Electronic Supplementary Material (ESI) for Molecular BioSystems.

Synthetic data generation and quality check

1. SynTReN parameter settings

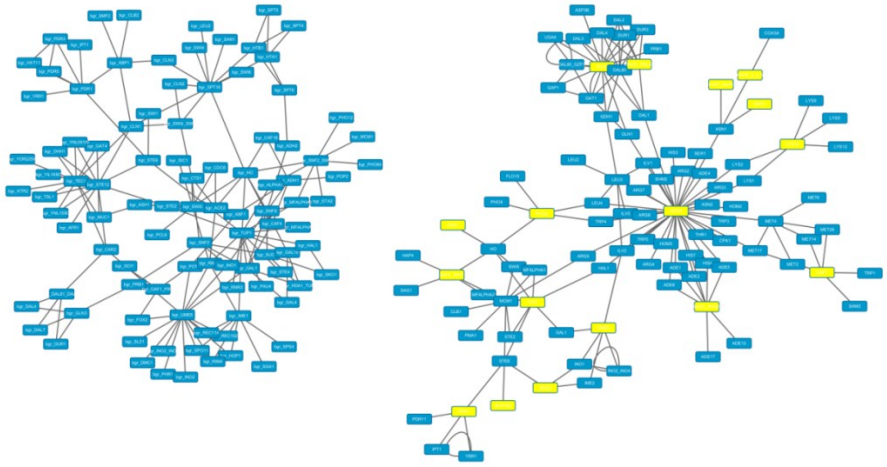

Fig S1-1. SynTReN simulated gene networks (*background* left, *foreground* right) and true regulators. True regulators are highlighted in yellow and listed in Table S1-1.

Table S1-1. True regulators of the synthetic data (yeast gene regulatory network).

LYS14
GCN4
DAL81_DAL82
GLN3
BAS1_PHO2
CBF1
PDR1
PHO2
SDS3
ALPHA2
SNF2_SWI1
HAP3
SWI3
TUP1
UME6
HAP2_HAP4
HAP2_3_4_5

Table S1-2. SynTReN settings for the synthetic data generation.

Parameter	Acronym	Range
number of replicates in each experimental group	*ngroup*	5
number of replicates per group	*nreplicate*	2,10,20,50,100
biological noise level in gene expressions	*bn*	0.1, 0.2, 0.3, 1.0
foreground gene number	-	100
background gene number	-	100
true regulator number	-	17

2. Quality check

AR (average correlation) is defined as the Pearson correlation coefficient (PCC) of the expressions across samples E_{g_i} and the sample-wise average expressions $\bar{E}$, where $E_g = (E_{g1}, E_{g2}, \ldots, E_{gS})$, S is the sample size, $\bar{E} = (\bar{E}_1, \bar{E}_2, \ldots, \bar{E}_S)$, $\bar{E}_s = \frac{1}{G}\sum_g E_{gs}$, G is gene number.

First, we tested the coexpression levels of the *foreground* using AR (Fig S1-2). Generally the *foreground* exhibits positive ARs, which fades gradually as the biological noise level rises. In comparison, the *background* genes do not show AR significantly deviating from zero. We hence regarded the *foreground* regulated genes as a single module and the *background* as false regulators.

Second, we checked noise levels of the *background* and *foreground* using the coefficient of variance (CV). As shown Fig S1-3, both have increased CVs as noise level rises. Specifically, in *foreground* (Fig S1-3A), the monotony of the CV-noise increment is restricted to the same *nreplicate* group (e.g. *nreplicate* = 20 or 100), and not preserved across multiple *nreplicate* groups; this is on the contrary to CVs in *background* (Fig S1-3B). This reflects SynTReN's limitation in controlling appropriately the *foreground* noise across different replicate groups. We also observed that the CVs of true regulators do not increase in correspondence to the noise levels as specified across *nreplicate* groups (Fig S1-4). These limitations reflect the fact that SynTReN cannot fully control expression noise, which may explain our observation of a few exceptions in comparisons of LemonTree performances when we varied replicate numbers or noise levels. Despite these limitations, our AR/CV profiling of the synthetic data confirms that this data simulator does offer a level of control on correlated input to output and hence warrants to be a reasonable LemonTree benchmark.

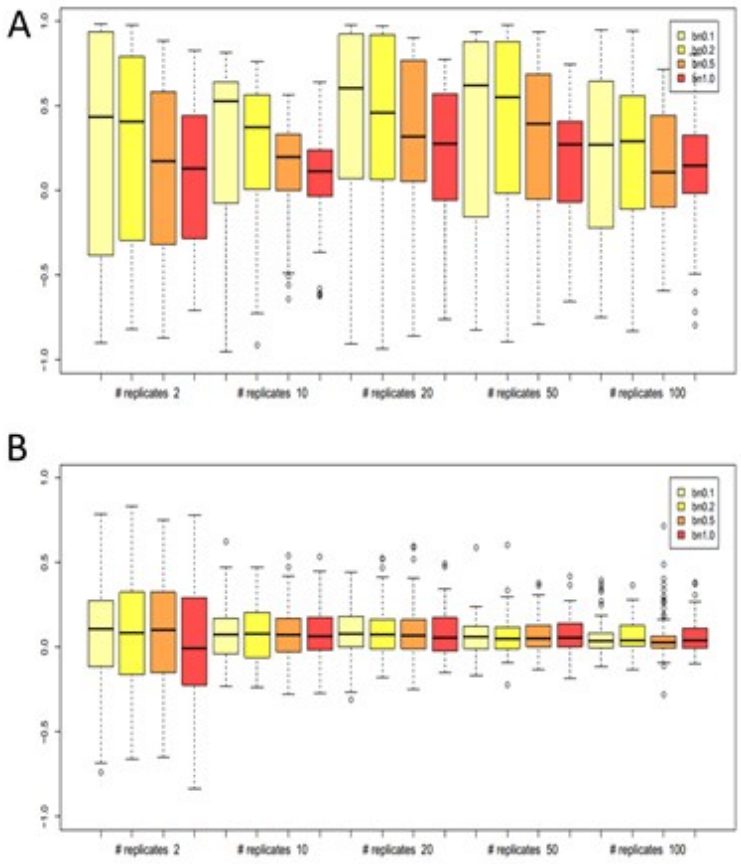

Fig S1-2. ARs of the *foreground* and *background* networks. (A) *foreground*. (B) *background*. *x*-axis *nreplicate*, *y*-axis AR, colors *bn*.

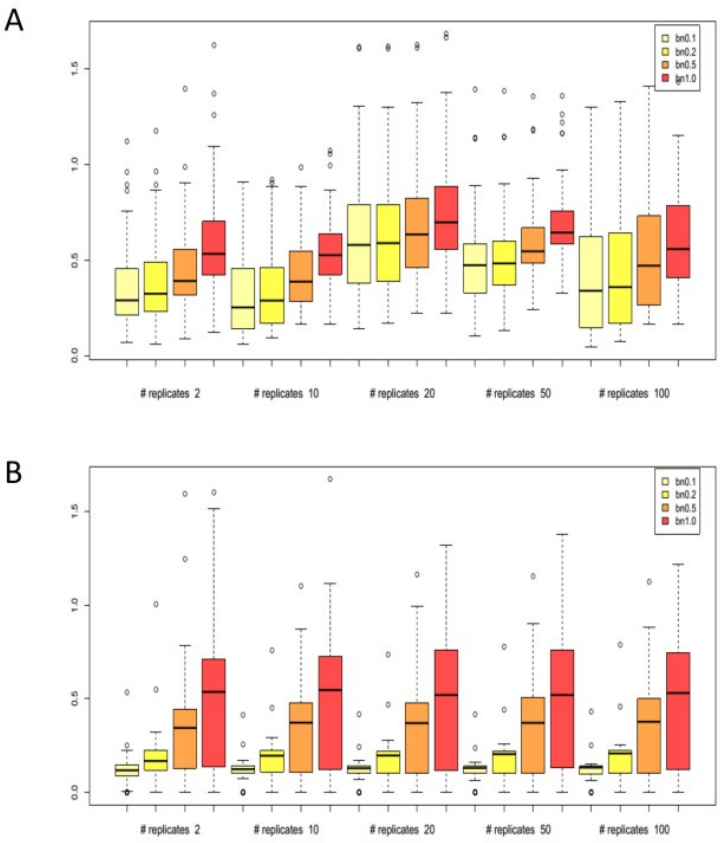

Fig S1-3. CVs of the *foreground* and *background* networks. (A) *foreground*. (B) *background*. *x*-axis *nreplicate*, *y*-axis CV, colors *bn*.

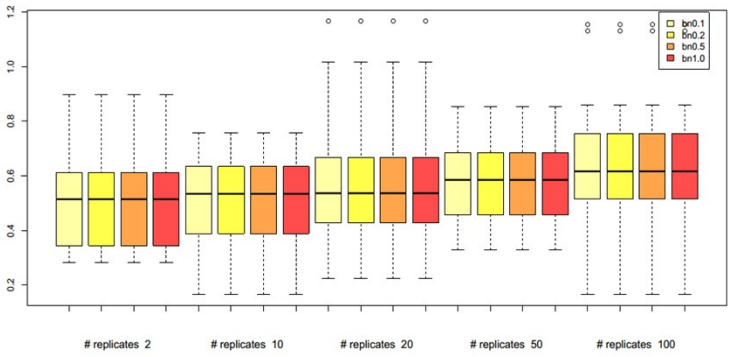

Fig S1-4. CVs of the true regulators in the *foreground* network. *x*-axis *nreplicate*, *y*-axis CV, colors *bn*.