

Network Modeling Project

Team

Daniel Masami Nishimura d.m.nishimura@medisin.uio.no

Malin Eriksen malin.eriksen@fys.uio.no

Simon Beyer simon.beyer@fdm.uni-freiburg.de

Mentors

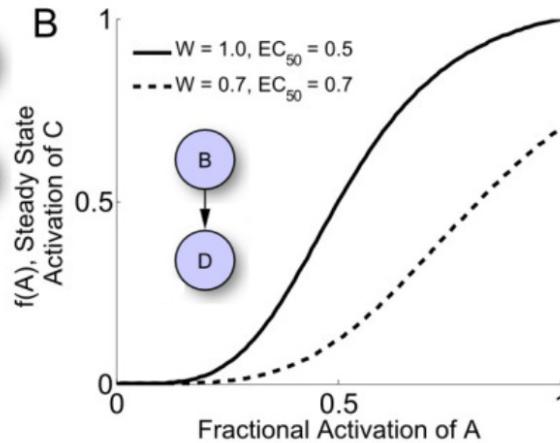
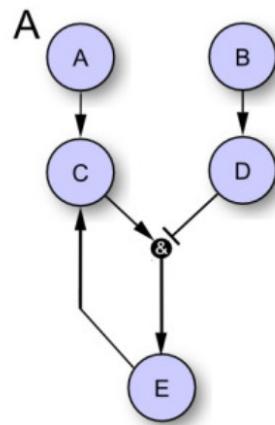
Joakim Sundnes sundnes@simula.no

Andrew McCulloch amcculloch@ucsd.edu

Logic-based Hill-type ordinary differential equations: Summary

Reaction Parameters

- n (Hill coefficient): steepness of the curve
- EC₅₀: half maximal activation
- W_{xy} (reaction weight , defaults to 1)



Species Parameters

- τ : time constant (related to half life)
- Y_{max} : (Species weight, defaults to 1)
- $Y_{initial}$ (initial condition)

OR: $X \vee Y \equiv X + Y - X \cdot Y$ **AND:** $X \wedge Y \equiv X \cdot Y$

$$\frac{dD}{dt} = \frac{1}{\tau_D} [(W_{BD}f_{act}(B)D_{max} - D)]$$

$$\frac{dC}{dt} = \frac{1}{\tau_C} \left[(W_{AC}f_{act}(A) + W_{EC}f_{act}(E) - W_{AC}f_{act}(A)W_{EC}f_{act}(E)) C_{max} - C \right]$$

$$\frac{dE}{dt} = \frac{1}{\tau_E} [(W_{CE}f_{act}(C)W_{DE}f_{inh}(D)E_{max} - E)]$$

$$f_{act}(X) = \frac{bX^n}{K^n + X^n}$$

$$f_{inh}(X) = 1 - \frac{bX^n}{K^n + X^n}$$

where $X \in [0,1]$

and n is the Hill coefficient

$$b = K^n + 1, K^n = \left(\frac{EC_{50}^n}{1 - 2EC_{50}^n} \right)$$

so that $f_{act}(X) = 1$ when $X = 1$

$$EC_{50}^n < \frac{1}{2} \text{ or } EC_{50} < 2^{-\frac{1}{n}}$$

$$J_{max}$$

Y_{max} and W_{IJ} default to unity

Kraeutler et al, BMC Sys Bio 2010

Network Model of Mechanoregulated Gene Expression in Cardiac Myocytes

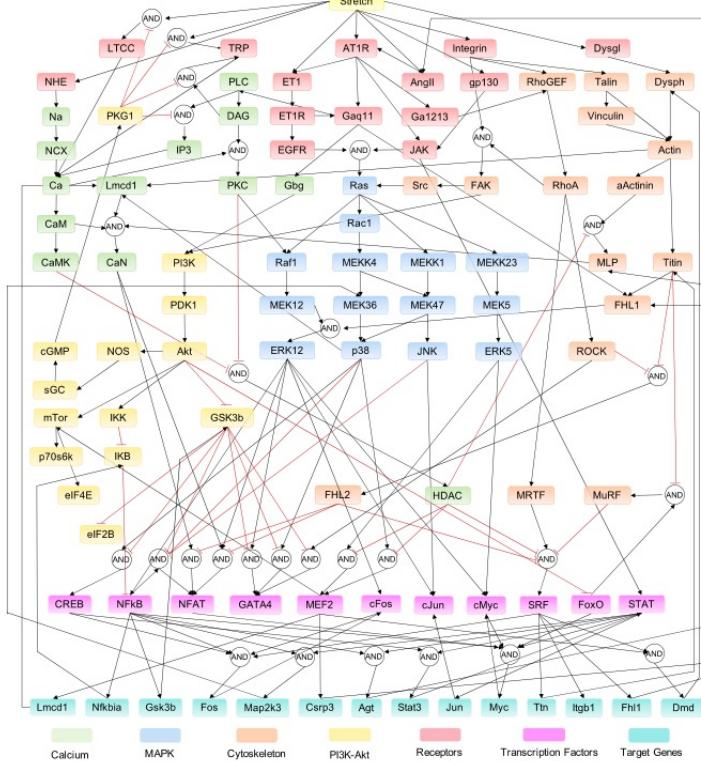
Cao S, Buchholz KS, Tan P, Stowe JC, Wang A, Fowler A, Knaus KR, Khalilimeybodi A, Zambon AC, Omens KH, Saucerman JJ, McCulloch AD (2024) [Differential sensitivity to longitudinal and transverse stretch mediates transcriptional responses in mouse neonatal ventricular myocytes](#). *Am J Physiol-Heart Circ Physiol* 2024;326:2, H370-H384

AMERICAN JOURNAL OF PHYSIOLOGY



HEART AND CIRCULATORY PHYSIOLOGY®

FEBRUARY 2024 VOLUME 326 NO 2



Calcium

MAPK

Cytoskeleton

PI3K-Akt

Receptors

Transcription

Genes

11 transcription factors

784 target genes

(293 experimentally validated
in mouse or in heart)

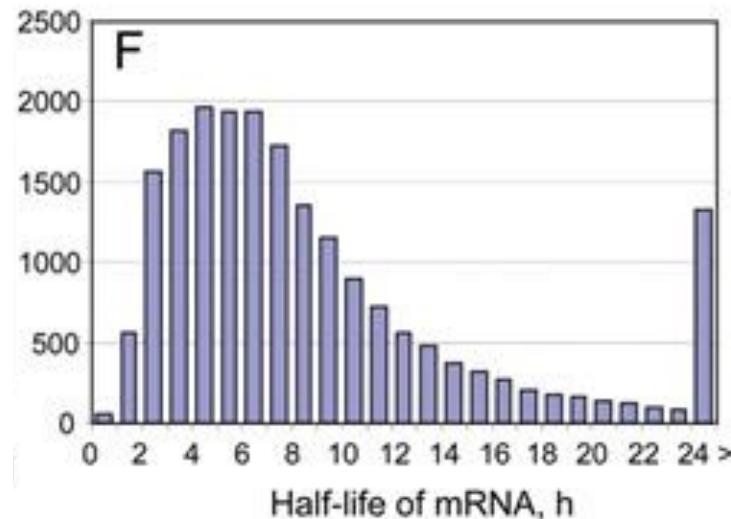
921 reactions

14 translation reactions

115 additional references (45
heart or myocyte studies)

Model Parameters for mRNA Expression

$k_{deg} = \frac{1}{\tau_Y} = \frac{\ln 2}{t_{\frac{1}{2}}}$ is the mRNA degradation rate, and $t_{\frac{1}{2}}$ is mRNA half-life



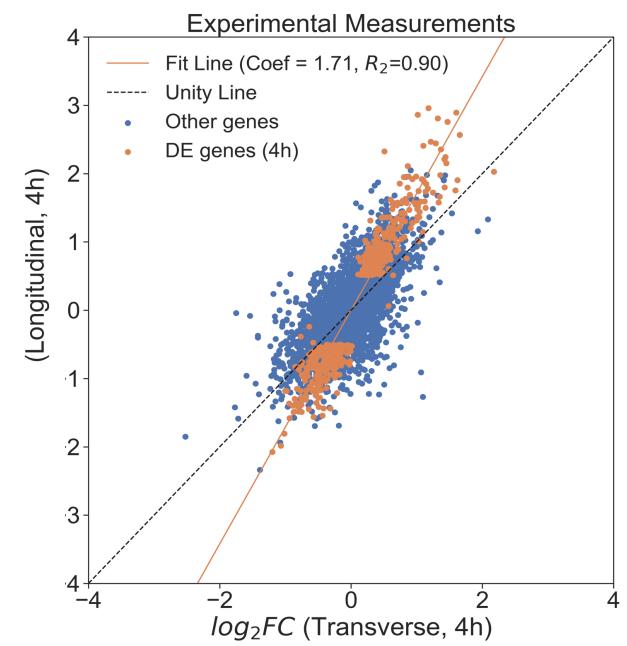
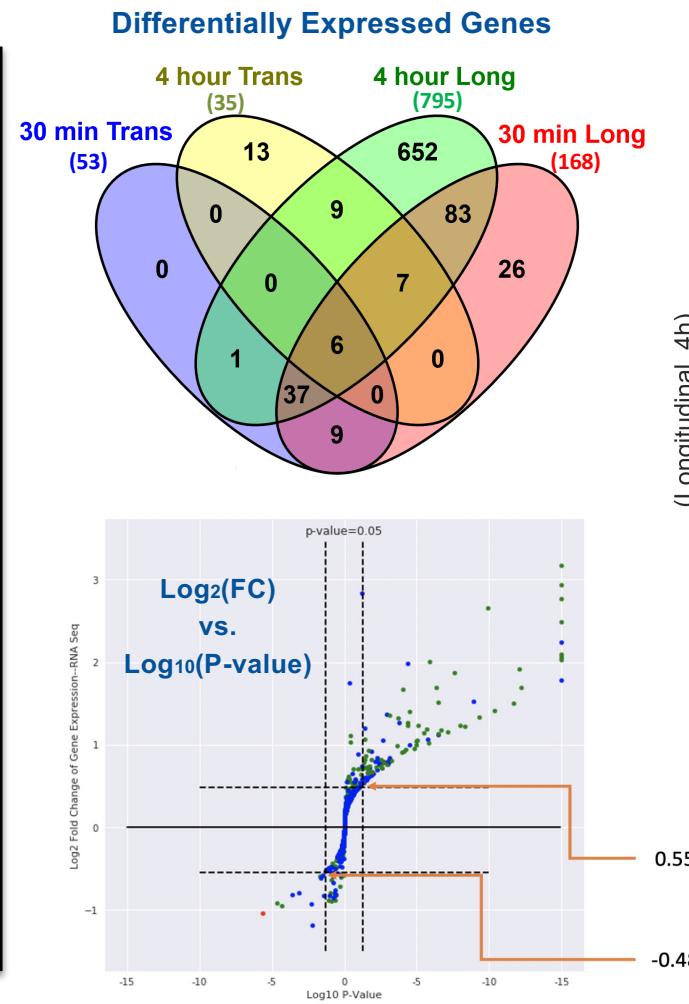
LV Sharova *et al.* (2009)
[Database for mRNA Half-Life of 19,977 Genes](#)
[Obtained by DNA Microarray Analysis of Pluripotent and Differentiating Mouse Embryonic Stem Cells](#), *DNA Research* **16**(1):45–58

- mRNA fold changes often have saturating and cooperative dependencies on promotor activity [1 ,2]: $n \sim 1 - 4$

1. Litovco P, Barger N, Li X, Daniel R (2021) [Topologies of synthetic gene circuit for optimal fold change activation](#). *Nucleic Acids Res* 49(9):5393–5406. PMID: 34009384.
2. Shi W, Ma Y, Hu P, Pang M, Huang X, Dang Y, Xie Y, Wu D (2024) [Hill Function-based Model of Transcriptional Response: Impact of Nonspecific Binding and RNAP Interactions](#). *arXiv* 2403.01702v1
3. Shi X (2021) [A Hill type equation can predict target gene expression driven by p53 pulsing](#). *FEBS Open Bio*, 11(6):1799–1808
4. Loewer A, Batchelor E, Gaglia G, Lahav G (2010). [Basal dynamics of p53 reveal transcriptionally attenuated pulses in cycling cells](#). *Molecular Systems Biology* 6:466

Stretch-Induced Gene Expression: Experimental Data

- We stretched cardiac myocytes for 30 min and 4 h, longitudinally or transversely to their myofiber axis
- Only 4h longitudinal stretch** produced a robust response
- Gene expression is typically expressed as a **fold change (FC)** from control or baseline
- Differential Gene Expression** analysis uses stringent statistical tests to reduce type I errors (e.g. $FDR < 0.05$ and $\log_2 FC > 1$)
- Experimental **replicates** and hence **power** are typically low, so the risk of type II errors is high
- We found large **quantitative difference** in gene expression in response to longitudinal vs. transverse stretch but no categorical (e.g. pathway) differences



Cao S, et al. Am J Physiol-Heart Circ Physiol 2024;326:2, H370-H384

**>90% of model predicted fold changes
were within ± 1 on a log₂ scale**

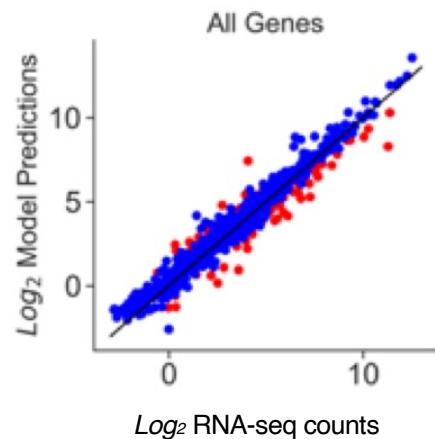


TABLE 1 Validation of Model-Predicted Gene Expression from Three Independent Experiments

	4h Longitudinal Stretch		1h Cyclic Stretch		2d TAC		Combined
	n	accuracy	n	accuracy	n	accuracy	accuracy
All Model Genes	784	69%	697	91%	740	63%	74%
Experimentally Confirmed Genes	288	77%	262	87%	269	76%	80%

Getting Started

1. **Unlike Cao et al, we will use the standard Netflix logic-based equations to predict quantitative gene expression**
 - Define a non-zero baseline stimulus corresponding to “zero stretch” and run to steady state. This solution will be used to compute fold changes. (The choice is arbitrary but if you use 0 as your baseline input, then some transcripts may be zero at baseline, so you couldn’t calculate a fold change)
 - Although the baseline and stretch stimulus magnitudes are arbitrary, there is reason to try and position the experimental stimulus (14% stretch) relative to a realistic maximum of say 20 or 25%. Cao et al, used a baseline of ~0.3 and a stimulus of 0.7. These are perfectly reasonable.
 - Note also that Cao *et al* included feedback from a model output (cell area) to the stretch input. That is because experimentalists have observed that the cells grow within a few hours to accommodate the stretch. This feedback is nice but not critical. You could also just prescribe a decaying stretch time course. Studies suggest that the cells have grown by the amount of the stretch in 4-6 hours.
 - The model also includes feedback from 14 genes which code for proteins in the network itself. You can ignore this feedback because translation is difficult to model accurately, and the time-courses are longer than the experiment itself anyway.
 - You can then check to see if the distribution of model-predicted (log) fold changes is reasonably consistent with the data.

Getting Started

2. The model predicts the expression of 784 genes

- These are predicted by their sequences using bioinformatics (ENCODE)
- A subset (288) has been validated experimentally, e.g. using ChIP-Seq
- Newer available data may increase these numbers somewhat
- The paper included a set of experimentally measured transcripts for 30 min and 4h longitudinal and transverse stretch. **Since only 4 h longitudinal stretch** induced a large number of genes, ignore the other groups.
- The paper (Table 1) also used two other public datasets (in GEO, <https://www.ncbi.nlm.nih.gov/geo/>) for comparison. Note that here a $\log_2(\text{FC})$ of ± 0.5 was used to predict a significant difference ($p < 0.05$) determined using false discovery rate.
- **You could also try using quantitative comparisons of gene expression between model and data**

Table 1. Validation of model-predicted gene expression from 3 independent experiments

X

	Longitudinal Stretch (4 h)		Cyclic Stretch (1 h)		TAC (2 Days)		Combined
	n	Accuracy, %	n	Accuracy, %	n	Accuracy, %	Accuracy, %
All model genes	784	69	697	91	740	63	74
Experimentally confirmed genes	288	77	262	87	269	76	80

TAC, transverse aortic constriction.

Getting Started

3. Model parameters have not been optimized

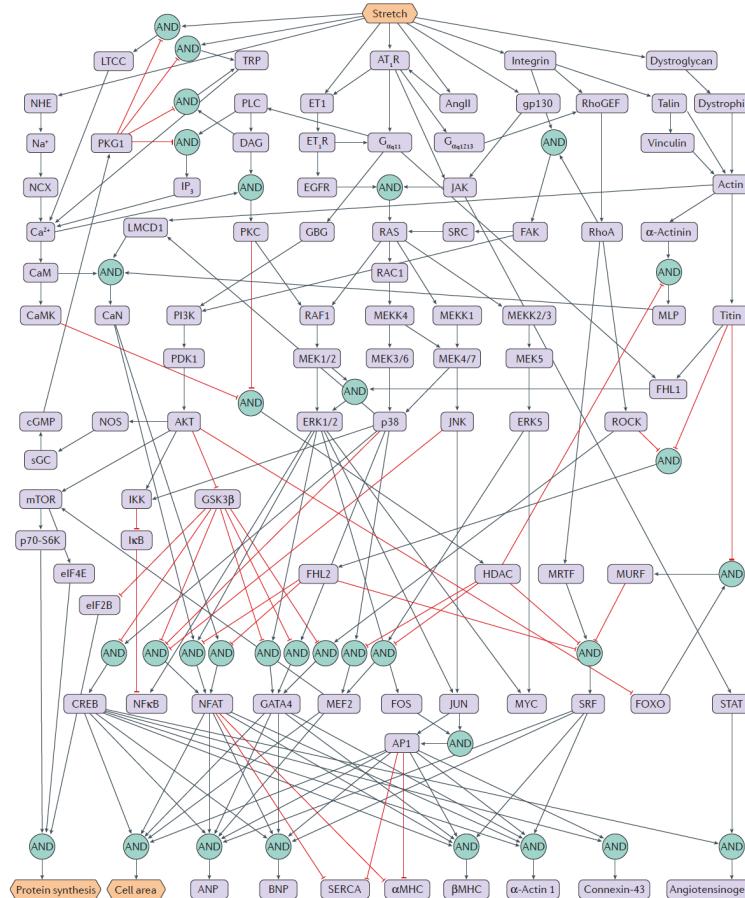
- $EC50$ and n are the same for every reaction by default
- $W=0.9$ and $Y_{max}=1$ for all except for sensitivity analyses.
- Time constants are the same for all signaling reactions
- Time constants for gene expression are based on a published dataset for **mRNA half lives**.
- Many genes are regulated by multiple transcription factors. Decreasing $EC50$ and replacing AND logic with OR logic as the default when more than one transcription factor regulates a gene (unless published evidence suggests co-regulation) caused model-predicted expression to saturate for fewer regulators in closer agreement with experiments (Fig. S4)
- Because there is a gene regulatory reaction for every gene, there are more model parameters than model outputs, so overfitting is easily possible.
- However, all genes are regulated by the same 9 upstream receptors, 4 pathways and 11 transcription factors. All the gene specific reactions are downstream of the transcription factors and could be left unchanged.
- The paper also includes some data on the effects of blocking specific pathways, but these experiments only measured a few genes by PCR.

Getting Started

4. Parameters to optimize

- For optimizing reactions, either hold n and $EC50$ constant and optimize W or hold W and n constant and optimize $EC50$. Note that you would need data at more times or doses (stretches) to optimize n . Since W is a linear scaling parameter between 0 and 1, it is probably ideal for optimizing.
- Similarly, you would need time-course data to optimize τ , so best not to change it either.
- While you could optimize Y_{max} , it will probably be better to keep these at 1 and save them for sensitivity analysis or simulating a knockout, knockdown or inhibition experiment.
- There are good formal methods for determining which parameters or parameter combinations exert the greatest effect on model outputs (“stiff” parameters) and which have the least effect (“sloppy” parameters). These methods could help reduce the number of parameters to optimize. The “sloppy models” approach uses the Fisher Information Matrix (FIM) to find parameter combinations that model outputs are most sensitive to. See
<https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.0030189>
- Two intuitive starting cases could be: optimize only the weights of reactions between stretch and the upstream receptors or only the weights of reactions immediately upstream of the transcription factors. Some genes are only regulated by one transcription factor while some are regulated by 2 or more. If optimizing TF activation to match the expression of genes regulated by 1 TF is a good strategy, then predicted expression of co-regulated genes should also be more accurate.
- The most robust optimizations trained on the stretch data in the paper should also improve prediction accuracy for other mRNA data sets, but it is also possible that different data sets will have to be optimized separately.

ODE Network Models are similar to some Neural Network Architectures



- Variables (**nodes**) in **dimensionless** units
 - Reactions (**arrows**) use **saturating functions**
 - **Directed graph structure**
 - Governing by continuous, differentiable **ODEs**
 - This model has characteristics of **neural network architectures**,
e.g.
 - **CTRNN:** Continuous-Time Recurrent Neural Networks
 - **NeuralODE:** Neural Ordinary Differential Equations
 - **GNN-ODE:** Graph Neural Networks with Neural ODEs

Tan et al, (2017) Systems analysis identifies key network regulators of cardiomyocyte mechano-signalling. *PLoS Comp Biol*

Neural Network Architectures

Continuous-Time Recurrent Neural Networks (CTRNNs)

$$\tau_i \frac{dx_i}{dt} = -x_i + \sum_j w_{ij} f(x_j)$$

Where τ_i are time constants, w_{ij} are weights, and $f(x)$ is a saturating activation function.

Neural Ordinary Differential Equations (Neural ODEs)

$$\frac{dx}{dt} = f(x, \theta)$$

- Smooth, continuous-time transformations governed by ODEs

Graph Neural Networks with Neural ODEs (GNN-ODEs)

- If the network has a graph structure, GNNs with Neural ODEs can learn the system.
- Allow for acyclic topologies, e.g. feedback loops.

Reservoir Computing (Echo State Networks / Liquid State Machines)

- If feedback not explicitly known, reservoir computing models can approximate the system.

Summary

- Known ODEs** governing dynamics → CTRNN or Neural ODEs.
- Graph structure is important** → GNN-ODEs.
- Highly nonlinear and complex** → Reservoir Computing (ESNs/LSMs)

Most Papers Use These Models to Impute Networks from Data

1. Biologically Informed NeuralODEs for Genome-Wide Regulatory Dynamics

“Biologically Informed NeuralODEs for Genome-Wide Regulatory Dynamics” Qian et al., *Genome Biol*, 2023. Framework that integrates biological priors into NeuralODEs to model genome-wide regulatory dynamics. Uses large-scale time-series gene expression datasets enhances accuracy of GRN modeling.

2. Inferring Gene Regulatory Networks from Single-Cell Multiome Data with LINGER

“Inferring Gene Regulatory Networks from Single-Cell Multiome Data with LINGER” by Chen et al., Nat Biotech, 2024. Lifelong neural network that infers GRNs from single-cell RNA-seq and ATAC-seq data in thousands of cells

3. From Graph Topology to ODE Models for Gene Regulatory Networks

“From Graph Topology to ODE Models for Gene Regulatory Networks” Villaverde et al., BMC Sys Biol, 2019. ML methodology to transform qualitative graph-based representations of gene regulatory networks into quantitative ODE models, using transcriptomic data.

4. Scalable Optimal Bayesian Classification of Single-Cell Trajectories

“Scalable Optimal Bayesian Classification of Single-Cell Trajectories Under Regulatory Model Uncertainty” by Hajiramezanali et al., BMC Genomics, 2019. Bayesian inference and particle-based methods to infer gene regulatory networks from single-cell RNA sequencing data.

SCIENCE ADVANCES | RESEARCH ARTICLE

SYSTEMS BIOLOGY

DeepVelo: Single-cell transcriptomic deep velocity field learning with neural ordinary differential equations

Zhanlin Chen¹, William C. King², Aheyon Hwang³, Mark Gerstein^{1,4,5*}, Jing Zhang^{6*}

Deep Neural Networks for Predicting Single-Cell Responses and Probability Landscapes

Published as part of the ACS Synthetic Biology *virtual special issue* "AI for Synthetic Biology".

Heidi E. Klumpe,^{||} Jean-Baptiste Lugagne,^{*||} Ahmad S. Khalil, and Mary J. Dunlop^{*}



Cite This: *ACS Synth. Biol.* 2023, 12, 2367–2381

Hossain et al. *Genome Biology* (2024) 25:127
<https://doi.org/10.1186/s13059-024-03264-0>

Genome Biology

Inferring gene regulatory networks from single-cell multiome data using atlas-scale external data

Received: 4 August 2023

Qiuyue Yuan & Zhana Duren

Accepted: 26 February 2024

RESEARCH

Open Access

Biologically informed NeuralODEs for genome-wide regulatory dynamics

Intekhab Hossain^{1*} , Viola Fanfani¹, Jonas Fischer¹, John Quackenbush^{1*} and Rebekka Burkholz²



Getting Started

5. A starting point could be to use model to train a NeuralODE

- If the input layer is the stimulus and parameters and the output layer is the model-predicted gene expression, then the neural network becomes a surrogate model
- If the input layer is the predicted genes and the output layer is the input parameters, then the neural network becomes a parameter optimizer
- It is likely that a combination of the two with additional regularization and a well formulated loss function will give the best results., e.g.

Input: Observed gene expression



Inverse network: predicts reaction weights



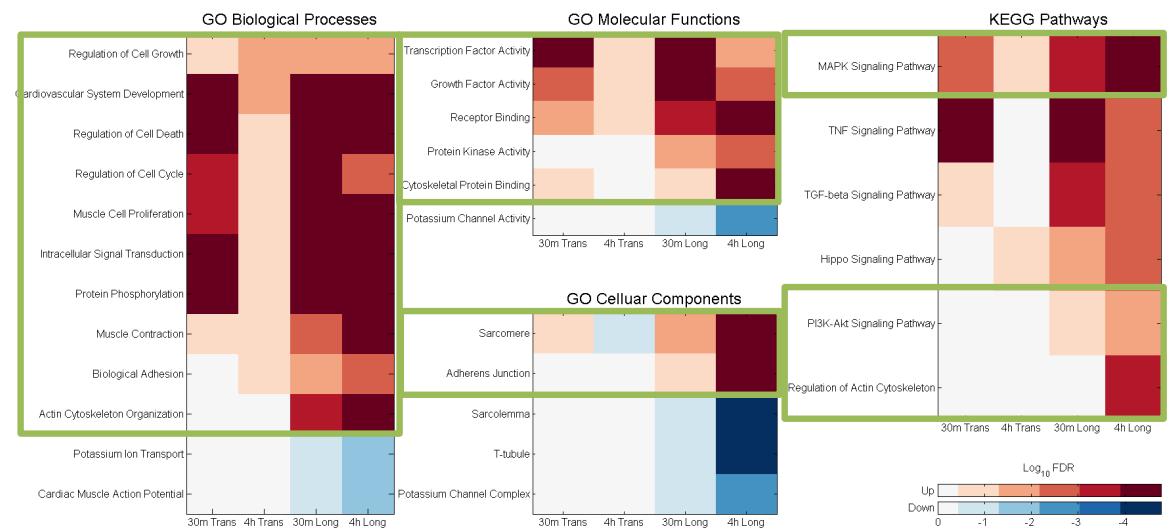
Forward ODE or surrogate model: predicts gene expression from weights



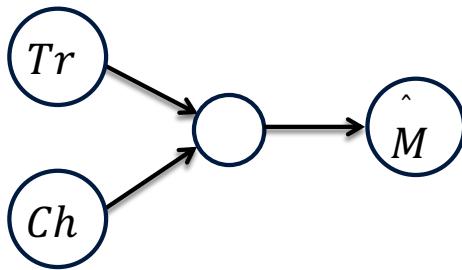
Loss = RMS or absolute difference between predicted and original gene expression

Some ways the model could be improved

- There are many genes differentially expressed in the data that are not included in the model, notably a cluster of K channel subunit genes that are downregulated by stretch. NFAT, MEF2, GATA4, oxidative stress-responsive factors, HDACs and miRs have all been implicated in K channel expression.
- There are ways to measure chromatin accessibility (epigenetic regulation) and include it in the model, though we don't have these measurements yet
- Better models of protein translation from mRNA are possible including feedback proteins



Chromatin-Mediated (Epigenetic) Regulation of Gene Expression



e.g., regulation of *IL2RA* (CD25)

gene expression M in T cell activation requires transcription factor activation Tr and chromatin accessibility Ch , which is low in unactivated cells.



Cell Reports Methods

Volume 2, Issue 3, 28 March 2022, 100182

Article

PeakVI: A deep generative model for single-cell chromatin accessibility analysis

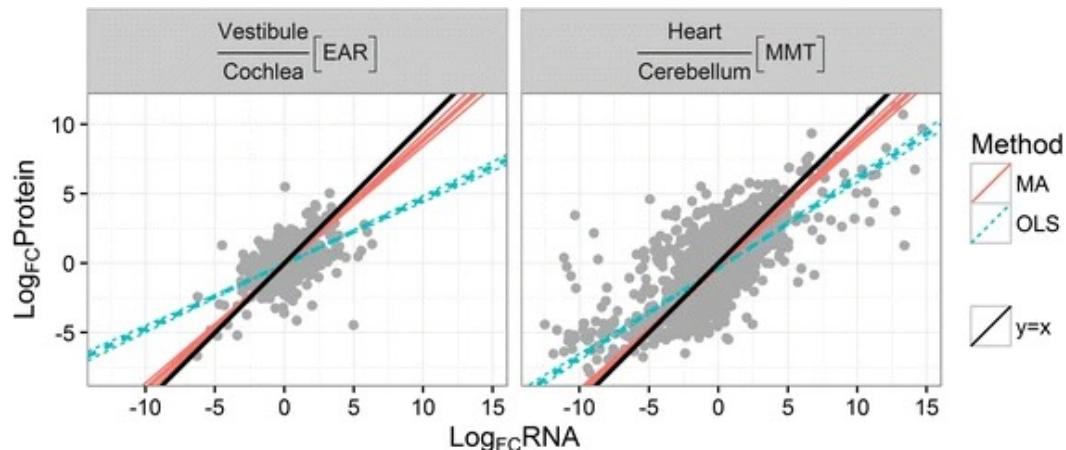
Tal Ashuach¹, Daniel A. Reidenbach², Adam Gayoso¹, Nir Yosef^{1 2 3 4 5}

Chromatin accessibility can be measured using ATAC-seq, DNase-seq or SMAC-seq and expressed as a normalized “accessibility score”

mRNA:Protein Expression is Nonlinear and Variable

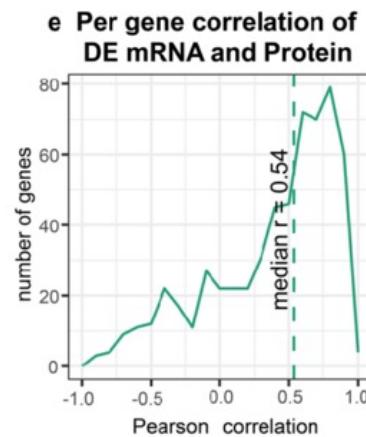
Mouse

Perl, K., Ushakov, K.,
Pozniak, Y. et al. [Reduced changes in protein compared to mRNA levels across non-proliferating tissues](#). *BMC Genomics* **18**, 305 (2017)



Dicty

Bart Edelbroek, Jakub Westholm, Jonas Bergquist, Fredrik Söderbom (2024) [Multi-omics analysis of aggregative multicellularity](#), *iScience* 27(9):110659



- Differences in mRNA and protein **stability**
- Protein **degradation and post-translational modifications**
- Ribosome **saturation** (limits on translation efficiency at high [mRNA])
- **Translational regulation** (ribosome binding efficiency, microRNA repression)
- **Noise** and dynamic regulation