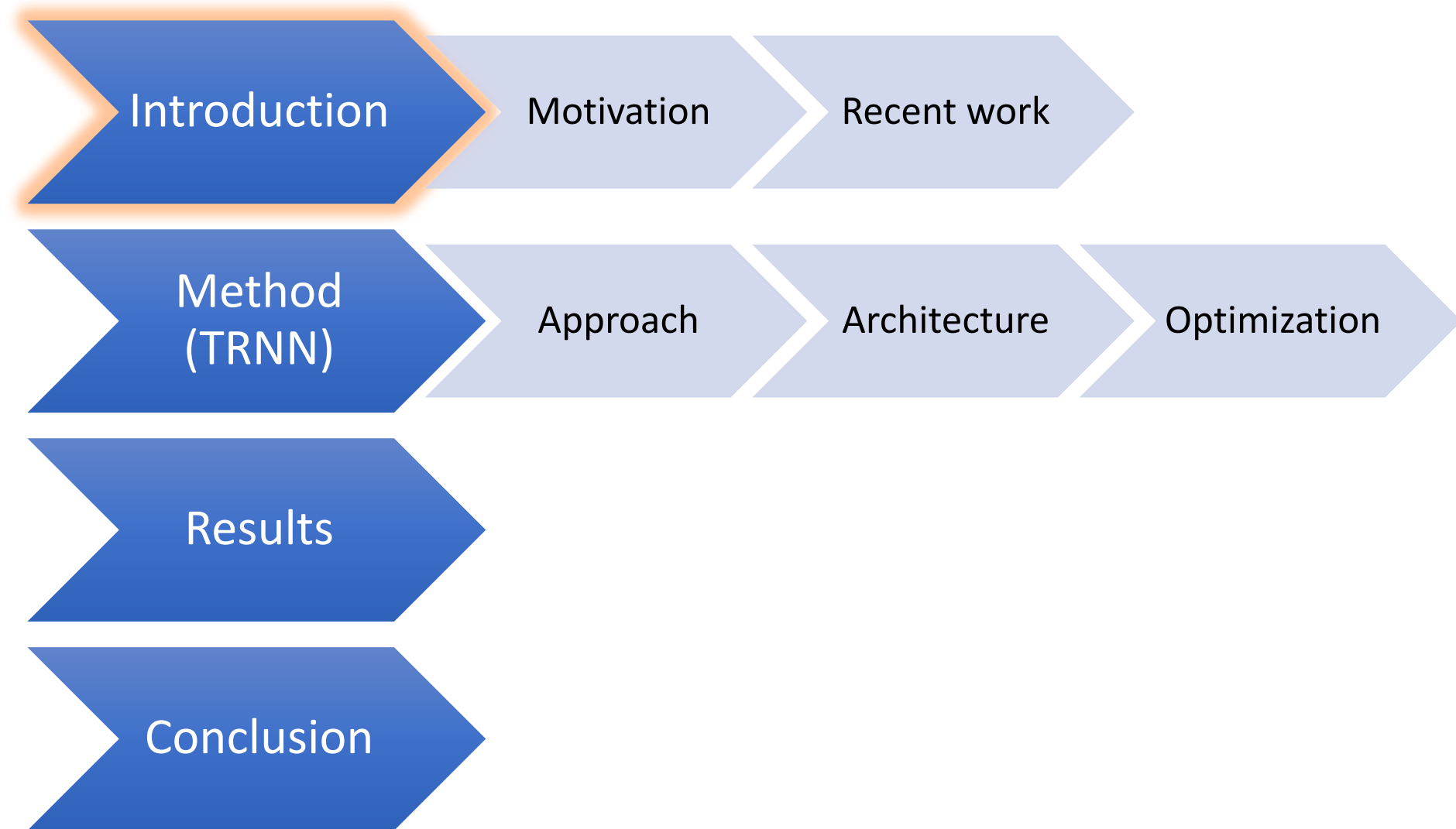


Towards deep mechanistic neural networks for multi-omics modeling.

Ameen Eetemadi
Tagkopoulos Lab, UC Davis

Outline



Motivation

GOAL: Predict cellular state and behavior from environmental settings and genetic background of the cell.

Genome-scale model



Image credit: Erik Jacobsen, Covert Lab

Movivation

GOAL: Predict cellular state and behavior from environmental settings and genetic background of the cell.

Synthetic biology

Test gene circuit in host cell
without experiment

Genome-scale model

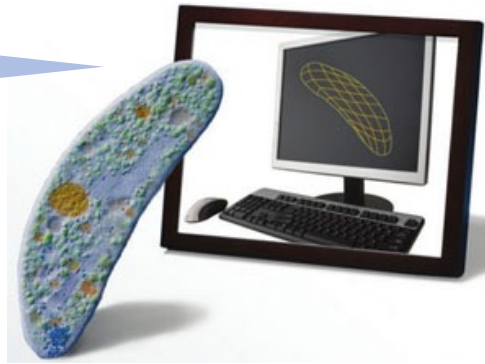


Image credit: Erik Jacobsen, Covert Lab

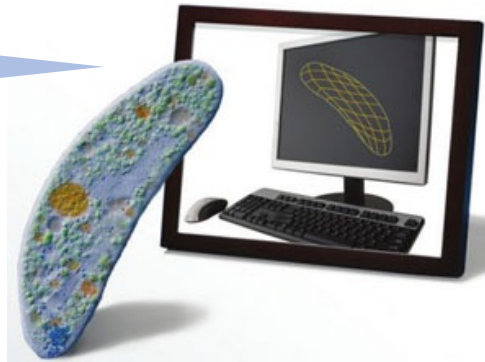
Movivation

GOAL: Predict cellular state and behavior from environmental settings and genetic background of the cell.

Synthetic biology

Test gene circuit in host cell
without experiment

Genome-scale model



Metabolic engineering

Find environment that maximizes
engineered products

Image credit: Erik Jacobsen, Covert Lab

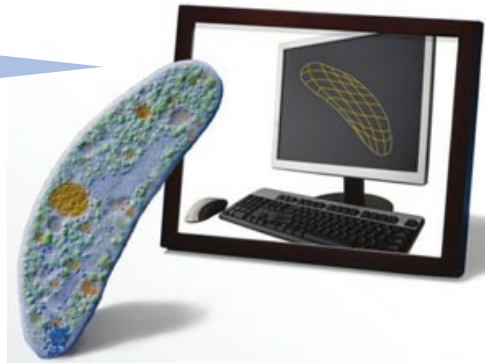
Movivation

GOAL: Predict cellular state and behavior from environmental settings and genetic background of the cell.

Synthetic biology

Test gene circuit in host cell
without experiment

Genome-scale model



Metabolic engineering

Find environment that maximizes
engineered products

Predictive medicine

Optimization of antibiotic effect
from ~100 existing antibiotics

Image credit: Erik Jacobsen, Covert Lab

Movivation

GOAL: Predict cellular state and behavior from environmental settings and genetic background of the cell.

Synthetic biology

Test gene circuit in host cell
without experiment

Genome-scale model

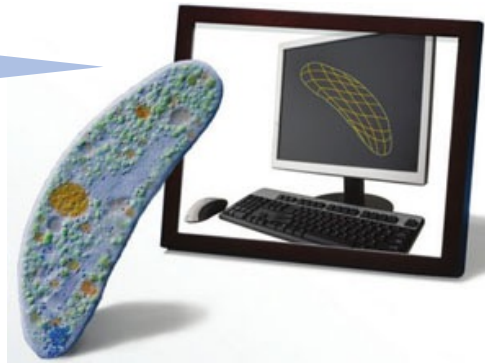


Image credit: Erik Jacobsen, Covert Lab

Metabolic engineering

Find environment that maximizes
engineered products

Predictive medicine

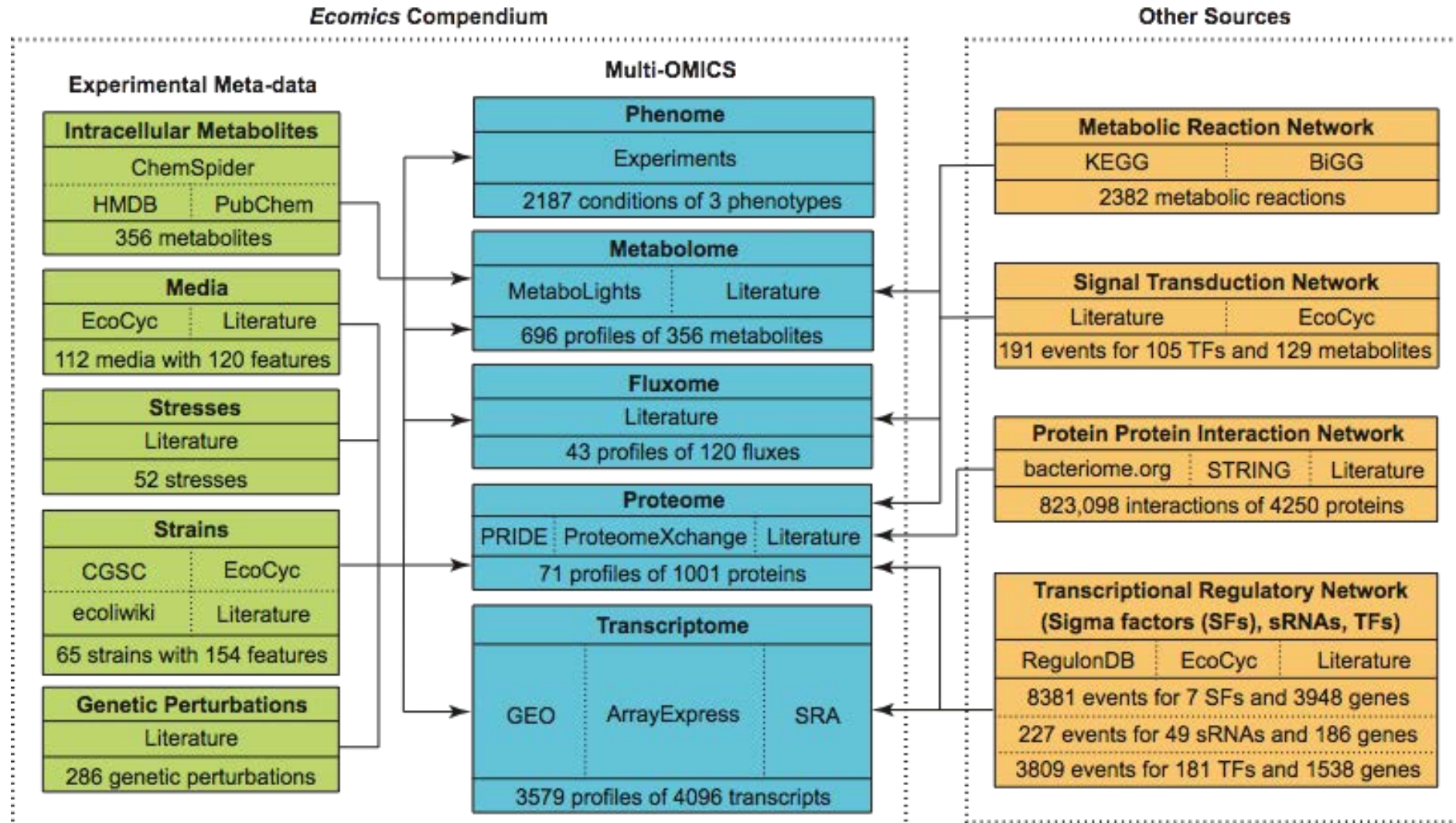
Optimization of antibiotic effect
from ~100 existing antibiotics

Guided experimentation

Find the conditions
with highest uncertainty

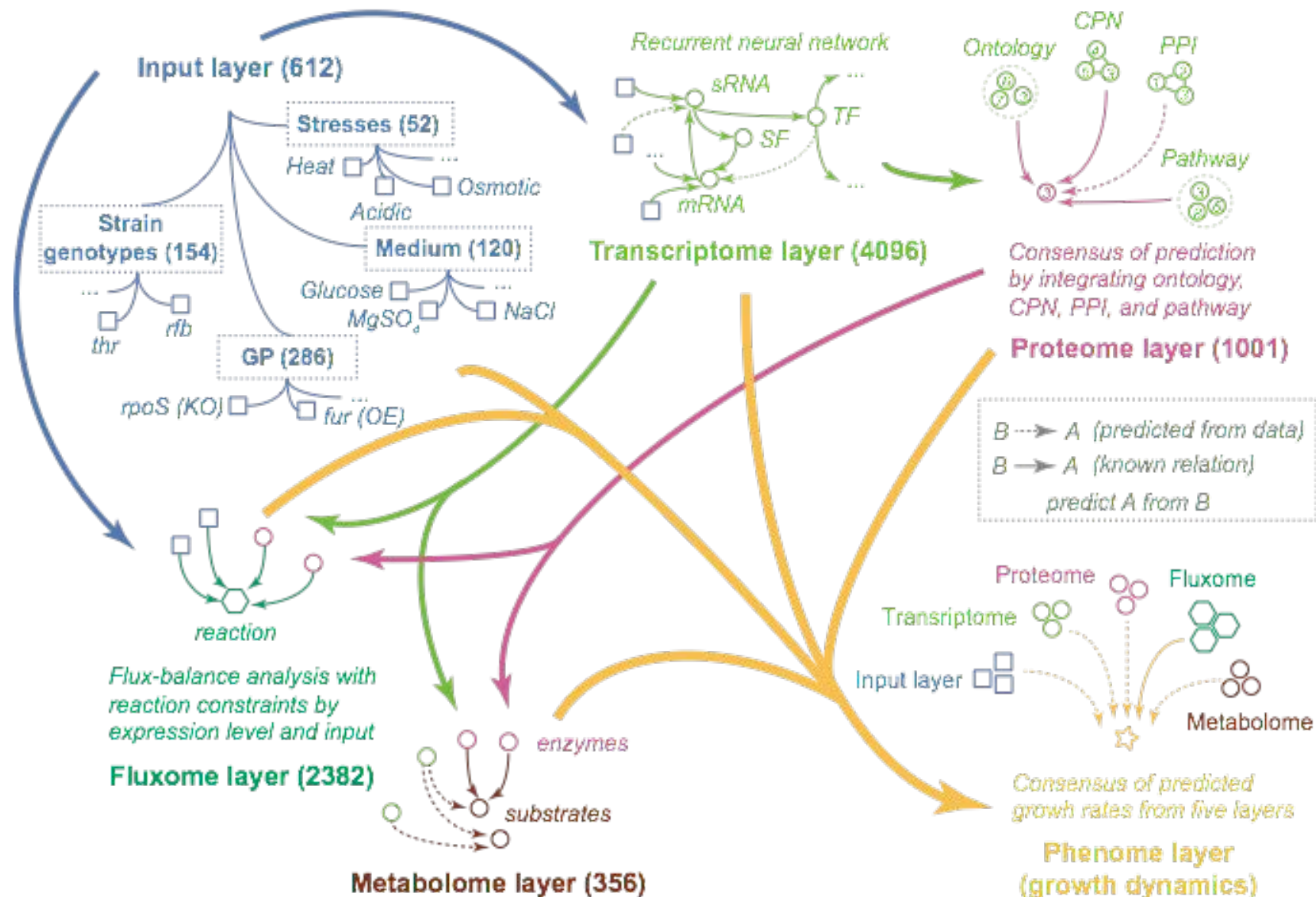
Recent work (ECOMICS) - Data

Most comprehensive compendium for an organism



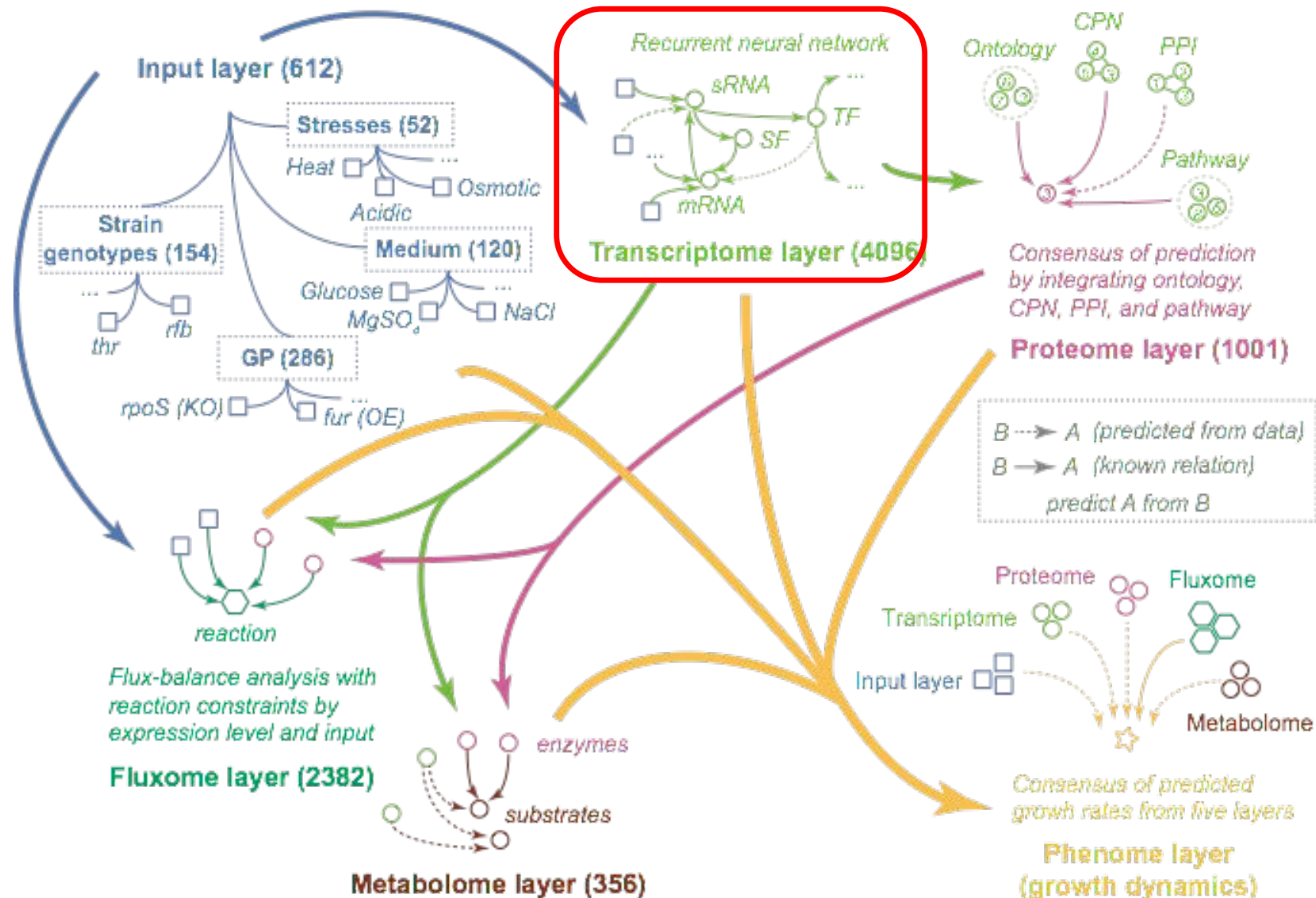
M. Kim, N. Rai, V. Zorraquino, and I. Tagkopoulos, "Multi-omics integration accurately predicts cellular state in unexplored conditions for *Escherichia coli*", *Nature communications* (2016)

Recent work (ECOMICS) – Predictive Model



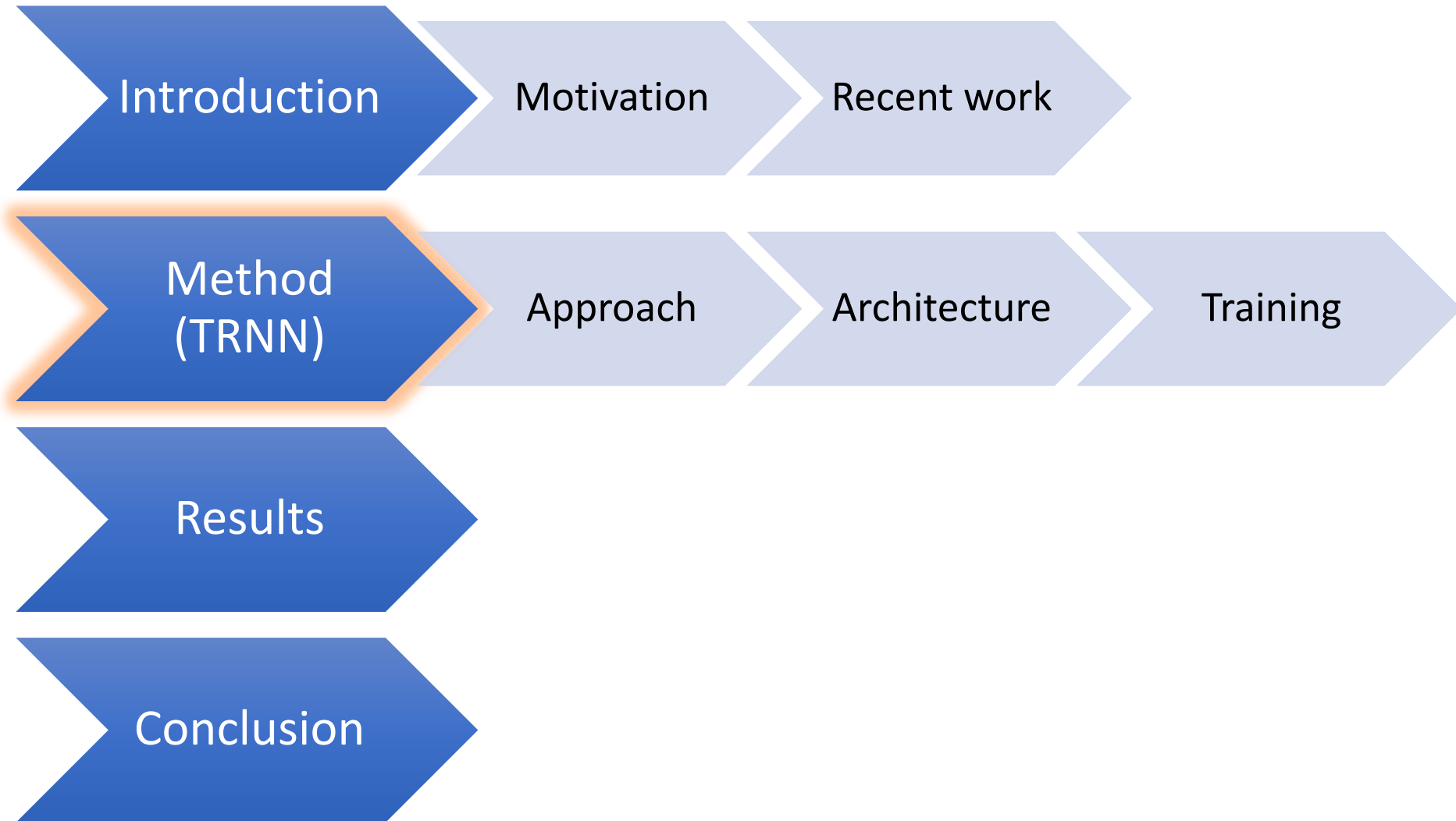
M. Kim, N. Rai, V. Zorraquino, and I. Tagkopoulos, "Multi-omics integration accurately predicts cellular state in unexplored conditions for *Escherichia coli*", *Nature communications* (2016)

Recent work (ECOMICS) – Predictive Model



M. Kim, N. Rai, V. Zorraquino, and I. Tagkopoulos, "Multi-omics integration accurately predicts cellular state in unexplored conditions for *Escherichia coli*", *Nature communications* (2016)

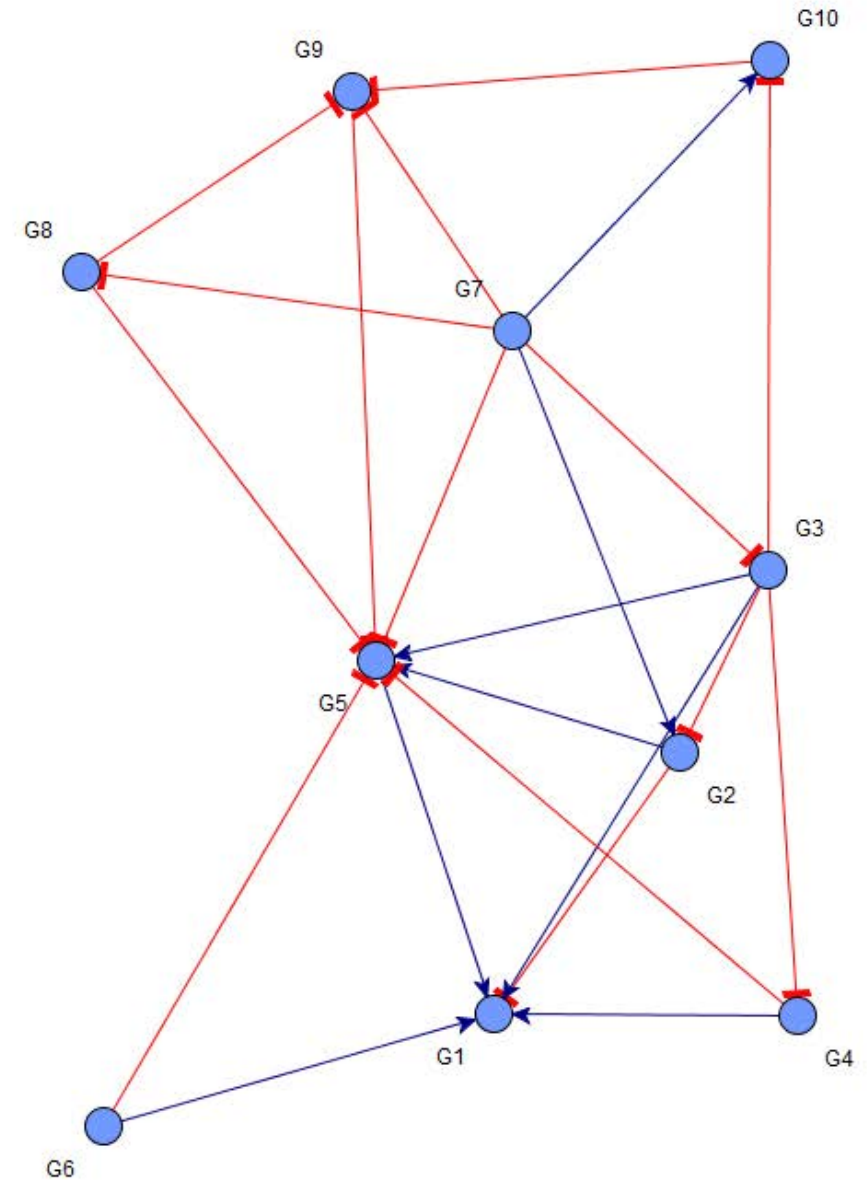
Outline



Approach – Transcription Regulatory Neural Network (TRNN)

How to Integrate:

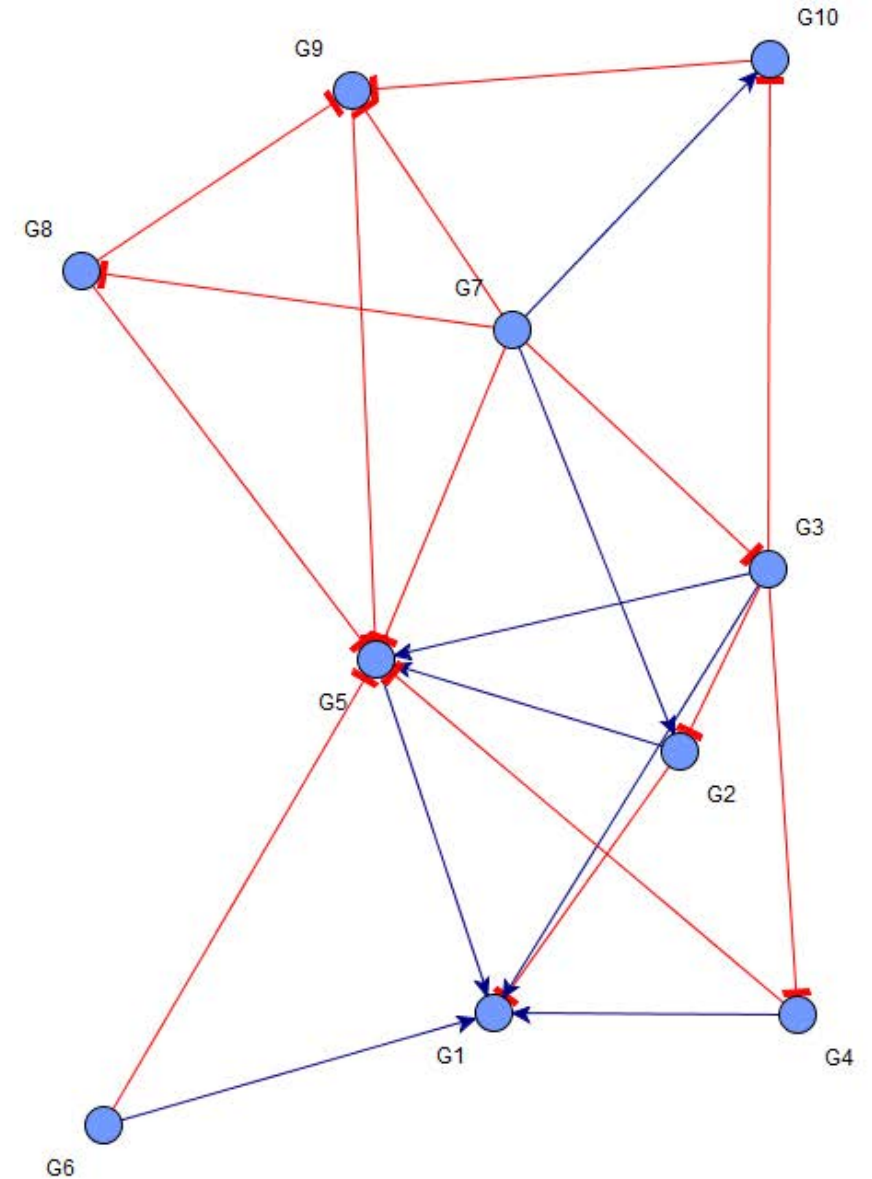
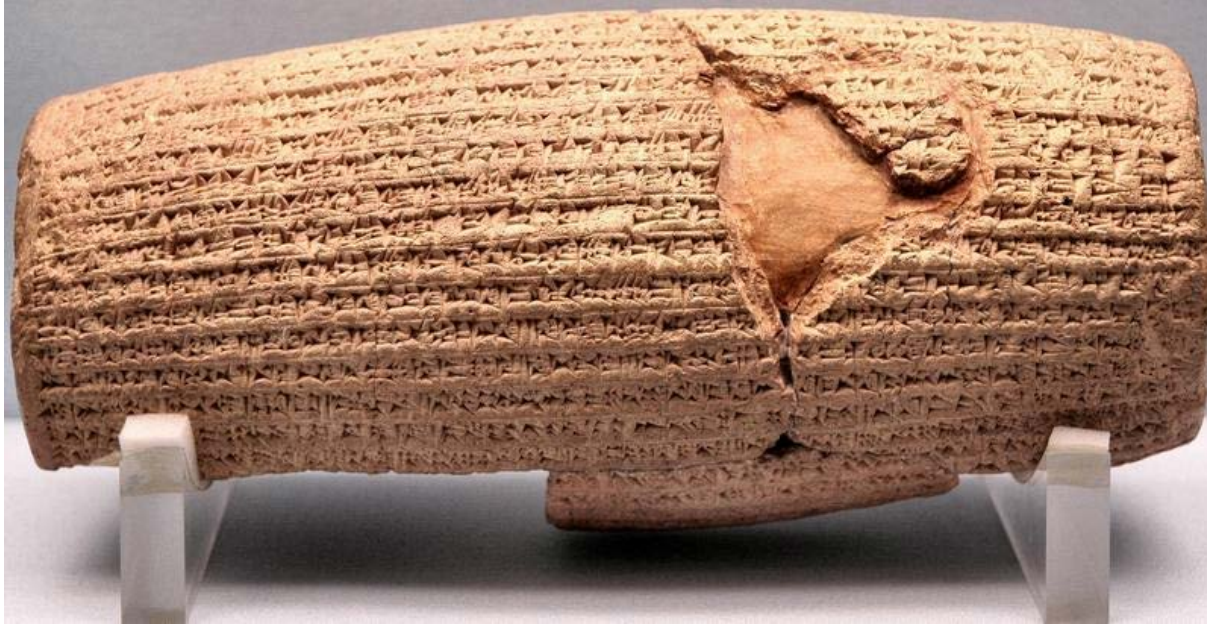
- 1) Transcription regulatory network
- 2) Transcription Thermodynamics
- 3) Transcription profiles



Approach – Transcription Regulatory Neural Network (TRNN)

How to Integrate:

- 1) Transcription regulatory network
- 2) Transcription Thermodynamics
- 3) Transcription profiles



Approach – Transcription Regulatory Neural Network (TRNN)

How to Integrate:

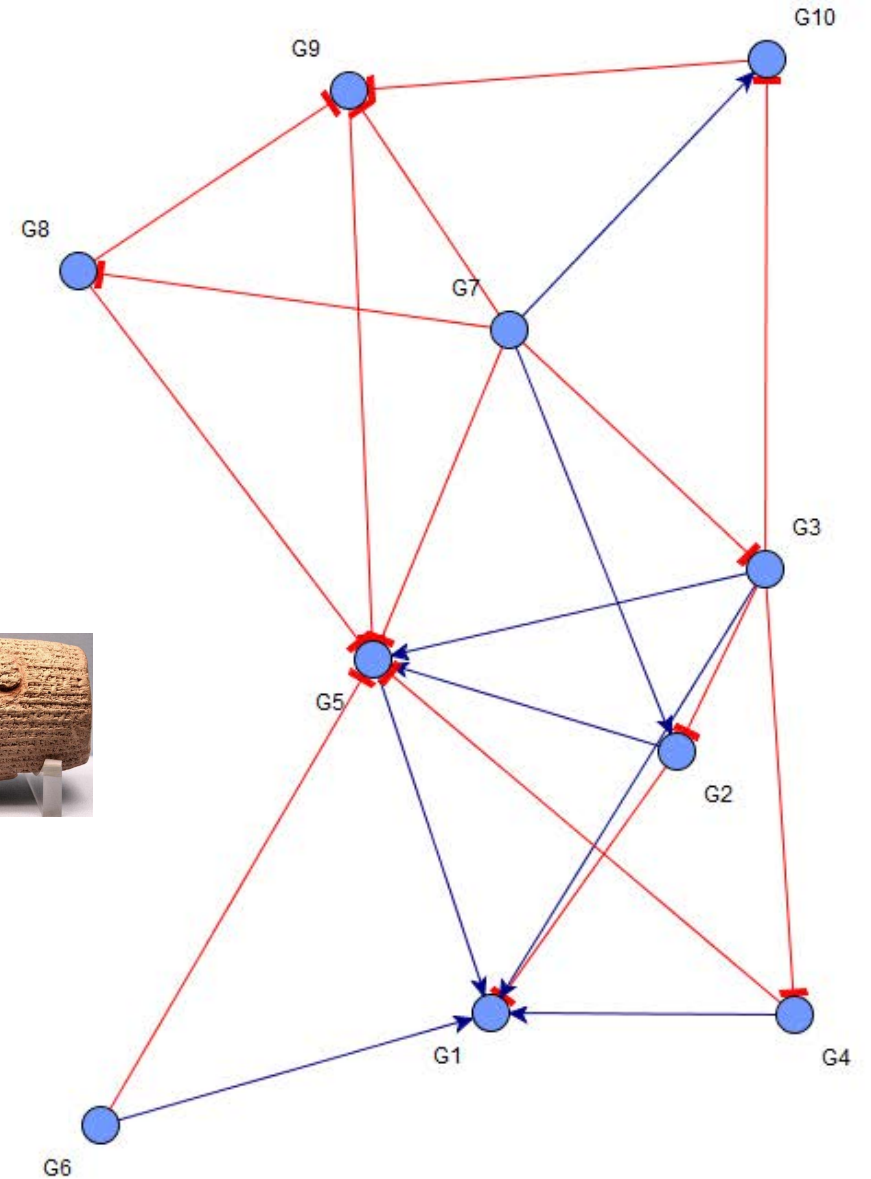
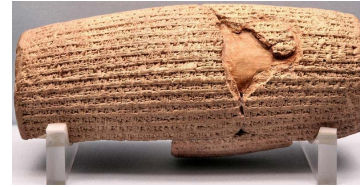
- 1) Transcription regulatory network
- 2) Transcription Thermodynamics
- 3) Transcription profiles

Understanding

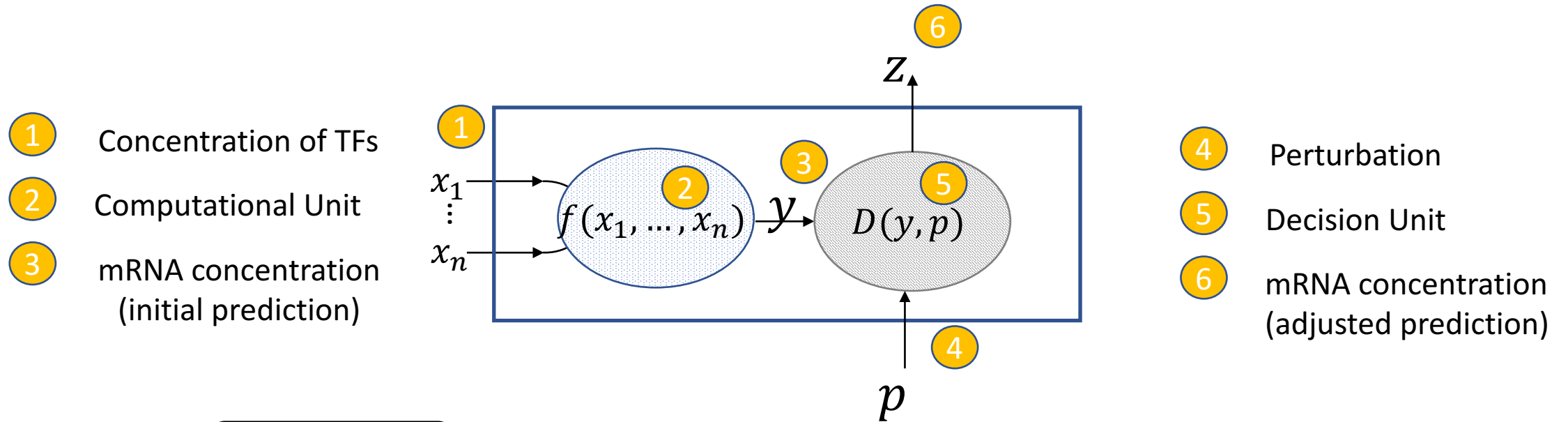
Sentences

Words

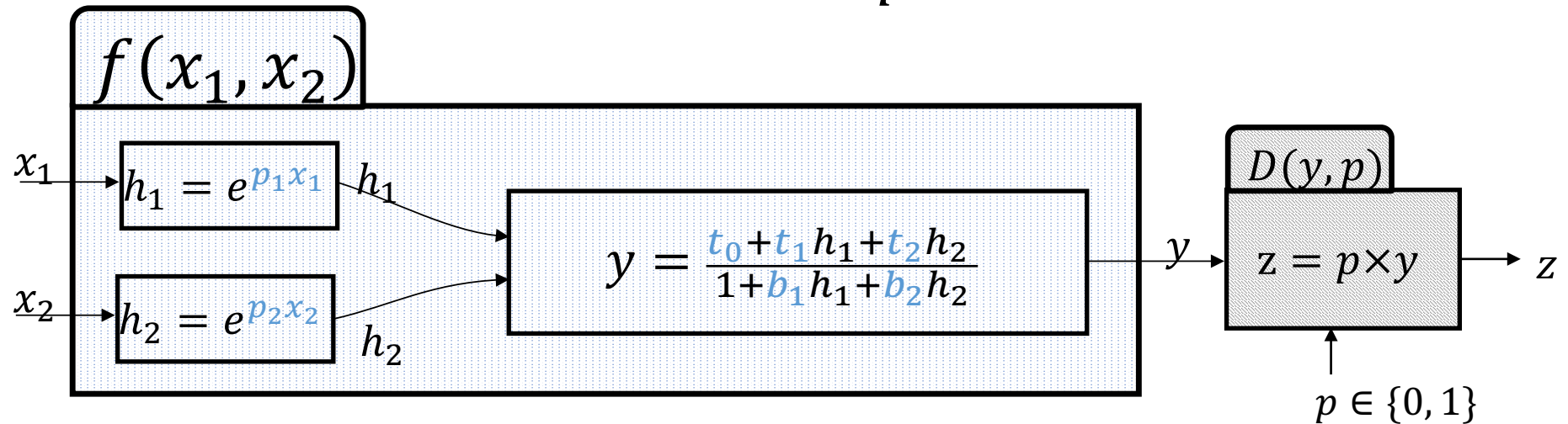
Alphabet



Architecture – TRNN Activation Function (*Alphabet*)

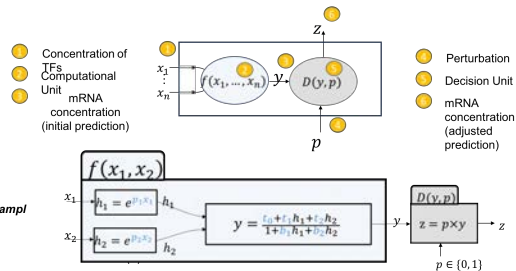
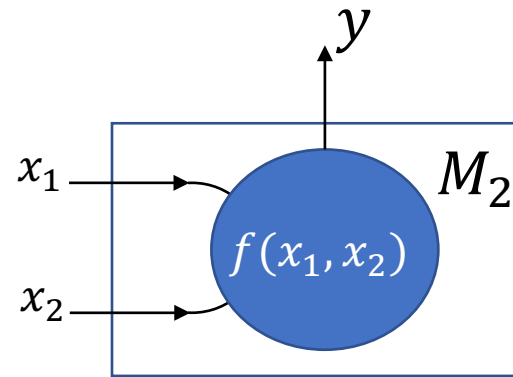
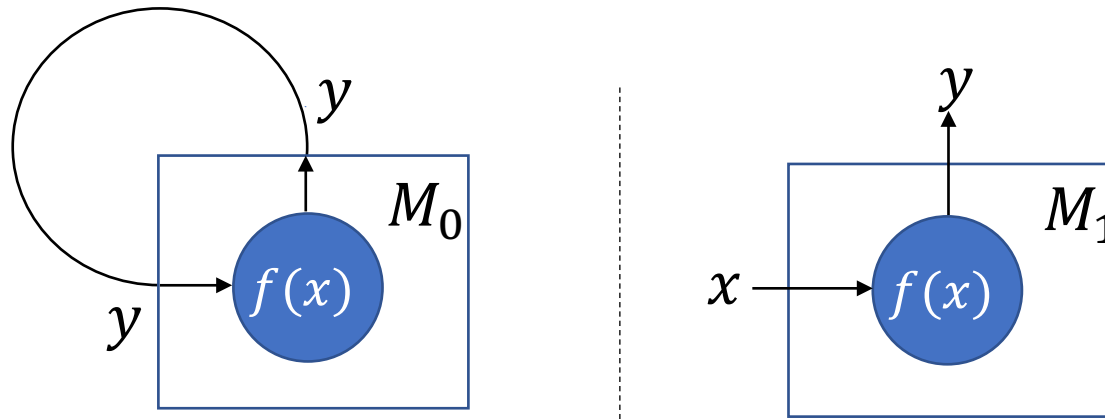


Example:



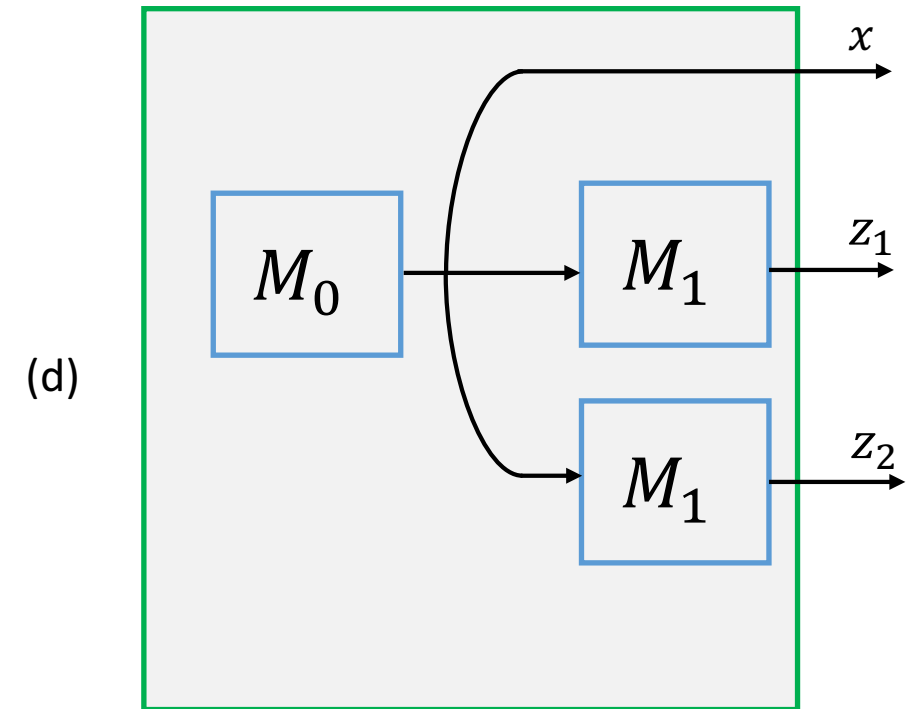
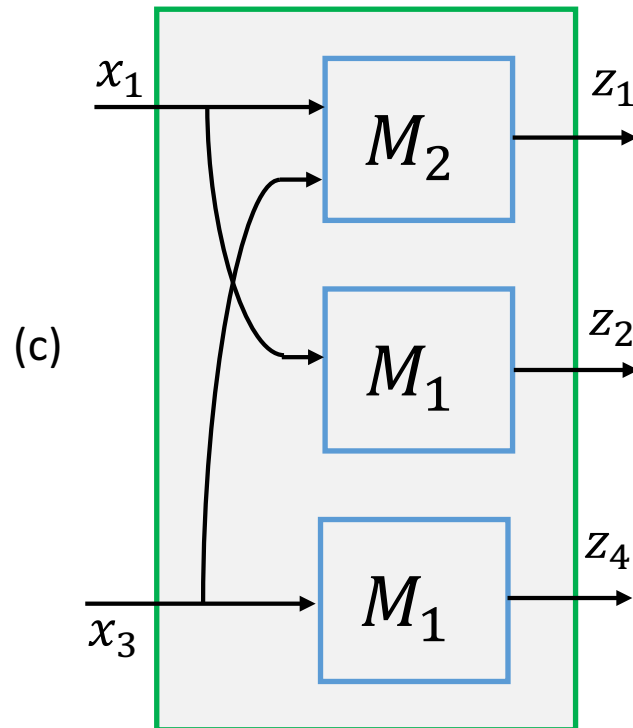
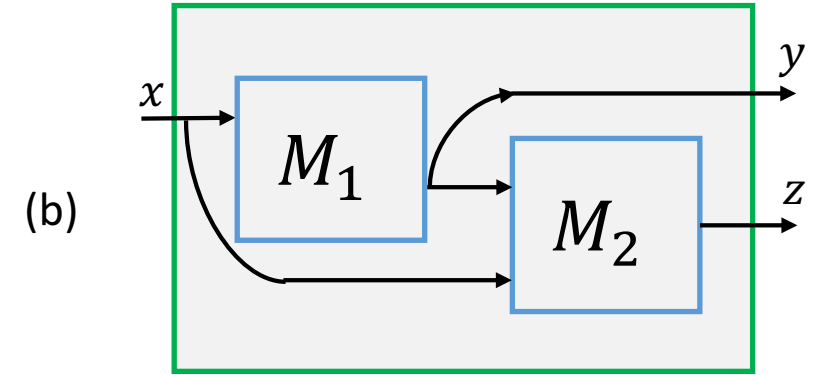
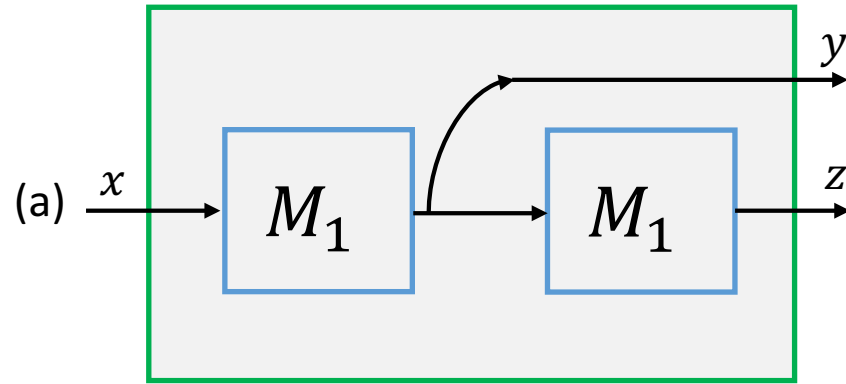
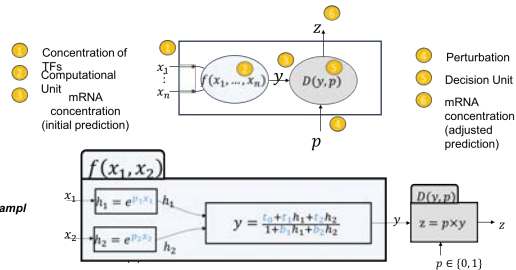
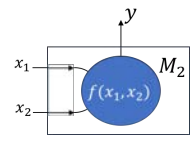
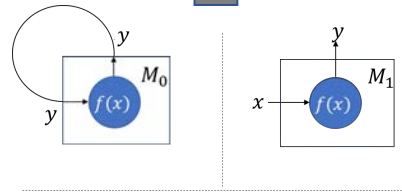
Architecture – TRNN Gene Module (*Word*)

Words



Architecture – TRNN Gene Circuit (*Sentence*)

Sentences



Training Strategy

Issue: backpropagation is very sensitive to initial values

Strategy:

- **Step1:** train individual modules separately through a customized LP based method
- **Step2:** train full network together

Use full-batch, conjugate gradient

Training Individual Modules

Algorithm 1: Fit f_θ , given C

Input : $C = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{y}\}$

Output: $\theta^* = \{t_0^*, t_1^*, t_2^*, b_1^*, b_2^*, p_1^*, p_2^*\}$

```
1 initialize  $p^* = [p_1^*, p_2^*]$ ;
2 while  $loss_{\theta^*}(C)$  is not converged do
3    $w^* \leftarrow \underset{w}{\text{ArgMin}} \|A \cdot w - \mathbf{y}\|_1^1, \text{s.t. } b_i \geq 0$ ;
4    $\theta^* \leftarrow \begin{bmatrix} w^* & p^* \end{bmatrix}$ ;
5   for  $j \leftarrow 1 : 2$  do
6      $\Delta_{p_j} \leftarrow \frac{\partial}{\partial p_j} loss_{\theta^*}(C)$ ;
7      $p_j^* \leftarrow p_j^* - \alpha \cdot \Delta_{p_j}$ ;
8   end
9    $\theta^* \leftarrow \begin{bmatrix} w^* & p^* \end{bmatrix}$ ;
10 end
11 return  $\theta^*$ 
```

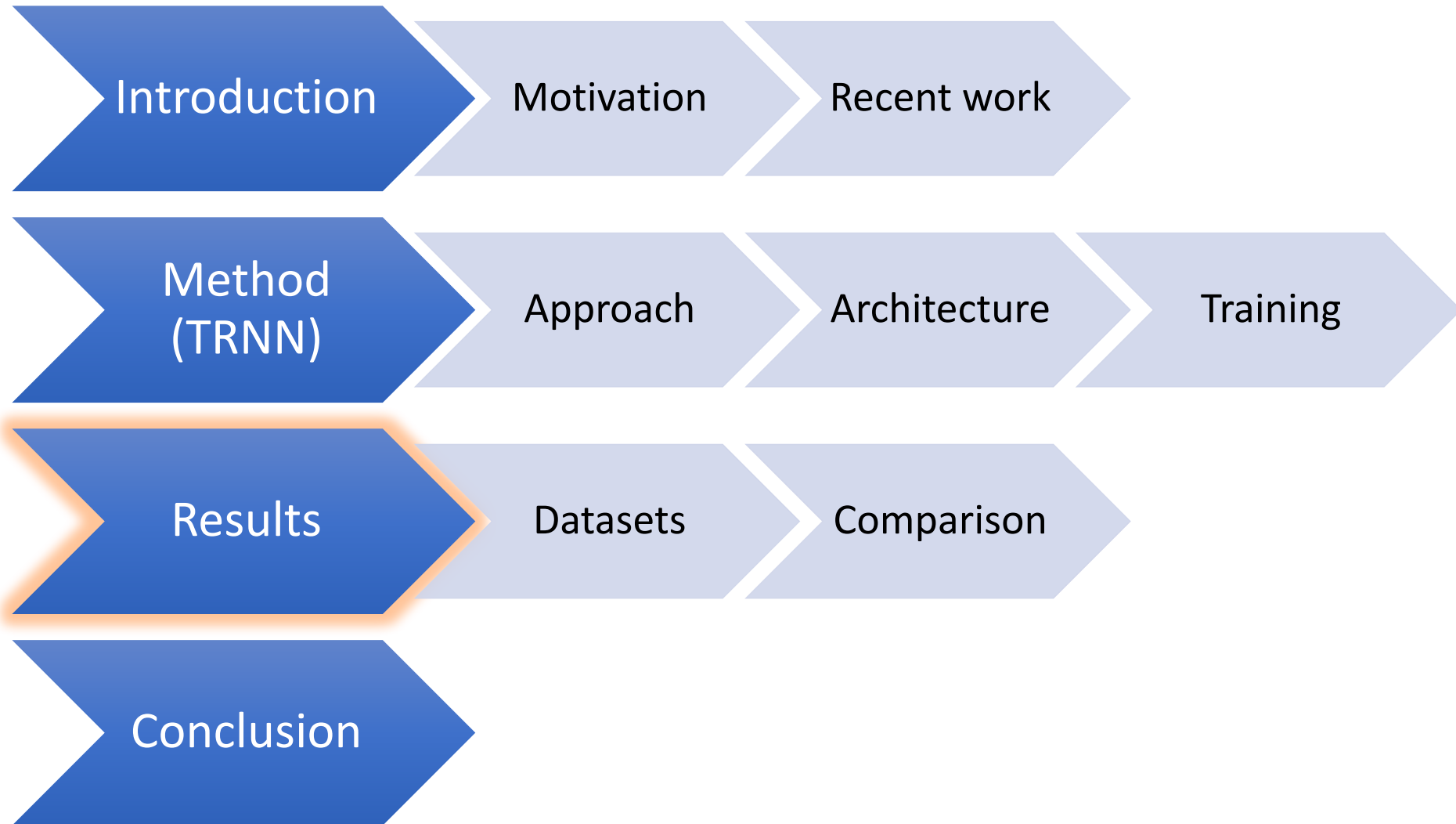
$$loss_\theta(C) = \sum_{i=1}^m [f_\theta(x_1^i, x_2^i) - y^i]^2$$

$$h_1^i = e^{p_1 x_1^i}, h_2^i = e^{p_2 x_2^i}$$

$$A = \begin{bmatrix} \mathbf{1} & h_1 & (-h_1 \odot \mathbf{y}) & h_2 & (-h_2 \odot \mathbf{y}) \end{bmatrix}$$

\odot : entry-wise vector multiplication,

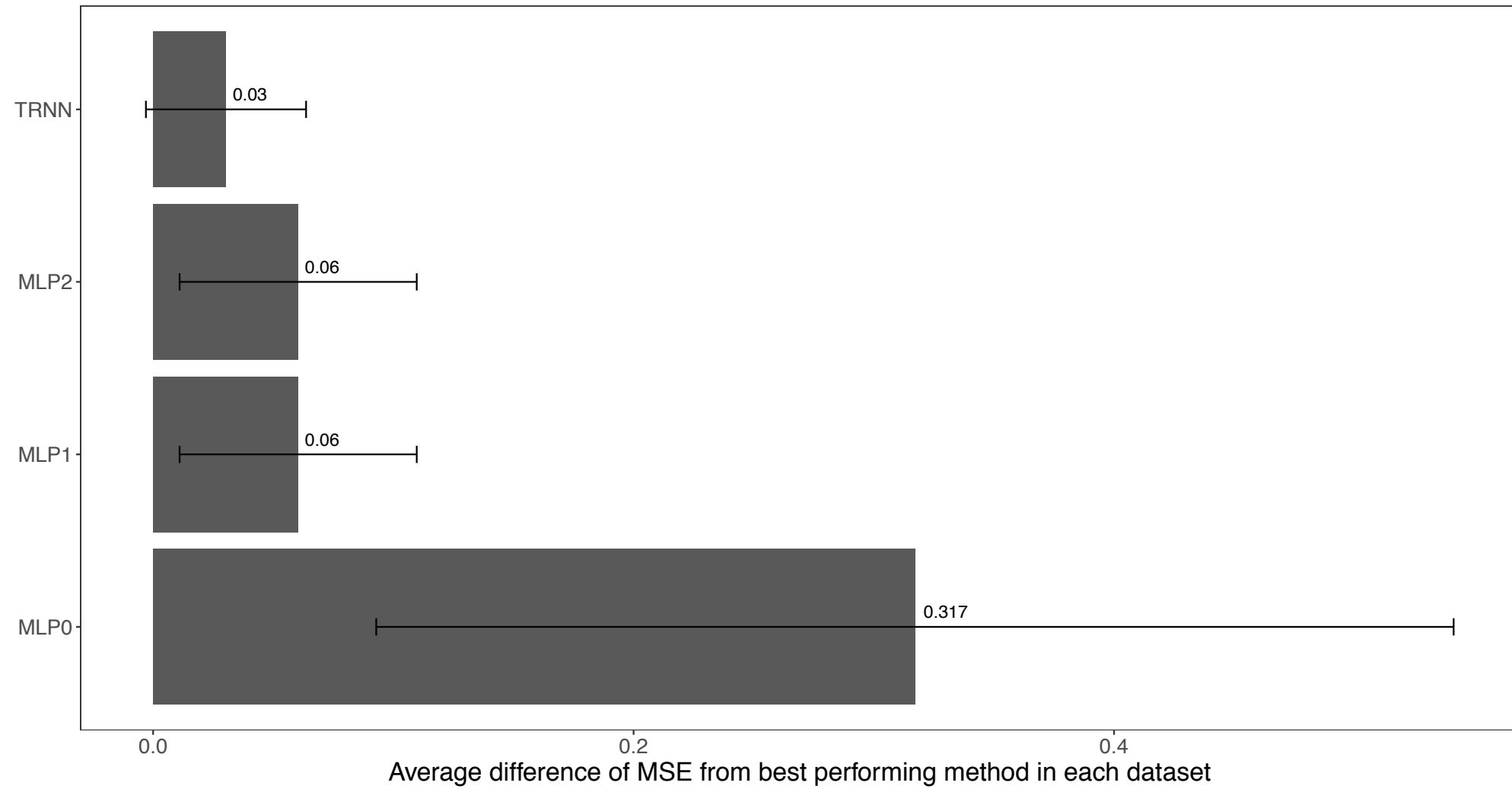
Outline



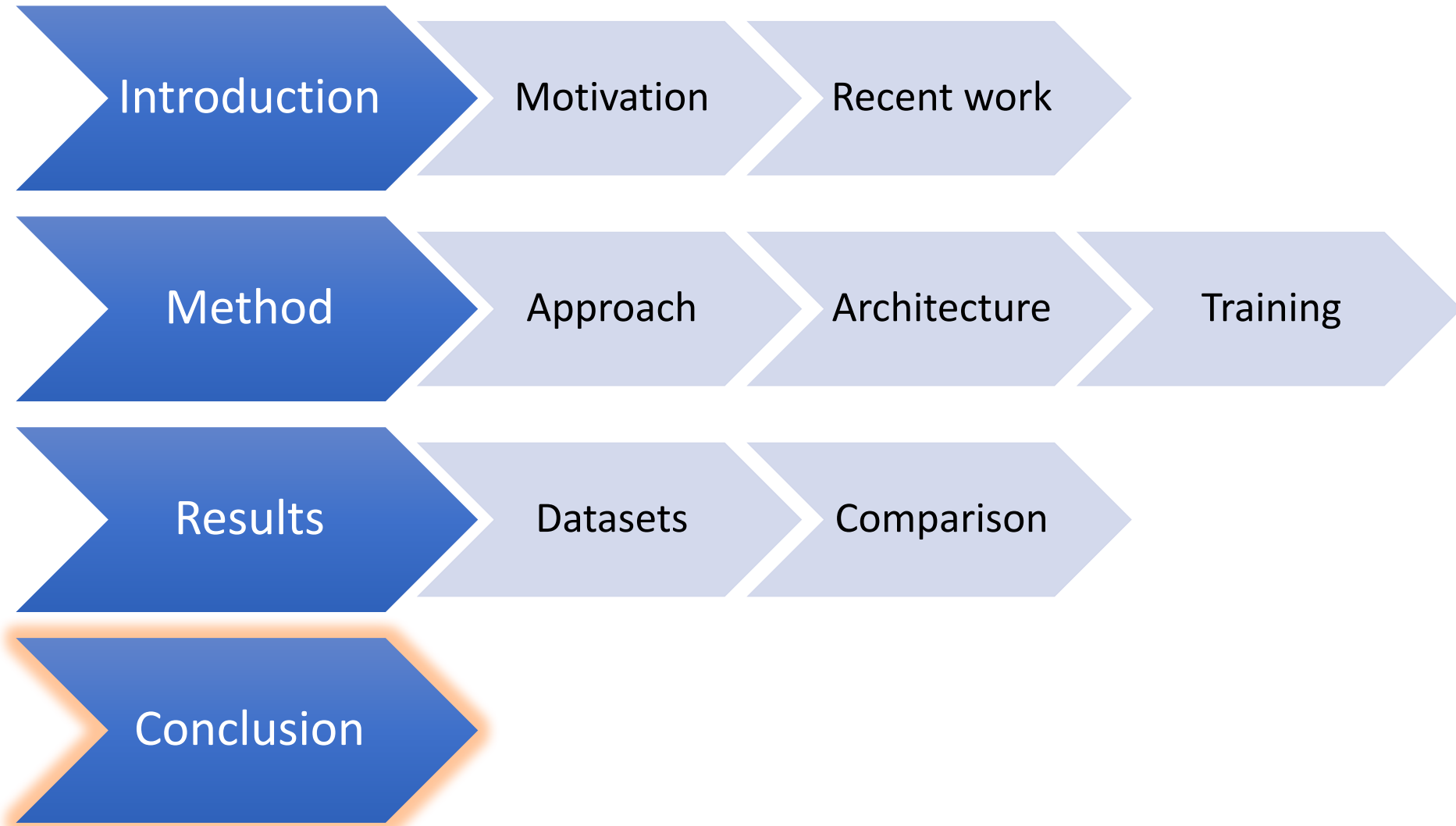
Dataset

- 46 different datasets:
 - Multiple candidate networks extracted from DREAM4 challenge
 - Number of genes for each network range from 3 to 9.
 - Synthetic data generated thermodynamic simulation
 - Gene Knockouts
 - Parameter perturbations (e.g. transcription rate, binding affinity of TFs, etc.)
 - Different noise levels
- Validation
 - Mean Squared Error
 - 5-Fold cross validation
 - Comparison with common ANN architectures

Results – Overall MSE Comparison



Outline



Conclusion

- TRNN, is a deep learning based framework for prediction of steady state mRNA concentration levels under perturbation.
- TRNN, requires specialized training due to it's architecture.
- Evaluation of TRNN on real data is key next step.
- For scaling up TRNN to genome scale modelling, we will make use of highly parallel computation provided by the cluster.