www.sciencemag.org/cgi/content/full/329/5999/1656/DC1

Supporting Online Material for

A General Mechanism for Network-Dosage Compensation in Gene Circuits

Murat Acar,* Bernardo F. Pando, Frances H. Arnold, Michael B. Elowitz, Alexander van Oudenaarden

*To whom correspondence should be addressed. E-mail: acar@caltech.edu

Published 24 September 2010, *Science* **329**, 1656 (2010)
DOI: 10.1126/science.1190544

This PDF file includes:

Materials and Methods
Figs. S1 to S8
Tables S1 to S5
References

A general mechanism for network-dosage compensation in gene networks

Murat Acar, Bernardo F. Pando, Frances H. Arnold, Michael B. Elowitz & Alexander van Oudenaarden

SUPPLEMENTARY INFORMATION

Contents

I	Constructions of plasmids and strains	1
II	Yeast strains used in this study	2
III	Growth conditions and media	3
IV	Additional view of the data and quantification of the effect of each gene	3
V	Model specification	4
VI	Constraints on model parameters	6
VII	Analytical approximation	8
VIII	Fitting procedure and best fit results	10
IX	Mathematical analysis of the properties of one-gene and two-gene networks	11

I. Constructions of plasmids and strains

*Kpn*I−P_{GAL1}−*Bam*HI, *Bam*HI−YFP−*Eco*RI fragments were cloned into pRS402 backbone upstream of *CYC1* transcriptional terminator. The P_{GAL1} promoter sequence corresponds to the 669 base-pair region directly upstream of the start codon of the *GAL1* gene. The *KanMX4* and *NatMX4* gene deletion cassettes (*1*, *2*) were used to delete specific GAL network genes. Integrations were verified by polymerase chain reaction (PCR). All strains are diploid and were derived from W303. Complete descriptions of the strains used in this study can be found in Table S1.

In Fig. S1 we show an schematic representing the construction of the first-order dosage-varied yeast strains. Each rectangle denotes a diploid strain that was dosage-halved in one of the four regulatory genes of the network.

II. Yeast strains used in this study

Strain	Genotype
MA0491	*MAT*a/α, *leu2*/*leu2*::*LEU2*, *his3*::*HIS3*/*his3*, *ade2*::*ADE2*-P_{MYO2}-*rtTA*/*ade2*::*ADE2*-P_{GAL1}-*YFP*, *gal2*Δ::*KanMX*/*GAL2*
MA0492	*MAT*a/α, *leu2*/*leu2*::*LEU2*, *his3*::*HIS3*/*his3*, *ade2*::*ADE2*-P_{MYO2}-*rtTA*/*ade2*::*ADE2*-P_{GAL1}-*YFP*, *gal3*Δ::*KanMX*/GAL3
MA0493	*MAT*a/α, *leu2*/*leu2*::*LEU2*, *his3*::*HIS3*/*his3*, *ade2*::*ADE2*-P_{MYO2}-*rtTA*/*ade2*::*ADE2*-P_{GAL1}-*YFP*, *gal4*Δ::*KanMX*/*GAL4*
MA0494	*MAT*a/α, *leu2*/*leu2*::*LEU2*, *his3*::*HIS3*/*his3*, *ade2*::*ADE2*-P_{MYO2}-*rtTA*/*ade2*::*ADE2*-P_{GAL1}-*YFP*, *gal80*Δ::*KanMX*/*GAL80*
MA0496	*MAT*a/α, *leu2*/*leu2*::*LEU2*, *his3*::*HIS3*/*his3*, *ade2*::*ADE2*-P_{MYO2}-*rtTA*/*ade2*::*ADE2*-P_{GAL1}-*YFP*
MA0614	*MAT*a/α, *leu2*/*leu2*::*LEU2*, *his3*::*HIS3*/*his3*, *ade2*::*ADE2*-P_{MYO2}-*rtTA*/*ade2*::*ADE2*-P_{GAL1}-*YFP*, *gal2*Δ::*KanMX*/*GAL2*, *GAL3*/*gal3*Δ::*KanMX*
MA0615	*MAT*a/α, *leu2*/*leu2*::*LEU2*, *his3*::*HIS3*/*his3*, *ade2*::*ADE2*-P_{MYO2}-*rtTA*/*ade2*::*ADE2*-P_{GAL1}-*YFP*, *gal2*Δ::*KanMX*/*GAL2*, *GAL4*/*gal4*Δ::*KanMX*
MA0616	*MAT*a/α, *leu2*/*leu2*::*LEU2*, *his3*::*HIS3*/*his3*, *ade2*::*ADE2*-P_{MYO2}-*rtTA*/*ade2*::*ADE2*-P_{GAL1}-*YFP*, *gal2*Δ::*KanMX*/*GAL2*, *GAL80*/*gal80*Δ::*KanMX*
MA0617	*MAT*a/α, *leu2*/*leu2*::*LEU2*, *his3*::*HIS3*/*his3*, *ade2*::*ADE2*-P_{MYO2}-*rtTA*/*ade2*::*ADE2*-P_{GAL1}-*YFP*, *gal3*Δ::*KanMX*/*GAL3*, *GAL4*/*gal4*Δ::*KanMX*
MA0618	*MAT*a/α, *leu2*/*leu2*::*LEU2*, *his3*::*HIS3*/*his3*, *ade2*::*ADE2*-P_{MYO2}-*rtTA*/*ade2*::*ADE2*-P_{GAL1}-*YFP*, *gal3*Δ::*KanMX*/*GAL3*, *GAL80*/*gal80*Δ::*KanMX*
MA0619	*MAT*a/α, *leu2*/*leu2*::*LEU2*, *his3*::*HIS3*/*his3*, *ade2*::*ADE2*-P_{MYO2}-*rtTA*/*ade2*::*ADE2*-P_{GAL1}-*YFP*, *gal4*Δ::*KanMX*/*GAL4*, *GAL80*/*gal80*Δ::*KanMX*
MA0620	*MAT*a/α, *leu2*/*leu2*::*LEU2*, *his3*::*HIS3*/*his3*, *ade2*::*ADE2*-P_{MYO2}-*rtTA*/*ade2*::*ADE2*-P_{GAL1}-*YFP*, *gal2*Δ::*KanMX*/*GAL2*, *gal3*Δ::*NatMX*/*GAL3*, *GAL4*/*gal4*Δ::*KanMX*
MA0621	*MAT*a/α, *leu2*/*leu2*::*LEU2*, *his3*::*HIS3*/*his3*, *ade2*::*ADE2*-P_{MYO2}-*rtTA*/*ade2*::*ADE2*-P_{GAL1}-*YFP*, *gal3*Δ::*KanMX*/*GAL3*, *gal4*Δ::*NatMX*/*GAL4*, *GAL80*/*gal80*Δ::*KanMX*
MA0622	*MAT*a/α, *leu2*/*leu2*::*LEU2*, *his3*::*HIS3*/*his3*, *ade2*::*ADE2*-P_{MYO2}-*rtTA*/*ade2*::*ADE2*-P_{GAL1}-*YFP*, *GAL2*/*gal2*Δ::*KanMX*, *gal3*Δ::*NatMX*/*GAL3*, *gal80*Δ::*KanMX*/*GAL80*
MA0623	*MAT*a/α, *leu2*/*leu2*::*LEU2*, *his3*::*HIS3*/*his3*, *ade2*::*ADE2*-P_{MYO2}-*rtTA*/*ade2*::*ADE2*-P_{GAL1}-*YFP*, *GAL2*/*gal2*Δ::*KanMX*, *gal4*Δ::*KanMX*/*GAL4*, *gal80*Δ::*NatMX*/*GAL80*
MA0625	*MAT*a/α, *leu2*/*leu2*::*LEU2*, *his3*::*HIS3*/*his3*, *ade2*::*ADE2*-P_{MYO2}-*rtTA*/*ade2*::*ADE2*-P_{GAL1}-*YFP*, *GAL2*/*gal2*Δ::*KanMX*, *gal3*Δ::*NatMX*/*GAL3*, *GAL4*/*gal4*Δ::*NatMX*, *gal80*Δ::*KanMX*/*GAL80*

Table S1. Yeast strains used in this study.

III. Growth conditions and media

Cultures were grown in synthetic dropout media with the appropriate amino-acid supplements. During the overnight growth period (20 hours in 30°C shaker), 2% raffinose was used as the carbon source. The overnight growth period was followed by the induction period (20 hours in 30°C shaker), with cultures containing 0.1% glucose and 0-0.4% galactose as carbon sources. 0.1% glucose instead of raffinose was used as a background carbon source to ensure a constant growth rate across different galactose concentrations. Glucose at this concentration does not fully repress *GAL* network activity. After the induction period, the expression distributions were determined by flow cytometry (FACScan; Becton Dickinson). The OD_{600} values at the end of the induction period were kept low ($OD_{600} \leq 0.322$) to prevent nutrient depletion. The culture volume was 10ml during both overnight growth and induction periods.

IV. Additional view of the data and quantification of the effect of each gene

In Fig. S2, we show another view of the experimental data presented in Fig. 3, comparing the inducibility profile of each strain analyzed against that of the wild-type.

In Fig. 3C, we quantified the average contribution of the second copy of each regulatory gene to network inducibility in the following manner. Separately for each gene, all genetic backgrounds in which that gene was dosage-halved were identified, corresponding to 8 out of 16 strain backgrounds. Inducibility values of the strains that carry both copies of a specific gene were subtracted from the inducibility values of the strains carrying one copy of that gene in an otherwise identical genetic background and these differences were averaged for each gene. To complement Fig. 3C, in Fig. S3 we report the average contribution of halving each of the regulatory genes involved using a metric based on the average squared differences (denoted χ^2, fig. S3A) across different galactose concentrations and genetic backgrounds. This measure quantifies the average contribution of each gene regardless of its sign and it complements the quantification based on an average of the differences (denoted Δ) presented in Fig. 3C and fig. S3B.

V. Model specification

We consider an effective stochastic model in which a given promoter site of each of the GAL regulated genes (*GAL2*, *GAL3* and *GAL80*) can be in either a state of active transcription (ON-state) or in a state in which transcription occurs less often (OFF-state) (*3, 4*). Each of these states is characterized by its typical transcription rate. We chose to parameterize the system so that one copy of the *GALi* gene will produce the corresponding proteins at rate θ_i (which coarse-grains the processes of transcription, translation and protein folding) when in the ON-state and at rate $\lambda\theta_i$, when in the OFF-state. So λ represents the relative transcriptional strength of the OFF-state compared to the ON-state.

We consider that slow stochastic transitions between these transcriptional states are possible and that the total concentration of the different regulatory proteins affects the rate at which the $\mathrm{OFF} \rightarrow \mathrm{ON}$ transition takes place.

$$\mathrm{OFF} \underset{h}{\overset{h\,\rho(x_2, x_3, x_4, x_{80})}{\rightleftharpoons}} \mathrm{ON} \tag{1}$$

In this scheme the parameter h represents a typical timescale at which these transitions take place and ρ is a function that quantifies how the total concentrations of the different GAL proteins (x_2, x_3, x_4, x_{80}) affect the $\mathrm{OFF} \rightarrow \mathrm{ON}$ transition. This description is valid as long as the molecular interactions that shape the regulating function ρ occur much more rapidly than the typical timescale at which protein concentrations change due to the processes of transcription, translation, and protein dilution/degradation.

We parameterized ρ by taking into account what is known about the way the different GAL proteins interact with each other and affect transcription. In Fig. S4 we represent schematically the main interactions between the proteins involved.

First, it is known that the *GAL4* protein is the main transcriptional activator when it is not bound by *GAL80* proteins, so we proposed the form

$$\rho = \left(\frac{x_4^*}{K_4}\right)^{\eta} \tag{2}$$

where K_4 represents the effective typical concentration scale of the interaction, $\eta > 0$ is its typical effective nonlinearity, and x_4^* is the concentration of *GAL4* that is not bound by *GAL80* and can therefore freely activate transcription. Instead of writing a set of

reactions for describing how x_4^* depends on the total concentrations of the GAL network proteins, we propose to use simple functional forms that effectively describe the main nature of the interactions. In these functional forms, the molecular interactions are characterized by typical concentration scales of action for each protein and by typical degrees of nonlinearity quantified by positive exponents. In this case, we know that the amount of free *GAL4* proteins will be a decreasing function of the concentration of *GAL80* proteins in the nucleus and an increasing function of the total concentration of *GAL4* proteins. Therefore, we propose to use the form

$$x_4^* = \frac{x_4}{1 + \left(\frac{x_{80}^*}{K_{80}}\right)^{\beta}} \tag{3}$$

where x_{80}^* is the concentration of *GAL80* proteins in the nucleus. This quantity, in turn, is regulated by the active *GAL3* proteins due to sequestration

$$x_{80}^* = \frac{x_{80}}{1 + \left(\frac{x_3^*}{K_3}\right)^{\alpha}} \tag{4}$$

where x_3^* is the concentration of active *GAL3* proteins. The internal galactose concentration, g^*, regulates the activation of *GAL3* proteins and therefore we propose to write

$$x_3^* = \frac{x_3}{1 + \left(\frac{g^*}{K_g}\right)^{-\nu}}. \tag{5}$$

Note that in this case the number of active *GAL3* proteins is an increasing function of the concentration of internal galactose because we assume $\nu > 0$. Finally, the concentration of internal galactose is regulated by the concentration of galactose with which the cells were grown, g, and the amount of *GAL2* proteins (the galactose permease) and so we write

$$g^* = \frac{g}{1 + \left(\frac{x_2}{K_2}\right)^{-\mu}}. \tag{6}$$

Equations (2) to (6) describe how the rate of the $\mathrm{OFF} \rightarrow \mathrm{ON}$ transition is regulated by the total concentrations of the different proteins involved as well as the concentration of external galactose.

To finalize the specification of the model, we also assumed that the protein degradation rates were slow compared to the growth rate and so we only included the effect of the dilution of proteins at rate γ, the average growth rate of yeast in the laboratory.

We simulated this system by using a custom-written C++ implementation of the Gillespie algorithm, which considered the production of proteins, their dilution due to cell growth, and the transitions between the transcriptional states as first-order stochastic reactions. In Fig. S5 we show sample trajectories for the different variables in a simulation corresponding to the parameters reported in Sections VI and VIII.

VI. Constraints on model parameters

To keep our model realistic, we constrained the values of several parameters to previously measured quantities. However, some quantities (especially the effective parameters we used) were difficult to estimate based on published work and therefore we extracted them out by fitting the model to our data.

On one hand, the doubling time of yeast in the environments used in this study is about 90 minutes, which imposes the constraint $\gamma \approx 0.46$ h^{-1}.

Previous high-throughput studies identified fold-differences in transcript levels for several yeast genes under two different growth conditions. More specifically, yeast cells were grown in two separate environments, one promoting the expression of the GAL genes and one repressing it. Average differences of about 5.5-fold and 3.7-fold were reported for *GAL3* and *GAL80*, respectively (*5*). A high-throughput study that quantified the amount and localization of different yeast proteins reported that there were about 800 *GAL3* and 700 *GAL80* proteins per cell when the GAL genes were repressed (*6*), which in the context of the proposed model would correspond to the situation in which these genes are in the OFF transcriptional state. Considering these observations and with the aim of simplifying the description further, we assumed that all GAL-regulated genes in the network follow a similar regulation scheme and fixed the values of the parameters θ_3, θ_{80} and λ to obtain basal expression levels of about 750 proteins per cell and a 5-fold increase in protein levels when the network is fully induced. So, taking the dilution rate into account, this implies the constraints $\theta_3 \cong \theta_{80} \cong 1725$ proteins/h and $\lambda \cong 0.2$. Related experimental evidence for *GAL2* is more elusive and for the sake of

simplicity we assumed the same transcription rate as for *GAL3* and *GAL80*. The same studies also reported that the level of *GAL4* transcripts does not change significantly between galactose-free and galactose-rich media (*5*) and that *GAL4* proteins are present at a concentration of about 200 proteins per cell (*6*), which implies the constraint $\theta_4 \cong 92$ proteins/h.

We note that in the model at hand the stochasticity in the change of the transcriptional plan is described by slow transitions between two different states. In this framework, the fluctuations in protein expression levels play a secondary role in establishing the fraction of active cells under a given condition, though they still play a major role in shaping the distributions associated with OFF and ON expression states. Furthermore, the way we set up the regulation scheme indicated by equations (2)-(6) allows us to interpret the constants K_i as being expressed in terms of the typical concentrations of the associated proteins. Therefore, we don't lose generality by using parameter values θ_i that might deviate slightly from experimental measurements.

The values of the parameters used in simulations and fits presented in this study can be found in Table S2. These parameters were constrained with respect to the previous studies described in this section.

Parameter	Value
θ_2	1500 proteins/h
θ_3	1500 proteins/h
θ_4	100 proteins/h
θ_{80}	1500 proteins/h
λ	0.2
γ	0.46 h^{-1}

Table S2. Parameters fixed based on previous observations.

In order to obtain OFF and ON expression states that are well-separated from each other, the time that it takes for protein levels to equilibrate has to be shorter than the typical timescale of the transitions between the two transcriptional states. We quantified the timescale of transitions through the parameter h in equation (1). Exploring a range of values for this quantity, we found reasonable agreement with the experimental results for $h = 2.5$ h^{-1} (Fig. S5 and Fig. S6A-C). If the value of h is too high (Fig. S6C) the distribution of protein numbers becomes monomodal; on the other hand if

this parameter is too low (Fig. S6A), the dynamics of establishment of fractions would be too slow compared to the experimental observations and in the case of multiple promoters it would lead to the appearance of three distinct expression states, which is something that is not observed experimentally. We also note that all inferences presented in this work are based on the analytical approximation described in the next section, where the exact value of the parameter h becomes immaterial.

VII. Analytical approximation

The stochastic simulation of the model described above is computationally time consuming. To simplify the exploration of model parameters and/or alternative models, we developed an approximation for the steady-state fraction of actively transcribing cells in a macroscopic population. We note that the presentation of the approximation proposed here is not a rigorous derivation. We based it on intuition and heuristic observations, and we eventually confirmed its power by comparing inferences drawn from it to those obtained from detailed stochastic simulations (Fig. S6D).

For one cell, we can approximate the time-evolution of the number of proteins associated with each GAL-network-regulated gene with a set of Langevin equations (*7-9*) of the form

$$\dot{x}_i = \theta_i[\phi + \lambda(1-\phi)] - \gamma x_i + \xi_i \qquad (7)$$

where x_i represents the concentration of the protein associated with the GALi gene, ϕ is a random binary variable that indicates whether the cell is transcribing or not, and ξ_i is a random variable that approximates the intrinsic stochasticity associated with the processes of protein production and dilution.

An equation for the evolution of mean protein numbers across a population of cells, $\langle x_i \rangle$, can be obtained by averaging equation (7) above. If we assume that the intrinsic noise in protein expression can be neglected we obtain a set of equations of the form:

$$\dot{x}_i = \theta_i[\lambda + (1-\lambda)\langle\phi\rangle] - \gamma\langle x_i\rangle \qquad (8)$$

where $\langle\phi\rangle$ represents the fraction of cells that are actively transcribing. Following a mean-field approximation approach, we estimate the fraction $\langle\phi\rangle$ with the value that we

would infer from assuming a constant background of protein concentrations equal to their average values, namely:

$$\langle\phi\rangle \cong \langle\frac{k_{\mathrm{OFF}\to\mathrm{ON}}(x_2,x_3,x_4,x_{80})}{k_{\mathrm{OFF}\to\mathrm{ON}}(x_2,x_3,x_4,x_{80})+k_{\mathrm{ON}\to\mathrm{OFF}}}\rangle \cong \frac{1}{1+\frac{k_{\mathrm{ON}\to\mathrm{OFF}}}{k_{\mathrm{OFF}\to\mathrm{ON}}(\langle x_2\rangle,\langle x_3\rangle,\langle x_4\rangle,\langle x_{80}\rangle)}}$$
$$= \frac{1}{1+[\rho(\langle x_2\rangle,\langle x_3\rangle,\langle x_4\rangle,\langle x_{80}\rangle)]^{-1}} \equiv f(\langle x_2\rangle,\langle x_3\rangle,\langle x_4\rangle,\langle x_{80}\rangle) \tag{9}$$

where we have explicitly incorporated the parameterization proposed in equation (1) and where we have assumed that the different copies of each promoter act in a correlated way due to the effect of the different proteins involved.

Taking into account that *GAL4* is not subject to regulation, the argument above implies that the dynamics of the system can be approximated by the system

$$\begin{cases} \dot{x_2} = \theta_2[\lambda+(1-\lambda)f(x_2,x_3,x_4,x_{80})]-\gamma x_2 \\ \dot{x_3} = \theta_3[\lambda+(1-\lambda)f(x_2,x_3,x_4,x_{80})]-\gamma x_3 \\ \dot{x_4} = \theta_4-\gamma x_4 \\ \dot{x_{80}} = \theta_{80}[\lambda+(1-\lambda)f(x_2,x_3,x_4,x_{80})]-\gamma x_{80} \end{cases} \tag{10}$$

where we have dropped the angled brackets to simplify notation.

In steady state, in order to obtain a self-consistent solution, the following set of algebraic equations must be satisfied:

$$\begin{cases} 0 = \theta_2[\lambda+(1-\lambda)f(x_2,x_3,x_4,x_{80})]-\gamma x_2 \\ 0 = \theta_3[\lambda+(1-\lambda)f(x_2,x_3,x_4,x_{80})]-\gamma x_3 \\ 0 = \theta_4-\gamma x_4 \\ 0 = \theta_{80}[\lambda+(1-\lambda)f(x_2,x_3,x_4,x_{80})]-\gamma x_{80} \end{cases} \tag{11}$$

We note that these equations imply that in equilibrium the concentrations of *GAL2*, *GAL3*, and *GAL80* will be proportional to each other through their relative transcriptional strengths,

$$\frac{x_2}{\theta_2}=\frac{x_3}{\theta_3}=\frac{x_{80}}{\theta_{80}}, \tag{12}$$

which allows us to reduce this set of relations to just one equation:

$$0 = \theta_3\left[\lambda+(1-\lambda)f\left(\frac{\theta_2}{\theta_3}x_3,x_3,\frac{\theta_4}{\gamma},\frac{\theta_{80}}{\theta_3}x_3\right)\right]-\gamma x_3. \tag{13}$$

Solving this equation for x_3 and then computing $f\left(\frac{\theta_2}{\theta_3}x_3,x_3,\frac{\theta_4}{\gamma},\frac{\theta_{80}}{\theta_3}x_3\right)$ allows us to obtain an approximation for the fraction of actively transcribing cells in a given population. We

solved this equation by bisection search considering as extrema the minimum and maximum possible *GAL3* concentrations that can be achieved under this scheme ($\lambda\theta_3/\gamma$ and θ_3/γ respectively).

In Fig. S6D, we compare the results obtained from detailed simulations of the stochastic process specified in Section V to the analytical approximation proposed in this section for a range of galactose values that includes those used in the experiments presented in this article. We observe reasonable agreement, which supports the usefulness of the approximation proposed as a proxy for studying the behavior of the system in a manner that is less taxing from the computational point of view.

VIII. Fitting procedure and best fit results

We determined the set of parameters that best describes the data by confronting the measurements with the model predictions using a Bayesian inference approach (*10*). Briefly, we assumed that for a given set of parameters, the likelihood of observing each measurement follows a normal distribution centered on the value indicated by the model and with an estimated uncertainty of 10% which is representative of the repeatability of the experiments. Applying Bayes theorem, this defines a distribution over parameter space where each parameter set gets weighed according to its likelihood of representing the data. We sampled this distribution using a Metropolis-Hastings algorithm, which allowed us to obtain estimates of the parameter set that has the highest likelihood of being a good description of the data as well as the corresponding uncertainties.

The sampling algorithm was run by following 10 independent Markov chains starting from the point indicated in Table S3. For each parameter, normal distributions with widths as indicated in the same table and centered in the previous point were used as jump distributions. We also imposed lower and upper bounds as indicated but the chains stayed away from the boundaries except in the case of K_2, which we relate to the fact that the experimental system does not exhibit much sensitivity to changes in the dosage of *GAL2* in the conditions explored. Each chain was followed for 10000 iterations and only the second half of the simulations was used to draw inferences.

Parameter		Sampling parameters			
symbol	Unit	starting point	Jump width	lower bound	upper bound
K_2	proteins	1000	200	0.1	4000
K_g	proteins	0.03	0.01	0.002	1
K_3	proteins	2.0	0.5	0.1	10
K_{80}	proteins	8.0	0.5	0.3	200
K_4	proteins	1.0	2.5	0.1	200
μ	-	0.50	0.25	0.05	20
ν	-	1.00	0.20	0.05	20
α	-	0.85	0.02	0.04	20
β	-	5.0	0.5	0.05	50
η	-	1.5	0.2	0.05	20

Table S3. Summary of the parameters used in the fitting procedure.

In Fig. S7, we show the inferred distributions for each parameter and in Table S4 we report first order statistics that describe the inferred values for each fit parameter.

Parameter		Inferences		
Symbol	Unit	Best	Mean	std
K_2	proteins	600	1700	1000
K_g	proteins	0.052	0.040	0.008
K_3	proteins	4.1	2.9	0.7
K_{80}	proteins	8	8	2
K_4	proteins	2	8	6
μ	-	0.9	0.6	0.2
ν	-	1.3	1.5	0.2
α	-	0.85	0.82	0.02
β	-	6.1	6.6	0.7
η	-	1.6	1.8	0.2

Table S4. Summary of the inferred parameters statistics.

IX. Mathematical analysis of the properties of one-gene and two-gene networks

In order to investigate the necessary features that can make natural gene networks display dosage invariance, we consider a set of genes that are subject to a common regulation scheme. The general question that we are interested in addressing is: what conditions on the regulation scheme guarantee the activity of the transcriptional

center to be invariant to proportional changes in the transcription rate of all the genes involved? Such changes would be produced as an organism undergoes a change in ploidy, as chromosomes are replicated throughout cell cycle, in genome-wide duplication or loss events and/or by global noise in gene expression.

We consider that each gene is transcribed and then translated proportionally to the activity of its transcription center (a number between 0 and 1 that might represent, for instance, the fractional occupancy of active promoter sites), the proportionality constant being the maximal transcriptional rate associated with the gene. We assume that all proteins are effectively degraded at the cell-division rate (γ), thinking about a situation in which the lifetime of the proteins is much longer than the cell-division time. We further consider that proteins generated off each gene interact with each other on fast timescales and that this interaction defines the state of the transcriptional center. Finally, we consider a mean field approximation in the sense defined in Section VII.

Under these conditions, we describe the time evolution of the concentrations of the relevant proteins by the following set of differential equations:

$$\begin{cases} \dfrac{dx_1}{dt} = \theta_1 f(\rho, x_1, \cdots, x_N) - \gamma x_1, \\ \qquad\qquad \vdots \\ \dfrac{\mathrm{d}x_N}{\mathrm{d}t} = \theta_N f(\rho, x_1, \cdots, x_N) - \gamma x_N. \end{cases} \tag{14}$$

In these equations, x_i represents the average total concentration of the protein coded by the i-th gene, θ_i is the transcriptional strength associated with it, γ is the dilution rate, ρ is some external control parameter (*e.g.*, galactose in the case of GAL network) and $f(\rho, x_1, \cdots, x_N)$ is a dimensionless quantity that takes values in [0, 1] and represents the activity level of the transcriptional system under consideration. We can think that f represents the fraction of active promoter sites. Under the context of this framework, one approach to formalize the question we are interested in addressing is to ask: what family of functions f describe systems in which the steady state value of f is invariant to proportional changes in all the θs for a wide range of the control parameter ρ (*i.e.*, exploring the full range of values of f, so as to get an inducible system).

One-dimensional case

Let's first consider the simplest possible case: a network with just one gene. At steady state, we have

$$\theta f(\rho, x) = \gamma x. \tag{15}$$

Mathematically, in order for a function to be invariant with respect to a variable, its derivative with respect to that variable has to be zero. Taking the derivative of the above expression with respect to θ we get

$$f + \theta \frac{\partial f}{\partial x}\frac{\mathrm{d}x}{\mathrm{d}\theta} = \gamma \frac{\mathrm{d}x}{\mathrm{d}\theta} \tag{16}$$

from where we obtain that $\frac{\mathrm{d}f}{\mathrm{d}\theta} = \frac{\partial f}{\partial x}\frac{\mathrm{d}x}{\mathrm{d}\theta} = \frac{f\frac{\partial f}{\partial x}}{\gamma - \theta\frac{\partial f}{\partial x}}$.

We conclude that for the system to be invariant for nontrivial fractions ($f \neq 0$) we need $\frac{\partial f}{\partial x} = 0$ at the value of x that solves the steady state equation. But if the system is to be inducible, we should assume that the value of x will change as we change ρ and therefore in order to get invariance across a range of induction conditions we need f to be independent of x, i.e., we cannot have feedback at all.

This means that the only possible way of getting an invariant system with just one species is if the system is not autoregulated, which makes the situation trivial: if the state of a promoter is not affected by the proteins it codes for, its fractional occupancy will be invariant to changes in its transcriptional strength. This situation corresponds to the case of a constitutively regulated gene. Having more copies of that gene in the cell is not expected to impose any change in the state of its constitutive promoter.

Two-dimensional case

Now we consider a network composed of two genes. The system under consideration is represented by the following set of differential equations:

$$\begin{cases} \frac{dx_1}{dt} = \theta_1 f(\rho, x_1, x_2) - \gamma x_1, \\ \frac{\mathrm{d}x_2}{\mathrm{d}t} = \theta_2 f(\rho, x_1, x_2) - \gamma x_2. \end{cases} \tag{17}$$

and we are interested in studying how the steady state value of f will be affected by proportional changes in θ_1 and θ_2.

Let's first note that using the same regulation scheme for the two genes imposes the condition that at steady state we must have (all variables represent steady state values from now on)

$$\frac{\theta_1}{\theta_2} = \frac{x_1}{x_2} \tag{18}$$

which implies that whatever change x_1 might undergo, x_2 is going to suffer a proportional modification as well. To study system behavior with respect proportional changes in θ_1 and θ_2, we introduce an additional parameter (δ) in the following way:

$$\begin{cases} (1+\delta)\theta_1 f(\rho, x_1, x_2) = \gamma x_1, \\ (1+\delta)\theta_2 f(\rho, x_1, x_2) = \gamma x_2, \end{cases} \tag{19}$$

which allows us to vary the transcriptional rates in a proportional manner and to explore how the value of f is affected by such changes.

Taking derivatives of both sides of the first equation in (19) with respect to δ, we obtain

$$\theta_1 f + (1+\delta)\theta_1 \left[\frac{\partial f}{\partial x_1}\frac{\mathrm{d}x_1}{\mathrm{d}\delta} + \frac{\partial f}{\partial x_2}\frac{\mathrm{d}x_2}{\mathrm{d}\delta}\right] = \gamma \frac{\mathrm{d}x_1}{\mathrm{d}\delta}. \tag{20}$$

Using equation (18) relating x_1 to x_2 at steady state, we can write $\frac{\mathrm{d}x_2}{\mathrm{d}\delta} = \frac{\theta_2}{\theta_1}\frac{\mathrm{d}x_1}{\mathrm{d}\delta}$ and plugging this expression into (20) we can solve the resulting equation for $\frac{\mathrm{d}x_1}{\mathrm{d}\delta}$:

$$\frac{\mathrm{d}x_1}{\mathrm{d}\delta} = \frac{\theta_1 f}{\gamma - (1+\delta)\left(\theta_1 \frac{\partial f}{\partial x_1} + \theta_2 \frac{\partial f}{\partial x_2}\right)} \tag{21}$$

where everything is evaluated at steady state. This implies that the change in f due to some small change in δ is proportional to

$$\frac{\mathrm{d}f}{\mathrm{d}\delta} = \frac{\left(\theta_1 \frac{\partial f}{\partial x_1} + \theta_2 \frac{\partial f}{\partial x_2}\right) f}{\gamma - (1+\delta)\left(\theta_1 \frac{\partial f}{\partial x_1} + \theta_2 \frac{\partial f}{\partial x_2}\right)}. \tag{22}$$

We conclude that for the system to be invariant with generality we need to satisfy

$$\theta_1 \frac{\partial f}{\partial x_1} + \theta_2 \frac{\partial f}{\partial x_2} = 0 \tag{23}$$

at steady state, but this implies that the signs of $\frac{\partial f}{\partial x_1}$ and $\frac{\partial f}{\partial x_2}$ have to be different; *i.e.*, we need one activator and one inhibitor.

Therefore, a gene circuit with two components that are regulated by the same transcriptional machinery requires components of opposite sign for the activity of the system to be invariant to network dosage. Contrary to the one-dimensional case, the genes here do not have to give up their feedback regulation schemes. This describes a minimal condition necessary to build dosage-invariant phenotypes into gene networks. In the main body of the paper we discuss additional requirements related to the topology of the underlying network and the effective stoichiometry of intermediate interactions.

Analysis of alternative two-dimensional topologies

As described in the main text, to further explore if certain wiring topologies of generic 2-component generic networks would make it easier or harder for cells to display network dosage invariance, we performed numerical investigations on the possible network topologies in which an activator and an inhibitor are controlled by similar transcriptional machineries and analyzed their inducibility properties.

In order to do so we computed numerical approximations to the dynamical system representing the evolution of the overall concentrations of activator (a) and inhibiting (i) agents for systems with different regulation schemes. *i.e.* we numerically solved the equations

$$\begin{cases} \dot{a} = \theta_a[\lambda + (1-\lambda)f(g,a,i)] - \gamma a \\ \dot{i} = \theta_i[\lambda + (1-\lambda)f(g,a,i)] - \gamma i \end{cases} \tag{24}$$

over a time interval of $t = 24$ h and computed the value of f corresponding to the values of a and i achieved at that point. In these equations the functional form of f quantifies the regulation scheme of the system. We considered the possibilities described in Fig. 4A, in each of which the function f is parameterized by 4 numbers that quantify the scale of action of the activator (S_a), the scale of action of the inhibitor (S_i), the nonlinearity with which the activator acts on downstream targets (α) and a corresponding parameter for the inhibitor (β). The parameter g in these equations represents the concentration of an

external activator of the system and, in analogy to how galactose affects the *GAL* system in yeast, we introduced its effect by the effective rescaling of the scale of action of the activator. We note that in the chosen parameterization this value is a dimensionless quantity. In the case of the network topology presented in the center column of Fig. 4, under some conditions this procedure yielded multiple solutions and we averaged them across ten random initial conditions (a_0, i_0) distributed uniformly across a range of values enclosing the possible physiological steady states that a and i could attain ($a_0 \in [0, \theta_a/\gamma], i_0 \in [0, \theta_i/\gamma]$) to obtain a mean inducibility level for each galactose concentration.

To investigate which of these systems had the property of both network dosage invariance and inducibility we randomly sampled the parameters characterizing the regulating functional forms over large ranges (Table S5) and fed them into the quantitative model to obtain numerical inducibility curves corresponding to the networks carrying one or two copies of the network genes. For each pair of these numerical curves, we calculated the level of dosage invariance by quantifying the area between the two curves, large areas corresponding to large penalties to network-dosage invariance, and vice versa. We also quantified the relative inducibility levels of our numerical curves relative to a reference induction profile (Fig. S8A). Large differences from the reference curve corresponded to large penalties to inducibility. A comparative examination of the dot-plots for each network configuration reveals that the topologies at left and right allow their host networks to be both dosage-invariant and inducible (Fig. 4B). The specific interaction scheme in the two networks is essential for the systems to display such behavior (Fig. 4A, left and right panels). However, the choice between activator and inhibitor in directly influencing the transcription is not essential as long as the effect of the other is indirect.

Sampled parameters			
Parameter	Sampling	Lower bound	Upper bound
S_a	logarithmic	10^{-2}	10^{2}
S_i	logarithmic	10^{-3}	10^{1}
α	linear	0.5	5
β	linear	0.5	5

Fixed parameters	
Parameter	Value
θ_a	1500
θ_i	1500
γ	0.46
λ	0.2

Table S5. Parameters used for sampling generic two-component functional forms.

The green areas in Fig. 4B enclose the parameters corresponding to dosage-invariant and inducible networks (low penalties in both axes). For each point populating the green regions, we extracted out the values of the set of 4 parameters (Fig. S8B, S8C). The parameter that quantifies the nonlinearity of the interaction between the inhibiting and activating agents (α in Fig. 4C and β in Fig. 4D) was the only one severely restricted in the values it could take; with values displaying a narrow distribution centered around 1. This finding suggests a further requirement on the network architecture: the effective stoichiometry of the interaction between the activating and inhibiting agents has to be 1-to-1 in order to produce a system that is both inducible and network-dosage invariant.

To further understand the stoichoimetry requirement, it proves useful to look at how the functional form in the left panel of Fig. 4A becomes compensated for parallel changes in a and i when α equals 1. In this case the functional form reduces to

$$f = \frac{1}{1 + \left[\frac{S_i i}{1 + S_a g a}\right]^{\beta}}. \tag{25}$$

In the limit $S_a g a \gg 1$, which we can interpret as identifying a situation with a sufficient amount of activator, we are allowed to make the approximation

$$f \cong \frac{1}{1 + \left[\frac{S_i i}{S_a g a}\right]^{\beta}}. \tag{26}$$

Let's recall that for values of λ of order 1, a is typically of order θ_a/γ and so the condition above can also be expressed in terms of just parameters of the system as $g \gg \frac{\gamma}{S_a \theta_a}$.

Let's now note that in steady state, the ratio between the overall levels of the inhibitor and activator is fixed to be $i/a = \theta_i/\theta_a$ as both players are regulated by the same scheme (see equation (18)). Therefore we obtain the simple relationship

$$f \cong \frac{1}{1 + \left[\frac{S_i}{S_a g}\frac{\theta_i}{\theta_a}\right]^{\beta}}. \tag{27}$$

and we can see that the value of f only depends on the ratio of the transcriptional strengths, *i.e.* the system is network dosage invariant. We further note that the system is inducible as it is still fully dependent on g, with potential for achieving the full dynamic

range of f. In other words by changing the value of g it is possible to explore values of f in the full range between 0 and 1. This expression also reveals the typical scale at which the external inducer has a significant effect on the state of the system:

$$f \cong \frac{1}{1+\left[\frac{g_{\text{typ}}}{g}\right]^{\beta}} \text{ with } g_{\text{typ}} = \frac{S_i}{S_a}\frac{\theta_i}{\theta_a}. \tag{28}$$

In Fig. S8D we show how this restriction appears in the numerical exploration we performed, by showing the distribution of g_{typ} over the points that belong to the green regions in Fig. 4B. We note that the distribution is centered close to the midpoint of the curve used as a reference to quantify inducibility (Fig. S8A).

Finally we would like to point out that the condition expressed by equation (23) is satisfied by any function of the form $f = f(g, x_1/x_2)$, *i.e.* any function that depends on the concentration of the two relevant players through their ratio. This allows one to understand why the regulation schemes depicted at the left and at the right of Fig. 4A can be both inducible and network-dosage invariant and why the system in the center does not have this property. In the case in which the regulation is such that the activator and the inhibitor act independently of each other the only way in which such a dependence on the ratio of the concentrations can be achieved is if $S_a g a \ll 1$, $S_i i \gg 1$ and $\alpha = \beta$, which yields

$$f \cong \left[\frac{S_a g a}{S_i i}\right]^{\alpha=\beta}. \tag{29}$$

We can see how in this limit even though the system would be network dosage invariant it can not be fully inducible as this approximation can not be valid for large values of g, as that would yield unphysical values for f.

SUPPLEMENTARY INFORMATION REFERENCES

1 Guldener, U. *et al*. A new efficient gene disruption cassette for repeated use in budding yeast. *Nucleic Acids Res* **24**, 2519 (1996).

2 Goldstein, A. L. & McCusker, J. H. Three new dominant drug resistance cassettes for gene disruption in Saccharomyces cerevisiae. *Yeast* **15**, 1541 (1999).

3 Kaufmann, B. B., Yang, Q., Mettetal, J. T. & van Oudenaarden, A. Heritable stochastic switching revealed by single-cell genealogy. *PLoS Biol* **5**, e239 (2007).

4 Raj, A., Peskin, C. S., Tranchina, D., Vargas, D. Y. & Tyagi, S. Stochastic mRNA synthesis in mammalian cells. *PLoS Biol* **4**, e309 (2006).

5 Gasch, A. P. *et al*. Genomic expression programs in the response of yeast cells to environmental changes. *Mol Biol Cell* **11**, 4241-4257 (2000).

6 Ghaemmaghami, S. *et al*. Global analysis of protein expression in yeast. *Nature* **425**, 737-741 (2003).

7 Gardiner, C. W. *Handbook of stochastic methods for physics, chemistry, and the natural sciences*. 3rd edn, (Springer-Verlag, 2004).

8 van Kampen, N. G. *Stochastic processes in physics and chemistry*. Rev. and enl. edn, (North-Holland, 1992).

9 Pedraza, J. M. & van Oudenaarden, A. Noise propagation in gene networks. *Science* **307**, 1965-1969 (2005).

10 Gelman, A. *Bayesian data analysis*. 2nd edn, (Chapman & Hall/CRC, 2004).

SUPPLEMENTARY FIGURE CAPTIONS

Fig. S1. Construction scheme of the first-order dosage-varied yeast strains. Each rectangle denotes a diploid strain that was dosage-halved in one of the four regulatory genes of the network.

Fig. S2. Measurements of the network inducibility level as a function of galactose concentration for all the different strains analyzed. Inducibility is quantified as the fraction of ON-cells in a bimodal population. The genetic background of each strain is specified by a square on the upper left corner. Each big square contains four small squares that represent the four regulatory genes of the GAL network (top-left: *GAL2*, top-right: *GAL3*, bottom-left: *GAL4*, bottom-right: *GAL80*). Grey (white) color marks the presence of two (one) copies of a specific gene. A line between two strains indicates that the two genetic backgrounds differ by a single copy of a specific gene and the color of the line codifies that gene (blue for *GAL3*, red for *GAL80*, green for *GAL2*, and orange for *GAL4*). In all cases except for the wild type, the measurements that correspond to the wild type were included as a reference (gray trace). The strongest second-order compensation was observed when both *GAL3* and *GAL80* were dosage-perturbed in the same strain. At the third-order, irrespective of the second-order genetic background on which the dosage of a third gene was reduced, halving the dosage of *GAL3* (*GAL80*) always decreased (increased) the inducibility levels. On the other hand, varying the dosage of *GAL2* had a neutral effect on inducibility and halving the dosage of *GAL4* decreased it slightly.

Fig. S3. Quantification of the effect of each gene. (**A**) Contribution of the second copy of each regulatory gene to network inducibility quantified as the squared difference (χ^2) between inducibility levels in strains that differ in one copy of the corresponding gene, averaged across different galactose concentrations and genetic backgrounds. (**B**) Signed contribution of the second copy of each regulatory gene to network inducibility quantified as the signed difference (Δ) between inducibility levels in strains that differ in one copy of the corresponding gene, averaged across different galactose concentrations and genetic backgrounds. Fig. S3B is the same as Fig. 3C and it is included here for reference purposes.

Fig. S4. The genetic switch between the OFF and ON states of the GAL network. ON-cells express (blue dotted arrows) both the positive (Gal2p and Gal3p) and the negative (Gal80p) regulators of the network while the OFF-cells' gene expression does not exceed basal levels. Gal4p is a constitutively expressed protein. Red arrows with pointed and blunted ends reflect the positive and negative effect of one network component on another, respectively.

Fig. S5. Stochastic simulations of the proposed model. (**A-F**) Left subpanels: traces of the different variables (number of promoter in an ON state, P_{GALi}ON, and protein concentrations, *galip*) corresponding to one realization of the stochastic model proposed in Section V for a time of 24 h. Right subpanels: Distribution of the different variables after 24 h (left) and 48 h (right) across 100 independent realizations. The fact that these distributions are similar to each other indicates that the process has reached a steady state. In all cases, simulations were started from initial conditions that correspond to an OFF state, namely, all promoters were OFF initially and the initial protein concentrations were chosen as $\theta_i\lambda/\gamma$ for *GAL2*, *GAL3* and *GAL80* and as θ_4/γ for *GAL4*. The parameters used were those indicated in Sections VI and VIII and the value of h used in this set of simulations was $h \cong 2.5$ h^{-1}.

Fig. S6. Effect of h and comparison between stochastic simulations and the proposed analytical approximation. (**A-C**) Simulations similar to those described in fig. S5 but for a fixed galactose concentration of 0.1% w/v and for three different values of h. (**D**) Fraction of ON cells according to the analytical approximation (solid line) and the stochastic simulations (dots). In the case of the stochastic simulations, the fraction of ON cells was determined by dividing the *GAL3* expression profile to two regions separated by the gray line (fig. S5, fig. S6A-C, right subpanels) and counting the fraction of cells in the region that corresponds to higher expression levels. The gray line lies exactly in between the maximal and basal expression levels. Blue dots are the fractions obtained from 100 simulations at 24 h and the gray dots correspond to the results at 48 h.

Fig. S7. Inferences on the parameters that best describe the data. Metropolis-Hastings samples of the parameter values that best describe the data, according to a Bayesian-inference approach (see Section VIII).

Fig. S8. Numerical exploration of network dosage invariance and inducibility in two-component systems. (**A**) Inducibility curve used as a reference for the numerical exploration described in the last part of Section IX. (**B, C**) Histograms of the parameter values corresponding to the green regions shown in Fig. 4B for the left (B) and right (C) configurations described in Fig. 4A. The dotted lines show what one would expect if the parameters had no effect in determining whether the system was in the green region or not (uniform distributions over the sampled values). (**D**) Distribution of the ratio $g_{\mathrm{typ}} = \frac{\theta_i S_i}{\theta_a S_a}$ for the points in the green regions shown in Fig. 4B.

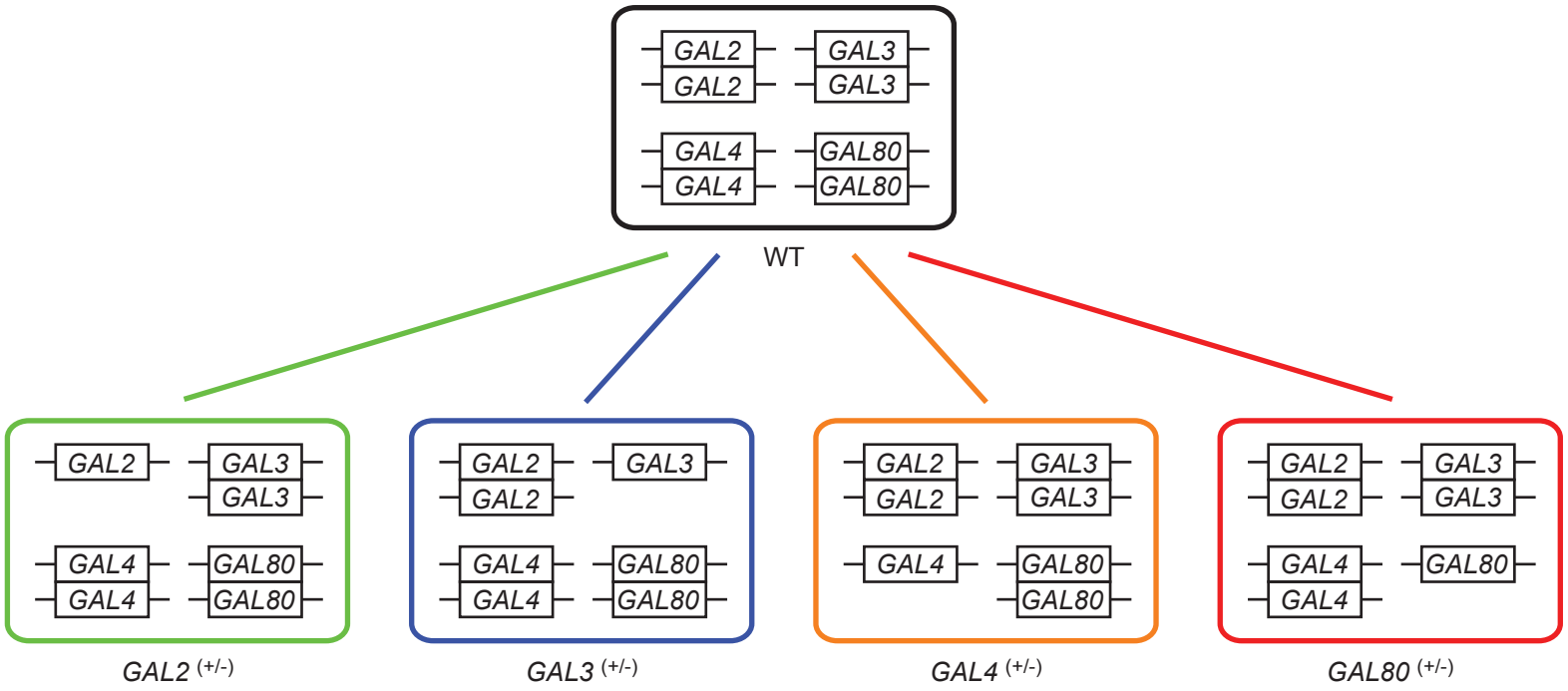
GAL2
GAL2
GAL3
GAL3
GAL4
GAL4
GAL80
GAL80
WT
GAL2
GAL3
GAL3
GAL4
GAL4
GAL80
GAL80
GAL2 (+/-)
GAL2
GAL2
GAL3
GAL4
GAL4
GAL80
GAL80
GAL3 (+/-)
GAL2
GAL2
GAL3
GAL3
GAL4
GAL80
GAL80
GAL4 (+/-)
GAL2
GAL2
GAL3
GAL3
GAL4
GAL4
GAL80
GAL80 (+/-)

Figure S1

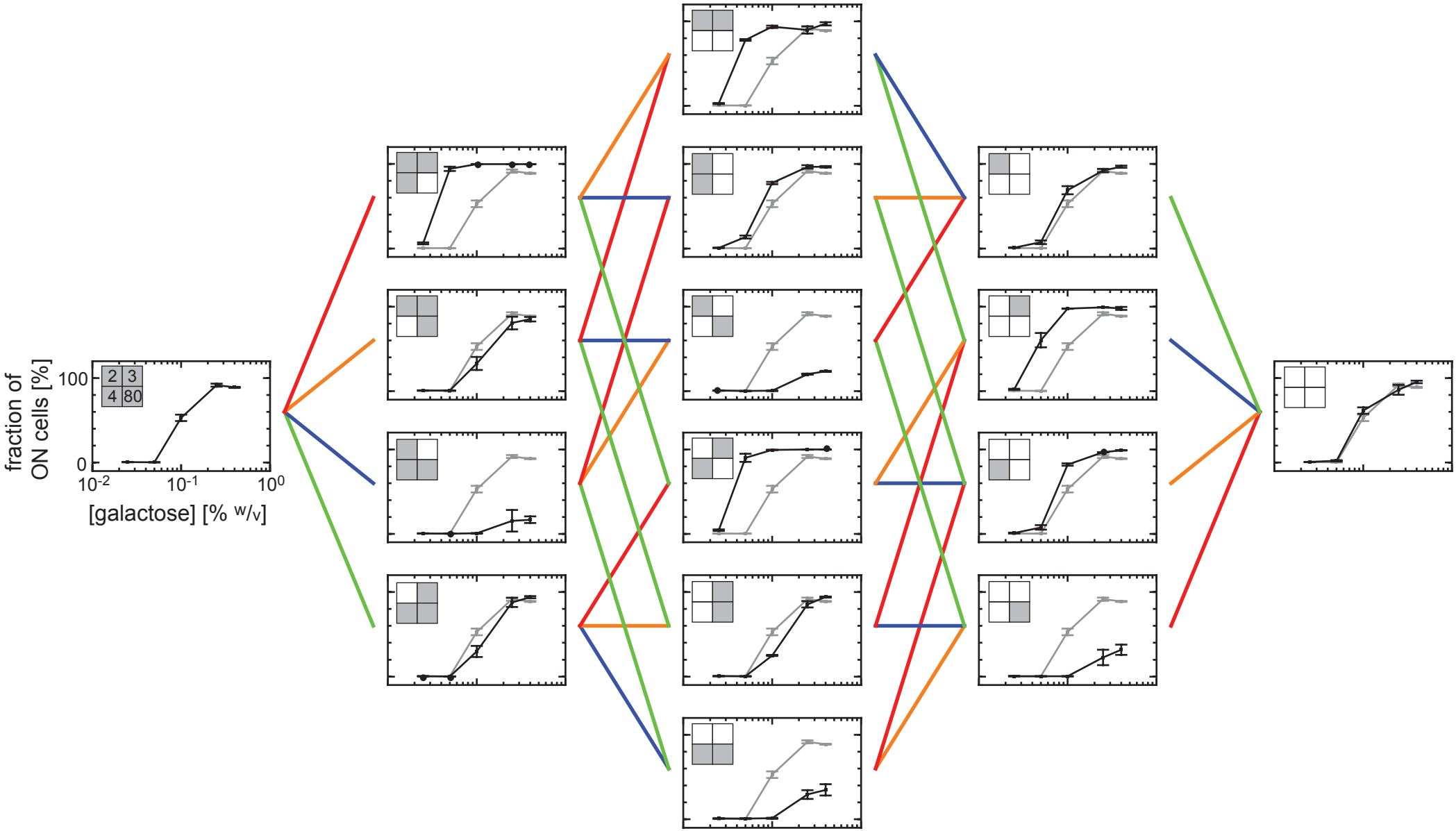

Figure S2

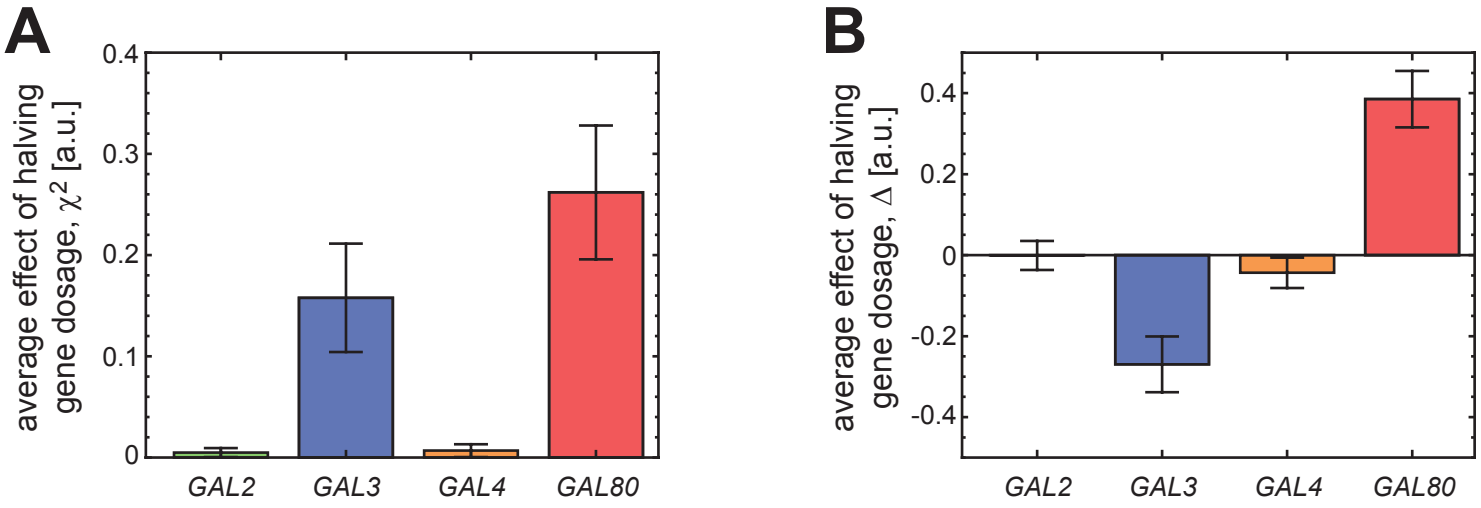
A
average effect of halving gene dosage, χ^2 [a.u.]
0.4
0.3
0.2
0.1
0
GAL2
GAL3
GAL4
GAL80
B
average effect of halving gene dosage, Δ [a.u.]
0.4
0.2
0
-0.2
-0.4
GAL2
GAL3
GAL4
GAL80

Figure S3

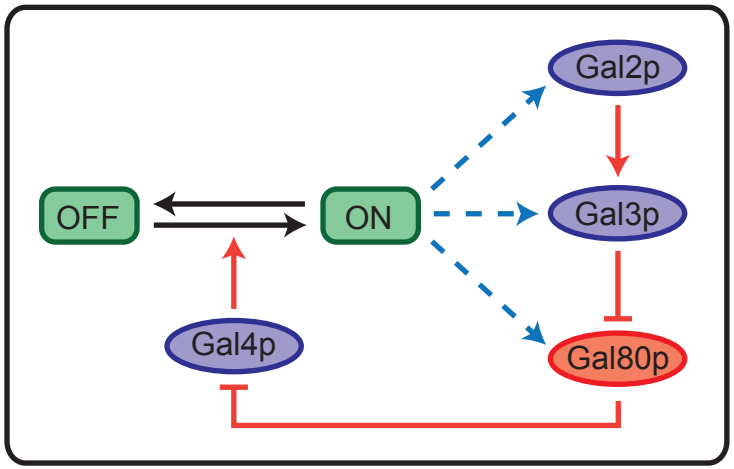

Figure S4

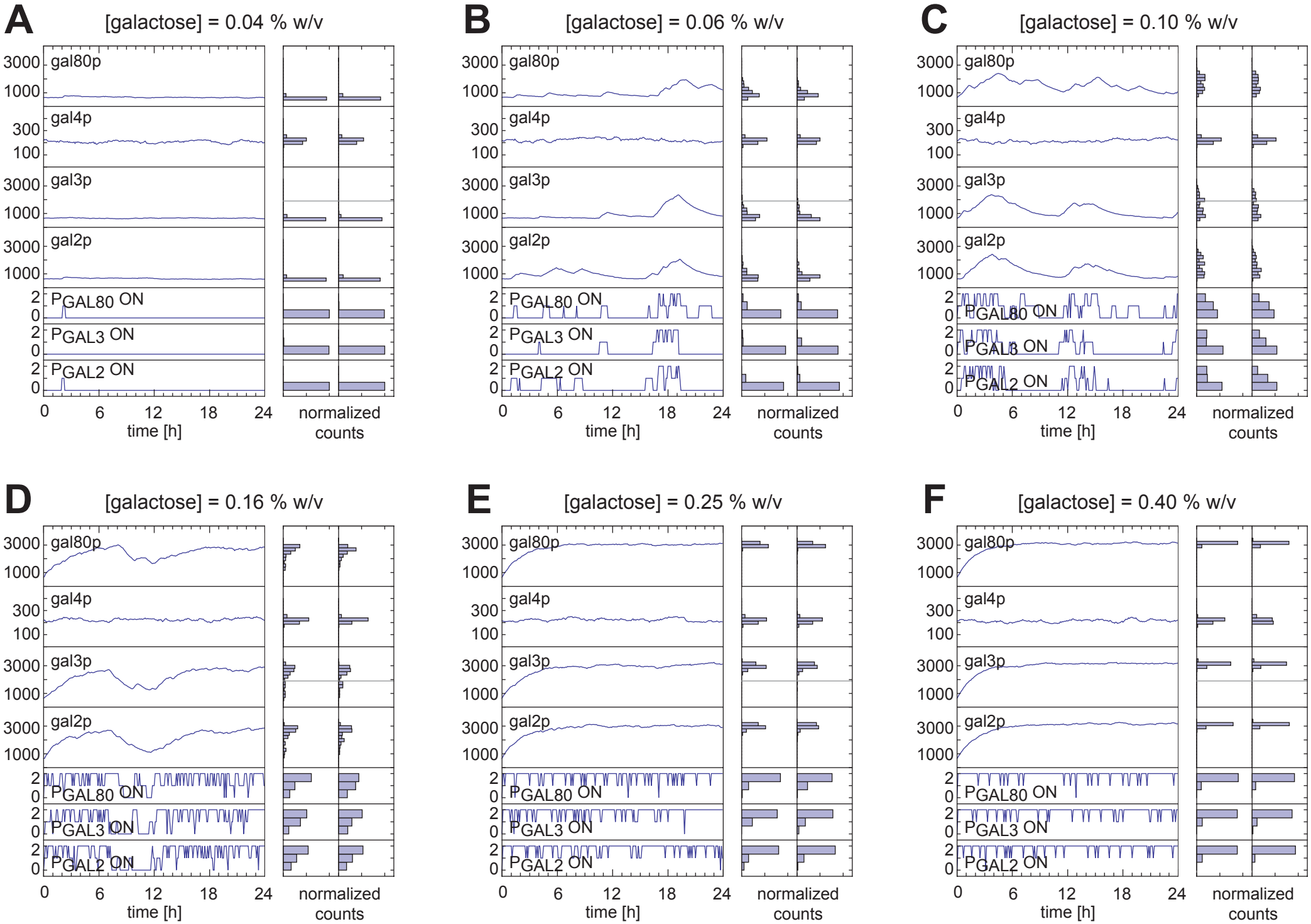
A
[galactose] = 0.04 % w/v
B
[galactose] = 0.06 % w/v
C
[galactose] = 0.10 % w/v
D
[galactose] = 0.16 % w/v
E
[galactose] = 0.25 % w/v
F
[galactose] = 0.40 % w/v
gal80p
gal4p
gal3p
gal2p
PGAL80 ON
PGAL3 ON
PGAL2 ON
3000
1000
300
100
2
0
0
6
12
18
24
time [h]
normalized counts

Figure S5

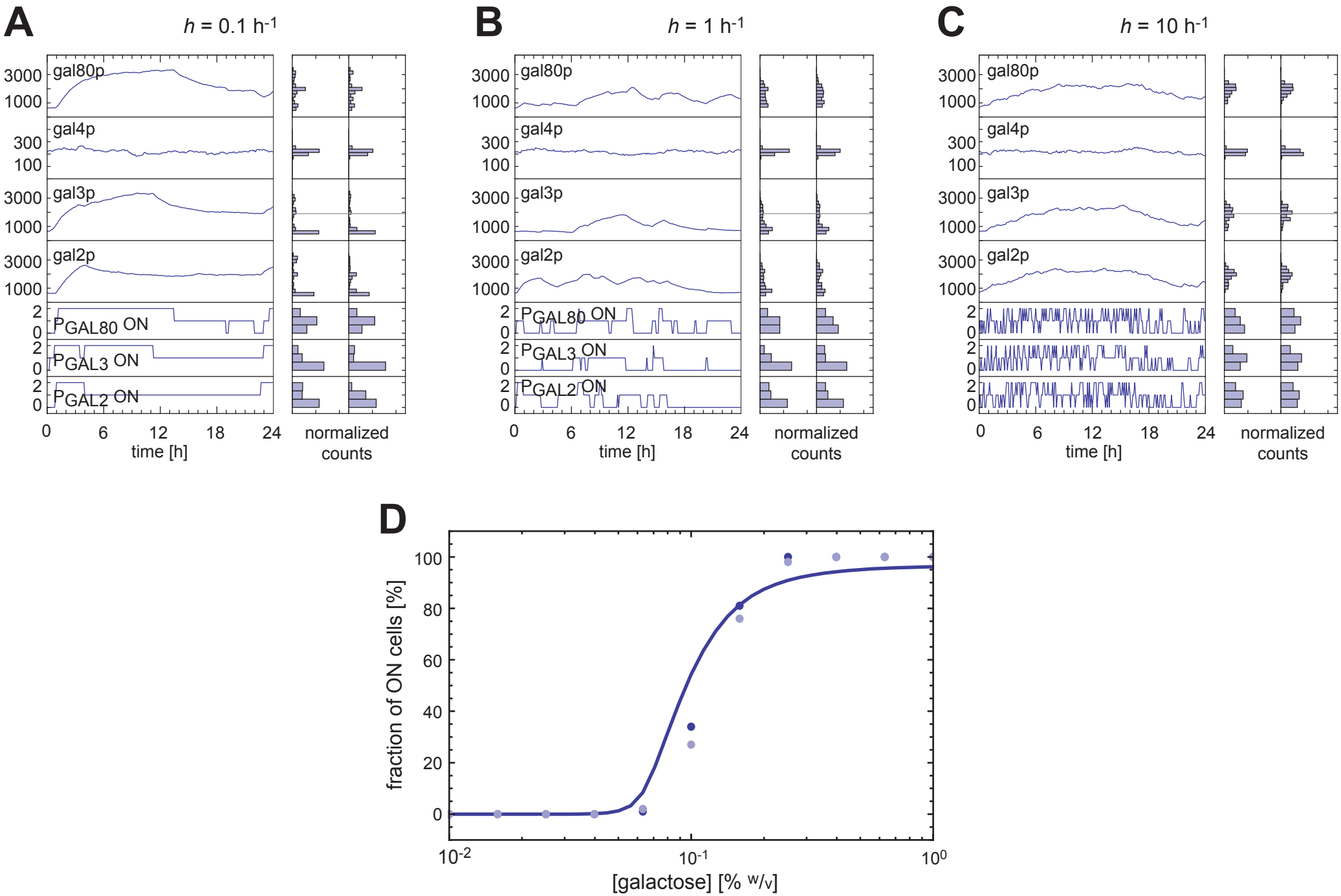
A
h = 0.1 h-1
B
h = 1 h-1
C
h = 10 h-1
gal80p
gal4p
gal3p
gal2p
PGAL80 ON
PGAL3 ON
PGAL2 ON
time [h]
normalized counts
D
fraction of ON cells [%]
[galactose] [% w/v]

Figure S6

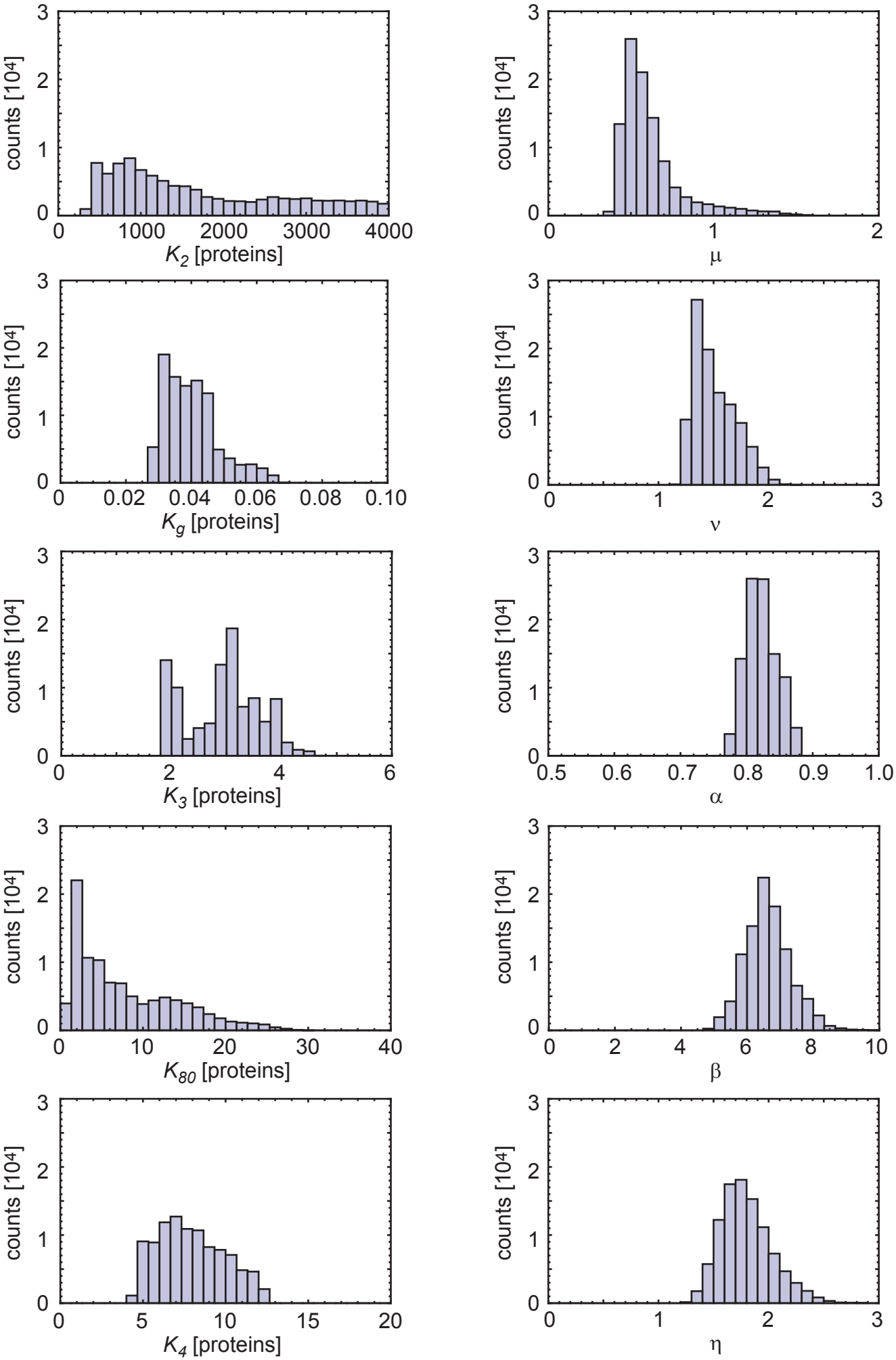
counts [10^4]
K2 [proteins]
μ
Kg [proteins]
ν
K3 [proteins]
α
K80 [proteins]
β
K4 [proteins]
η

Figure S7

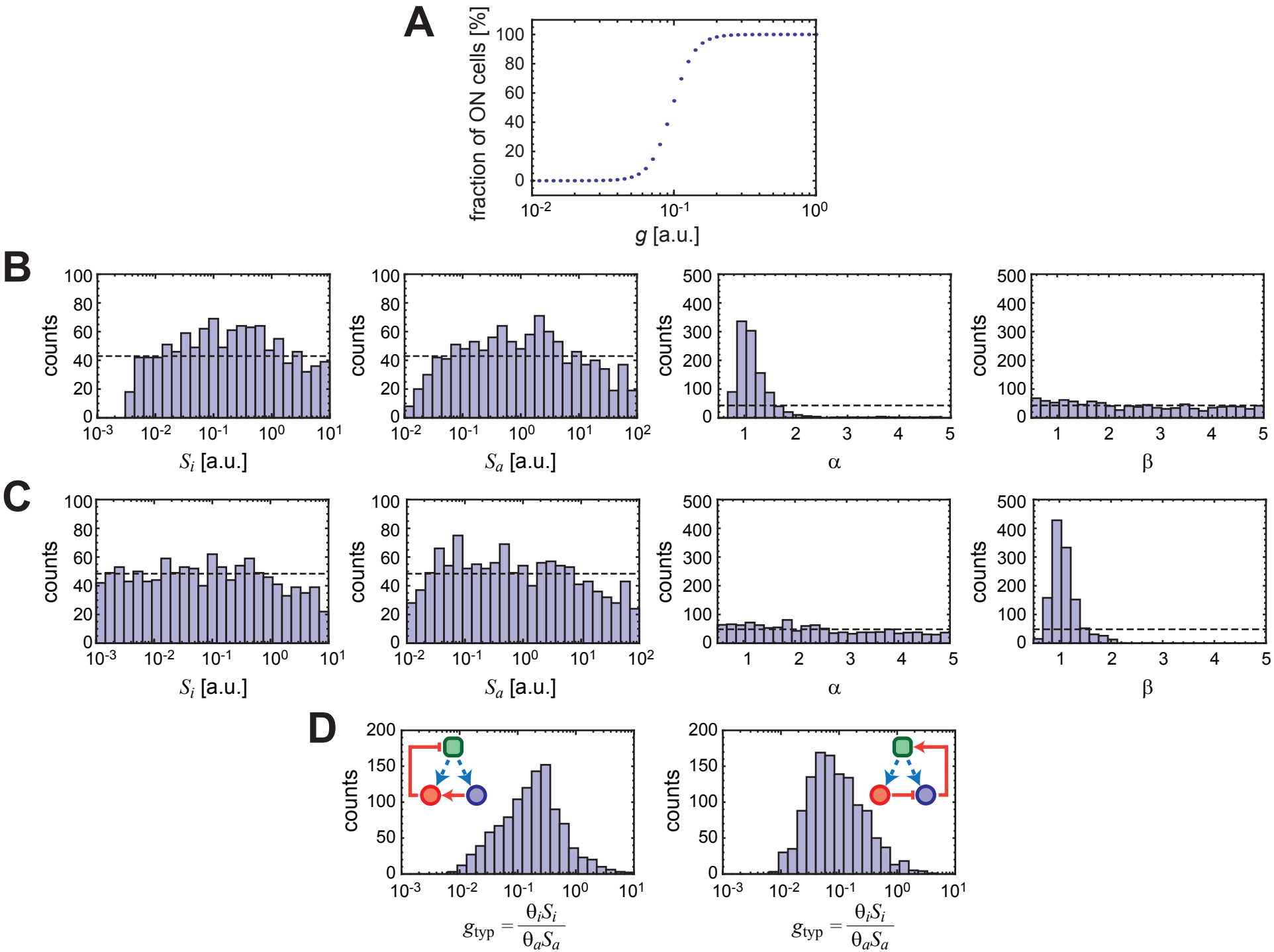
A
fraction of ON cells [%]
g [a.u.]
B
counts
S_i [a.u.]
S_a [a.u.]
α
β
C
counts
S_i [a.u.]
S_a [a.u.]
α
β
D
counts
$g_{typ} = \frac{\theta_i S_i}{\theta_a S_a}$
$g_{typ} = \frac{\theta_i S_i}{\theta_a S_a}$

Figure S8