BIOZENTRUM
Universität Basel
The Center for Molecular Life Sciences

SIB
Swiss Institute of Bioinformatics

Benefits of an erratic boss:

Expression noise facilitates the evolution of gene regulation

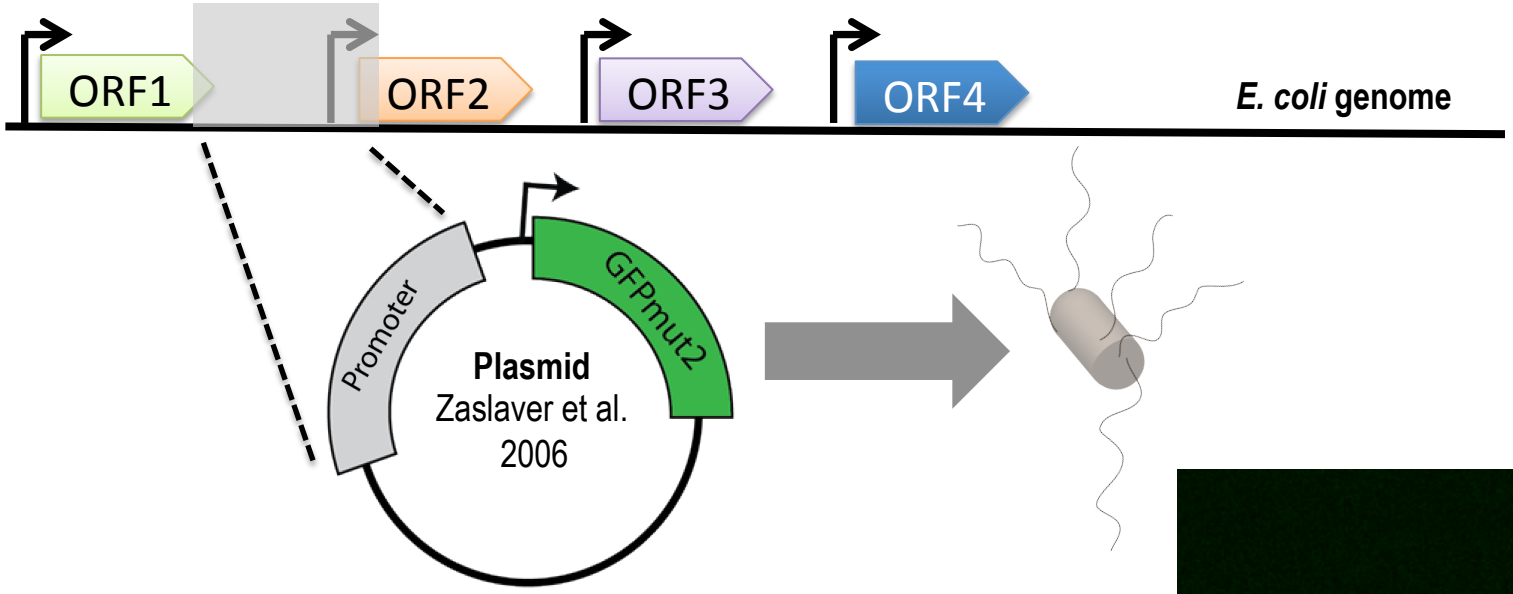
Luise Wolf **Olin Silander**

Outline

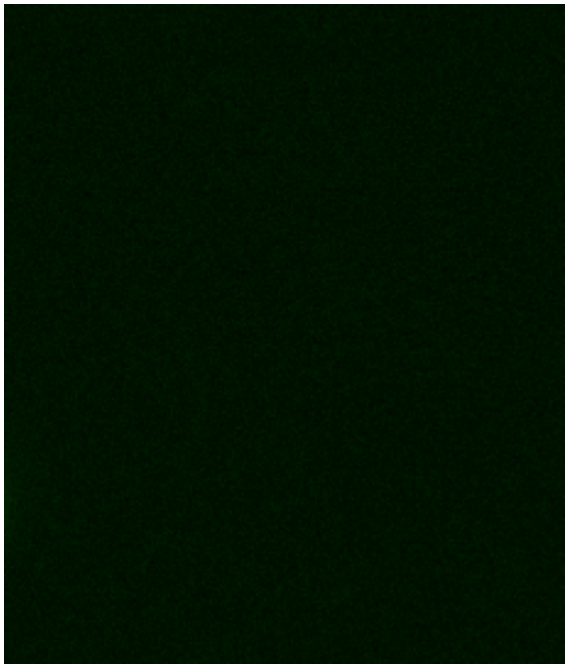
- Different *E. coli* promoters display different noise levels.
- How has natural selection acted on promoter noise?
- Evolved *synthetic promoters* under controlled selective conditions and measured their noise.
- Comparison implies selection has acted to *increase noise* in native *E. coli* promoters.
- General correlation between promoter noise and amount of regulation of the promoter.
- Bet hedging or unavoidable side-effect of regulation? Answer: both.
- General theory: Transmission of noise from regulator to target is often functional.
  Allows for smooth evolution of finely tuned regulation.

http://biorxiv.org/content/early/2014/07/18/007237

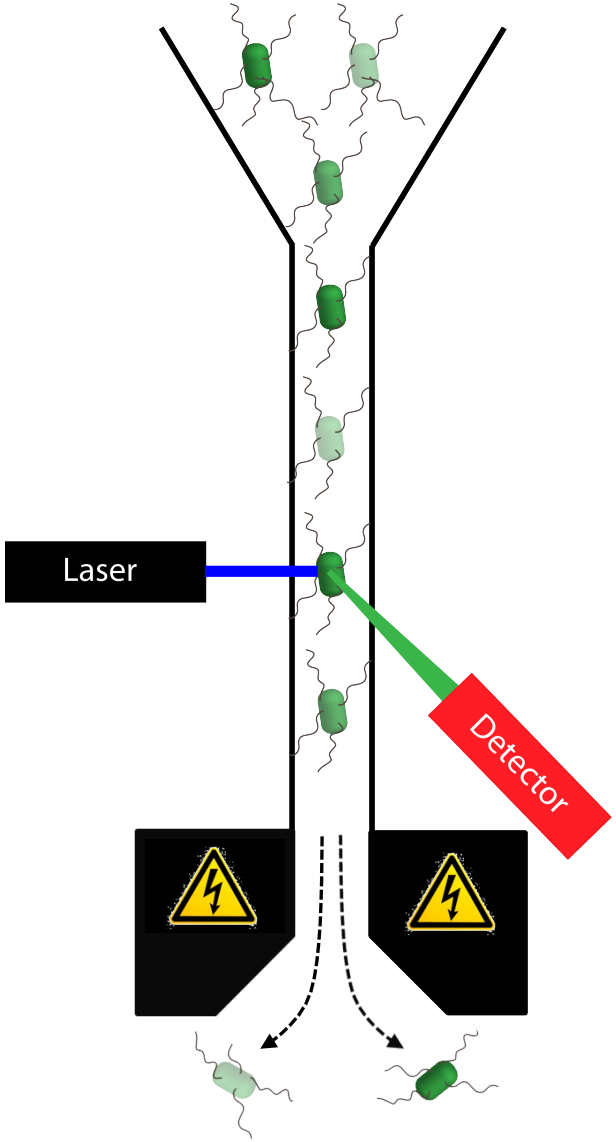
Measuring transcription from each *E. coli* promoter in single cells



- GFP fluorescence per cell proportional to protein number.
- Besides microscopy, GFP levels of single cells can be **measured in high-throughput using FACS**.
- Quantitatively characterize the distribution of expression levels across single cells, for all *E. coli* promoters.

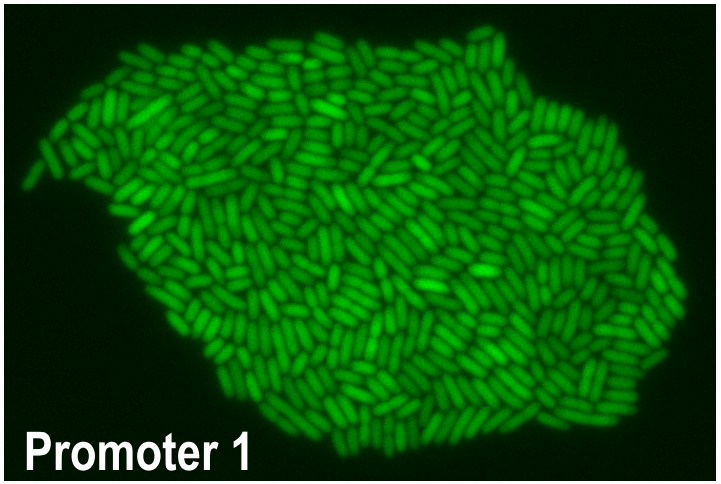


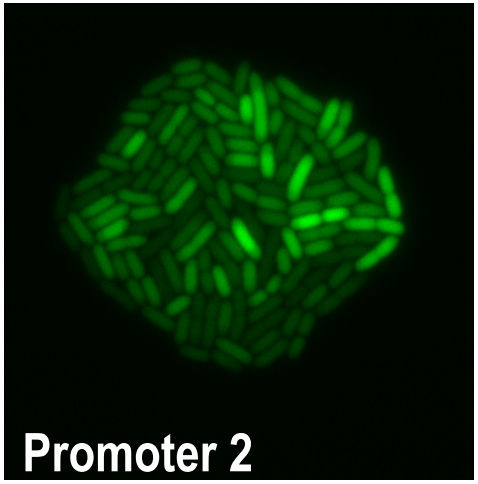
FACS:
Measuring and selecting single cells

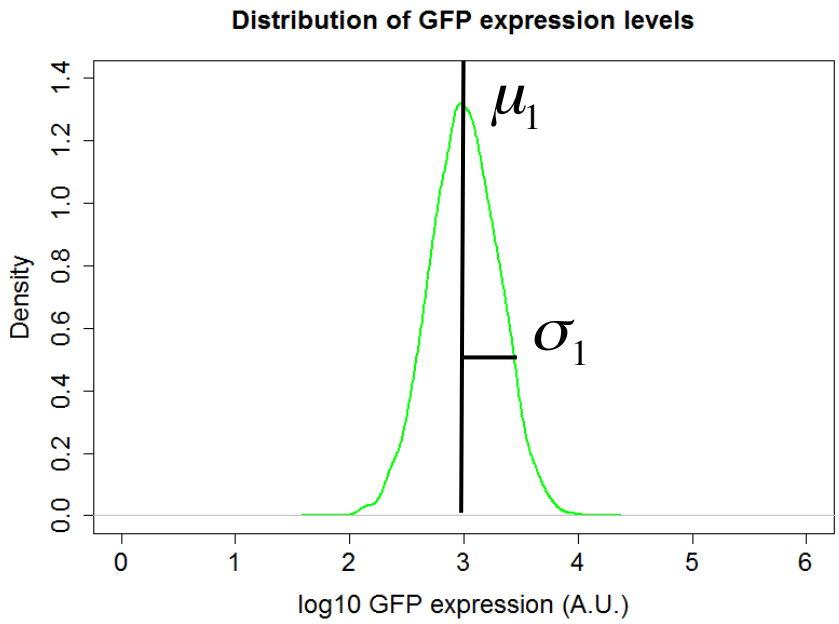


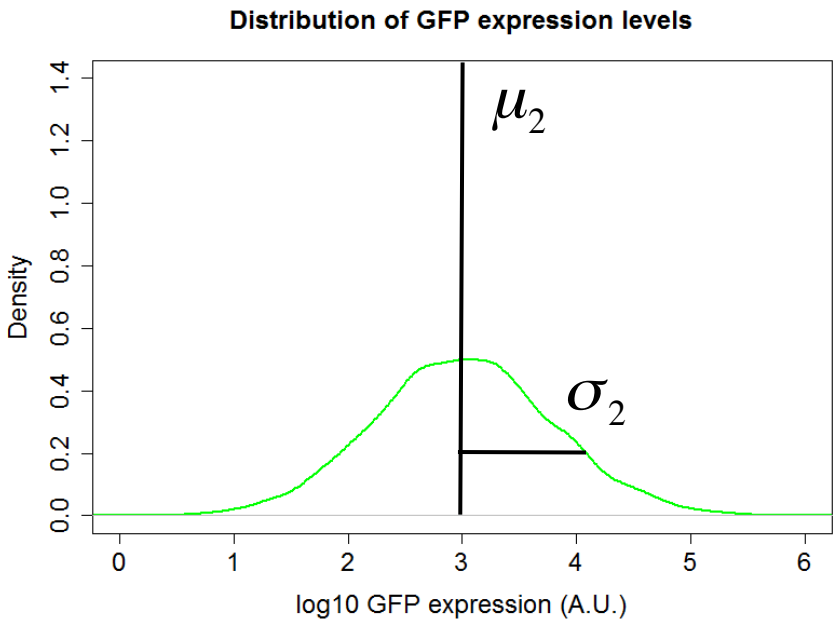
- Cells move one-by-one in a flow channel.
- Each cell passes in front of a laser and its fluorescence is measured.
- By selectively charging particles based on their measured fluorescence, they can be sorted into different subpopulations.
- Important for later: One can set the FACS to only `select' the cells whose fluorescence lies in a certain range.

Gene expression distributions for different promoters



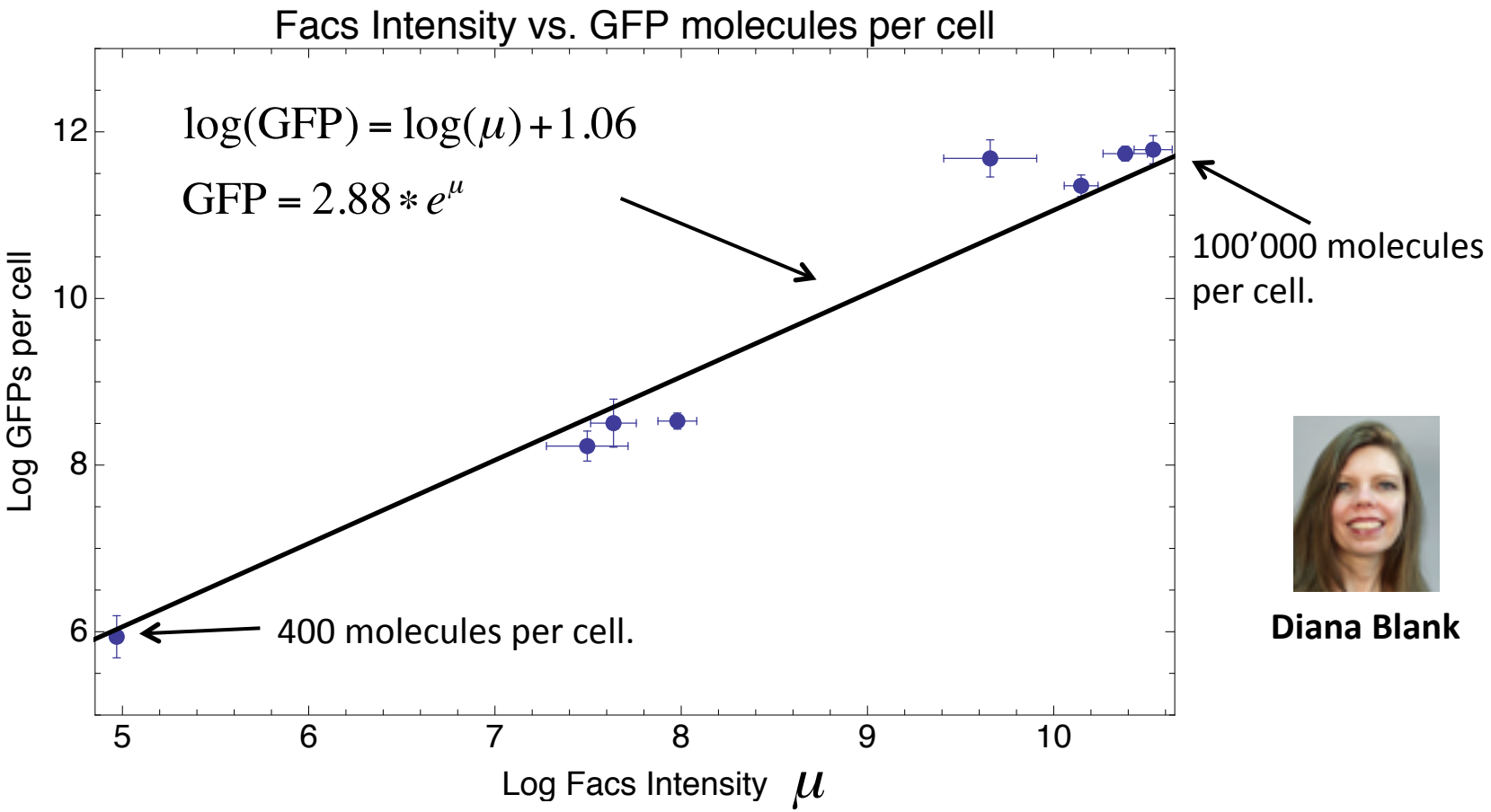






From FACS intensity to GFP molecules per cell

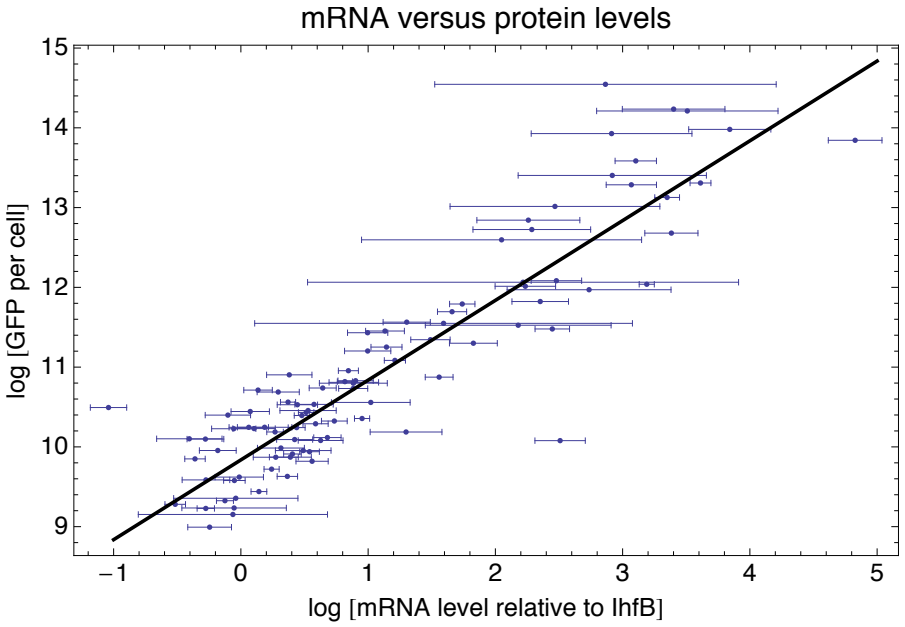
Quantitative Westerns together with reference amounts of GFP (6 replicates).

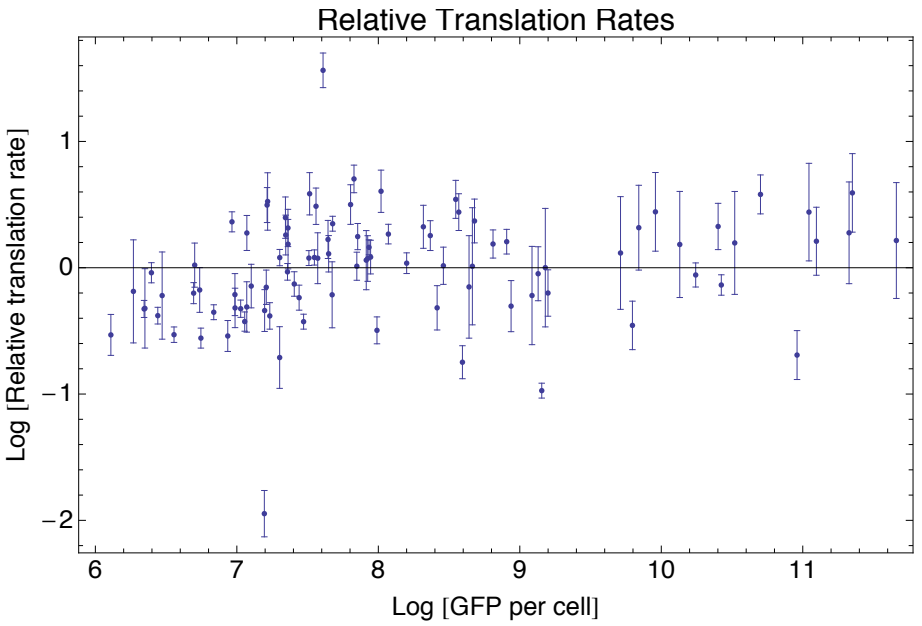


Diana Blank

GFP levels reflect transcription rates

- 96 Strains with different GFP levels.
- qPCR in triplicate using IhfB mRNA as a reference in each.





Mean protein $\langle p \rangle$ and mRNA numbers $\langle m \rangle$ are related through: $\langle p \rangle = \frac{r_{\text{translation}}}{r_{\text{protdecay}}} \langle m \rangle$

GFP is very stable: $r_{\text{protdecay}}$ is dominated by the dilution. Consequently:

$$\log\left[\langle p \rangle / \langle m \rangle\right] = \log[r_{\text{translation}}] + \text{constant}$$

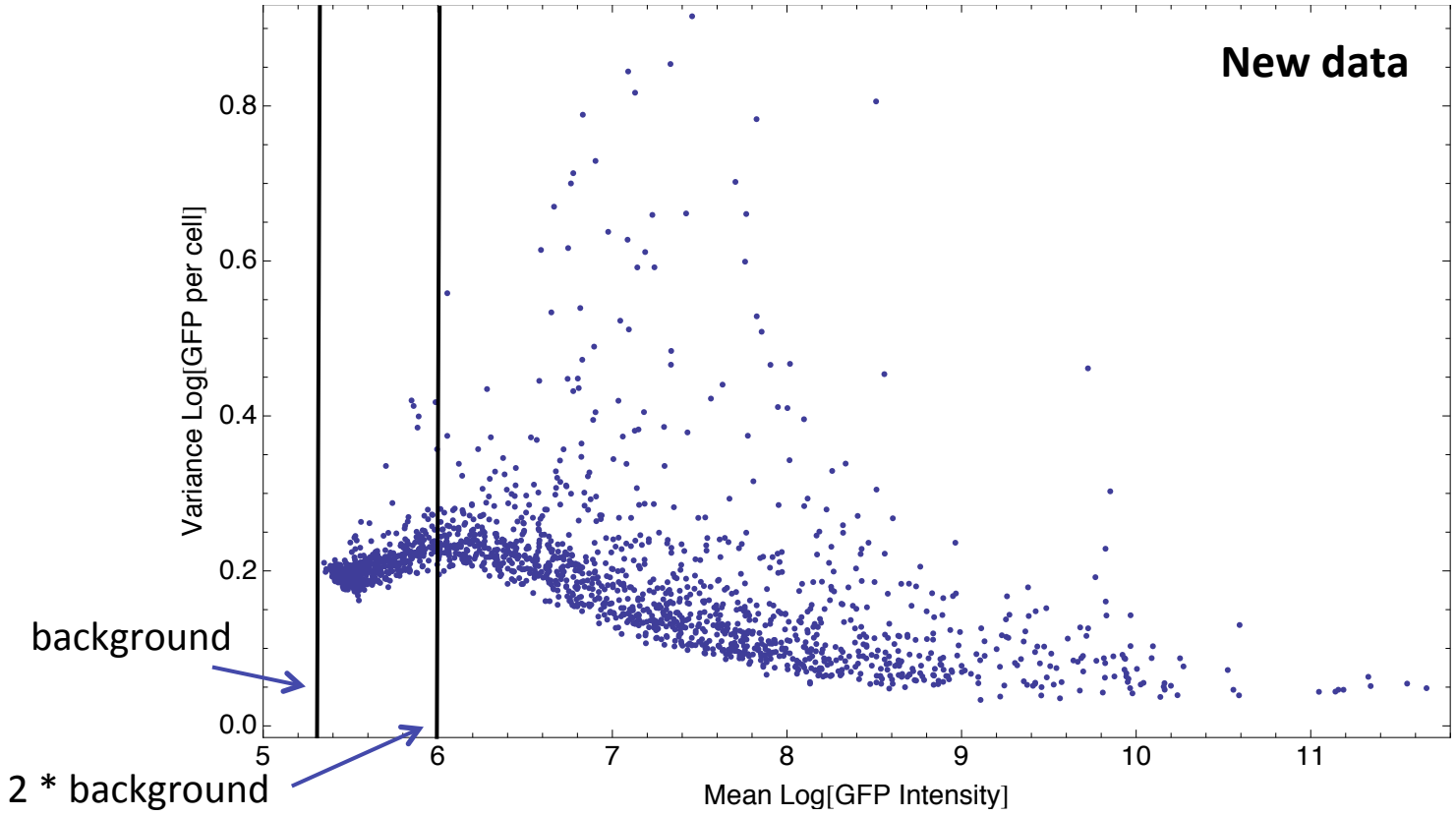
Means and variances of native E. coli promoters

PLoS Genet. 2012 Jan;8(1):e1002443. doi: 10.1371/journal.pgen.1002443. Epub 2012 Jan 19.

A genome-wide analysis of promoter-mediated phenotypic noise in Escherichia coli.

Silander OK, Nikolic N, Zaslaver A, Bren A, Kikoin I, Alon U, Ackermann M.

Computational and Systems Biology, Biozentrum, University of Basel, Basel, Switzerland. olinsilander@gmail.com



Characteristic lower bound on variance as a function of mean expression.

Noise model

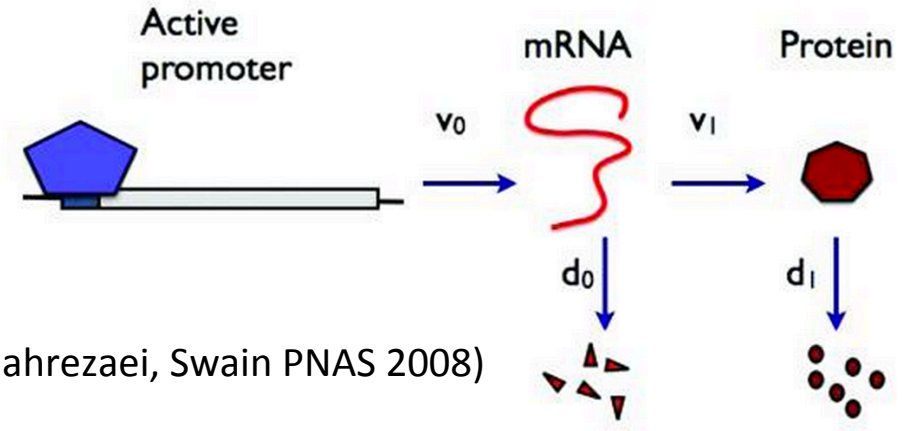
The simplest model of gene expression assumes *constant rates*:

v_0 transcription

d_0 mRNA decay

v_1 translation

d_1 protein decay



(Shahrezaei, Swain PNAS 2008)

The ratios: $a = \frac{v_0}{d_1}$ $b = \frac{v_1}{d_0}$ "burst size"

determine mean and variance: $\langle n \rangle = ab, \quad \mathrm{var}(n) = (b+1)\langle n \rangle$

Additionally, the 4 rates *fluctuate* from cell to cell. Assuming the product *ab* is log-normally distributed, the total variance becomes:

$$\sigma_n^2 = \sigma_{ab}^2 \langle n \rangle^2 + (b+1)\langle n \rangle$$

Note: for large means the first term will dominate.

Excess variance

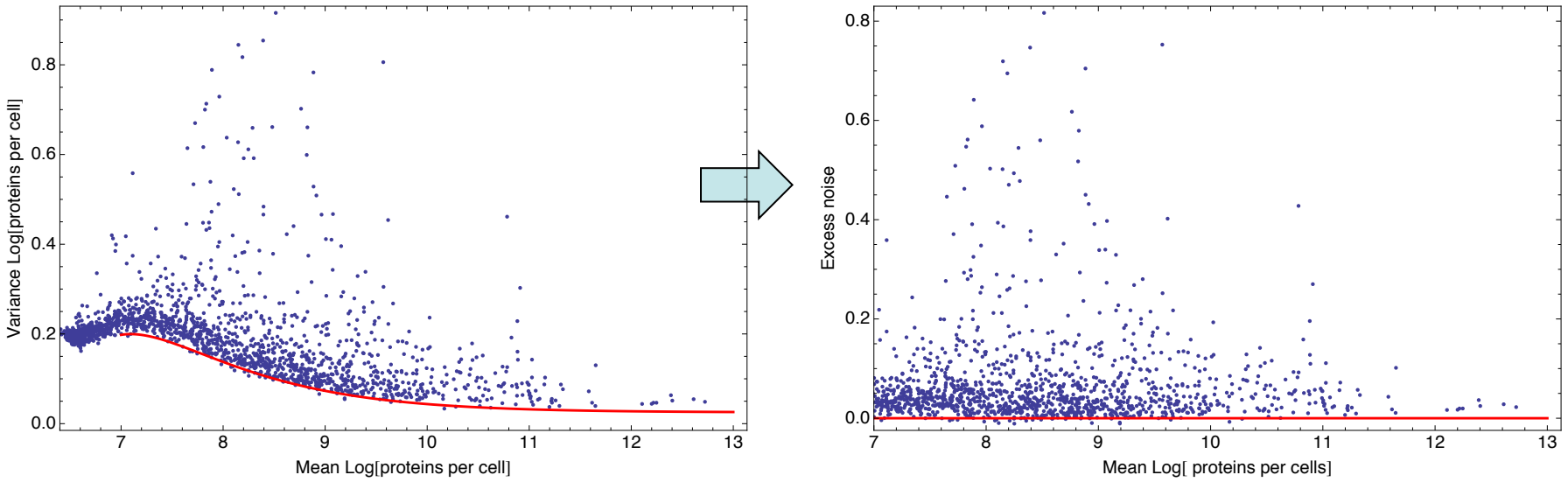
Total GFP signal: $g = g_{bg} + \langle n \rangle + \varepsilon\sigma_n^2$ with ε a fluctuating variable with mean 0 and variance 1.

If the mean $\langle g \rangle = g_{bg} + \langle n \rangle$ is large relative to the fluctuations, we have in *log-scale*:

$$\mathrm{var}\left[\log(g)\right] = \sigma_{ab}^2 \left(1 - \frac{g_{bg}}{\langle g \rangle}\right)^2 + \frac{(b+1)}{\langle g \rangle}\left(1 - \frac{g_{bg}}{\langle g \rangle}\right)$$

Red curve: $\sigma_{ab}^2 = 0.025, \quad b = 450$

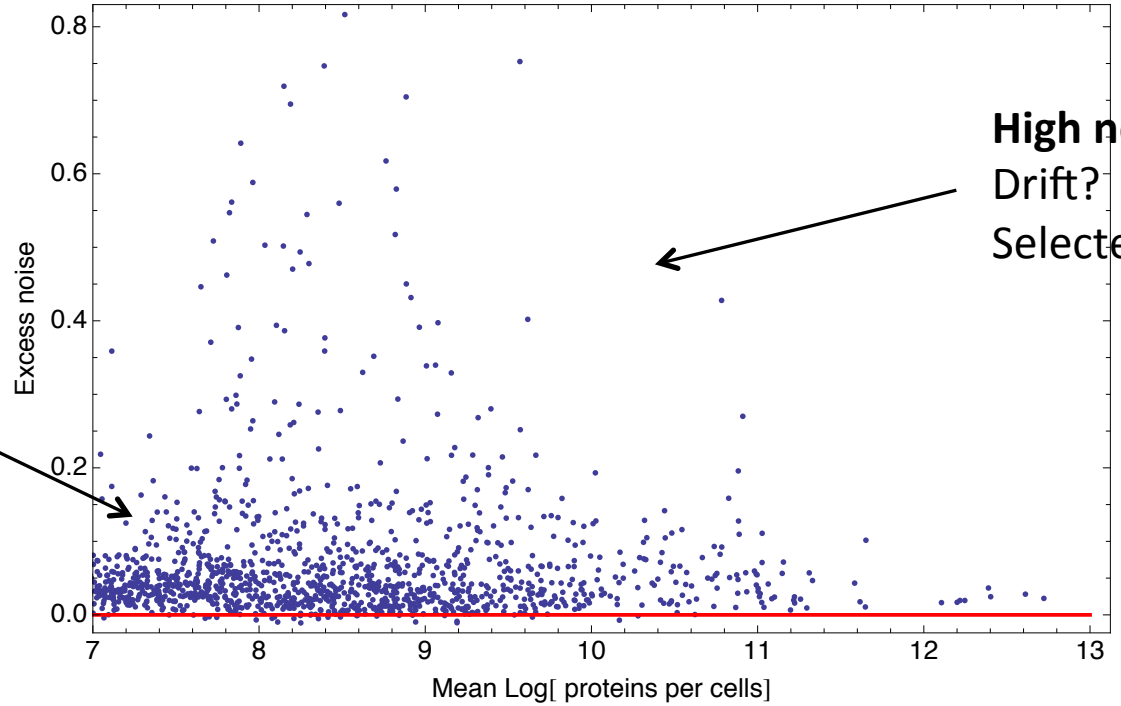
Excess variance (red curve subtracted)



By subtracting out the mean-dependent minimal noise, we can compare the *excess noise* levels of promoters with different means.

Why do different promoters have different noise levels?

Low noise.
Selection toward an intrinsic lower bound?

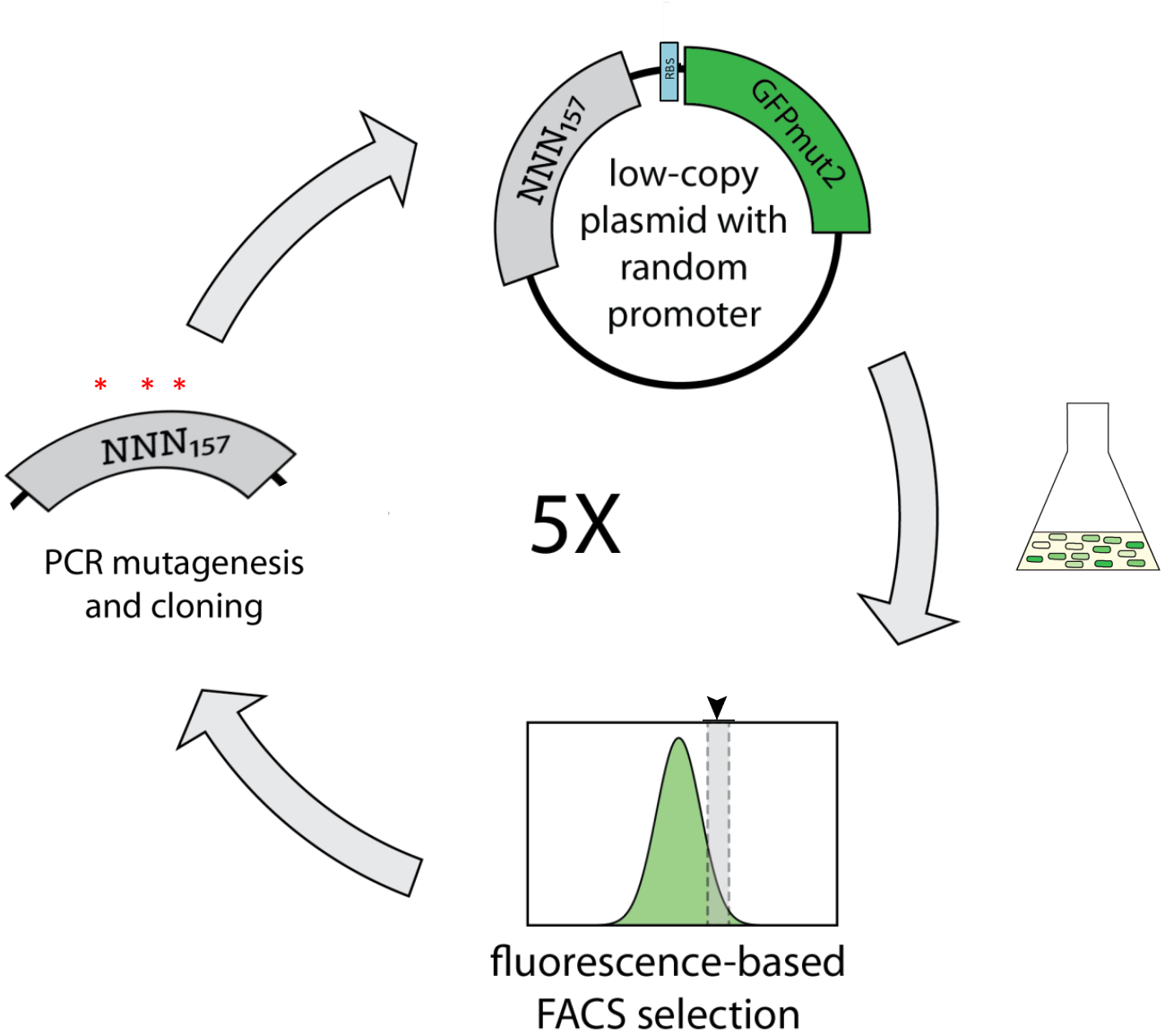


High noise
Drift?
Selected for noise?

- Assuming expression means are tuned to optimal levels, fluctuations would per definition be detrimental.
- This would imply selection acts to minimize noise.
- Indeed, low noise promoters are enriched for essential and well-conserved genes.

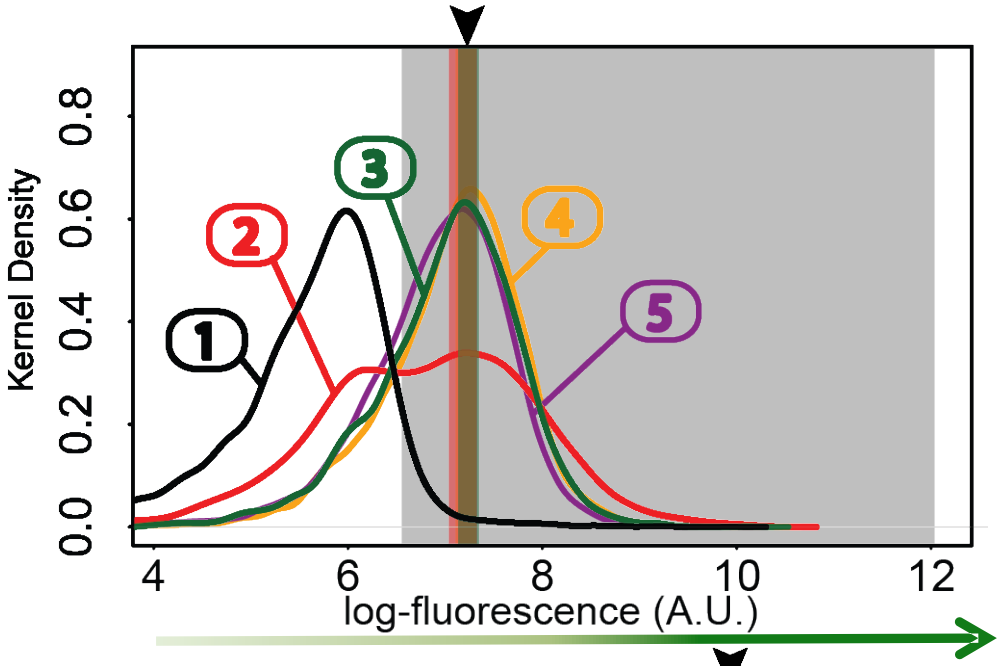
- **To assess the role of selection, we will compare the behavior of native promoters, with *synthetic promoters*, evolved in a precisely controlled selective environment.**

Directed evolution of promoters that express at a desired level

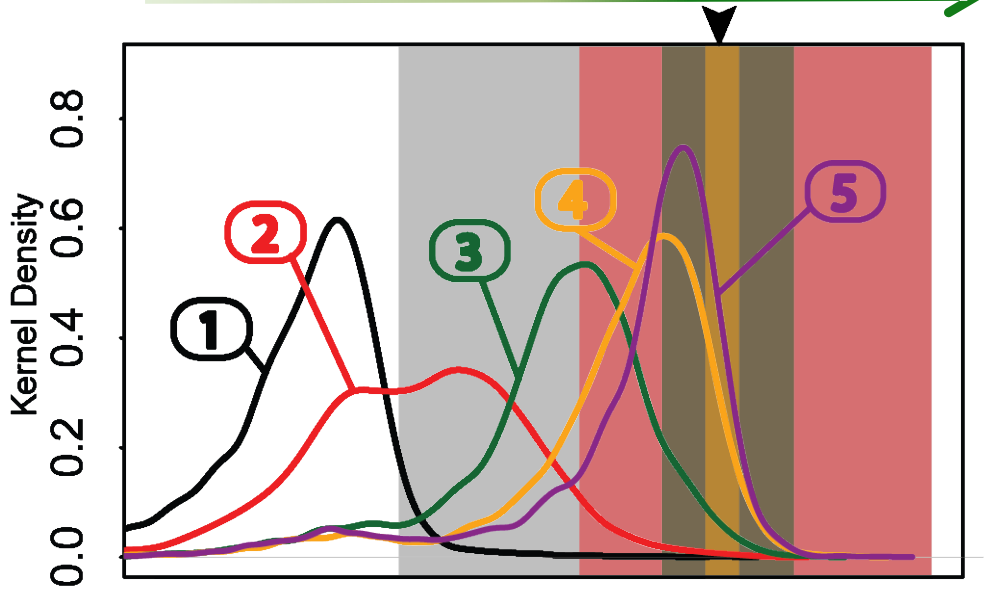


Evolution of population expression levels

Selecting for
Medium expression

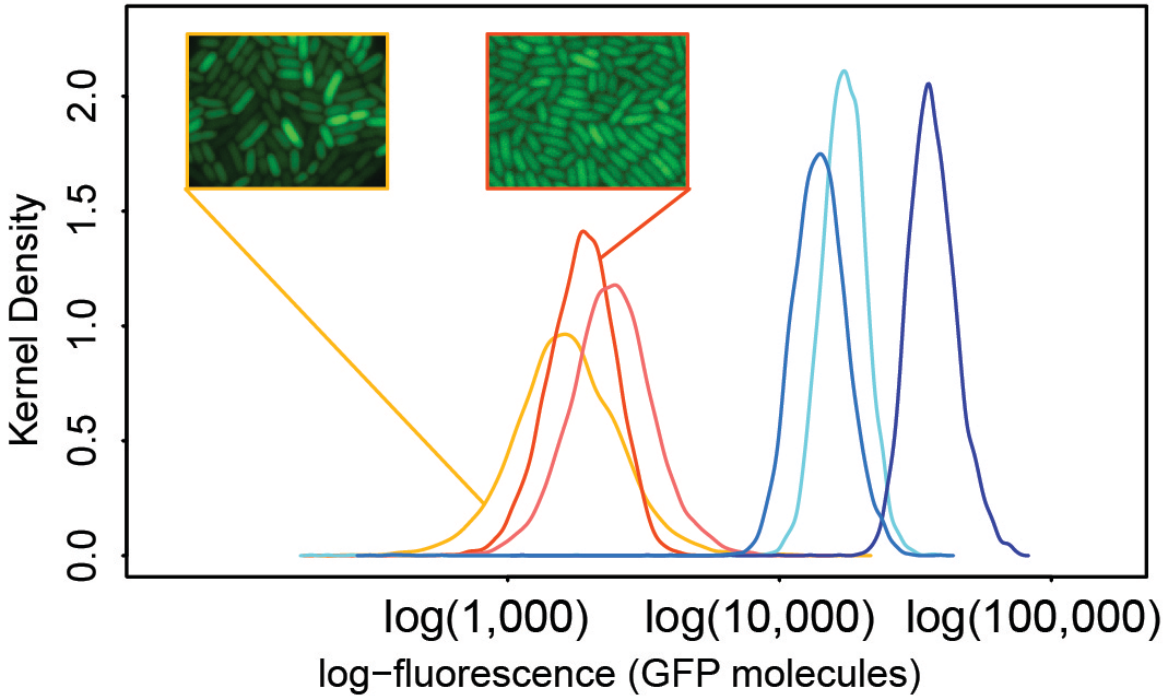


Selecting for
High expression



Expression distributions of individual promoter clones

- We isolated 378 clones from the 3rd and 5th generation and sequenced their promoters.
- The sequences were highly diverse.
- We measured expression distributions for all 378 sequenced promoters.

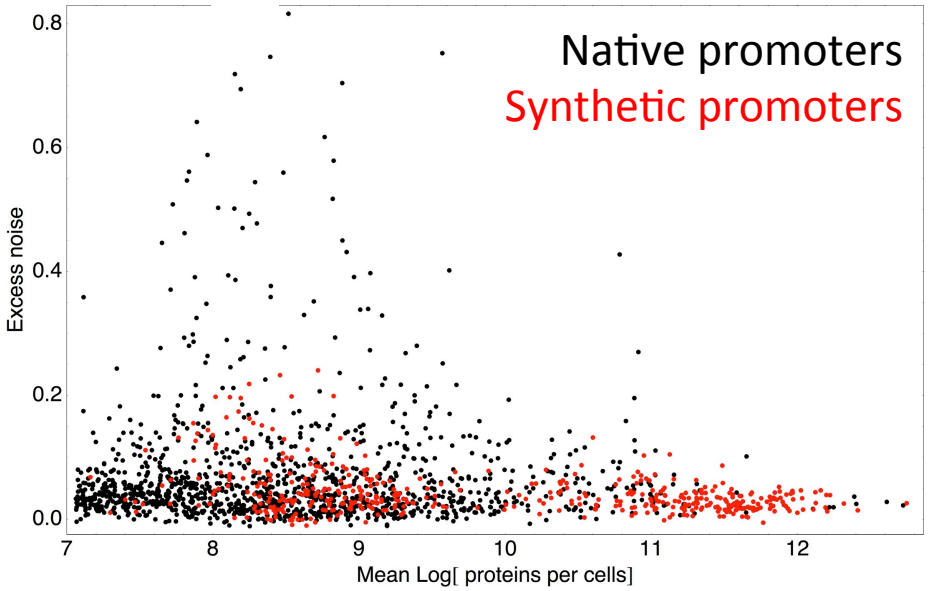


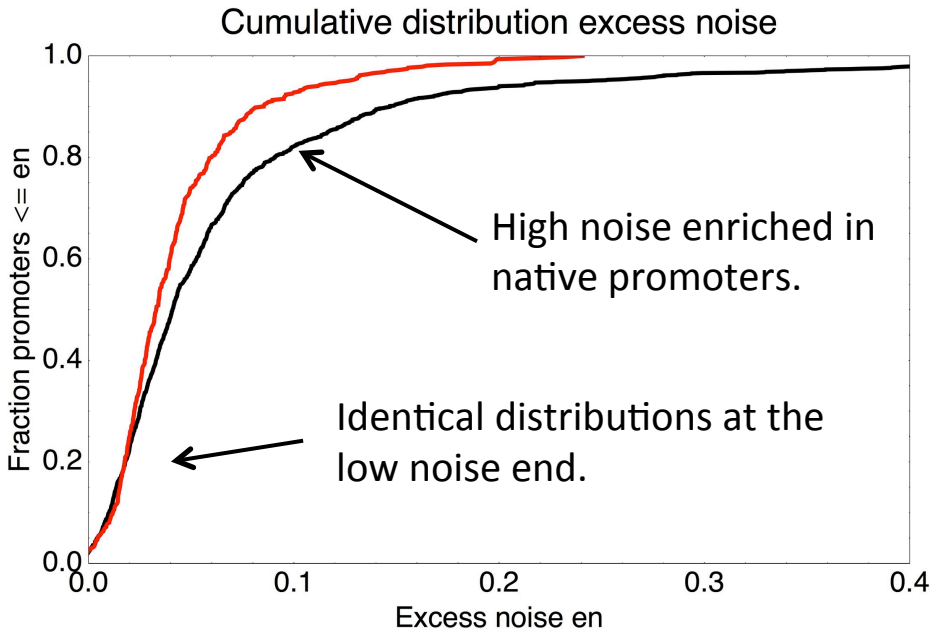
How do the noise levels of the synthetic promoters compare with those of the native promoters?

BIOZENTRUM
Universität Basel
The Center for Molecular Life Sciences

Native promoters have higher excess variance

SIB
Swiss Institute of Bioinformatics





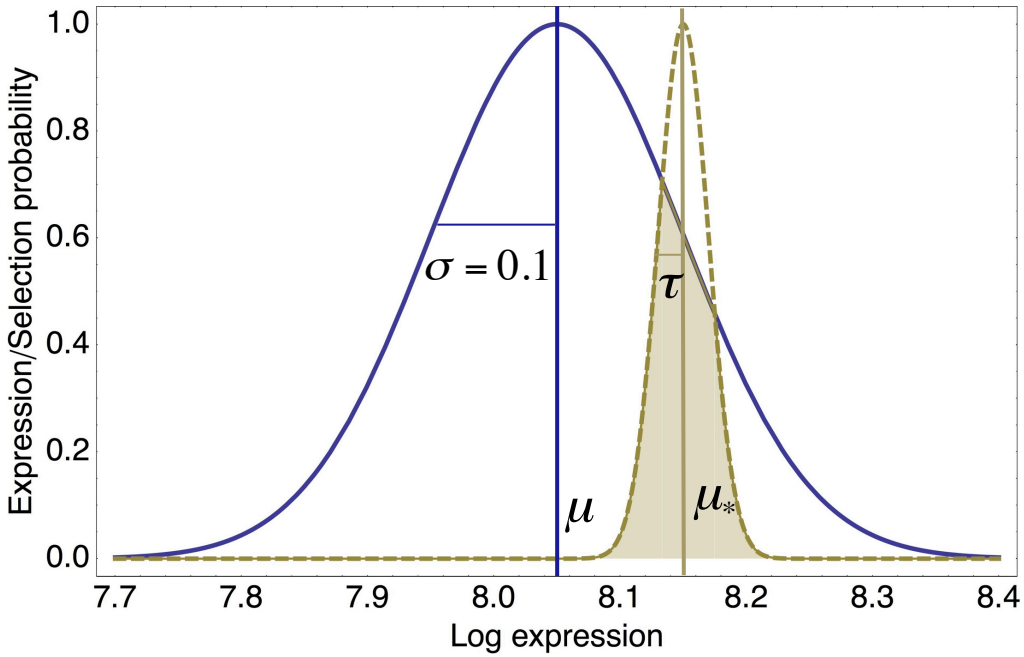
Possible interpretation:

Maybe our scheme strongly and efficiently selects against noise in promoters.

⇩

We need to quantify selection for/against noise in our system.

The FACS fitness function



Fitness (probability to be selected):

$$f(x \mid \mu_*, \tau) = \exp\left(-\frac{(x-\mu_*)^2}{2\tau^2}\right)$$

Distribution of expression levels:

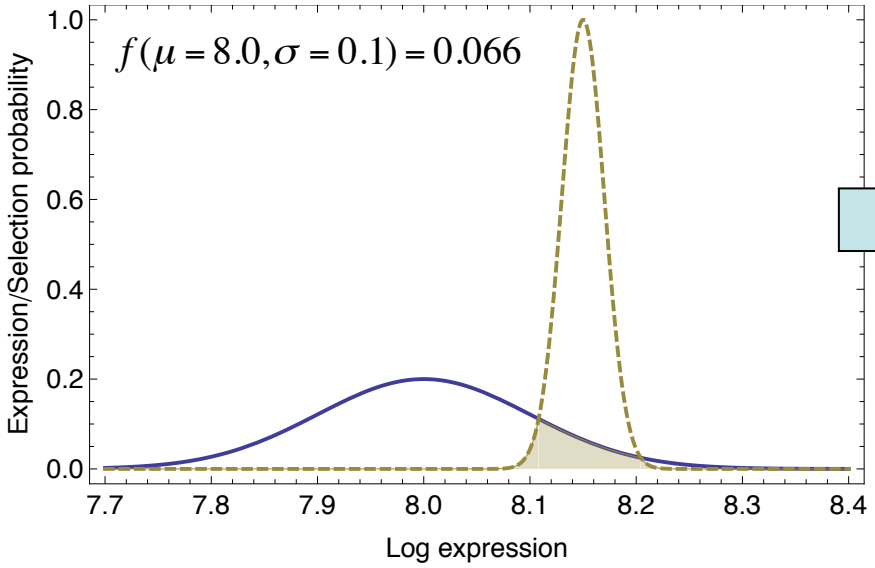
$$p(x \mid \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

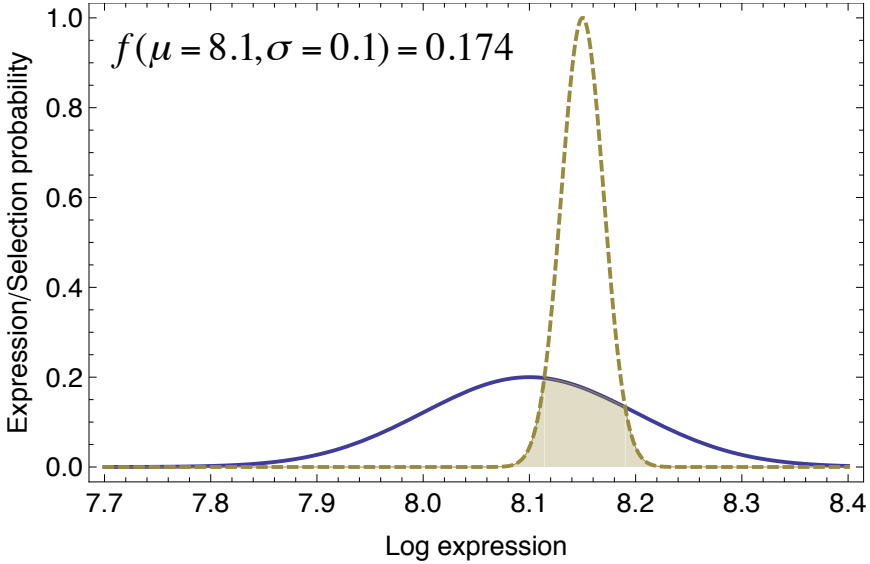
The fitness of a promoter `genotype' (fraction of its cells selected) as a function of its mean and variance:

$$f(\mu, \sigma \mid \mu_*, \tau) = \int dx p(x \mid \mu, \sigma) f(x \mid \mu_*, \tau) = \sqrt{\frac{\tau^2}{\tau^2+\sigma^2}}\exp\left(-\frac{(\mu-\mu_*)^2}{2(\tau^2+\sigma^2)}\right)$$

Moving the mean toward the desired level always increases fitness

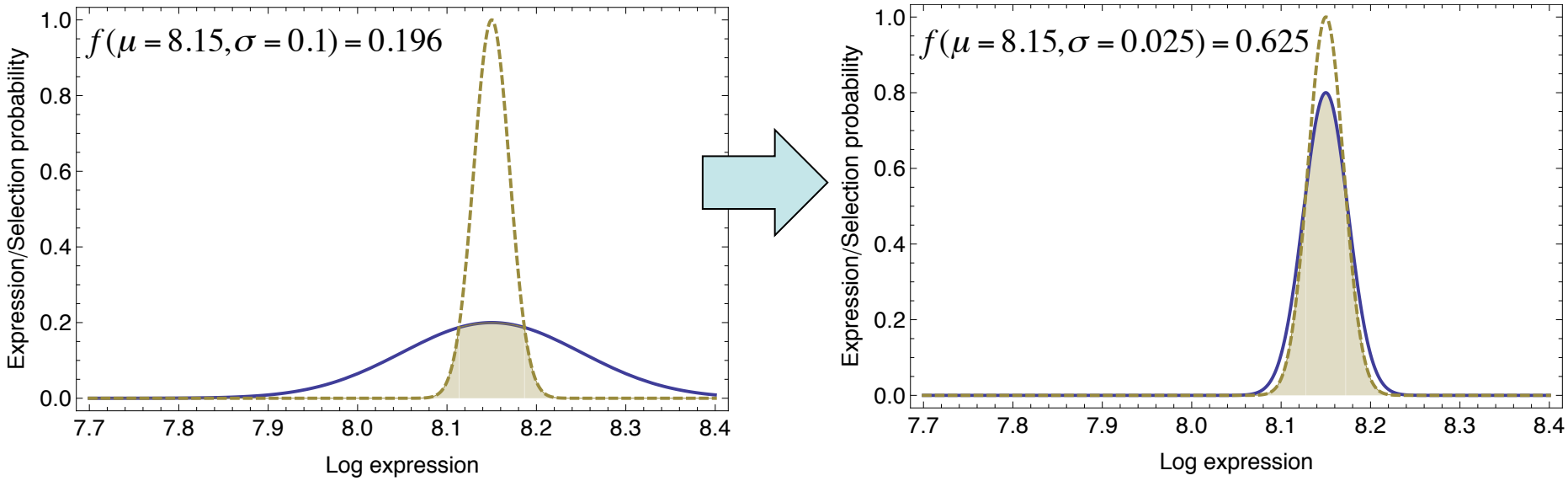
$$f(\mu,\sigma \mid \mu_*,\tau) = \sqrt{\frac{\tau^2}{\tau^2+\sigma^2}}\exp\left(-\frac{(\mu-\mu_*)^2}{2(\tau^2+\sigma^2)}\right)$$



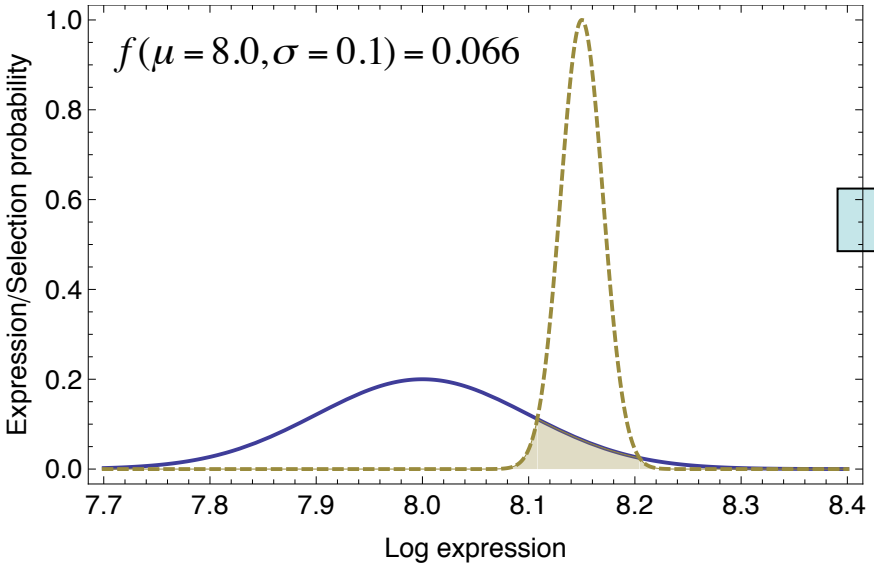


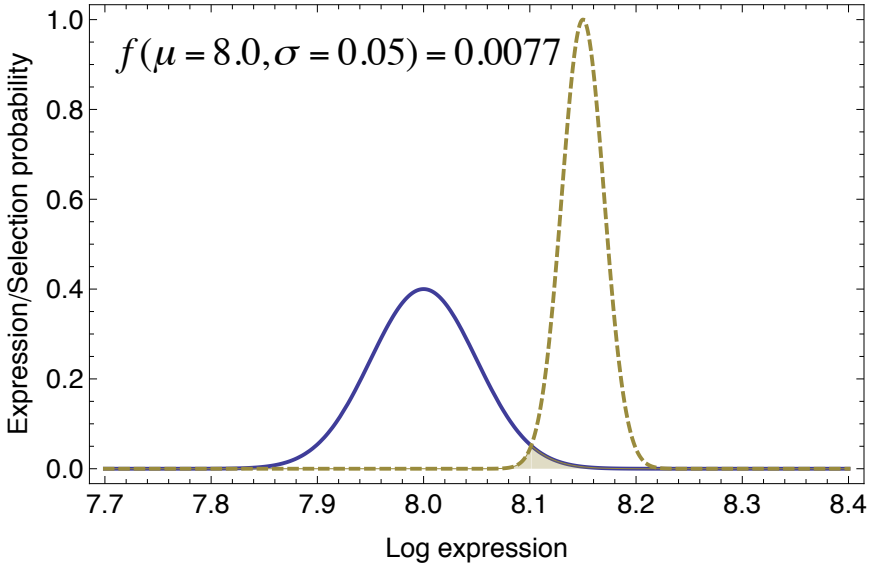
At optimal mean minimal noise is preferred

$$f(\mu,\sigma \mid \mu_*,\tau) = \sqrt{\frac{\tau^2}{\tau^2+\sigma^2}}\exp\left(-\frac{(\mu-\mu_*)^2}{2(\tau^2+\sigma^2)}\right)$$



As mean moves away from the optimum there is a bifurcation to nonzero noise



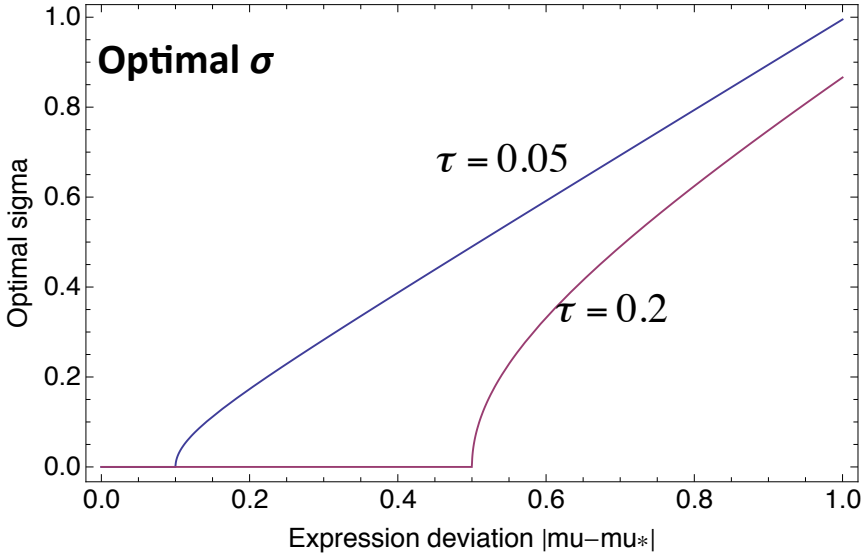


$$f(\mu, \sigma \mid \mu_*, \tau) = \sqrt{\frac{\tau^2}{\tau^2 + \sigma^2}} \exp\left(-\frac{(\mu - \mu_*)^2}{2(\tau^2 + \sigma^2)}\right)$$

`Bifurcation' in optimal σ

When $|\mu - \mu_*| \geq \tau$, the optimal noise level is non-zero:

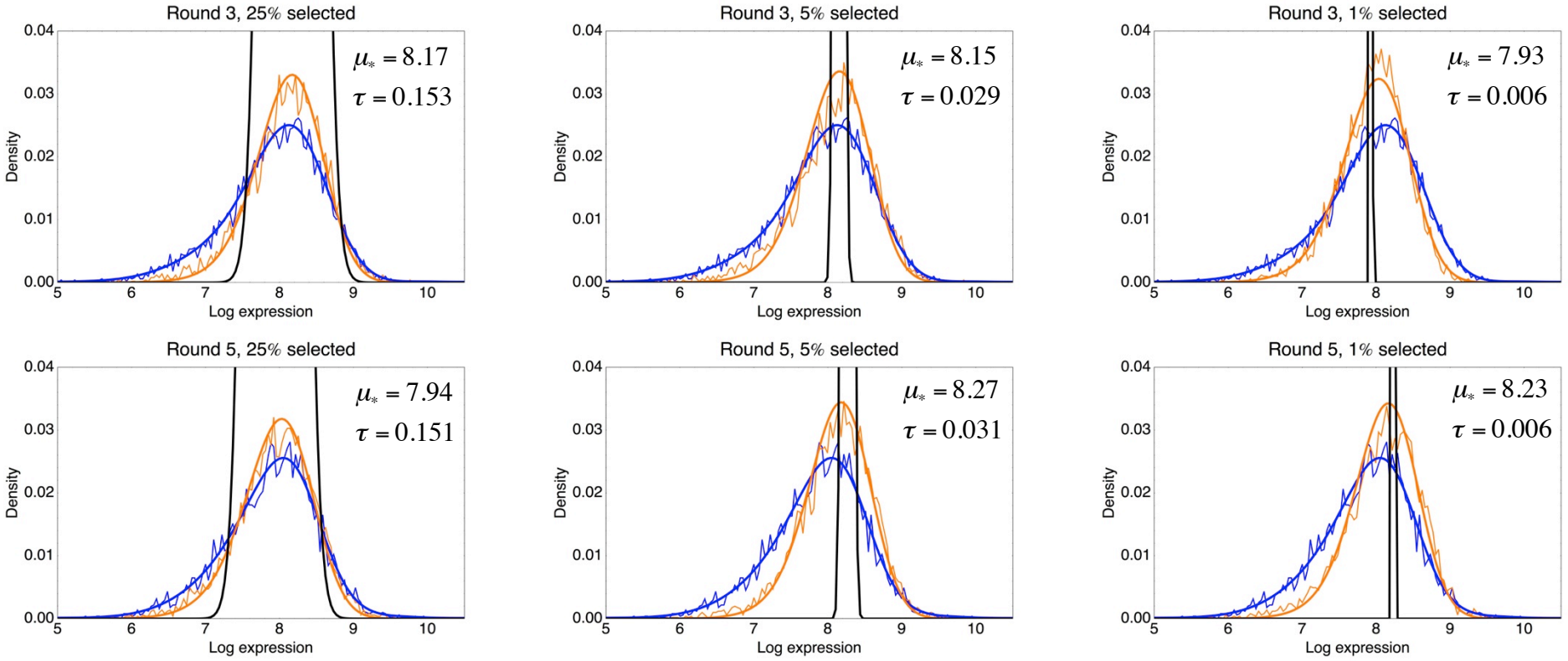
$$\sigma_* = \sqrt{(\mu - \mu_*)^2 - \tau^2}$$



BIOZENTRUM
Universität Basel
The Center for Molecular Life Sciences

Inferring the parameters of our FACS fitness function

SIB
Swiss Institute of Bioinformatics



Population expression distribution *before* selection.

Population expression distribution *after* selection.

Inferred selection window.

Note

The inferred width τ is proportional to fraction of the population that was selected.

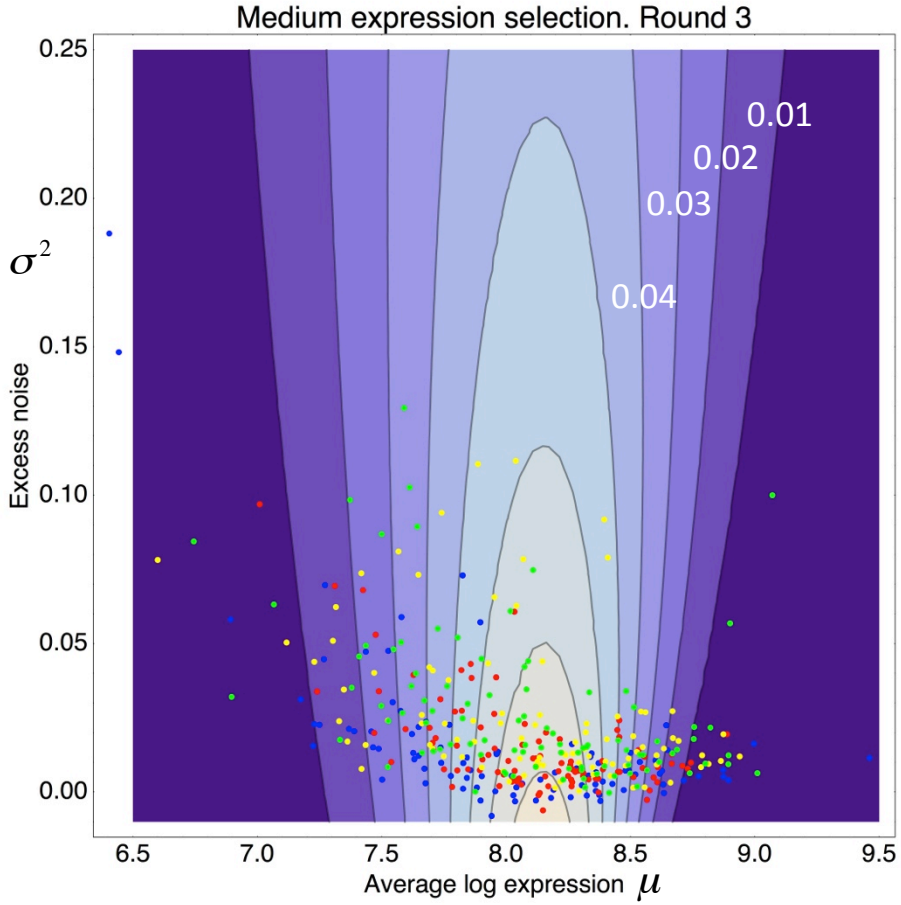
The desired mean fluctuates a bit in replicates.

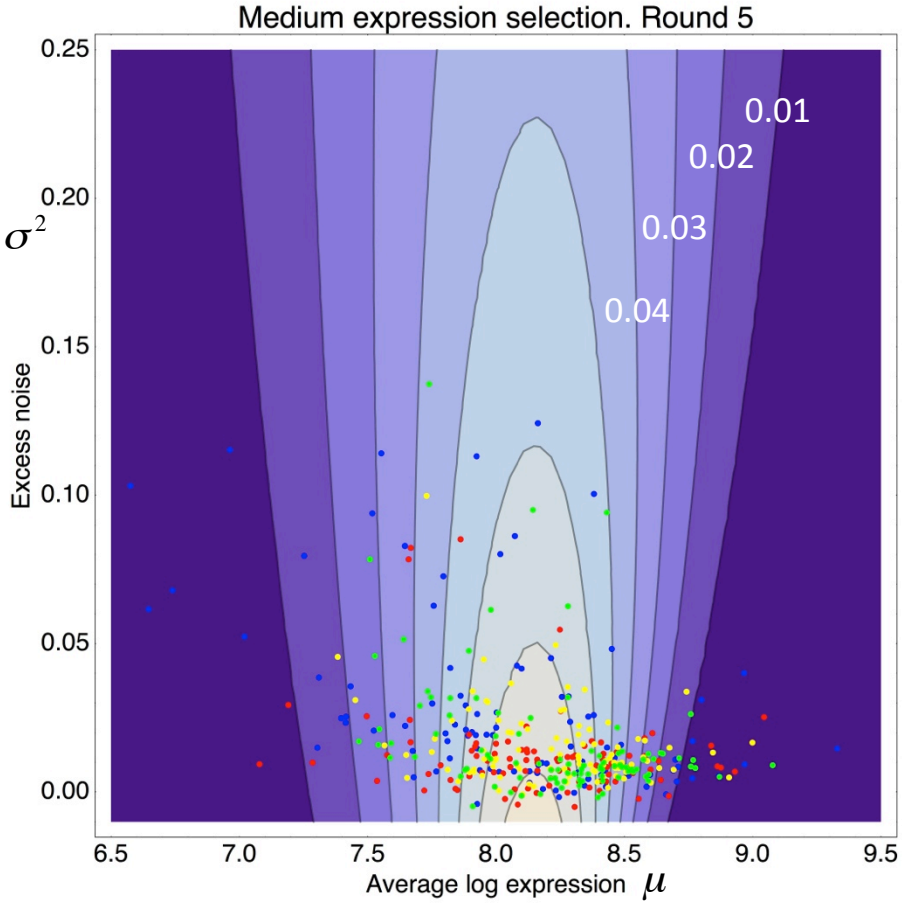
BIOZENTRUM
Universität Basel
The Center for Molecular Life Sciences

FACS selection acts predominantly on mean and not variance

SIB
Swiss Institute of Bioinformatics

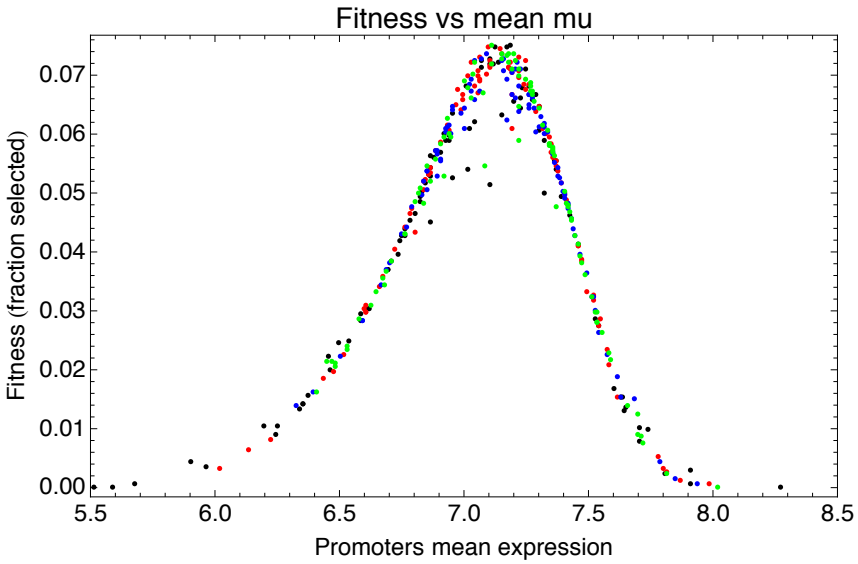
$\mu_* = 8.11, \quad \tau = 0.03$

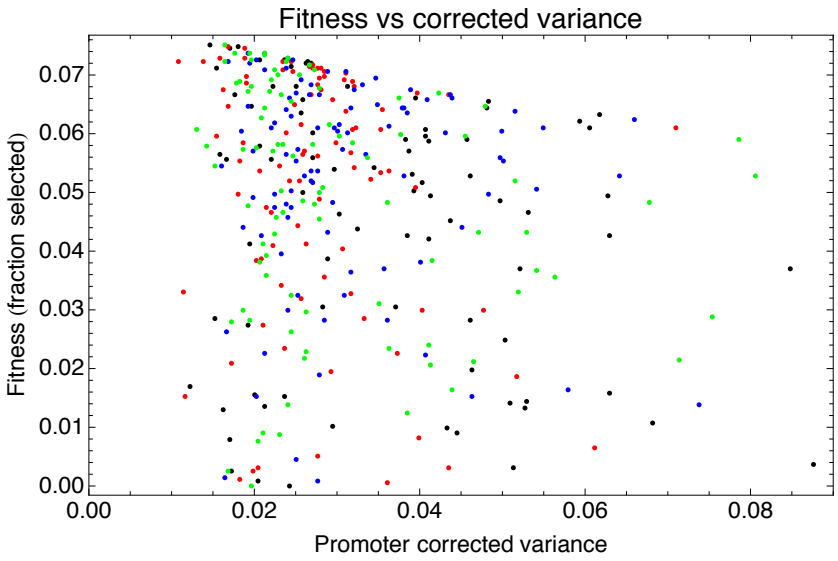




- = Promoters from the evolutionary lineage (left: round 3, right: round 5)
- = Promoters after an additional round of stringent selection (1%).
- = Promoters after an additional round of `normal' selection (5%).
- = Promoters after an additional round of `normal' selection (25%).

For our synthetic promoters fitness is a function of mean expression not noise





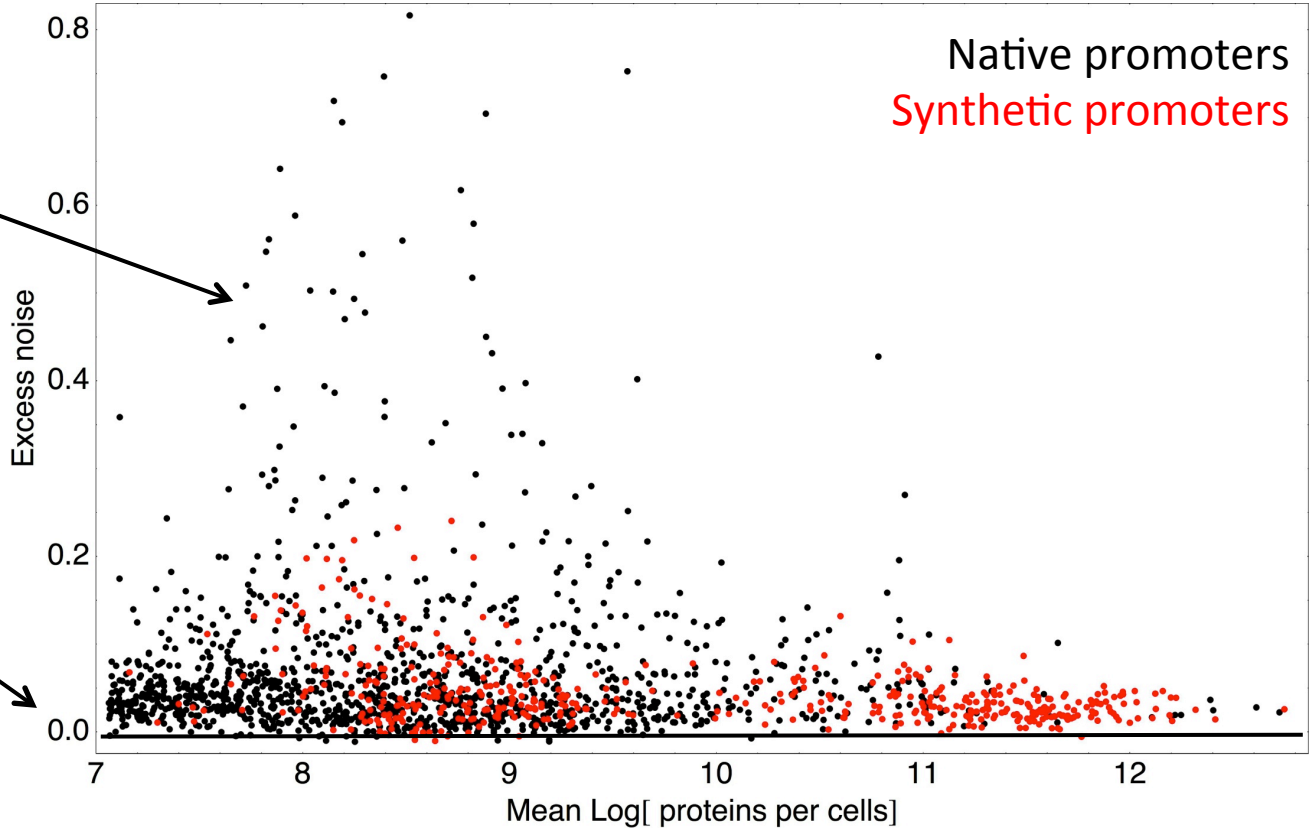
- Mean expression level of a promoter is highly predictive of its fitness.
- Almost no correlation between mean-corrected variance and fitness.

Low noise is the *default* state of a promoter selected for a given expression level

Selection caused increased noise in a substantial fraction native promoters

High noise driven by selection

Intrinsic lower bound $\sigma_{\min}$

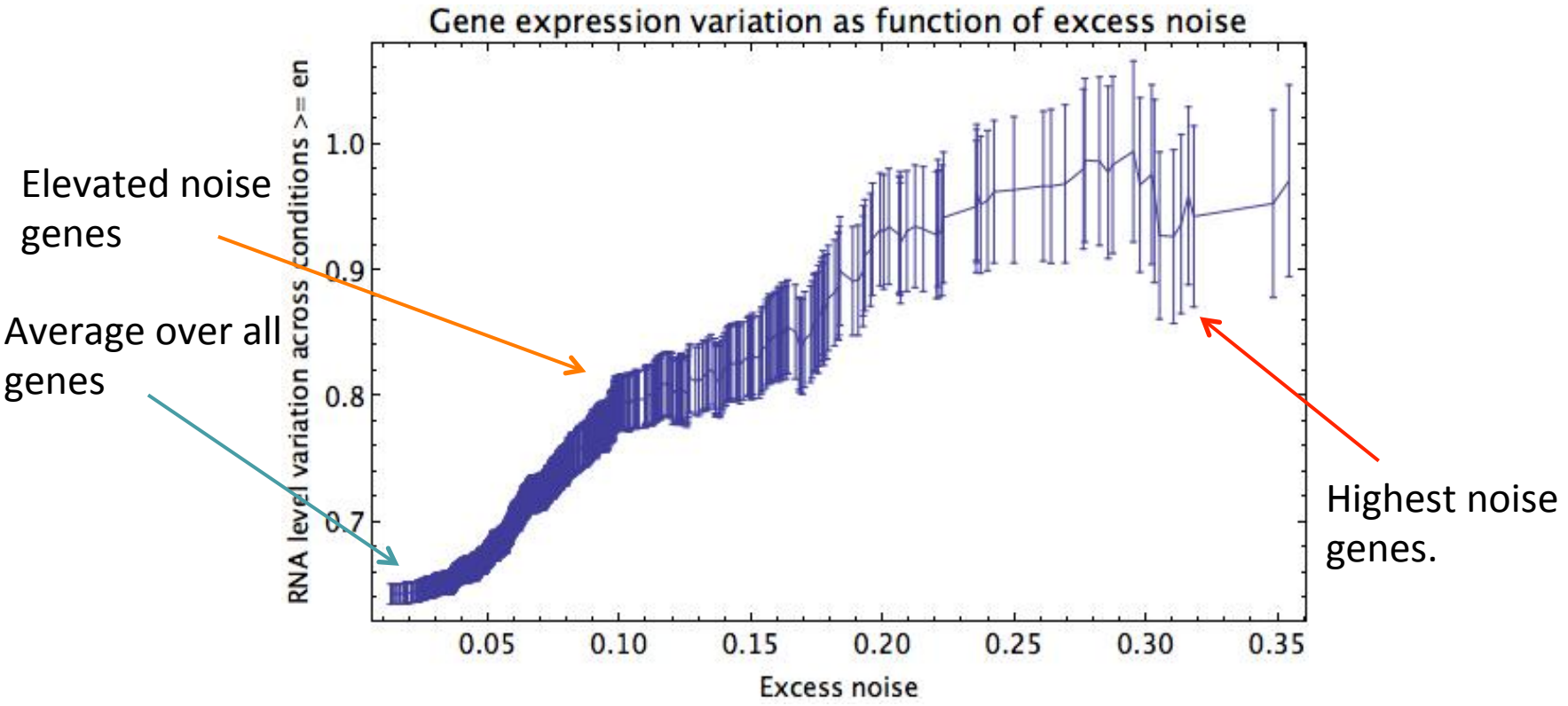


What is `special' about native promoters that show high noise?

Noisy genes have more expression variation across environments

GenExpDB revealing the essence of life
http://genexpdb.ou.edu

Fold-changes in average mRNA levels of *E. coli* genes across 240 different conditions.

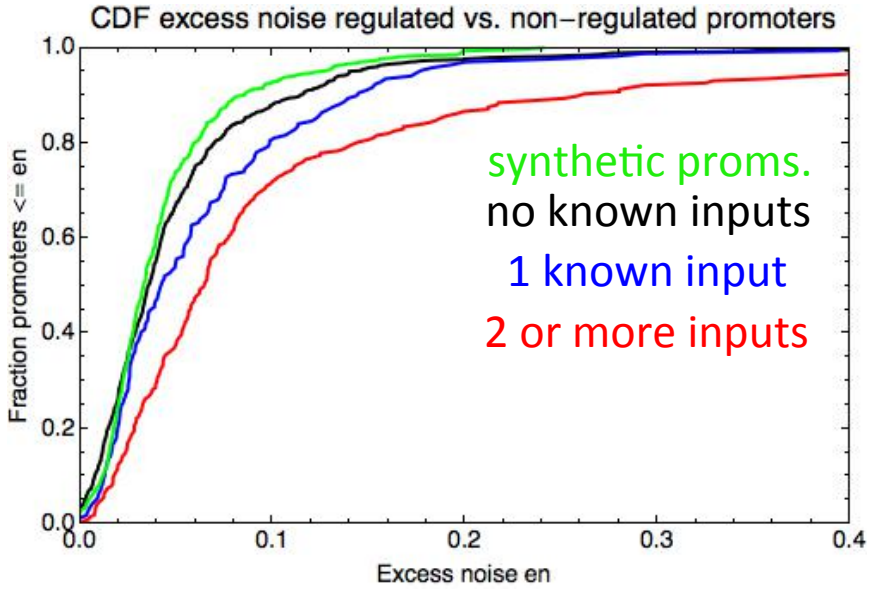


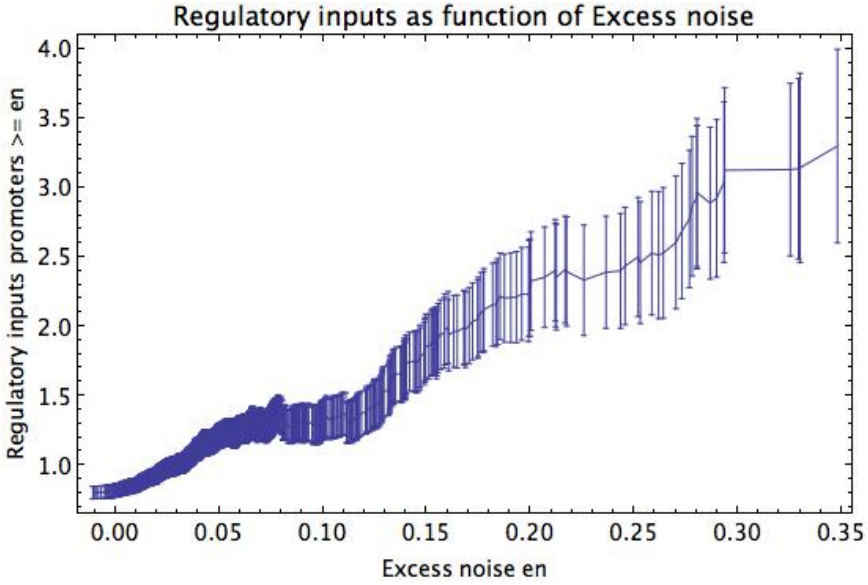
Higher noise in our condition correlates with increased variation in mean *across* conditions.
(This has been observed previously in yeast).

Noisy genes have more regulatory inputs

RegulonDB
Escherichia coli K12 Transcriptional Network

- 185 *E. coli* transcription factors (TFs).
- 4123 known regulatory interactions TF → promoter.





Genes with known regulatory inputs tend to have higher noise.

Genes with higher noise have (on average) higher numbers of known regulatory inputs.

Why is there a general association between noise and regulation?

Why is there a general association between noise and regulation?

Noise as an unavoidable side effect of regulation

- Any regulator has itself some unavoidable noise: This noise is *transmitted* to its regulators.

  Proc Natl Acad Sci U S A. 2001 Jul 17;98(15):8614-9. Epub 2001 Jul 3.

  Intrinsic noise in gene regulatory networks.

  Thattai M[1], van Oudenaarden A.

- Genes that need complex regulation unavoidably couple to the noise in their regulators.

  Cell Mol Life Sci. 2011 Mar;68(6):1005-10. doi: 10.1007/s00018-010-0589-y. Epub 2010 Nov 30.

  Fluctuation and response in biology.

  Lehner B[1], Kaneko K.

Expression noise as a bet hedging strategy

- Phenotypic diversity can generally be selected for in fluctuating environments.

Evolution of Phenotypic Variance
J. J. Bull
Evolution, Vol. 41, No. 2 (Mar., 1987), pp. 303-315

Science. 2005 Sep 23;309(5743):2075-8. Epub 2005 Aug 25.

Phenotypic diversity, population growth, and information in fluctuating environments.

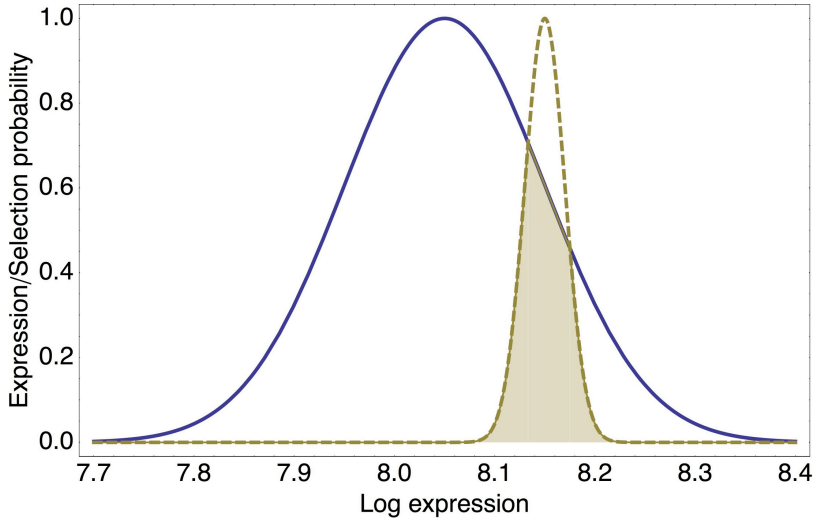
Kussell E[1], Leibler S.

Can't it be both?

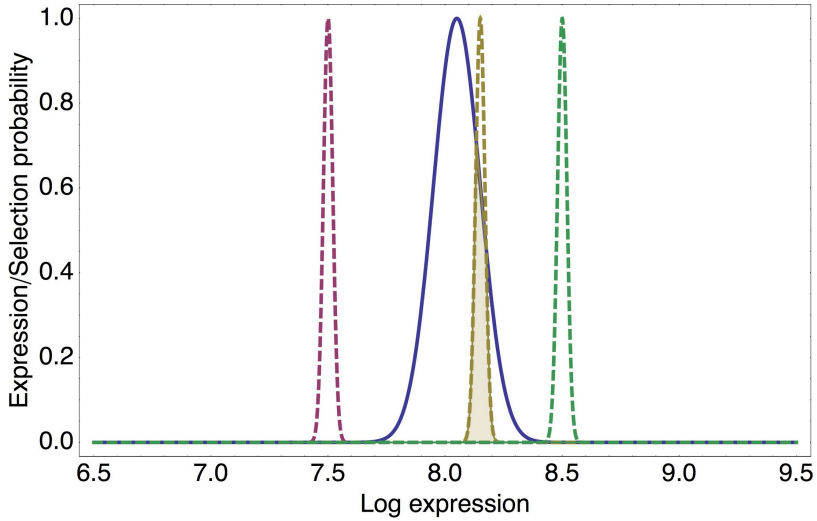
Fitness in a variable environment

In its 'natural habitat', the organisms experience different environments *e*.
Assume that in each environment *e* there is an optimal expression level μ_e.

One environment



Different environments



In each environment fitness is assumed Gaussian:

$$f(\mu,\sigma \mid \mu_e,\tau) = \sqrt{\frac{\tau^2}{\tau^2+\sigma^2}}\exp\left(-\frac{(\mu-\mu_e)^2}{2(\tau^2+\sigma^2)}\right)$$

Overall log-fitness is average of log-fitness across environments:

$$\log\left[f(\mu,\sigma)\right] = -\frac{\left\langle(\mu-\mu_e)^2\right\rangle}{2(\tau^2+\sigma^2)} + \frac{1}{2}\log\left[\frac{\tau^2}{\tau^2+\sigma^2}\right]$$

Fitness of an unregulated gene

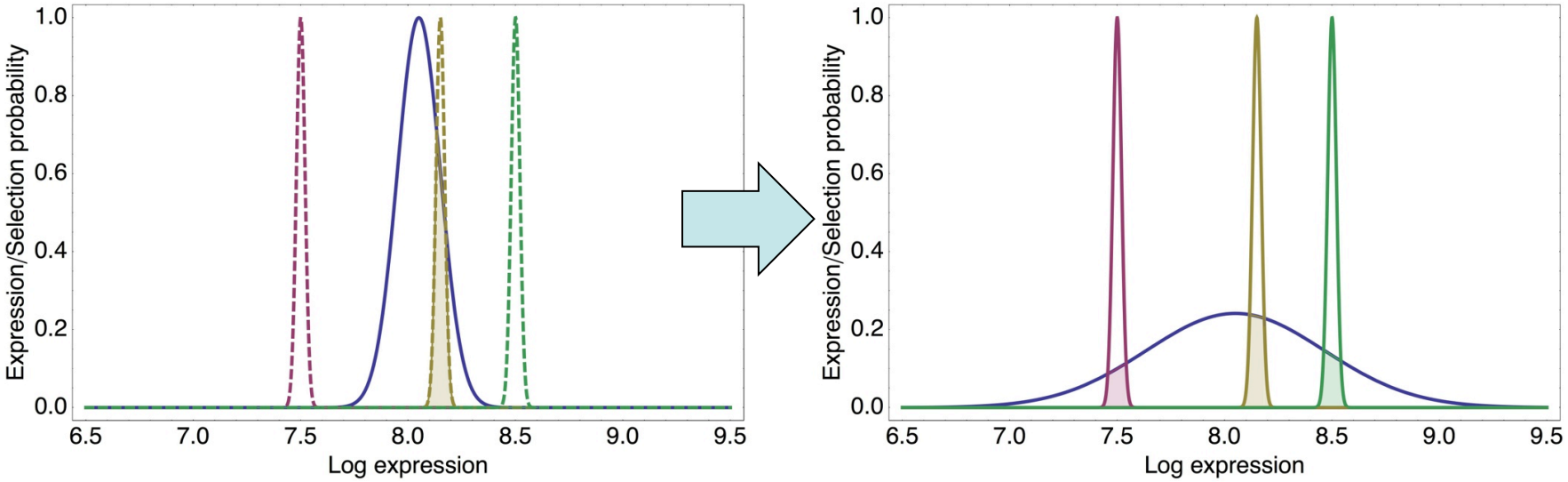
Log-fitness in a variable environment: $\log\left[f(\mu,\sigma)\right] = -\frac{\left\langle(\mu-\mu_e)^2\right\rangle}{2(\tau^2+\sigma^2)} + \frac{1}{2}\log\left[\frac{\tau^2}{\tau^2+\sigma^2}\right]$

Assuming no regulation (constant mean), optimal mean equals $\mu = \langle\mu_e\rangle$

Log-fitness becomes:

$$\log\left[f(\mu,\sigma)\right] = -\frac{\mathrm{var}(\mu_e)}{2(\tau^2+\sigma^2)} + \frac{1}{2}\log\left[\frac{\tau^2}{\tau^2+\sigma^2}\right]$$

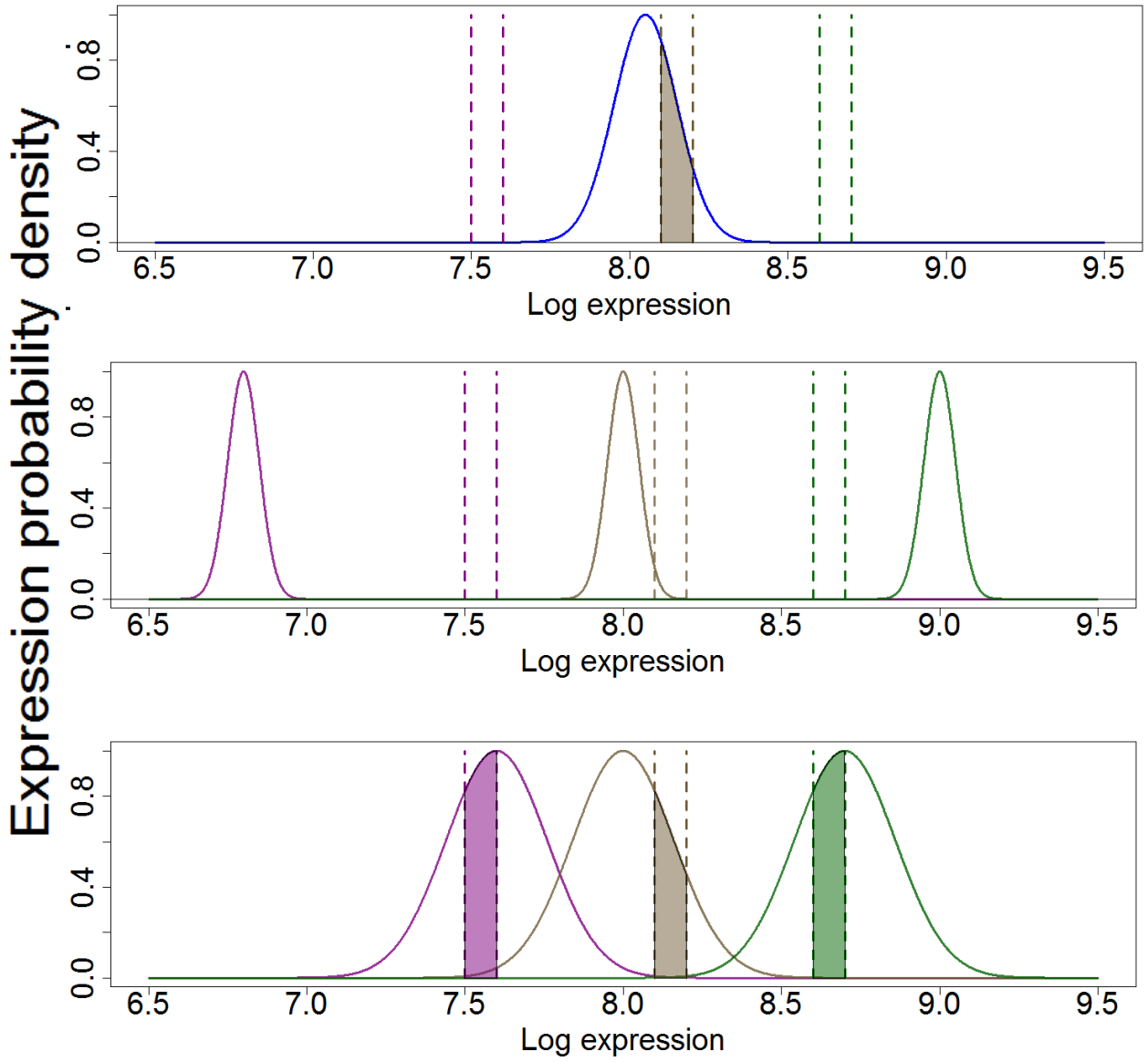
Optimal noise matches the variation in desired expression levels: $\sigma^2_{\mathrm{opt}} = \mathrm{var}(\mu_e) - \tau^2$



This is the basic **bet hedging** scenario. But:

How to genetically encode higher noise? Wouldn't it be better to evolve gene regulation?

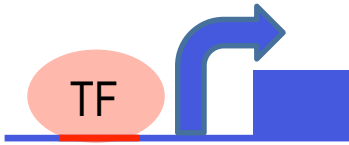
Evolving gene regulation



Gene without regulation

Regulator's expression

Gene coupled to the regulator



Including gene regulation in the model

$$\log\left[f(\mu,\sigma)\right] = -\frac{\left\langle(\mu-\mu_e)^2\right\rangle}{2(\tau^2+\sigma^2)} + \frac{1}{2}\log\left[\frac{\tau^2}{\tau^2+\sigma^2}\right]$$

- Assume in environment *e* a regulator has mean expression r_e with variance σ_r^2
- **Assume linear coupling**: When the gene evolves a binding site for regulator *r,* its expression across environments *e* are affected in 2 ways:

  Mean: $\mu(e) = \mu + cr_e$ **condition-response effect**

  Variance: $\sigma_{\text{tot}}^2 = \sigma^2 + c^2\sigma_r^2$ **noise-transmission effect**

- *c* is a coupling constant between the gene and the regulator.

Note

- A `good' regulator is one whose means r_e accurately track the desired levels μ_e.
- Coupling to a regulator always increases the noise of the gene.
- The gene's noise is increased a lot when the regulator is *itself* noisy and the coupling is strong.

Fitness of a regulated gene

$$\log\left[f(\mu,\sigma,c)\right] = -\frac{\left\langle(\mu + cr_e - \mu_e)^2\right\rangle}{2(\tau^2+\sigma^2+c^2\sigma_r^2)} + \frac{1}{2}\log\left[\frac{\tau^2}{\tau^2+\sigma^2+c^2\sigma_r^2}\right]$$

This can be rewritten in terms of 4 effective parameters:

$Y^2 = \frac{\mathrm{var}(\mu_e)}{\tau^2+\sigma^2}$ Expression mismatch *Y*

$S^2 = \frac{\mathrm{var}(r_e)}{\sigma_r^2}$ Regulator signal-to-noise *S*

$X^2 = \frac{c^2\sigma_r^2}{\tau^2+\sigma^2}$ Normalized coupling *X*

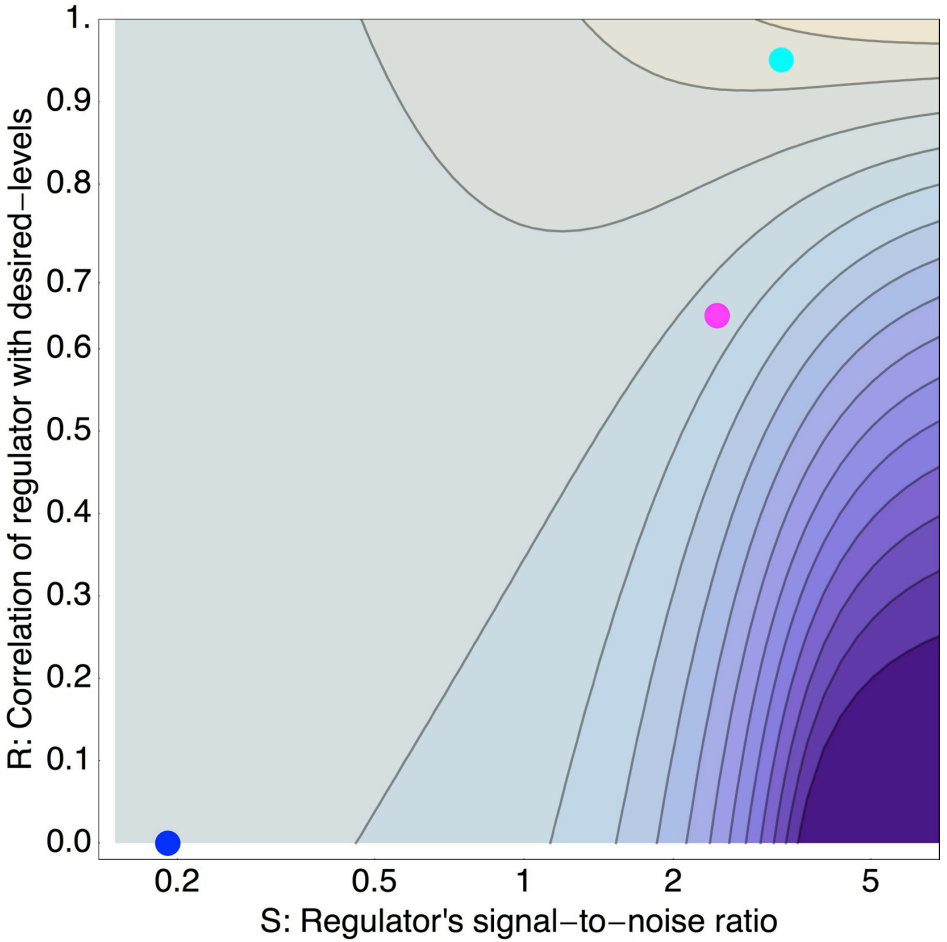
$R = \frac{\langle r_e \mu_e \rangle}{\sqrt{\mathrm{var}(r_e)\,\mathrm{var}(\mu_e)}}$ Correlation regulator-desired mean *R*

$$\log[f(X,Y,S,R)] = -\frac{1}{2}\frac{Y^2(1-R^2)+\left(SX-RY\right)^2}{(1+X^2)} - \frac{1}{2}\log\left[1+X^2\right]$$

Scenario:

- An unregulated promoter wants to improve its fitness.
- There are different regulators in the genome, each with a certain *R* and *S*.
- What fitness can be achieved by *optimally* coupling to a promoter with given *R* and *S*?

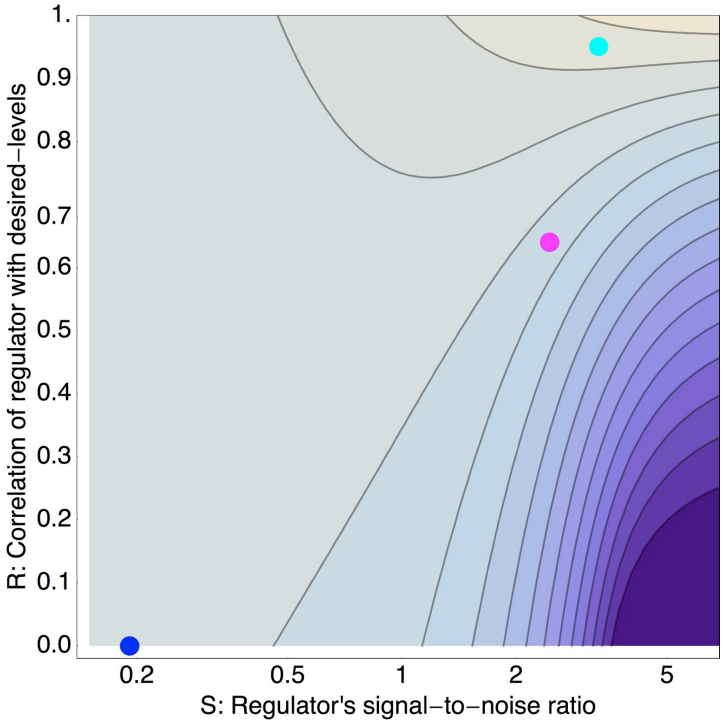
Fitness with optimal coupling to a regulator of given correlation *R* and signal-to-noise *S*



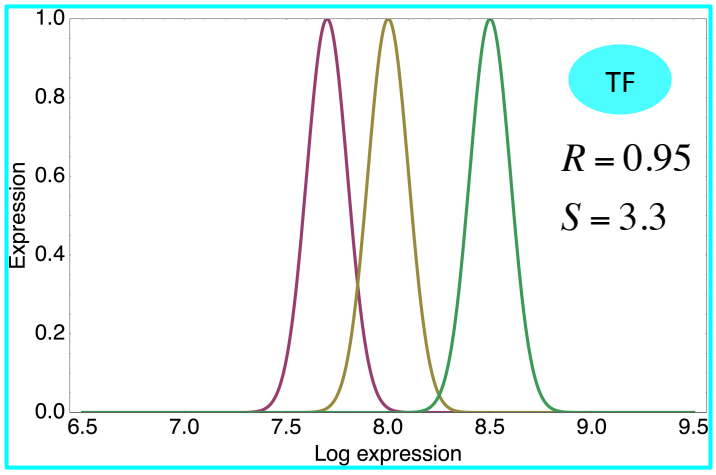
Y=4

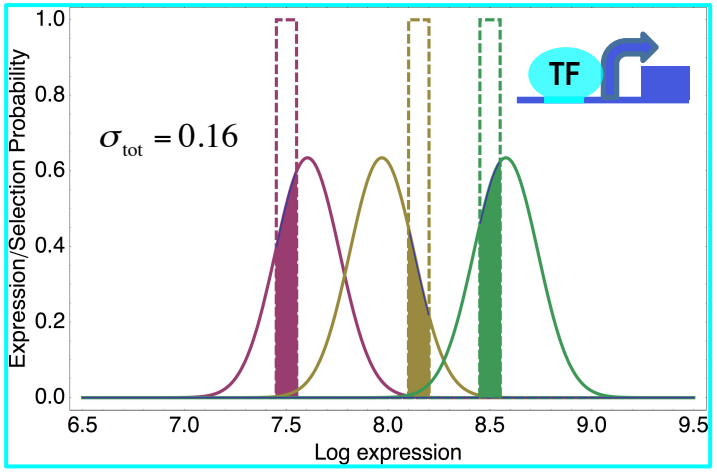
Fitness of an unregulated promoter.

Coupling to a very accurate regulator: noise as a side-effect of regulation

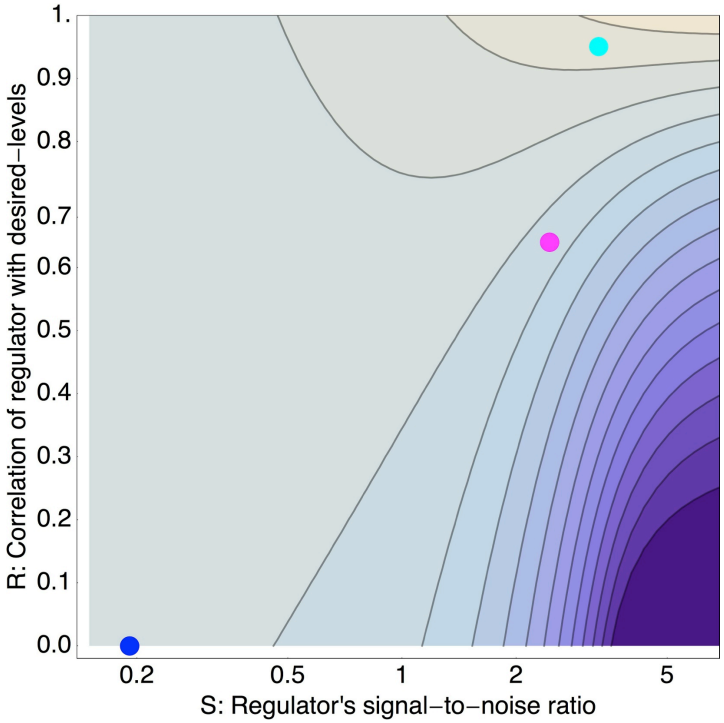


Y=4

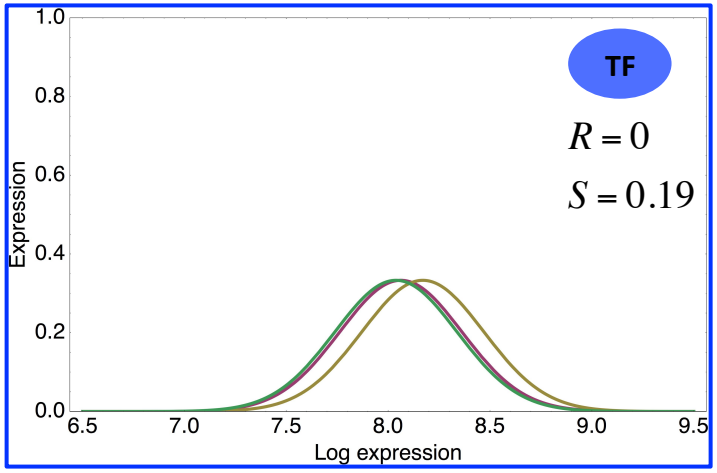


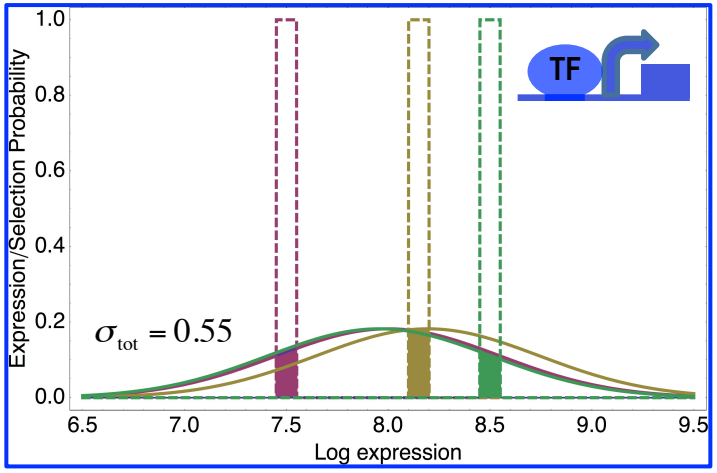


Coupling to a noisy uncorrelated regulator: implementing a bet hedging strategy

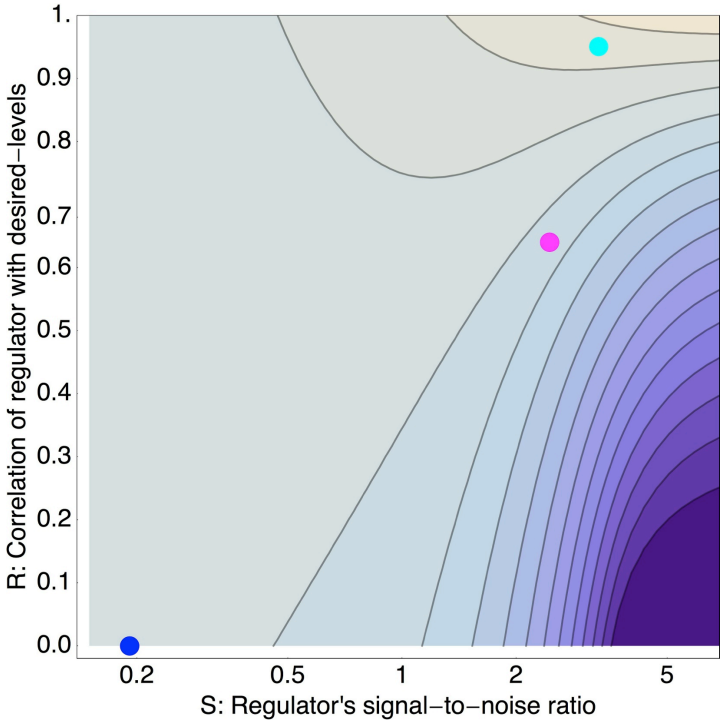


Y=4

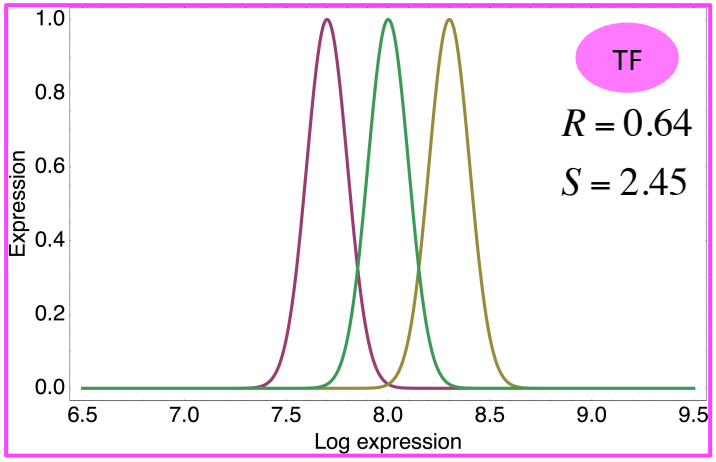


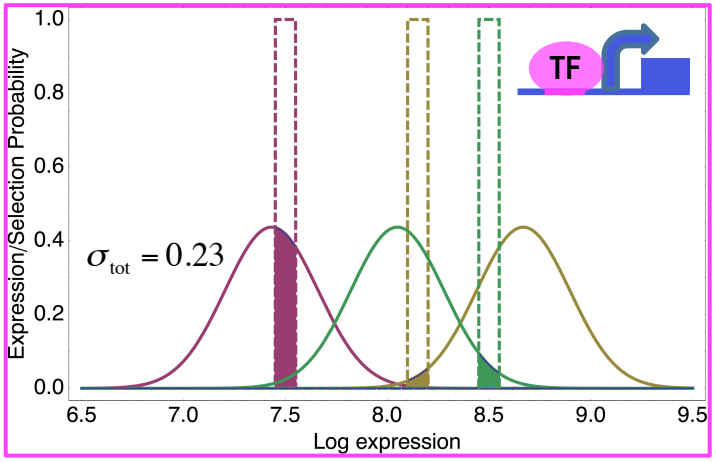


Intermediate case: a moderately correlated regulator



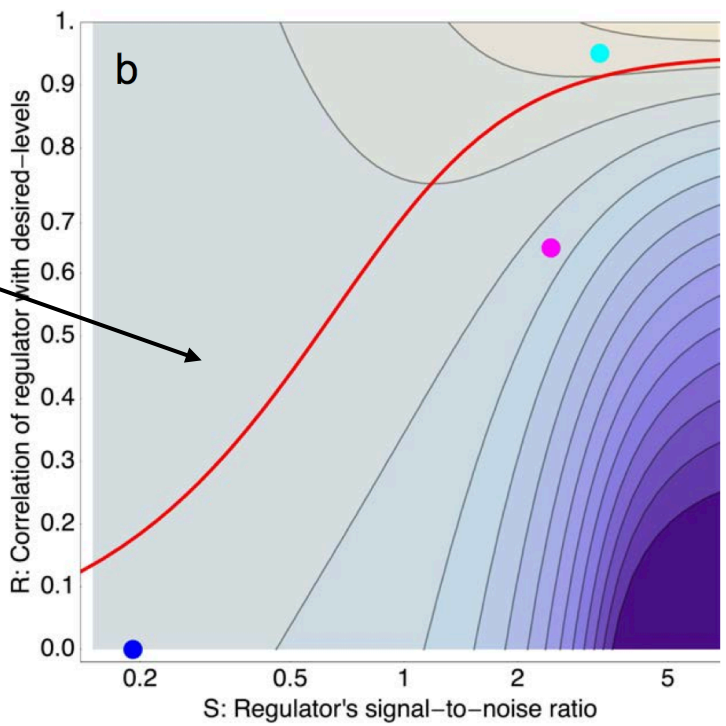
Y=4



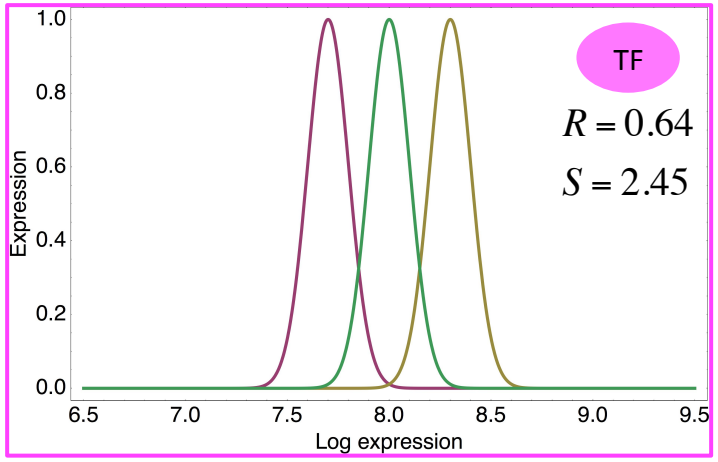


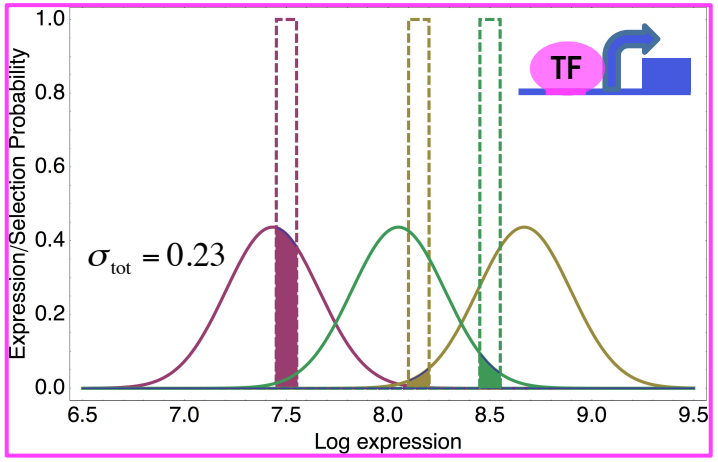
Intermediate case: a moderately correlated regulator

Optimal *S* at a given *R*.



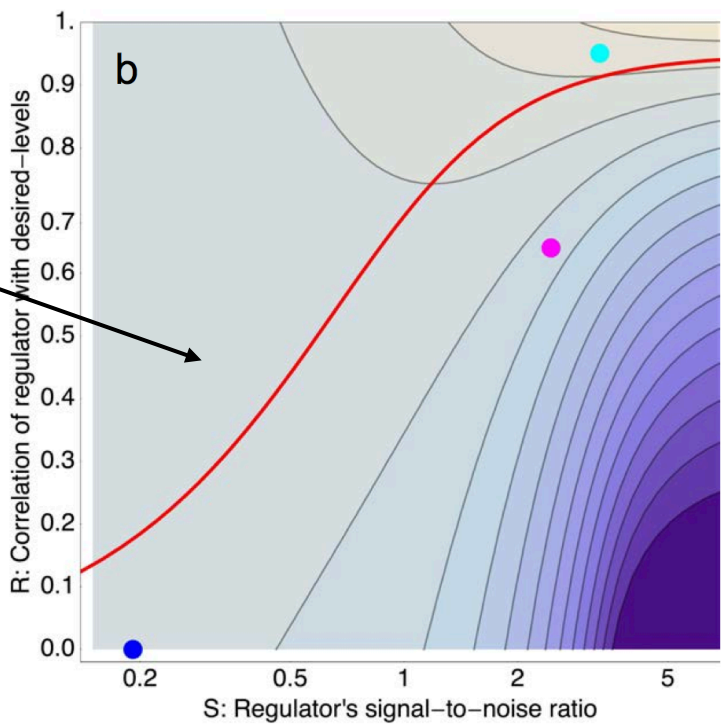
Y=4





Intermediate case: a moderately correlated regulator

Optimal *S* at a given *R*.



Y=4

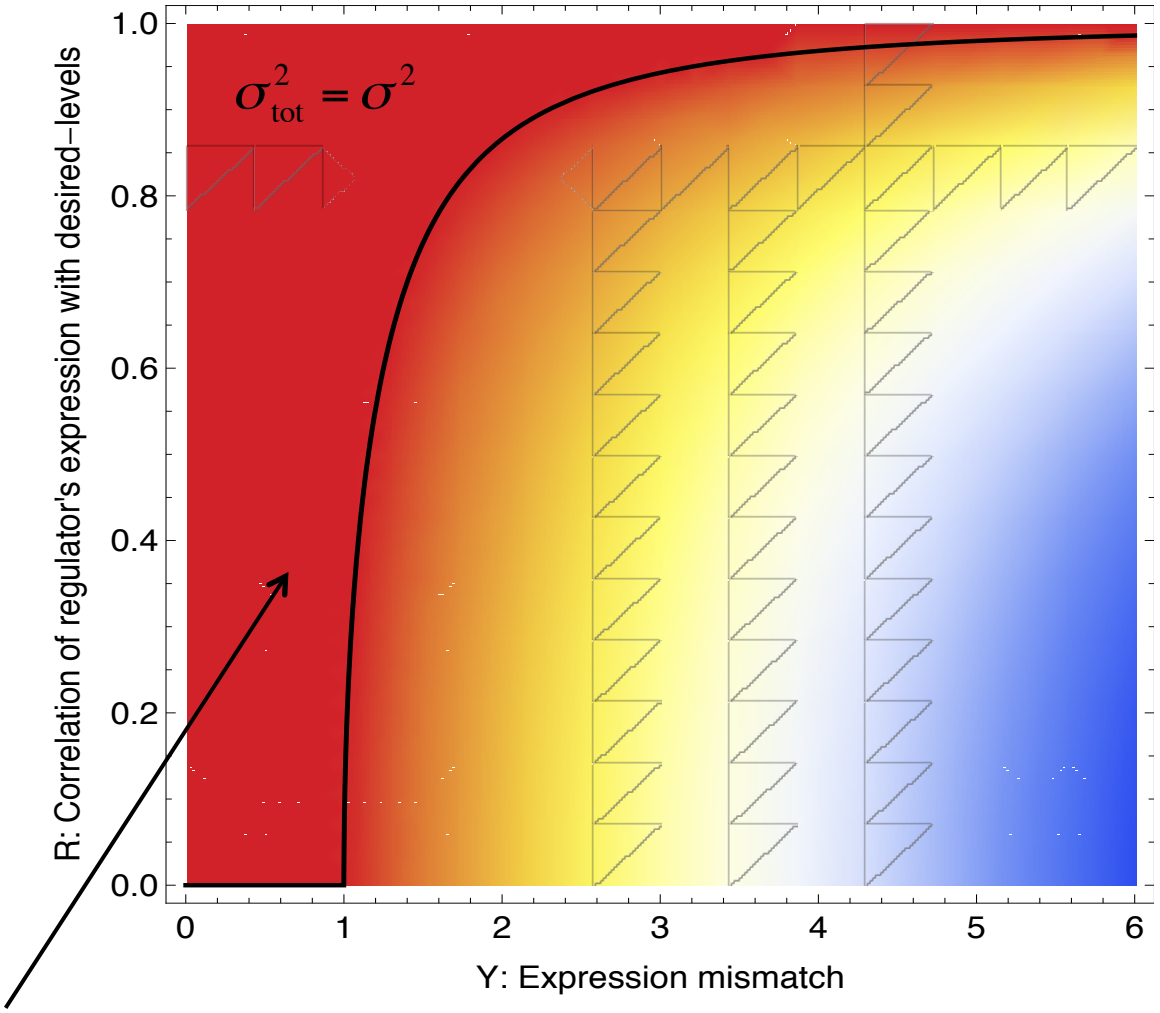
Summary

Coupling to an uncorrelated (*R*=0) and noisy (*S* low) regulator is beneficial.

To outcompete coupling to a `noise regulator', high *R* and high *S* are required.

Regulators that only moderately correlate with the desired levels of their `client' promoters, will be under selection to become noisy.

Optimizing coupling *X* and signal-to-noise *S*: Phase diagram for total noise

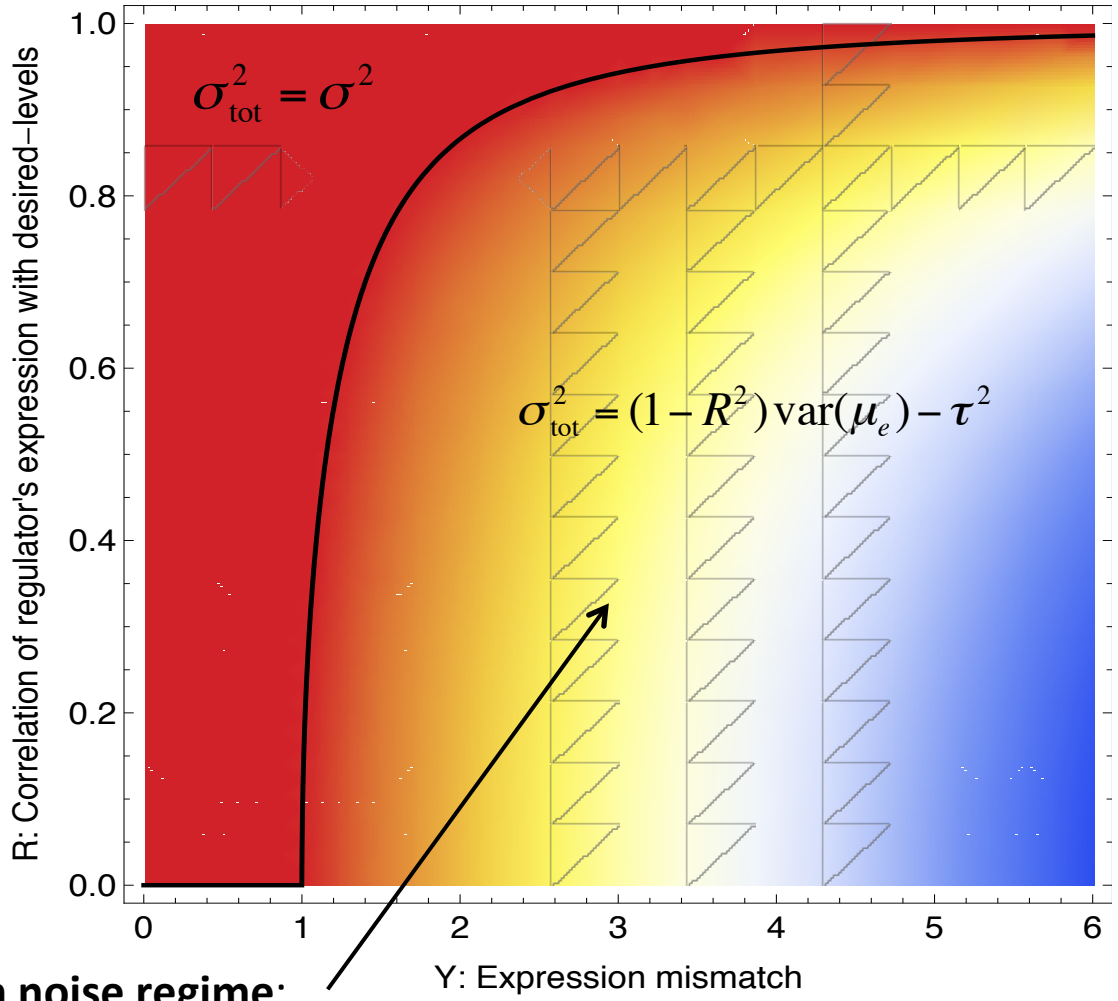


Basal noise regime:

Promoters with low expression mismatch *Y*<1 `do not bother' to be regulated.

For extremely precise regulation, noise remains low.

Optimizing coupling *X* and signal-to-noise *S*: Phase diagram for total noise

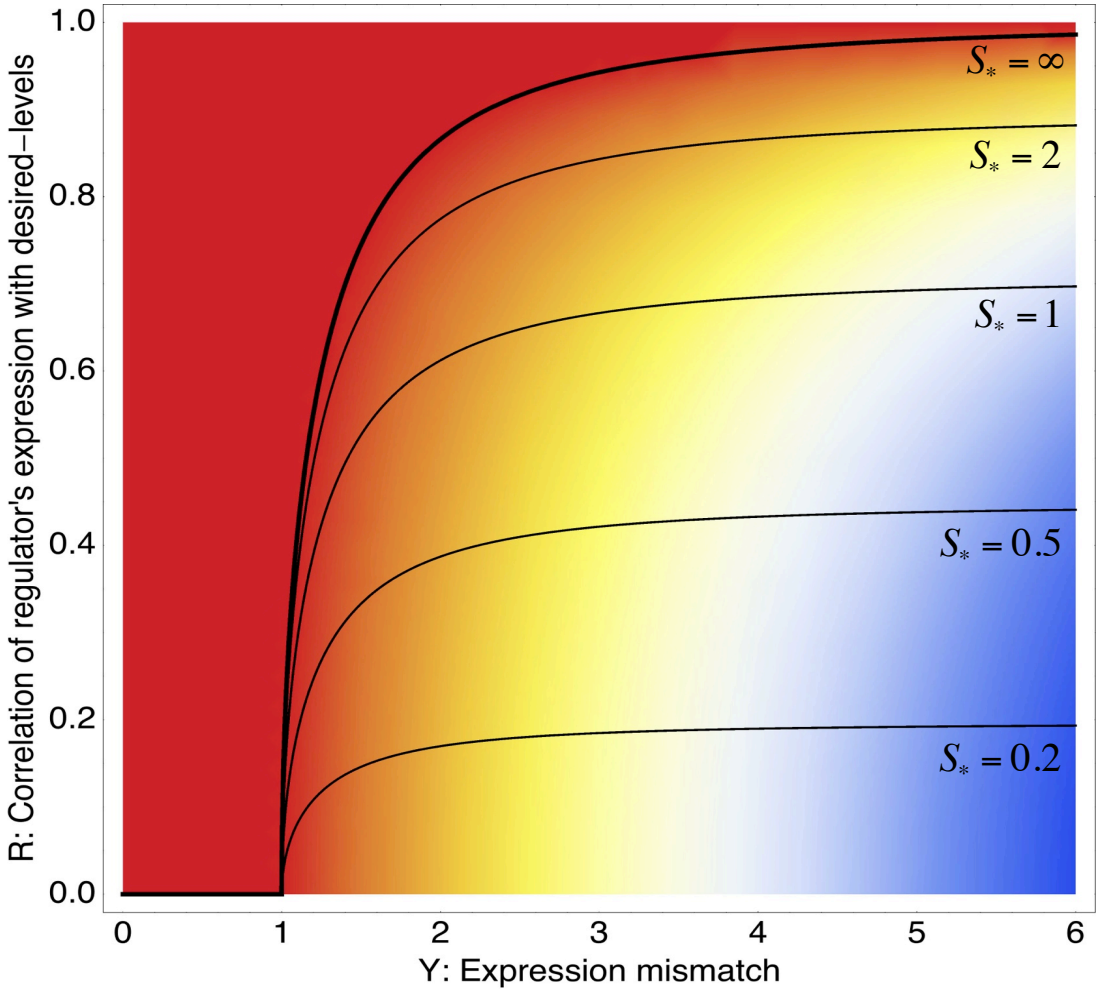


Environment-driven noise regime:

The noise level matches the fraction of variance in desired levels *not tracked* by the condition-dependence.

Unless regulation is very precise, noise increases with var(μ_e).

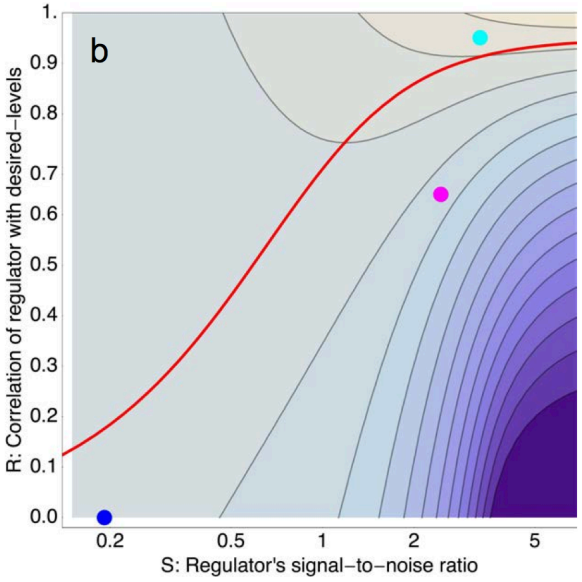
Optimizing coupling *X* and signal-to-noise *S*: Phase diagram for total noise



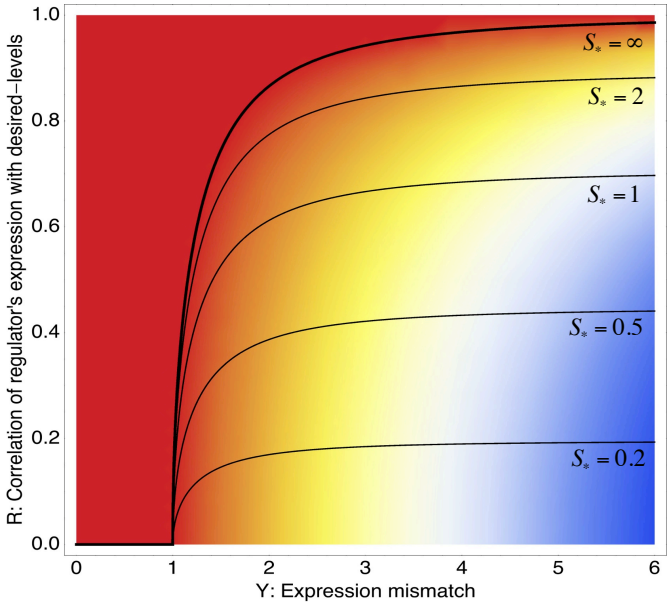
Noisy regulators are often preferred:
Optimal signal-to-noise of the regulator *S*<1 for most of the phase space.

Summary of the theory

Fitness with optimal coupling to regulator of given (*S*,*R*).



Gene noise with optimal coupling to regulator of optimal noise level *S*.



- Not only the condition-dependence but also the noise-transmission are *functional* consequences of coupling to a regulator.
- Whenever regulation is imperfect noisy regulators are preferred.
- The higher the variation in desired levels:
  - the higher the coupling to regulators, even *random* regulators.
  - The higher the final noise level of the promoter.
- Final noise level matches the variation not tracked by condition-dependence.

Implications for the evolution of gene regulation

- Regulators are often implicitly assumed to be finely tuned to implement the desired condition-dependent expression levels of their targets.
- It is unclear how this would be accomplished through small evolutionary tinkering starting from a condition without regulation.

Predictions of the theory

- Genes often benefit from coupling to regulators that only act to make the gene's expression more noisy.

- A regulator may `start out' by doing nothing more than increasing expression noise in its targets.

- Over time a regulator may then optimize its activity to increase correlation with the requirements of its targets.

- However, genes with very different regulatory requirements may benefit from coupling to the same regulator (that only weakly correlates with each gene's requirements).

- Such regulators can never optimize the correlations with all their targets and will experience selective pressure to increase their own noise.

Outlook for the theory

- In principle straight-forward to extend to many targets and many regulators.

- Can one say something about the statistics of the solutions one expects?

Challenges:

- What is correlation structure of the `expression desires' of different genes in the genome?

- To what extent can regulators match their activities to desires of their targets?
  - There must be a signal to respond to.

- Condition-dependence of noise properties.

- How does noise and regulation relate to other properties of genes such as expression level, substitution rate, etcetera?

Collect a set of Gene statistics

1. dN: Amino acid subsitution rate
2. dS: Synonymous site subtitution rate

Cell. 2008 Jul 25;134(2):341-52. doi: 10.1016/j.cell.2008.05.042.

Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution.

Drummond DA, Wilke CO.

FAS Center for Systems Biology, Harvard University, Cambridge, MA 02138, USA. dadrummond@cgr.harvard.edu

3. Average protein level

paxdb 3.0 **Von Mering lab** PaxDb: Protein Abundance Across Organisms

4. Average RNA level

Science. 2010 Jul 30;329(5991):533-8. doi: 10.1126/science.1188308.

Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells.

Taniguchi Y, Choi PJ, Li GW, Chen H, Babu M, Hearn J, Emili A, Xie XS.

Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA 02138, USA.

5. Promoter conservation

Genome Res. 2008 Jan;18(1):148-60. Epub 2007 Nov 21.

Universal patterns of purifying selection at noncoding positions in bacteria.

Molina N, van Nimwegen E.

Biozentrum, the University of Basel, and Swiss Institute of Bioinformatics, 4056-CH, Basel, Switzerland.

6. Regulatory inputs

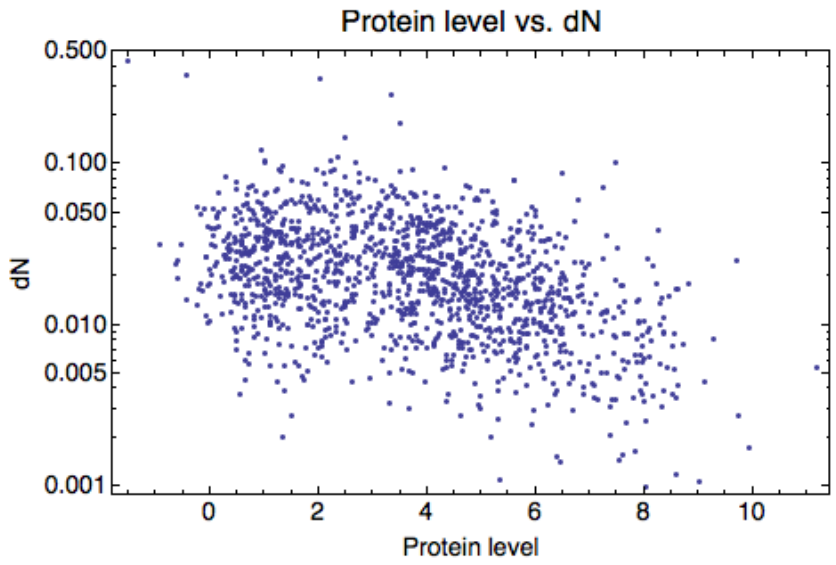
RegulonDB
Escherichia coli K12 Transcriptional Network

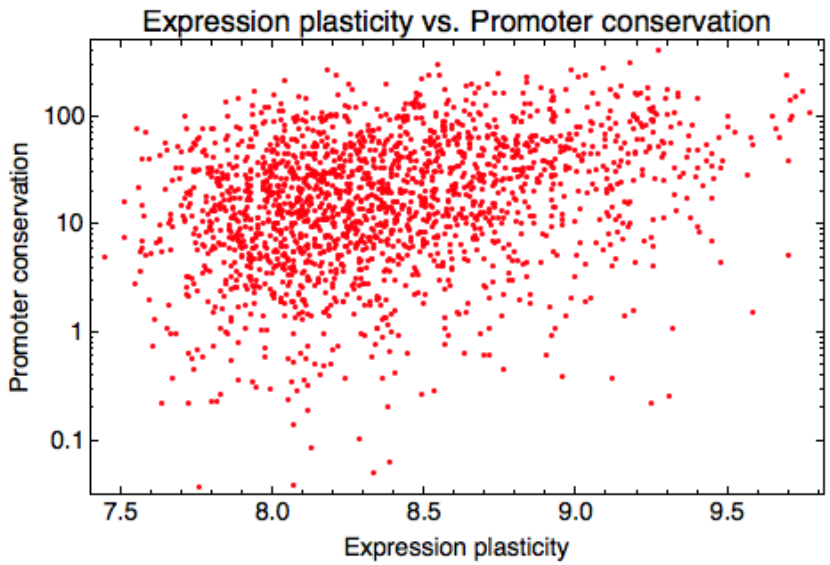
7. Expression plasticity

GenExpDB revealing the essence of life **http://genexpdb.ou.edu**

8. Promoter mean expression (our data)
9. Promoter excess noise (our data)

Calculate matrix of correlations





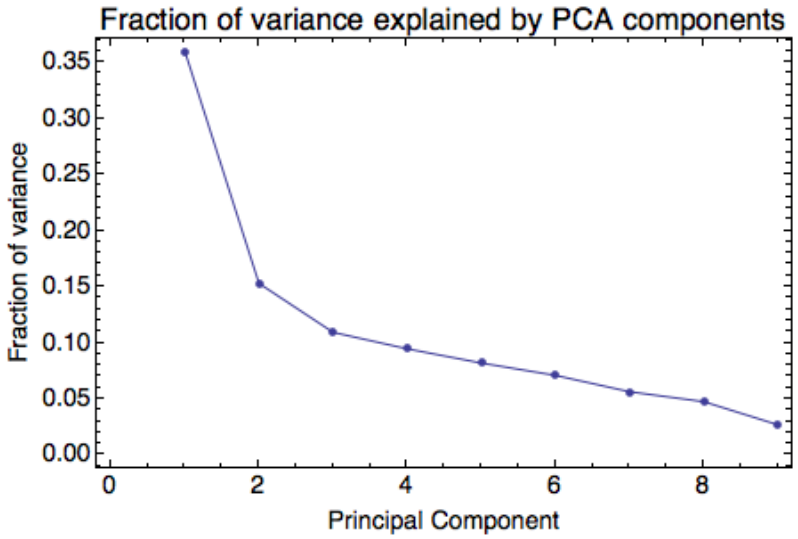
For each pair of gene stats we calculate their correlation.

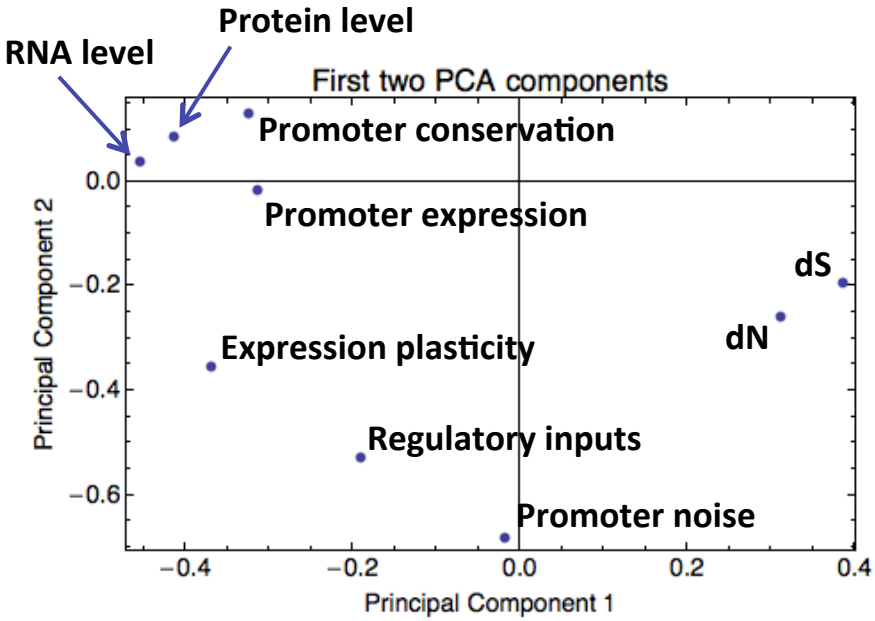
Perform *Principal Component Analysis* on the matrix of correlations.
(linear combinations of statistics that are mutually uncorrelated)

BIOZENTRUM
Universität Basel
The Center for Molecular Life Sciences

Principal component analysis of gene statistics

SIB
Swiss Institute of Bioinformatics





First principal Component

RNA level	dS
Protein level	dN
Expression plasticity	Promoter conservation
Promoter conservation	Expression plasticity
dN	Protein level
dS	RNA level

Second principal Component

Promoter noise	Promoter conservation
Regulatory inputs	Protein level
Expression plasticity	RNA level
RNA level	Expression plasticity
Protein level	Regulatory inputs
Promoter conservation	Promoter noise

BIOZENTRUM
Universität Basel
The Center for Molecular Life Sciences

SIB
Swiss Institute of Bioinformatics

Thank you!

UNI BASEL

SystemsX.ch
The Swiss Initiative in Systems Biology

FNSNF
Swiss National Science Foundation

SIB
Swiss Institute of Bioinformatics