

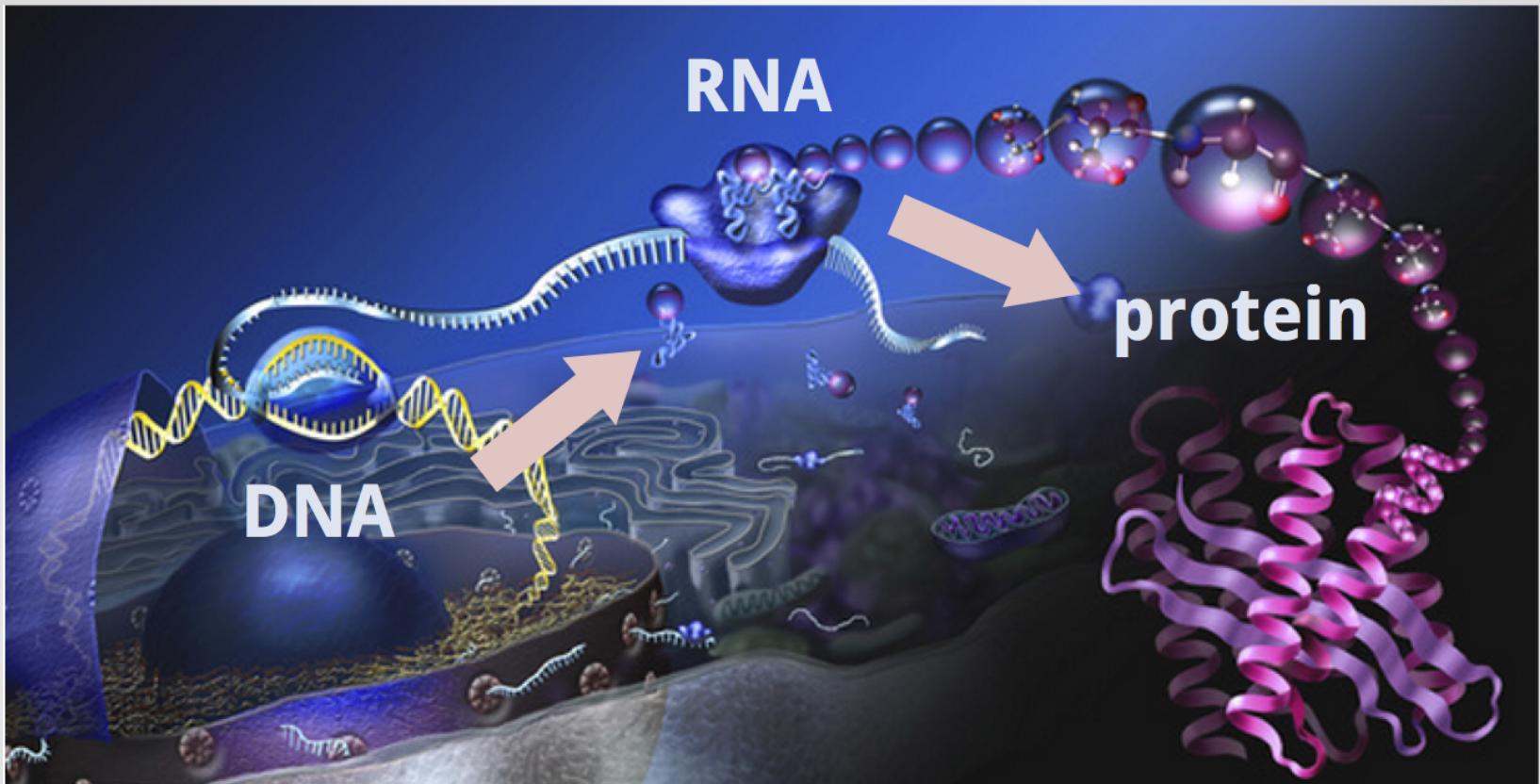
Bacterial genetics & Transcriptional dynamics in *E. coli*

Keith Hughitt

2014/06/12

The Big Picture...

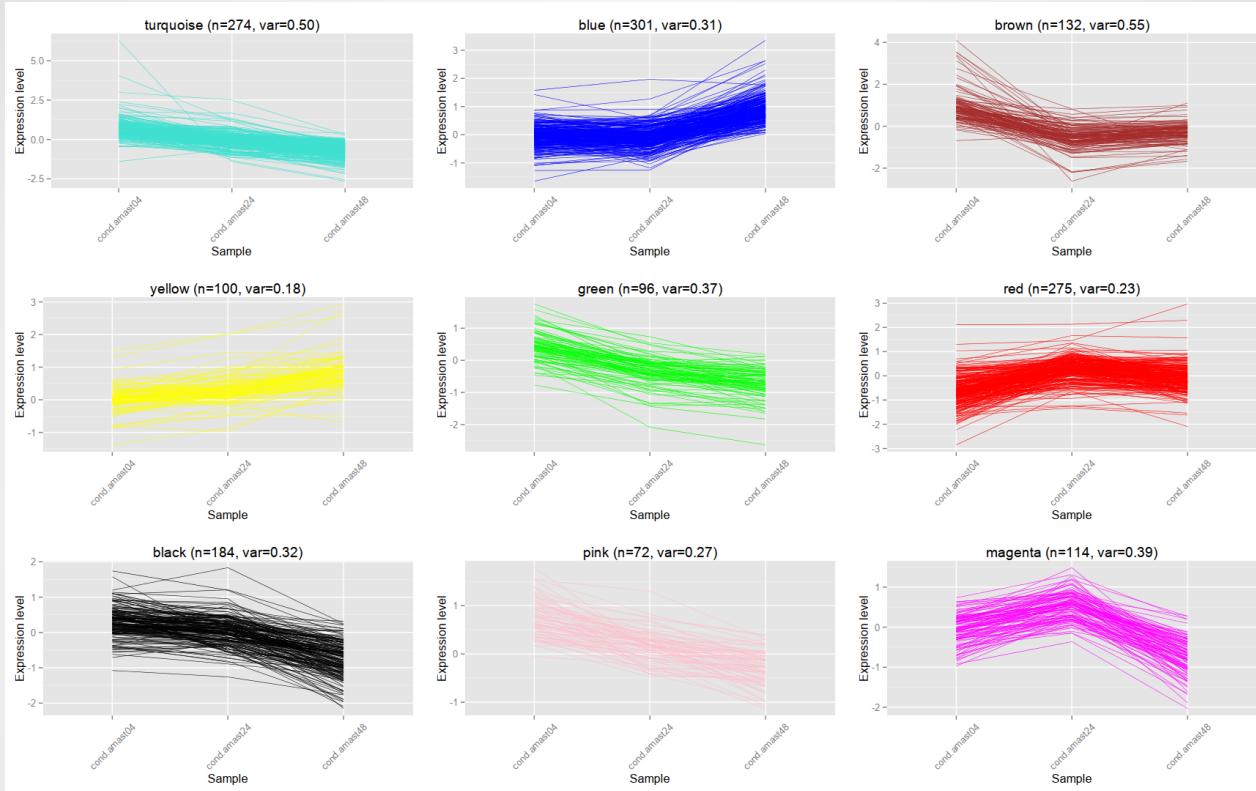
What we want to know...



<http://thisisartlab.com/2012/11/26/radiobiology/>

- We would like to know the quantities of various *proteins* produced in a cell under a given set of conditions (e.g. time during development, infection, etc.)
- However, mRNA levels turn out to be much easier to quantify, and provides a (reasonable) proxy for protein levels.
- Also, although we are most interested in understanding the expression levels of individual cells in a population, it also turns out to be easier to look at populations of cells.

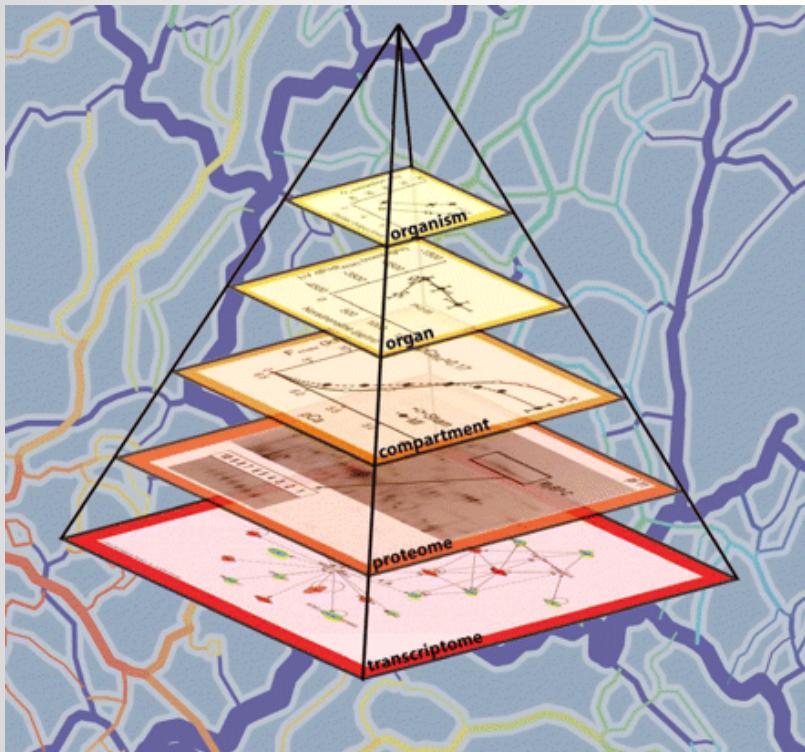
What we want to know...



(My own data; unpublished 2014)

- While it's interesting to consider the expression profiles for individual genes, and sometimes that is sufficient to understand the phenomena we are interested in, often, the levels of expression for a single genes, or even a small collection of genes, is not sufficient to explain what we are interested in.
- We need to start thinking about *systems* of interacting genes or proteins...

Our end goal...



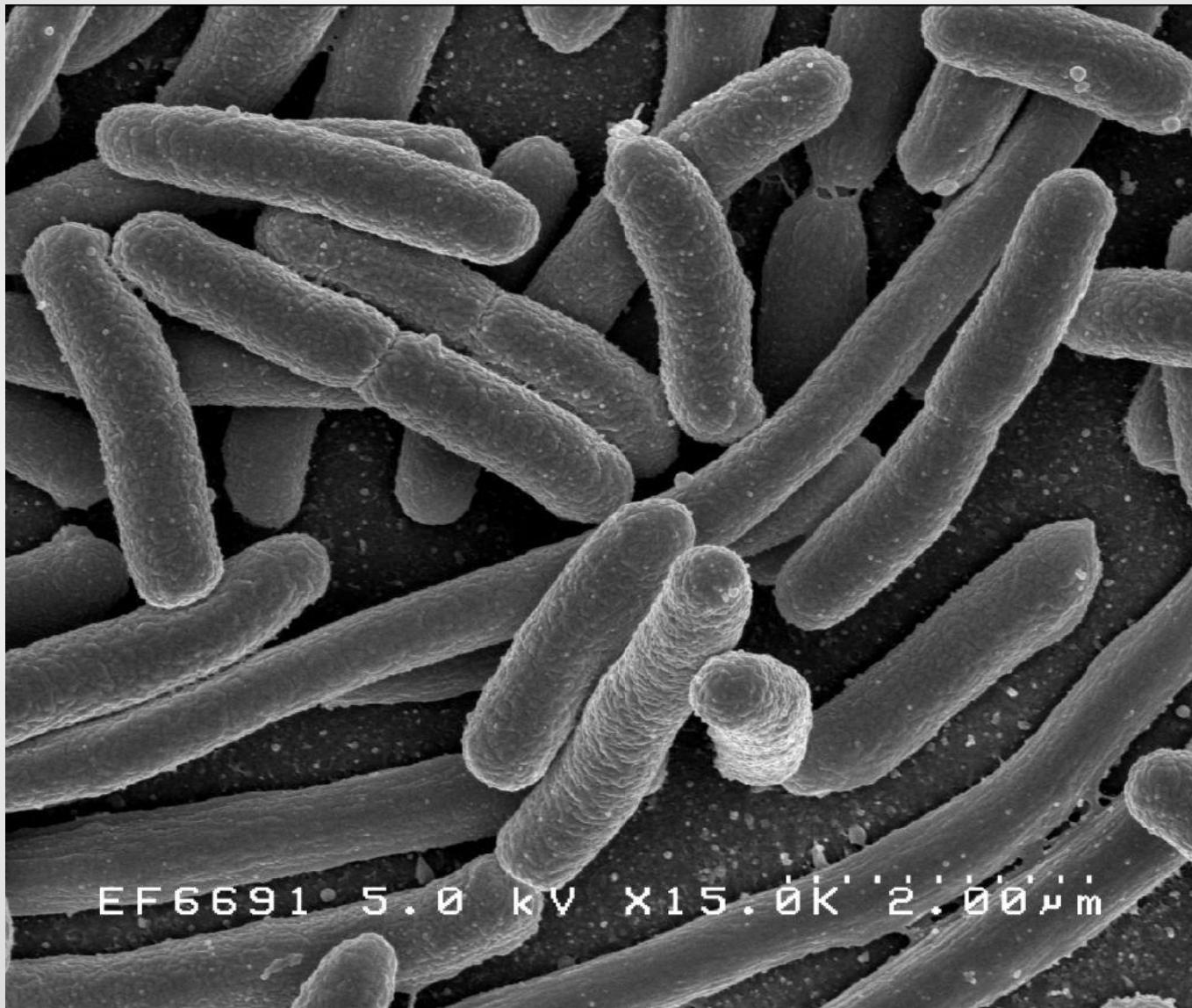
Ultimately, what we would like to do is attain a unified picture of the various systems in a cell, at all levels ranging from the entire organism (or collection of organisms) down to the level of individual genes and proteins.

Questions:

Given various types of data measurements (e.g. mRNA expression levels), how can we use that information to make inferences about the underlying networks that produced that data?

Escherichia coli

Why *E. coli*?



■ ■ ■



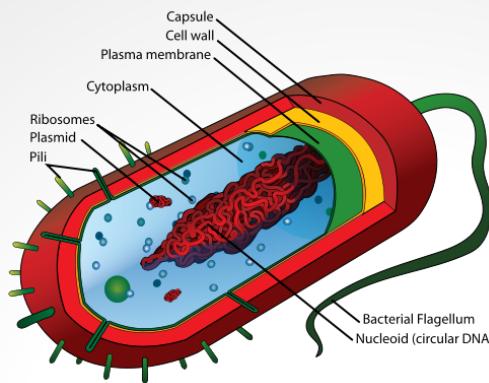
EF6691 5.0 kV x15.0K 2.00 μm

Escherichia coli

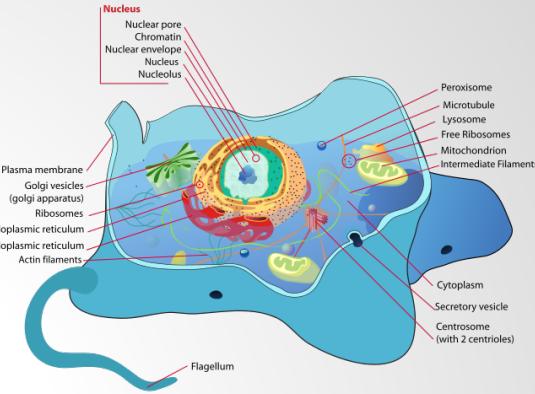
- Simple (~4000 genes)
- Well-annotated
- Abundance of microarray, RNA-seq, etc. data available
- Probably the most well-described networks of any organism
- Principles learned in *E. coli* can be applied directly to other prokaryotes and with some differences, give us insight into networks in Eukaryotes as well
- Makes validating networks more straight-forward!

Biology 101

Prokaryotes vs. Eukaryotes



Prokaryotes



Eukaryotes

nucleus	no nucleus	has nucleus
size	1-10 um	10-100um
chromosomes	single circular chromosome	multiple linear chromosomes
cell wall	cell wall	cell wall in plants and fungi
expression	simultaneous transcription & translation	separate transcription and translation
cytoskeleton	not present	present
membrane-bound organelles	not present	present
genome	small-ish / no histones	large-ish / histones

Transcriptional regulation in bacteria

- **Operons**
 - Bacterial genes commonly linked together in functional units known as **operons**.
 - The genes in a given operon are contiguous in the bacterial DNA.
 - All genes in a given operon may be controlled by a single promoter, thus allowing the collection of genes to be turned off/on.
 - Compare this to Eukaryotes where genes in related pathways are *not* usually co-located and are regulated individually (although the same TF may regulate a number of genes.)
- Polycistronic transcription
- No introns
- Transcriptional regulation is main source of control of gene expression in prokaryotes.

General properties of transcriptional time series in *Escherichia coli*

Lok-hang So, Anandamohan Ghosh, Chenghang Zong, Leonardo A Sepúlveda, Ronen Segev &
Ido Golding

Overview

Stochastic gene expression

- Transcription occurs in bursts which can be modelled as a stochastic process.
- Individual cells in a population exhibit variability in the timing of these bursts.

Study

- By looking at individual cells, it is possible to measure the dynamics of transcription.
- This can be modeled using a two state (on/off) model.
- In this model, there are multiple ways to produce a given expression level.
- What are the cells actually doing? Does this vary between genes or between cells?
- Single-cell FISH data used to measure expression across time and attempt to select the correct model.

Two-state model of gene expression

- **Model:**
 - k_{on} = rate of transcription initiation (rate of bursts)
 - k_{off} = rate of transcription stopping (duration of bursts)
 - k_{TX} = rate of transcription (magnitude of bursts)
 - k_d = rate of mRNA decay.
- Reaction rates used to determine probability / unit time.
- Simulated using the Gillespie algorithm
- 1000 stochastic simulations performed and for each one a time point was chosen randomly to mimic smFISH experiment.

Two-state model of gene expression

Fano factor

- $b = \text{var} / \text{mean}$ mRNA copy number
- Measurement of the dispersion of a probability distribution (noise to signal ratio.)
- Here, it is used to quantify the “burstiness” (b) of an expression pattern:
 - Represents how “bursty” a time series is relevant to a Poisson process.
 - $b = 1$ Nonbursty (Poissonian) mRNA production
 - Since we can calculate mean and var in the two-state model, we can infer b .
- If only one kinetic parameter is varying, profile of b can be predicted for a given n ; therefore can work backwards to determine which parameter is fluctuating.

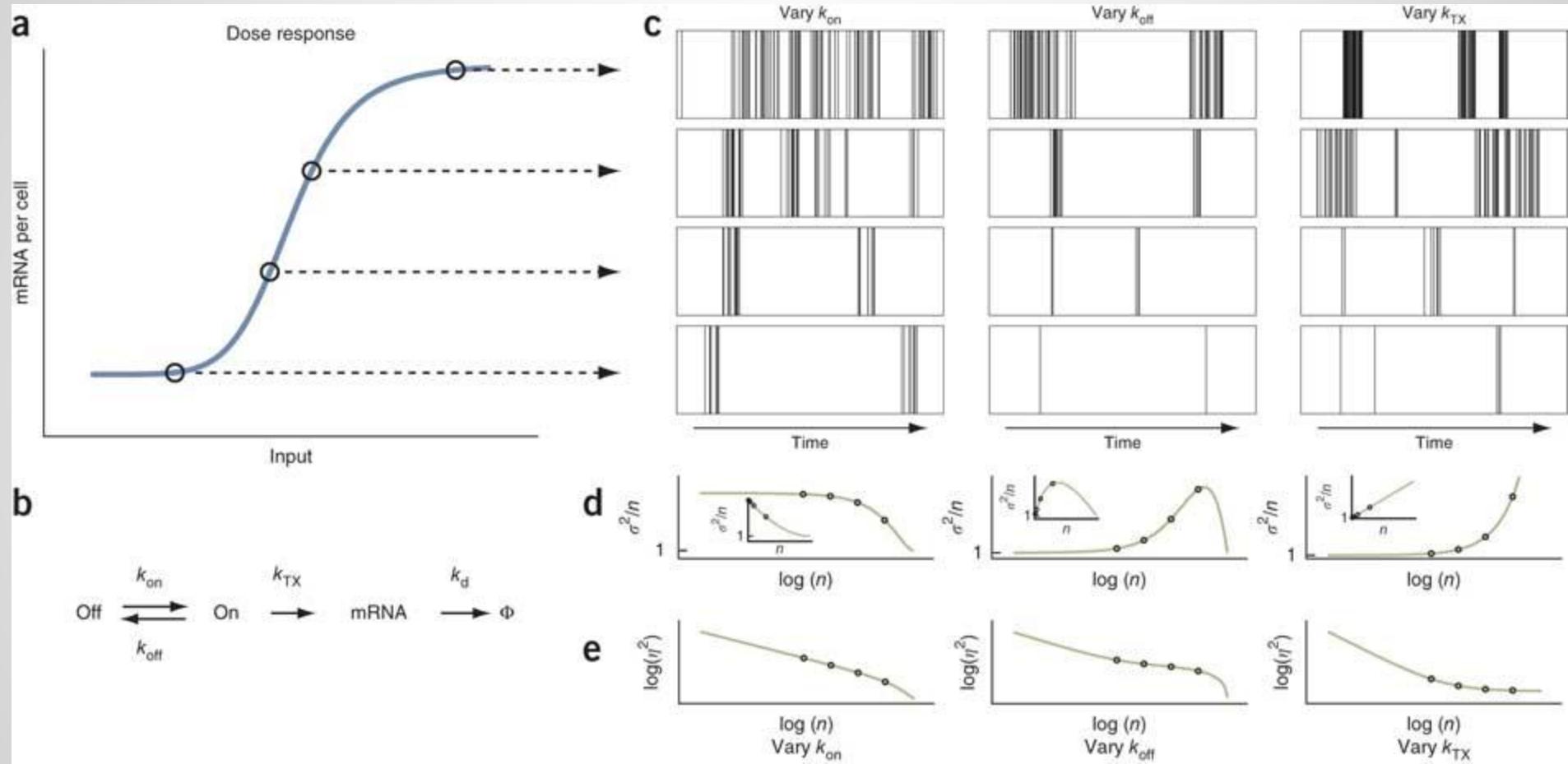


Figure 1: Different features of the transcriptional time series can be modulated to vary gene expression level.

Typical bacterial promoter activity profile

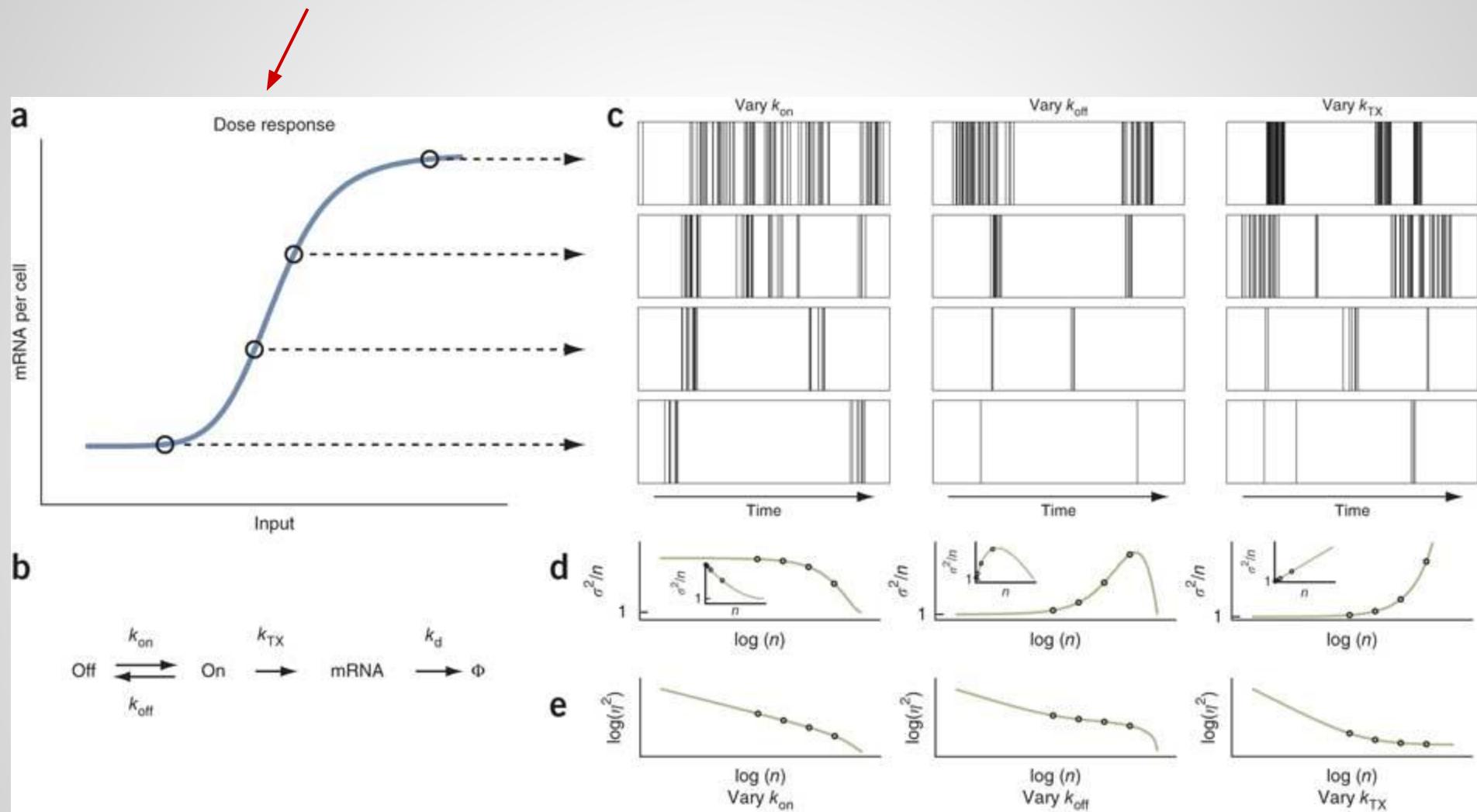


Figure 1: Different features of the transcriptional time series can be modulated to vary gene expression level.

Different ways of achieving the same expression output level, modifying only one parameter

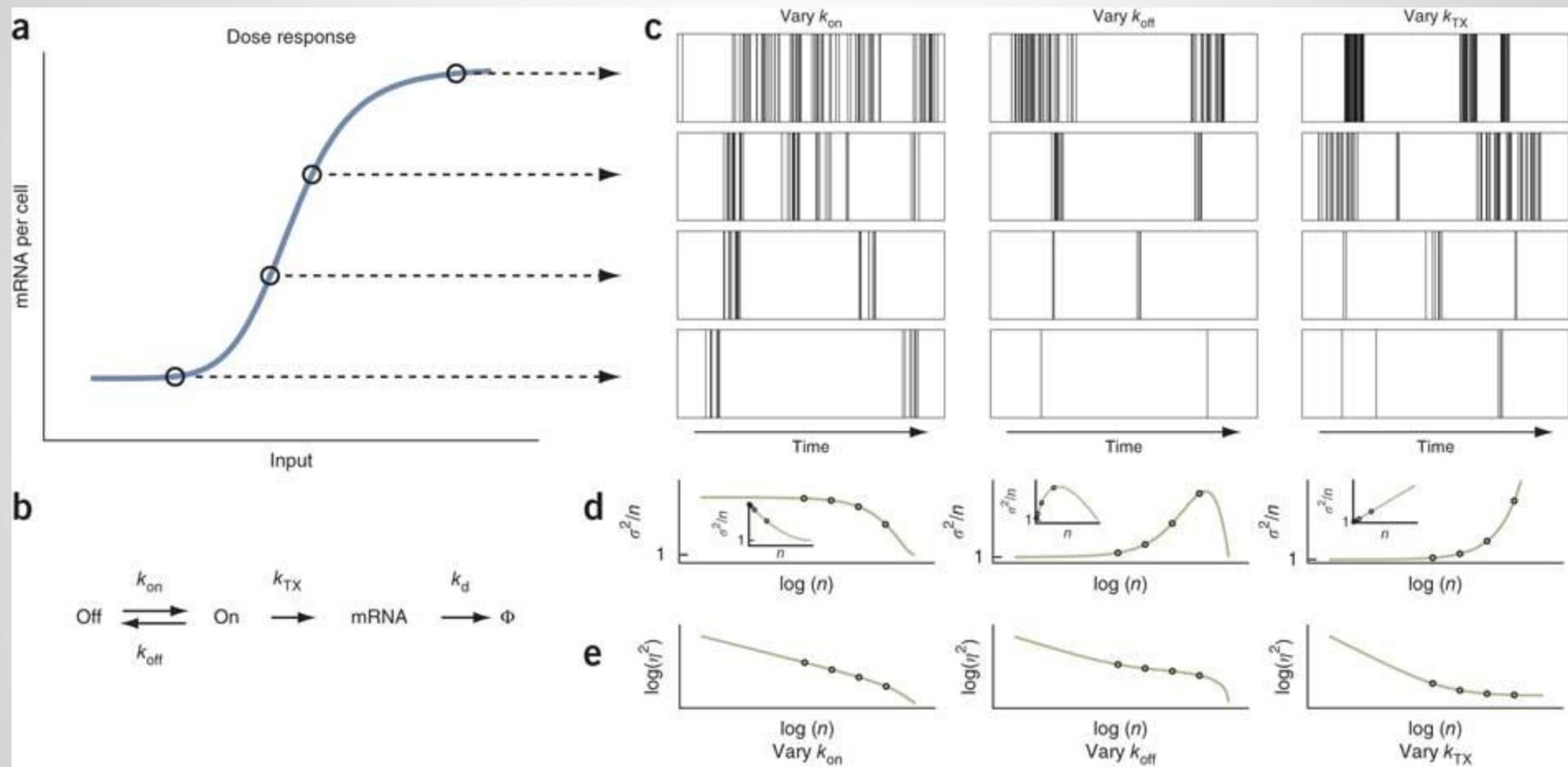
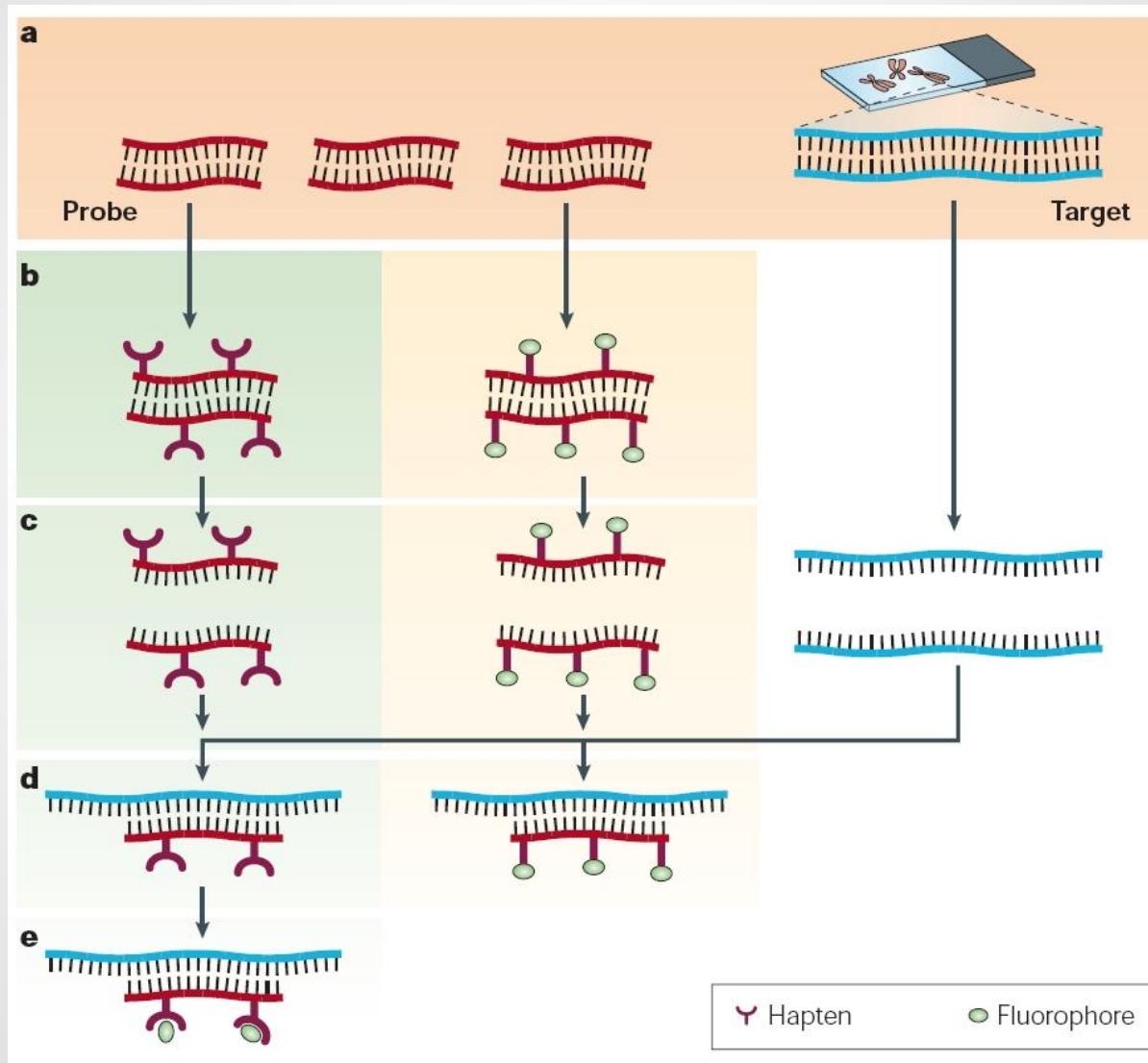


Figure 1: Different features of the transcriptional time series can be modulated to vary gene expression level.

Quantifying mRNA statistics

- Single molecule Fluorescence *in situ* hybridization (FISH) used to measure expression of individual cells
 - Limited dynamic range, but seems to be okay here.
 - Consistent with other expression assays (qPCR) and measurements from literature.
- By measuring $\langle n \rangle$ and var., can compute the burstiness (b) of a given transcriptional time series.
- mRNA histograms fit a negative binomial distribution.
 - Same distribution used in RNA-Seq differential expression analyses to model expression across variants for a given genes.

Fluorescence *in situ* hybridization (FISH)



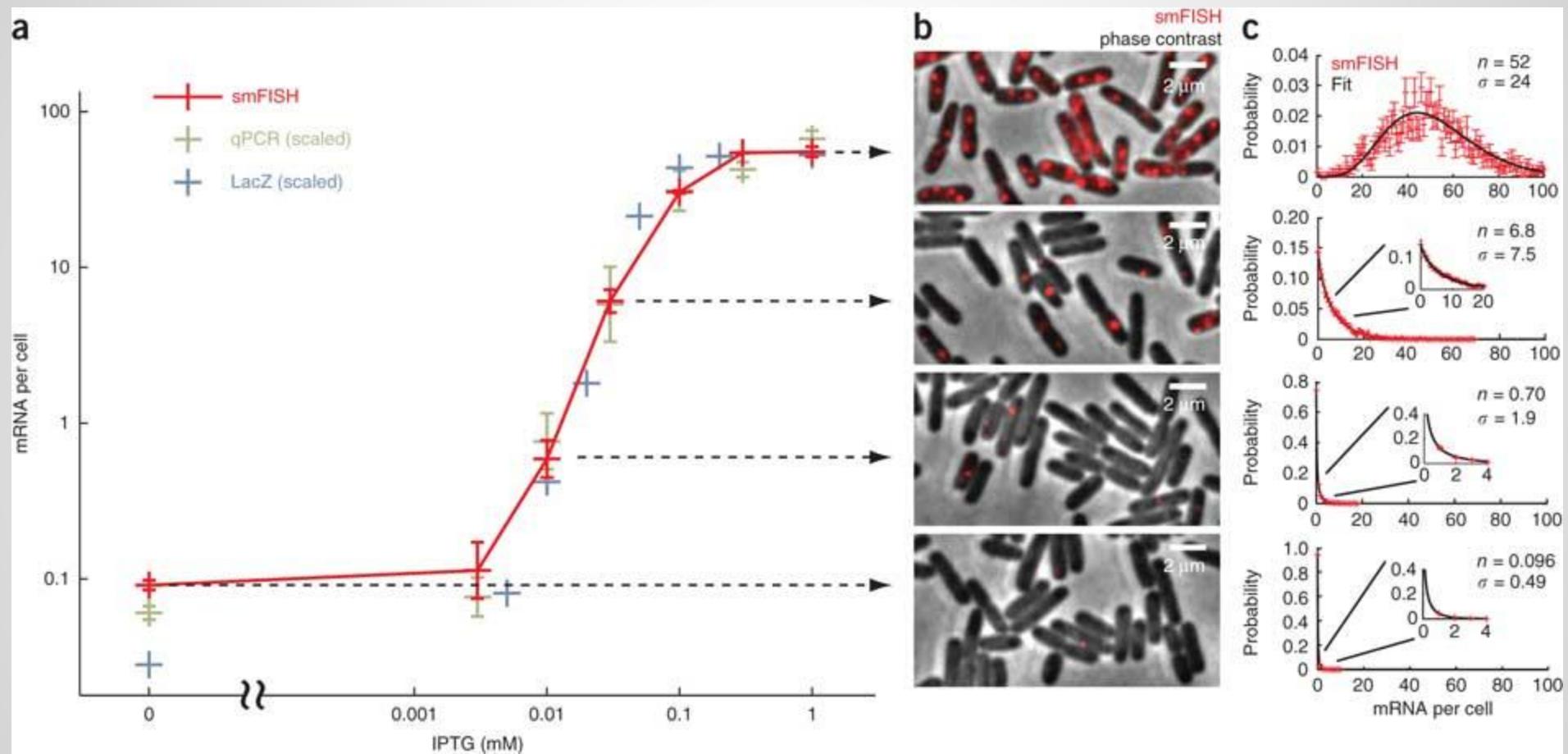


Figure 2: Single-molecule FISH (smFISH) used to characterize mRNA copy-number statistics.

Levels of mRNA expression as sugar is added.

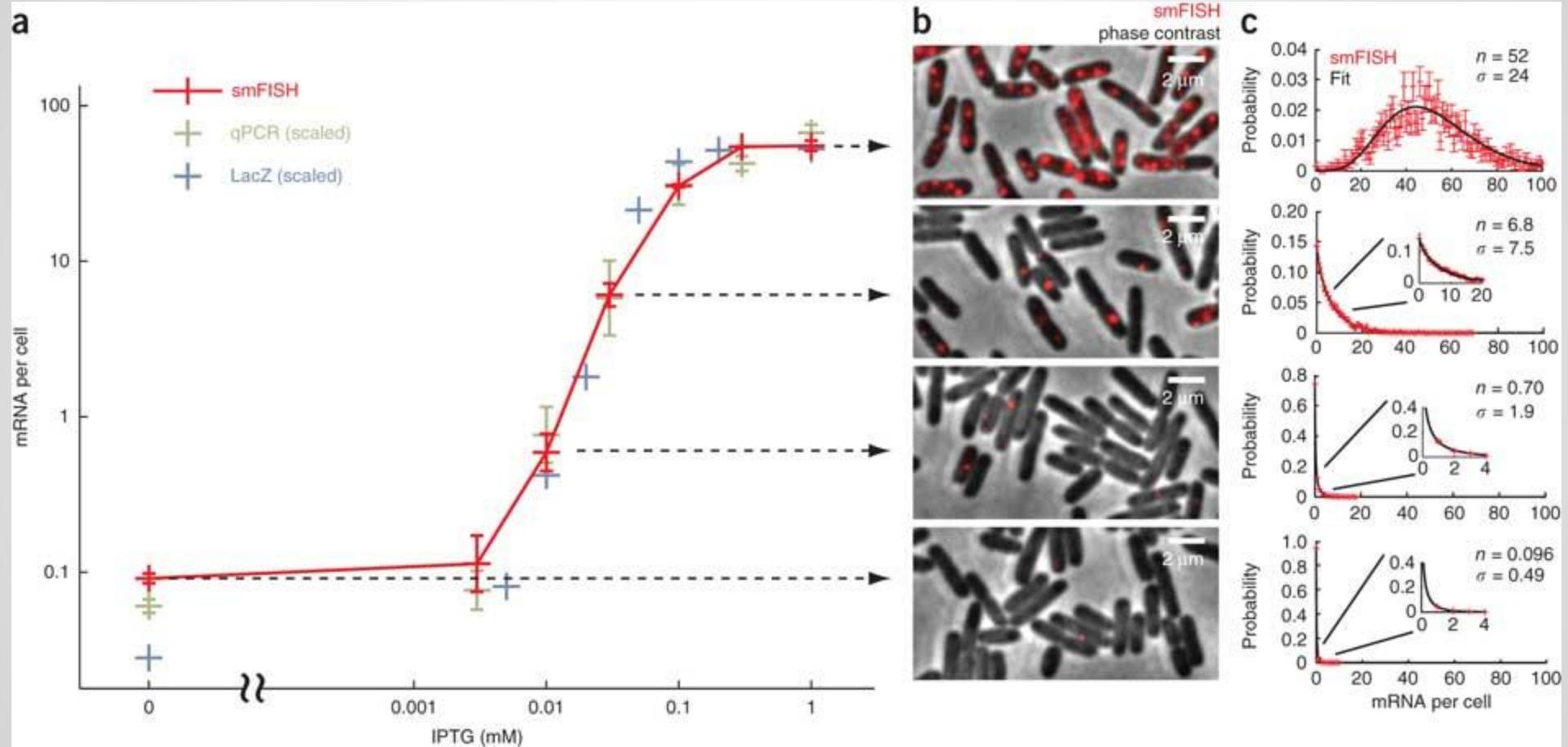


Figure 2: Single-molecule FISH (smFISH) used to characterize mRNA copy-number statistics.

blue = beta-gal (gene in the lac operon) activity
levels reported in literature...

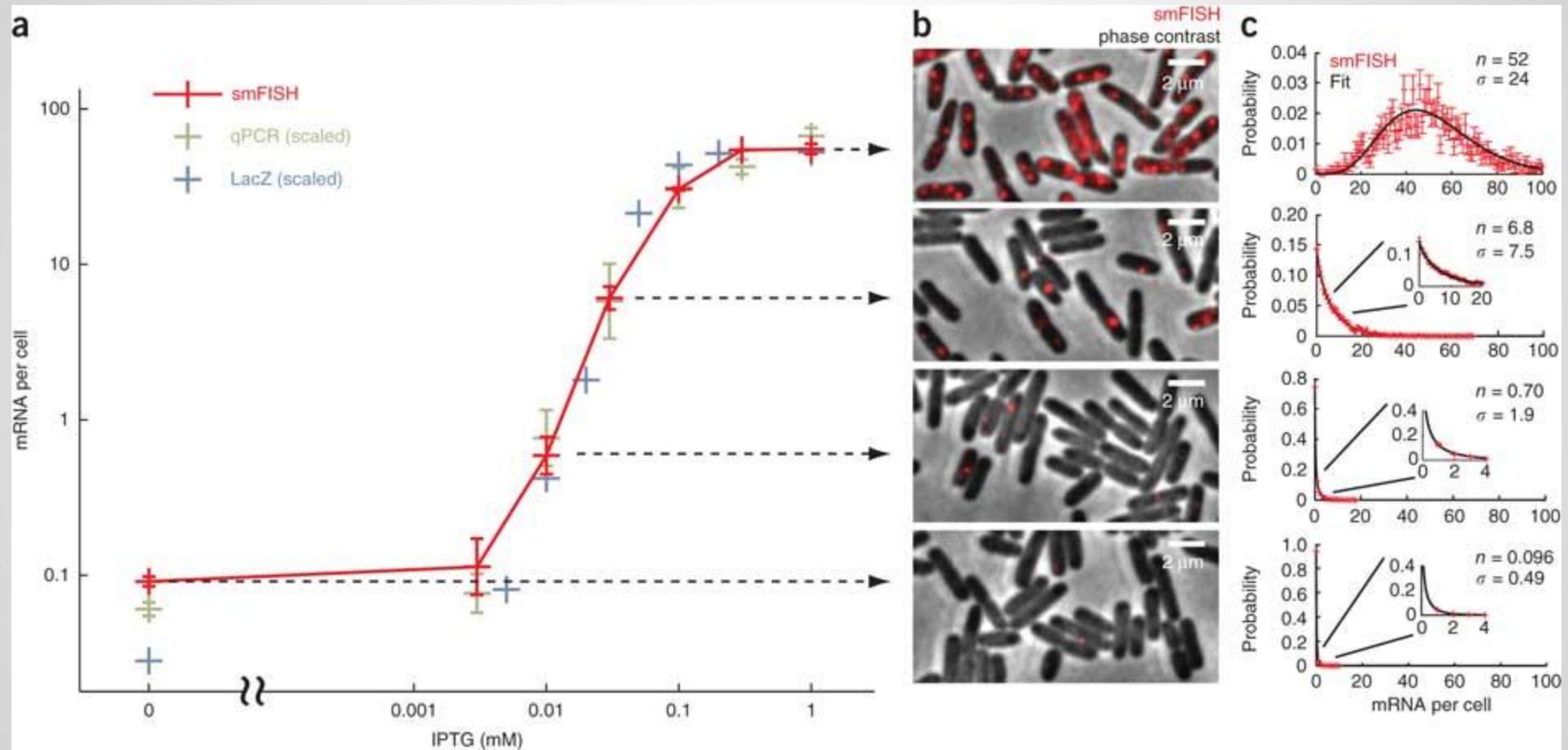


Figure 2: Single-molecule FISH (smFISH) used to characterize mRNA copy-number statistics.

smFISH measurements fit to negative binomial

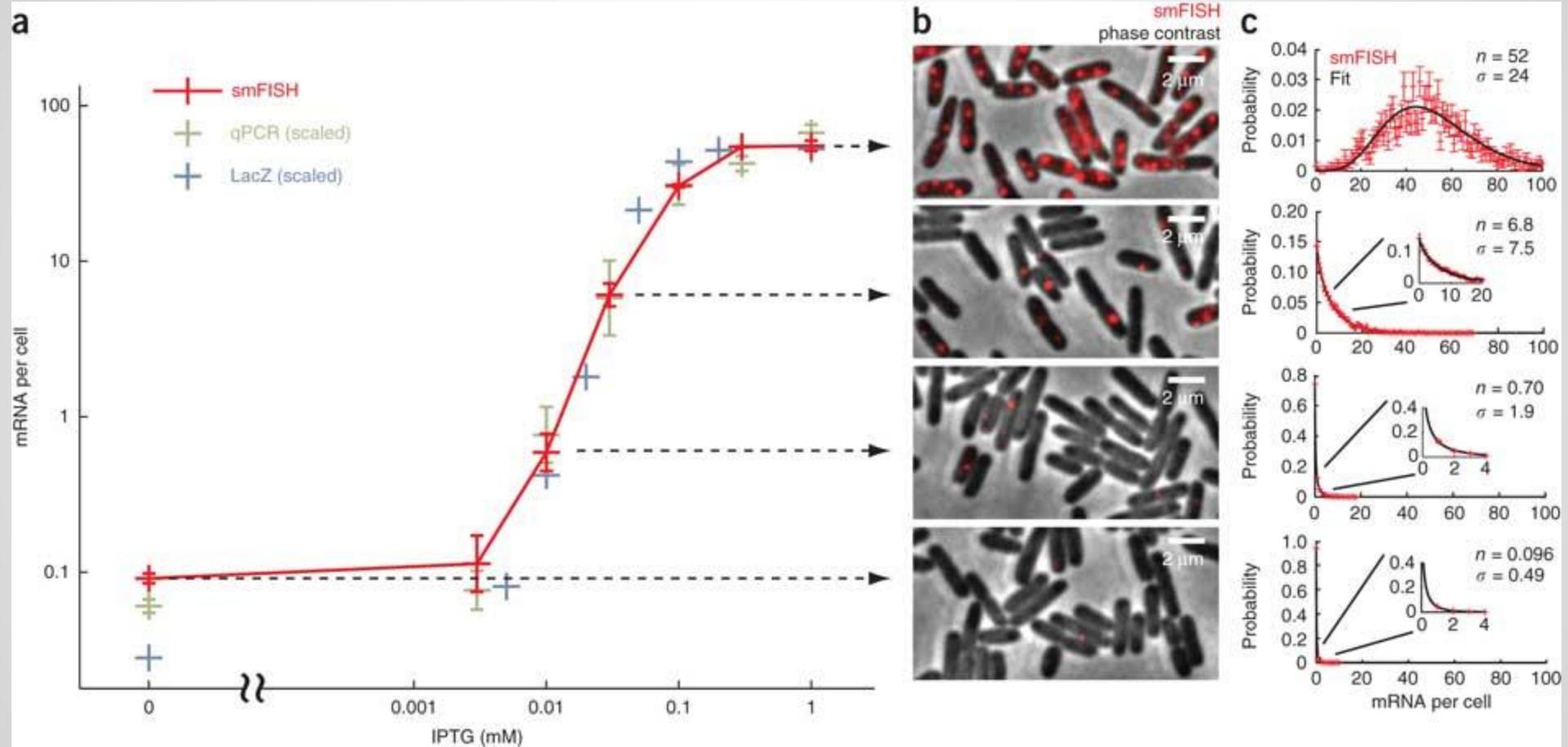


Figure 2: Single-molecule FISH (smFISH) used to characterize mRNA copy-number statistics.

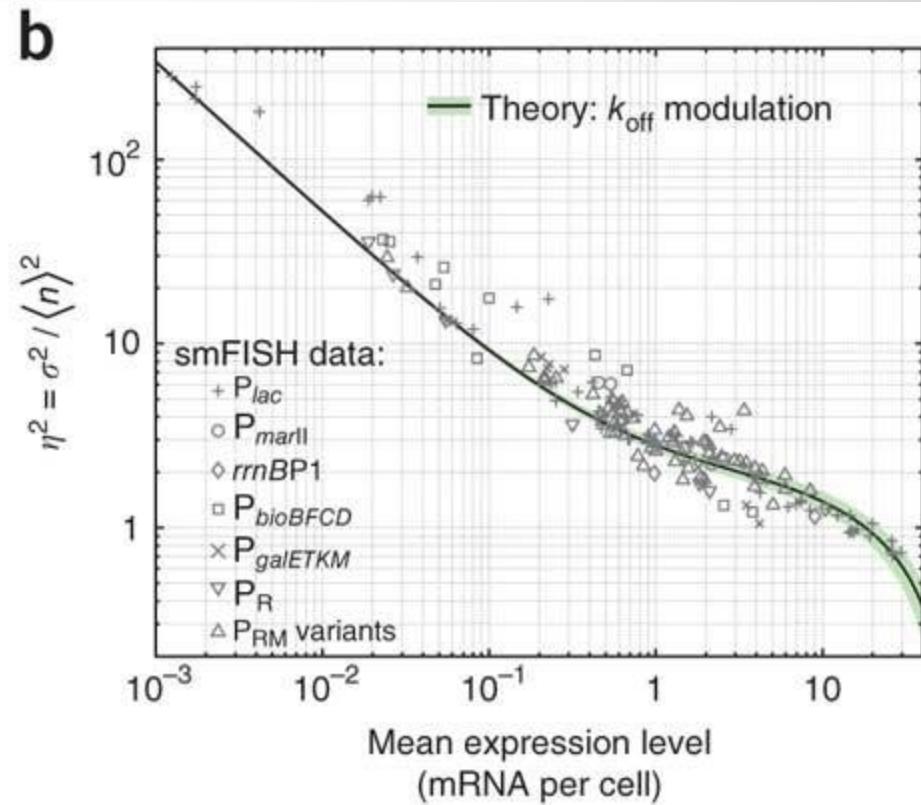
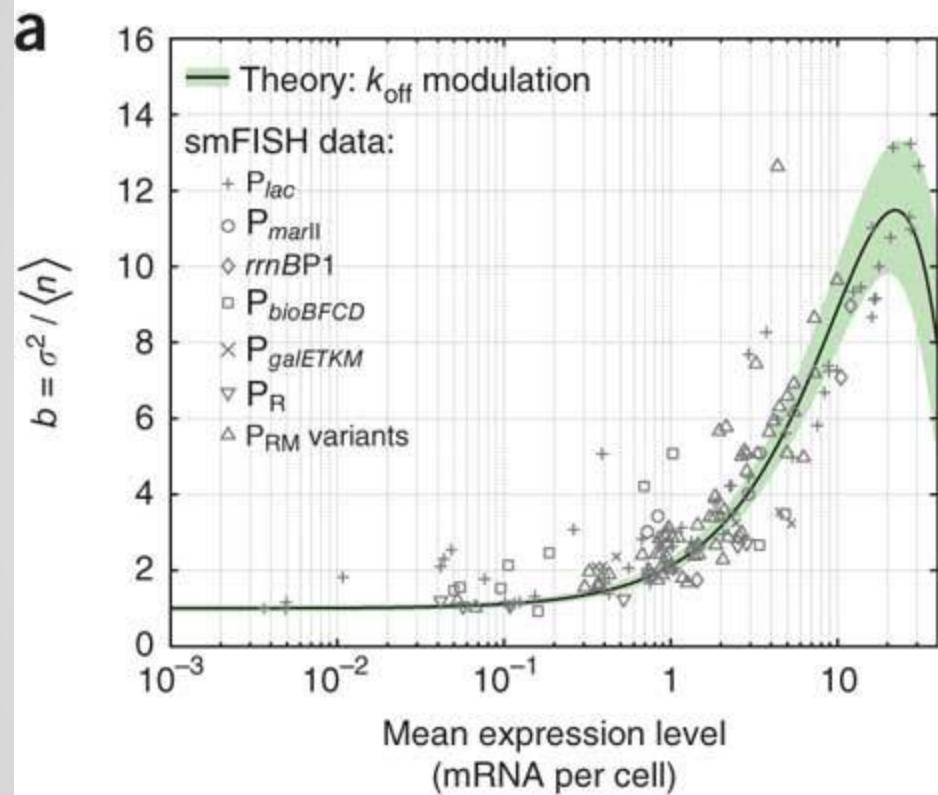


Figure 3: Gene expression level in *E. coli* is varied by changing the gene off rate.

burstiness increases as $\langle n \rangle$ increases;
pattern is similar for different promoters

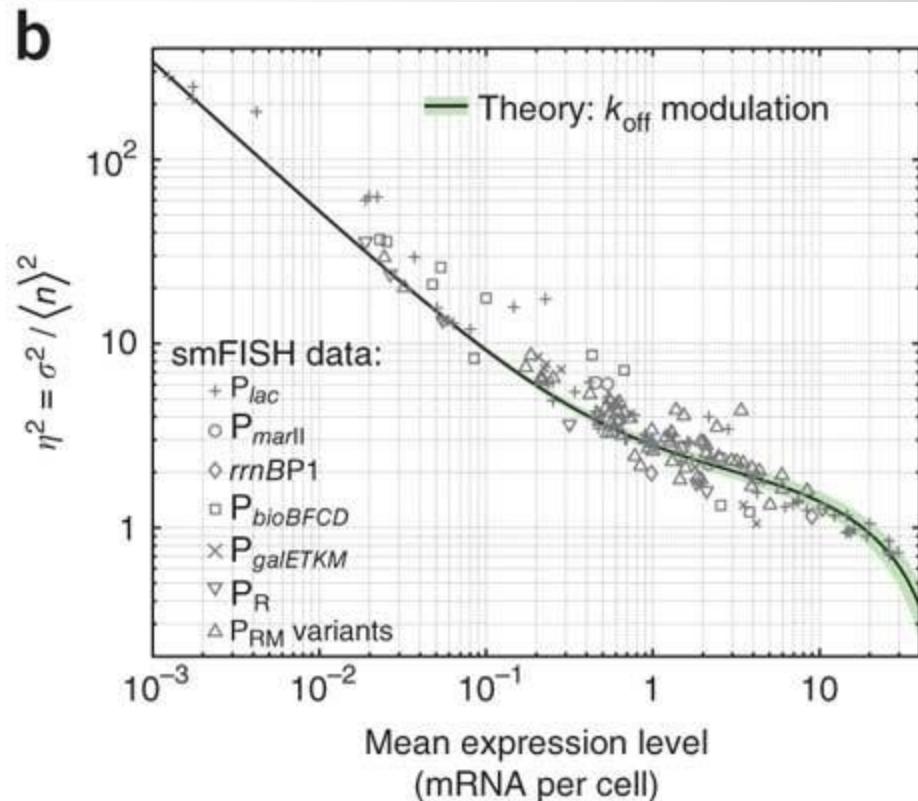
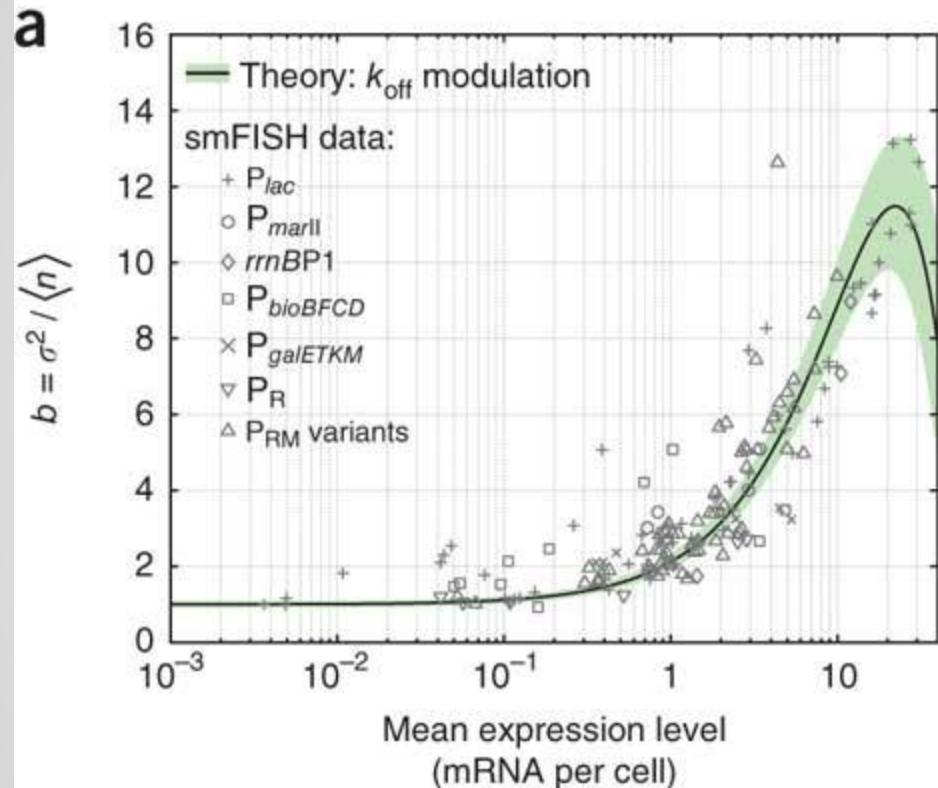


Figure 3: Gene expression level in *E. coli* is varied by changing the gene off rate.

noise as a function of $\langle n \rangle$

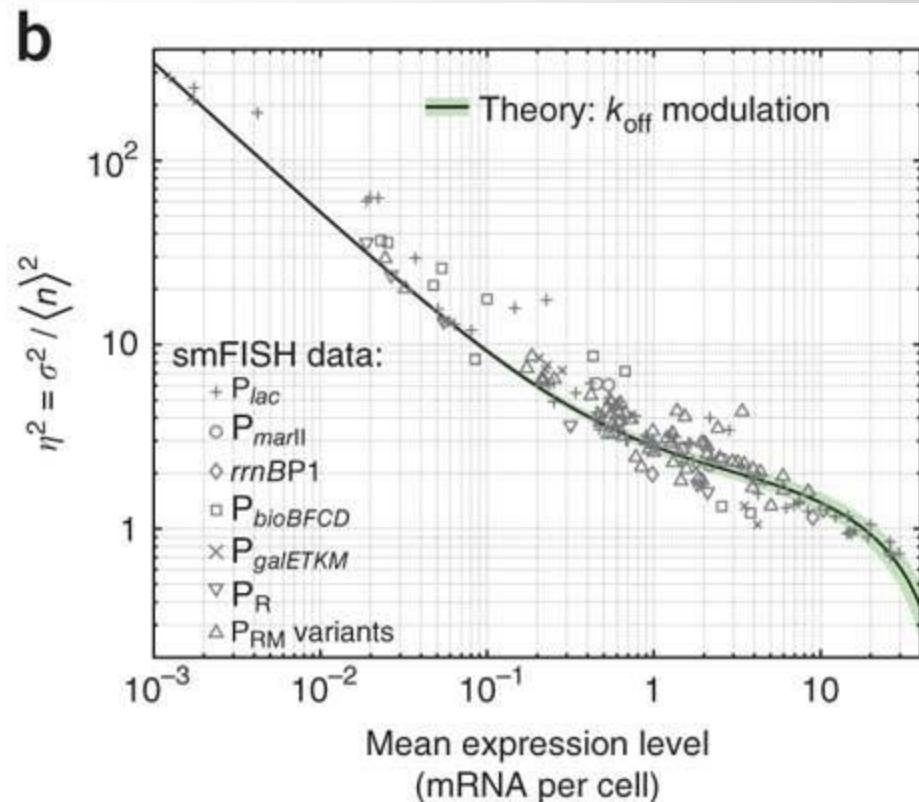
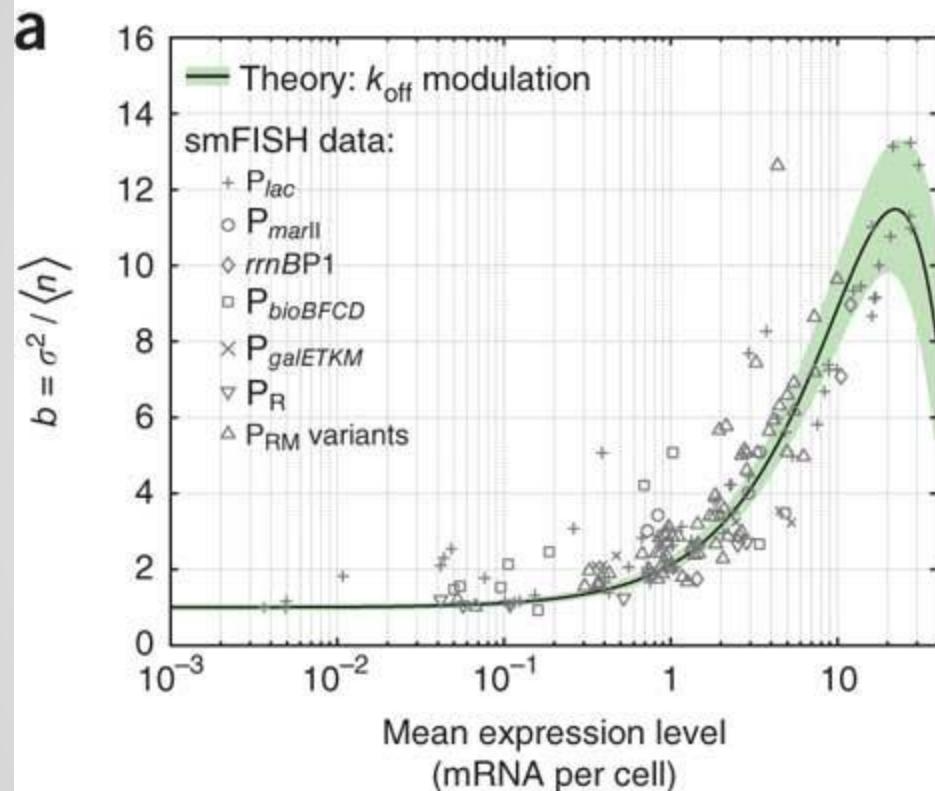


Figure 3: Gene expression level in *E. coli* is varied by changing the gene off rate.

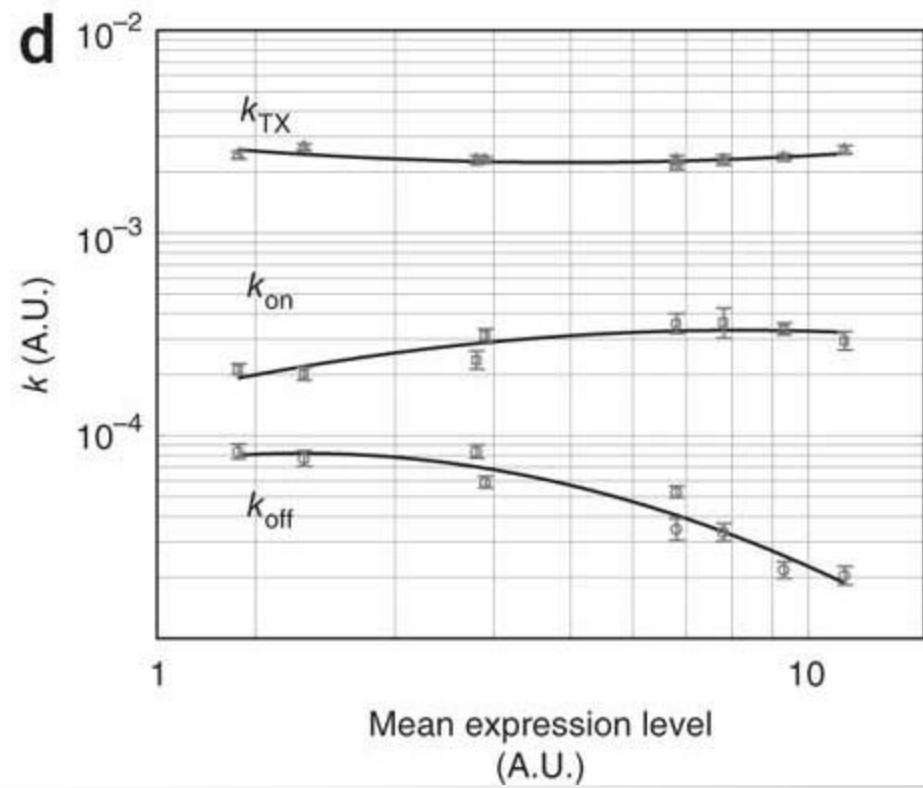
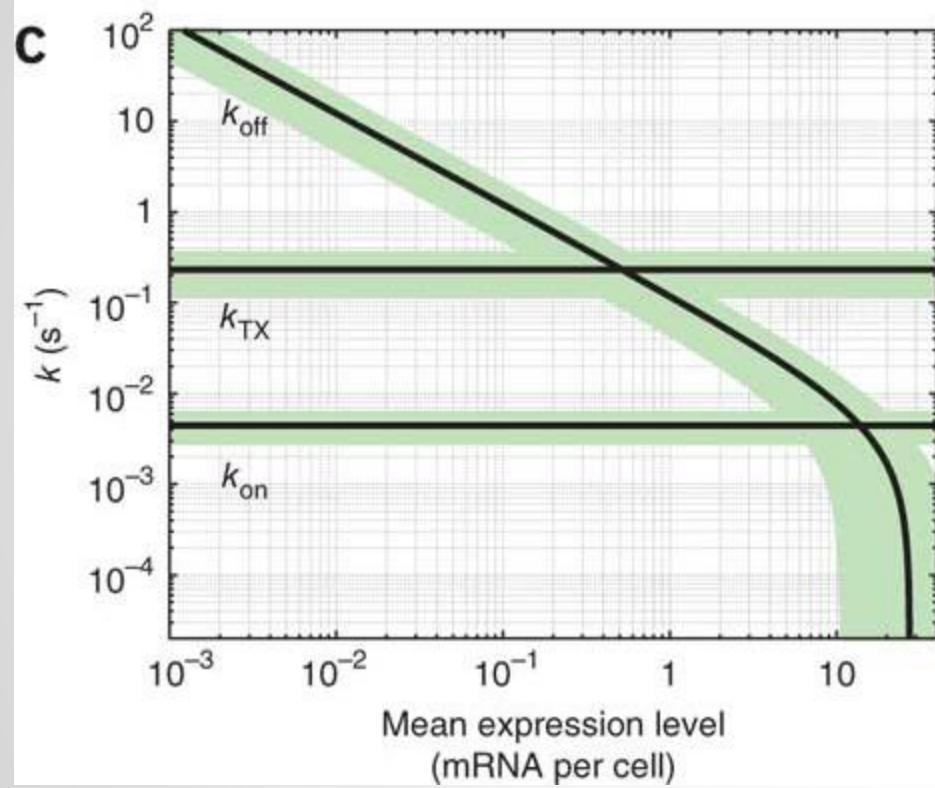


Figure 3: Gene expression level in *E. coli* is varied by changing the gene off rate.

Estimated rate parameters (green shaded region represents variability between promoters)

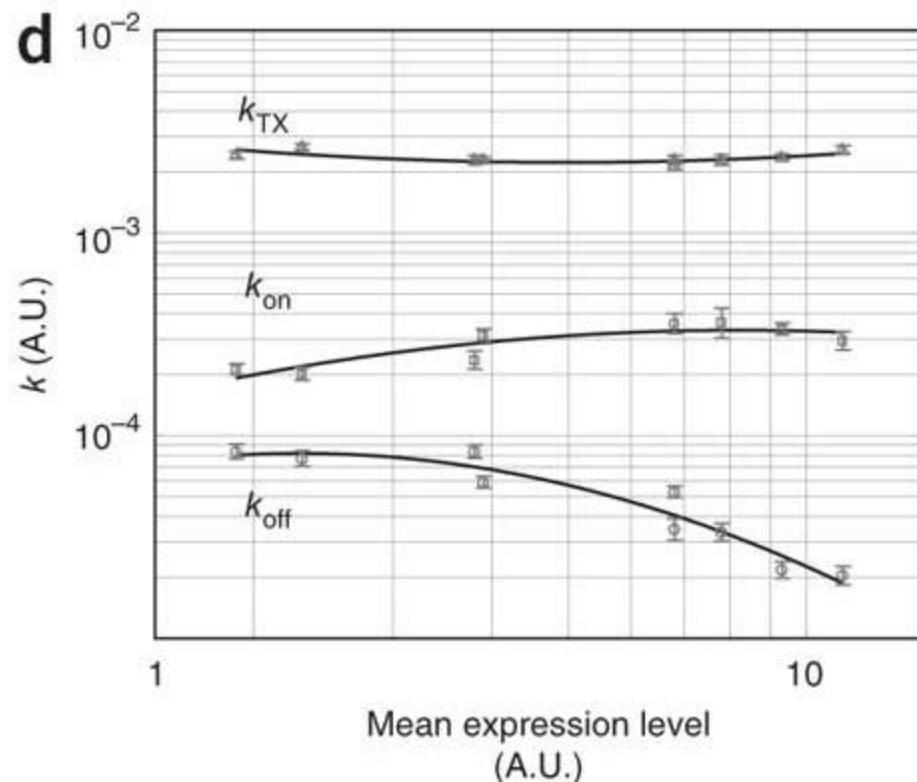
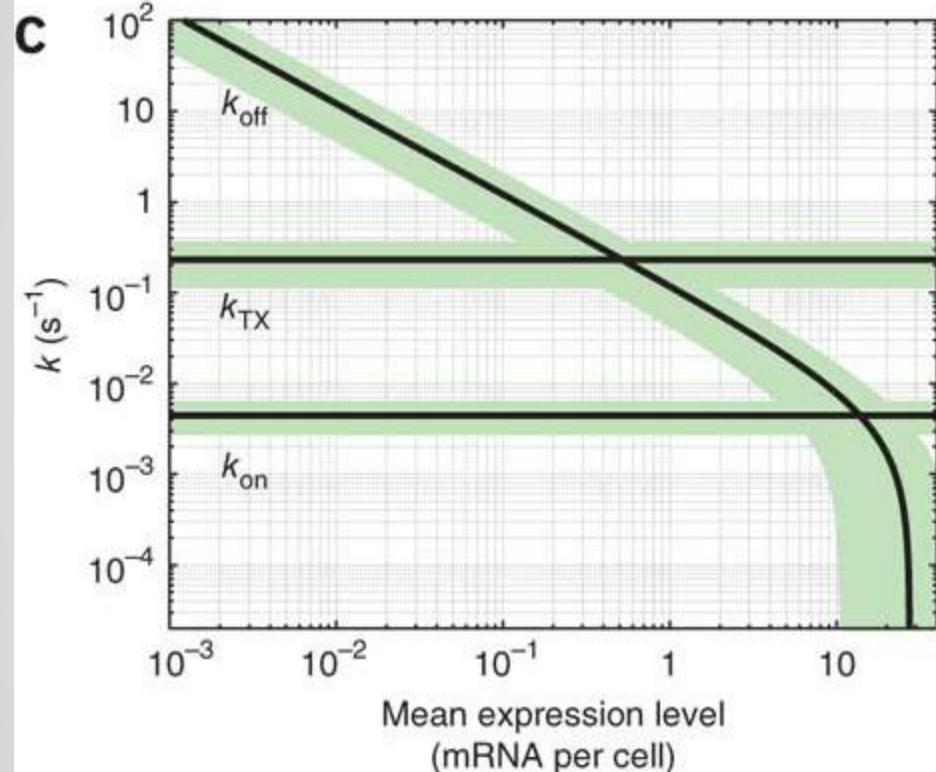


Figure 3: Gene expression level in *E. coli* is varied by changing the gene off rate.

Actual values based on MS2-GFP assay for a single promoter ($P_{lac/ara}$)

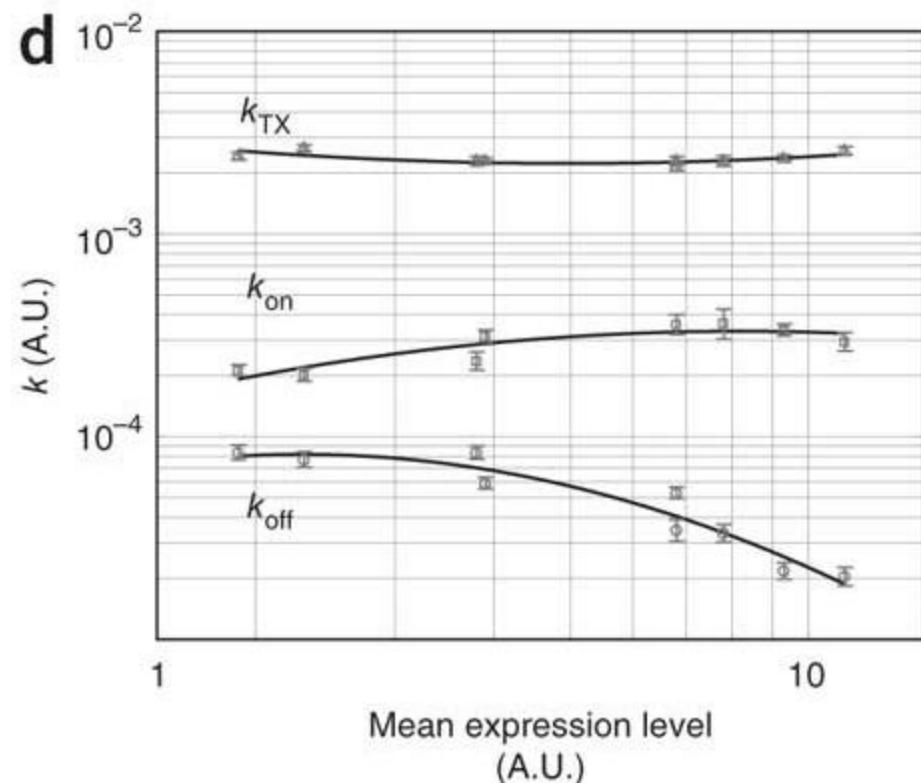
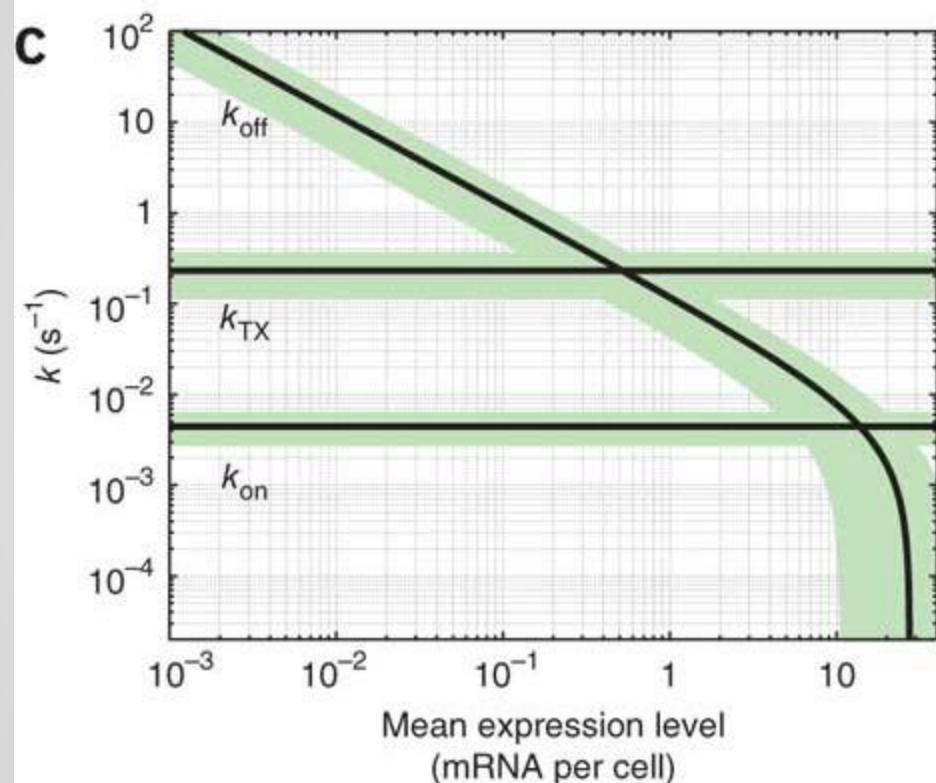


Figure 3: Gene expression level in *E. coli* is varied by changing the gene off rate.

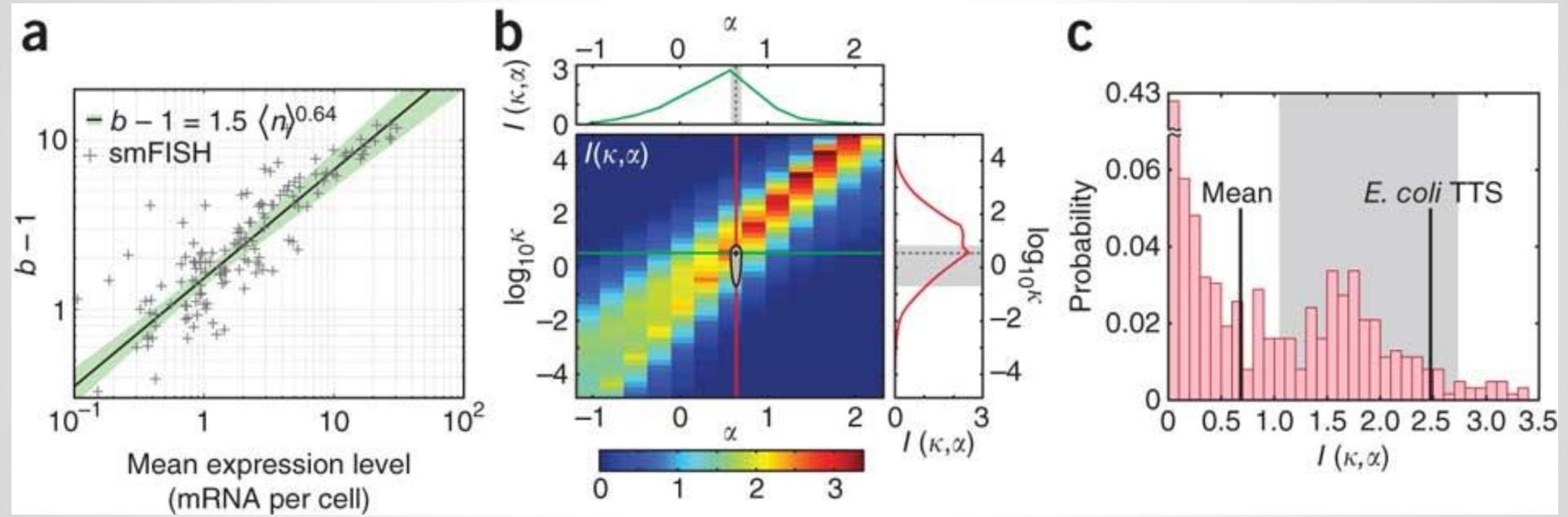


Figure 4: The transcriptional time series optimizes information representation by the cell.

Conclusions

- Cell-to-cell variability in mRNA levels is dominated by fluctuations in the two-state process, and not due to differing kinetic parameters.
- Burstiness increases with increasing gene activity; non-bursty around $\langle n \rangle = 1$.
- Kinetics of expression are gene-independent; depends only on the expression level $\langle n \rangle$.
 - All genes expressed at a given level have a similar transcriptional time-series.
- The duration of transcriptional bursts (k_{off}) appears to be the main kinetic parameter determining expression output; k_{on} and k_{TX} are relatively static. (maybe...)
- The transcriptional parameters are fairly optimized for responding to specific signals (e.g. environmental cues.)
- Still don't have a mechanistic understanding of what produces transcriptional bursts (in Eukaryotes, it's thought to be related to chromatin modifications)

References

- So, L.-H., Ghosh, A., Zong, C., Sepúlveda, L. a, Segev, R., & Golding, I. (2011). General properties of transcriptional time series in *Escherichia coli*. *Nature Genetics*, **43**(6), 554–60. doi:10.1038/ng.821