Learning Boolean Networks in HepG2 cells using ToxCast High-Content Imaging Data

Todor Antonijevic[1,2], Imran Shah[2]

1. The Oak Ridge Institute for Science and Education
2. U.S. Environmental Protection Agency, Office of Research and Development, National Center for Computational Toxicology, Research Triangle Park, NC 27711

www.epa.gov

This poster does not necessarily reflect U.S.EPA policy.

Todor Antonijevic | Antonijevic.Todor@epa.gov | ORCID ID 0000-0002-0248-8412

Background

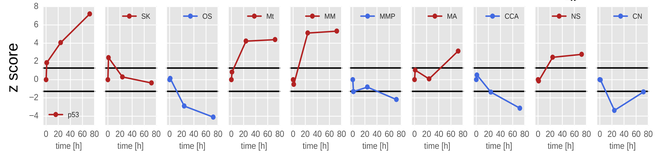
Objective: Cells adapt to their environment via homeostatic processes that are regulated by complex molecular networks. Our objective was to identify key elements of these networks in HepG2 cells using ToxCast High-content imaging (HCI) measurements taken over three time points (1, 24, and 72h) and across 10 concentrations (0.39-200μM) for 309 chemicals.

Approach: Cell states were monitored via phospho-p53 (p53), phospho-c-Jun (SK), phospho-Histone H2A.x (OS), phospho-Histone H3 (MA), phospho α-tubulin (Mt), mitochondrial membrane potential (MMP), mitochondrial mass (MM), cell cycle arrest (CCA), nuclear size (NS) and cell number (CN) endpoints. We used Boolean Networks (BNs) as a simple coarse-grained representation of biological regulatory networks. First, measured endpoints were standardized, and then discretized into perturbed/unperturbed values based on a noise threshold (z_0=1.28) and dynamic trends. Second, we inferred the best Boolean functions and constructed a set of 300 BNs for 2,193 trajectories with at least 1 perturbation using uniform sampling. The accuracy of initial 657,900 BNs was estimated as the number of errors between predicted and observed trajectories. 486,746 BNs with the smallest error, defined by the baseline error, were tested across analyzed trajectories. We defined "coverage" as the number of trajectories predicted by each BN with an accuracy ≤ to the baseline error.

Methods

1. **HCI data**[1] were used to study the effect of ToxCast I chemicals on HepG2 cell states by monitoring 10 endpoints across 3 time points (1, 24, and 72h) and 10 concentrations (0.4 to 200μM).

2. **Discretization of standardized data**. Data expressed as time vs. z sore ($z = \frac{x - x^*}{\sigma_x}$).



3. **Boolean functions[2] were inferred and BNs were construction for each trajectory.**

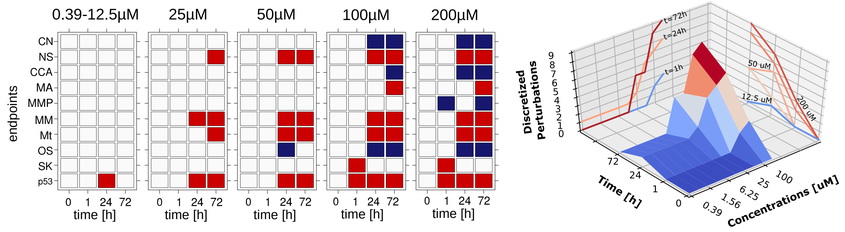
4. **Needleman-Wunsch[3] global alignment and error estimation** were calculated between observed discretized trajectory and BN prediction for:
   a) the trajectory for which BNs were sampled;
   b) all trajectories

5. **Boolean Network Coverage**

	trajectories						
	1	2	3	4	5	6	7
BN1	1	1	1	0	1	0	0
BN2	0	0	0	1	1	1	0
BN3	0	0	1	0	0	0	1

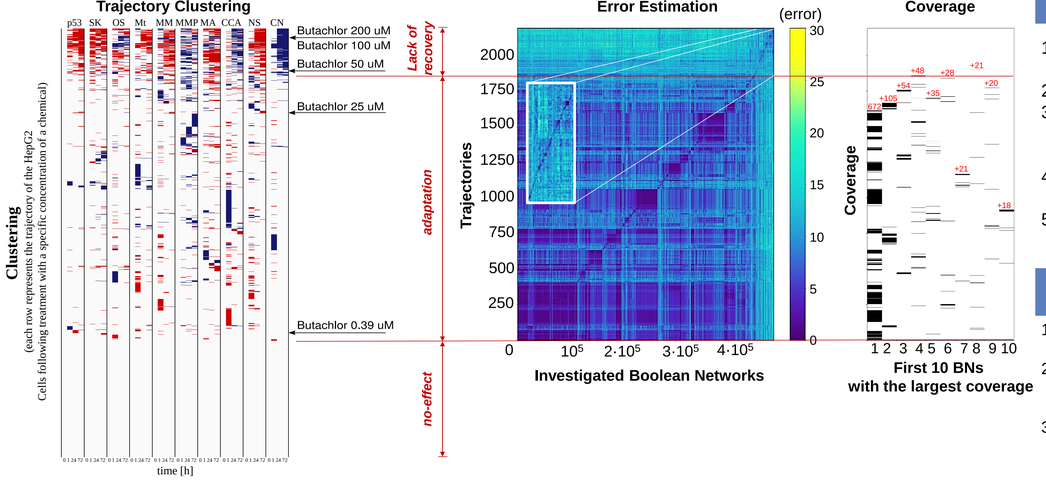
Discretized HepG2 Trajectories and Perturbations

Discretized HepG2 trajectories reveal increased endpoint perturbations as a response to increase in concentration. Here we show trajectories and perturbations induced by Butachlor.

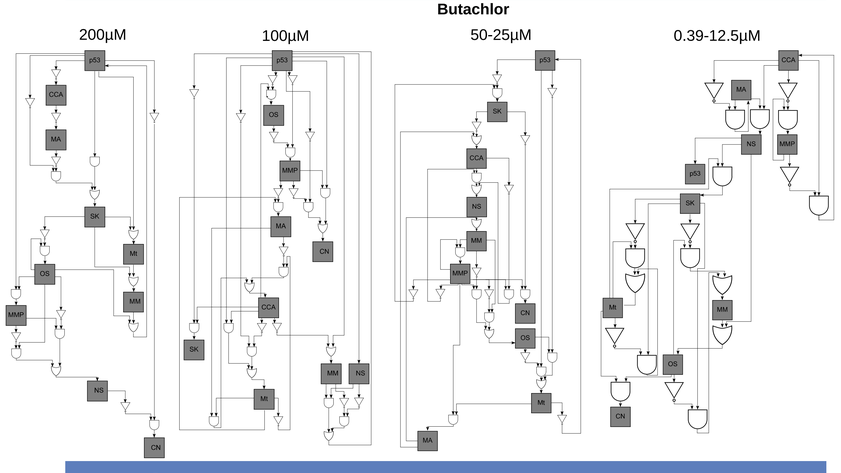


Clustering HepG2 Trajectories, Error Estimation and Coverage

Clustering of HCI trajectories shows 3 temporal trends: 1) no-effect, 2) adaptation, and 3) lack of recovery. It also suggests that similar perturbation patterns may indicate similar response mechanisms. Further, error estimation discriminates BNs that perform well in adaptation region from those that lead toward adverse effects.



Inferred Boolean Networks in case of Butachlor



Results

1. Response of HepG2 cells to concentration dependent chemical treatment shows three temporal trends: 1) no-effect, 2) adaptation, and 3) lack of recovery.
2. We have found that 573 BNs are needed to cover all trajectories.
3. BN with the greatest coverage explained 1,489 trajectories. These trajectories were produced by low treatment concentrations and we believe they represent cellular recovery processes.
4. Trajectories produced by high concentration treatments, that resulted in cell death, were predicted by a different set of BNs.
5. Our findings illustrate the utility of BNs that differentiate cellular programs involved in adaptation versus injury.

References

1. Shah, Imran, et al. "Using ToxCast™ Data to Reconstruct Dynamic Cell State Trajectories and Estimate Toxicological Points of Departure." *Environmental health perspectives* (2016) .
2. Müssel, Christoph, Martin Hopfensitz, and Hans A. Kestler. "BoolNet—an R package for generation, reconstruction and analysis of Boolean networks."*Bioinformatics* 26.10 (2010): 1378-1380.
3. Needleman, Saul B. Wunsch, Christian D. "A general method applicable to the search for similarities in the amino acid sequence of two proteins" Journal of Molecular Biology 1970; 48/3 443-453