RACIPE Sampling Methods

Parameters:
Maximum production rate (G)
Degradation rate (k)
Fold change (l)*
Threshold (X_0)
Hill coefficient (n)#

1.1 Defining the ranges of parameters for randomization

The RACIPE method generates an ensemble of models with different set of parameters, which are randomly sampled in a certain range following a certain distribution. The tool provides options to choose from three types of distributions: uniform distribution, rectified Gaussian distribution (its negative elements is reset to zero) and exponential distribution. Users can select one of them using the *"-dist"* option. The default ranges of parameters for randomization are as follows:

Parameters	Min to Max Values (Uniform)	Mean, Standard Deviation (Rectified Gaussian)+	Mean (Exponential)
Maximum production rate (G)	1-100	50.5, 49.5	50.5
Degradation rate (k)	0.1-1	0.55, 0.45	0.55
Fold change (l)*	1-100	50.5, 49.5	50.5
Threshold (X_0)	The ranges, which depend on the inward regulations, are estimated by a Monte Carlo simulation.		
Hill coefficient (n)#	1-6	3.5, 2.5	3.5

Table S1. Default ranges of the parameters for randomization. The ranges of threshold levels are estimated by a mean-field approximation considering all the inward regulations. More details are available in our previous paper (Huang *et al.*, 2017).

*For Uniform distribution, the fold change l of an inhibitory regulation ranges from 0.01 to 1. But the inverse of l is uniformly sampled from 1 to 100 instead of sampling l uniformly from 0.01 to 1. By doing so, we ensure that the average l is about 0.046, instead of ~ 0.5. Similar approaches are adopted for the Gaussian and exponential distributions.

#The value of the Hill coefficient is an integer from 1 to 6.

+The negative part is trimmed.

In addition, two assumptions are made in RACIPE to ensure that it generates a representative ensemble of models for a specific circuit topology. First, the maximum production rate of each gene should lie roughly within the same range (from 1 to 100 in this study, see S1 Table), as the maximum rate is determined by how fastest the transcriptional machinery can work. For a gene regulated by only one activator, the maximum production rate (G) is achieved when the activator is abundant, and thus the basal production rate of the gene $g = G/l_+$. For a gene regulated by only one inhibitor, the maximum rate (G) is achieved in the absence of the inhibitor, i.e. $g = G$. This criterion can be generalized to genes regulated by multiple regulators. Therefore, in

practice, we directly randomize the maximum production rate (G) instead of the basal production rate (g), and then calculate the value of g according to the above criterion. The ranges of these parameters are summarized in details in S1 Table. The RACIPE randomization procedure allows a gene to have a relative expression ratio of up to 1,000 for two sets of parameters, even when it is not regulated by other genes.

Now comes the question of how they sample the Threshold. This appears to be where the half-functional rule comes in.

Half-functional rule:

Second, we also assume that, for all the members of the RACIPE model ensemble, each regulatory link in the circuit should have roughly equal chance of being functional or not functional, referred to as the *half-functional* rule. For example, in the case that gene A regulates gene B, all the threshold parameters are selected in such a way that, for the RACIPE ensemble, the level of A at the steady states has roughly 50% chance to be above and 50% chance to be below its threshold level. Otherwise, if the threshold level is too large or too small, the regula- tory link is either not functional most of the time or constitutively active, thereby changing the effective circuit topology, and limiting the comprehensive understanding of circuit function (S2 Fig).

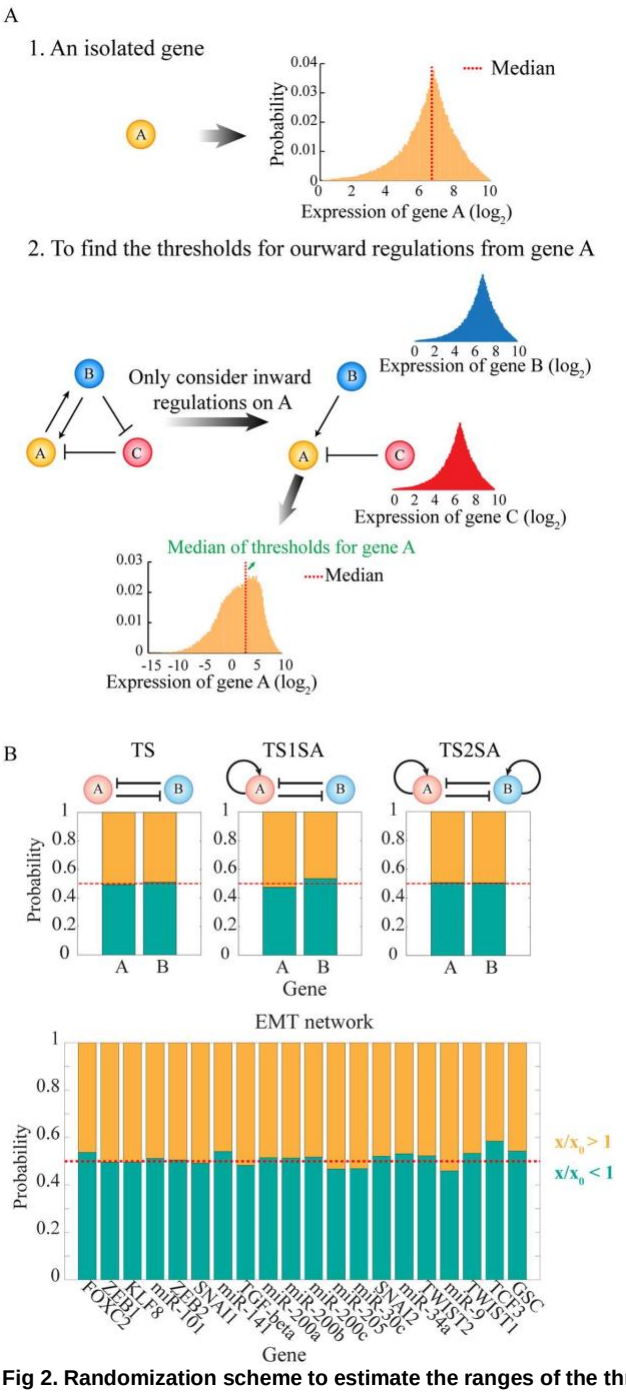
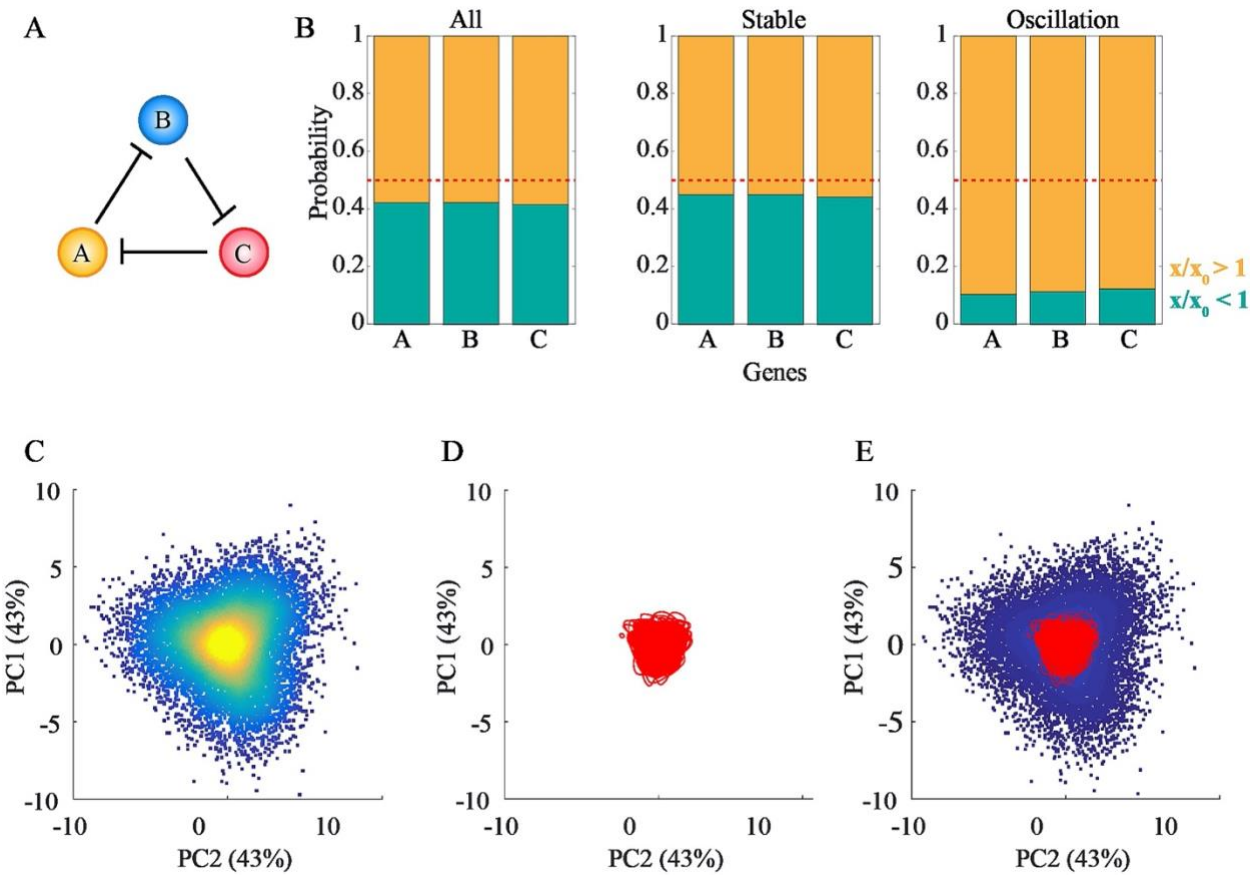
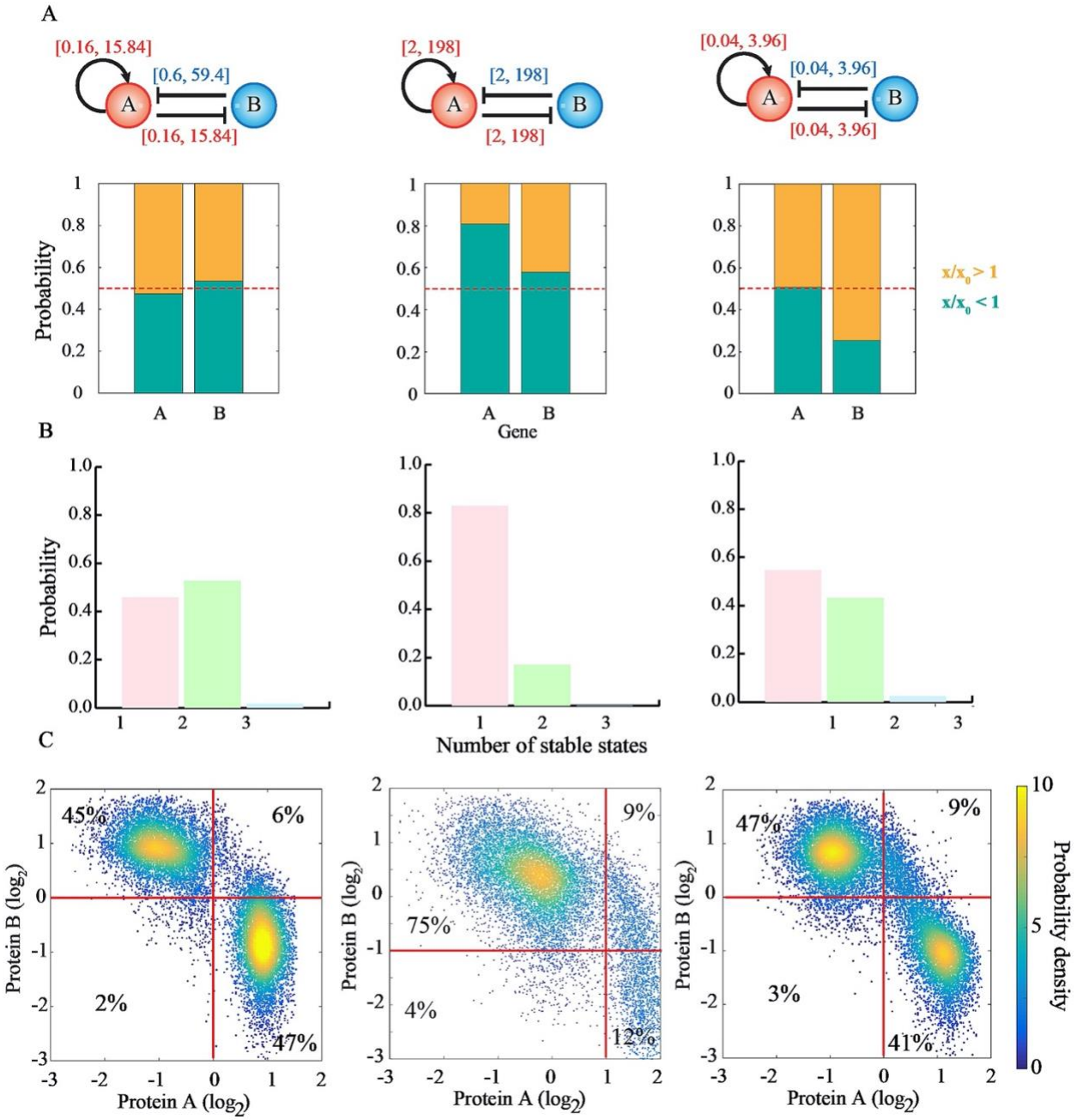


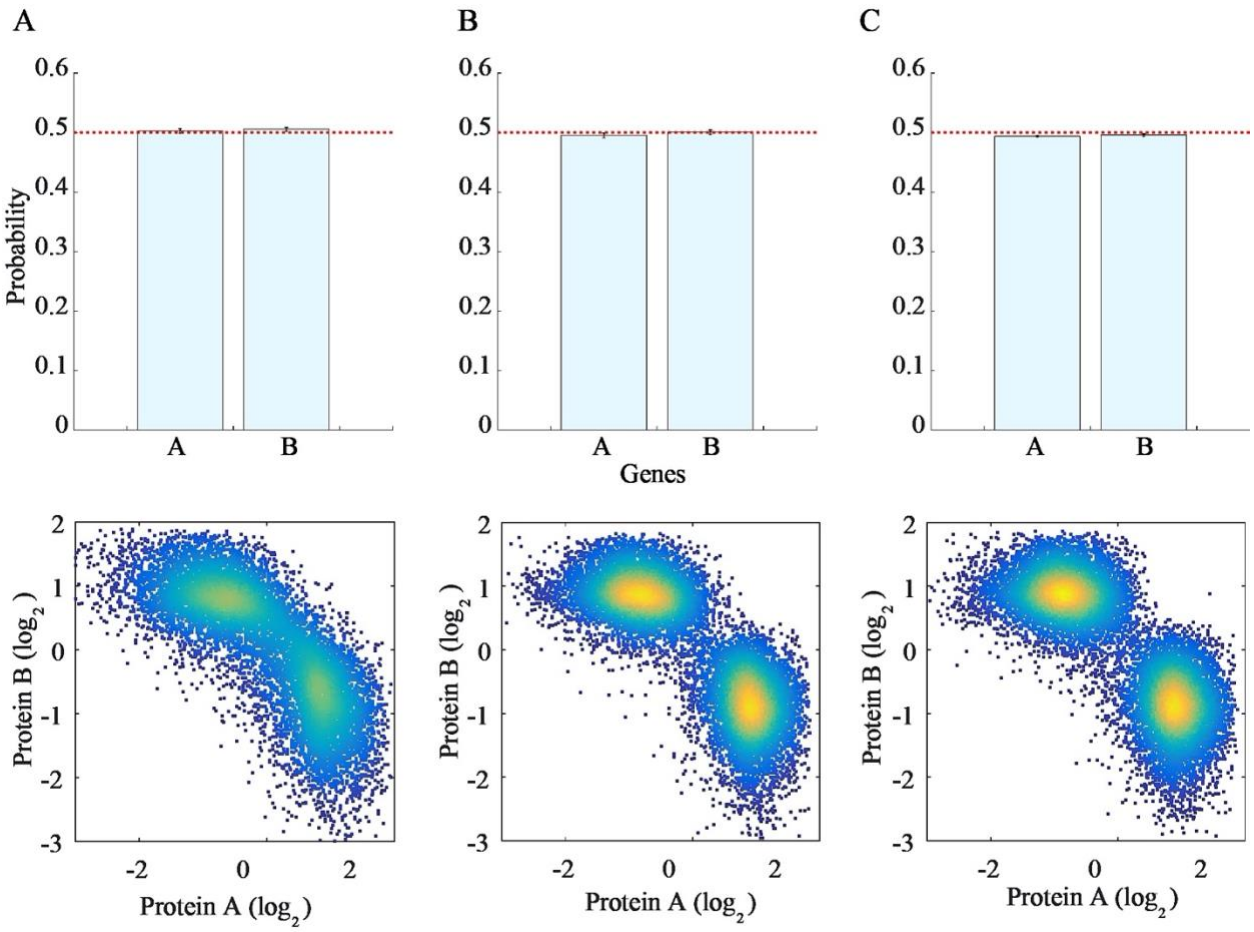
Fig 2. Randomization scheme to estimate the ranges of the threshold parameters. (A) Schematic of the procedure to estimate the ranges of the threshold parameters, so that the level of a regulator has 50% chance to be above or below the threshold level of each regulatory link ("half-functional rule"). First, for a gene A without any regulator, the RACIPE models are generated by randomizing the maximum production rate and the degradation rate according to S1 Table. The distribution of A level is obtained from the stable steady state solutions of all the RACIPE models (top left panel, yellow histogram). Second, for a gene A in a gene circuit, the distribution of A level is estimated only on the basis of the inward regulatory links (i.e. the B to A activation and the C to A inhibition in the bottom left panel). The distributions of the levels of the inward regulators B and C are assumed to follow the same distributions as a gene without any regulator (bottom left panel, blue and red distribution); the threshold levels for these inward links are chosen randomly from (0.02M to 1.98M), where M is the median of their gene expression distributions. Finally, the distribution of A level is obtained by randomizing all the relevant parameters. That includes the levels of B and C, the strength of the inward regulatory links (i.e., the threshold level, the Hill coefficient and the fold change), the maximum production rate and the degradation rate of A, and the threshold for any regulatory link starting from A is chosen randomly from (0.02M to 1.98M), where M is the median level of the new distribution of A level (orange in the bottom panel). The same procedure is followed for all other genes.



S1 Fig. Application of RACIPE to study oscillatory dynamics of a repressilator gene cir- cuit. (A) Illustration of a repressilator circuit with three genes, where each gene represses the next gene in the circuit. (B) Tests of the half-functional rule for all RACIPE models (leftmost panel), the models with stable steady states (middle panel), and the models with stable oscilla- tion (right panel).



S2 Fig. Tests of several random sampling schemes with and without the half-functional rule. (A) Test of the half-functional rule of a toggle-switch with one-sided self-activation where different ranges were used to randomize the threshold parameters. The leftmost panel shows the circuit and the sampling ranges of the threshold parameters by RACIPE. The middle and the rightmost panels show two examples where same ranges of the threshold parameters are used for all regulatory links. (B) Probability distributions of the number of stable steady states for each circuit. (C) Probability density maps of the gene expression data from all the RACIPE models, where the fraction of stable gene expressions in each quadrant is shown. (TIF)



S5 Fig. The effects of Hill coefficients on the robustness of the sampling scheme and gene states of a toggle-switch circuit motif. Test of the half-functional rule (top-panels) and 2D probability density map (bottom-panels) of RACIPE-generated gene expression data are shown for cases where Hill coefficients are randomized with different ranges—A: 1–3; B: 4–6; C: 7–9.

Regarding the half-functional rule, there are two functions in RACIPELIB.c of note. These are:

void estimate_threshold(int num, int ID, double minP, double maxP, double minK, double maxK, double minN, double maxN, double minF, double maxF, double *minT, double *maxT, struct topo *topoinfo, int dist, double SF)

and

void generate_random_range(FILE *f_paras, struct topo *topoinfo, struct opts *simu_opts)

estimate_threshold details the estimation of the thresholds. This is where the half-functional rule is employed. In the generate_random_range function, the estimate_threshold function is used in a for loop for each gene in the input circuit topology. These two functions represent the "confusing" parts of their randomization. For both functions, we need only look at Case 1, as this represents the uniform distribution employed for the results we are trying to reproduce.

Results:

7> _T_test.dat storing the test of the half functional rule for each RACIPE model.

Format of _T_test.dat:

Model_index Over_threshold_A Below_threshold_A ...

Over_threshold_A : The number of stable states for the current RACIPE model whose expression of gene A is larger than its threshold parameter of A.

Below_threshold_A : The number of stable states for the current RACIPE model whose expression of gene A is smaller than its threshold parameter of A.

For each model, the probability for gene A's expression to be larger than its threshold equals to the sum of Over_threshold_A across all the models divided by the sum of both Over_threshold_A and Below_threshold_A across all the models.

Sources

Huang, B., Jia, D., Feng, J. *et al.* RACIPE: a computational tool for modeling gene regulatory circuits using randomization. *BMC Syst Biol* **12,** 74 (2018). https://doi.org/10.1186/s12918-018-0594-6

Huang B, Lu M, Jia D, Ben-Jacob E, Levine H, et al. (2017) Interrogating the topological robustness of gene regulatory circuits by randomization. PLOS Computational Biology 13(3): e1005456. https://doi.org/10.1371/journal.pcbi.1005456