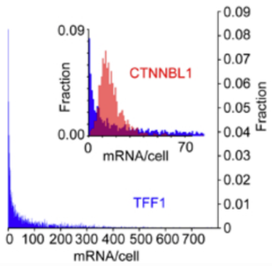
HARVARD MEDICAL SCHOOL

SIMULATING THE TRANSCRIPTIONAL BURSTING
of estrogen-responsive TFF1 gene

Neha Anand, Joseph Nasser and Jeremy Gunawardena

BLAVATNIK INSTITUTE
SYSTEMS BIOLOGY

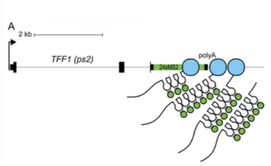
BACKGROUND



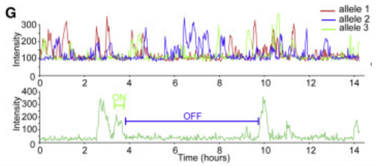
F1. Rodriguez, J et al., 2019
Despite being genetically identical in similar external environments, the gene regulation differs.

Rodriguez, J et al., Cell 2019. - the paper that forms the base crux of this project describes the observed heterogeneity in gene expression in the highly induced, estrogen-responsive TFF1 gene as a consequence of the stochastic nature of biochemical reactions.

(F1) shows a comparative histogram of #mRNAs transcribed per cell of two genetically identical genes: CTNNBL1, a housekeeping gene and the TFF1 gene.



F2. Rodriguez, J et al., 2019: 24 MS2 stem looping into 3' UTR. As MS2 loops transcribe - live cell visualisation at the site of transcription is enabled.



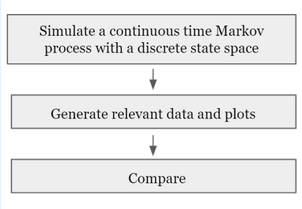
F3. Rodriguez, J et al., 2019 : Noise in transcription: long repressive states, short periods of transcriptional bursting- period of RNA synthesis activity when multiply polymerases initiate

Using live-cell RNA imaging and MS2 tagging, the paper studies TFF1 transcription in real-time for a single cell, over multiple alleles and at varying E2 (estradiol) concentrations and reports that heterogeneity in inactive times explains the 'noise' in human gene expression and distribution of protein levels in human tissue.

METHODS

The overarching goal of the project (F4) was simulating a mathematical model that could explain the experimental results collated in Rodriguez, J et al., Cell 2019.

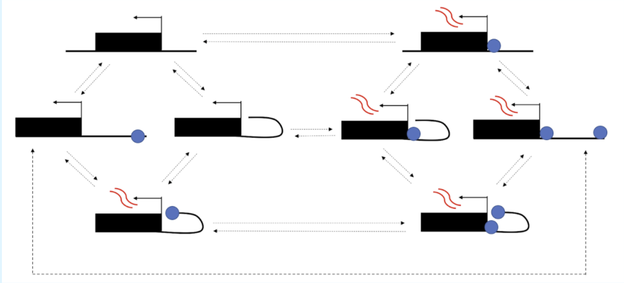
Analysis of the TFF1 gene reveals three binding sites for the transcription factors - one proximal and two distal locations, as depicted in (F5). A simple ON-OFF model with discrete state-space (F6) was used to simulate the enhancer-promoter transcription dynamics of the TFF1 gene at the steady-state. For the simplicity of the base model, the distal binding locations were lumped together as one binding site in the further analysis.



F4. Project Outline

F5. Rodriguez, J et al., 2019: An enhancer of TFF1 is located ~10kb upstream, and contains two EREs: Estrogen Response Elements (grey).

The simulation algorithm - a continuous-time Markov Chain generates data of the time held in each state and the ON-OFF data. The simulation runs for a total steady-state time duration of 14.2 hours taking an input combination of six parameters of which five are free parametres.



F6. Proposed Model for Simulation; The blue dots represent the transcription factor binding at the proximal and distal sites, and the red wiggly lines depict mRNA generation due to transcription.

The generated data is further sampled every 100 seconds to replicate the experimental conditions of (F8) obtained from Rodriguez, J et al., Cell 2019. An additional parameter 'p' scales down the observed events to account for the binding of the transcription factor leading to mRNA generation.

The simulation is run for 100 allele traces at saturating E2 conditions, as per Rodriguez, J et al., Cell 2019 to generate the ON-OFF data for plotting (F9).

FUTURE DIRECTIONS

Being an early-stage exploratory project, the work presents a wide scope of possible pathways that can be explored further in future.

Some of the interesting questions that can be addressed are:

- Is the model an accurate biological representation of the experimental conditions?
- Which combinations and subsets of the five free parameters outlined in this project result in simulations closest to the experimental results?
- Is the first passage time analysis with the study of the first ON-times of the alleles a sufficient metric for comparing the experiment and simulation?
- Is this approach reproducible and extendable to applications in other MS2 datasets?

RESULTS

The simulation input Q: transition rate matrix for the continuous-time Markov Process simulation of (F6) modelled specifically for the TFF1 gene is given below.

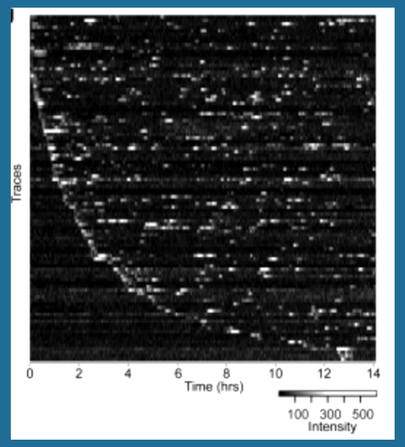
$$\begin{pmatrix} -0.21 & 0.1 & 0.01 & 0 & 0.1 & 0 & 0 & 0 \\ 2 & -2.11 & 0 & 0.01 & 0 & 0 & 0.1 & 0 \\ 0.1 & 0 & -0.3 & 0.1 & 0 & 0.1 & 0 & 0 \\ 0 & 0.1 & 2 & -2.2 & 0 & 0 & 0 & 0.1 \\ 5 & 0 & 0 & 0 & -5.11 & 0.01 & 0.1 & 0 \\ 0 & 0 & 5 & 0 & 0.1 & -5.2 & 0 & 0.1 \\ 0 & 5 & 0 & 0 & 2 & 0 & -7.01 & 0.01 \\ 0 & 0 & 0 & 5 & 0 & 0.1 & 0.1 & -5.2 \end{pmatrix}$$

F7. Transition Rate Matrix

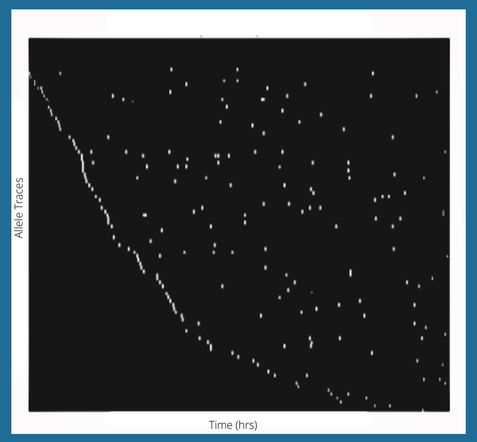
The parametres for Q have been chosen to be biologically plausible estimates. The initial looping and unlooping rate of the DNA in the simulation was set as 0.01/s and 0.1/s computed by modeling the enhancer as freely diffusing relative to the gene promoter, using Hi-C data - taking a radius around the promoter (r = 100nm) and a diffusion rate of 10^-2 um^2/s.

The binding rate of the transcription factor was set as 0.1/s for both the proximal and distal binding sites. For the unbinding rate at the proximal site and the distal site: 5/s and 2/s was used respectively. The scaling parameter 'p' was chosen as 0.01.

The experimental and simulation plots were compared on the first-passage times. Manhattan distance was computed to quantify the error in the first ON time of the 100 simulated alleles.



F8. Rodriguez, J et al., Cell 2019: Raw intensity traces of 100 alleles plotted as a heatmap.



F9. Simulation data generated plot to replicate the experimental conditions of (F5)

The sum of the computed Manhattan distances between the curve data points of the experimental plot (F8) and simulation generated plot (F9) was 121.22 hours.

REFERENCES

1. Rodriguez J, Ren G, Day CR, Zhao K, Chow CC, Larson DR. Intrinsic Dynamics of a Human Gene Reveal the Basis of Expression Heterogeneity. Cell. 2019; doi:10.1016/j.cell.2018.11.026
2. Pan YF, Wansa KD, Liu MH, et al. Regulation of estrogen receptor-mediated long range transcription via evolutionarily conserved distal response elements. doi:10.1074/jbc.M802024200
3. A. Amitai, D. Holcman. Polymer physics of nuclear organization and function. doi:https://doi.org/10.1101/076661
4. K.J. Pawlak, J.P. Wiebe, Regulation of estrogen receptor (ER) levels in MCF-7 cells by progesterone metabolites, The Journal of Steroid Biochemistry andMolecular Biology, Volume 107, Issues 3–5, 2007
5. Leonid Mirny et al, How a protein searches for its site on DNA: the mechanismof facilitated diffusion, 2009 J. Phys. A: Math. Theor. 42 434013
6. Fudenberg G, Mirny LA. Higher-order chromatin structure: bridging physics and biology. doi:10.1016/j.gde.2012.01.006

ACKNOWLEDGEMENTS

I would like to thank Dr Jeremy Gunawardena for graciously accepting me into his laboratory and for the opportunity to work on such exciting prospects. Further, I wish to thank Dr Jospeh Nasser for consistent mentoring, guidance and support. Lastly, a vote of thanks to all the members of the Gunawardena Lab who made even the virtual summer interactions exciting!