

## Supplementary Note

Contributions of low molecule number and chromosomal positioning to stochastic gene expression (Becskei *et al.*)

### 1. Mathematical analysis of noise amplification by chromatin mediated cooperativity.

When a transcription factor ( $x_1$ , e.g. rtTA) binds cooperatively to a promoter, the protein production ( $x_2$ , e.g. YFP) driven by this promoter is described by

$$\frac{dx_2}{dt} = l + m \frac{x_1^n}{x_1^n + K^n} - \gamma_d x_2$$

$l$  and  $m$  denote the basal and maximal induced transcription rates,  $K^n$  is the equilibrium constant for transcription factor - promoter binding with a Hill-coefficient of  $n$ .  $\gamma_d$  is the protein (YFP) decay rate. The logarithmic gain<sup>1</sup>  $H_{21}$  for this transcriptional step is

$$\begin{aligned} H_{21} &= \frac{n}{x_2(x_{2\max} - x_{2\text{bas}})} (x_{2\text{bas}} - x_2)(x_{2\max} - x_2) \\ x_{2\text{bas}} &= \frac{l}{\gamma_d} \\ x_{2\max} &= \frac{l + m}{\gamma_d} \end{aligned} \tag{S1}$$

where  $x_{2\text{bas}}$  and  $x_{2\max}$  are the protein concentrations at zero (basal transcription) and maximal induction, respectively.  $|H_{21}|$  reaches a maximum when  $x_2 = \sqrt{x_{2\text{bas}} \cdot x_{2\max}}$ . In this case,

$$H_{21\max} = n \left( \frac{2\sqrt{x_{2\text{bas}} \cdot x_{2\max}}}{x_{2\max} - x_{2\text{bas}}} + 1 \right)$$

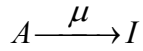
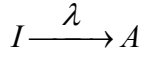
If  $x_{2\max} \gg x_{2\text{bas}}$ ,  $|H_{21\max}| \approx n$ . When intrinsic noise at the output is low, the output noise  $\eta_2$  is proportional to the product of input noise  $\eta_1$  and the logarithmic gain of the promoter  $H_{21}$  (see Ref. 1):  $\eta_2 \propto \eta_1 |H_{21}|$ .

## 2. Time-dependent behavior of the noise amplification system

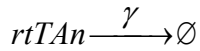
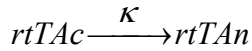
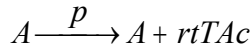
The kinetics of noise transmission was studied by stochastic simulation of the reactions using the Gillespie algorithm<sup>2</sup>:

### Steps simulated in the noise transmission module

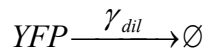
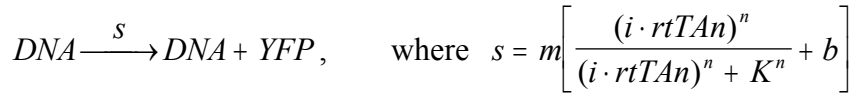
1. Promoter activation and inactivation:



2. rtTA production, transport and decay:



3. YFP production and decay:



rtTA is produced when the promoter switches from the inactive ( $I$ ) to the active ( $A$ ) state. The posttranscriptional processing of eukaryotic mRNAs is strongly coupled to activation of transcription<sup>3,4</sup>; therefore, rtTA mRNA and protein production is described in a single step. Transport / diffusion of cytoplasmic rtTA, denoted as  $rtTAc$ , into the nucleus is a relatively slow process<sup>5</sup>. Upon entry of rtTA into the nucleus,  $rtTAn$ , is ubiquitinated and degraded<sup>6</sup>. The rate of YFP expression at the tetO<sub>7</sub> promoter depends on the equilibrium concentration of active rtTA.

The concentration  $i$  denotes the relative doxycycline concentration, so that  $i=1$  if all rtTA molecules are in the active form.  $i=0$  (doxycycline is absent) if all rtTA molecules are inactive. The rate  $b$  denotes the basal expression rate at the tetO<sub>7</sub> promoter. The rate  $s$  involves fast reactions (doxycycline binding to rtTA, and rtTA binding to the tet operators), which are considered to be in equilibrium compared to the other reaction steps.  $m$  is a proportionality constant for YFP production.

### Parameter values

The average yeast mRNA half-life is around 17 min (e.g. *SWI6* mRNA  $t_{1/2}=21$  min); therefore a steady-state level of 0.1 mRNA/cell corresponds to  $\lambda=0.0045$  min<sup>-1</sup> if one mRNA molecule is transcribed at a single round of gene activation. The kinetics of signal transmission by rtTA was studied in Ref. 5 based on which, the following values were taken for rtTA transport and decay rates:  $\kappa=0.03$  min<sup>-1</sup>,  $\gamma=0.08$  min<sup>-1</sup>.  $n = 1.5$ ,  $K=0.5$ ,  $b=0.0005$ . The GFP decay rate is determined by dilution, which is inversely proportional to the cell division time,  $\gamma_{dil}=\ln 2/90$  min<sup>-1</sup>. For the example presented in Fig. S2a the experimental data are fitted well with  $\mu=0.1$  min<sup>-1</sup> and  $p=80$  min<sup>-1</sup>. One GFP molecule corresponds to an experimentally measured fluorescence intensity of 0.195 arbitrary units.

### Simulation

The coefficient of variation of YFP concentration ( $CV_{YFP}$ ) was obtained by sampling the simulated YFP molecule number over a long period of time. The final output noise  $\eta_2(x_2)$  was calculated by adding a constant noise background ( $\eta_b = 0.2$ ) to the  $CV_{YFP}$  to account for the experimentally observed residual cell variability at maximal induction of the tetO<sub>7</sub> promoter. The resulting  $\eta_2(x_2)$  relation fits well to the experimental data (Supplementary Fig. 2a).

**3. Description of the auxiliary function for fitting the input noise.** Fitting the input noise to experimental data, using the Gillespie algorithm, requires intensive computation. We therefore looked for an auxiliary analytical expression that can be used to directly fit the experimental data without extensive computation. This auxiliary function must contain a parameter representing input noise that, when varied, allows a fit of the auxiliary function to the  $\eta_2(x_2)$  relation obtained using the Gillespie algorithm.

Since the output noise is determined by the logarithmic gain of promoter response, we argued that convolution of a probability distribution with a Hill-type transfer function may approximate the output YFP distribution for a given range of parameter values; in particular, when the rate of mRNA production is slow compared to the GFP decay rate. The convolved function describes a process where the response in the concentration is immediate. It has to be noted that this formalism does not describe actual cellular processes in noise transmission, since it is only an auxiliary fitting function. Since rare events of mRNA/protein production often give rise to right-skewed probability distributions (such as lognormal<sup>7</sup>), we have chosen the lognormal distribution  $f(x)$  as the input distribution. The lognormal distribution of the input variable  $x$  was convolved with a Hill-type function  $g(x_1)$  involving a basal expression  $x_{2bas}$ , and  $x_{2max} = 1$ .

$$f(x_1) = \frac{1}{x_1 s \sqrt{2\pi}} e^{-\frac{1}{2s^2}(\ln(x_1)-m)^2}$$

$$g(x_1) = x_{2bas} + (1 - x_{2bas}) \frac{x_1^n}{x_1^n + K^n}$$

where  $m$  and  $s$  are functions of the mean  $\mu$  and the standard deviation  $\sigma$  of the distribution:

$$m = \ln\left(\frac{\mu^2}{\sqrt{\mu^2 + \sigma^2}}\right)$$

$$s = \sqrt{\ln\left(\frac{\sigma}{\mu}\right)^2 + 1}$$

The resulting output distribution of  $x_2$  ( $x_{2bas} \leq x_2 \leq 1$ ) is given by

$$f_{x_2}(x_2) = \frac{1 - x_{2bas}}{(x_2 - x_{2bas})(1 - x_2)} \frac{1}{ns\sqrt{2\pi}} e^{-\frac{1}{2s^2}\left(\ln\left(K^n \frac{x_2 - x_{2bas}}{1 - x_2}\right) - m\right)^2} \quad [\text{S2}]$$

The  $\eta_2(x_2)$  relation is calculated by varying the mean of the input probability distribution  $\mu$ , keeping  $\eta_1 = \frac{\sigma}{\mu}$  fixed. The output noise  $\eta_2$  is calculated by dividing the mean by the standard deviation of the convolved probability distribution Eq. S2.

### Validation of the auxiliary function

To justify the use of this particular function we implemented the following strategy: First, the input noise in the stochastic simulation was changed by varying the promoter activation rate between  $\lambda=0.0045$  and  $0.018 \text{ min}^{-1}$ . This range of promoter activation rate produces output noise values, which are similar to the range of experimentally observed noise (cf. Fig. 3b and Supplementary Fig. 2a). The auxiliary  $\eta_2(x_2)$  relation obtained using Eq. S2 was then fit to the output noise values from the Gillespie algorithm for a given  $\lambda$ , so that  $x_2$  in Eq. S2 equals the YFP expression level in stochastic simulation multiplied by 1.16. Each fit retrieves a fitted input noise value.

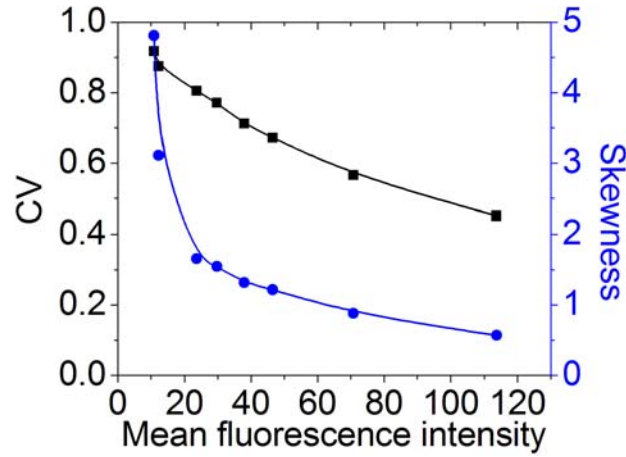
Finally, we compared these values with ones obtained using an exact relation obtained with a Master Equation approach (see Methods):

$$\eta_1^2 = \frac{\kappa}{\lambda} \left( \frac{\lambda + \mu}{p} + \frac{\mu}{\lambda + \mu + \kappa} \right) \quad [\text{S3}]$$

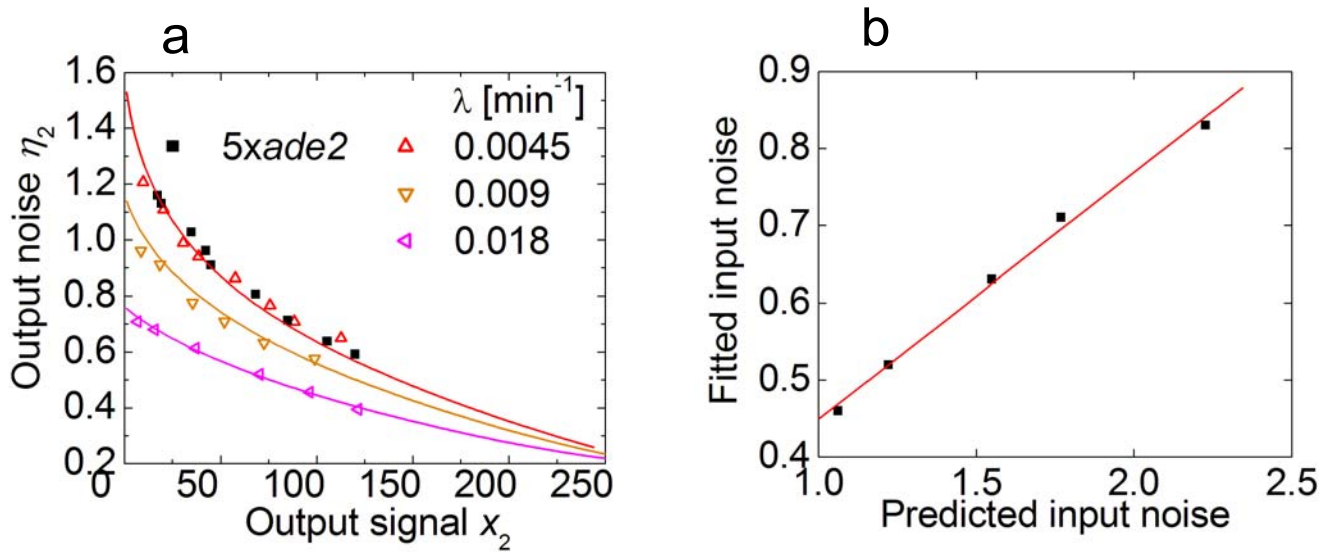
It is apparent that the fitted input noise, using the auxiliary equation, is linearly proportional to the predicted input noise from Eq. S3 (Supplementary Fig. 2b) for the entire range of  $\lambda$  examined. The auxiliary function is therefore a convenient tool to estimate the input noise without using extensive fitting procedure based computationally intensive Monte Carlo simulations.

## Literature cited

1. Paulsson, J. Summing up the noise in gene networks. *Nature* **427**, 415-8 (2004).
2. Gillespie, D. T. Exact stochastic simulation of coupled chemical-reactions. *J Phys Chem* **81**, 2340-2361 (1977).
3. Muratani, M., Kung, C., Shokat, K. M. & Tansey, W. P. The F box protein Dsg1/Mdm30 is a transcriptional coactivator that stimulates Gal4 turnover and cotranscriptional mRNA processing. *Cell* **120**, 887-99 (2005).
4. Ares, M., Jr. & Proudfoot, N. J. The spanish connection: transcription and mRNA processing get even closer. *Cell* **120**, 163-6 (2005).
5. Becskei, A., Boselli, M. G. & van Oudenaarden, A. Amplitude control of cell-cycle waves by nuclear import. *Nat Cell Biol* **6**, 451-7 (2004).
6. Salghetti, S. E., Caudy, A. A., Chenoweth, J. G. & Tansey, W. P. Regulation of transcriptional activation domain function by ubiquitin. *Science* **293**, 1651-3 (2001).
7. Krishna, S., Banerjee, B., Ramakrishnan, T. V. & Shivashankar, G. V. Stochastic simulations of the origins and implications of long-tailed distributions in gene expression. *Proc Natl Acad Sci U S A* **102**, 4771-6 (2005).



**Supplementary Figure 1.** Moments of distribution of YFP fluorescence induced in a population of ABY0542 cells ( $P_{SW14}\text{-rtTA}::[P_{TET07}\text{-YFP}]_3$ ) using different doxycycline concentrations. The coefficient of variation (CV) corresponds to noise as defined in this work. Skewness equals the third central moment normalized by the cube of standard deviation of the distribution. For comparison, distributions using  $P_{SW15}\text{-GFP}$  and  $P_{CLB2}\text{-GFP}$  constructs had skewness values of  $0.60\pm 0.07$  and  $0.64\pm 0.07$ , respectively.



**Supplementary Figure 2.** The auxiliary function fits the  $\eta_2(x_2)$  relation obtained from the Gillespie algorithm. **(a)** Triangles denote noise values calculated by the Gillespie algorithm varying the inducer concentrations for each value of  $\lambda$ . For comparison, black squares denote experimental data using the ABY0545 strain. Solid lines are fits to the triangles using Eq. S2 (see Supplementary Note). **(b)** Correlation between predicted input noise calculated by the master equation, Eq.S3, and fitted input noise values obtained using Eq. S2 (see Supplementary Note). Input noise was varied by altering the value of  $\lambda$  between 0.0045 and 0.018 min<sup>-1</sup>, as in (a). The difference in the absolute values of predicted input noise and the fitted input noise is due to neglecting time-averaging due to the fluorescent reporter in Eq. S2. However, this does not alter the linearity of this relation and Eq. S2 detects accurately the relative changes in the input noise in the noise amplification system.

**Supplementary Table 1. Yeast strains.**

| <i>Name</i> | <i>Genotype</i>  |
|-------------|--|
| ABY0511a    | <i>ade2 / ade2::ADE2-P<sub>CLN3</sub>-rtTA(S2) ura3 / ura3::[URA3-P<sub>TETO1</sub>-YFP]<sub>1</sub></i>   |
| ABY0519     | <i>ade2 / ade2::ADE2-P<sub>SWI6</sub>- rtTA(S2) ura3 / ura3::[URA3-P<sub>TETO7</sub>-YFP]<sub>3</sub></i>  |
| ABY0520     | <i>ade2 / ade2::ADE2-P<sub>MYO2</sub>- rtTA(S2) ura3 / ura3::[URA3-P<sub>TETO7</sub>-YFP]<sub>3</sub></i>  |
| ABY0521     | <i>ade2 / ade2::ADE2-P<sub>CLN2</sub>- rtTA(S2) ura3 / ura3::[URA3-P<sub>TETO7</sub>-YFP]<sub>3</sub></i>  |
| ABY0528a    | <i>ade2 / ade2::ADE2-P<sub>SWI4</sub>- rtTA(S2) his3 / his3::HIS3</i><br><i>ura3::[URA3-P<sub>TETO7</sub>-CFP]<sub>2</sub> / ura3::[URA3-P<sub>TETO7</sub>-YFP]<sub>1</sub></i>                    |
| ABY0528b    | <i>ade2 / ade2::ADE2-P<sub>SWI4</sub>- rtTA(S2) his3 / his3::HIS3</i><br><i>ura3::[URA3-P<sub>TETO7</sub>-CFP]<sub>1</sub> / ura3::[URA3-P<sub>TETO7</sub>-YFP]<sub>1</sub></i>                    |
| ABY0529     | <i>ade2 / ade2::ADE2-P<sub>SWI5</sub>- rtTA(S2) ura3 / ura3::[URA3-P<sub>TETO1</sub>-YFP]<sub>1</sub></i>  |
| ABY0530     | <i>ade2 / ade2::ADE2-P<sub>SWI5</sub>- rtTA(S2) ura3 / ura3::[URA3-P<sub>TETO2</sub>-YFP]<sub>1</sub></i>  |
| ABY0531     | <i>ade2 / ade2::ADE2-P<sub>SWI5</sub>- rtTA(S2) ura3 / ura3::[URA3-P<sub>TETO7</sub>-YFP]<sub>1</sub></i>  |
| ABY0535     | <i>ade2 / ade2::ADE2-P<sub>SWI4</sub>- rtTA(S2) ura3 / ura3::[URA3-P<sub>TETO7</sub>-YFP]<sub>1</sub></i>  |
| ABY0536     | <i>ade2::ADE2-P<sub>SWI5</sub>- rtTA(S2) / ade2::ADE2-P<sub>SWI5</sub>- rtTA(S2)</i><br><i>ura3 / ura3::[URA3-P<sub>TETO7</sub>-YFP]<sub>1</sub> his3 / his3::HIS3</i>                             |
| ABY0537     | <i>ade2 / ade2::ADE2-P<sub>CLN2</sub>- rtTA(S2) ura3 / ura3::[URA3-P<sub>TETO7</sub>-YFP]<sub>3</sub> SWI4 / swi4Δ::HIS5</i>   |
| ABY0538     | <i>ade2 / ade2::ADE2-P<sub>CLB2</sub>- rtTA(S2) ura3 / ura3::[URA3-P<sub>TETO7</sub>-YFP]<sub>3</sub></i>  |
| ABY0542     | <i>ade2 / ade2::ADE2-P<sub>SWI4</sub>- rtTA(S2) ura3 / ura3::[URA3-P<sub>TETO7</sub>-YFP]<sub>3</sub></i>  |
| ABY0543     | <i>ade2::[ADE2-P<sub>SWI6</sub>- rtTA(S2)]<sub>1</sub> / ade2::[ADE2-P<sub>SWI6</sub>- rtTA(S2)]<sub>1</sub></i><br><i>ura3 / ura3::[URA3-P<sub>TETO7</sub>-YFP]<sub>3</sub> his3 / his3::HIS3</i> |
| ABY0545     | <i>ade2 / ade2::[ADE2-P<sub>SWI6</sub>- rtTA(S2)]<sub>~5</sub> ura3 / ura3::[URA3-P<sub>TETO7</sub>-YFP]<sub>3</sub></i>   |
| ABY0547a    | <i>his3 / his3::[HIS3-P<sub>SWI6</sub>- rtTA(S2)]<sub>2</sub> ura3 / ura3::[URA3-P<sub>TETO7</sub>-YFP]<sub>3</sub></i>  |
| ABY0547b    | <i>his3 / his3::[HIS3-P<sub>SWI6</sub>- rtTA(S2)]<sub>1</sub> ura3 / ura3::[URA3-P<sub>TETO7</sub>-YFP]<sub>3</sub></i>  |
| ABY0548c    | <i>ade2 / ade2::ADE2-P<sub>BUD1</sub>- rtTA(S2) ura3 / ura3::[URA3-P<sub>TETO7</sub>-YFP]<sub>3</sub></i>  |
| ABY0549     | <i>ade2::ADE2-P<sub>SWI5</sub>- rtTA(S2) / ade2::[ADE2-P<sub>SWI5</sub>- YFP]<sub>2</sub></i><br><i>URA3 / ura3::[URA3-P<sub>TETO7</sub>-CFP]<sub>2</sub> his3 / his3::HIS3 trp1 / trp1::TRP1</i>  |
| ABY0551a    | <i>ade2 / ade2::ADE2-P<sub>BUD1</sub>- rtTA(S2) SWI6 / swi6:: P<sub>SWI5</sub>- SWI6- ADE2</i><br><i>ura3 / ura3::[URA3-P<sub>TETO7</sub>-YFP]<sub>3</sub> his3 / his3::HIS3</i>                   |
| ABY0557     | <i>ade2 / ade2::[ADE2-P<sub>SWI5</sub>- rtTA(S2)-CFP]<sub>3</sub></i><br><i>ura3 / ura3::[URA3-P<sub>TETO7</sub>-YFP]<sub>3</sub> his3 / his3::HIS3</i>  |
| ABY0559     | <i>ade2::[ADE2-P<sub>SWI5</sub>- rtTA(S2)-CFP]<sub>3</sub> / ade2::[ADE2-P<sub>SWI5</sub>- rtTA(S2)-YFP]<sub>3</sub> ura3 / URA3</i><br><i>his3 / his3::HIS3</i>                                   |