise could be an important mechanism to generate different cellular phenotypes from isogenic populations in a manner that can confer increased evolutionary fitness

- Nucleosomes and chromatin density around the TSS, in addition to its known functions in controlling expression levels and imparting cellular memory, may regulate that gene expression noise

Problems of the experiment

- A surprising aspect of these results is that earlier studies that were also performed using the HIV LTR system had pointed out that both burst sizes and frequencies determine the mean expression levels (Singh *et al*, 2010; Dar *et al*, 2012). A possible explanation for the apparent discrepancy is that the particular technique used in the earlier studies (half-life GFP) could have biased the analysis toward clones with higher expression, whereas the use of sm-FISH by Dey *et al* allows exploring a larger range of expression levels

- A limitation of DNAse I accessibility assay is that provides measurements of chromatin environment of averaged ensembles of cells, whereas the phenomenon of transcriptional noise is different in each single cell