# Supplementary information

**Supplementary Video I** . This supplementary video illustrates an interactive model exploration using our tools, as described in the example Jupyter notebook (which is destributed with the sources code). Video avaiable here:
https://www.dropbox.com/s/o0wszm7xdsnc7ri/paper1.mp4?dl=0

**Parameter sweep design of test GRN model** Table 1 outlines the parameter bounds used in the exploration of the oscillator and the robust parameter points for the validation points used in Results phase 3 and 4.

| Parameter | lower bound | upper bound | robust |
|---|---|---|---|
| $\theta_A$ | 0.5 | 5000 | 50 |
| $\theta_R$ | 1.0 | 10000 | 100 |
| $\alpha_A$ | 0.5 | 5000 | 50 |
| $\alpha_A^{'}$ | 5.0 | 50000 | 500 |
| $\alpha_R$ | 0.0001 | 1.0 | 0.01 |
| $\alpha_R^{'}$ | 0.5 | 5000 | 50 |
| $\beta_A$ | 0.5 | 5000 | 50 |
| $\beta_R$ | 0.05 | 500 | 5.0 |
| $\delta_A$ | 0.01 | 100 | 1.0 |
| $\delta_{MA}$ | 0.1 | 1000 | 10 |
| $\delta_{MR}$ | 0.005 | 50 | 0.5 |
| $\delta_R$ | 0.002 | 20 | 0.2 |
| $\gamma_A$ | 0.01 | 100 | 1.0 |
| $\gamma_C$ | 0.02 | 200 | 2.0 |
| $\gamma_R$ | 0.01 | 100 | 1.0 |

Table 1: Parameter Sweep design and robust settings for oscillations (24 h period).

**Summary statistics** Summary statistics (time series features from TSFRESH) used apart from the extreme values, mean, median, standard deviation, variance and sum.

- Absolute sum of changes:

$$\sum_{i=1}^{n-1} \mid x_{i+1} - x_i \mid$$

- Autocorrelation:

$$\frac{1}{(n-l)\sigma^2} \sum_{t=1}^{n-l} (X_t - \mu)(X_{t+l} - \mu)$$

where $n$ is the length of the time series $X_i$, $\sigma^2$ its variance and $\mu$ its mean. $l$ denotes the lag.

**Toggle switch example** Here we demonstrate our method on seperate problem involving a smaller stochastic biochemical network of a genetic toggle switch (Gardner et al. 2000). The model is a bistable toggle network consisting of two repressor and two promoter regions where the opposing repressors inhibit transcription of each other. Here we assume no knowledge of the bistability inherent in the model, and naively begin exploring the simulation outputs by randomly sample the parameter space using our workflow. The range of parameter space is presented in table 2. The summary statistics used for this example were extreme values, mean, median, standard deviation, variance and sum for both repressors and an additional summary being the Pearson correlation coefficient between repressors expression levels. Figure 1 A shows a reduced 2D TNSE plot of this feature space containig an inital bacth of 500 uniformly sampled points. The principal behavior of the model becomes apprent while exploring the data, showing both bistibility and monostable behaviors. While most points produces low copy number noise, the interesting points can be seperated by human-in-the-loop labeling and semi-supervised modelling. Here we have used the full set of summary statistics as input to label propagation using the RBF kernel as the similarity measure to contruct the graph Laplacian. The hyper-parameter optimization was then performed in the same manner as in our main example. Figure 2 B show the potential benefits of a downstream process using the learned model as an predictor. By continuously sample a total of 34000 points the label propagation model filtered out 5166 as beloning to interesting classes. Mapping these points back to their associated parameters we can get an initial qualitative understanding of which parameters give rise to these behaviors (Figure 1 B).

| Parameter | lower bound | upper bound |
|---|---|---|
| $\alpha_1$ | 0.01 | 100 |
| $\alpha_2$ | 0.01 | 100 |
| $\beta$ | 0.02 | 200 |
| $\gamma$ | 0.02 | 200 |
| $\mu$ | 0.01 | 100 |

Table 2: Parameter Sweep design of genetic toggle switch.

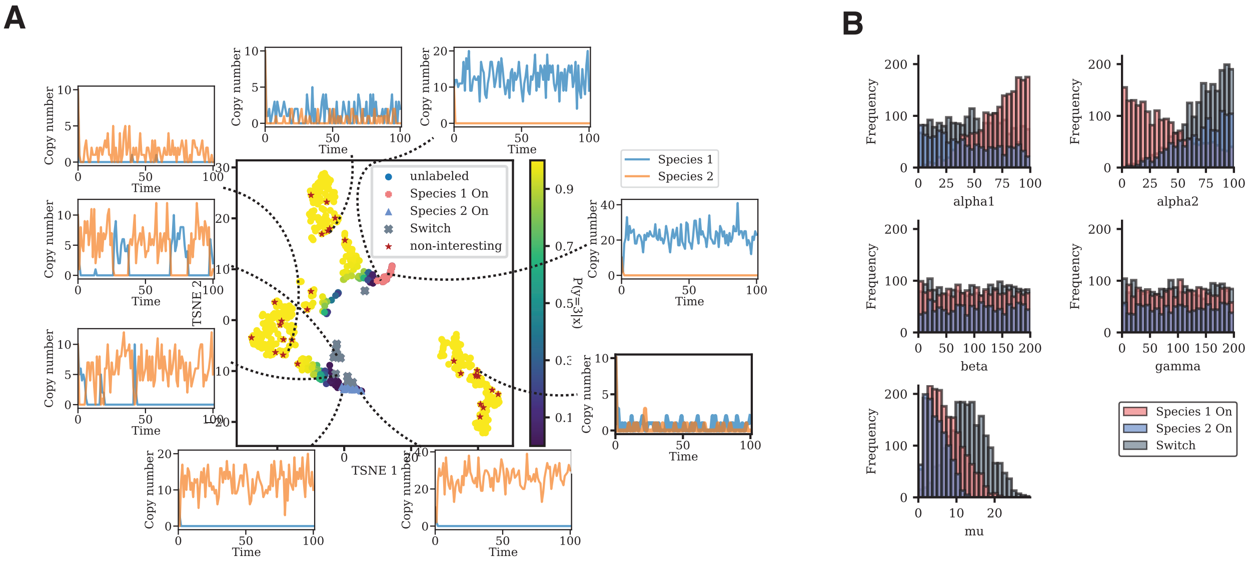

Figure 1: (A) Model exploration of genetic toggle switch. (B) Mapping of parameters after downstream prediction