Learning Boolean Networks from ToxCast High-Content Imaging Data

Todor Antonijevic

ORCID ID 0000-0002-0248-8412

Outline

I. **Introduction**

II. **Methods:**

   1. Dataset
   2. Data standardization, and Noise Threshold (z_0).
   3. Data Discretization.
   4. Learning Boolean Functions and Construction of Boolean Networks (BNs).
   5. Needleman-Wunsch (NW) optimal global alignment, and Error Estimation.
   6. Coverage.

III. **Results:**

   1. Discretized Trajectories and Total Perturbation.
   2. Clustering of discretized trajectories, Error Estimation, and Coverage (first 10 BNs).
   3. Learned BNs in case of Butachlor.

IV. **Summary**

Outline

I. Introduction

II. **Methods:**

1. Dataset
2. Data standardization, and Noise Threshold (z_0).
3. Data Discretization.
4. Learning Boolean Functions and Construction of Boolean Networks (BNs).
5. Needleman-Wunsch (NW) optimal global alignment, and Error Estimation.
6. Coverage.

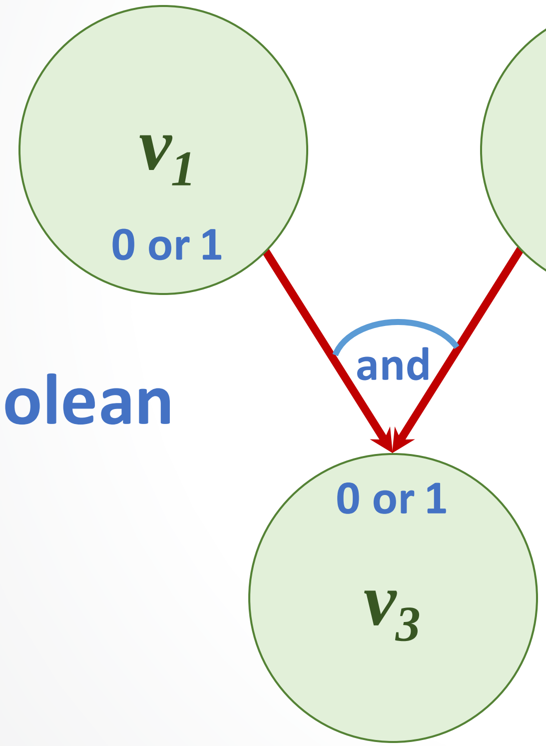
III. **Results:**

1. Discretized Trajectories and Total Perturbation.
2. Clustering of discretized trajectories, Error Estimation, and Coverage (first 10 BNs).
3. Learned BNs in case of Butachlor.

IV. **Summary**

Introduction

- Networks

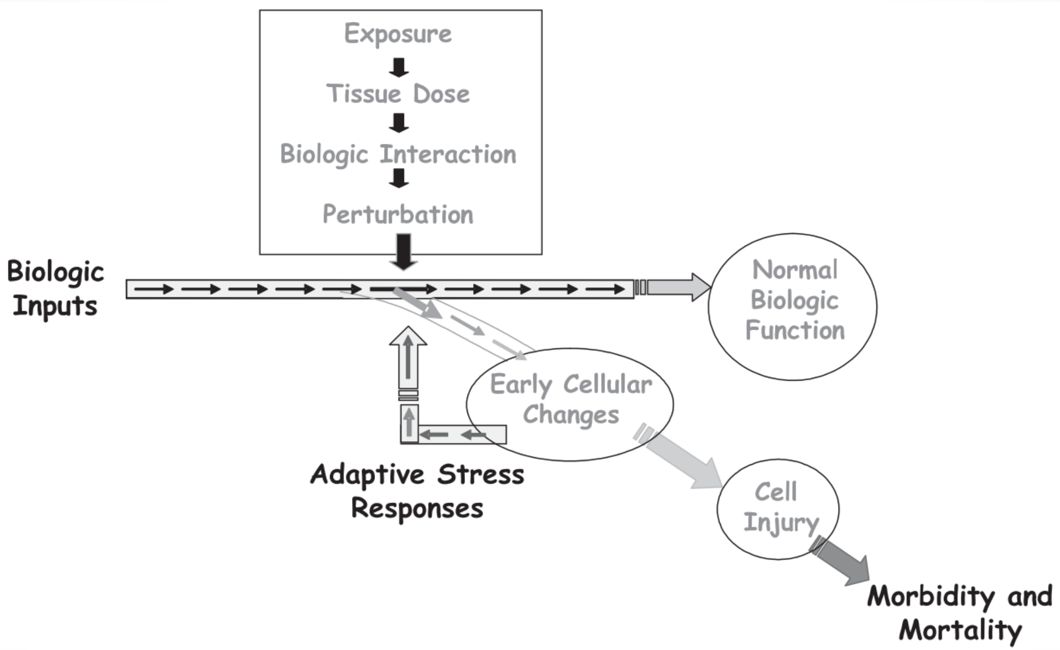


Nodes – proteins, genes,
small molecules, etc.

Boolean

Edges - *actions* of proteins,
genes, small molecules, etc.

Introduction



Krewski, Daniel, et al. "Toxicity testing in the 21st century: a vision and a strategy." *Journal of Toxicology and Environmental Health, Part B* 13.2-4 (2010): 51-138.

- **“Tipping point”** – system threshold between adaptation and adversity.
- Boolean networks (BN) are logical models of integrated cellular response pathways
- Here we reconstruct simple BN using high-content imaging data to analyze cellular tipping points

Shah, Imran, et al. "Using ToxCast™ Data to Reconstruct Dynamic Cell State Trajectories and Estimate Toxicological Points of Departure." *Environmental health perspectives* (2016)

Outline

I. Introduction

II. Methods:

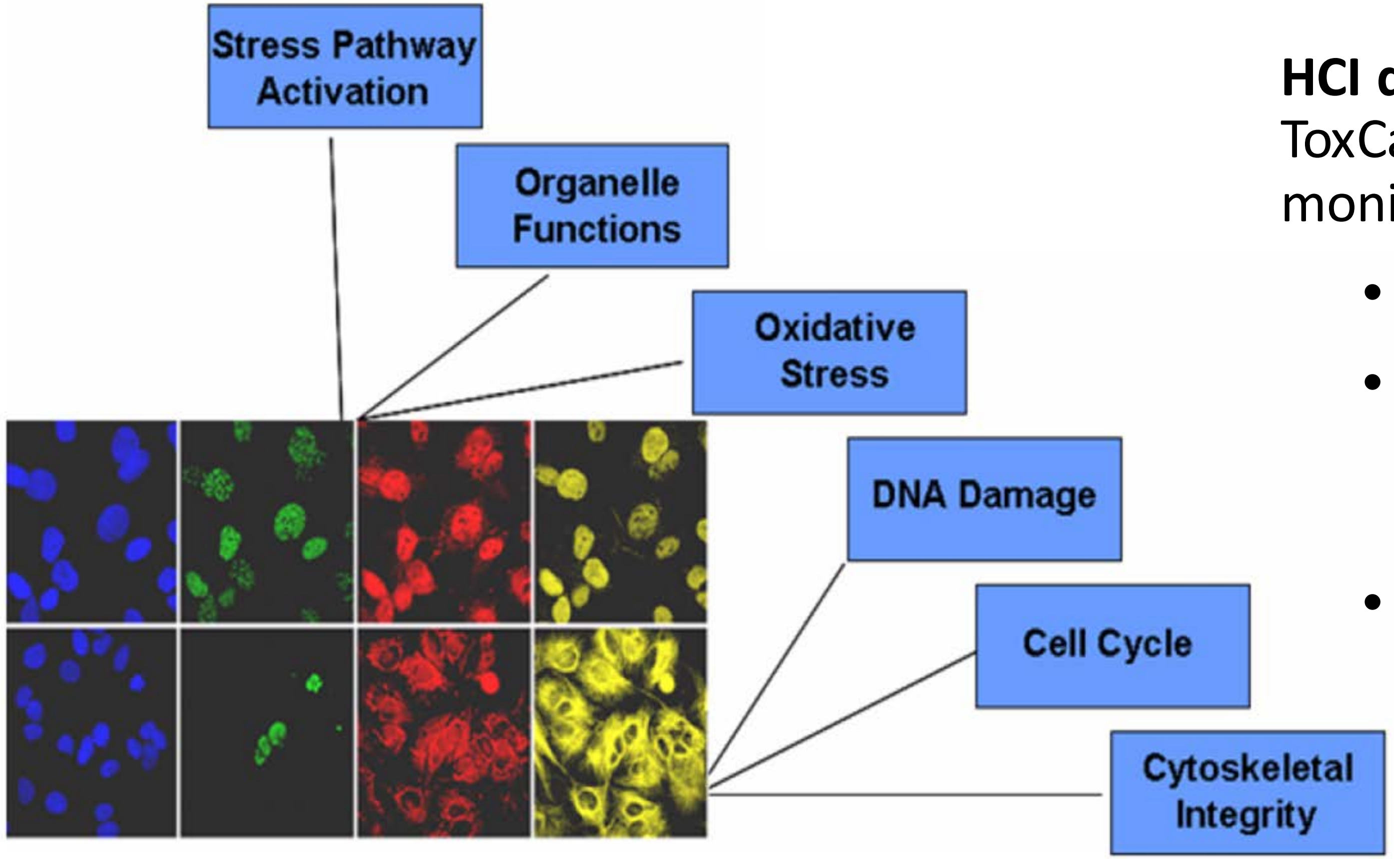
1. **Dataset**
2. **Data standardization, and Noise Threshold (z_0).**
3. **Data Discretization.**
4. **Learning Boolean Functions and Construction of Boolean Networks (BNs).**
5. **Needleman-Wunsch (NW) optimal global alignment, and Error Estimation.**
6. **Coverage.**

III. Results:

1. Discretized Trajectories and Total Perturbation.
2. Clustering of discretized trajectories, Error Estimation, and Coverage (first 10 BNs).
3. Learned BNs in case of Butachlor.

IV. Summary

1. Dataset - High Content Imaging (HCI)



Dataset:

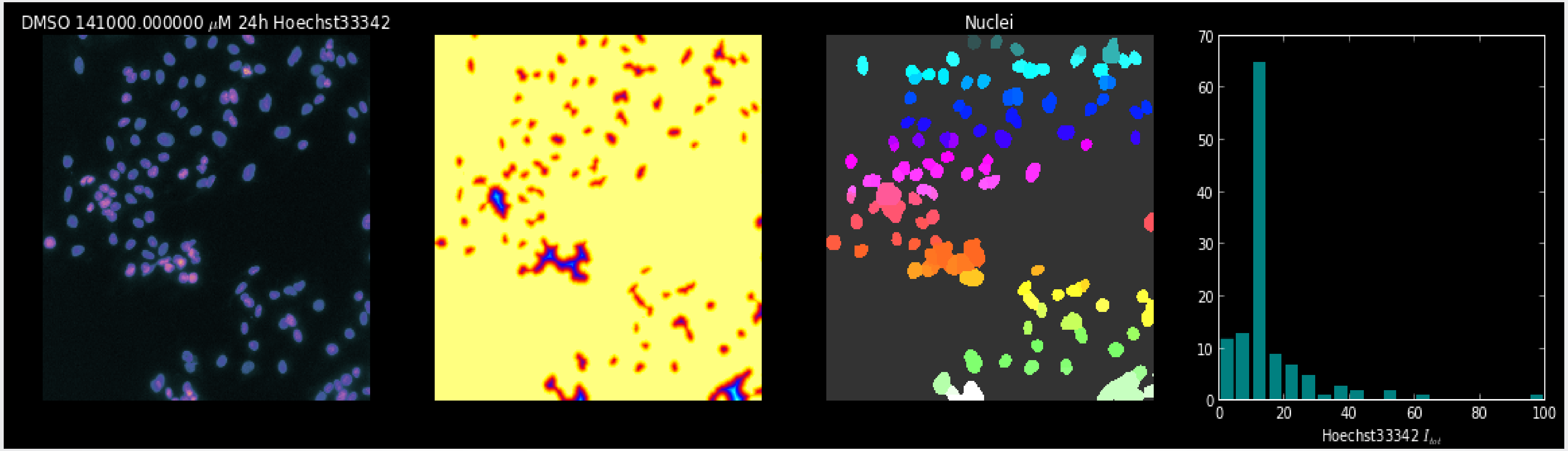
HCI data[1] were used to study the effect of 967 ToxCast chemicals on **HepG2** cell states by monitoring:

- **10 endpoints** across
- **multiple time points:**
  - **ToxCast I:** 1, 24, and 72h,
  - **ToxCast II**: 24 and 72h
- **10 concentrations** (0.4 to 200µM).

Shah, Imran, et al. "Using ToxCast™ Data to Reconstruct Dynamic Cell State Trajectories and Estimate Toxicological Points of Departure." *Environmental health perspectives* (2016)

1. Dataset - High Content Imaging (HCI)

- Image analysis and cell level features are conducted by Cyprotex Inc.



Raw Image (Hoechst) → Intensity Analysis → Object Identification → Nuclear intensity distribution

Shah, Imran, et al. "Using ToxCast™ Data to Reconstruct Dynamic Cell State Trajectories and Estimate Toxicological Points of Departure." *Environmental health perspectives* (2016)

1. Dataset - High Content Imaging (HCI)

- The following cellular endpoints were quantified:

1. phosphorylated p53 / p53 activation (**p53**),
2. phosphorylated c-Jun/c-Jun activation (**SK**),
3. phospho-Histone H2A.x (**OS**),
4. phospho-Histone H3 / mitotic arrest (**MA**),
5. phosphorylated α-tubulin / microtubules (**Mt**),
6. mitochondrial membrane potential (**MMP**),
7. mitochondrial mass (**MM**),
8. cell cycle arrest (**CCA**),
9. nuclear size (**NS**), and
10. cell number (**CN**).

Shah, Imran, et al. "Using ToxCast™ Data to Reconstruct Dynamic Cell State Trajectories and Estimate Toxicological Points of Departure." *Environmental health perspectives* (2016)

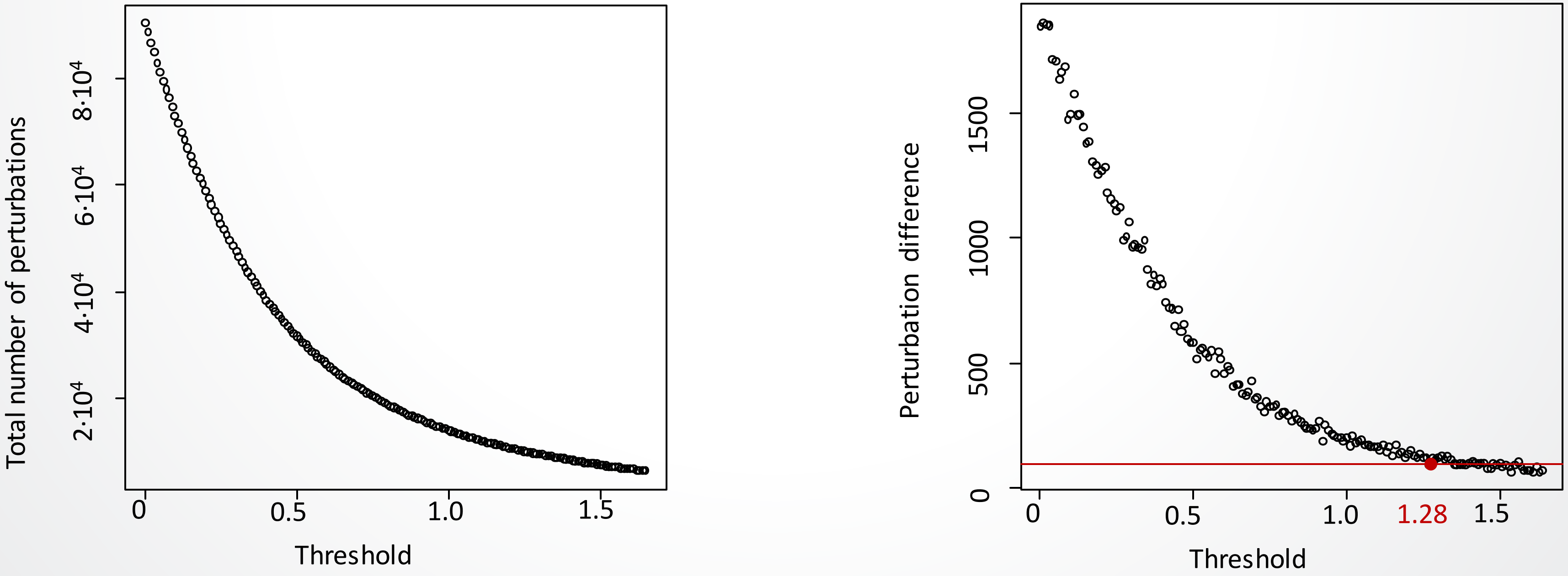
2. Data standardization, and Noise Threshold (z_0).

Data standardization:

$$z = \frac{x - x^*}{\sigma_x}$$

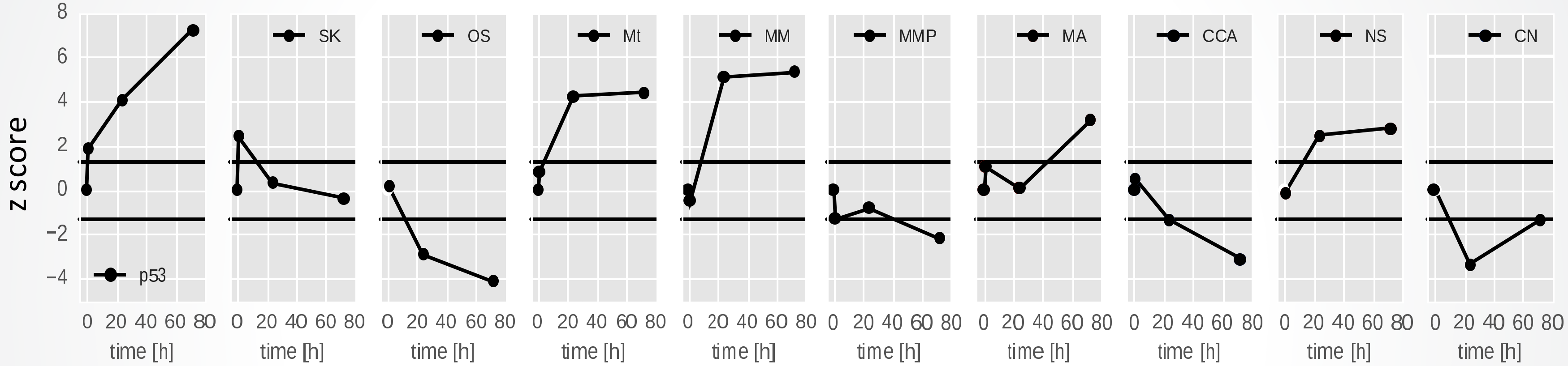
x – $\log_2$ transformed fold change
x^* – the median value
σ_x – the standard deviation

Noise Threshold z_0:



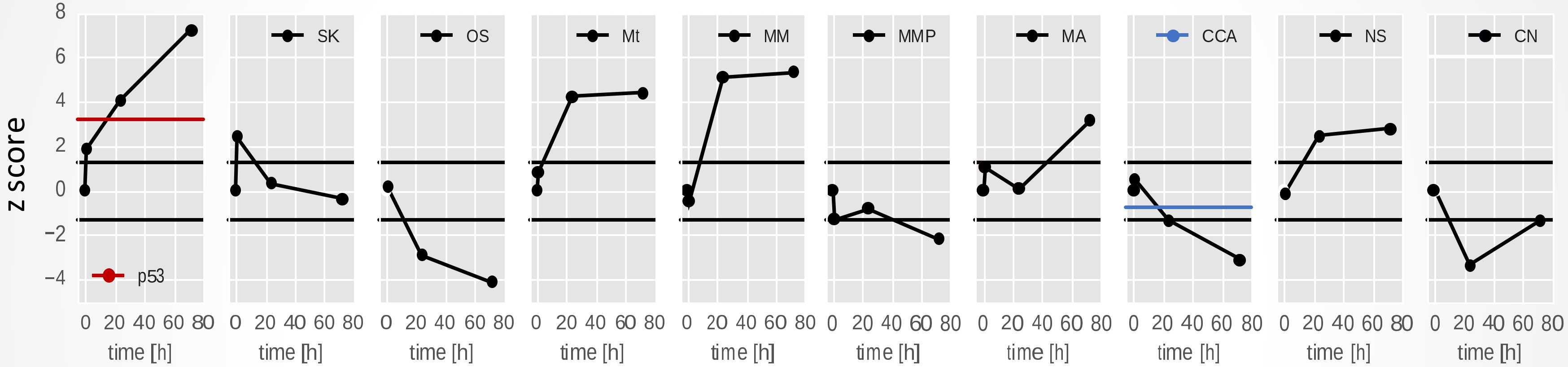
3. Data Discretization.

Motivation: Increase of p53 causes decrease in OS, CCA or CN

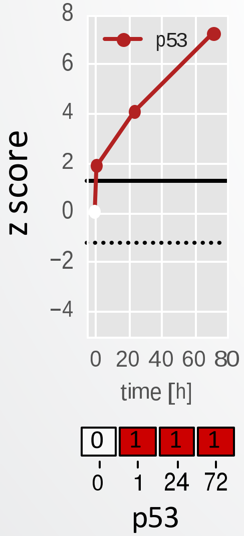


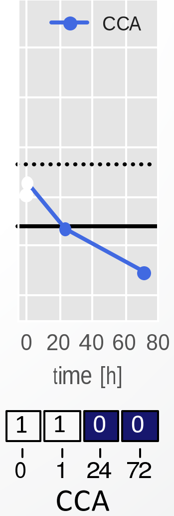
3. Data Discretization.

Endpoint Trend Assessment: Calculate an average perturbation value



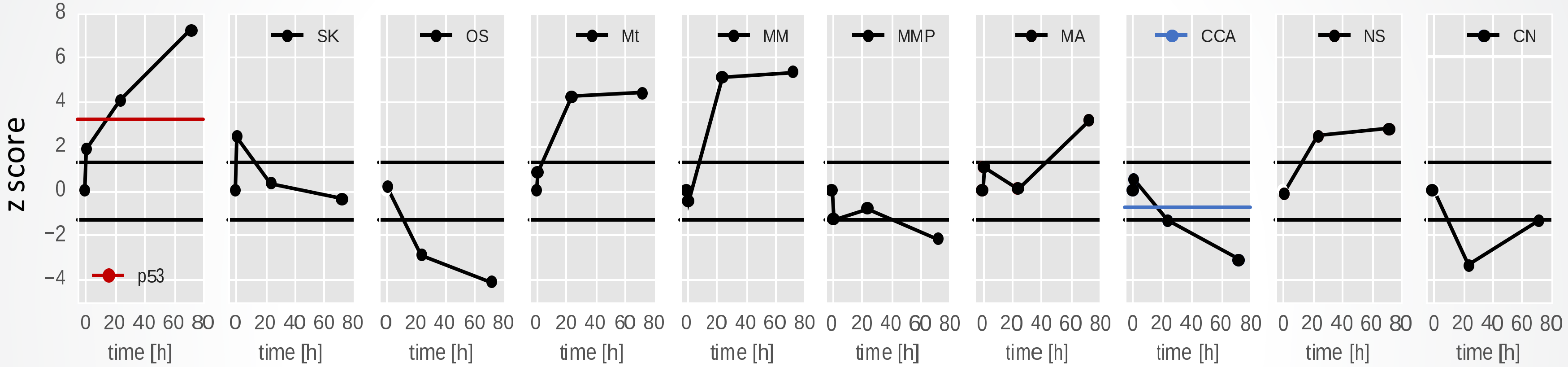
Discretization based on threshold value



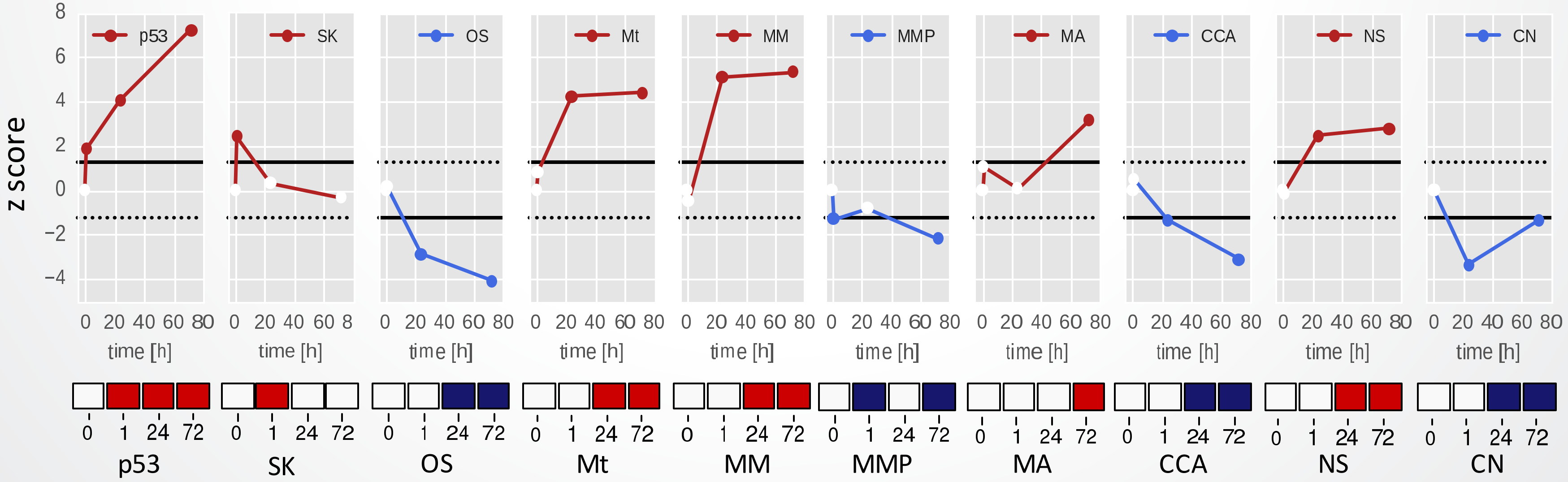


3. Data Discretization.

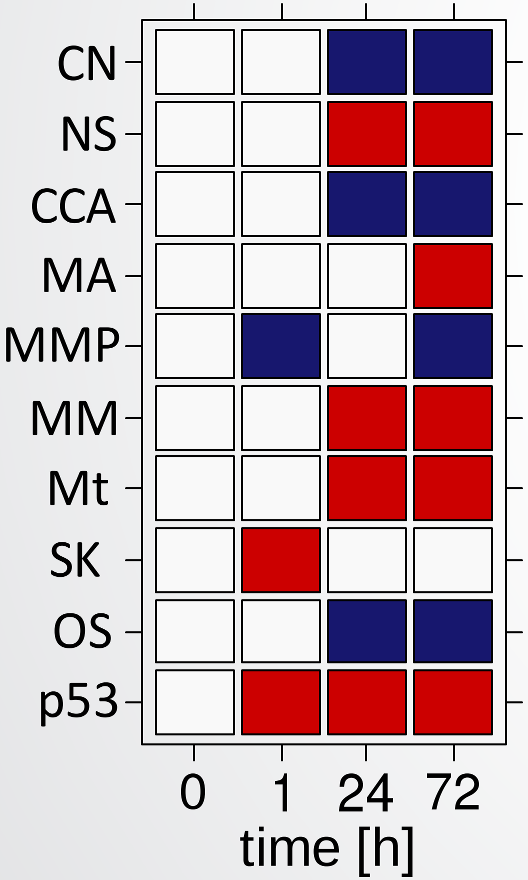
Endpoint Trend Assessment: Calculate an average perturbation value



Discretization based on threshold value

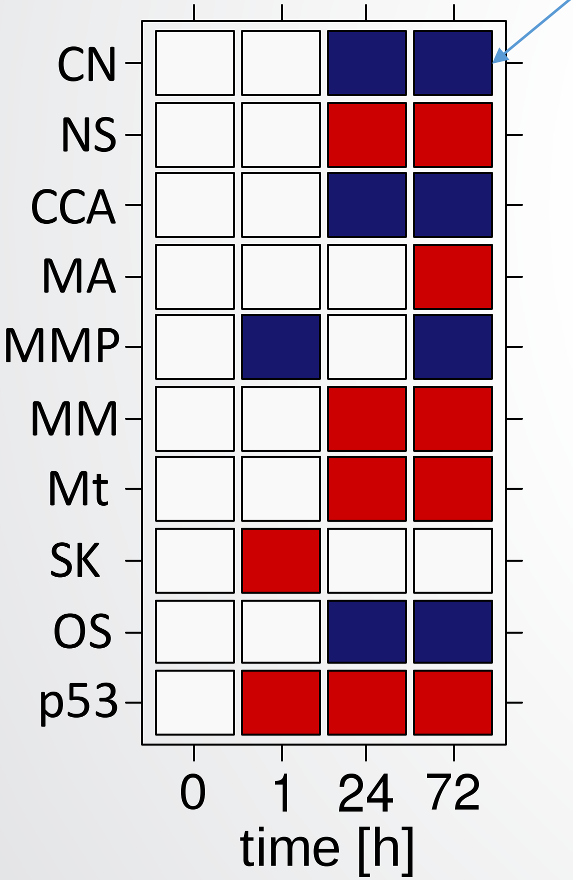


4. Learning Boolean Functions and Construction of Boolean Networks (BNs).



	0	1	24	72
CN	⬜	⬜	🟦	🟦
NS	⬜	⬜	🟥	🟥
CCA	⬜	⬜	🟦	🟦
MA	⬜	⬜	⬜	🟥
MMP	⬜	🟦	⬜	🟦
MM	⬜	⬜	🟥	🟥
Mt	⬜	⬜	🟥	🟥
SK	⬜	🟥	⬜	⬜
OS	⬜	⬜	🟦	🟦
p53	⬜	🟥	🟥	🟥

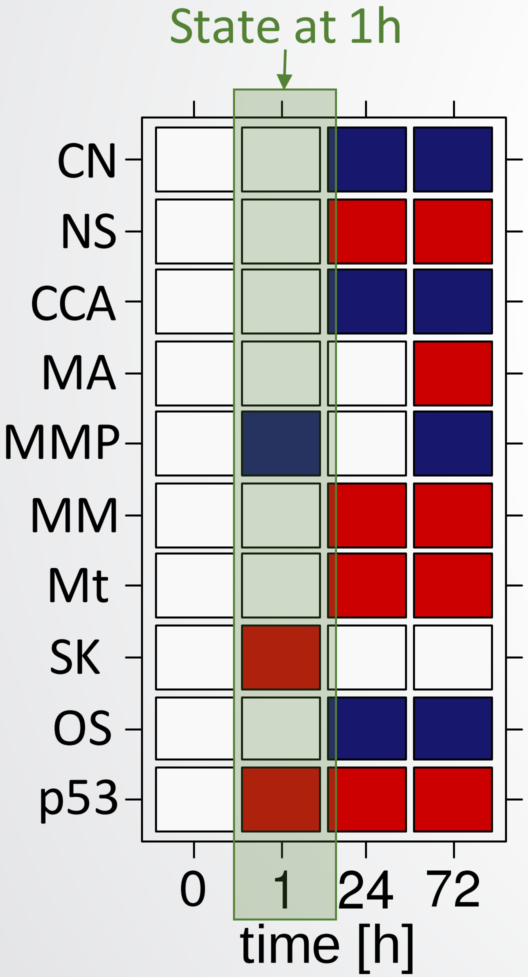
4. Learning Boolean Functions and Construction of Boolean Networks (BNs).



	0	1	24	72
CN	⬜	⬜	🟦	🟦
NS	⬜	⬜	🟥	🟥
CCA	⬜	⬜	🟦	🟦
MA	⬜	⬜	⬜	🟥
MMP	⬜	🟦	⬜	🟦
MM	⬜	⬜	🟥	🟥
Mt	⬜	⬜	🟥	🟥
SK	⬜	🟥	⬜	⬜
OS	⬜	⬜	🟦	🟦
p53	⬜	🟥	🟥	🟥

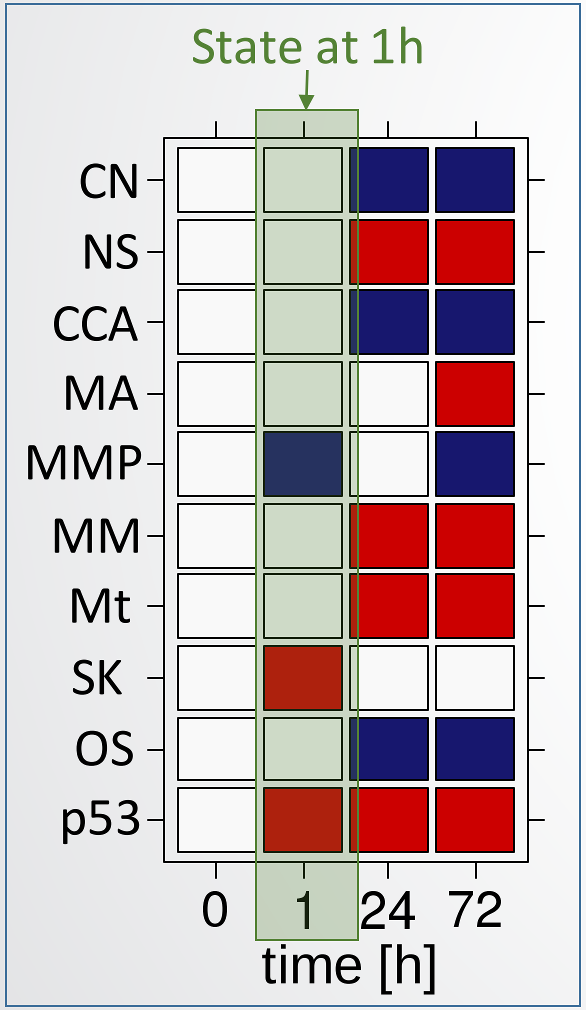
Endpoint perturbation

4. Learning Boolean Functions and Construction of Boolean Networks (BNs).



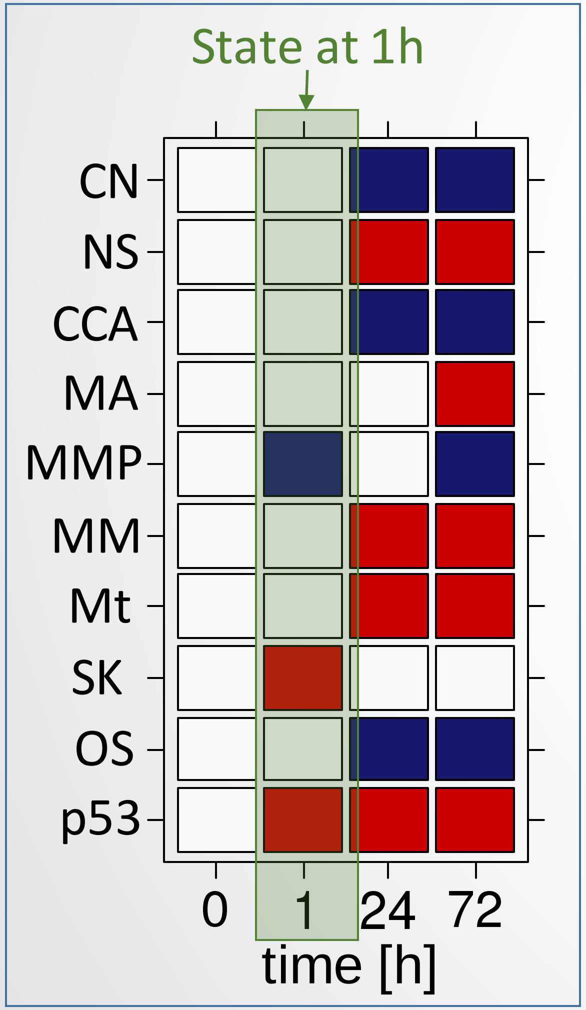
4. Learning Boolean Functions and Construction of Boolean Networks (BNs).

Discretized Trajectory of HepG2
after application of Butachlor at 200uM



4. Learning Boolean Functions and Construction of Boolean Networks (BNs).

Discretized Trajectory of HepG2
after application of Butachlor at 200uM



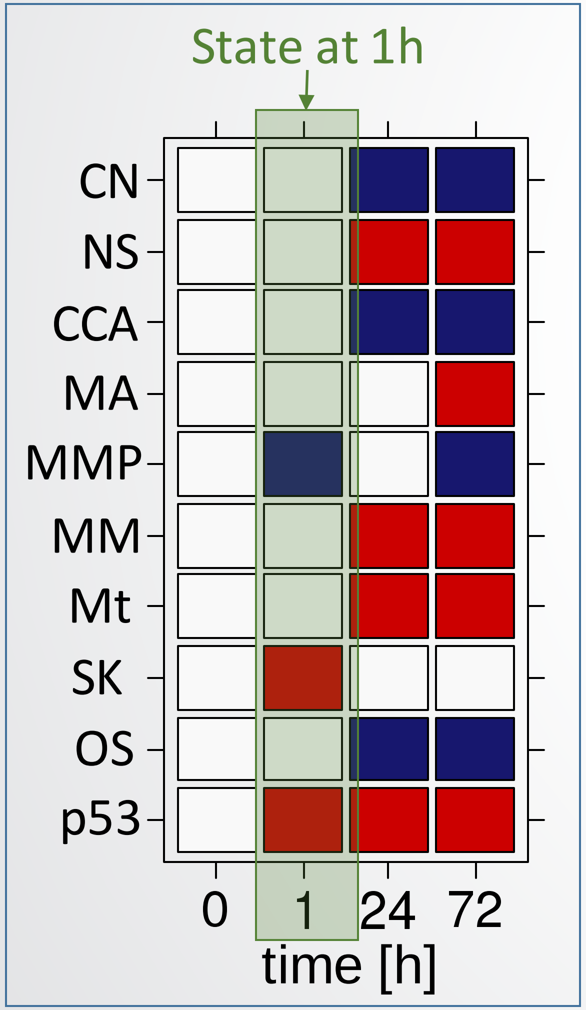
Find Boolean Functions

“Best-Fit Extension”

$$\mathbf{F}_{\mathbf{CN}} \in [f^{1}_{CN}, \ldots, f^{i}_{CN}]$$
$$\mathbf{F}_{\mathbf{NS}} \in [f^{1}_{NS}, \ldots, f^{j}_{NS}]$$
$$\mathbf{F}_{\mathbf{CCA}} \in [f^{1}_{CCA}, \ldots, f^{k}_{CCA}]$$
$$\mathbf{F}_{\mathbf{MA}} \in [f^{1}_{MA}, \ldots, f^{l}_{MA}]$$
$$\mathbf{F}_{\mathbf{MMP}} \in [f^{1}_{MMP}, \ldots, f^{m}_{MMP}]$$
$$\mathbf{F}_{\mathbf{MM}} \in [f^{1}_{MM}, \ldots, f^{n}_{MM}]$$
$$\mathbf{F}_{\mathbf{Mt}} \in [f^{1}_{Mt}, \ldots, f^{o}_{Mt}]$$
$$\mathbf{F}_{\mathbf{SK}} \in [f^{1}_{SK}, \ldots, f^{p}_{SK}]$$
$$\mathbf{F}_{\mathbf{OS}} \in [f^{1}_{OS}, \ldots, f^{q}_{OS}]$$
$$\mathbf{F}_{\mathbf{p53}} \in [f^{1}_{p53}, \ldots, f^{r}_{p53}]$$

4. Learning Boolean Functions and Construction of Boolean Networks (BNs).

Discretized Trajectory of HepG2
after application of Butachlor at 200uM

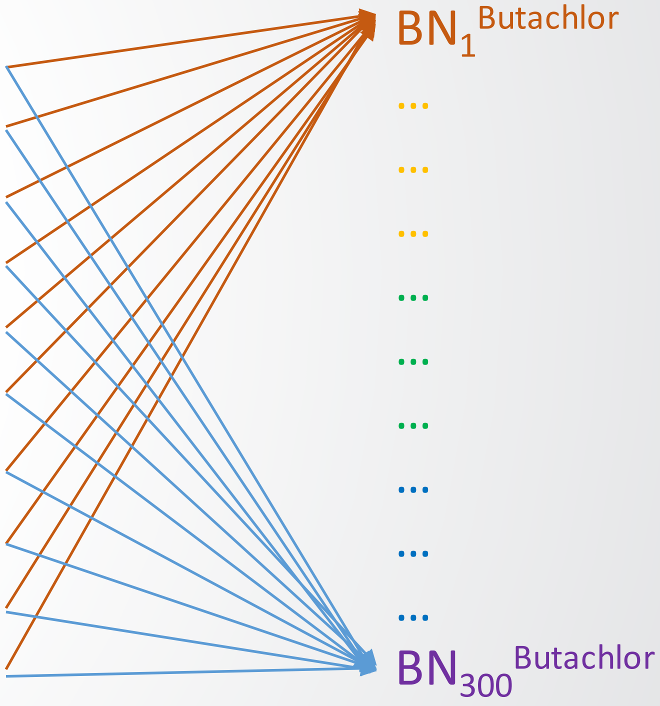


Find Boolean Functions

"Best-Fit Extension"

$$
\begin{aligned}
\mathbf{F}_{CN} &\in [f^{1}_{CN}, \ldots, f^{i}_{CN}] \\
\mathbf{F}_{NS} &\in [f^{1}_{NS}, \ldots, f^{j}_{NS}] \\
\mathbf{F}_{CCA} &\in [f^{1}_{CCA}, \ldots, f^{k}_{CCA}] \\
\mathbf{F}_{MA} &\in [f^{1}_{MA}, \ldots, f^{l}_{MA}] \\
\mathbf{F}_{MMP} &\in [f^{1}_{MMP}, \ldots, f^{m}_{MMP}] \\
\mathbf{F}_{MM} &\in [f^{1}_{MM}, \ldots, f^{n}_{MM}] \\
\mathbf{F}_{Mt} &\in [f^{1}_{Mt}, \ldots, f^{o}_{Mt}] \\
\mathbf{F}_{SK} &\in [f^{1}_{SK}, \ldots, f^{p}_{SK}] \\
\mathbf{F}_{OS} &\in [f^{1}_{OS}, \ldots, f^{q}_{OS}] \\
\mathbf{F}_{p53} &\in [f^{1}_{p53}, \ldots, f^{r}_{p53}]
\end{aligned}
$$

Find Boolean Networks



$BN_1^{Butachlor}$

...

...

...

...

...

...

...

...

...

$BN_{300}^{Butachlor}$

5. Needleman–Wunsch optimal global alignment, and Error Estimation

Error in BN prediction was estimated as the sum of the Hamming distances* between observed and predicted discretized trajectories.

		endpoints											Butachlor 25uM			
													0	1h	24h	72h
endpoints		p53											0	0	1	1
			SK										0	0	0	0
				OS									0	0	0	0
					Mt								0	0	0	1
						MM							0	0	1	1
							MMP						0	0	0	0
								MA					0	0	0	0
									CCA				0	0	0	0
										NS			0	0	0	1
											CN		0	0	0	0
												dec	0	0	544	610
BN4 prediction	1	0	0	0	0	0	0	0	0	0	0	0	10	10	8	6
	2	1	1	1	0	0	1	1	1	1	0	926	3	3	3	3
	3	0	0	0	0	1	0	0	0	1	0	34	8	8	8	8
	4	1	1	1	0	0	0	1	0	1	0	906	5	5	5	5
	5	1	0	0	0	1	0	0	1	1	0	550	6	6	8	8
	6	1	0	0	1	1	0	1	0	1	0	618	5	5	7	9
	7	1	0	0	0	1	0	1	0	1	0	554	6	6	8	8
		the Hamming distances:											0	2	2	1

---

* The Hamming distance between two states is the number of positions at which the states are different

6. Coverage

I. Error Estimation is performed:

1. **For each trajectory** – During this step we split BNs with the lowest error ("**the baseline error**") from BNs with higher error.
2. **Across all trajectories** – During this analysis we estimated the number of trajectories predicted by each BN with an accuracy ≤ to the baseline error ("**coverage**").

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

1 – BN covers traj.

0 – BN does not cover traj.

BN1 Coverage = 4 traj.

BN2 Coverage = 2 traj.

BN3 Coverage = 2 traj.

II. **The smallest set of BNs** that covers all trajectories was inferred by selecting BNs with the largest coverage.

Outline

I. **Introduction**

II. **Methods:**

   1. Dataset
   2. Data standardization, and Noise Threshold (z_0).
   3. Data Discretization.
   4. Learning Boolean Functions and Construction of Boolean Networks (BNs).
   5. Needleman-Wunsch (NW) optimal global alignment, and Error Estimation.
   6. Coverage.

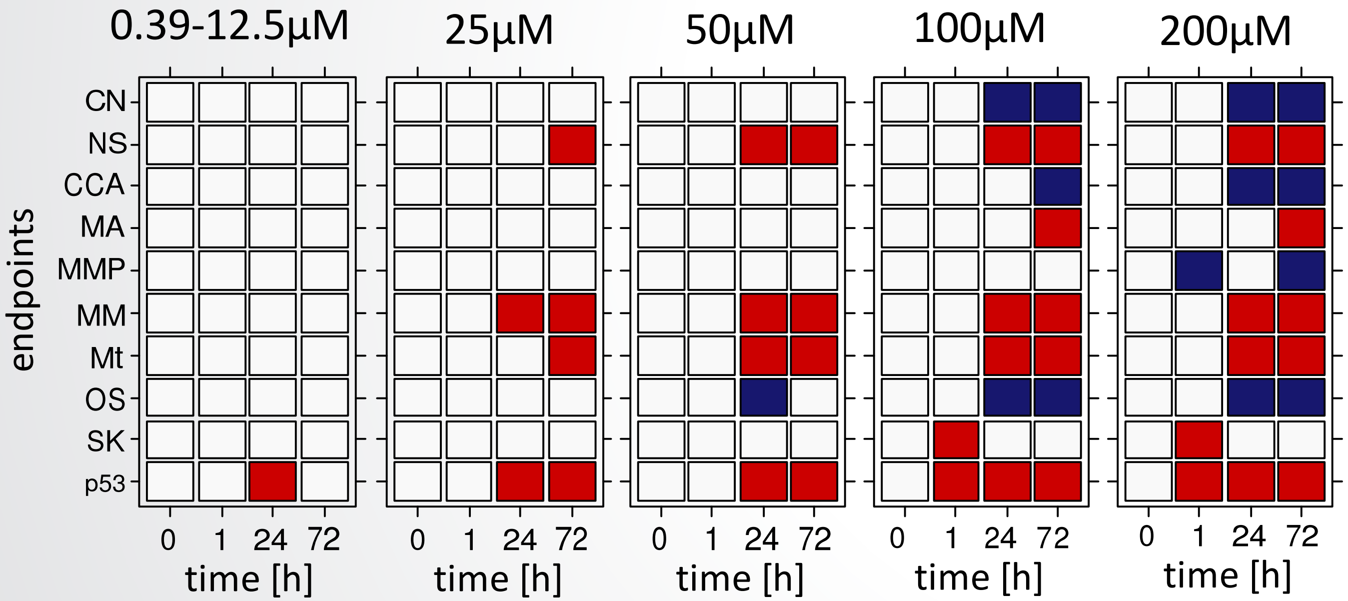
III. **Results:**

   1. **Discretized Trajectories and Total Perturbation.**
   2. **Clustering of discretized trajectories, Error Estimation, and Coverage (first 10 BNs).**
   3. **Learned BNs in case of Butachlor.**

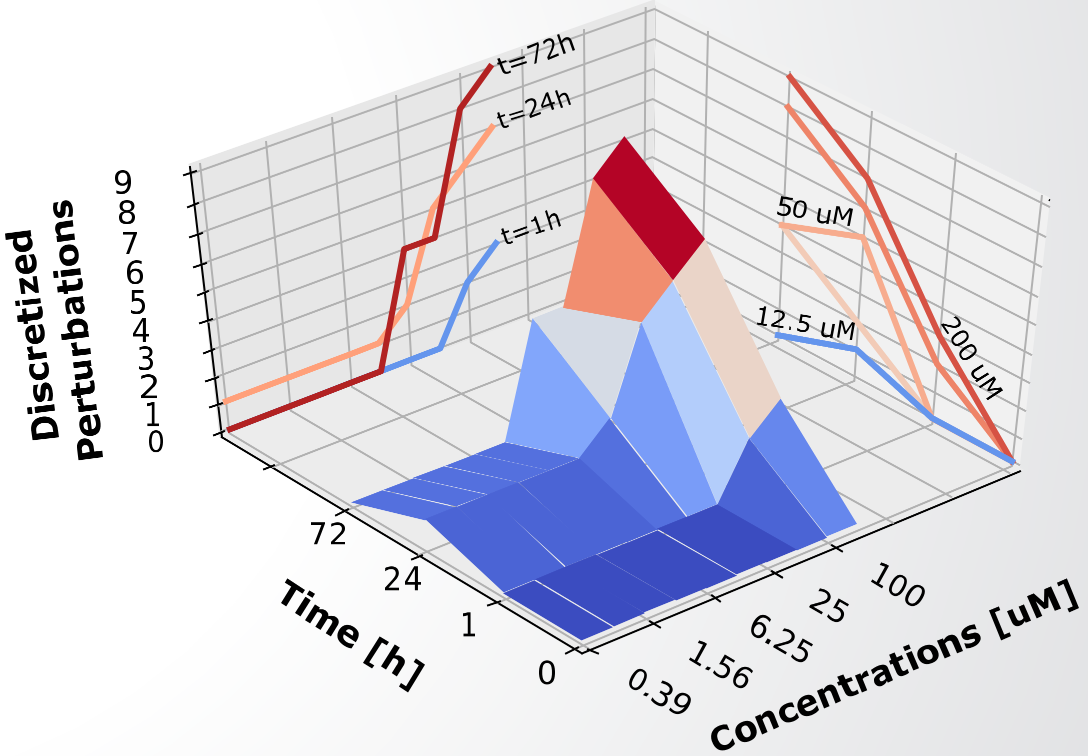
IV. **Summary**

1. Discretized Trajectories and Total Perturbation

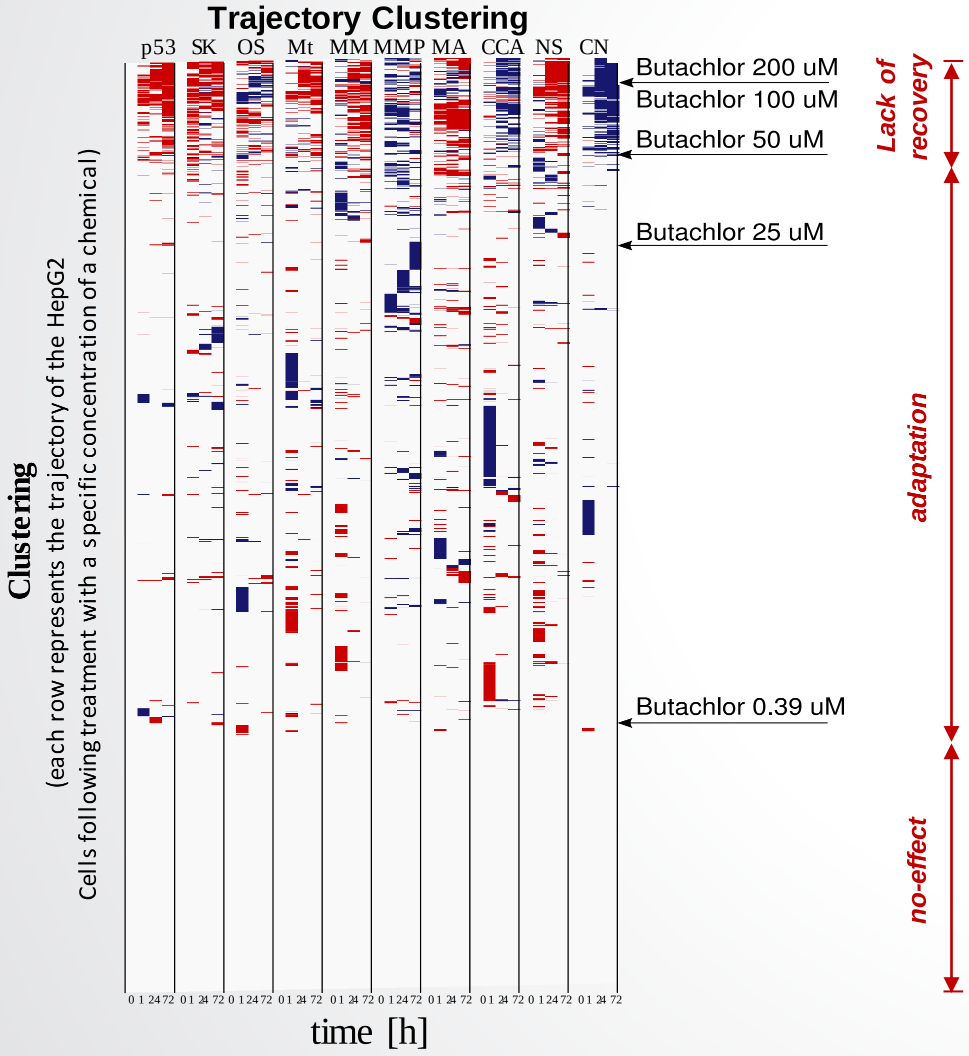
Example: Butachlor - one of the most commonly used herbicides in agriculture.



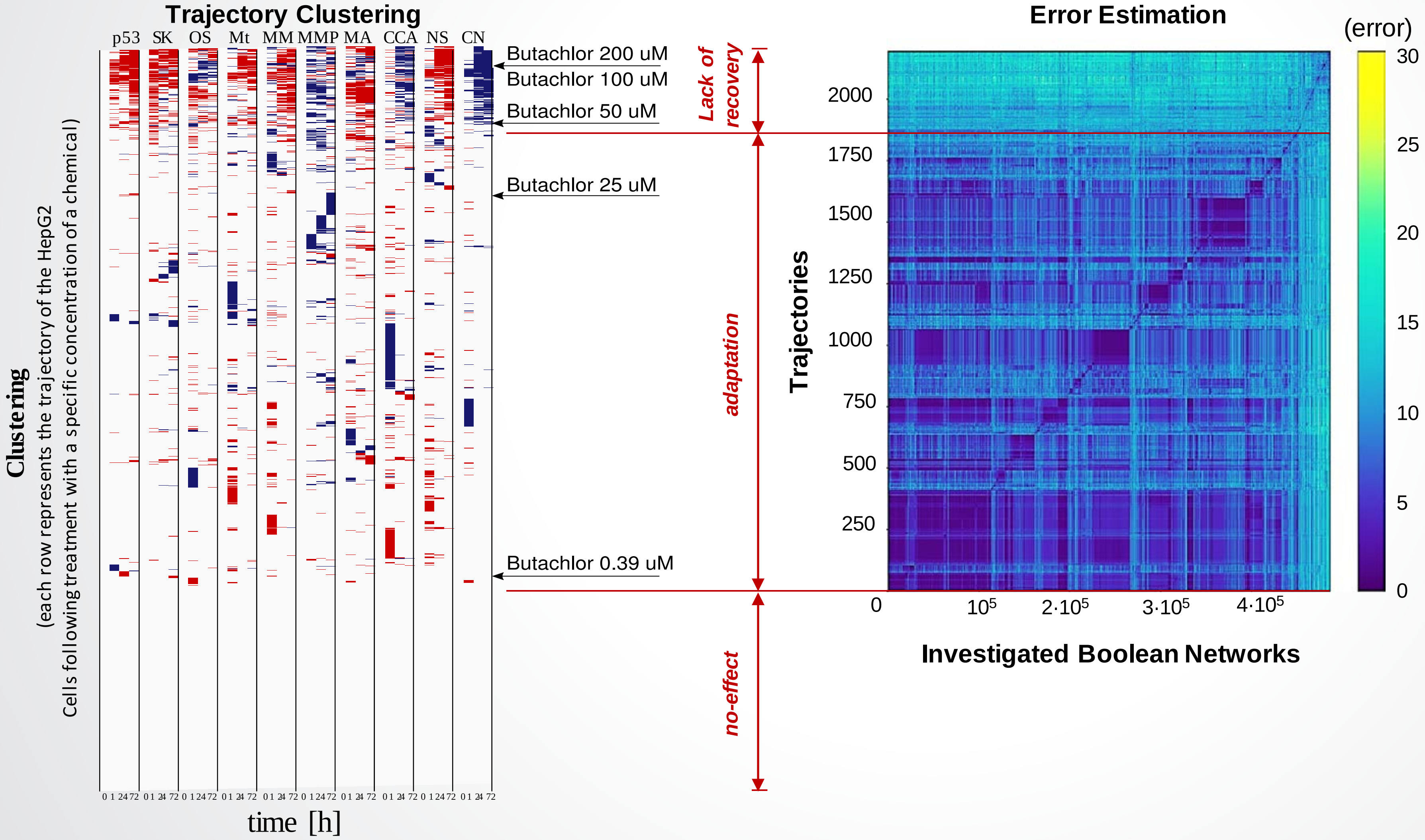
endpoints	0.39-12.5µM				25µM				50µM				100µM				200µM			
time [h]	0	1	24	72	0	1	24	72	0	1	24	72	0	1	24	72	0	1	24	72
CN															🟦	🟦			🟦	🟦
NS								🟥			🟥	🟥			🟥	🟥			🟥	🟥
CCA																🟦			🟦	🟦
MA																🟥				🟥
MMP																		🟦		🟦
MM							🟥	🟥			🟥	🟥			🟥	🟥			🟥	🟥
Mt								🟥			🟥	🟥			🟥	🟥			🟥	🟥
OS											🟦				🟦	🟦			🟦	🟦
SK														🟥				🟥		
p53			🟥				🟥	🟥			🟥	🟥		🟥	🟥	🟥		🟥	🟥	🟥



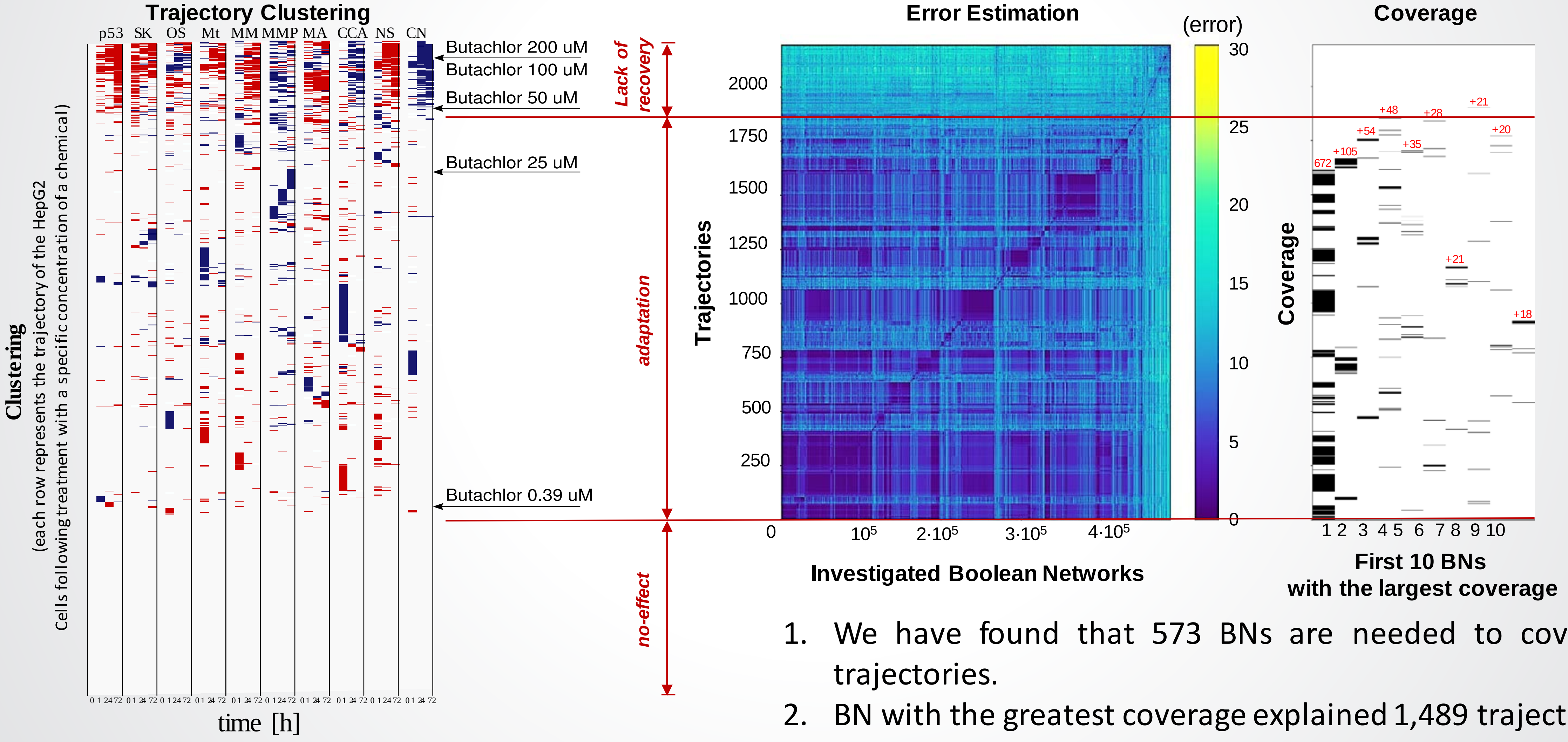
2. Clustering of discretized trajectories, Error Estimation, and Coverage



2. Clustering of discretized trajectories, Error Estimation, and Coverage



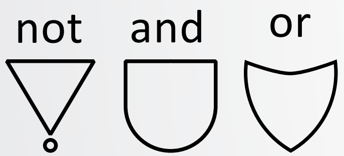
2. Clustering of discretized trajectories, Error Estimation, and Coverage

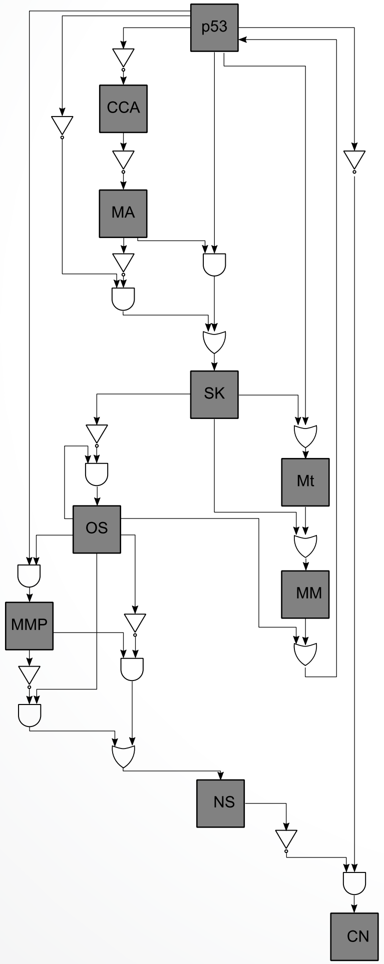


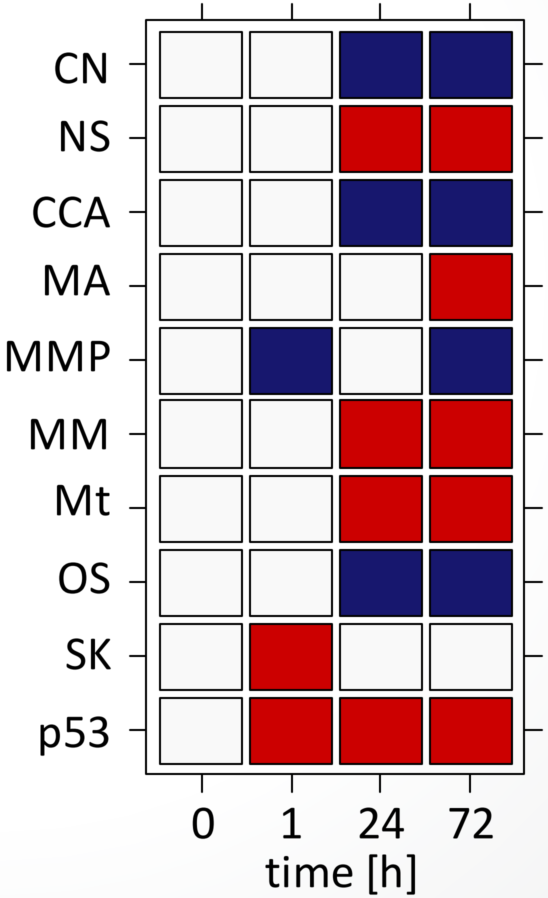
1. We have found that 573 BNs are needed to cover all trajectories.
2. BN with the greatest coverage explained 1,489 trajectories.

3. Learned BNs in case of Butachlor

Butachlor 200µM

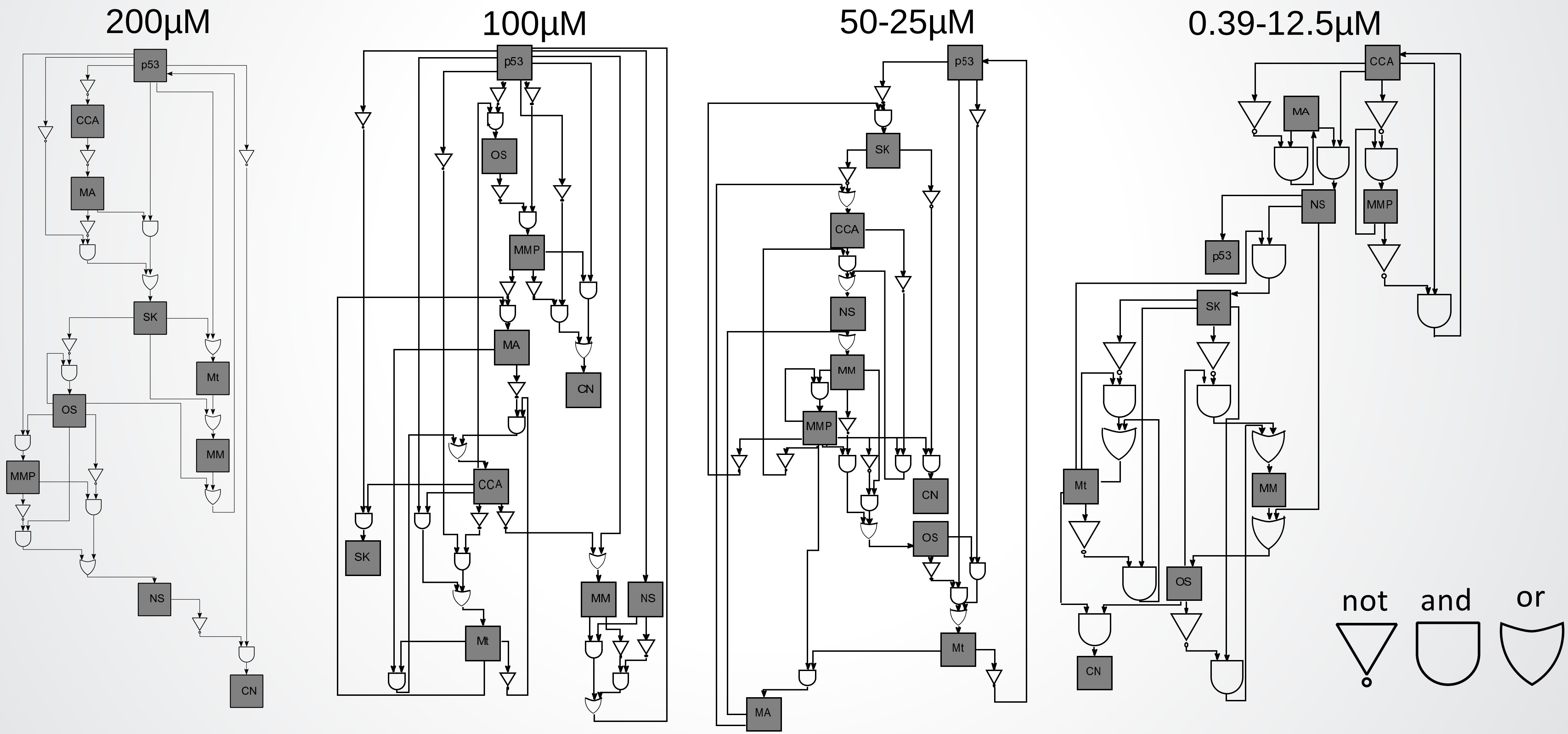






3. Learned BNs in case of Butachlor

Butachlor



200µM

100µM

50-25µM

0.39-12.5µM

Outline

- **I. Introduction**
- **II. Methods:**
  1. Dataset
  2. Data standardization, and Noise Threshold (z_0).
  3. Data Discretization.
  4. Learning Boolean Functions and Construction of Boolean Networks (BNs).
  5. Needleman-Wunsch (NW) optimal global alignment, and Error Estimation.
  6. Coverage.
- **III. Results:**
  1. Discretized Trajectories and Total Perturbation.
  2. Clustering of discretized trajectories, Error Estimation, and Coverage (first 10 BNs).
  3. Learned BNs in case of Butachlor.
- **IV. Summary**

IV. Summary

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.

Thank you