Development of a generative framework for design-based selection of mammalian genetic programs

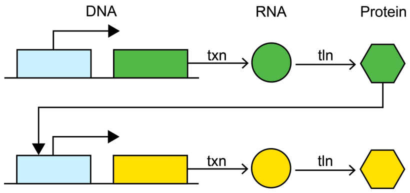
PSED Cluster 2020-2021

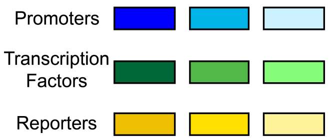
Kathleen Dreyer, Julie Anh Nguyen, Joshua N. Leonard*, and Niall Mangan*

*Faculty Advisor

Research Objective

A central goal of synthetic biology is to genetically engineer cells to perform customizable functions, which can be done using genetic circuits.





COMET toolkit enables construction of complex genetic circuits

There is a need in mammalian cellular engineering for *design* tool that enables high throughput construction and evaluation of genetic circuit designs.

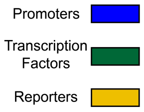
In pursuit of the ultimate development of a genetic circuit design tool, we propose to develop a generative algorithm for circuit design.

- Converts user-specified design objectives into candidate circuit topologies
- Utilizes model selection & parameter estimation to recover system of ODEs & parameters

Northwestern | ENGINEERING

Methods

LIbrary of Genetic Parts and Generalized ODEs

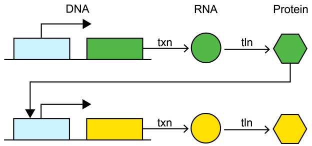


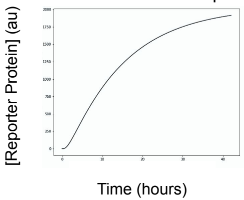
$$\frac{dX_i^R}{dt} = -k_{dR}X_i^R + \Phi_i(X^P)$$

$$\frac{dX_i^P}{dt} = -k_{dP}X_i^P + k_{tln}X_i^R$$

Generalized ODEs can be used to describe mRNA and protein states within the circuit.

Design Objective Test Cases





Circuit Design as a Model Selection Problem with Hidden Variables

We use 4DVar to mathematically derive the cost function and implement the LASSO-type reformulation using KNITRO as the optimizer.

$$\min \mathbf{J}(\mathbf{x_0}, \mathbf{p}) = \frac{1}{2}\sum_{j=0}^{t}(\mathbf{y_j} - \mathbf{H}[\mathbf{x_j}])^T \mathbf{R_j}^{-1}(\mathbf{y_j} - \mathbf{H}[\mathbf{x_j}]) \quad \text{s.t.} \quad ||\mathbf{p}||_1 \le \lambda$$

$\mathbf{x}_t$: true states at time t ($\mathbb{R}^n$)
$\mathbf{y}_t$: measurements at time t ($\mathbb{R}^m$)
H: a mapping operator($\mathbb{R}^n \rightarrow \mathbb{R}^m$)
$\mathbf{p}$: parameters in the model ($\mathbb{R}^d$)

To solve a parameter estimation problem, we can omit the constraint on the vector of parameters.

Results and Conclusions

Model Selection Algorithm Results

We use a time series of 600 time points obtained with a time step of 0.005 hours and apply the algorithm to a fully connected circuit. We want to down-select from 14 parameters to only 4 parameters.

	B_0	B_6	B_7	B_8
Ground truth	1	0.08	22	0.036
Model selection	0.39	0	32.5	0.096

Parameter Estimation Algorithm Results

We examine the robustness of the algorithm w.r.t. the effect of the time step/the number of time points by applying the algorithm to estimating parameters.

	B_1	B_2	B_3	$\|\|e\|\|_1$
Ground truth	0.25	54	0.018	
$\Delta_t = 0.25$	0.33	55	0.018	220
$\Delta_t = 0.05$	0.0056	55	0.017	47
$\Delta_t = 0.01$	0.051	54	0.018	11

The results we obtained so far are promising.

- Parameter estimation: results are highly accurate
- Model selection: almost all of the correct parameters, except one, are recovered.

We have some issues with recovering/estimating one particular parameter but this can be resolved with continued effort.