# INFERRING TRANSCRIPTION LOGIC FROM MULTIPLE DYNAMIC EXPERIMENTS

*Giorgos Minas,*
*School of Mathematics & Statistics,*
*University of St-Andrews, Scotland*
*@Barcelona, 09.07.2018*

University of St Andrews | FOUNDED 1413

*Scotland's first university*

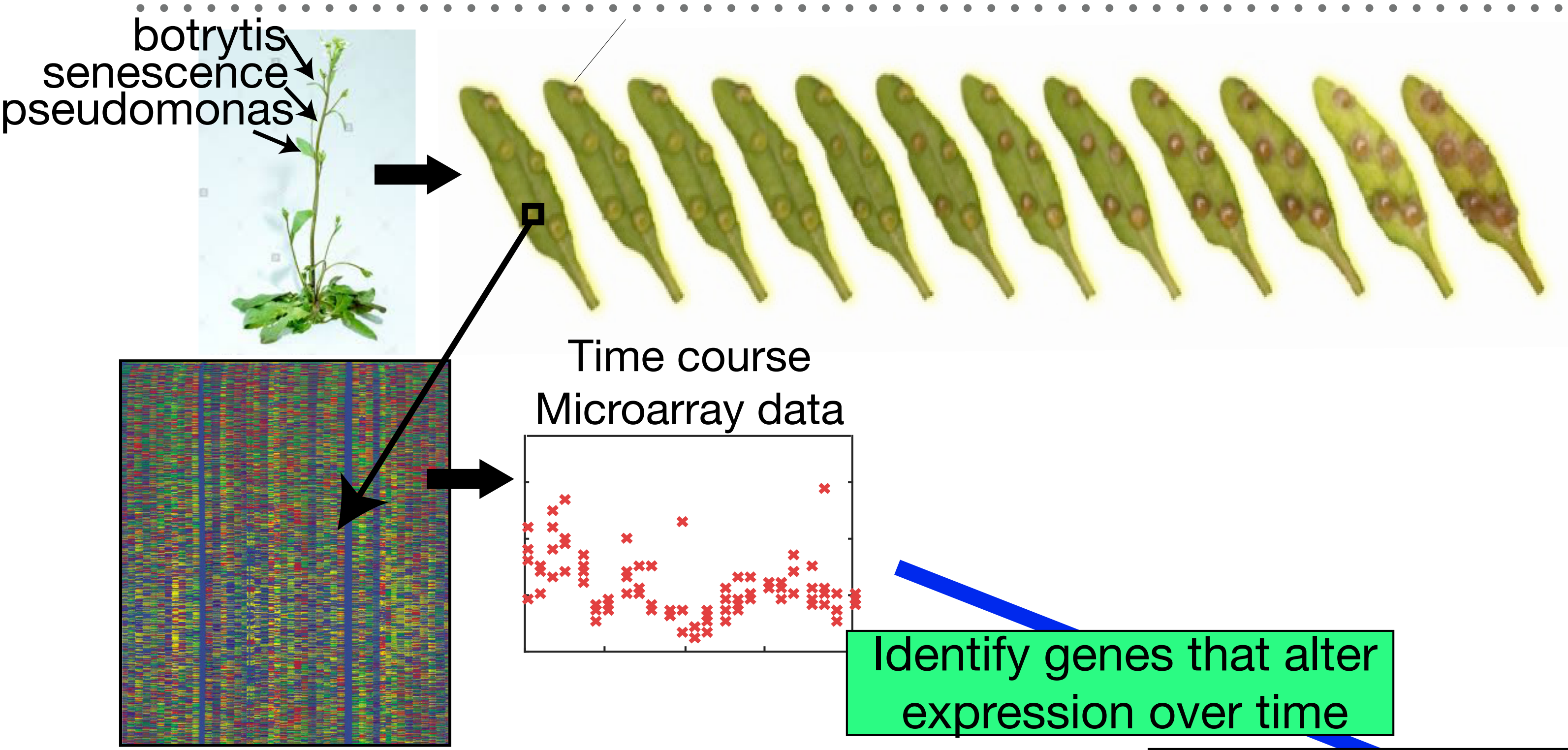
WARWICK
THE UNIVERSITY OF WARWICK

# PLAN OF THE TALK

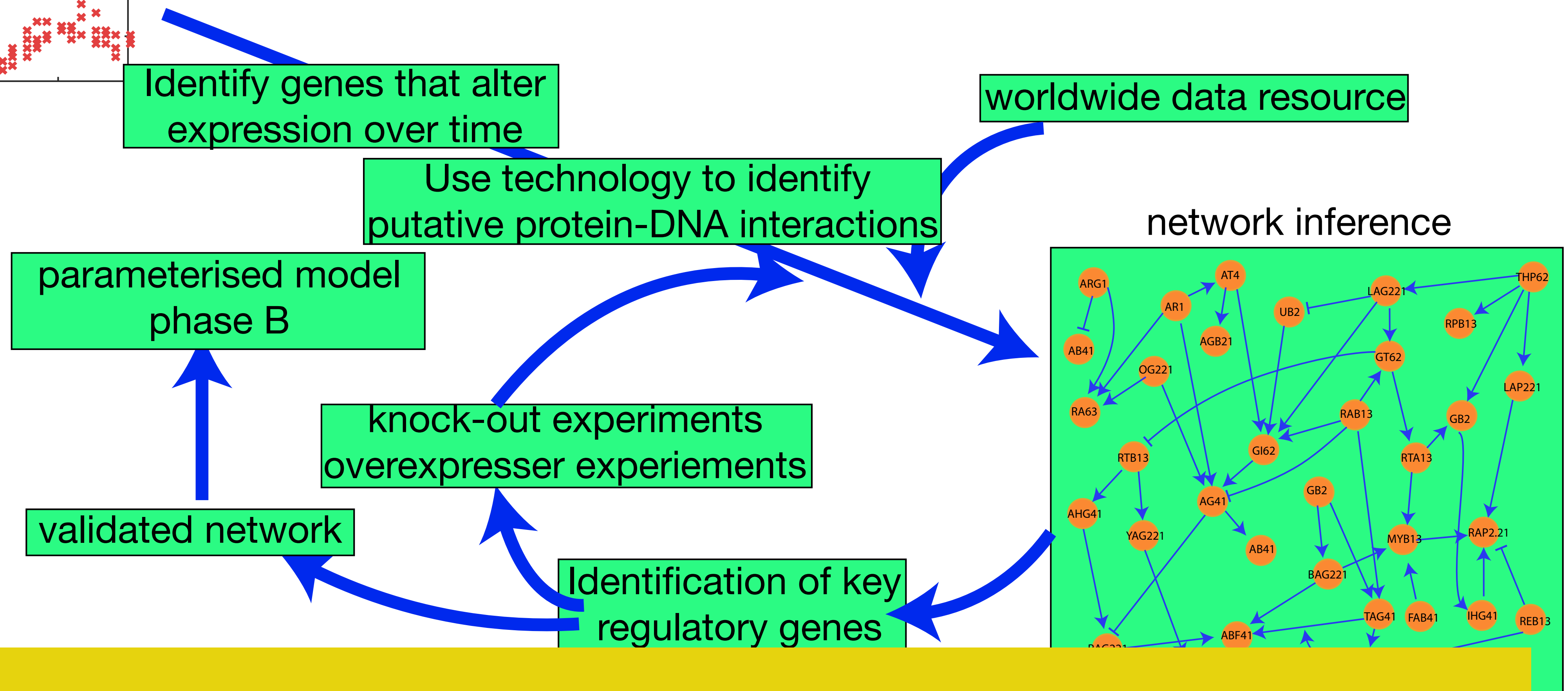
- ***P**redicting **R**esponses under **E**nvironmental **St**resses in **A**. thaliana*
- ***T**ranscriptional **R**egulation*
- ***TR** modelling*
- ***TR S**witch modelling*
- *Statistical inference for **TRS***


University of St Andrews FOUNDED 1413 WARWICK THE UNIVERSITY OF WARWICK

# Predicting Responses in Environmental Stresses in A. thaliana

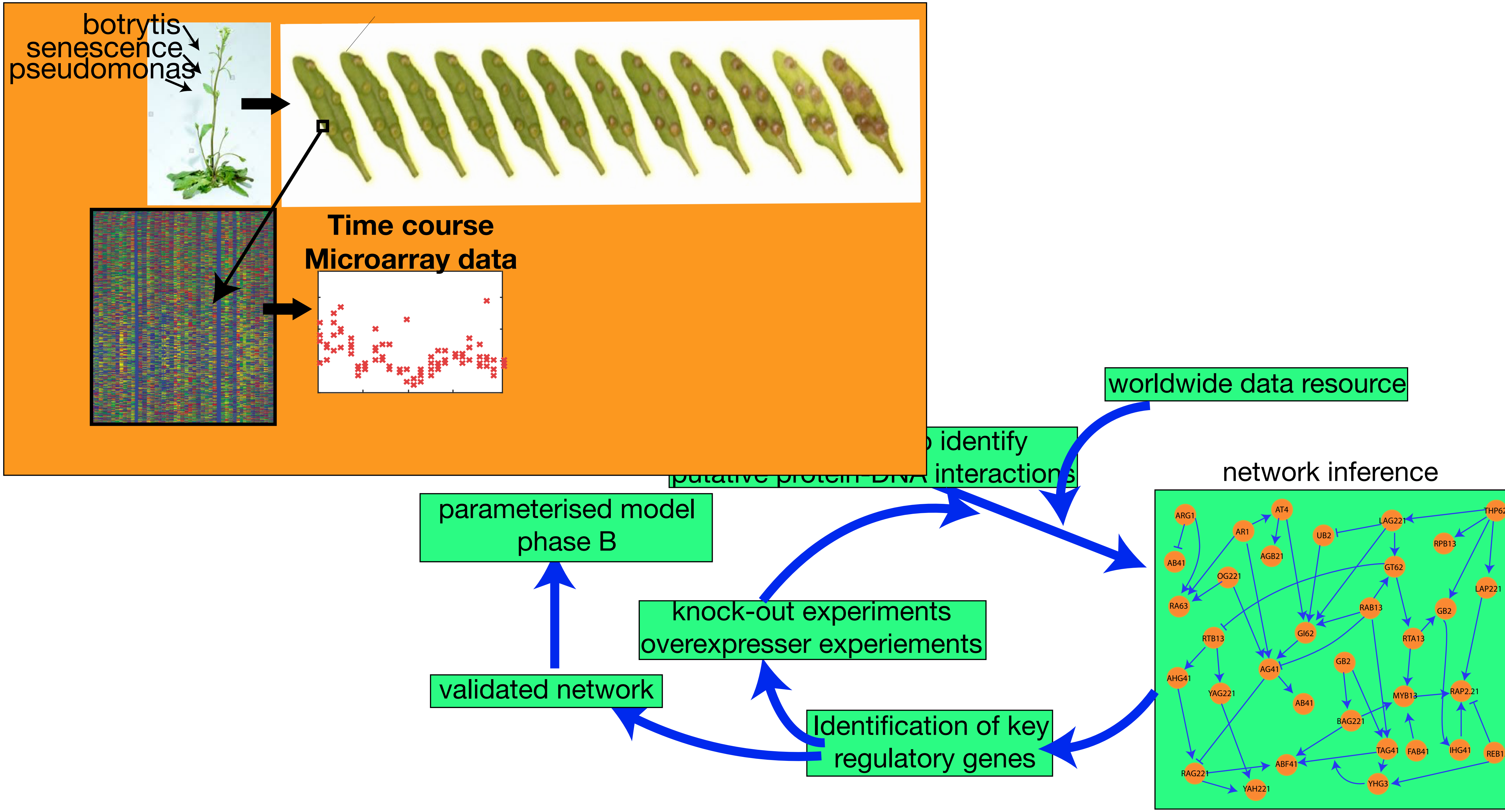



**General question:**
How plants respond to climate change?



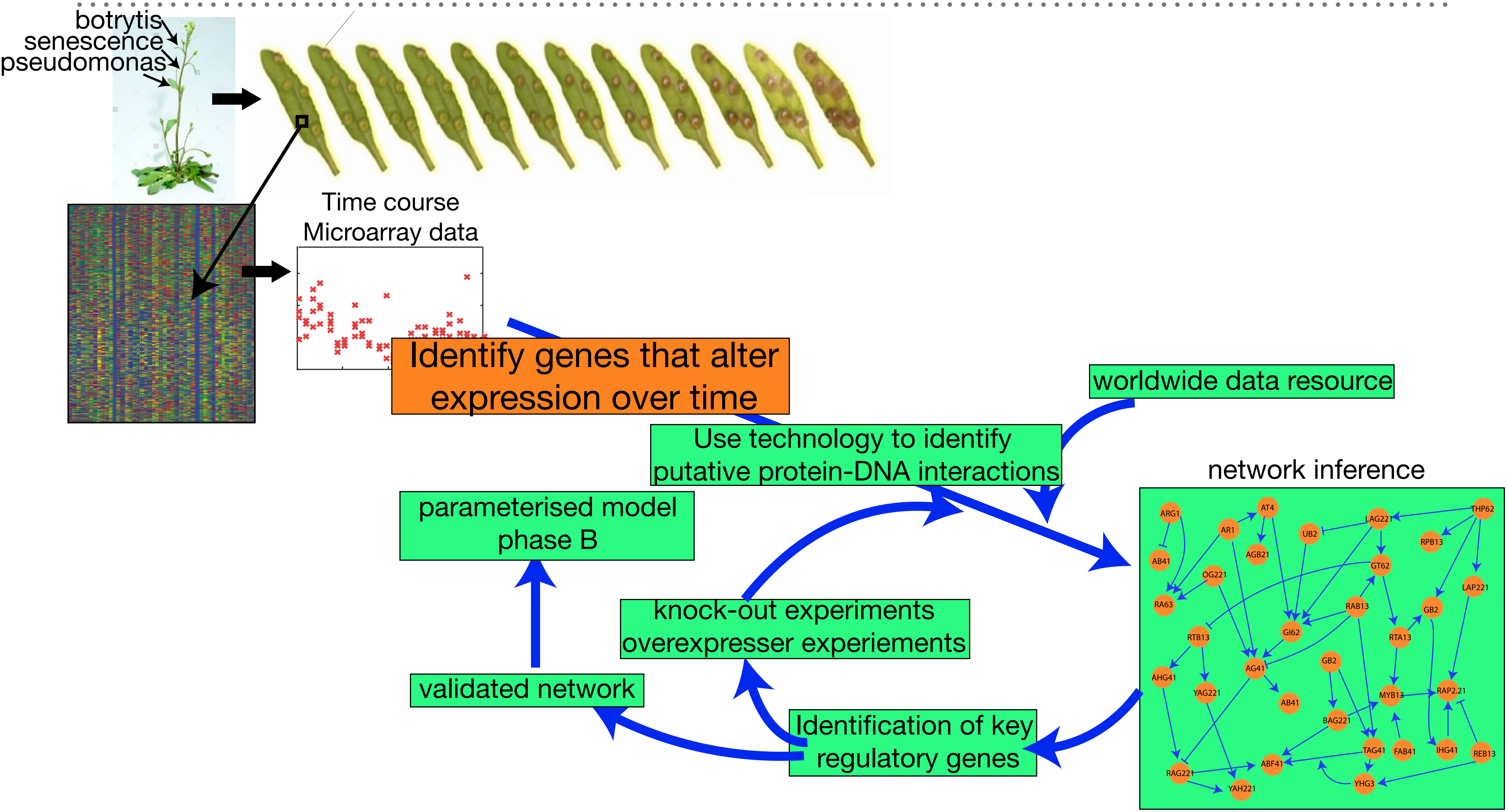

**Target question:**
What gene regulation networks are active under environmental stresses? Which are the key genes for the response to stresses?

# PREDICTING RESPONSES IN ENVIRONMENTAL STRESSES IN A. THALIANA

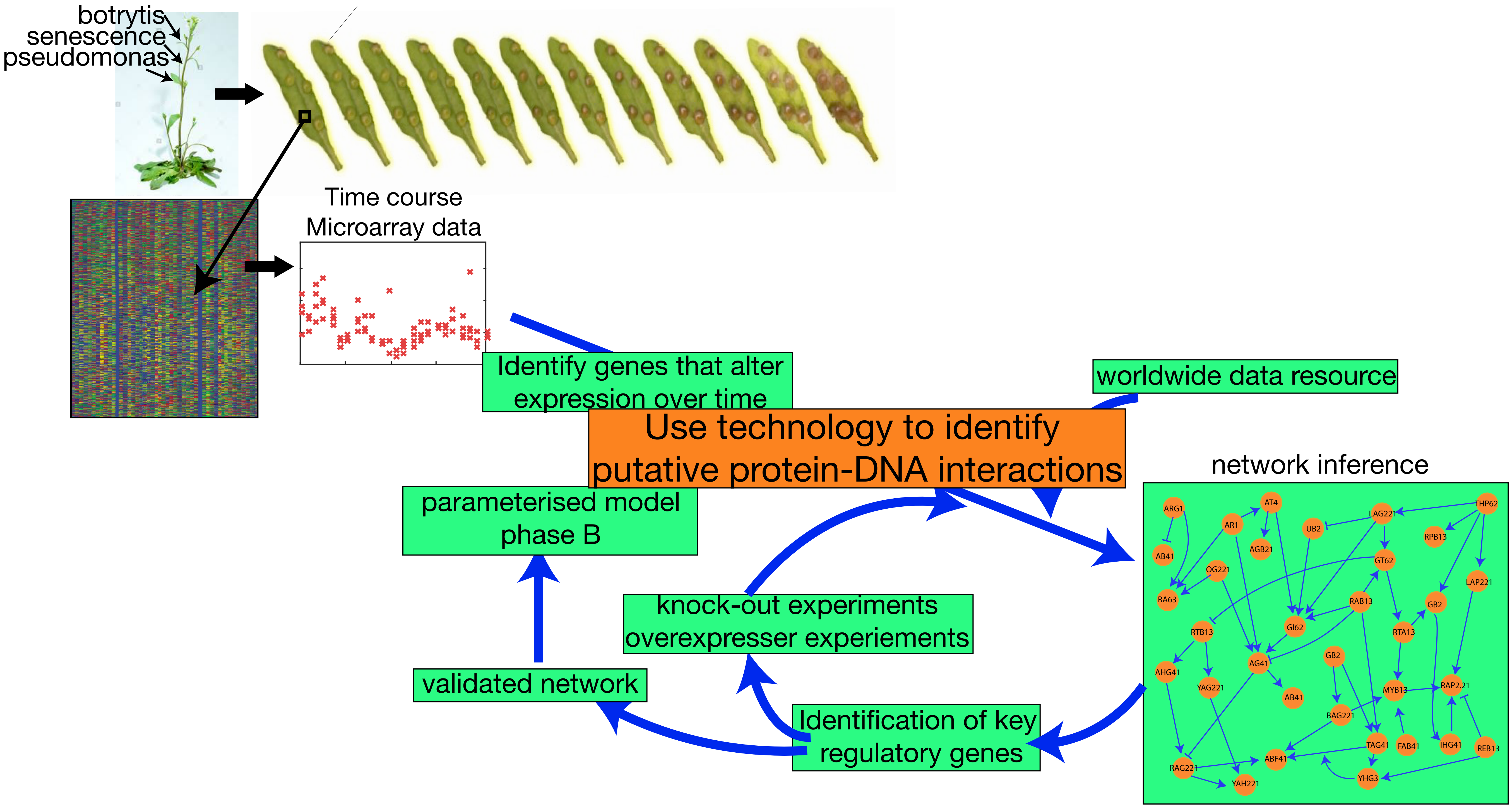



- Observe plants under various environmental stress
- Measure gene expression (microarray) in multiple time-points

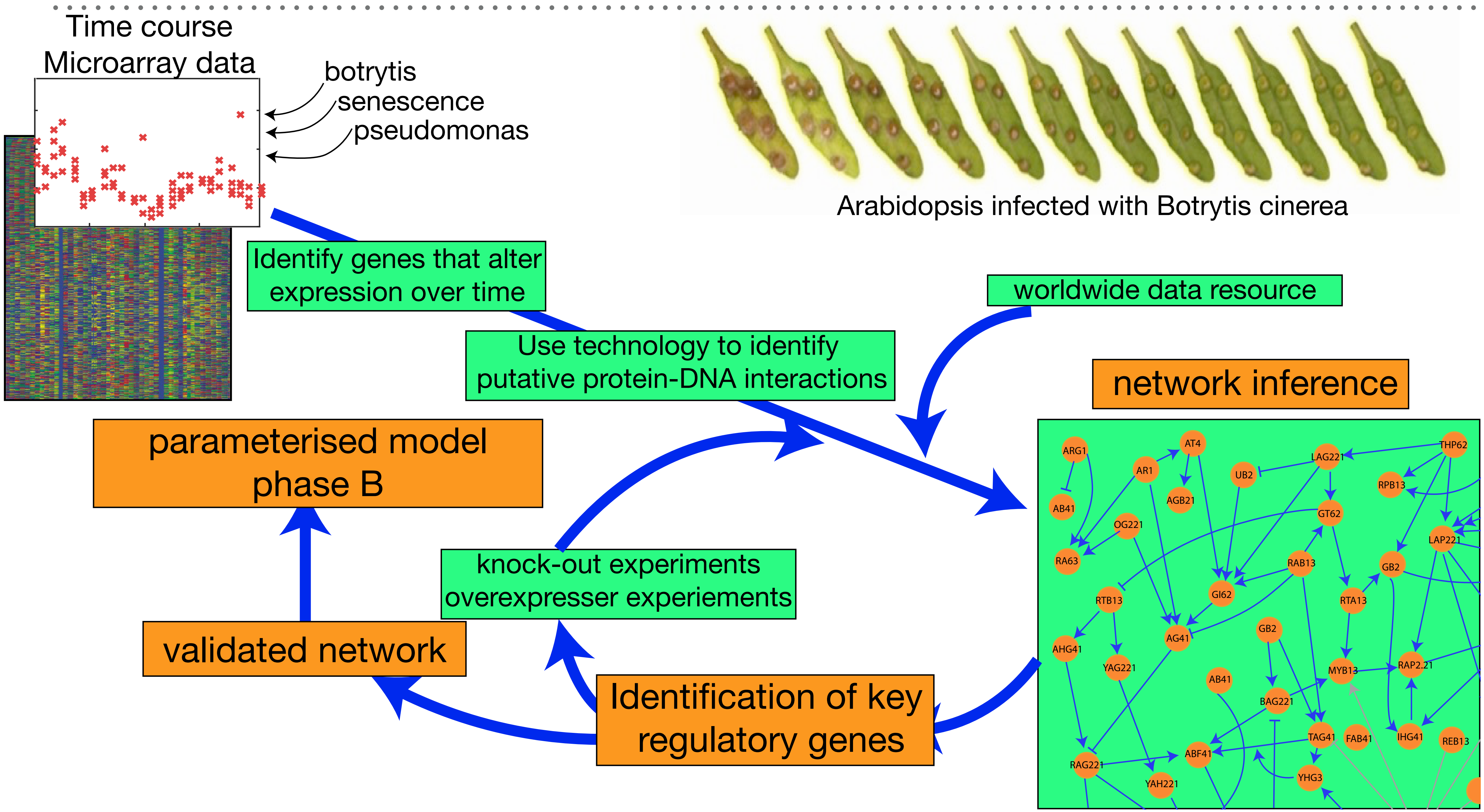
# PREDICTING RESPONSES IN ENVIRONMENTAL STRESSES IN A. THALIANA



# PREDICTING RESPONSES IN ENVIRONMENTAL STRESSES IN A. THALIANA



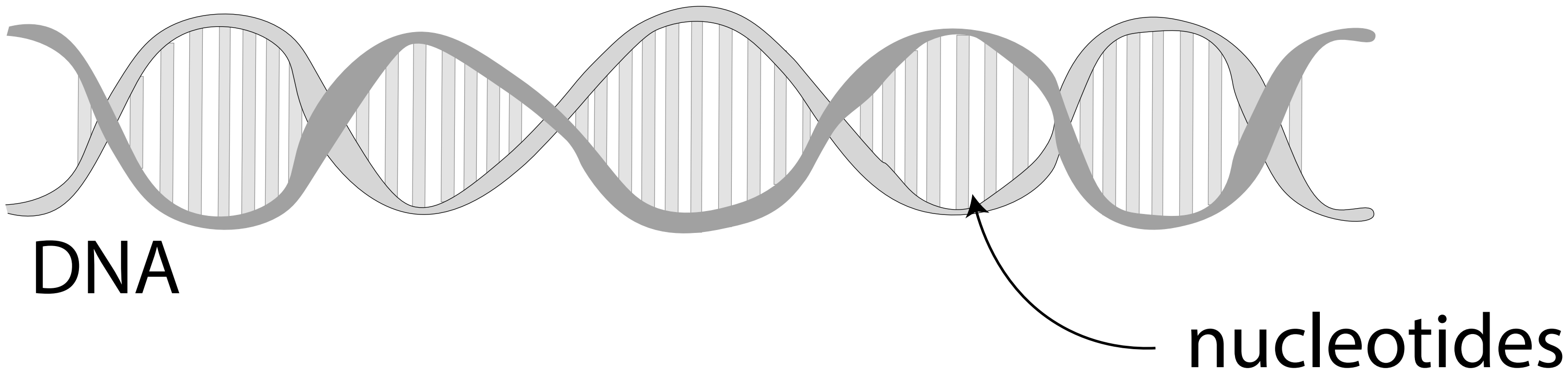
# PREDICTING RESPONSES IN ENVIRONMENTAL STRESSES IN A. THALIANA



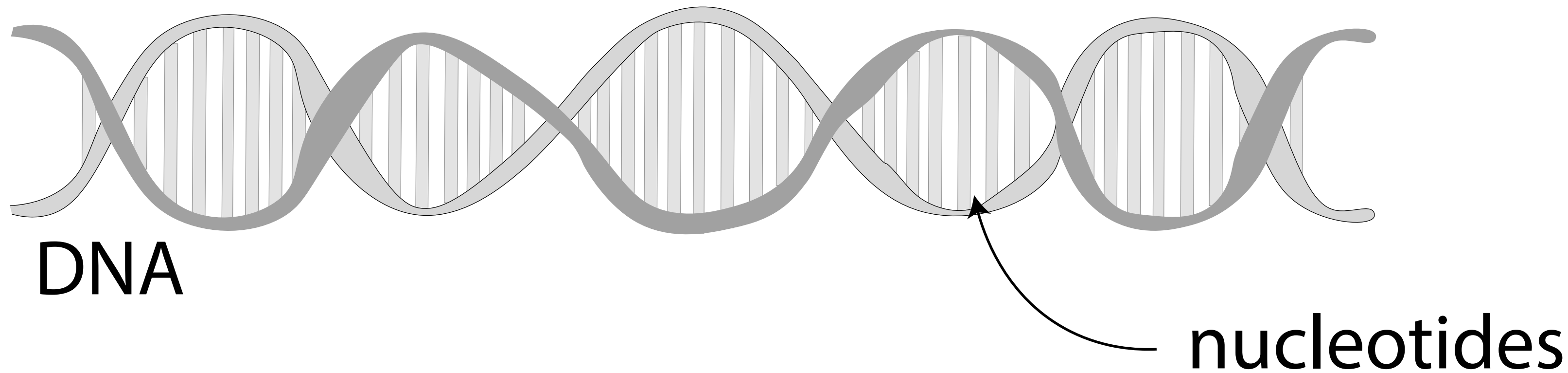

**My part:**

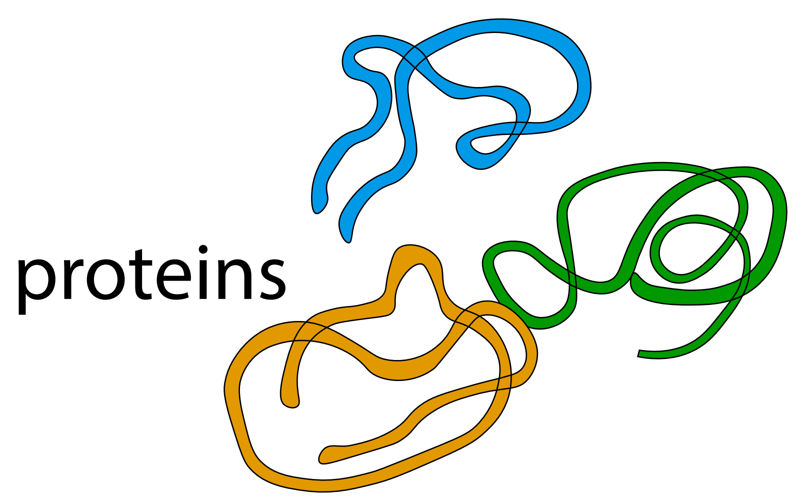
Develop a model of transcription regulation between the genes that are identified to be active and important under the stresses.

# FROM GENES TO PROTEINS

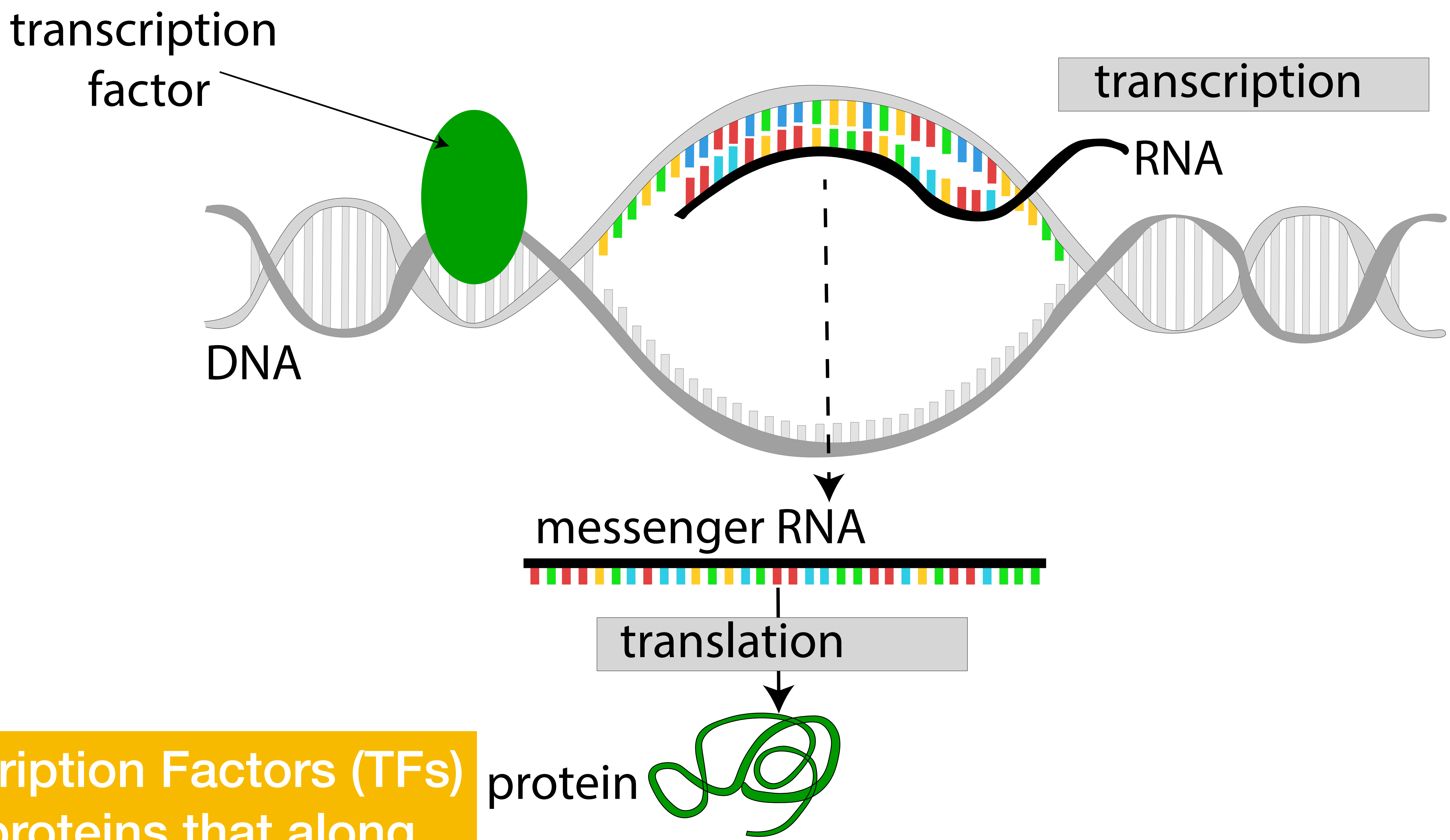


# FROM GENES TO PROTEINS



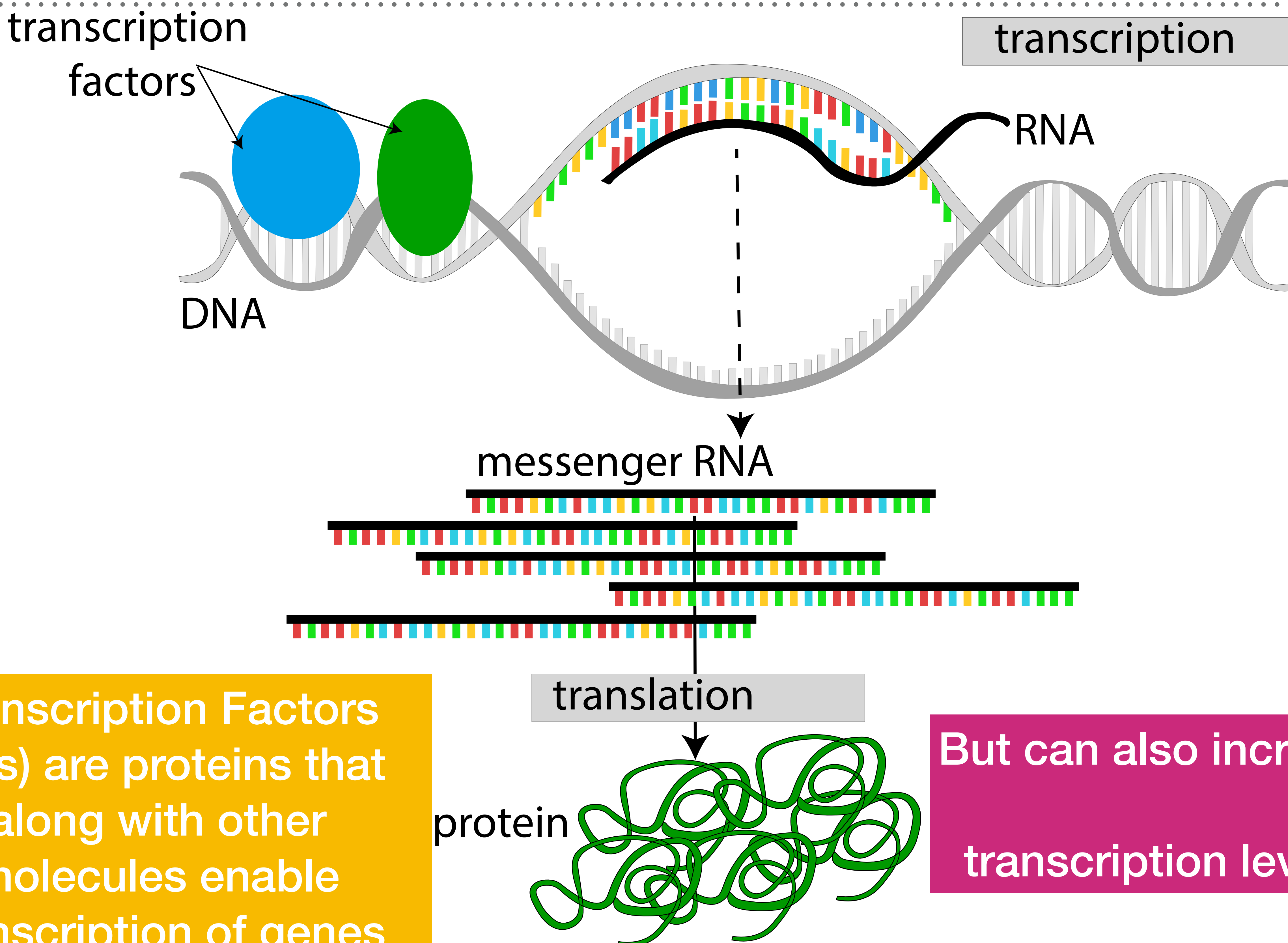



# FROM GENES TO PROTEINS




Transcription Factors (TFs) are proteins that along with other molecules enable transcription of genes

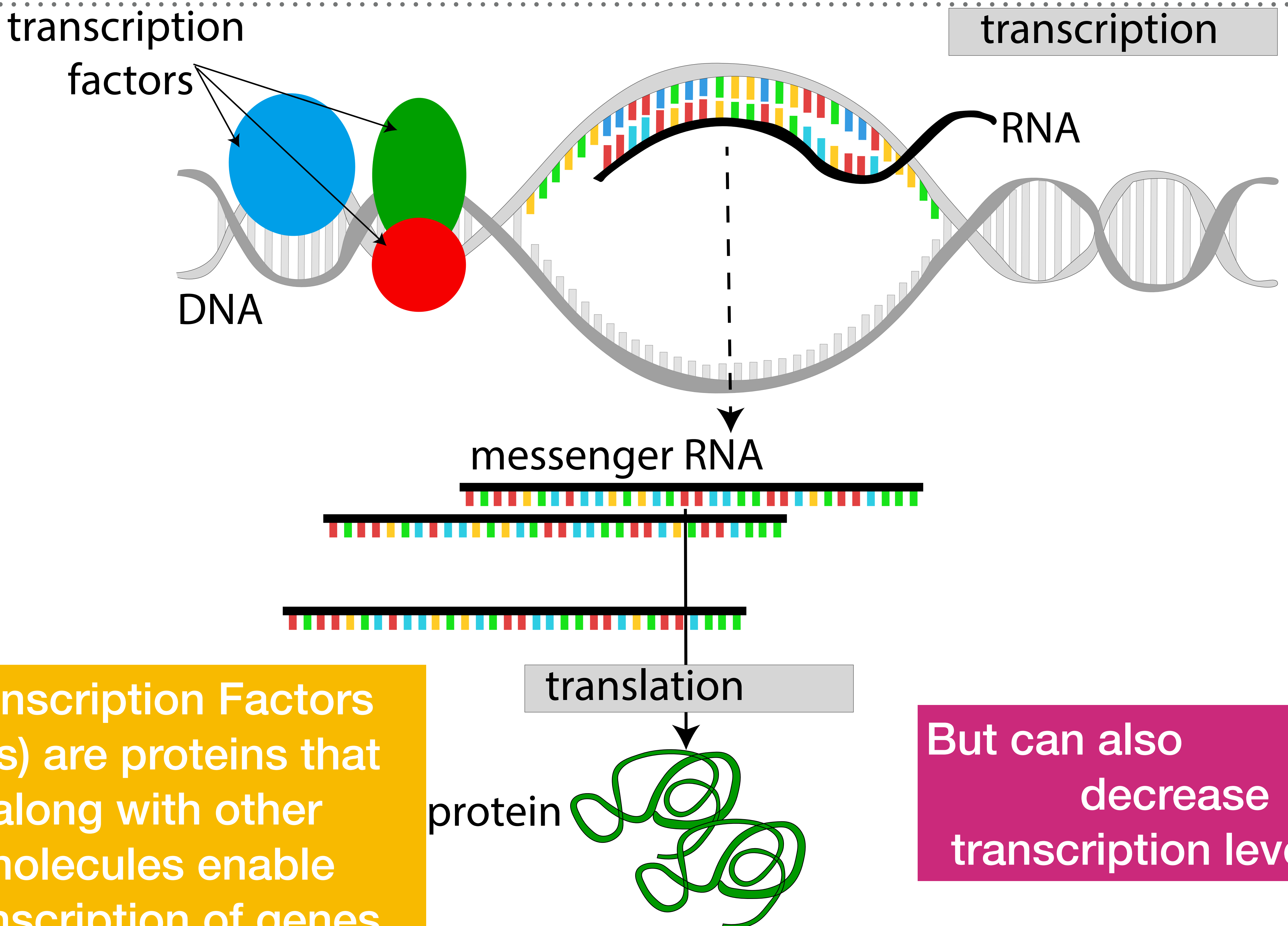
# TRANSCRIPTIONAL REGULATION-ACTIVATION




Transcription Factors (TFs) are proteins that along with other molecules enable transcription of genes

But can also increase transcription levels

# TRANSCRIPTIONAL REGULATION-REPRESSION

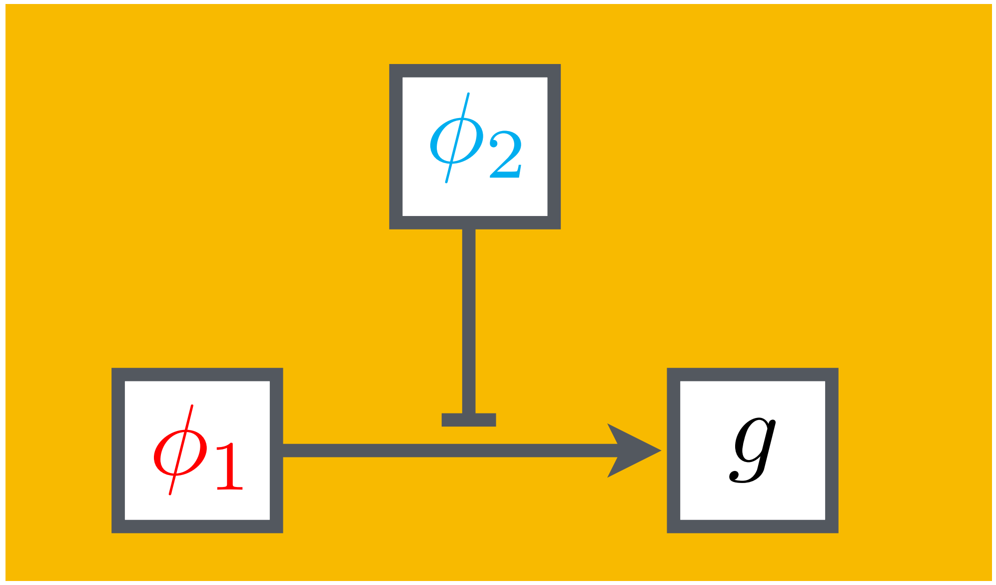



Transcription Factors (TFs) are proteins that along with other molecules enable transcription of genes

But can also decrease transcription levels

# TRANSCRIPTION REGULATION MODELLING

- *A large number of different models for dynamic & static observations, and for binary, discrete & continuous state spaces.*




Graph of a simple example of transcriptional regulation

# TRANSCRIPTION REGULATION MODELLING

- *A large number of different models for* ***dynamic*** *& static observations, and for binary, discrete &* ***continuous*** *state spaces.*
- *For dynamic (time-series) observations with continuous* $\mathbb{R}^+$ *state space, the mRNA expression is often modelled as the ODE*

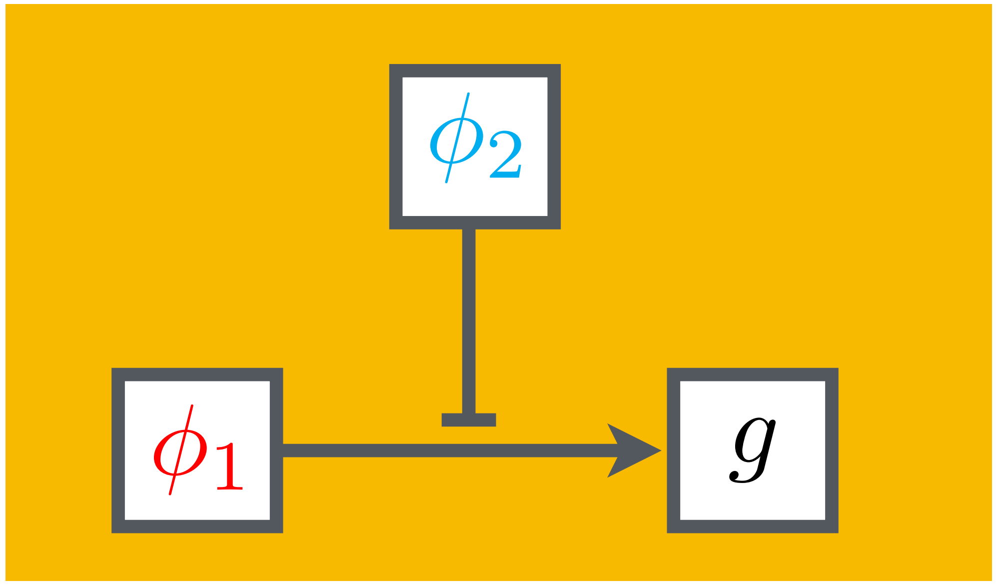
$$\frac{dM}{dt} = \tau(t) - \delta M(t)$$

$\tau(t)$: transcription (birth) rate

$\delta$: degradation (decay) rate

*which can be also written in* ***regression form,***

$$M(t) = M(0) + f(t, \tau, \delta)$$




Graph of a simple example of transcriptional regulation

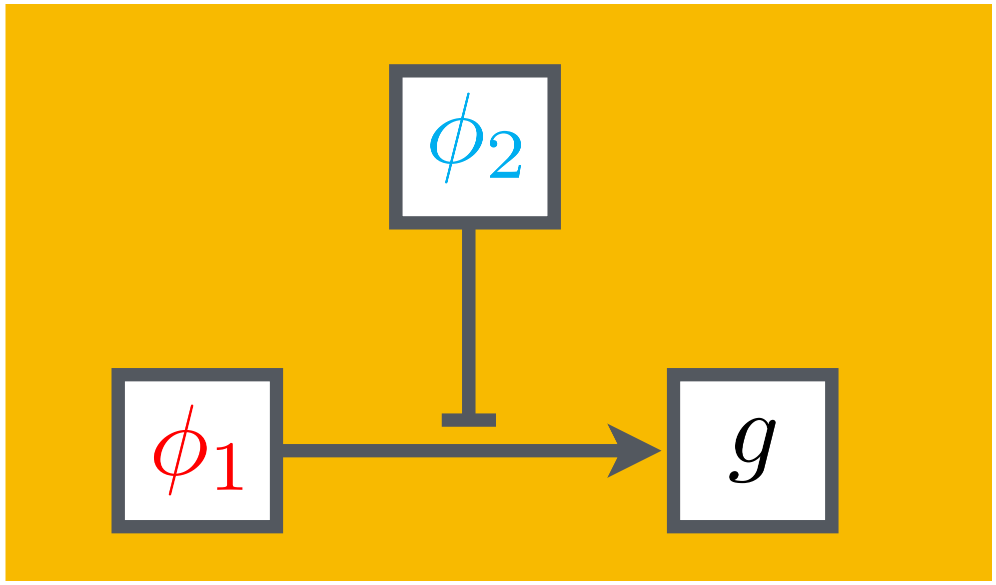
# TRANSCRIPTION REGULATION MODELLING

- *For dynamic (time-series) observations with continuous* $\mathbb{R}^+$ *state space, the mRNA expression is often modelled as the ODE*

$$\frac{dM}{dt} = \tau(t) - \delta M(t)$$

$\tau(t)$: transcription (birth) rate

$\delta$: degradation (decay) rate




Graph of a simple example of transcriptional regulation

*where some of the models used for transcription are*

$$\tau(t) = \alpha_0 + \alpha_1 P_{\phi_1}(t) + \alpha_2 P_{\phi_2}(t) + \ldots$$

$$\tau(t) = \alpha_0 + \alpha_1 P_{\phi_1}(t) + \alpha_2 P_{\phi_2}(t) + \alpha_{12} P_{\phi_1}(t) P_{\phi_2}(t) + \ldots$$

$$\tau(t) = \alpha_0 + \alpha_1 f_1(P_{\phi_1}(t), \theta_1) + \alpha_2 f_2(P_{\phi_2}, \theta_2)(t) + \ldots$$

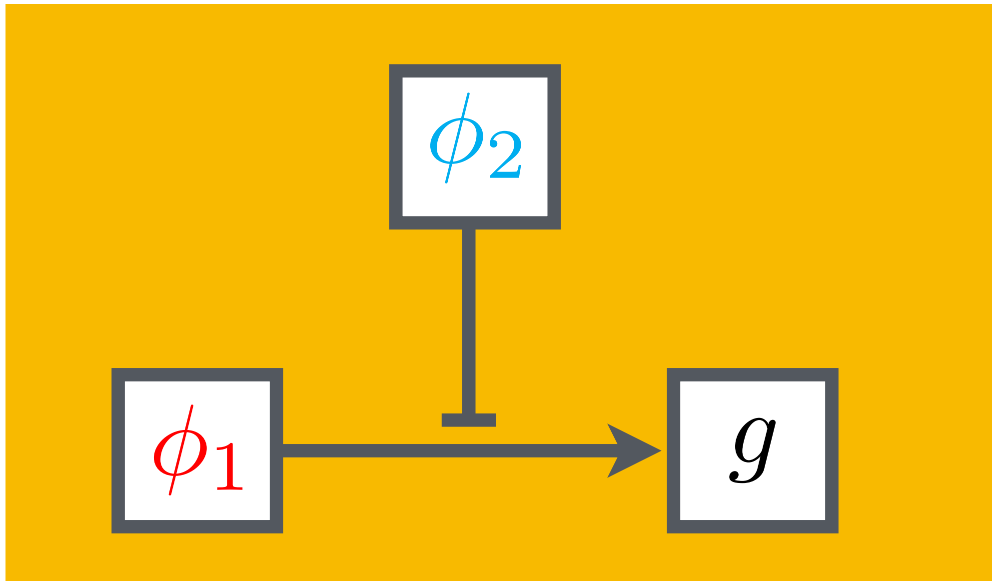
# TRANSCRIPTION REGULATION MODELLING

➤ *For dynamic (time-series) observations with continuous* $\mathbb{R}^+$ *state space, the mRNA expression is often modelled as the ODE*

$$\frac{dM}{dt} = \tau(t) - \delta M(t)$$

$\tau(t)$: transcription (birth) rate

$\delta$: degradation (decay) rate




Graph of a simple example of transcriptional regulation

*where some of the models used for transcription are*

$$\tau(t) = \alpha_0 + \alpha_1 P_{\phi_1}(t) + \alpha_2 P_{\phi_2}(t) + \dots$$

$$\tau(t) = \alpha_0 + \alpha_1 P_{\phi_1}(t) + \alpha_2 P_{\phi_2}(t) + \alpha_{12} P_{\phi_1}(t) P_{\phi_2}(t) + \dots$$

$$\tau(t) = \alpha_0 + \alpha_1 f_1(P_{\phi_1}(t), \theta_1) + \alpha_2 f_2(P_{\phi_2}, \theta_2)(t) + \dots$$

**Challenges: interpretability, identifiability, speed, accuracy/fit to data**

# TRANSCRIPTION REGULATION MODELLING

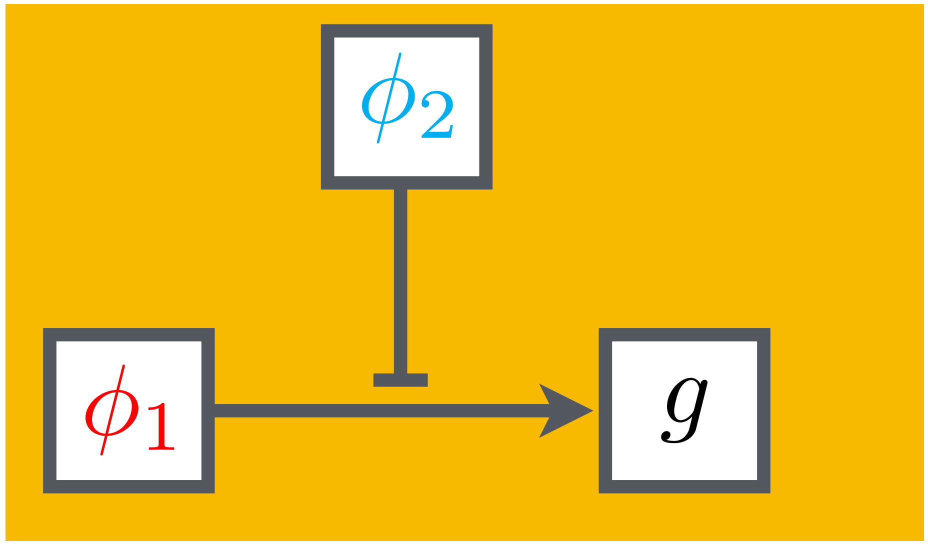
- *For dynamic (time-series) observations with continuous* $\mathbb{R}^+$ *state space, the mRNA expression is often modelled as the ODE*

k=1,2,...,K (experiment)

$$\frac{dM^{(k)}}{dt} = \tau^{(k)}(t) - \delta M^{(k)}(t)$$

$\tau(t)$: transcription (birth) rate

$\delta$: degradation (decay) rate




Graph of a simple example of transcriptional regulation

*where, some of the models used for transcription are*

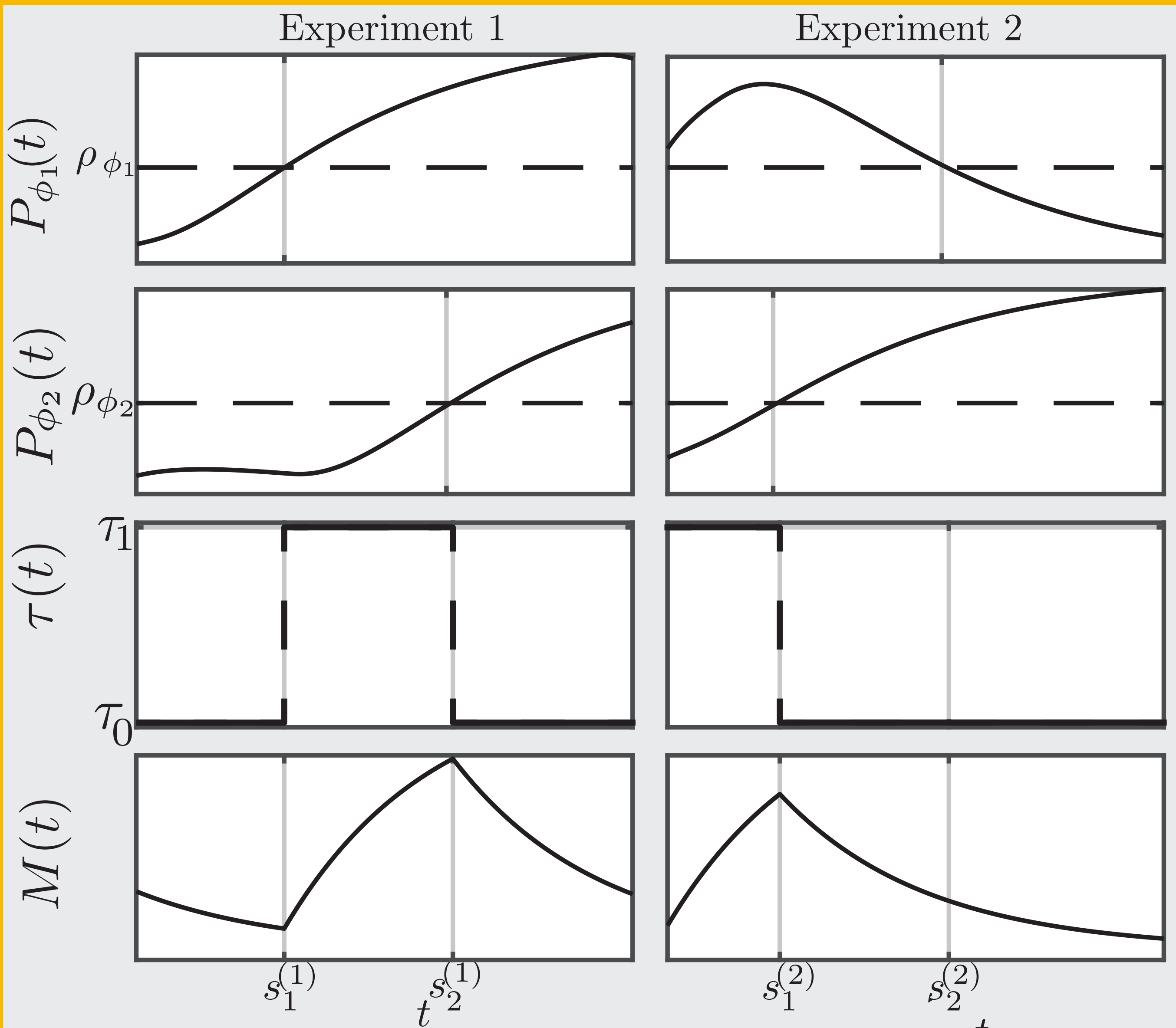
$$\tau^{(k)}(t) = \alpha_0^{(k)} + \alpha_1^{(k)} P_{\phi_1}^{(k)}(t) + \alpha_2^{(k)} P_{\phi_2}^{(k)}(t) + \dots$$

$$\tau^{(k)}(t) = \alpha_0^{(k)} + \alpha_1^{(k)} P_{\phi_1}^{(k)}(t) + \alpha_2^{(k)} P_{\phi_2}^{(k)}(t) + \alpha_{12}^{(k)} P_{\phi_1}^{(k)}(t) P_{\phi_2}^{(k)}(t) + \dots$$

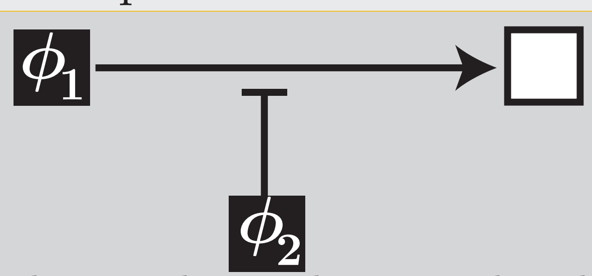
$$\tau^{(k)}(t) = \alpha_0^{(k)} + \alpha_1^{(k)} f_1^{(k)}(P_{\phi_1}^{(k)}(t), \theta_1^{(k)}) + \alpha_2^{(k)} f_2^{(k)}(P_{\phi_2}^{(k)}, \theta_2)(t) + \dots$$

**Challenges: interpretability, identifiability, speed, accuracy/fit to data**

# TRANSCRIPTION REGULATION SWITCH (TRS) MODEL




(a) Example of the TRS model




(b) Diagram of the regulation logic in the above model

- *The mRNA expression of the target gene*

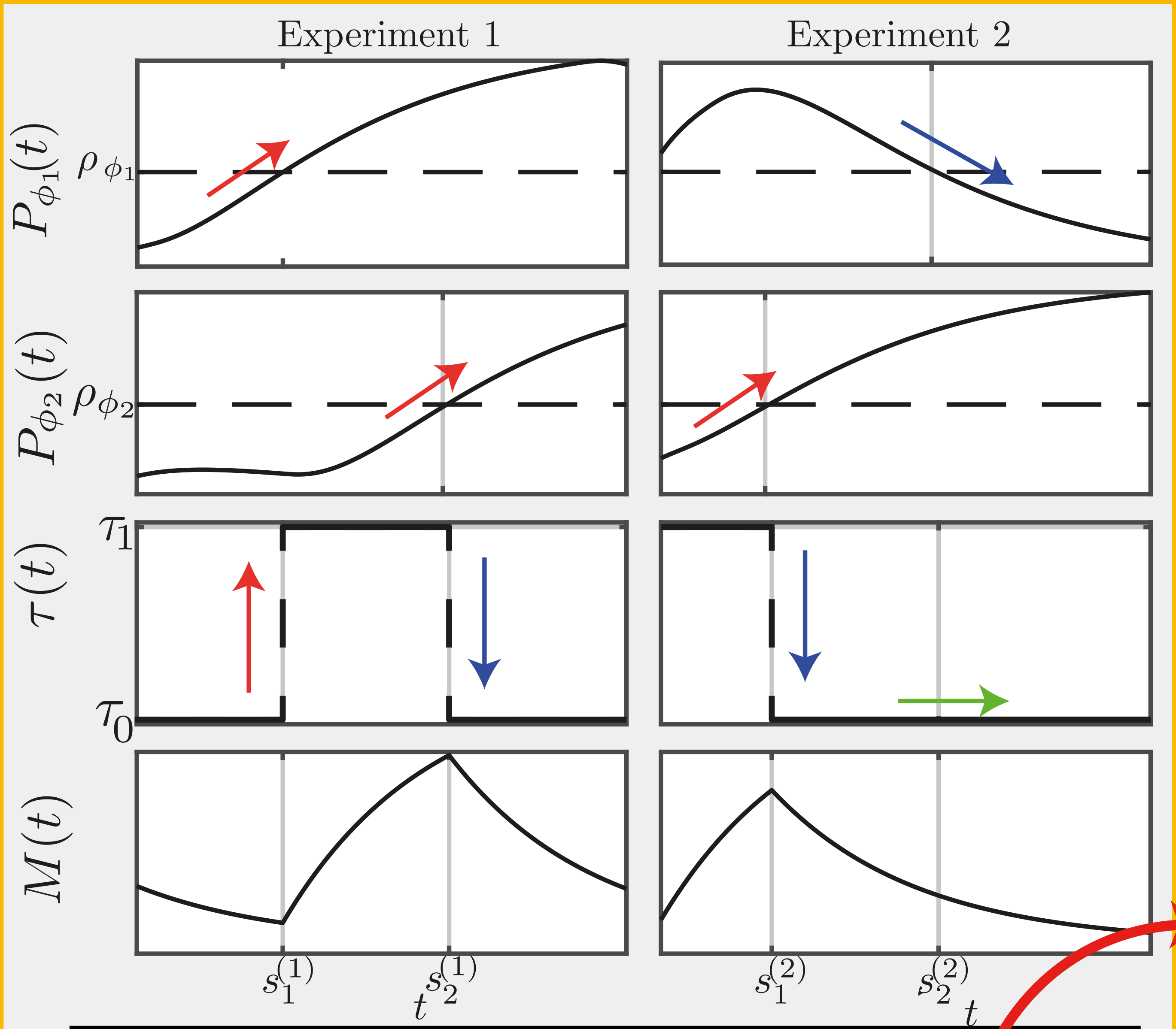
$$\frac{dM^{(k)}}{dt} = \tau(t) - \delta M^{(k)}(t)$$

*where the transcription rate*

$$\tau(t) = \begin{cases} \tau_0, & t \in [0, s_1], \\ \tau_1, & t \in [s_1, s_2), \\ \vdots & \\ \tau_q, & t \in [s_q, s_{q+1}], \end{cases}$$

- *The switches* $s_i$ *occur when one of regulators* $\phi_1, \phi_2, \ldots, \phi_n$ *crosses its threshold* $\rho_{\phi_i}$.
- *The* $\tau$*-rates are forced to be the same in time-intervals where all regulators are at the same activation state*

# TRANSCRIPTION REGULATION SWITCH (TRS) MODEL




**Biological assumption 1: transcriptional switches are caused by large changes in the TF's levels**

- *The mRNA expression of the target gene*

$$\frac{dM}{dt} = \tau(t) - \delta M(t)$$

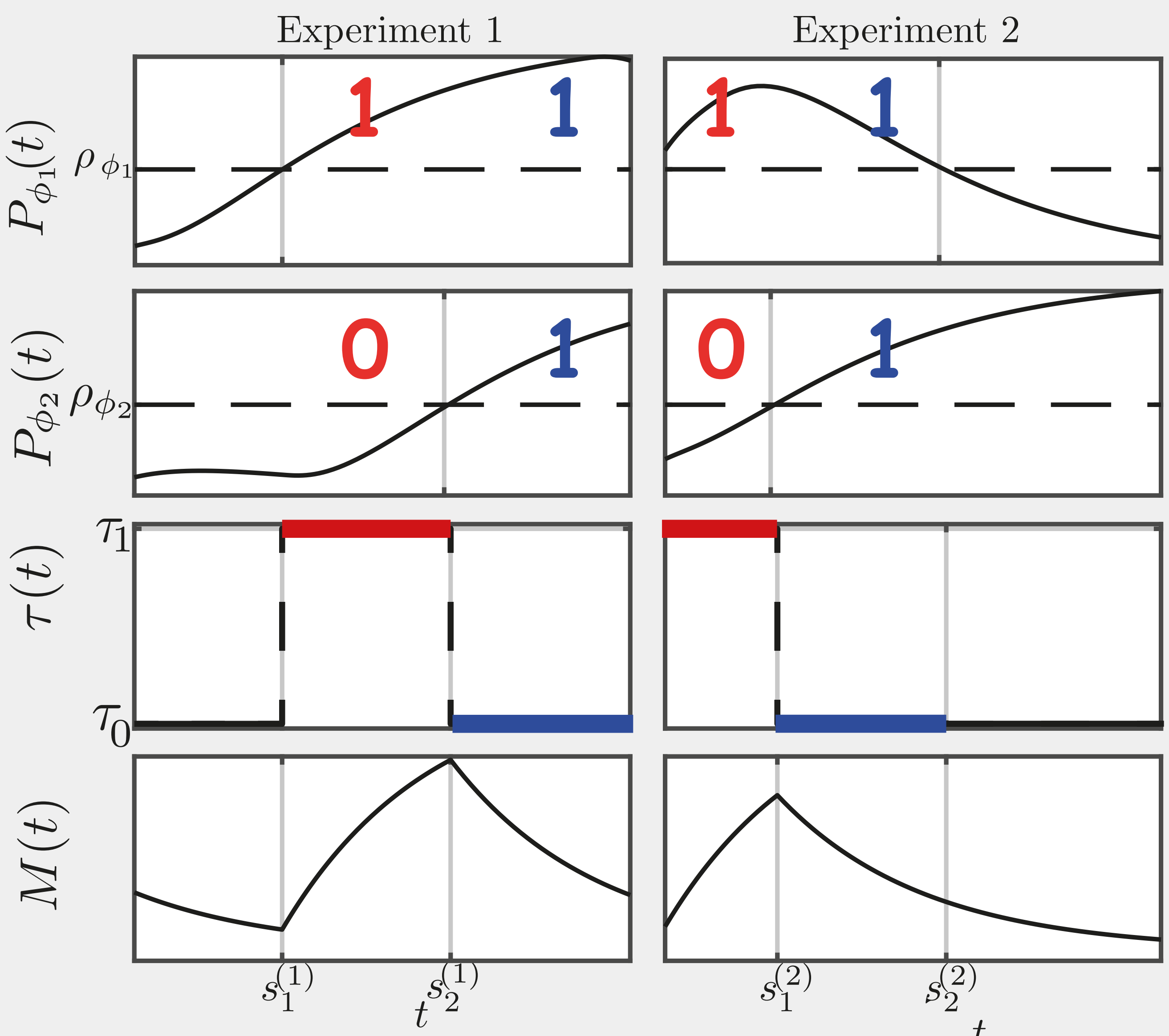
*where the transcription rate*

$$\tau(t) = \begin{cases} \tau_0, & t \in [0, s_1], \\ \tau_1, & t \in [s_1, s_2), \\ \vdots & \\ \tau_q, & t \in [s_q, s_{q+1}], \end{cases}$$

- ***The switches $s_i$ occur when one of regulators $\phi_1, \phi_2, \ldots, \phi_n$ crosses its $\rho_{\phi_i}$ threshold.***

- *The $\tau$-rates are forced to be the same in time-intervals where all regulators are at the same activation state*

# TRANSCRIPTION REGULATION SWITCH (TRS) MODEL




**Biological assumption 2: The gene network is always the same. Different parts of the network may be active in different times & conditions.**

- *The mRNA expression of the target gene*

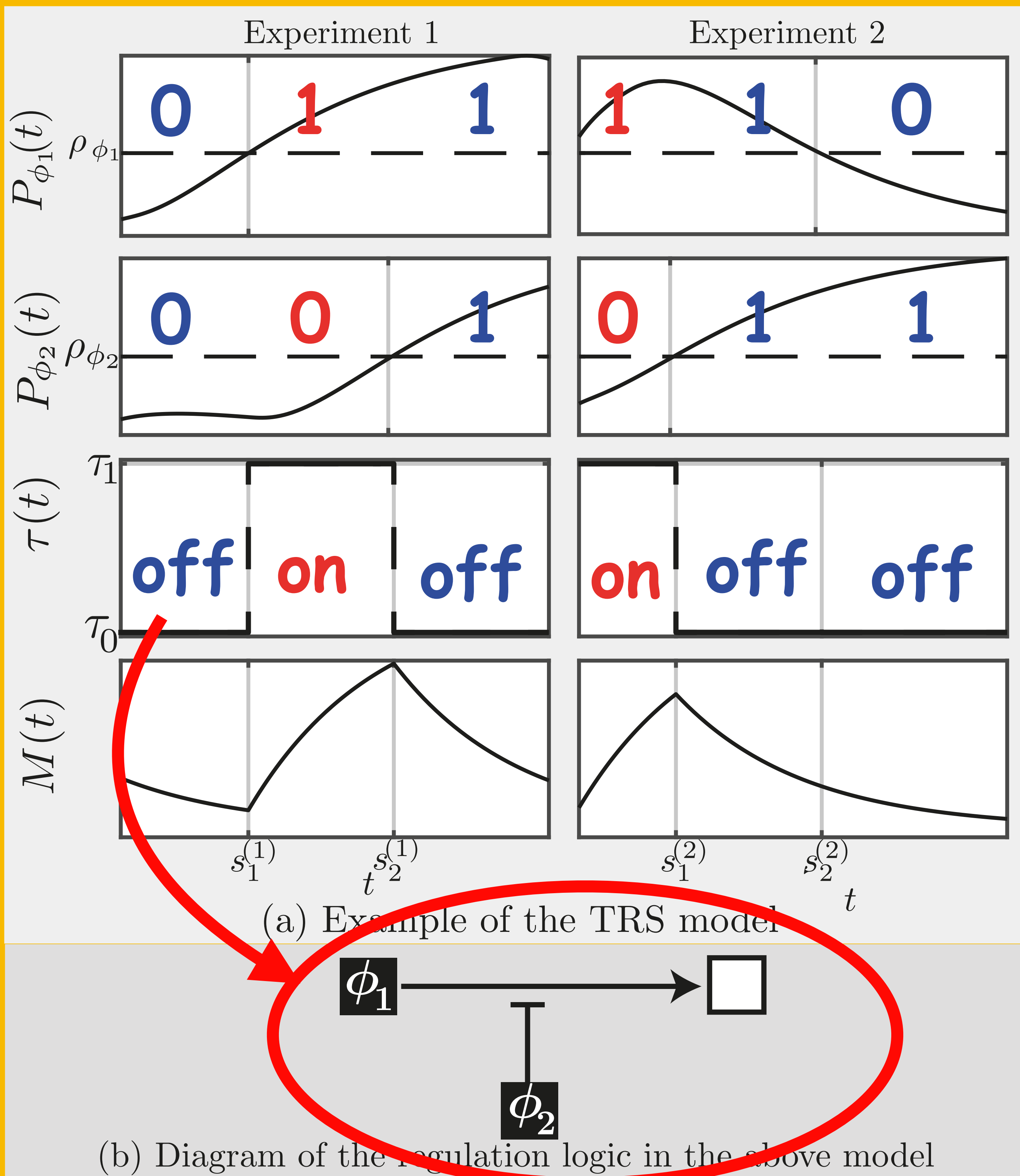
$$\frac{dM}{dt} = \tau(t) - \delta M(t)$$

*where the transcription rate*

$$\tau(t) = \begin{cases} \tau_0, & t \in [0, s_1], \\ \tau_1, & t \in [s_1, s_2), \\ \vdots & \\ \tau_q, & t \in [s_q, s_{q+1}], \end{cases}$$

- *The switches $s_i$ occur when one of regulators $\phi_1, \phi_2, \ldots, \phi_n$ crosses its threshold $\rho_{\phi_i}$*
- *The $\tau$-rates are forced to be the same in time-intervals where all regulators are at the same activation state.*

# TRANSCRIPTION REGULATION SWITCH (TRS) MODEL



(a) Example of the TRS model

(b) Diagram of the regulation logic in the above model

- *The mRNA expression of the target gene*

$$\frac{dM}{dt} = \tau(t) - \delta M(t)$$

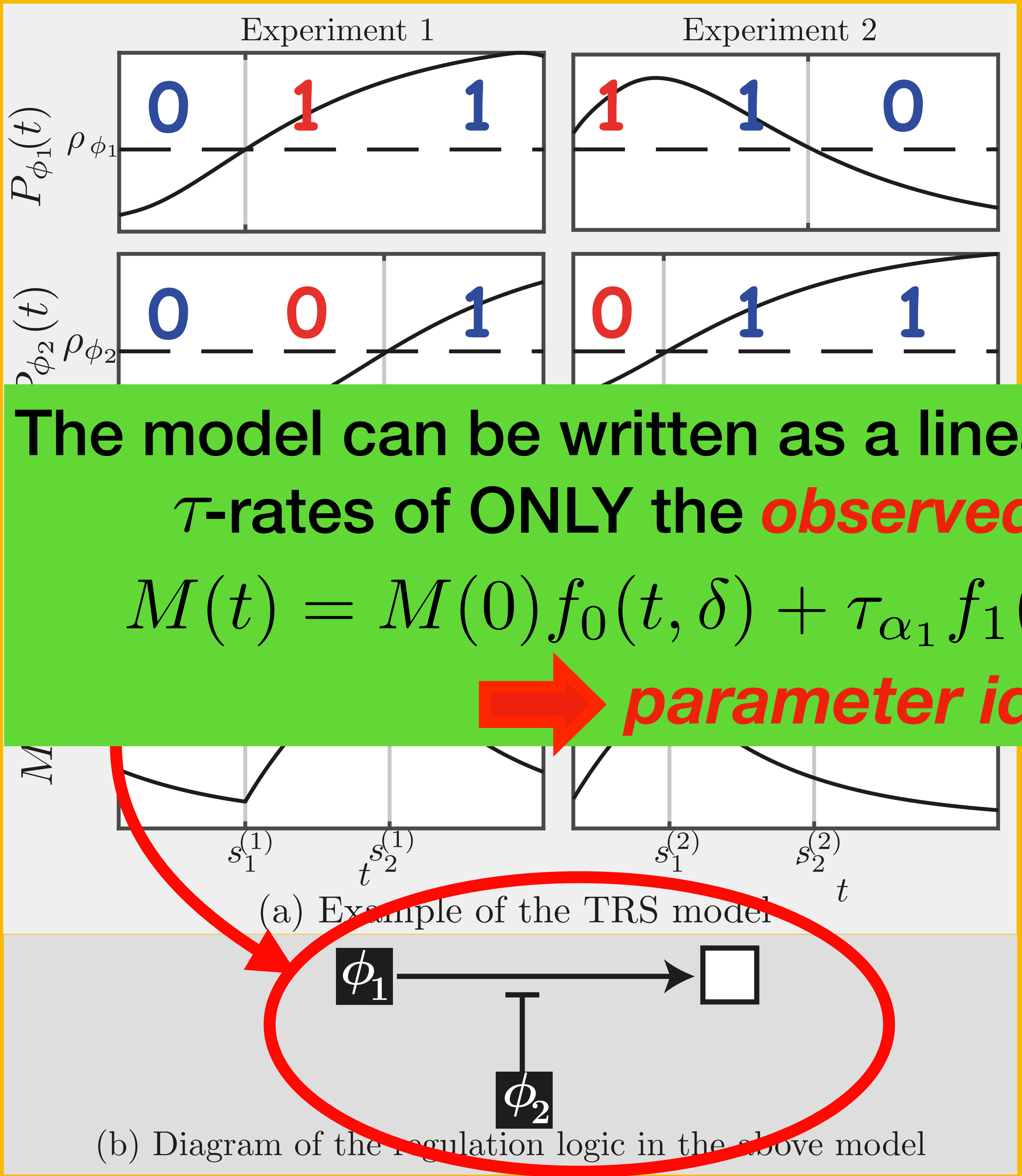
*where the transcription rate*

$$\tau_q, \quad t \in [s_q, s_{q+1}],$$

**The TRS model targets biologically interpretable descriptions of gene regulation**

- *The switches $s_i$ occur when one of regulators $\phi_1, \phi_2, \ldots, \phi_n$ crosses its threshold $\rho_{\phi_i}$.*
- *The $\tau$-rates are forced to be the same in time-intervals where all regulators are at the same activation state*

# TRANSCRIPTION REGULATION SWITCH (TRS) MODEL




(a) Example of the TRS model

(b) Diagram of the regulation logic in the above model

- *The mRNA expression of the target gene*

$$\frac{dM}{dt} = \tau(t) - \delta M(t)$$

*where the transcription rate*

$$\begin{cases} \tau_0, & t \in [0, s_1], \end{cases}$$

**The model can be written as a linear regression with coefficients the $\tau$-rates of ONLY the *observed* TF states $\alpha_1, \alpha_2, \ldots, \alpha_q$**

$$M(t) = M(0) f_0(t, \delta) + \tau_{\alpha_1} f_1(T_{\alpha_1}, \delta) + \cdots + \tau_{\alpha_q} f_q(T_{\alpha_q}, \delta)$$

➡ ***parameter identifiability***

- *The switches $s_i$ occur when one of regulators $\phi_1, \phi_2, \ldots, \phi_n$ crosses its threshold $\rho_{\phi_i}$.*
- *The $\tau$-rates are forced to be the same in time-intervals where all regulators are at the same activation state*

# INFERENCE FOR THE TRS MODEL

- *The model extends to multiple experimental conditions*

$$Y_j(t) = M_j(t) + \epsilon_j(t)$$

$$j = 1, 2, \ldots, J \ \text{(experiment)}$$

- *The parameter vector includes*
  - *number of regulators*
  - *identity of regulators*
  - *their activation threshold*
  - *and the $\tau$-rates*

*which imply the regulation mechanism*

- *The mRNA expression of the target gene*

$$\frac{dM}{dt} = \tau(t) - \delta M(t)$$

*where the transcription rate*

$$\tau(t) = \begin{cases} \tau_0, & t \in [0, s_1], \\ \tau_1, & t \in [s_1, s_2], \\ \vdots & \\ \tau_k, & t \in [s_k, s_{k+1}] \end{cases}$$

- *The switches $s_i$ occur when one of regulators $\phi_1, \phi_2, \ldots, \phi_n$ crosses its threshold $\rho_{\phi_i}$.*
- *The $\tau$-rates are forced to be the same in time-intervals where all regulators are at the same activation state*

# INFERENCE FOR THE TRS MODEL

- *The model extends to multiple experimental conditions*

$$Y_j(t) = M_j(t) + \epsilon_j(t)$$

$$j = 1, 2, \ldots, J \ \text{(experiment)}$$

- *The parameter vector includes*
  - *number of regulators*
  - *identity of regulators*
  - *their activation threshold*
  - *and the $\tau$-rates*

*that imply simple regulation mechanisms*

- *The mRNA expression of the target gene*

$$\frac{dM}{dt} = \tau(t) - \delta M(t)$$

*where the transcription rate*

$$\tau(t) = \begin{cases} \tau_0, & t \in [0, s_1], \\ \tau_1, & t \in [s_1, s_2], \\ \vdots & \\ \tau_k, & t \in [s_k, s_{k+1}] \end{cases}$$

- *The switches $s_i$ occur when one of regulators $\phi_1, \phi_2, \ldots, \phi_n$ crosses its threshold $\rho_{\phi_i}$.*
- *The $\tau$-rates are forced to be the same in time-intervals where all regulators are at the same activation state*

# INFERENCE FOR THE TRS MODEL

- *The model extends to multiple experimental conditions*

$$Y_j(t) = M_j(t) + \epsilon_j(t)$$

$$j = 1, 2, \ldots, J \ \text{(experiment)}$$

- *The parameter vector includes*
  - *number of regulators*
  - *identity of regulators*
  - *their activation threshold*
  - *and the $\tau$-rates*

*which imply the regulation mechanism*

- *The mRNA expression of the target gene*

$$\frac{dM}{dt} = \tau(t) - \delta M(t)$$

*where the transcription rate*

$$\tau(t) = \begin{cases} \tau_0, & t \in [0, s_1], \\ \tau_1, & t \in [s_1, s_2], \\ \vdots & \\ \tau_k, & t \in [s_k, s_{k+1}] \end{cases}$$

- *The switches $s_i$ occur when one of regulators $\phi_1, \phi_2, \ldots, \phi_n$ crosses its threshold $\rho_{\phi_i}$.*
- *The $\tau$-rates are forced to be the same in time-intervals where all regulators are at the same activation state*

# INFERENCE FOR THE TRS MODEL

- *The model extends to multiple experimental conditions*

$$Y_j(t) = M_j(t) + \epsilon_j(t)$$

$$j = 1, 2, \ldots, J \ \text{(experiment)}$$

- *The parameter vector includes*
  - *number of regulators*
  - *identity of regulators*
  - *their activation threshold*
  - *and the $\tau$-rates*

*which imply the regulation mechanism*

- *The mRNA expression of the target gene*

$$\frac{dM}{dt} = \tau(t) - \delta M(t)$$

*where the transcription rate*

$$\tau(t) = \begin{cases} \tau_0, & t \in [0, s_1], \\ \tau_1, & t \in [s_1, s_2], \\ \vdots & \\ \tau_k, & t \in [s_k, s_{k+1}] \end{cases}$$

- *The switches $s_i$ occur when one of regulators $\phi_1, \phi_2, \ldots, \phi_n$ crosses its threshold $\rho_{\phi_i}$.*
- *The $\tau$-rates are forced to be the same in time-intervals where all regulators are at the same activation state*

# INFERENCE FOR THE TRS MODEL

- *The model extends to multiple experimental conditions*

$$Y_j(t) = M_j(t) + \epsilon_j(t)$$

$$j = 1, 2, \ldots, J \ \text{(experiment)}$$

- *The parameter vector includes*
  - *number of regulators*
  - *identity of regulators*
  - ***their activation threshold***
  - *and the $\tau$-rates*

*which imply the regulation mechanism*

- *The mRNA expression of the target gene*

$$\frac{dM}{dt} = \tau(t) - \delta M(t)$$

*where the transcription rate*

$$\tau(t) = \begin{cases} \tau_0, & t \in [0, s_1], \\ \tau_1, & t \in [s_1, s_2], \\ \vdots & \\ \tau_k, & t \in [s_k, s_{k+1}] \end{cases}$$

- *The switches $s_i$ occur when one of regulators $\phi_1, \phi_2, \ldots, \phi_n$ crosses its threshold $\rho_{\phi_i}$.*
- *The $\tau$-rates are forced to be the same in time-intervals where all regulators are at the same activation state*

# INFERENCE FOR THE TRS MODEL

- *The model extends to multiple experimental conditions*

$$Y_j(t) = M_j(t) + \epsilon_j(t)$$

$$j = 1, 2, \dots, J \ \text{(experiment)}$$

- *The parameter vector includes*
  - *number of regulators*
  - *identity of regulators*
  - *their activation threshold*
  - ***and the $\tau$-rates***

*which imply the regulation mechanism*

- *The mRNA expression of the target gene*

$$\frac{dM}{dt} = \tau(t) - \delta M(t)$$

*where the transcription rate*

$$\tau(t) = \begin{cases} \tau_0, & t \in [0, s_1], \\ \tau_1, & t \in [s_1, s_2], \\ \vdots & \\ \tau_k, & t \in [s_k, s_{k+1}] \end{cases}$$

- *The switches $s_i$ occur when one of regulators $\phi_1, \phi_2, \dots, \phi_n$ crosses its threshold $\rho_{\phi_i}$.*
- *The $\tau$-rates are forced to be the same in time-intervals where all regulators are at the same activation state*

# INFERENCE FOR THE TRS MODEL

- *The model extends to multiple experimental conditions*

$$Y_j(t) = M_j(t) + \epsilon_j(t)$$

$$j = 1, 2, \ldots, J \text{ (experiment)}$$

- *The parameter vector includes*
  - *number of regulators*
  - *identity of regulators*
  - *their activation threshold*
  - *and the $\tau$-rates*

*which imply the regulation mechanism*

- *The mRNA expression of the target gene*

$$\frac{dM}{dt} = \tau(t) - \delta M(t)$$

*where the transcription rate*

$$\tau(t) = \begin{cases} \tau_0, & t \in [0, s_1], \\ \tau_1, & t \in [s_1, s_2], \\ \vdots & \\ \tau_k, & t \in [s_k, s_{k+1}] \end{cases}$$

- *The switches $s_i$ occur when one of regulators $\phi_1, \phi_2, \ldots, \phi_n$ crosses its threshold $\rho_{\phi_i}$.*

- *The $\tau$-rates are forced to be the same in time-intervals where all regulators are at the same activation state*

# INFERENCE FOR THE TRS MODEL

- *The model extends to multiple experimental conditions*

$$Y_j(t) = M_j(t) + \epsilon_j(t)$$

$$j = 1, 2, \ldots, J \ \ \text{(experiment)}$$

- *The parameter vector includes*
  - *number of regulators*
  - *identity of regulators*
  - *their activation threshold*
  - *and the $\tau$-rates*

*which imply the regulation mechanism*

- ***We use a trans-dimensional RJMCMC algorithm to infer the parameters***

- *The mRNA expression of the target gene*

$$\frac{dM}{dt} = \tau(t) - \delta M(t)$$

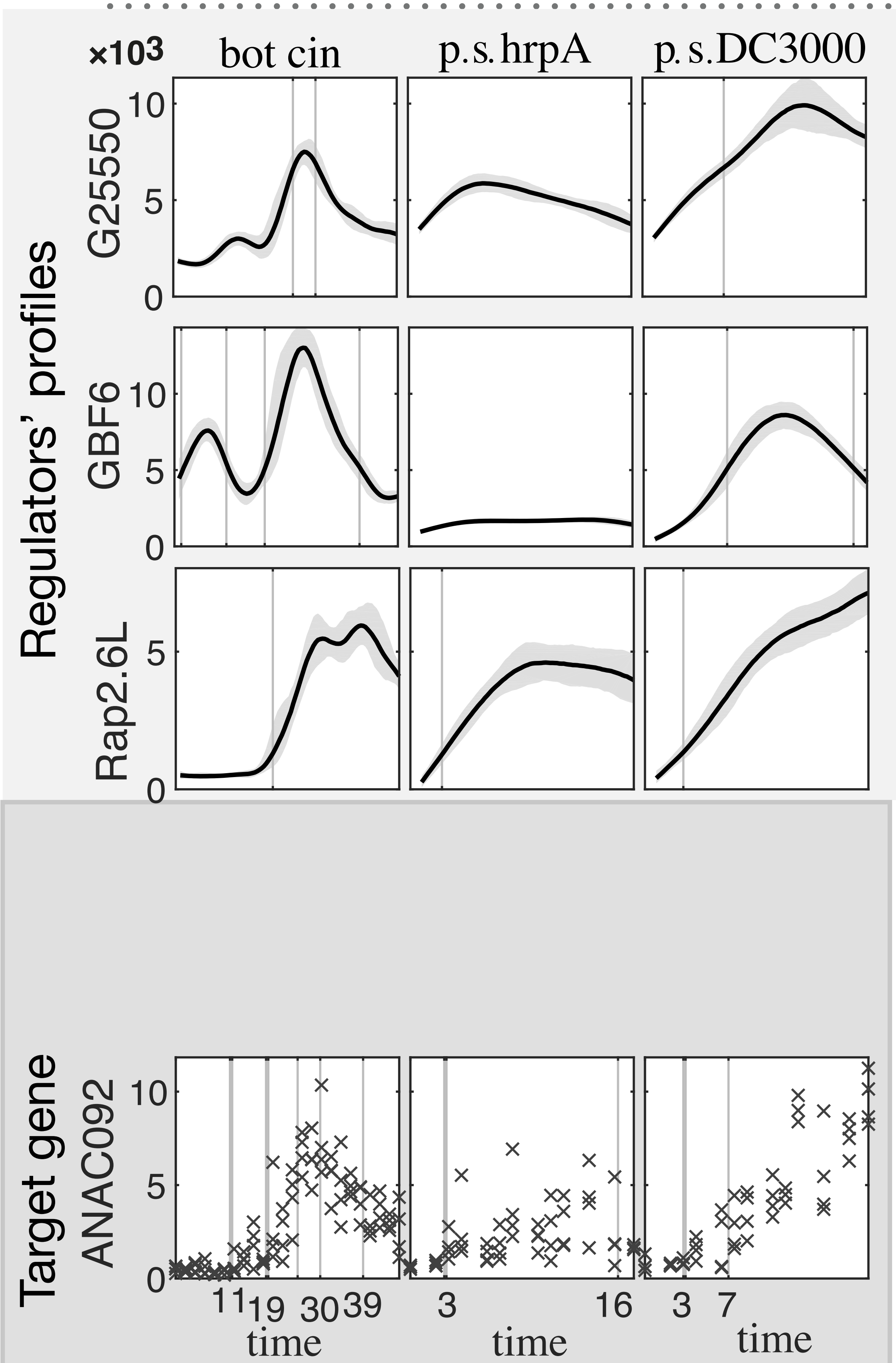
*where the transcription rate*

$$\tau(t) = \begin{cases} \tau_0, & t \in [0, s_1], \\ \tau_1, & t \in [s_1, s_2], \\ \vdots & \\ \tau_k, & t \in [s_k, s_{k+1}] \end{cases}$$

- *The switches $s_i$ occur when one of regulators $\phi_1, \phi_2, \ldots, \phi_n$ crosses its threshold $\rho_{\phi_i}$.*

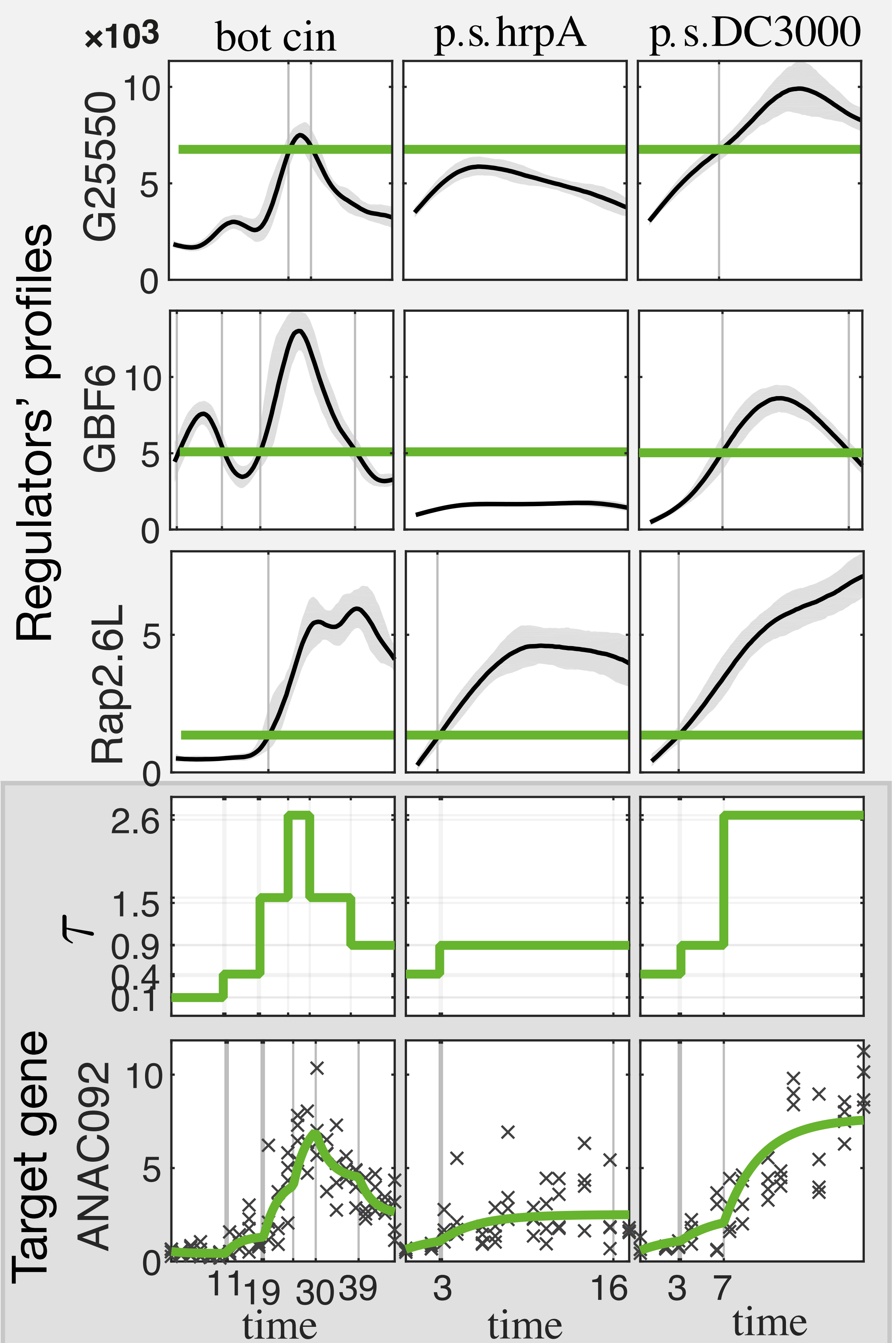
- *The $\tau$-rates are forced to be the same in time-intervals where all regulators are at the same activation state*

# APPLICATION OF TRS TO A. THALIANA MICROARRAY DATA




- *Microarray (time-series) data:*
  - ➤ *Three experimental conditions*
  - ➤ *Four replicates*
  - ➤ *Twenty candidate regulators*

# APPLICATION OF TRS TO A. THALIANA MICROARRAY DATA




- *TRS can infer the parameters and unravel* ***biologically interpretable*** *mechanisms that are* ***consistent across multiple experiments****.*
- *Good balance between model interpretability and model identifiability*
- *Output: posterior probabilities for all possible regulation models*

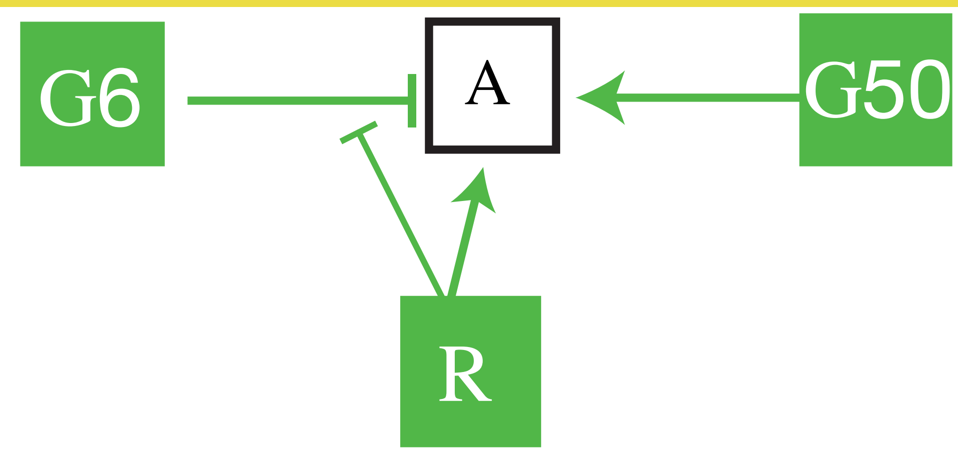
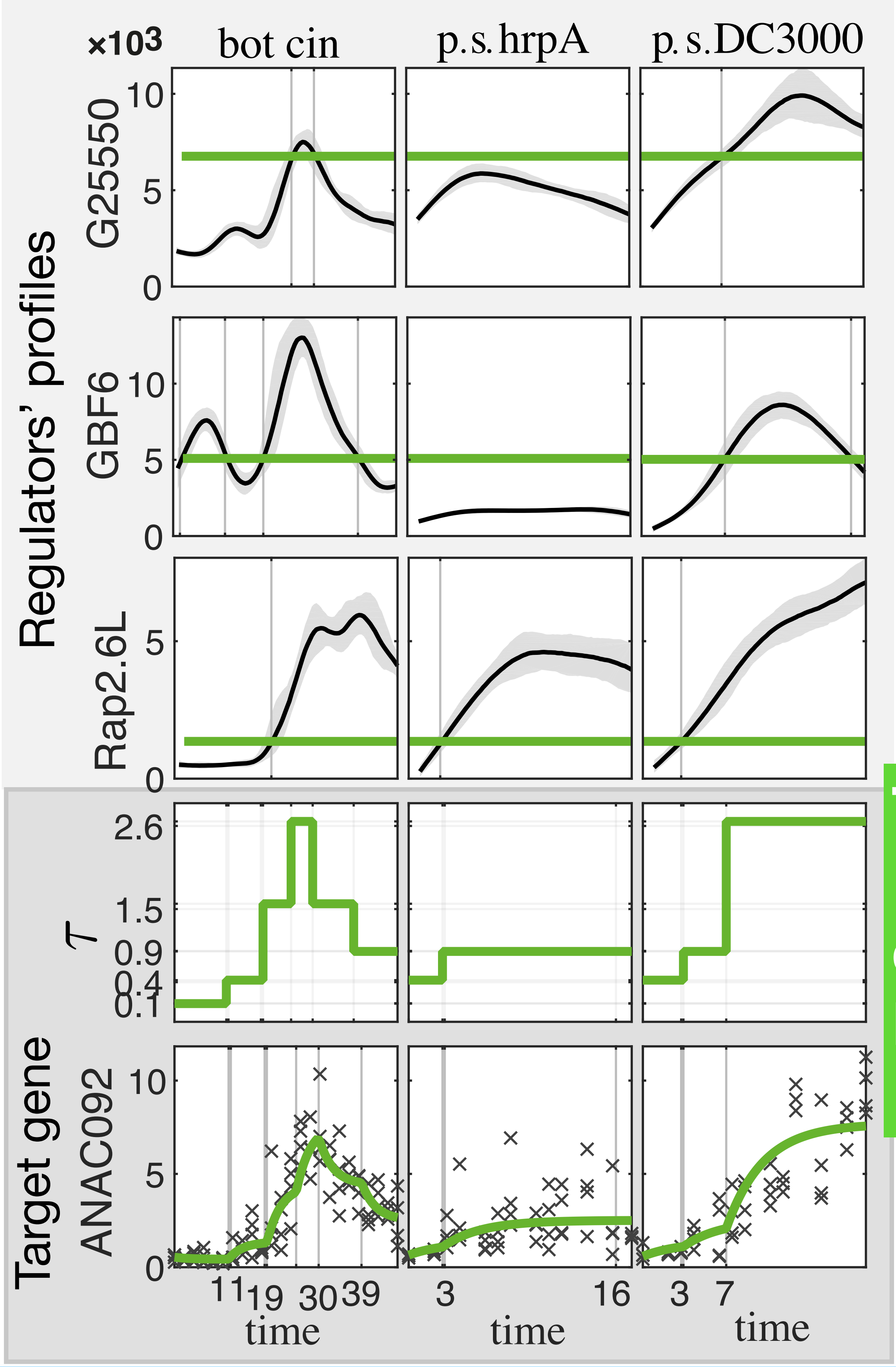



Diagram of the *a posteriori* most likely regulation mechanisms

# APPLICATION OF TRS TO A. THALIANA MICROARRAY DATA




- *TRS can infer the parameters and unravel* ***biologically interpretable*** *mechanisms that are* ***consistent across multiple experiments****.*
- *Good balance between model interpretability and model identifiability*
- *Output: posterior probabilities for all possible regulation models*

**TRS Method published in:**
Inferring transcriptional logic from multiple dynamic experiments,
Minas G, Jenkins D, Rand D, Finkenstadt B
Bioinformatics, 33, 21 (2017)

**TRS software:**
**implemented in MATLAB available at**
https://github.com/giorgosminas/TRS

# INSTEAD OF SUMMARY

**Data Driven Approaches**

**Hypothesis-driven approaches**
**Mechanistic models**

**more data**
**more simple models**

**more complex models**
**more knowledge**

PCA
Correlation networks
Multiple testing
Penalized Likelihoods
Clustering

Reaction Networks
Markov processes
ODEs
Bifurcation theory
Catastrophe theory
SDEs
PDEs

Bayesian Networks
Boolean networks
Gaussian Processes
regression models
ABC
MCMC
Information Theory


Sponsors:
BBSRC bioscience for the future

Collaborators:
Dafyd Jenkins
David Rand
Barbel Finkenstadt
Katherine Denby





University of St Andrews FOUNDED 1413
WARWICK THE UNIVERSITY OF WARWICK