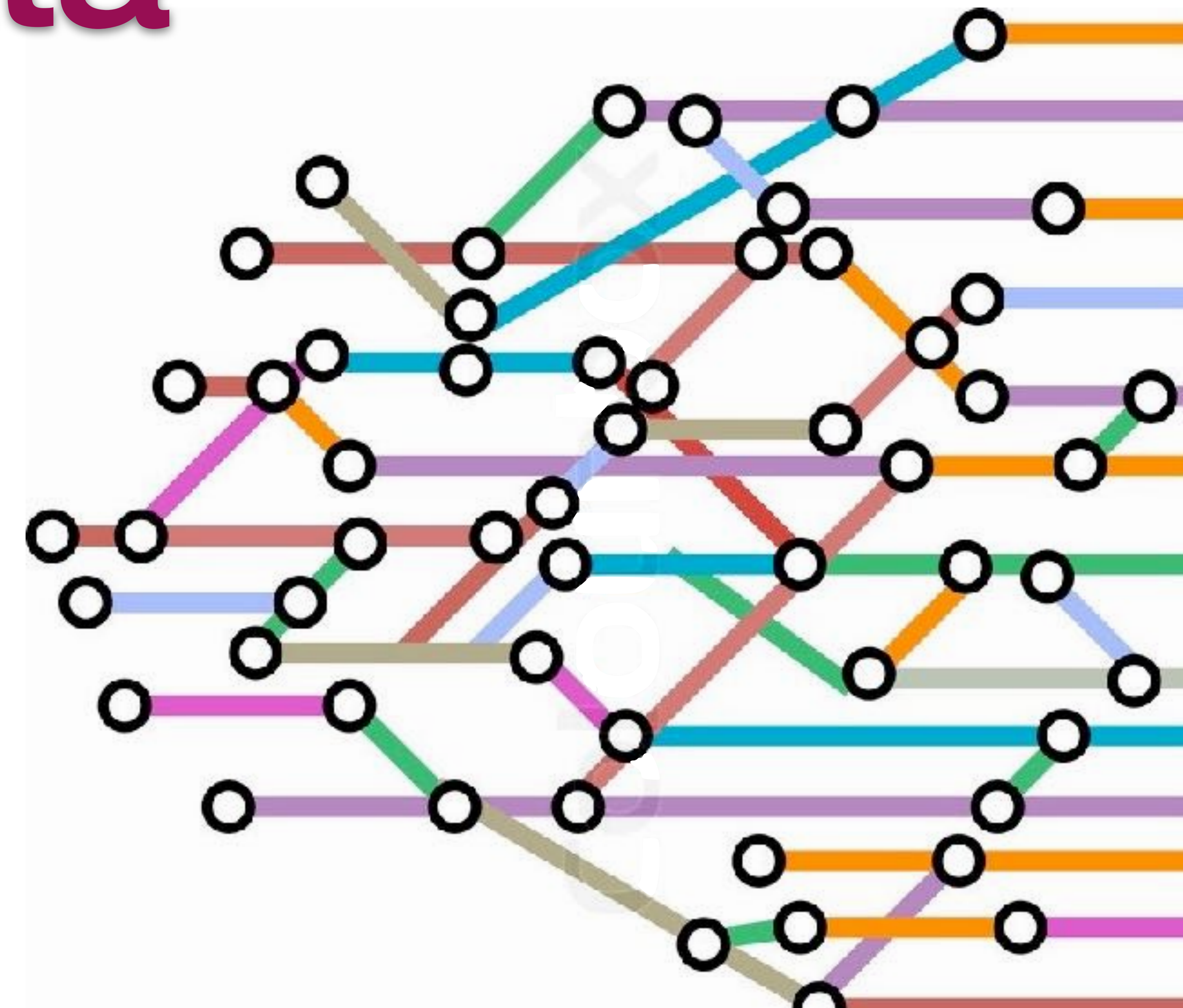
UNIVERSITÀ DEGLI STUDI
DEL SANNIO Benevento
DST
DIPARTIMENTO DI
SCIENZE E TECNOLOGIE

Reconstruct regulatory networks from trascriptomics data

With applications in functional genomics

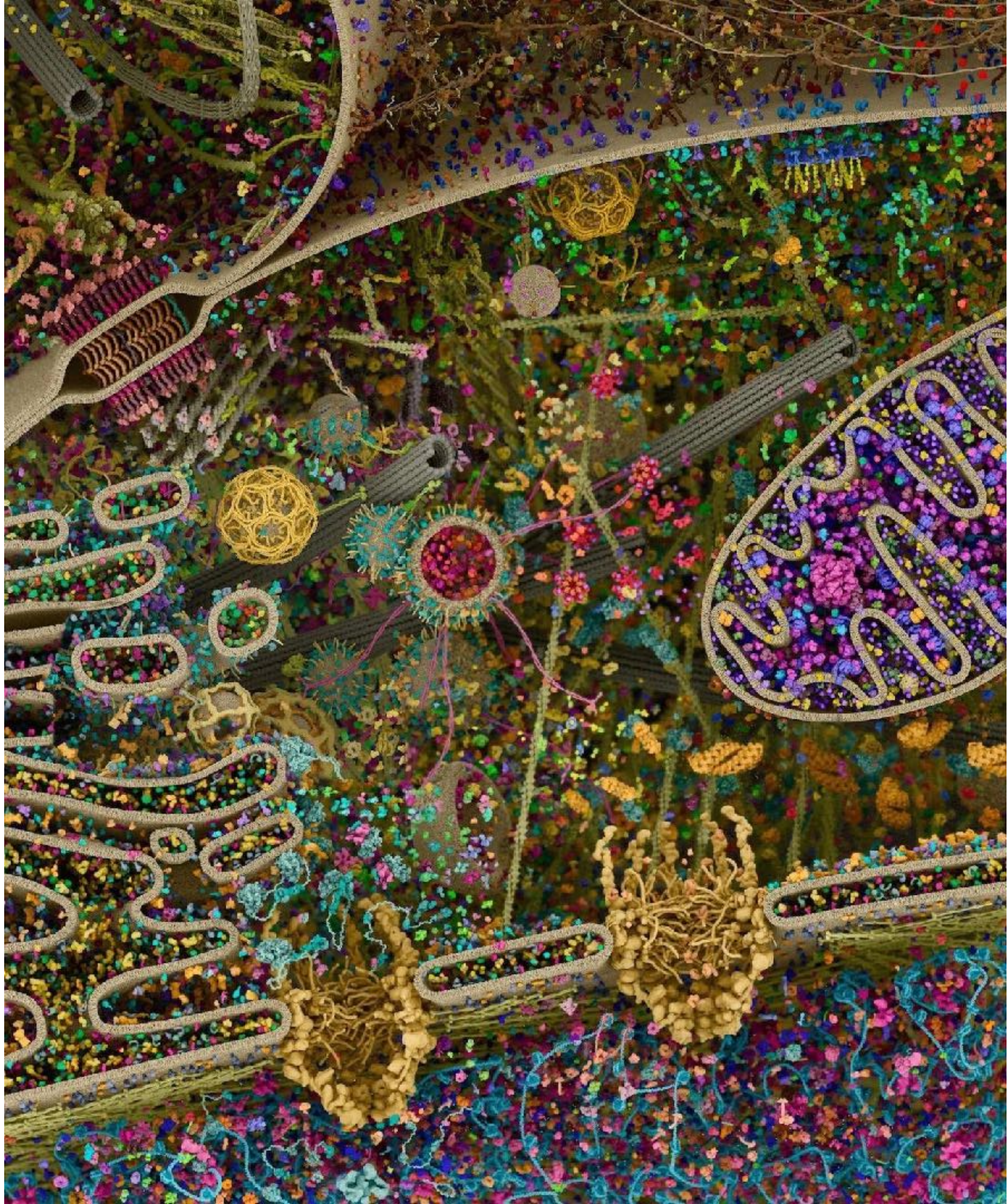
Luigi Cerulo



International Summer School on System Biology 2022 - SZN (Napoli)

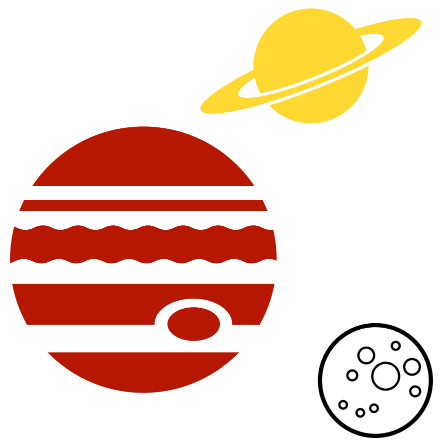
The complexity of living things

- Cell function and regulation depend on transient interactions among thousands of different macromolecules in the cell.
- It is necessary a systemic approach to understand how the cell is organized and how genes and proteins interact
  (Ludwig von Bertalanffy, 1934)



How we understand complex systems

Reverse engineering


Complex system

How we understand complex systems

Reverse engineering

Complex system

Measurements

Data

How we understand complex systems

Reverse engineering

Complex system

Measurements

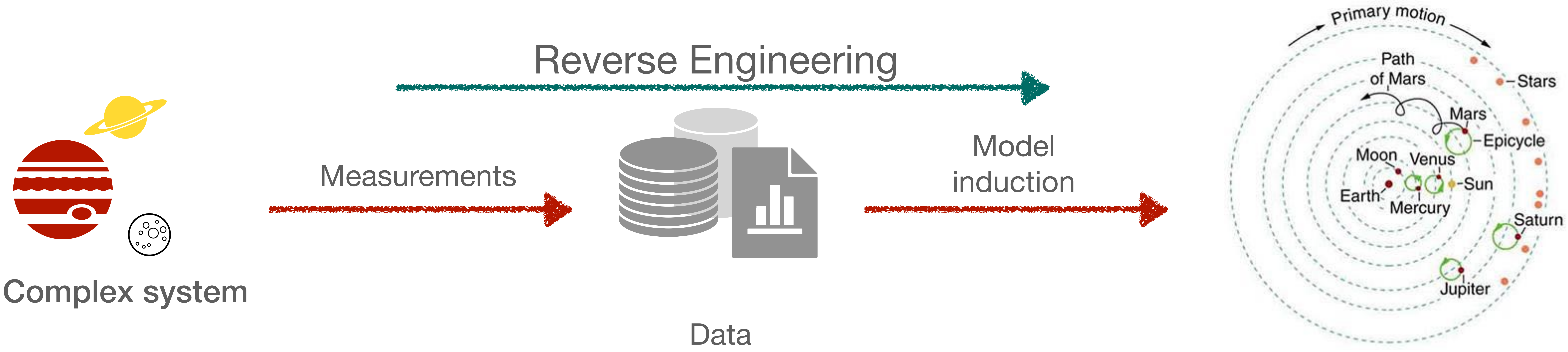
Data

Model induction

Model

How we understand complex systems

Reverse engineering



Reverse Engineering

Complex system

Measurements

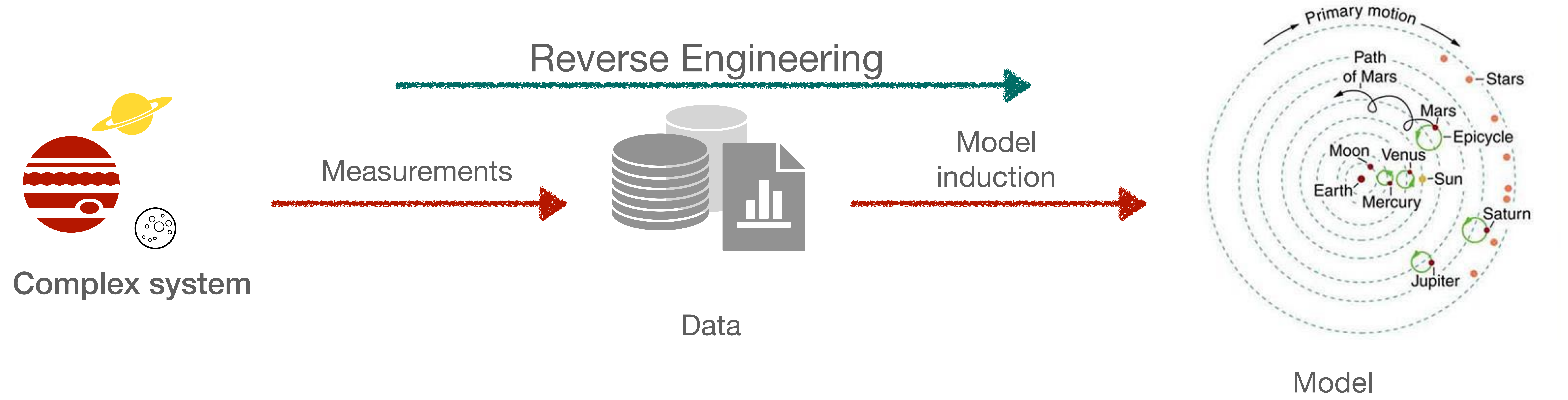
Data

Model induction

Model

How we understand complex systems

Reverse engineering



Reverse Engineering

Measurements

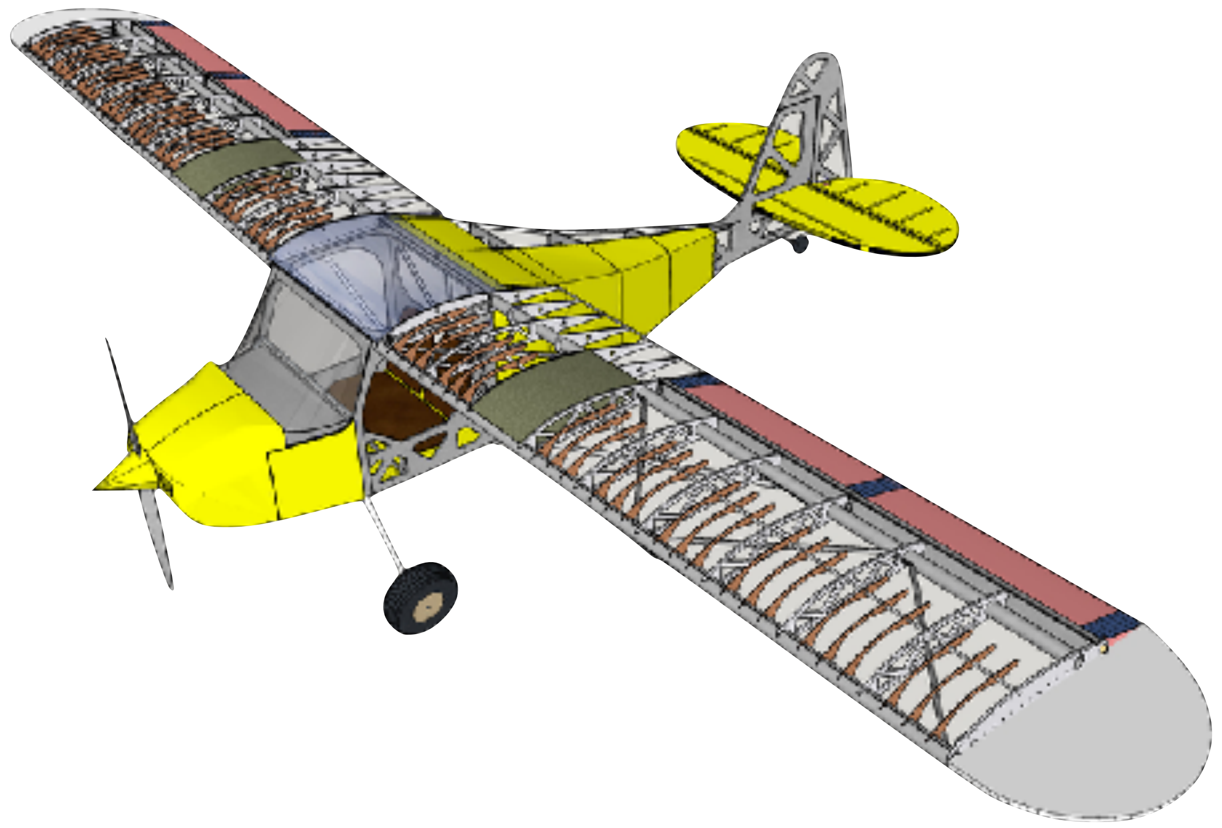
Model induction

Complex system

Data

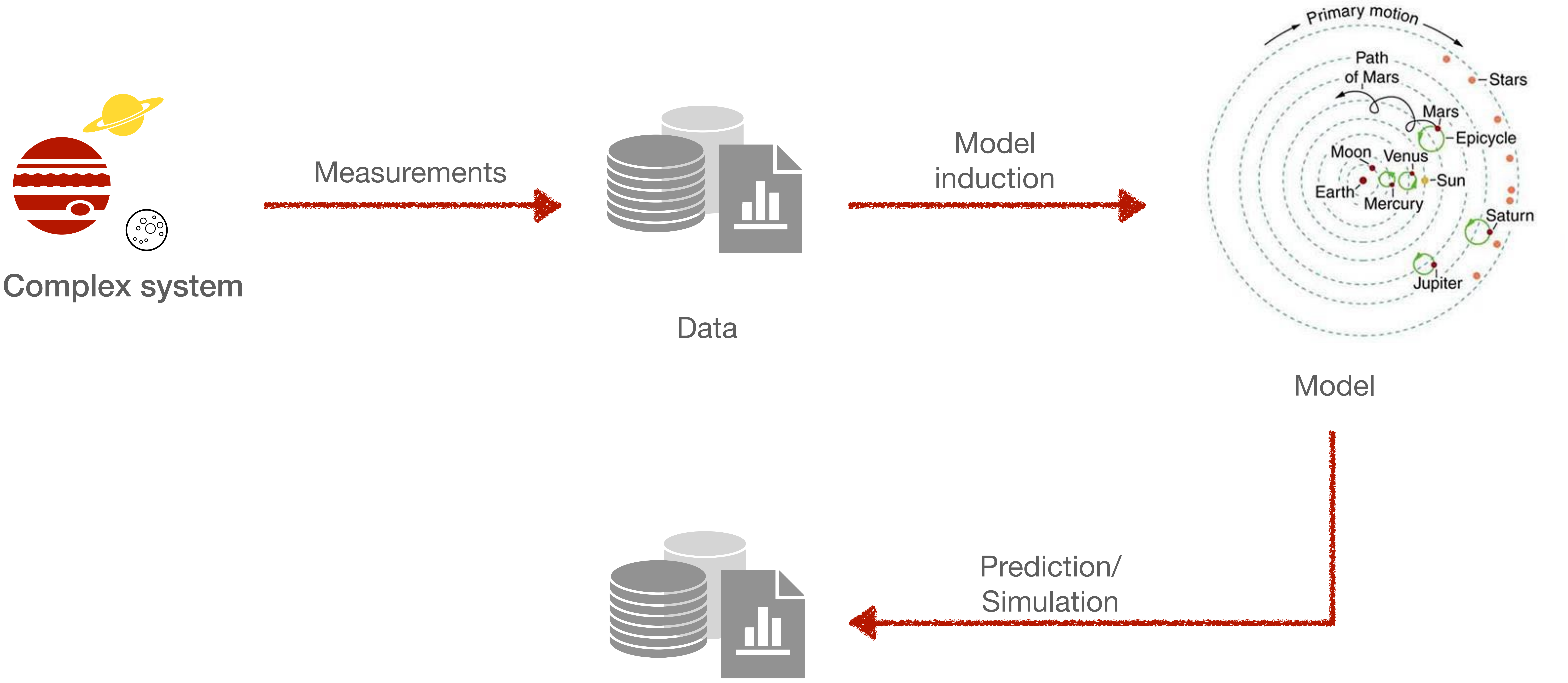
Model

Forward Engineering



How we understand complex systems

Reverse engineering



Complex system

Measurements

Data

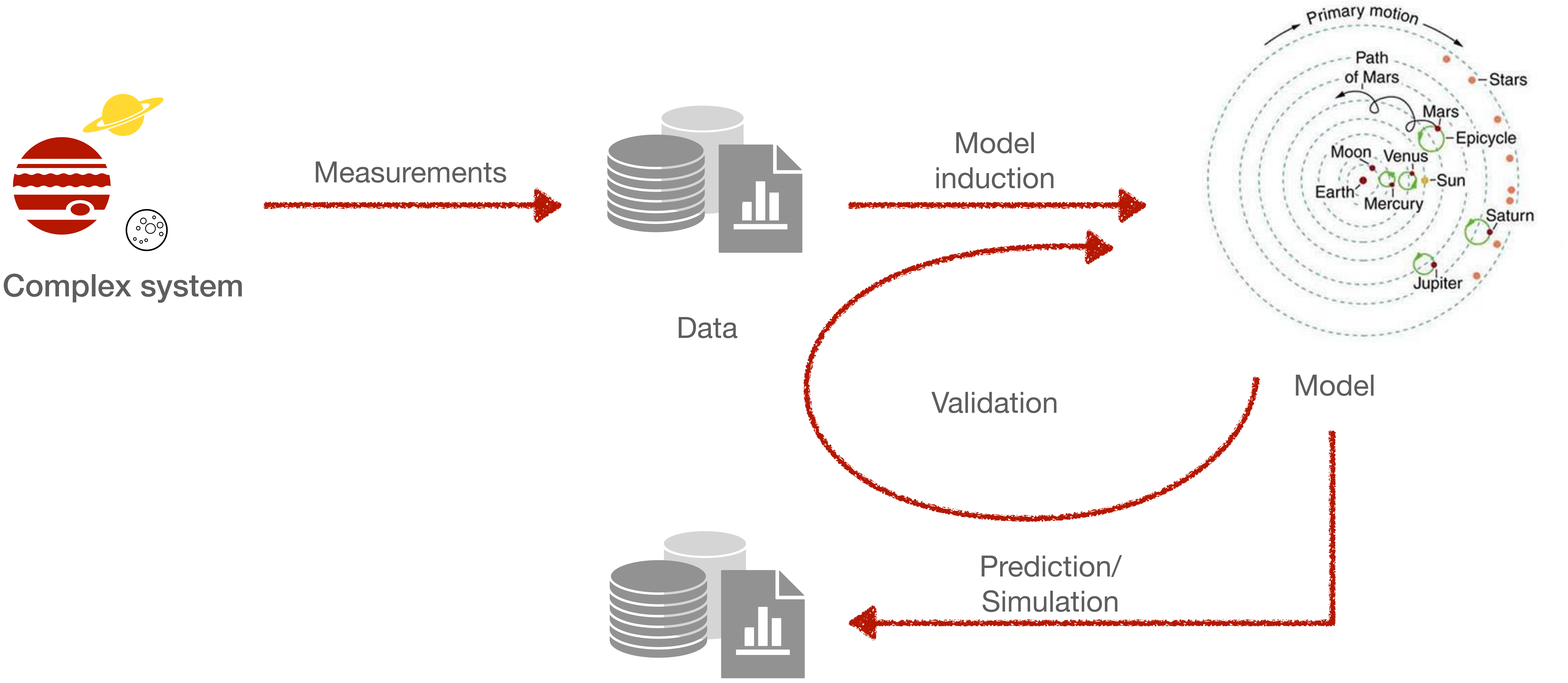
Model induction

Model

Prediction/ Simulation

How we understand complex systems

Reverse engineering



Complex system

Measurements

Data

Model induction

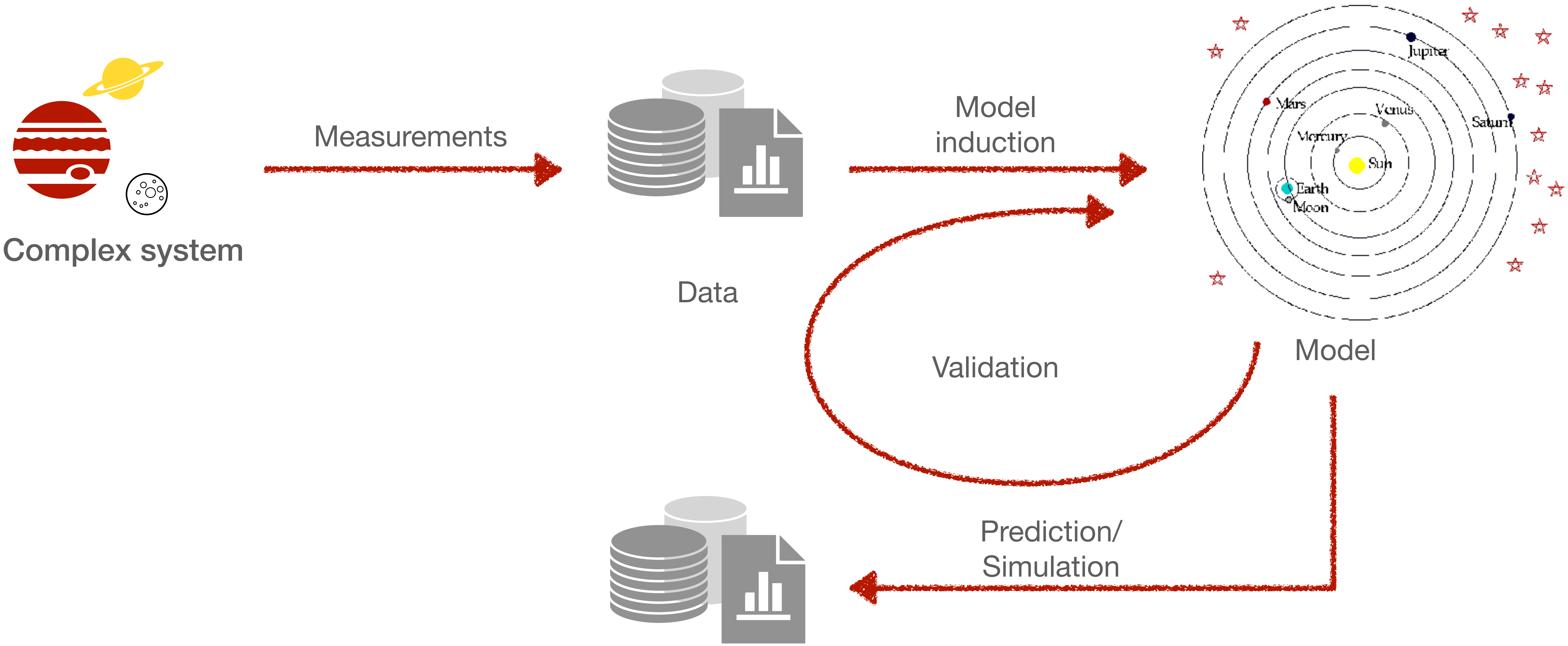
Model

Validation

Prediction/ Simulation

How we understand complex systems

Reverse engineering



Complex system

Measurements

Data

Model induction

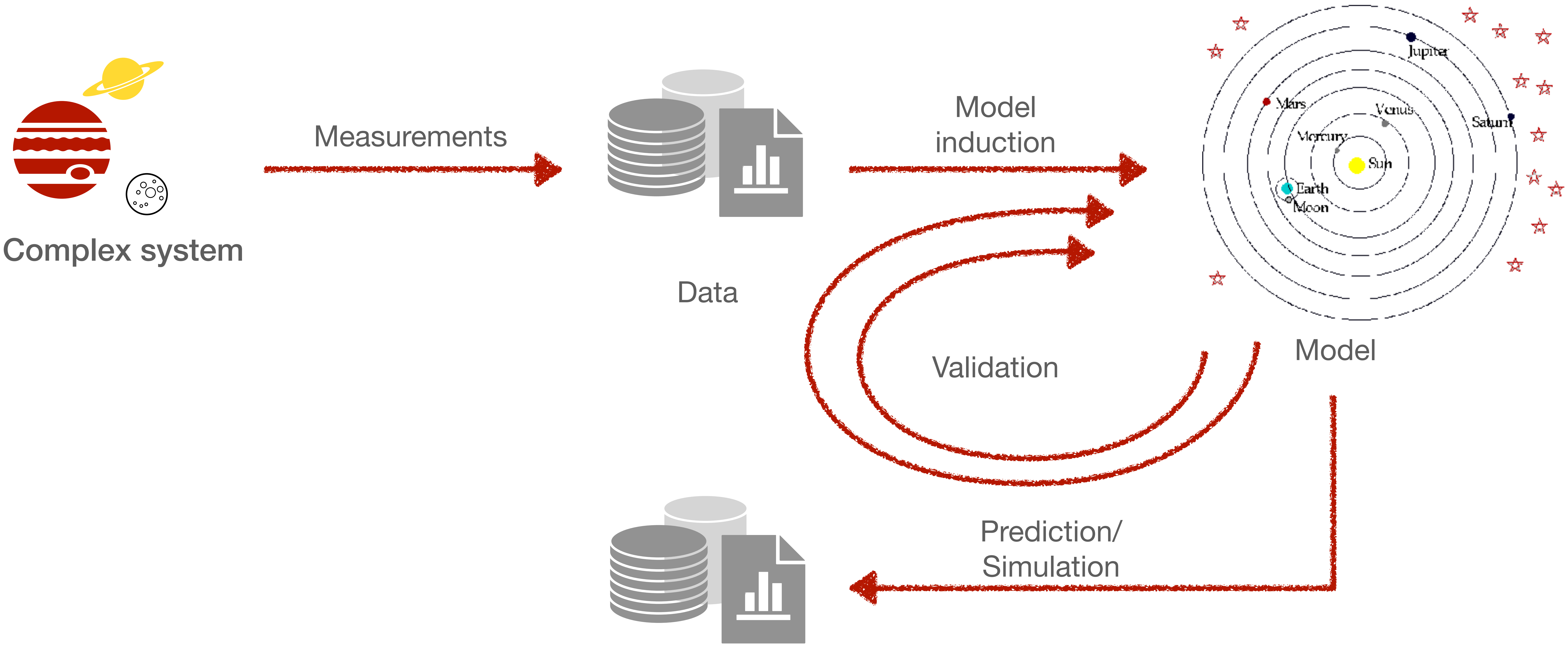
Model

Validation

Prediction/
Simulation

How we understand complex systems

Reverse engineering



Complex system

Measurements

Data

Model induction

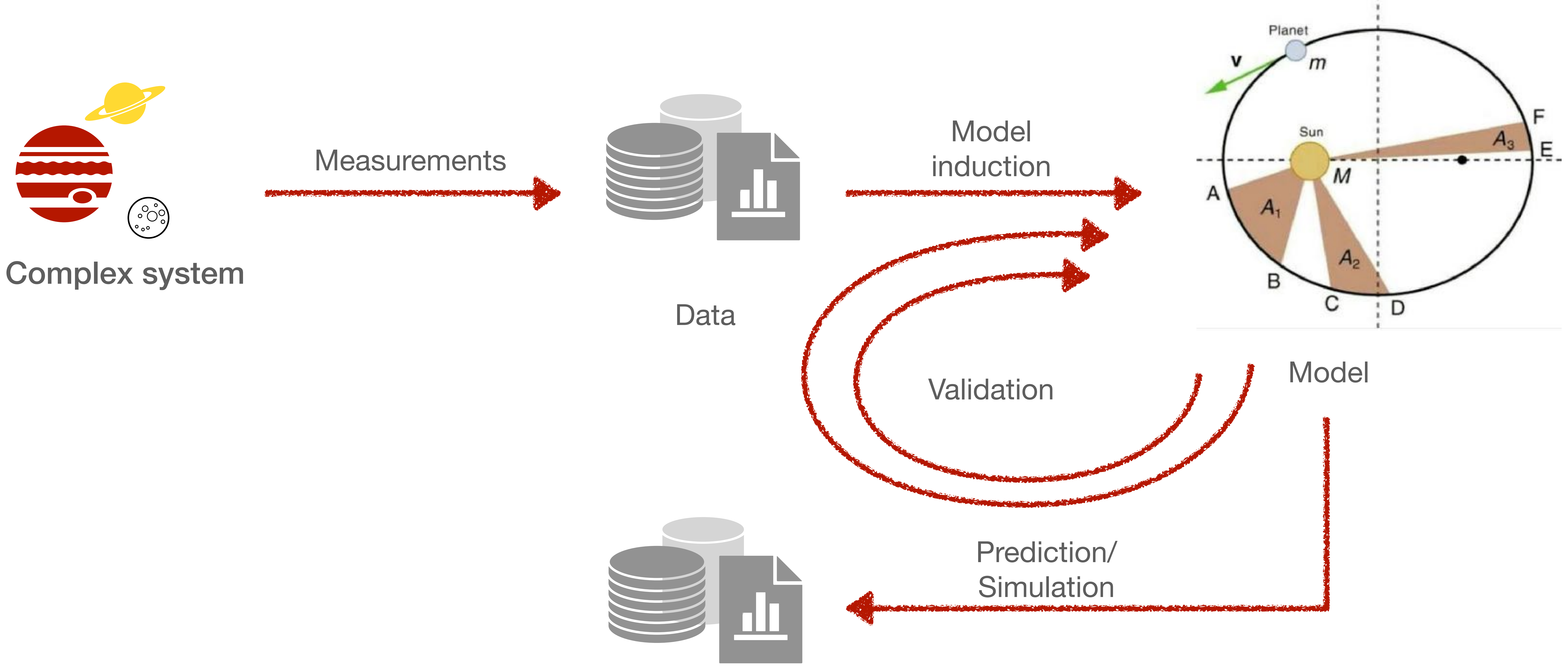
Validation

Model

Prediction/ Simulation

How we understand complex systems

Reverse engineering



Complex system

Measurements

Data

Model induction

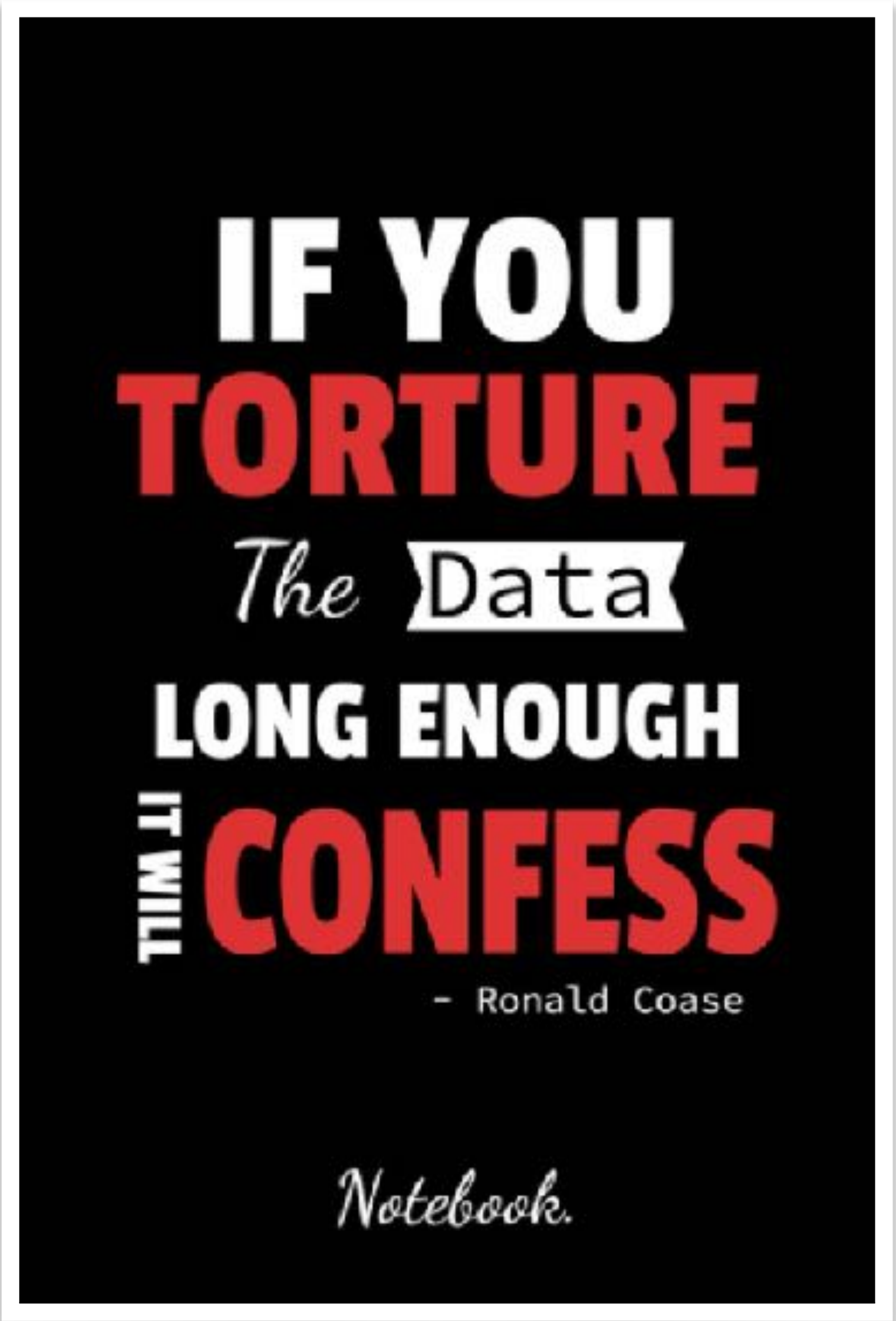
Model

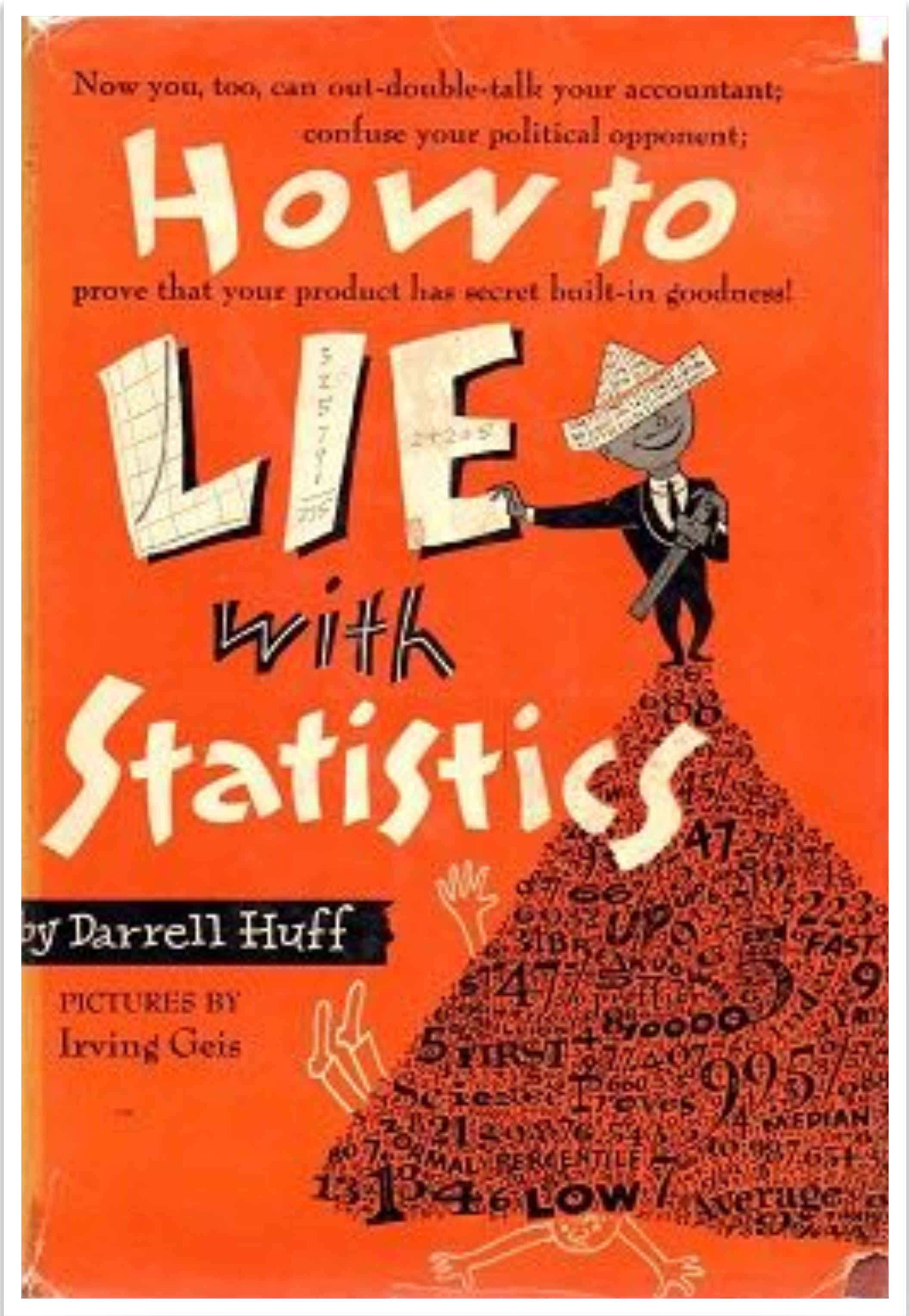
Validation

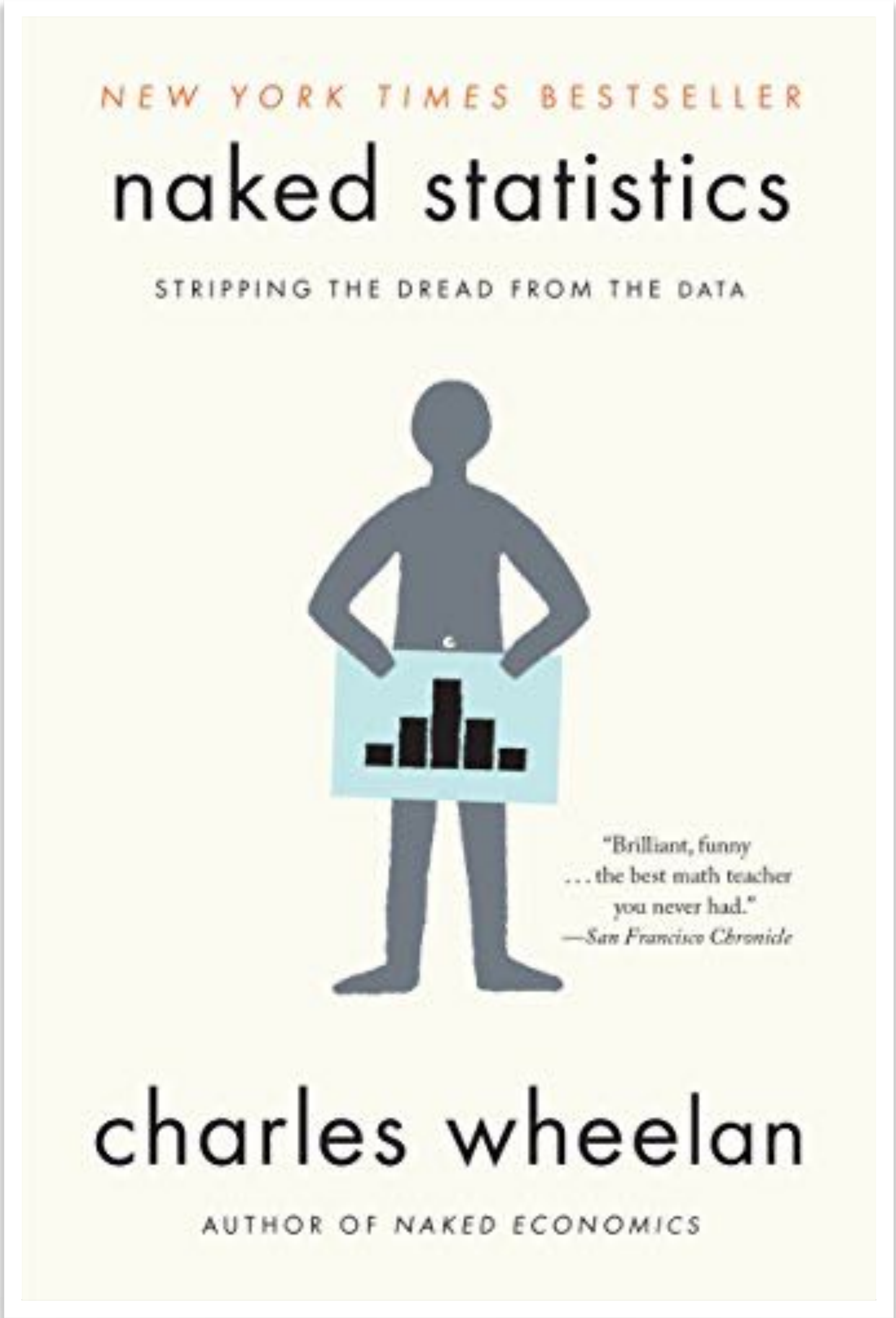
Prediction/ Simulation

Model induction

Numbers may lie

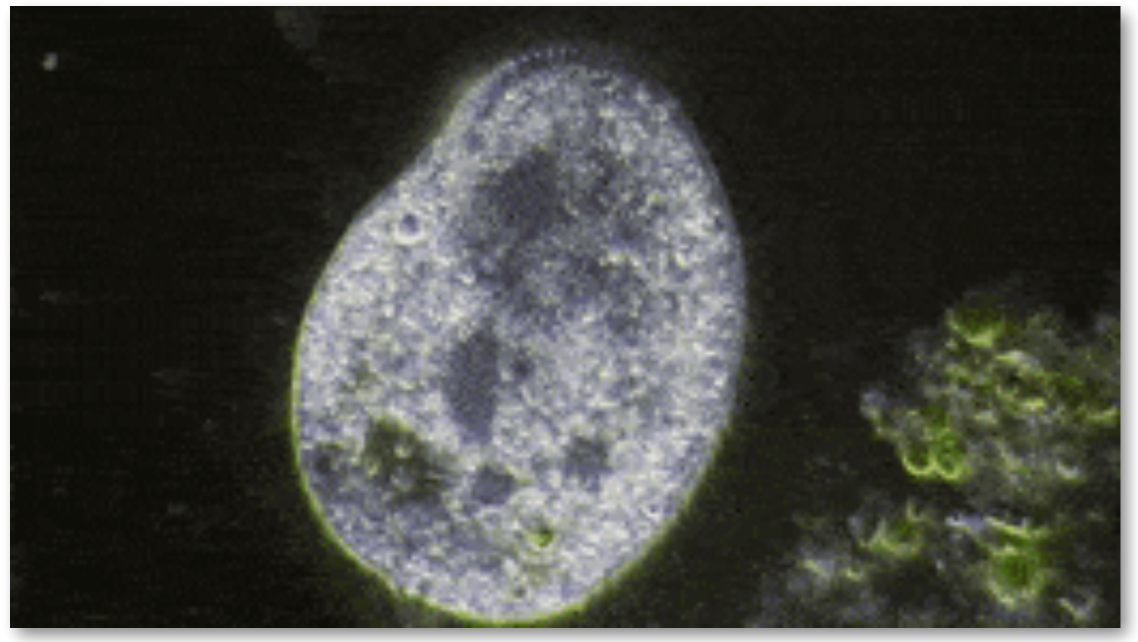






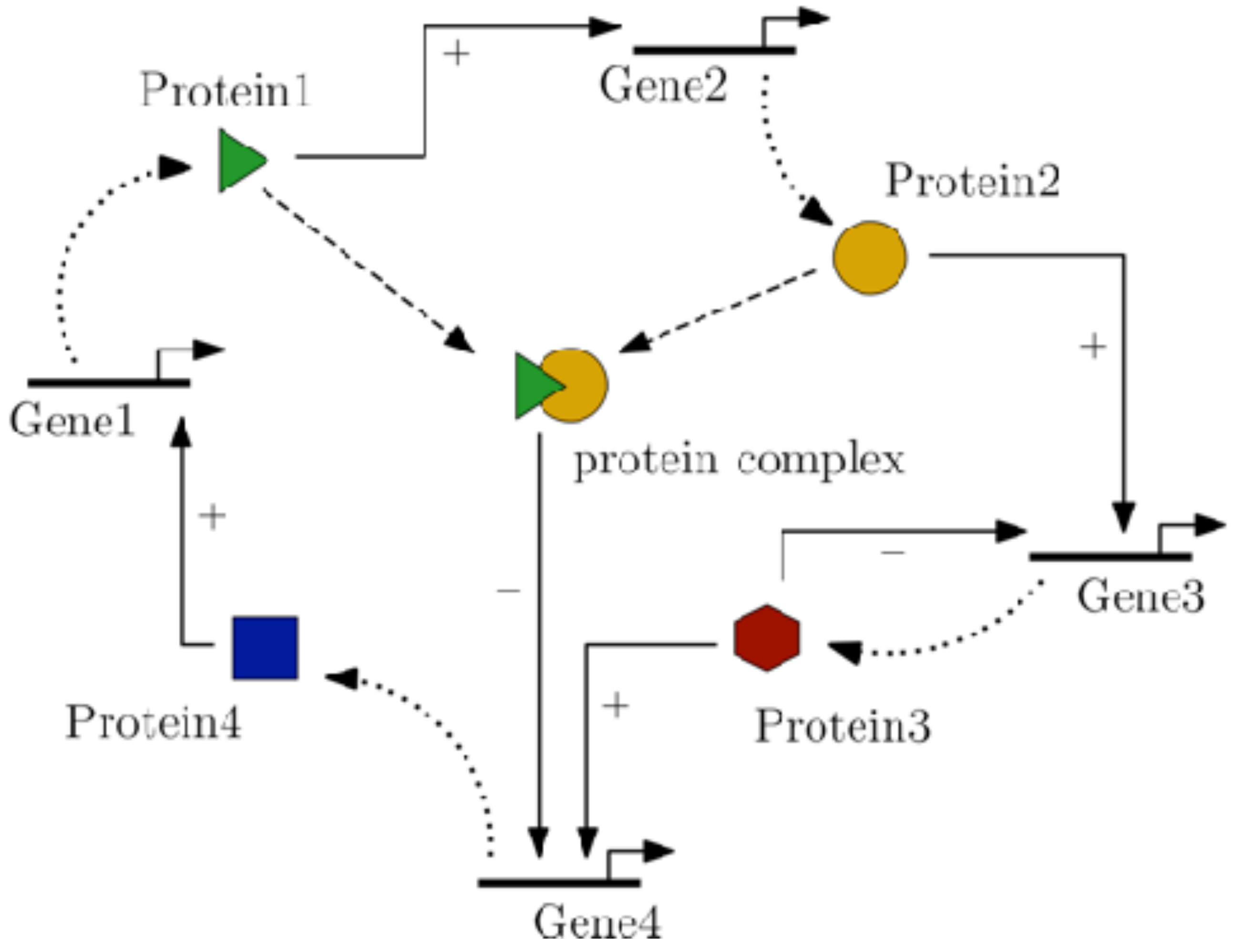
A photograph of the Sun taken at the same time every day for a year will yield the visual pattern seen here, known as analemma

Reverse engineering of gene regulatory nets



Living thing

Reverse Engineering



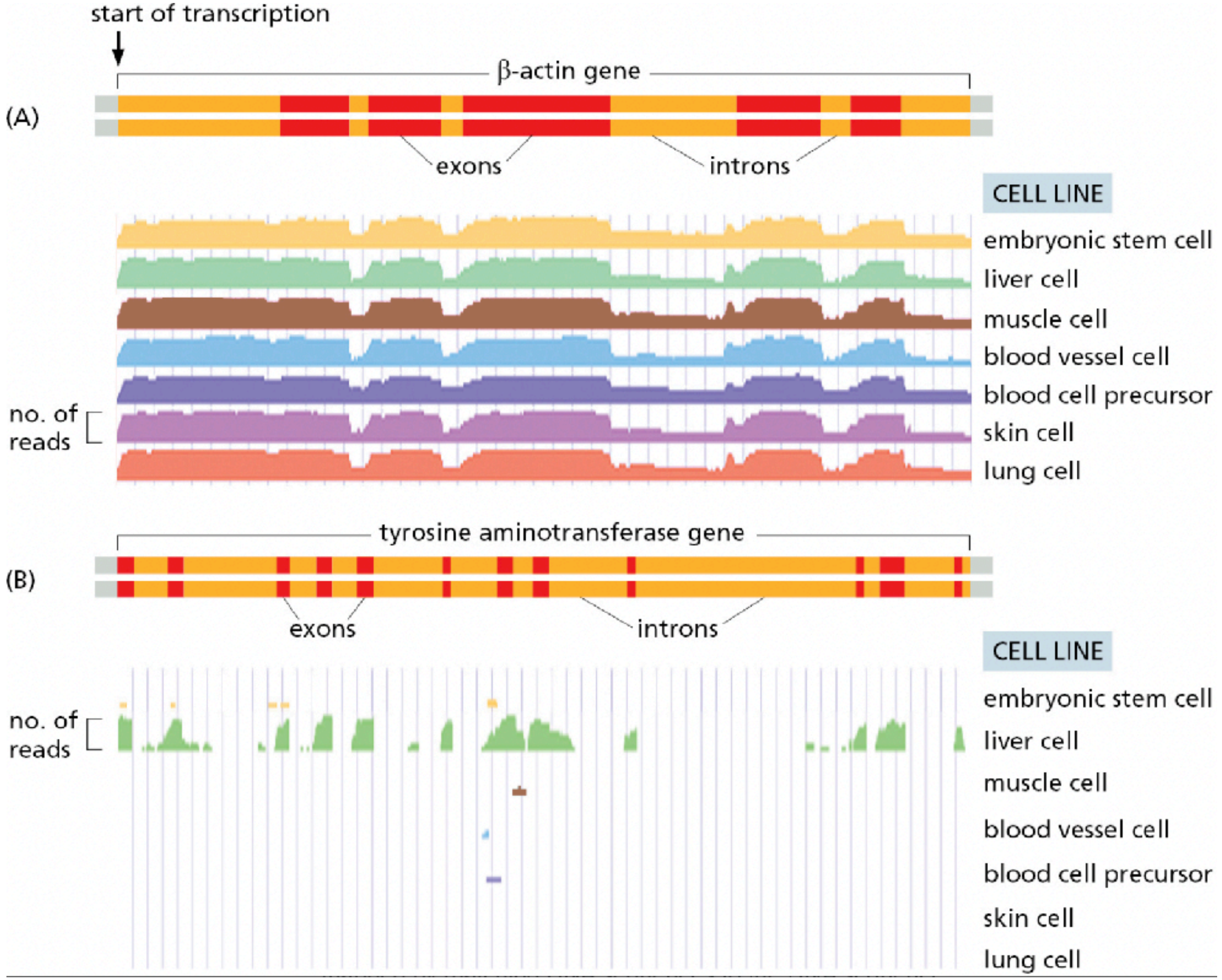
Gene regulatory Model

Reverse engineering of gene regulatory nets

What we know about gene regulation?

Regulation of gene expression

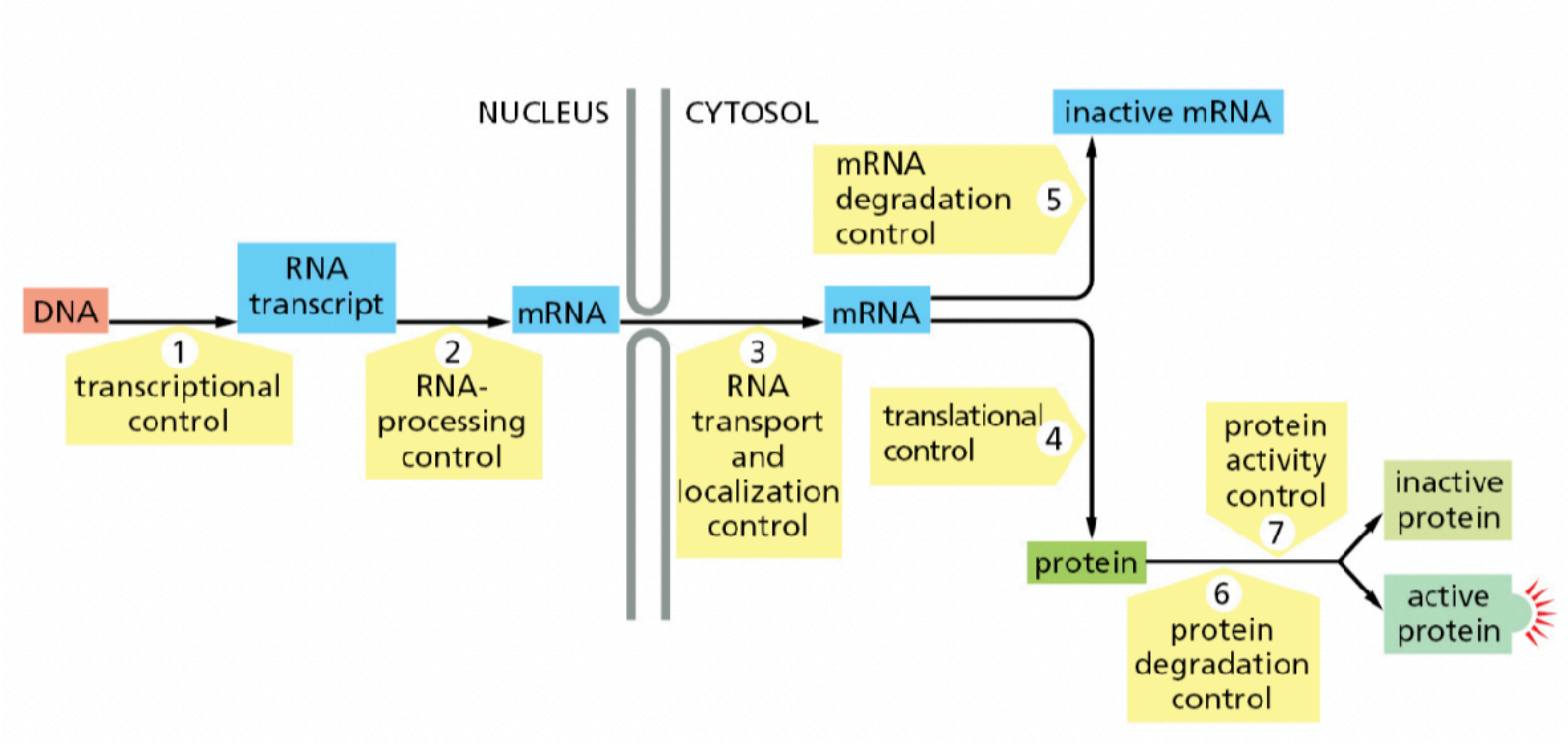
- Cells share the same DNA but different cell types synthesise different sets of RNAs and proteins
- Many processes are common to all cells and many other are specific for each cell type
- Moreover, external signals can cause a cell to change the expression of its genes



S. Djebali et al., *Nature* 489:101-109, 2012

Regulation of gene expression

Gene Expression Can Be Regulated at Many of the Steps in the Pathway from DNA to RNA to Protein. How can this occur?

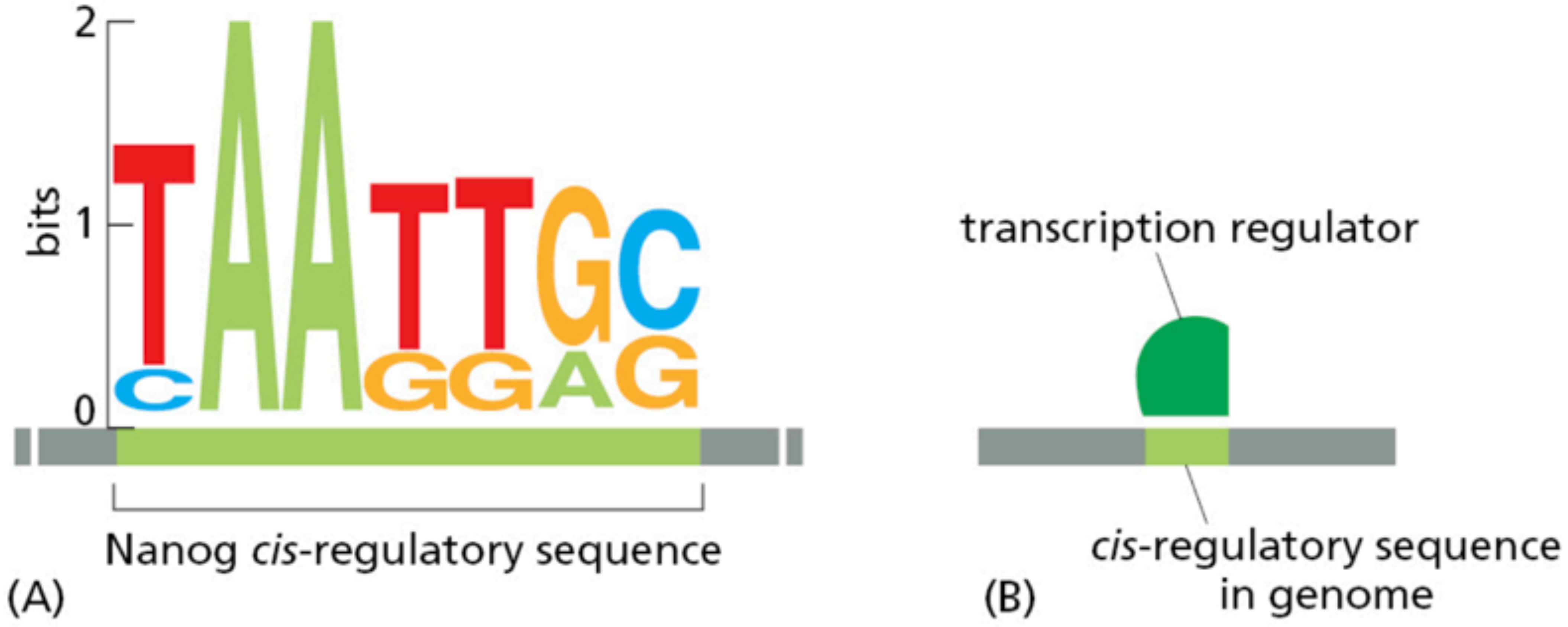


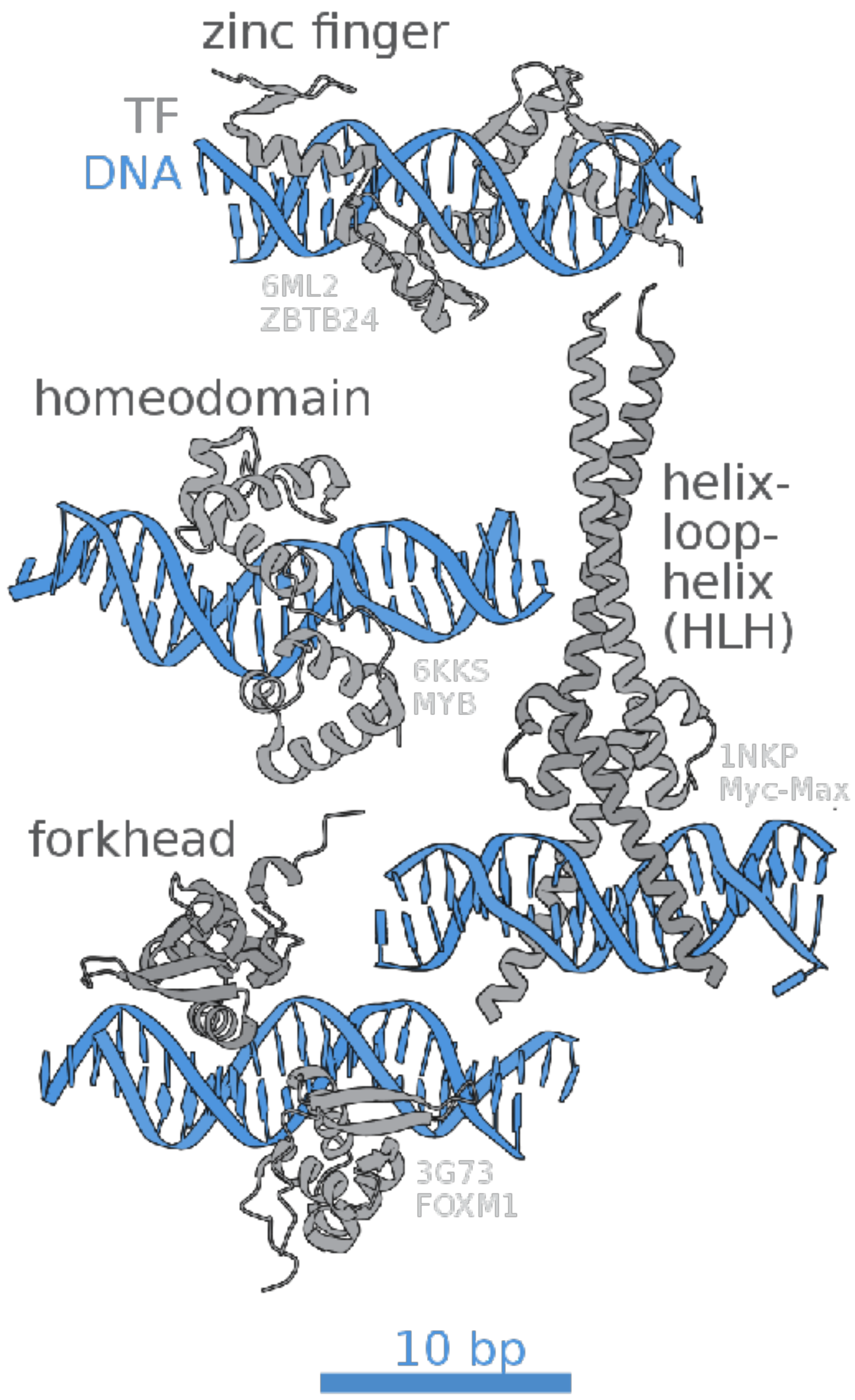
Transcriptional control

Transcription regulators

Recognize specific sequences of DNA (typically 5-12 nucleotides)

Approximately 10% of protein coding genes are devoted to transcription regulators

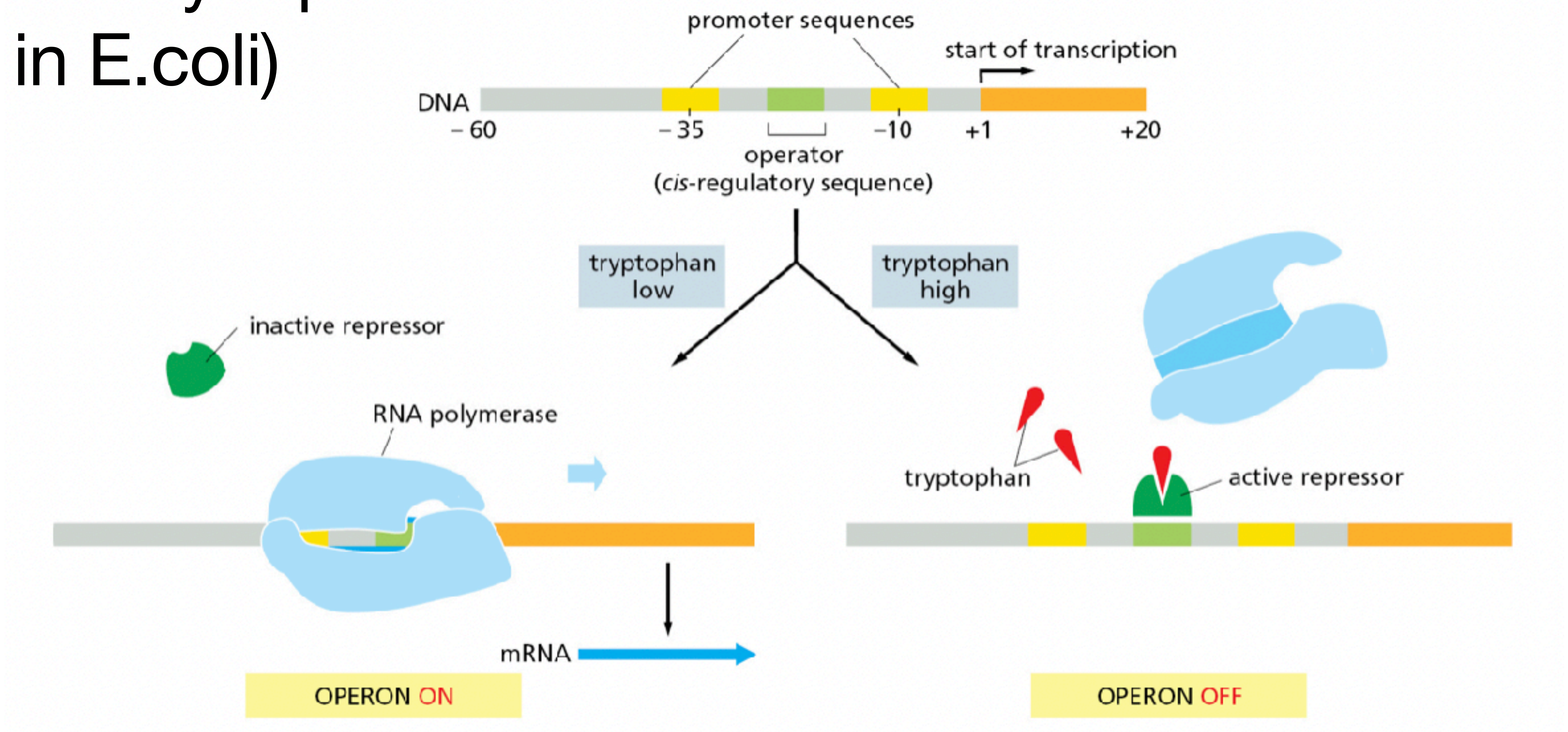




Transcriptional control

Transcription regulators

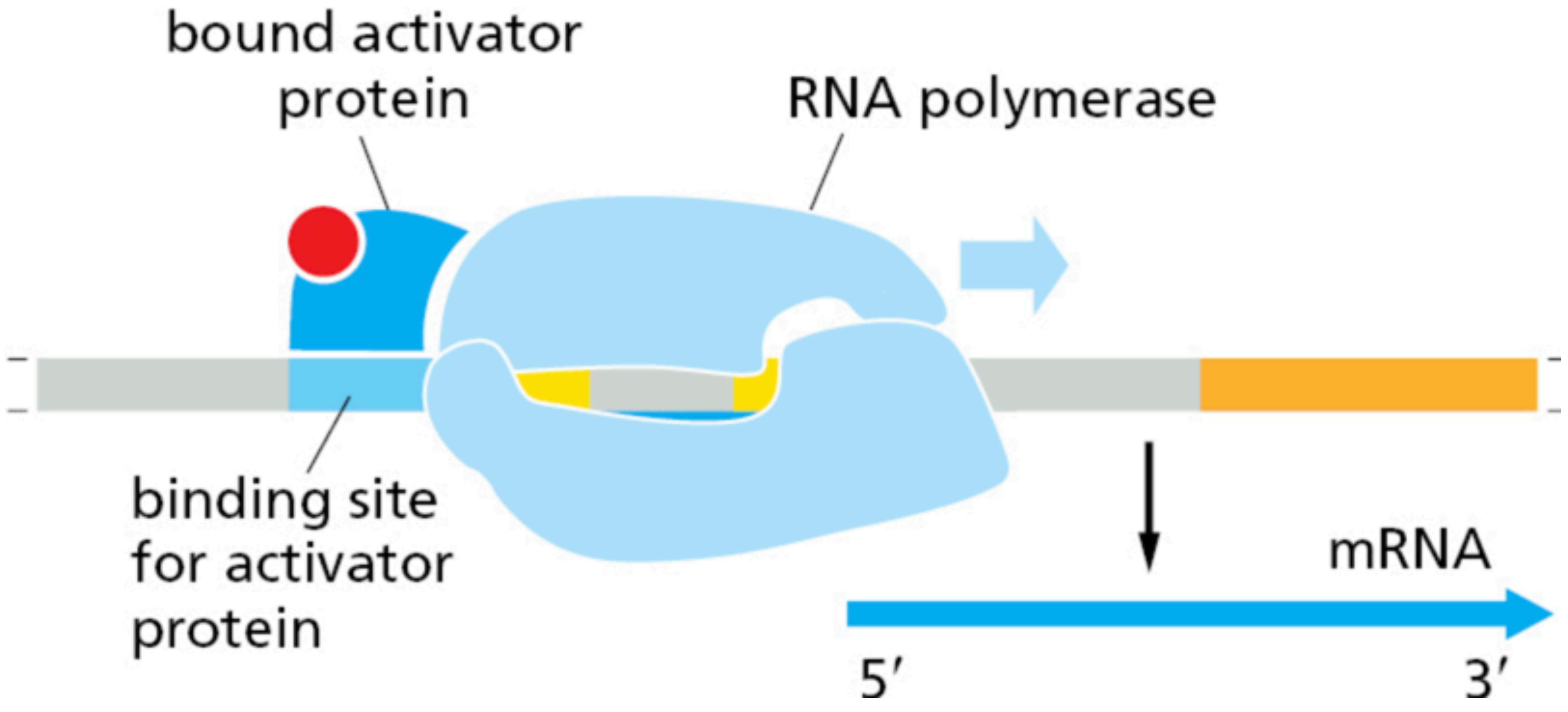
Genes can be switched OFF by repressor proteins (e.g. tryptophan in E.coli)



Transcriptional control

Transcription regulators

Genes can be switched ON by activator proteins

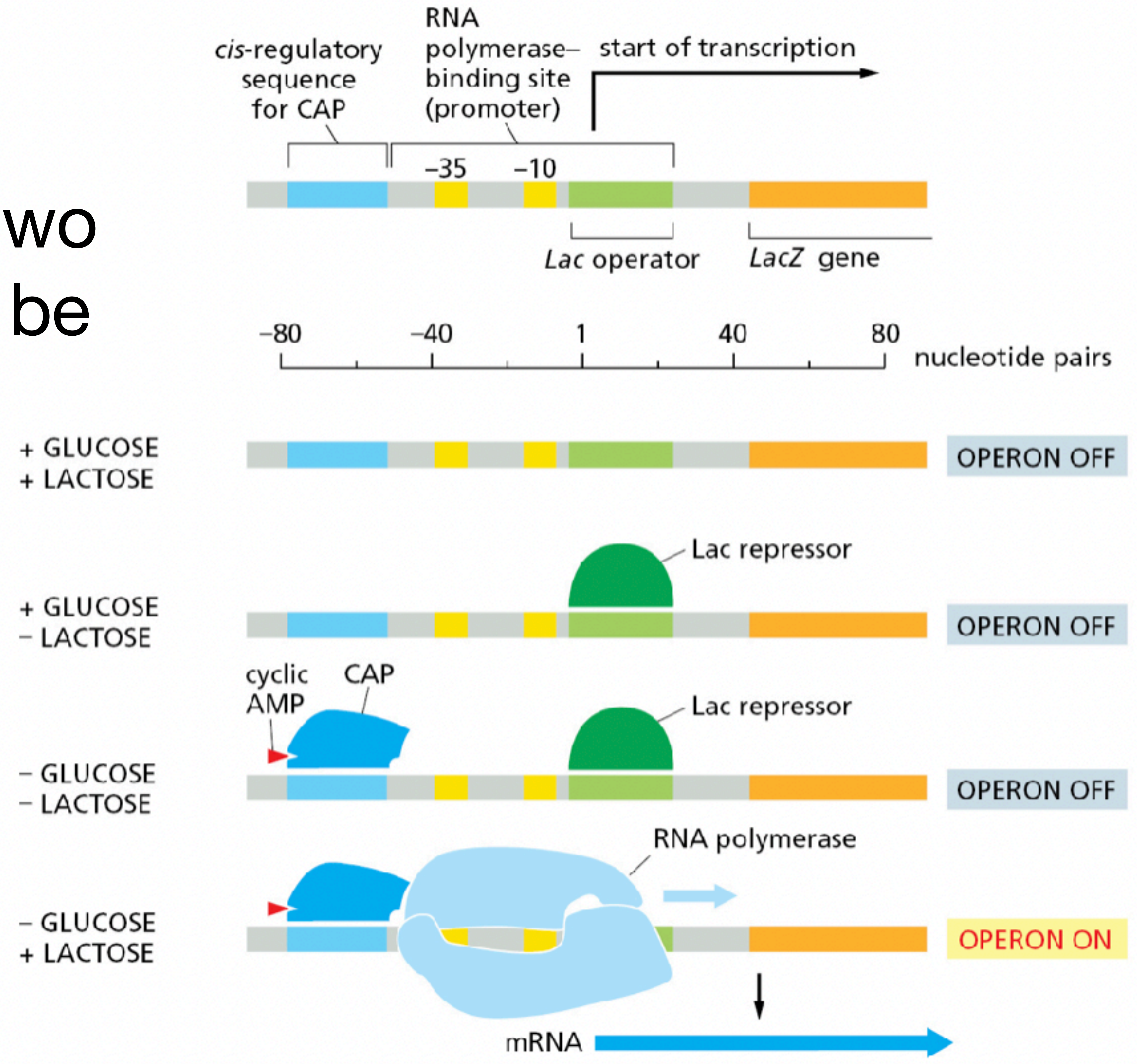


Transcriptional control

Transcription regulators

Lac operon in E.coli is controlled by two transcription regulators, causing it to be expressed only when needed,

i.e. lactose present AND glucose absent

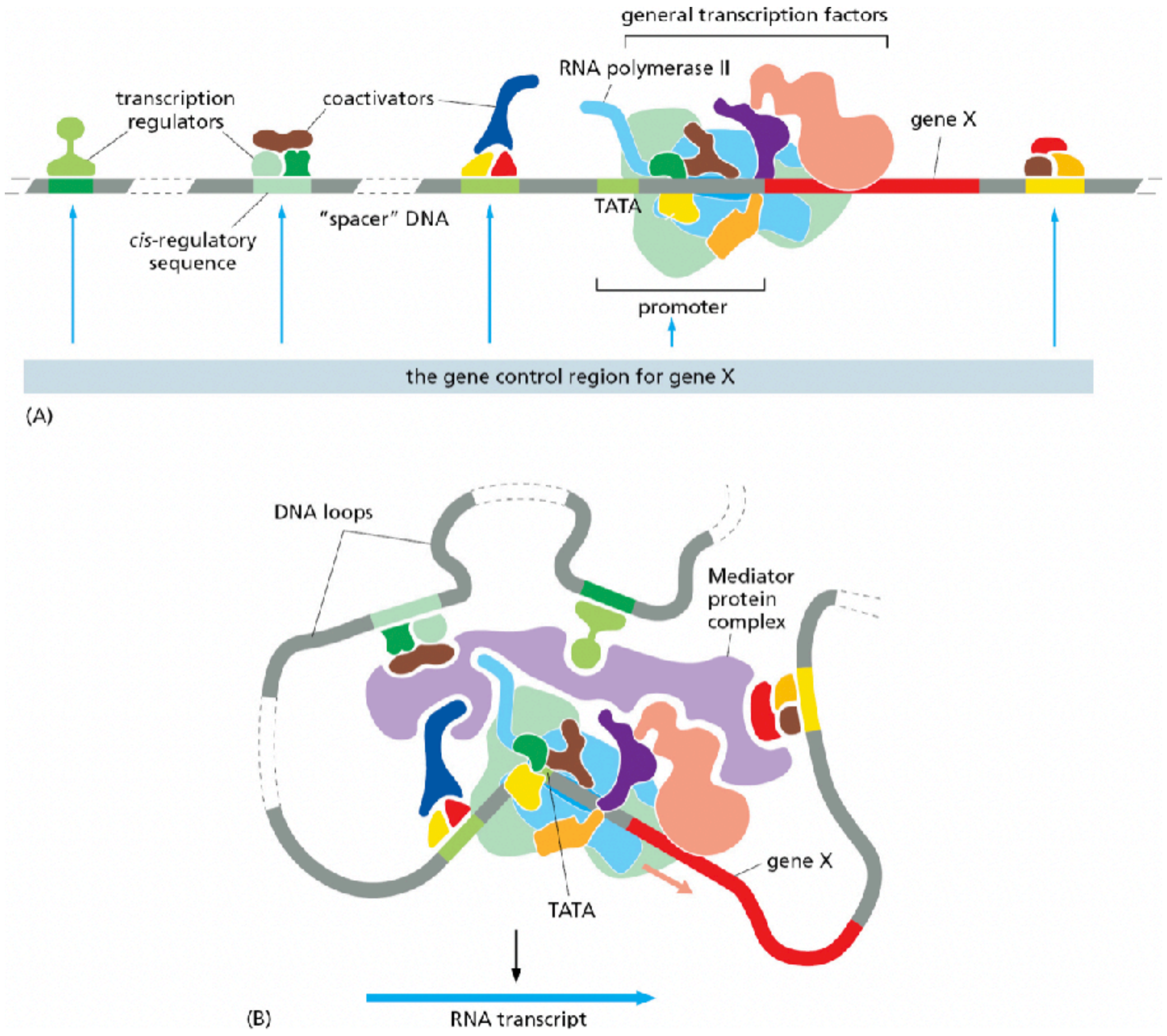


Transcriptional control

Transcription regulators

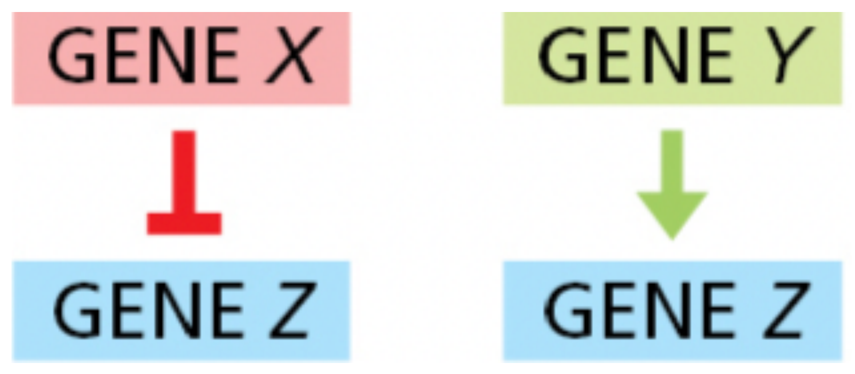
Eukaryotic gene control region includes many cis-regulatory sequences allowing transcription regulators to work in groups as co-activators and co-repressors.

Co-activators an co-repressors can acts in a variety numbers of ways



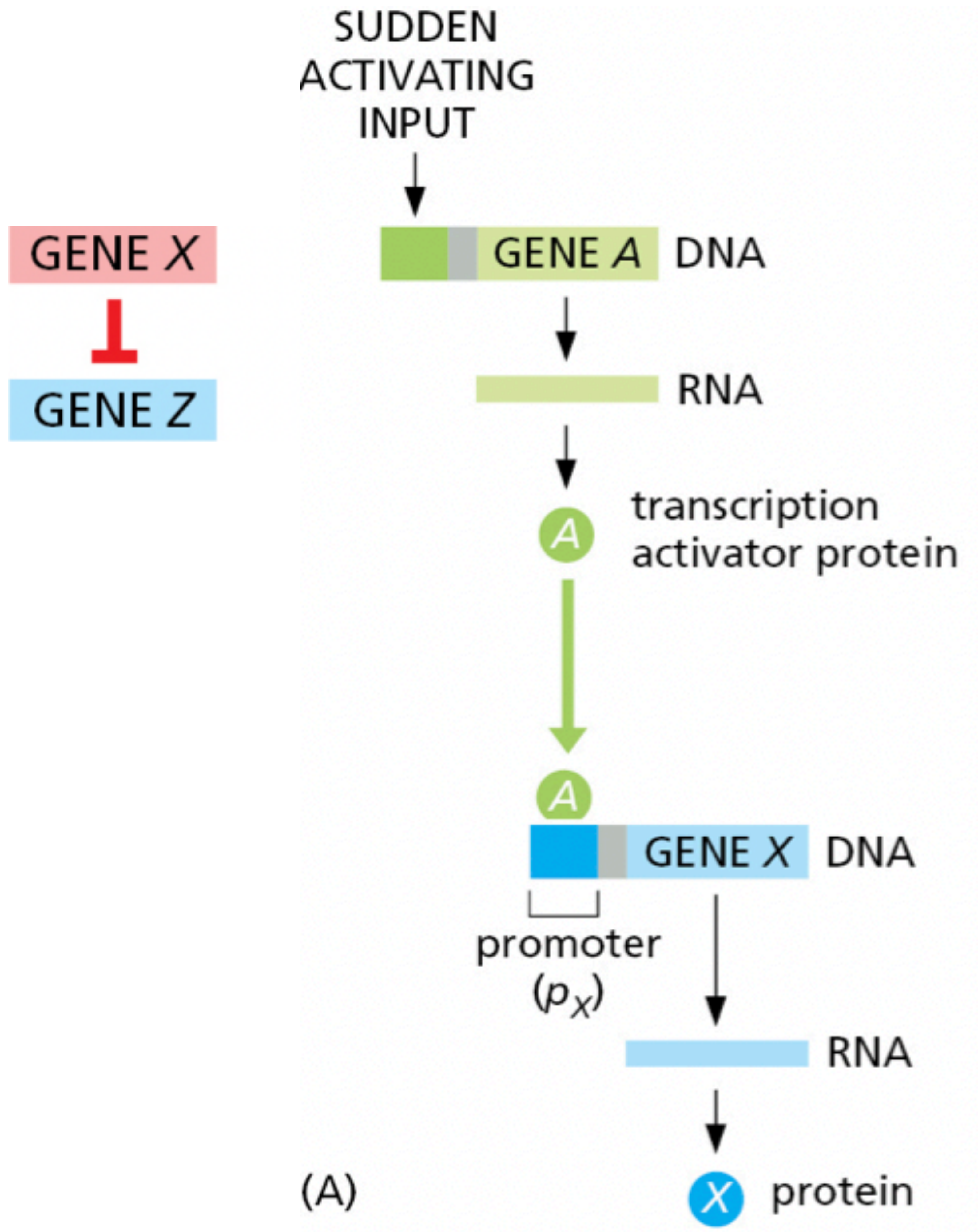
Mathematically analysis of cell function

Summary of biochemical relationships



Mathematically analysis of cell function

Summary of biochemical relationships

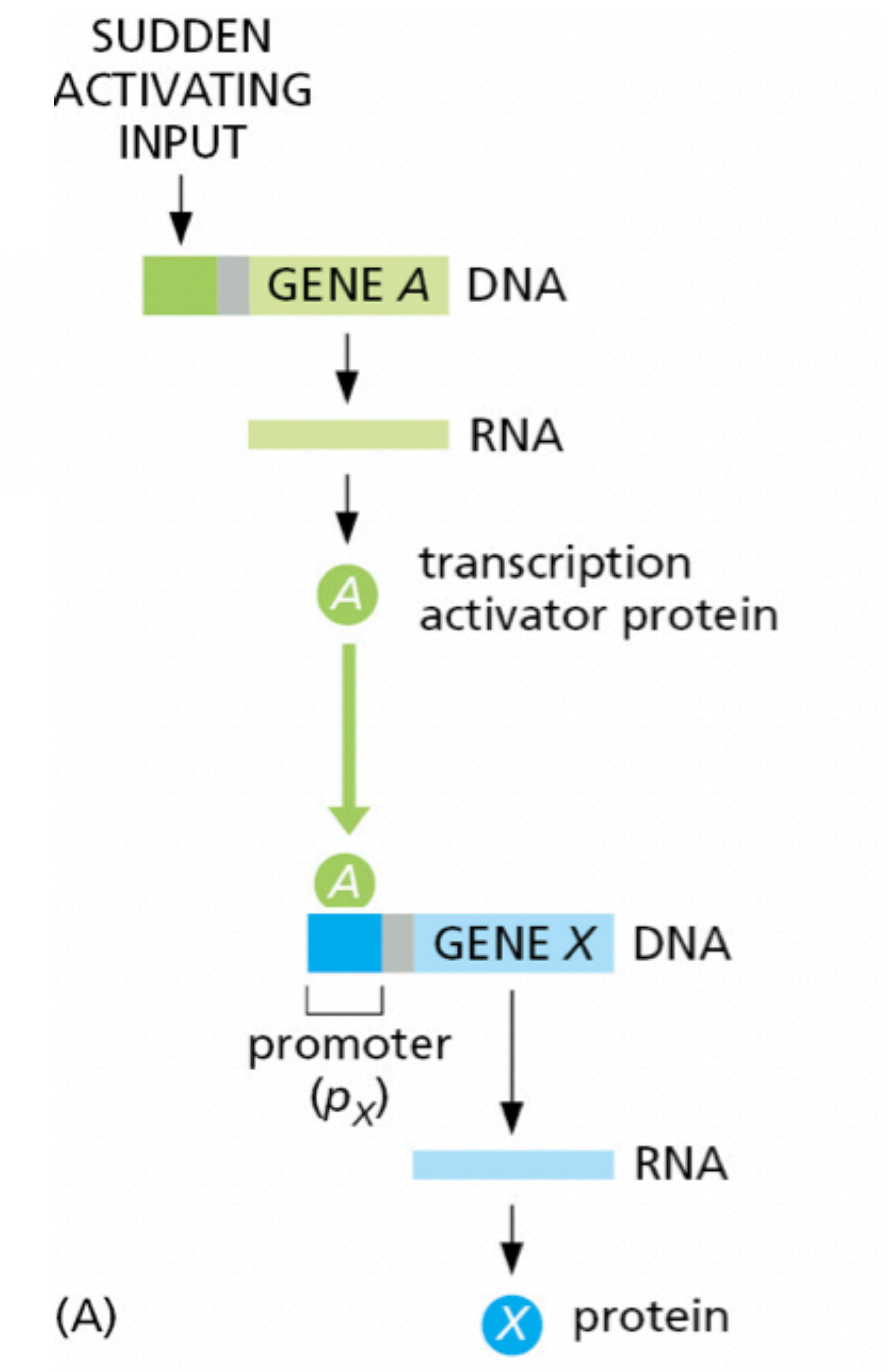


Mathematically analysis of cell function

Differential equation model

$$\frac{d[X]}{dt} = \textit{protein production rate} - \textit{protein degradation rate}$$

$$\frac{d[X]}{dt} = \beta \cdot m \frac{K[A]}{1 + K[A]} - \frac{[X]}{\tau_X}$$ **Equation 8–5**



Mathematically analysis of cell function

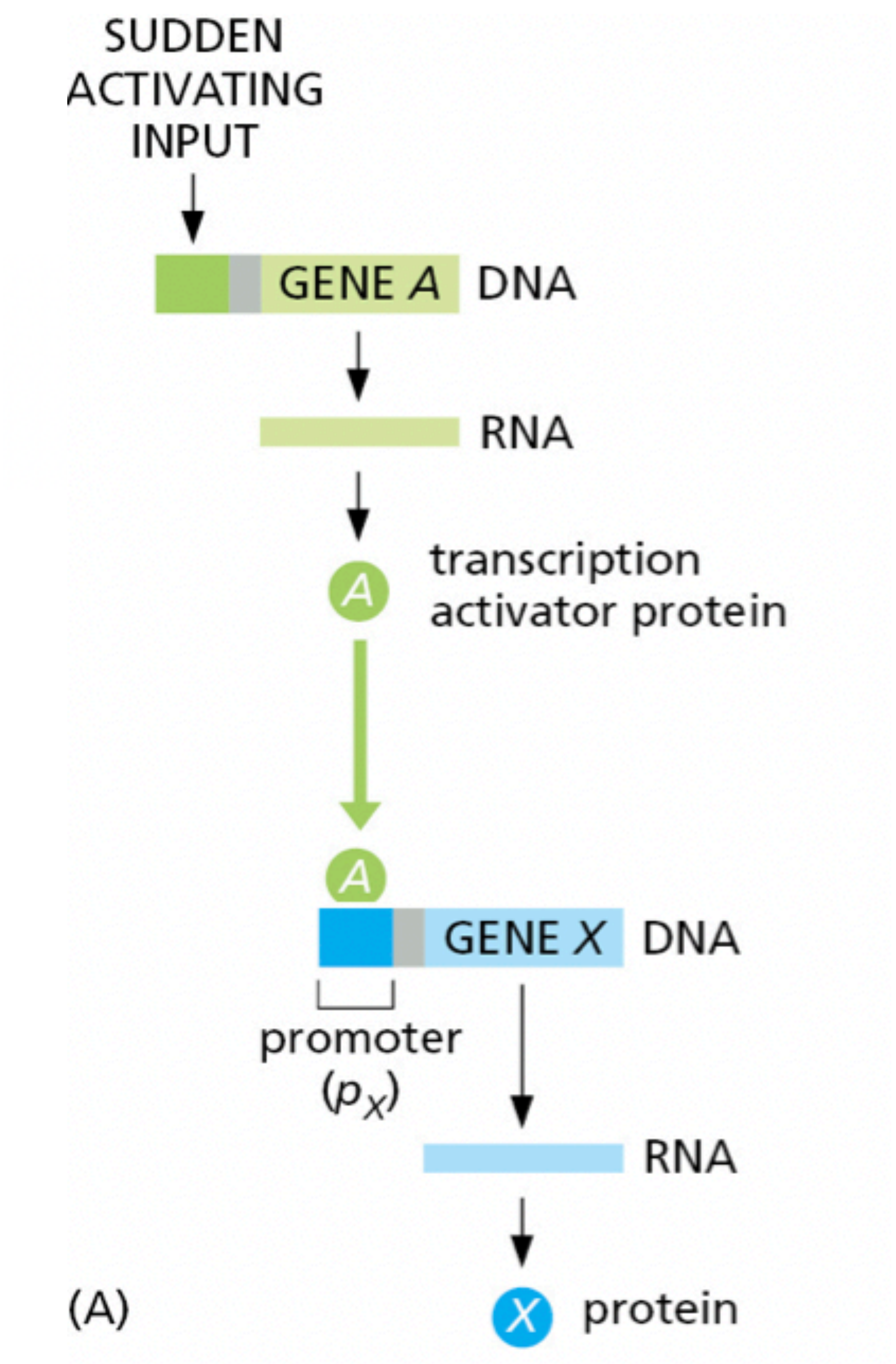
Differential equation model

$$\frac{d[X]}{dt} = \textit{protein production rate} - \textit{protein degradation rate}$$

$$\frac{d[X]}{dt} = \beta \cdot m \frac{K[A]}{1 + K[A]} - \frac{[X]}{\tau_X}$$

Equation 8–5

$$[X](t) = [X_{st}](1 - \exp(-t/\tau_X))$$



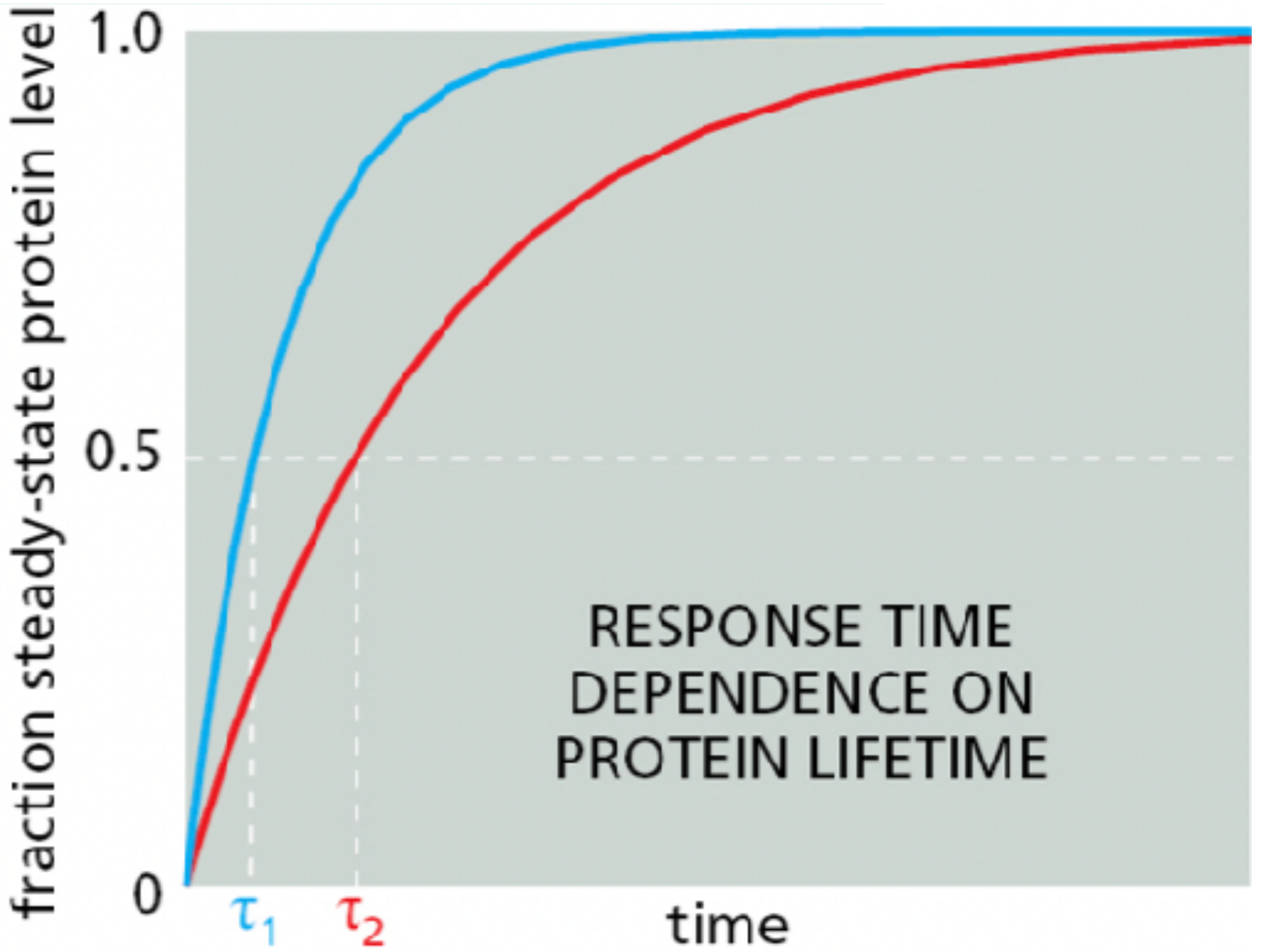
Mathematically analysis of cell function

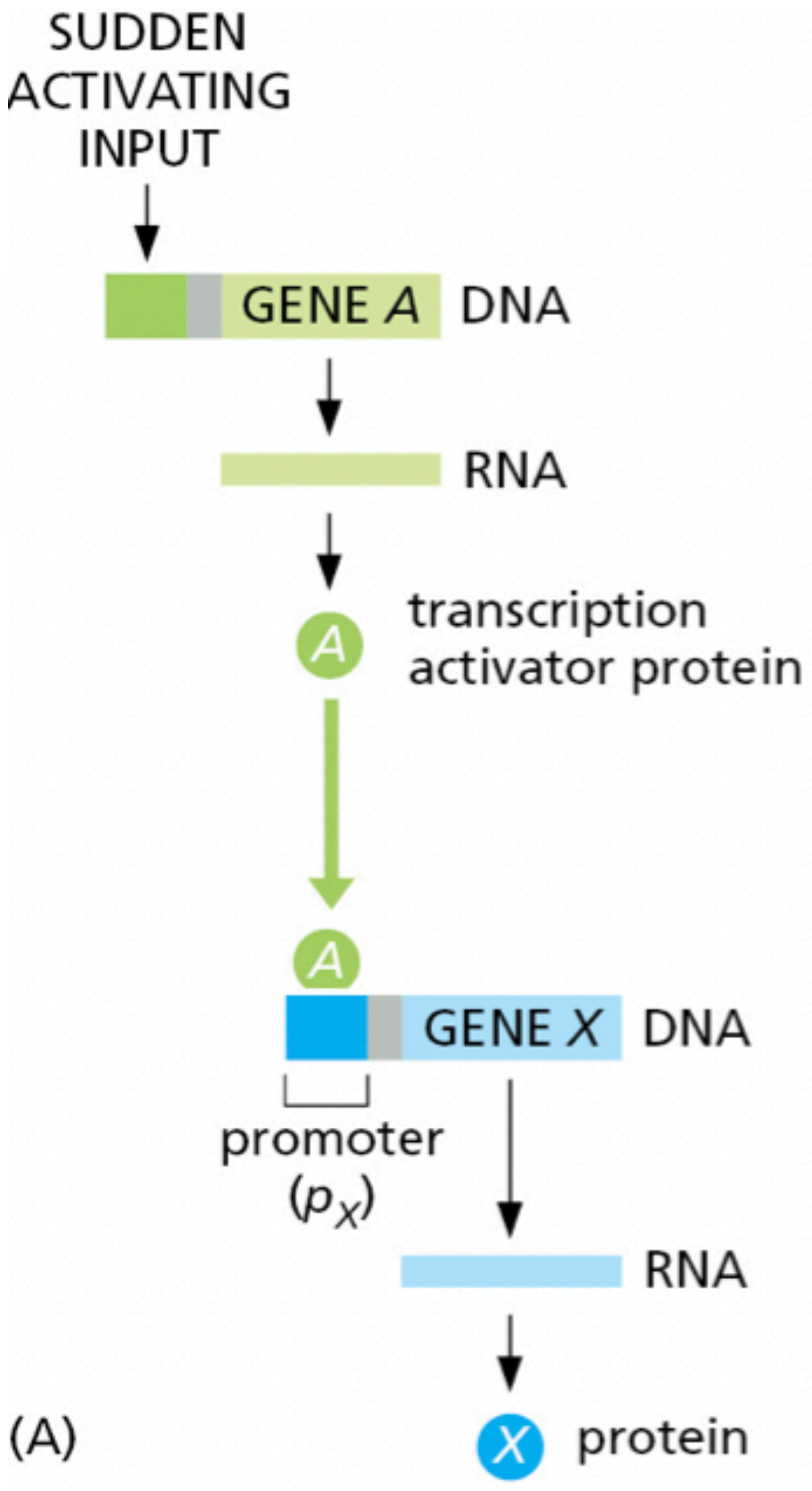
Differential equation model

$$\frac{d[X]}{dt} = \textit{protein production rate} - \textit{protein degradation rate}$$

$$\frac{d[X]}{dt} = \beta \cdot m \frac{K[A]}{1 + K[A]} - \frac{[X]}{\tau_X}$$ **Equation 8–5**

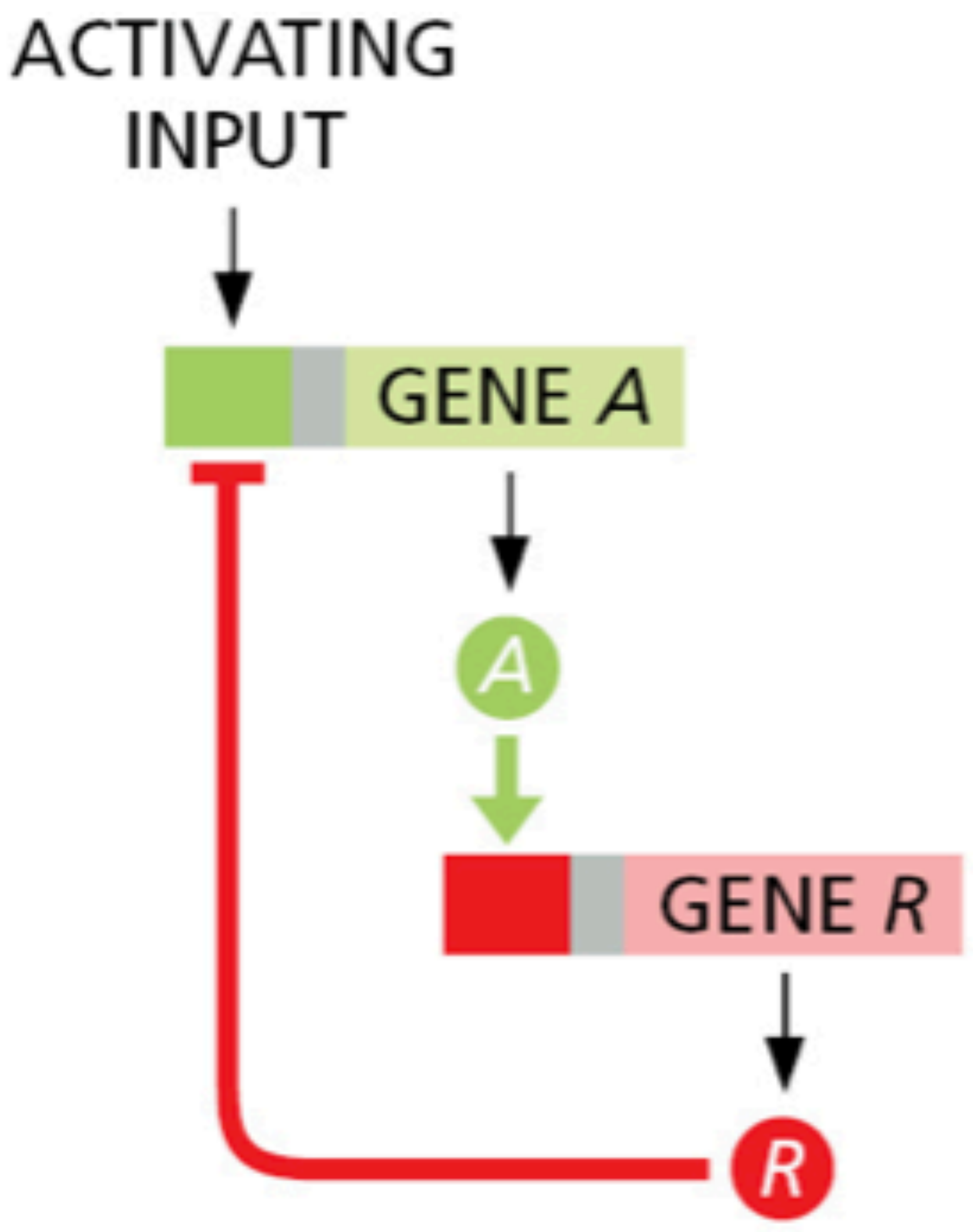
$$[X](t) = [X_{st}](1 - \exp(-t/\tau_X))$$





Mathematically analysis of cell function

Negative feedback is a powerful strategy in Cell Regulation



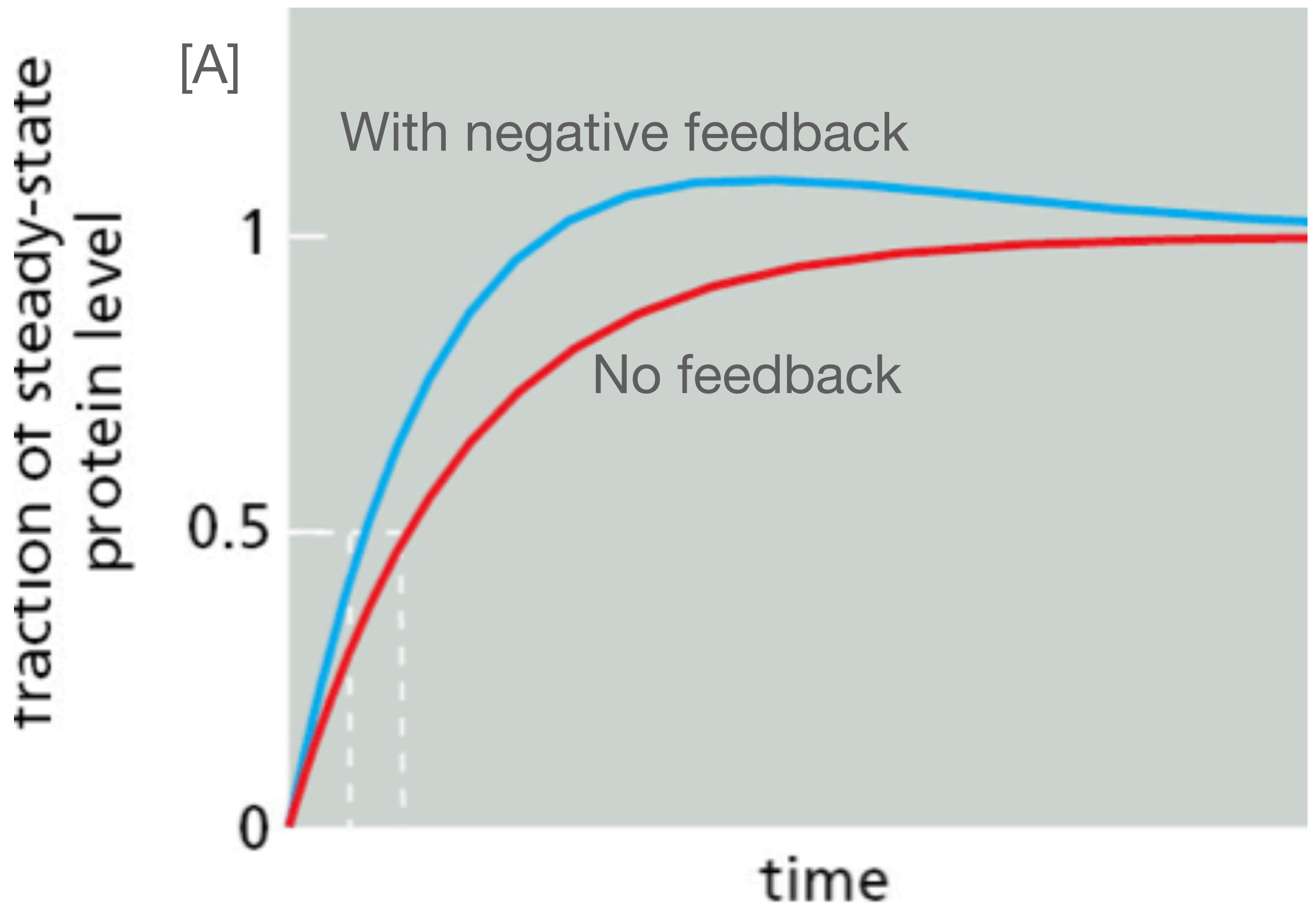
Mathematically analysis of cell function

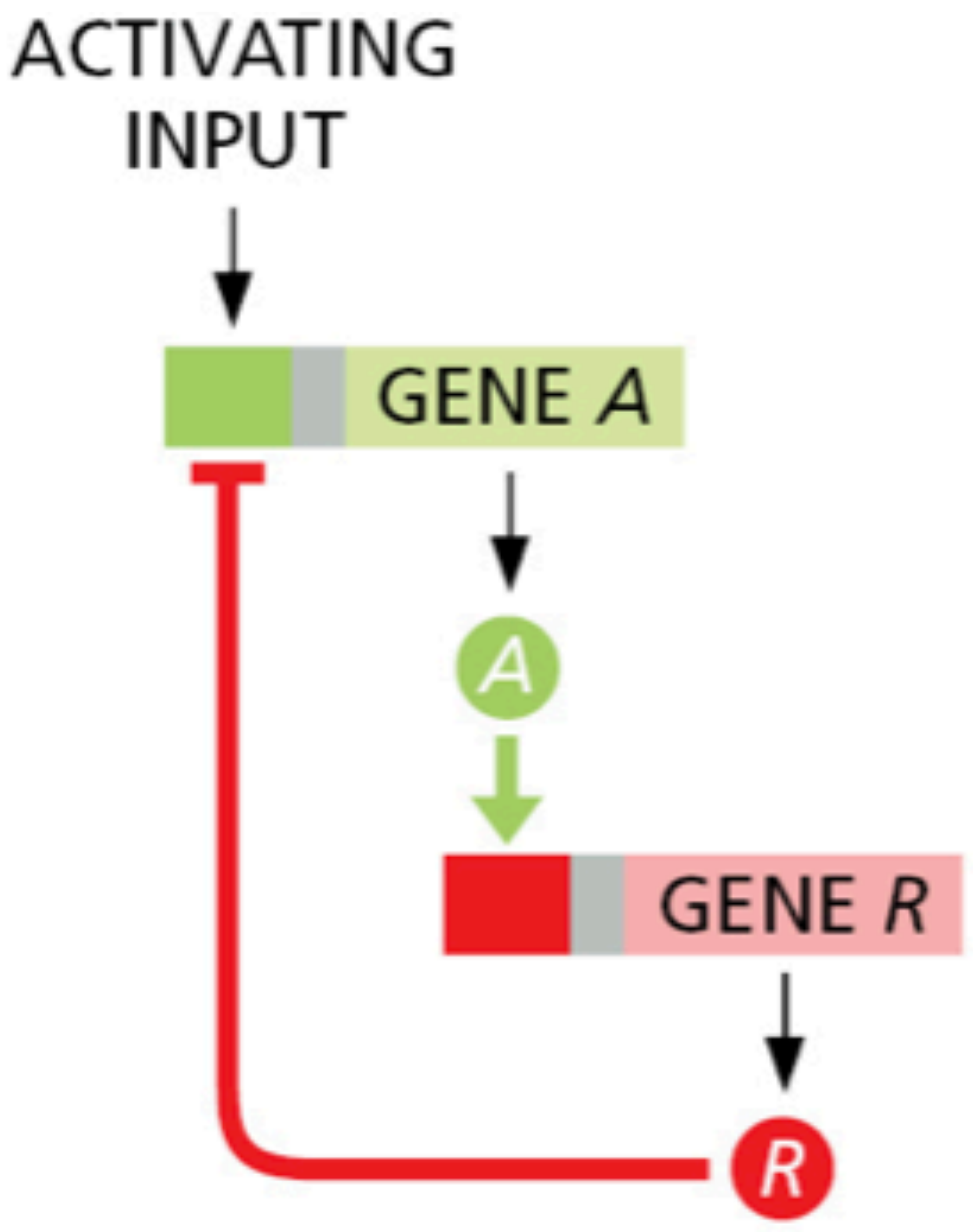
Negative feedback is a powerful strategy in Cell Regulation

$$\frac{d[A]}{dt} = \frac{\beta_A \cdot m_A}{1 + K_R[R]} - \frac{[A]}{\tau_A}$$

$$\frac{d[R]}{dt} = \beta_R \cdot m_R \frac{K_A[A]}{1 + K_A[A]} - \frac{[R]}{\tau_R}$$

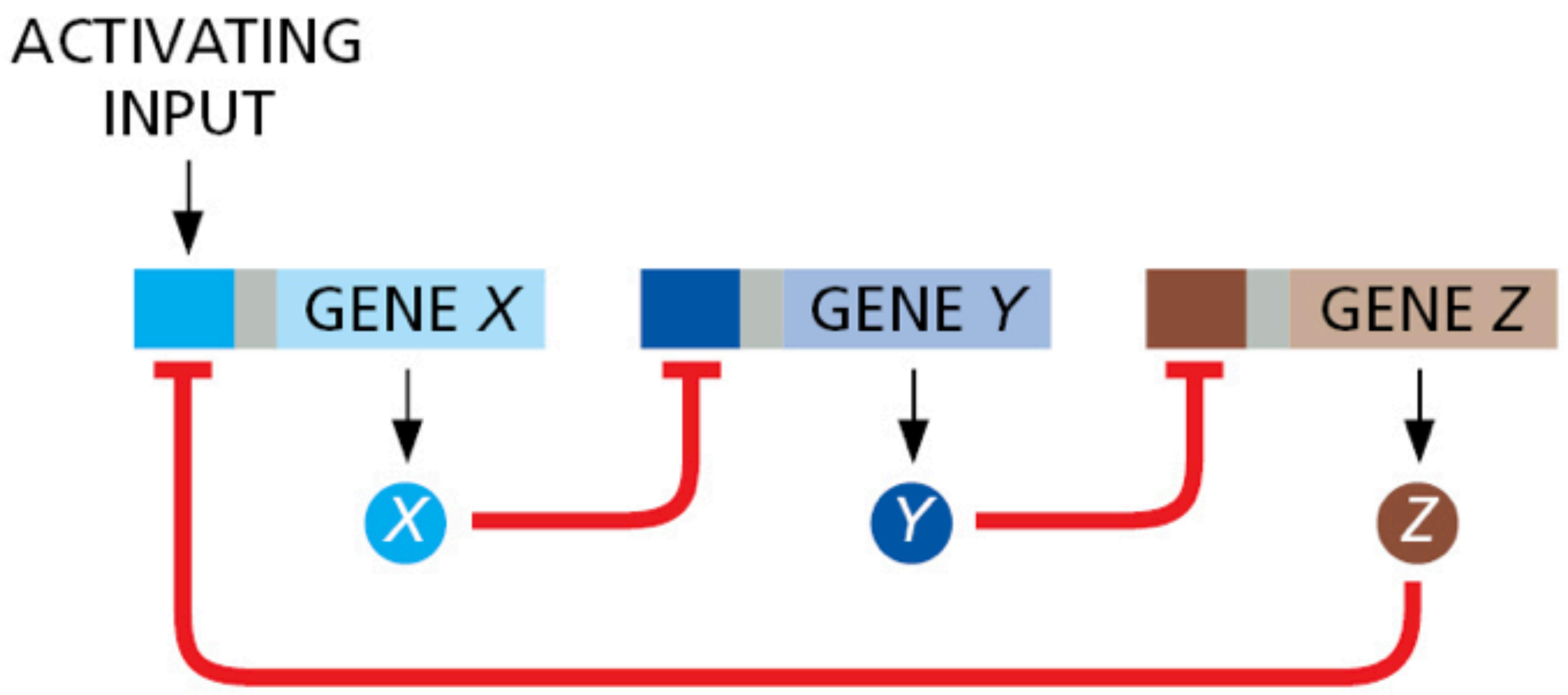
Equation set 8–8





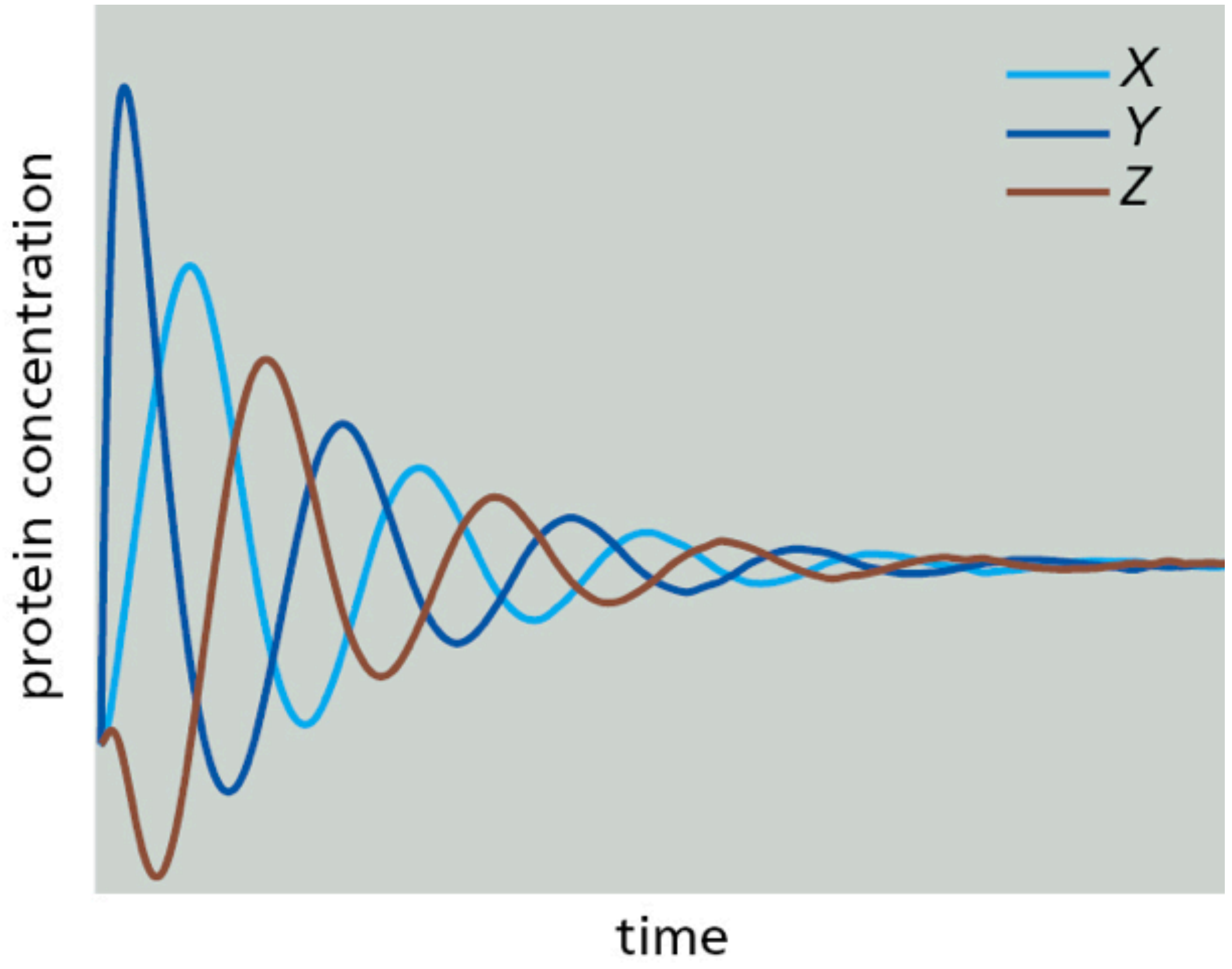
Mathematically analysis of cell function

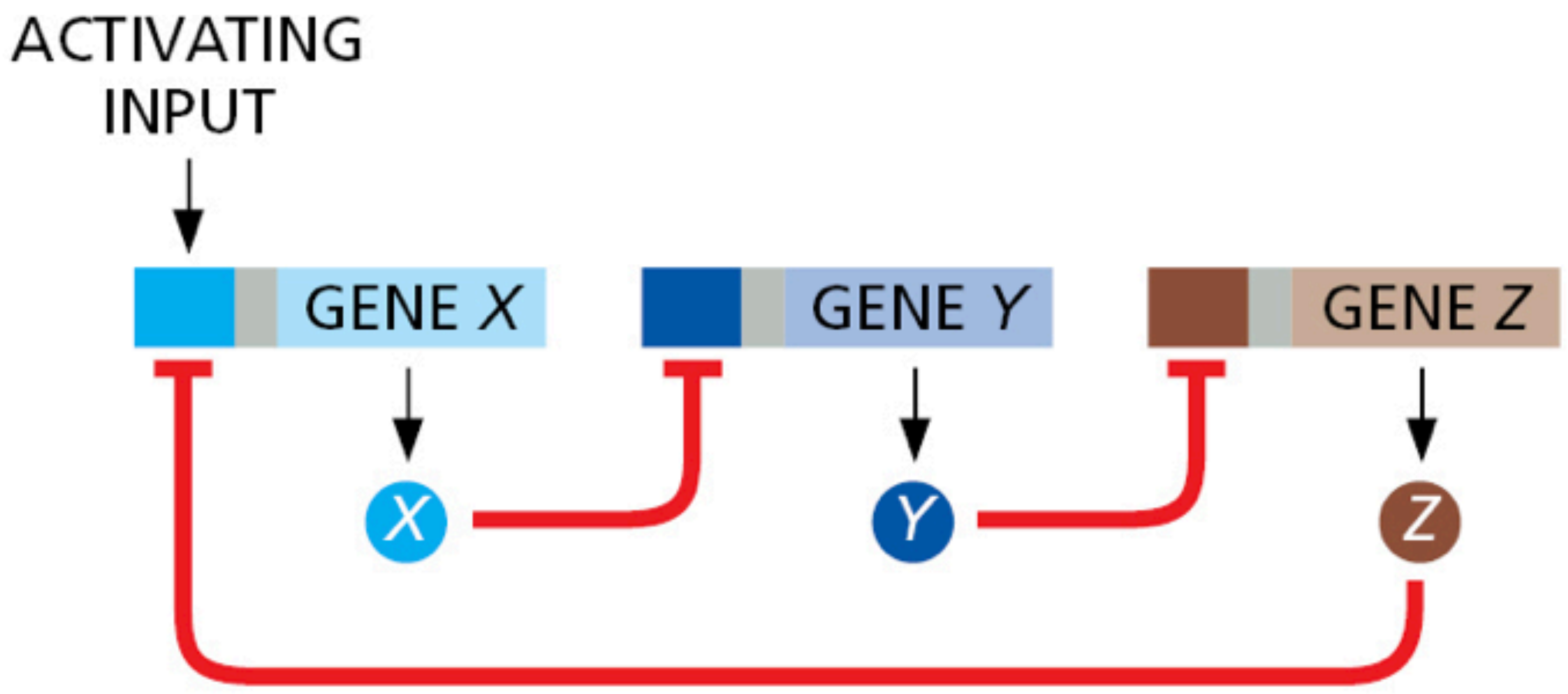
Delayed negative feedback can induce oscillations



Mathematically analysis of cell function

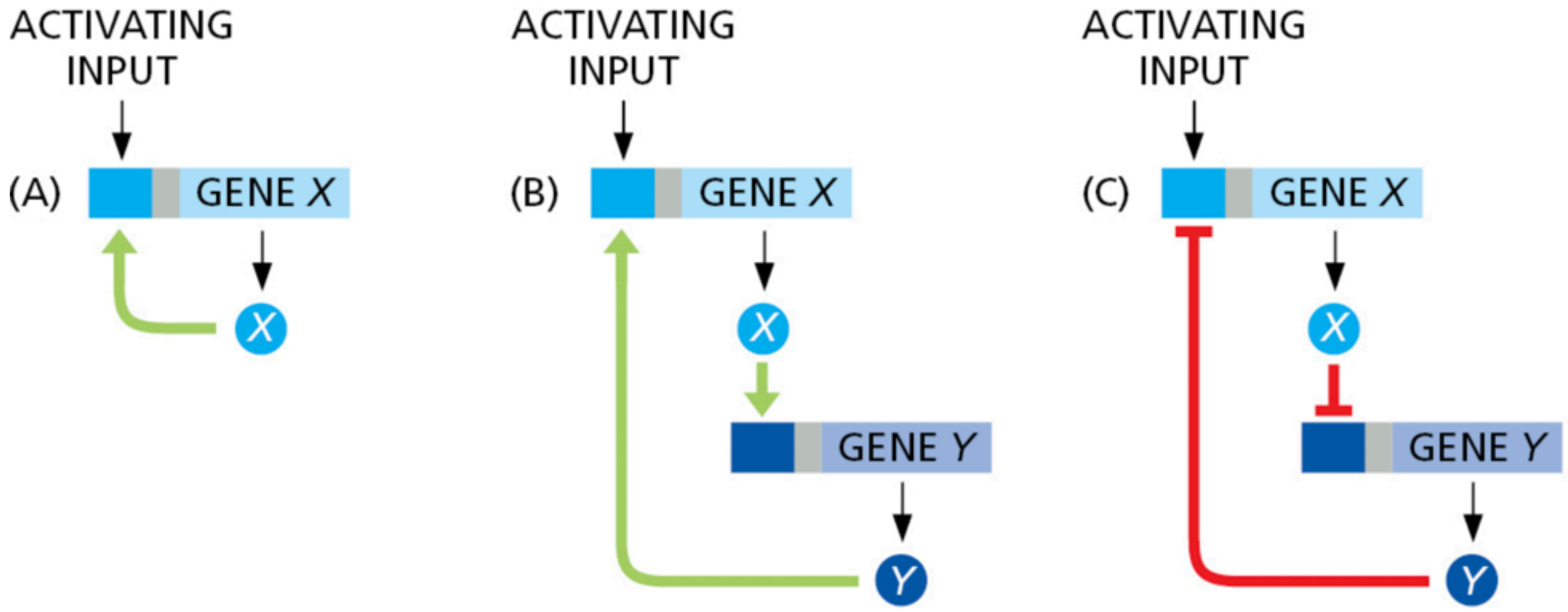
Delayed negative feedback can induce oscillations





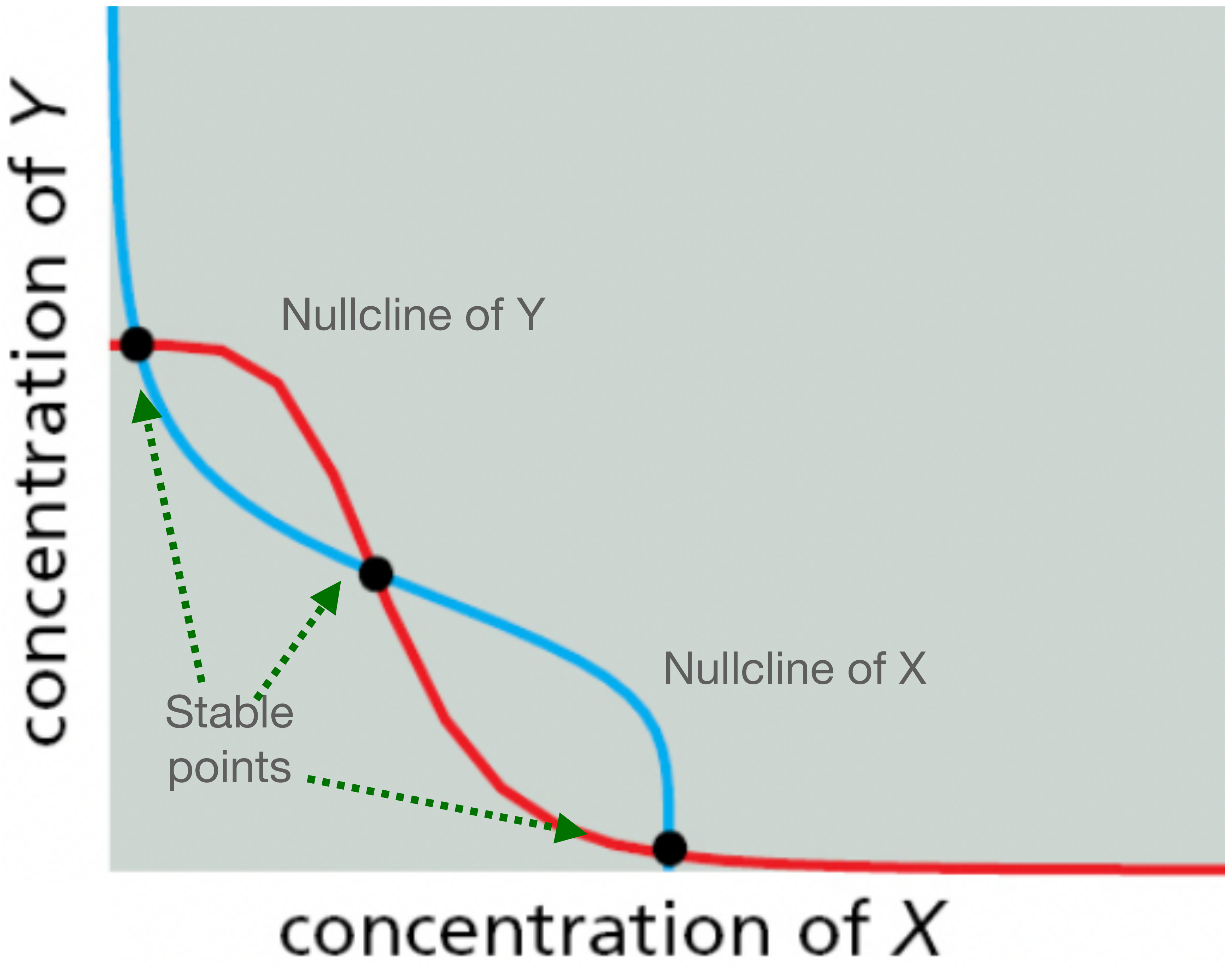
Mathematically analysis of cell function

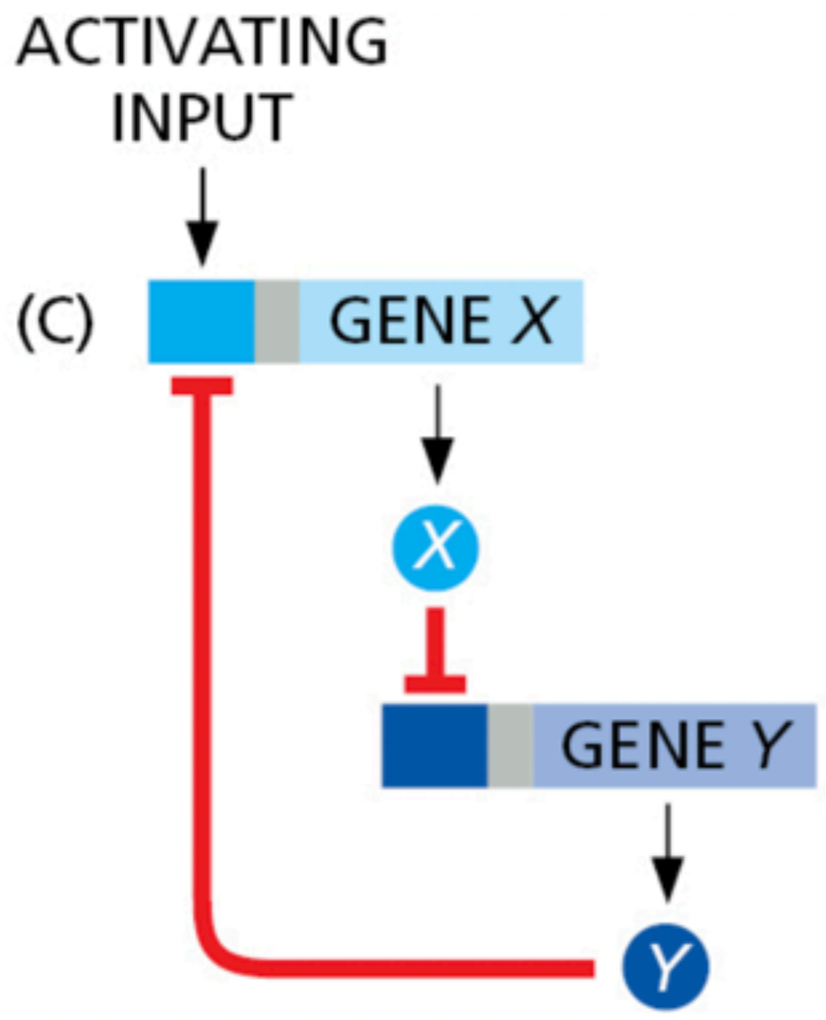
Positive feedback is important for switch-like responses and bistability



Mathematically analysis of cell function

Positive feedback is important for switch-like responses and bistability



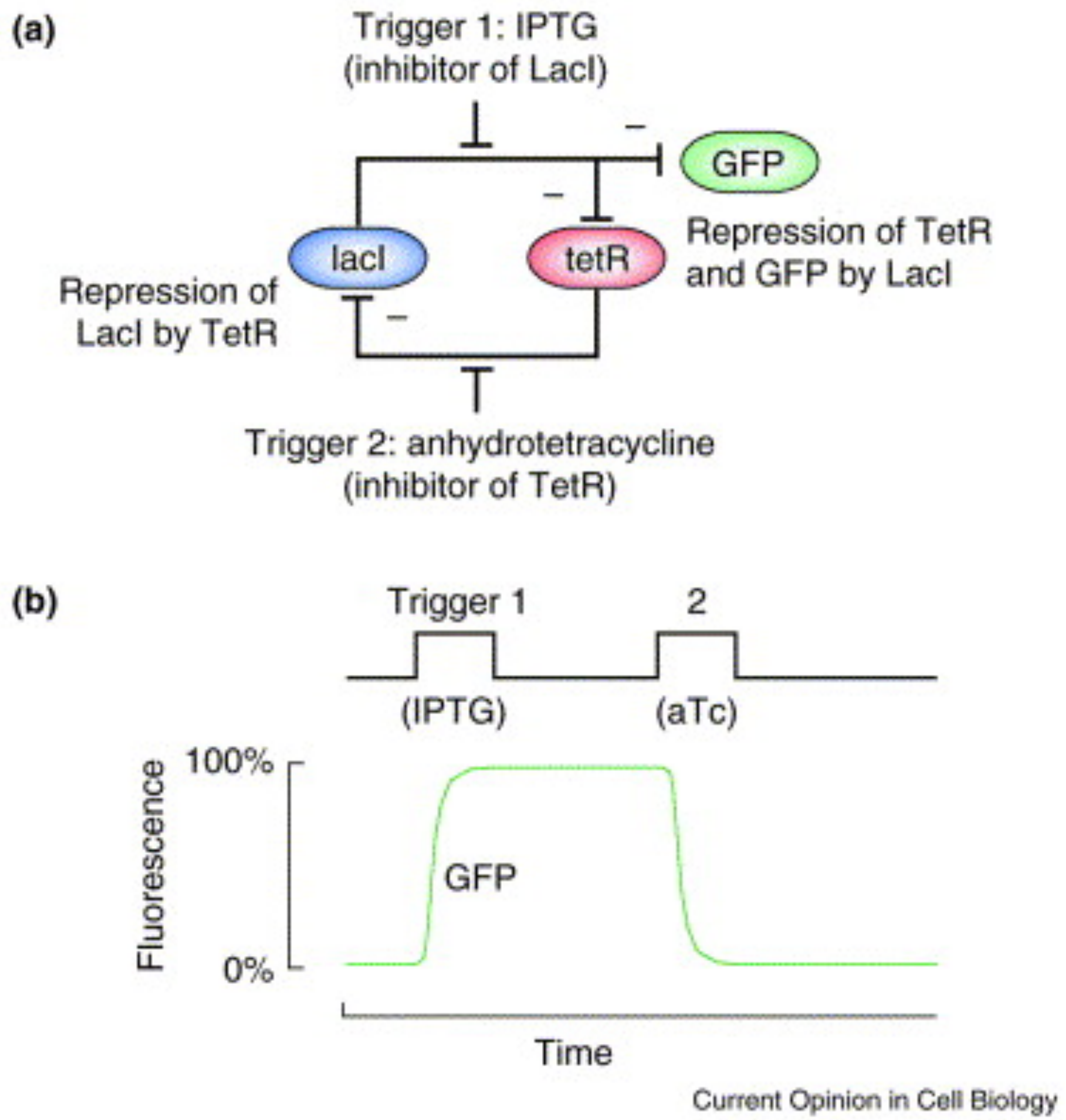


Mathematically analysis of cell function

E.g. an artificial bistable system in E. coli. (Gardner et al., Nature, 2000)

LacI represses the expression of TetR (and GFP, used as a reporter of the status of tetR transcription), and TetR represses the expression of LacI

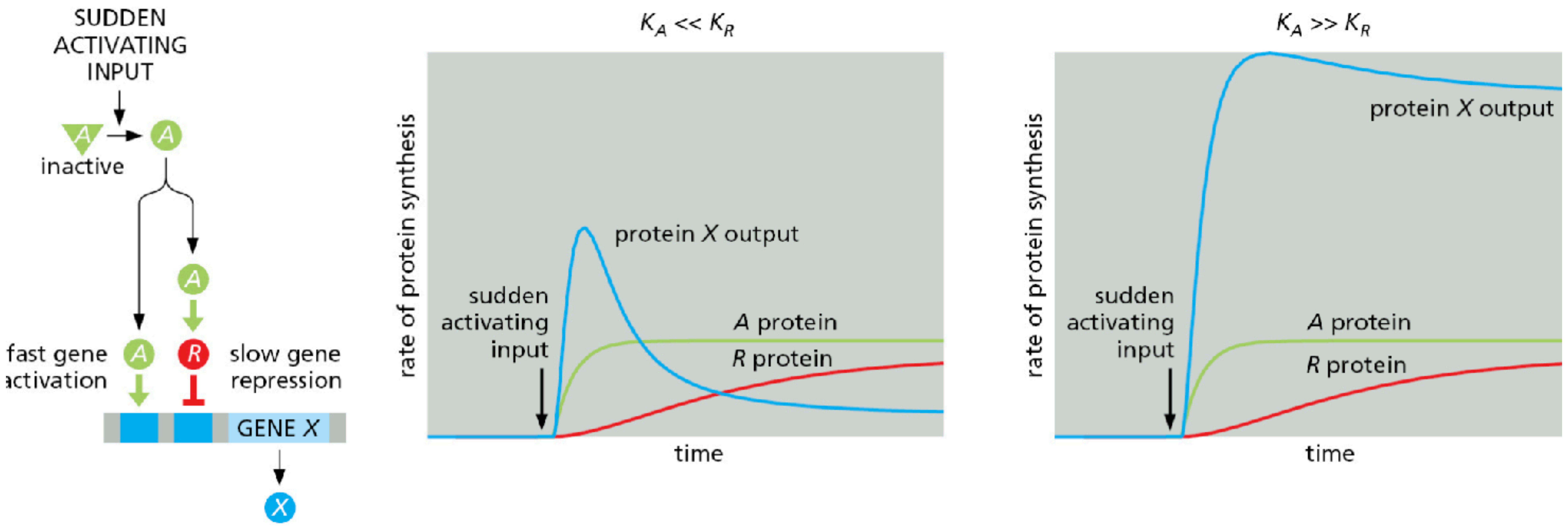
The system could toggle between the TetR-off and TetR-on states by the addition of external trigger stimuli



Current Opinion in Cell Biology

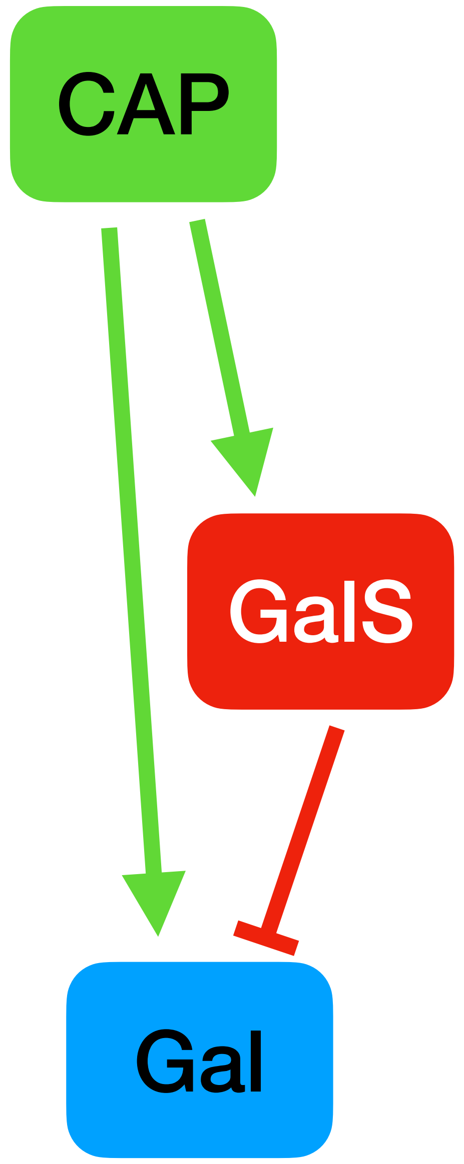
Mathematically analysis of cell function

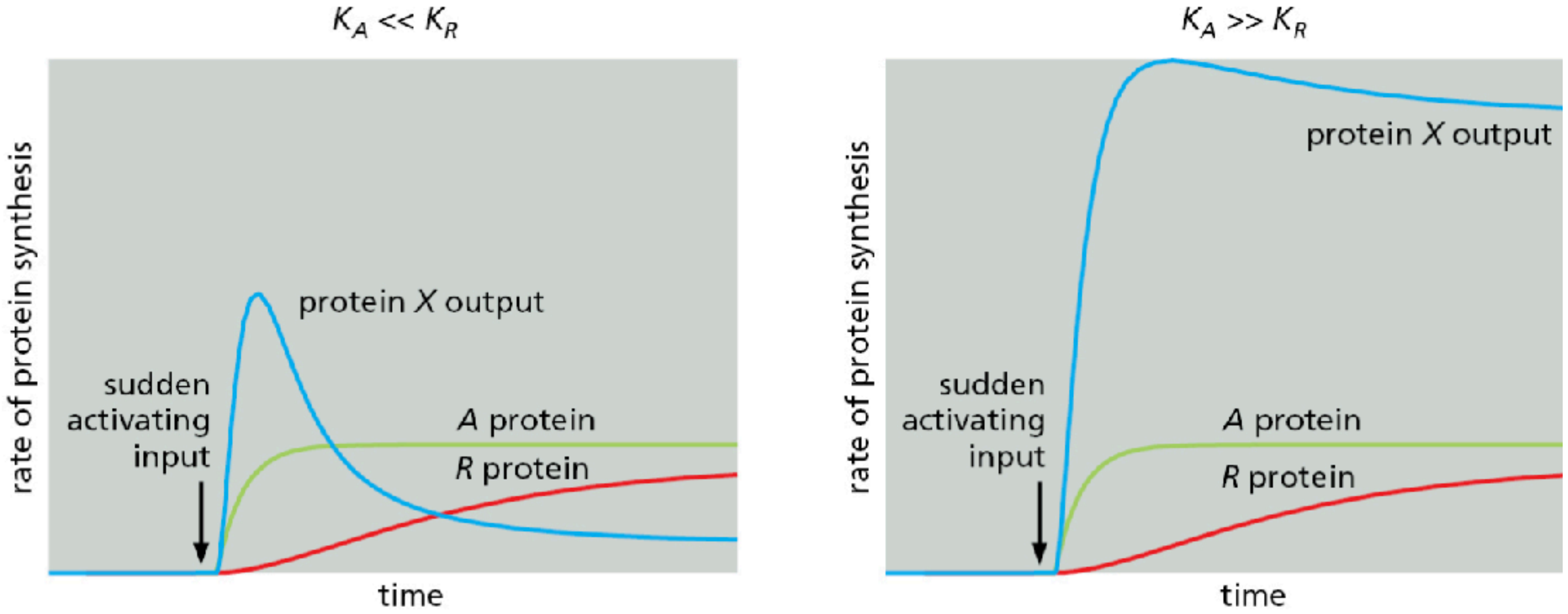
Incoherent feed-forward loops generates pulses



Mathematically analysis of cell function

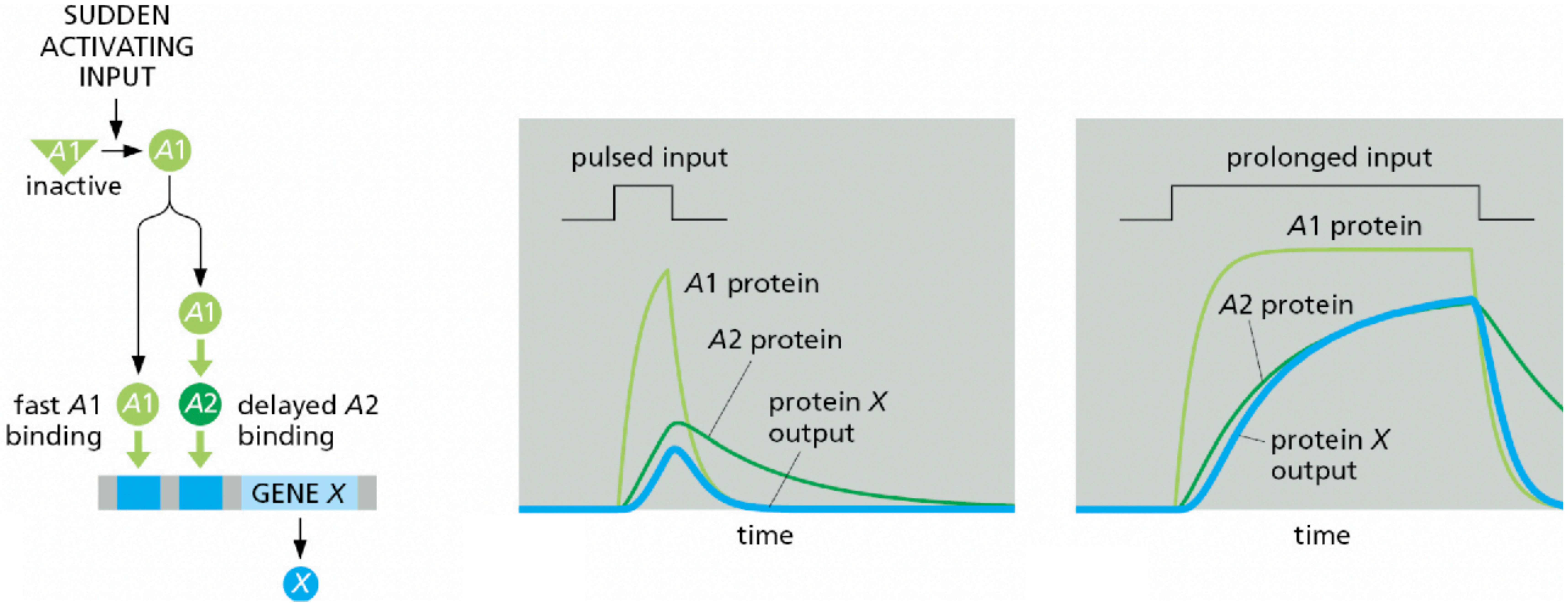
Incoherent feed-forward loops generates pulses





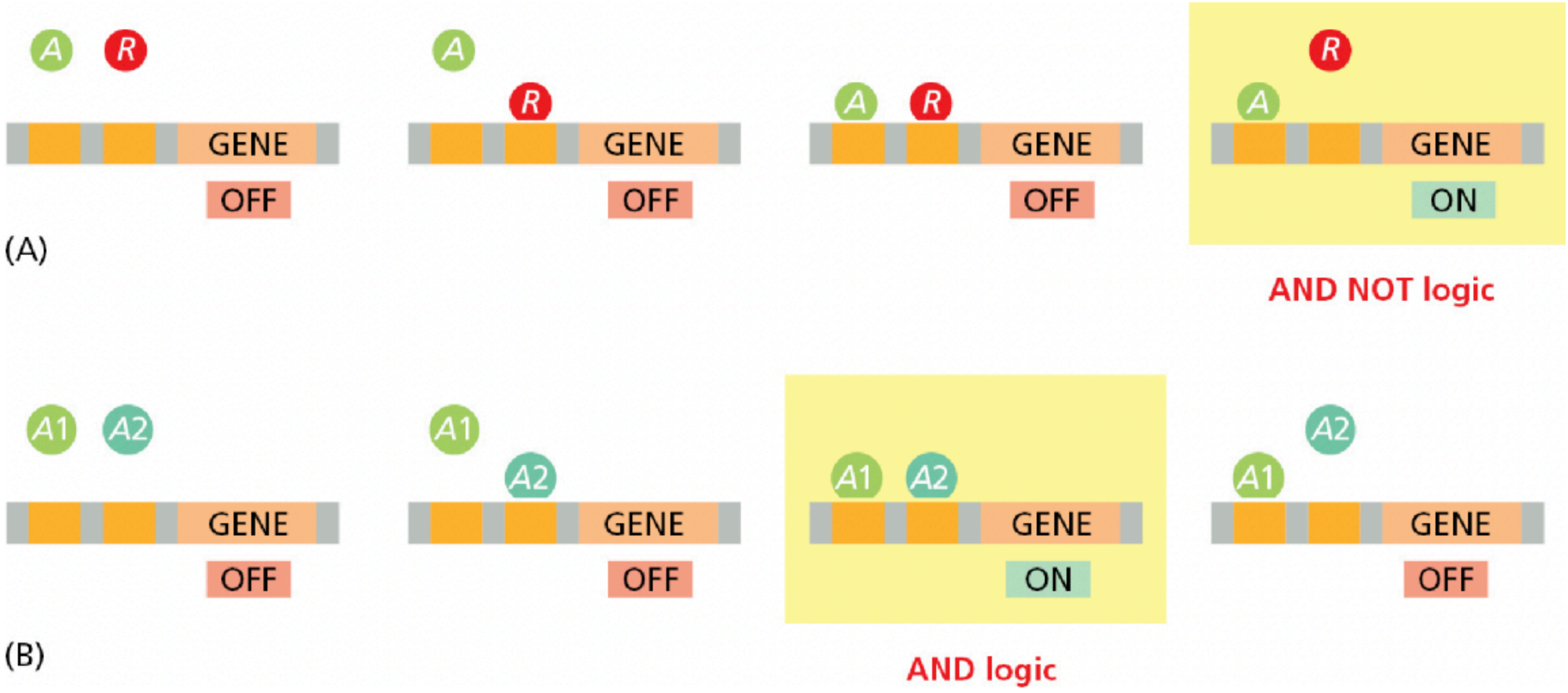
Mathematically analysis of cell function

Coherent feed-forward detects persistent inputs ignoring random fluctuations

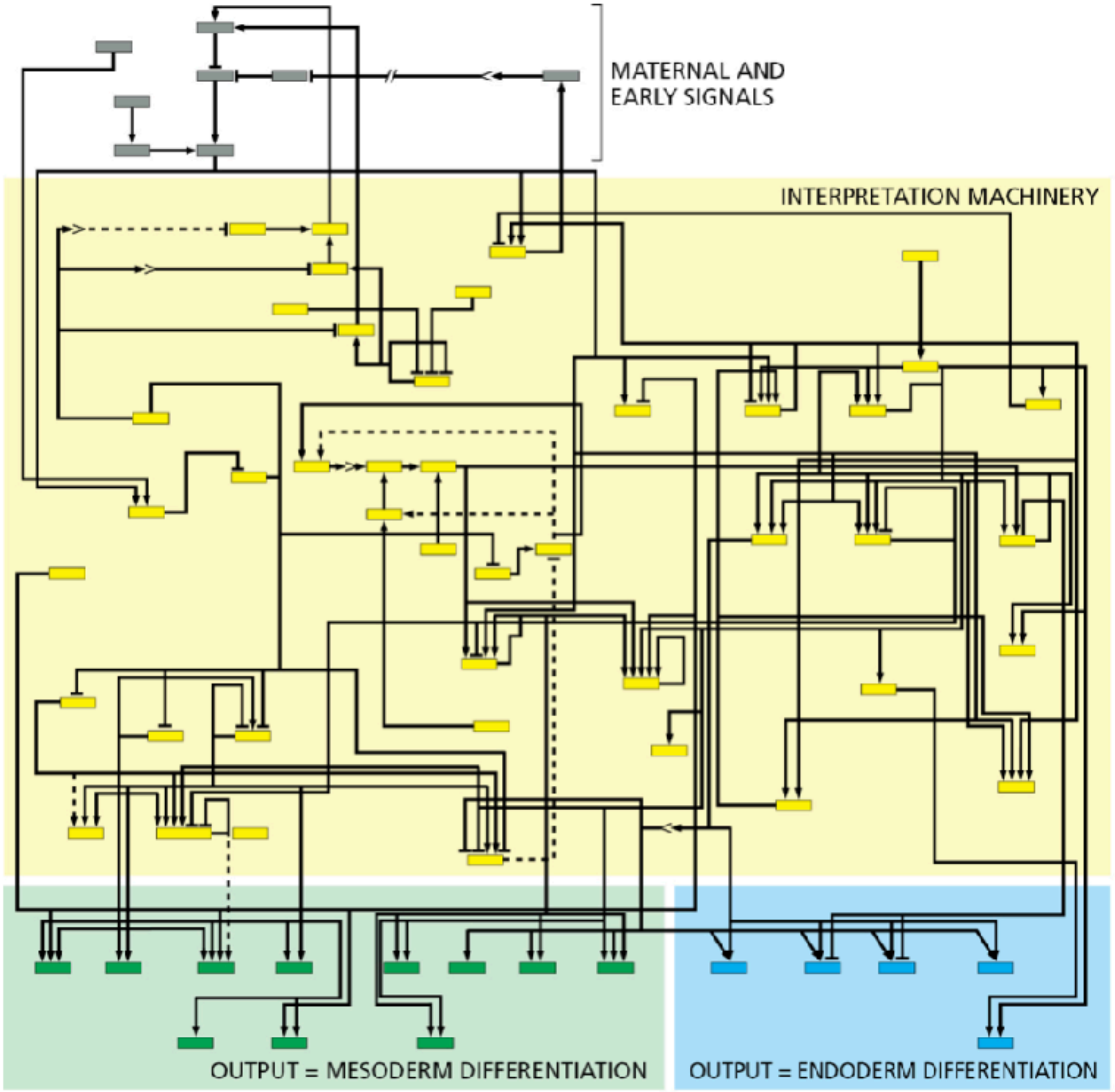


Mathematically analysis of cell function

Transcription regulators can exert combinatorial control

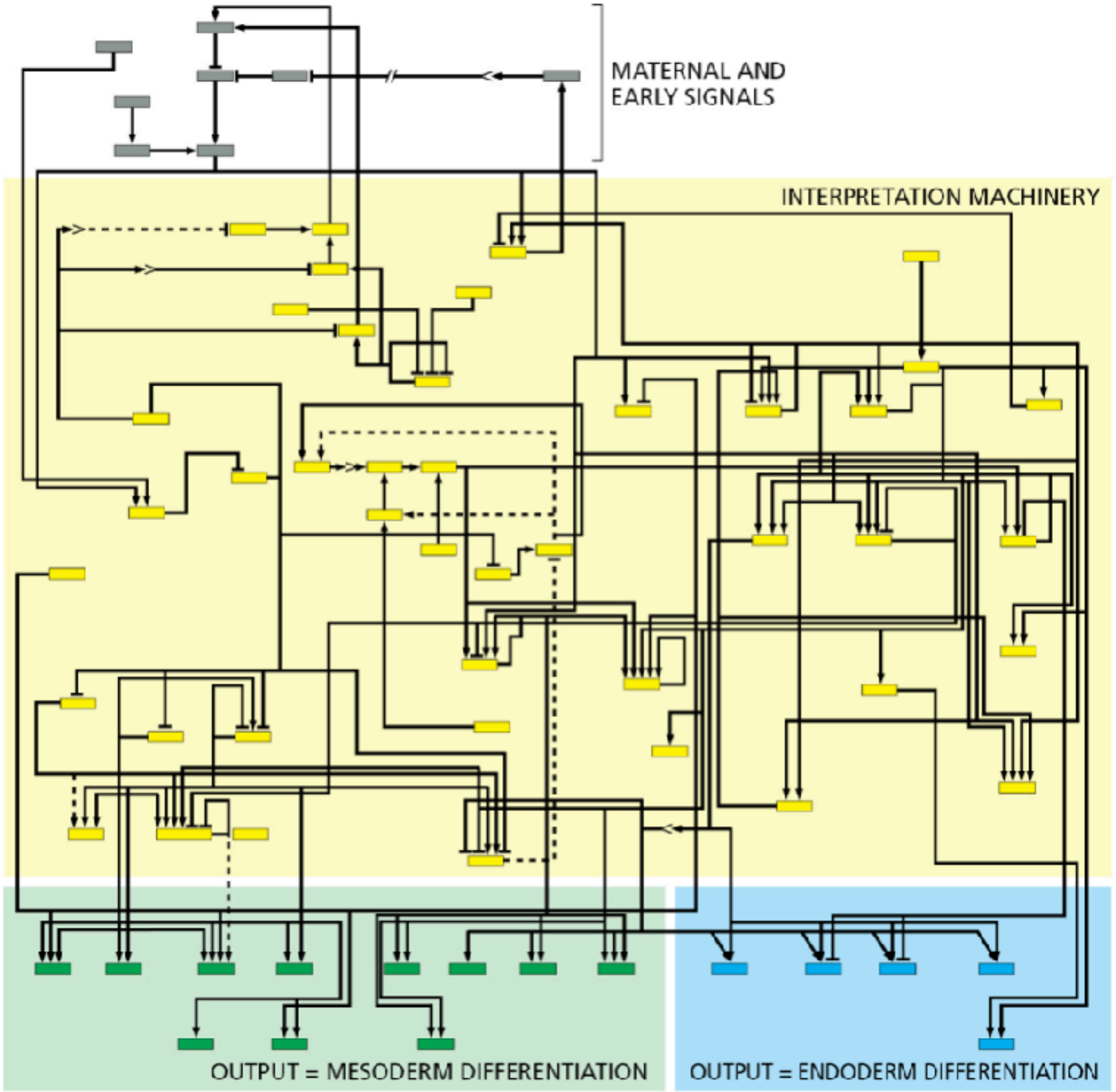


Genetic circuits



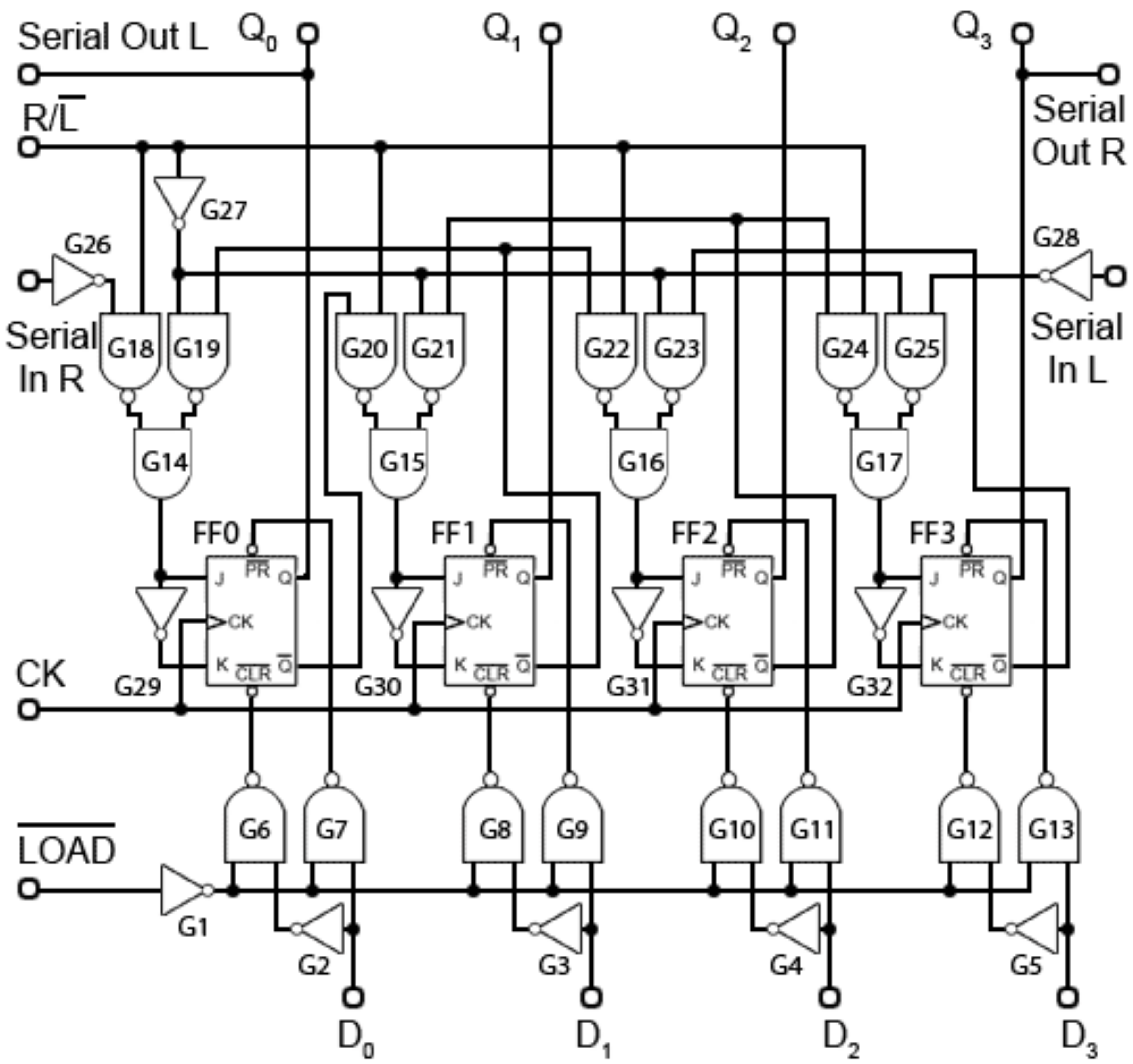
Par of the genetic circuit of the sea urchin developing embryo, (E.H. Davidson, *Nature* 474:635-639, 2011)

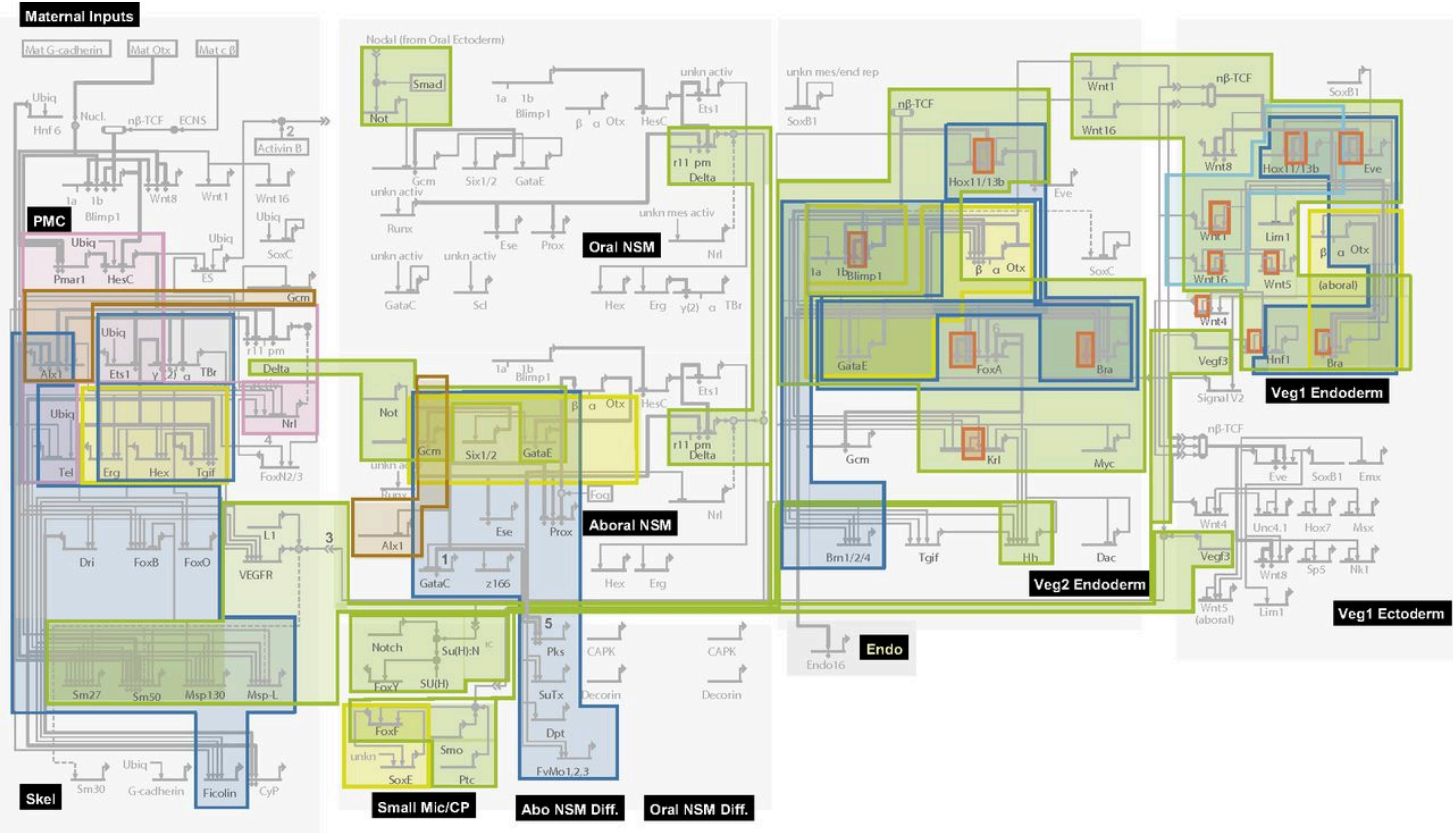
Genetic circuits



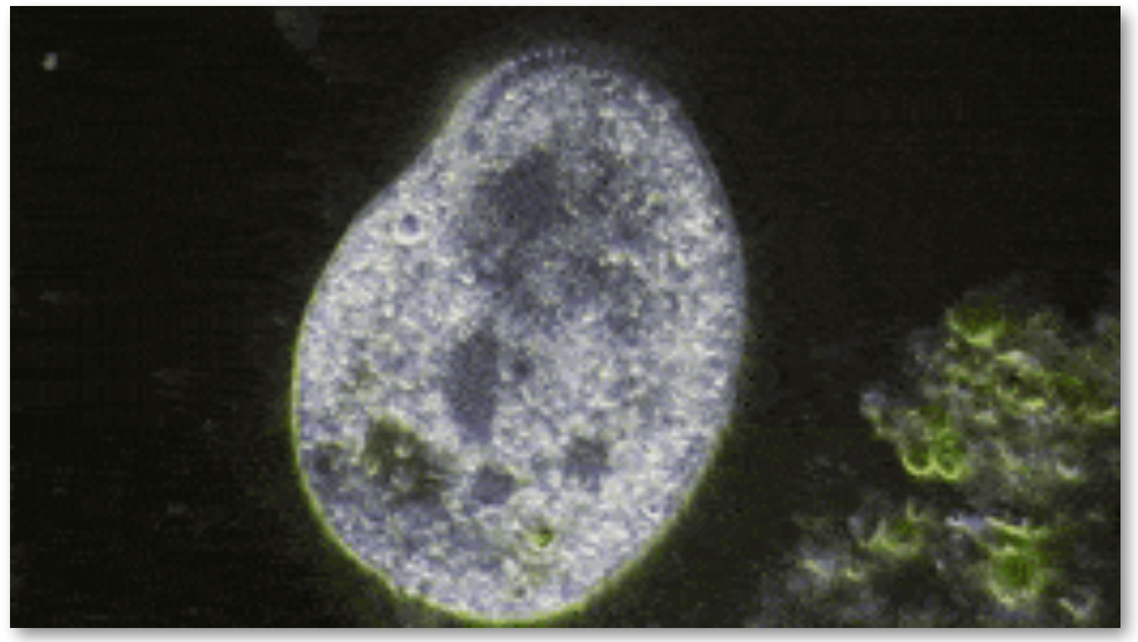
Par of the genetic circuit of the sea urchin developing embryo, (E.H. Davidson, *Nature* 474:635-639, 2011)

Digital electronics circuits



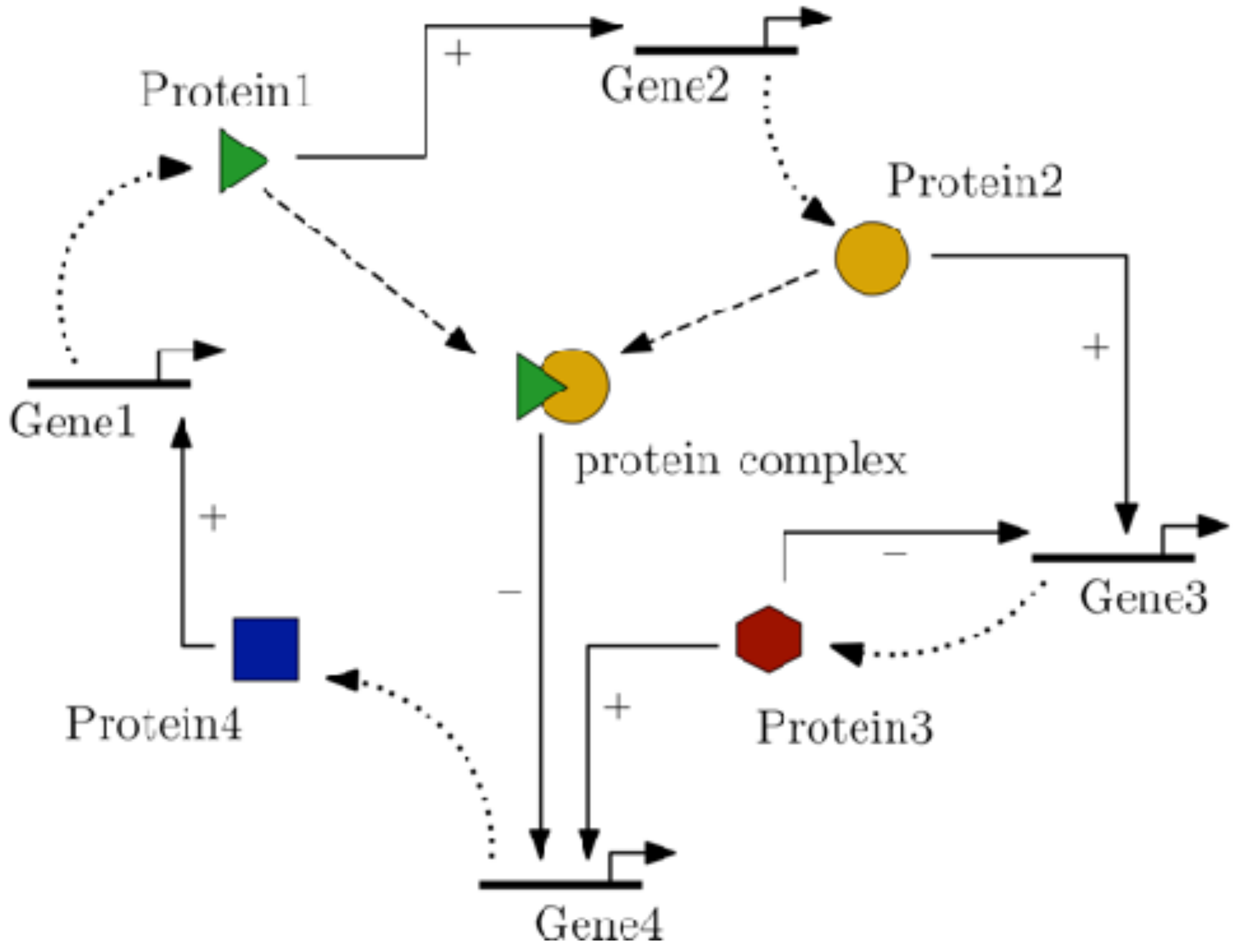

Maternal Inputs
Mat G-cadherin
Mat Otx
Mat cβ
Ubiq
Hnf 6
Nucl.
nβ-TCF
ECNS
Activin B
1a
1b
Blimp1
Wnt8
Wnt1
Wnt16
Ubiq
SoxC
ES
PMC
Pmar1
HesC
Gcm
Alx1
Ets1
γ(2)
α
TBr
r11 pm
Delta
Nrl
Tel
Erg
Hex
Tgif
FoxN2/3
Dri
FoxB
FoxO
L1
VEGFR
Sm27
Sm50
Msp130
Msp-L
Sm30
G-cadherin
Ficolin
CyP
Skel
Nodal (from Oral Ectoderm)
Smad
Not
Six1/2
GataE
unkn activ
Runx
Ese
Prox
Oral NSM
GataC
Scl
β
Otx
Ets1
r11 pm
Delta
unkn mes activ
Nrl
Erg
TBr
Aboral NSM
Fog
1
z166
5
Notch
Su(H):N
FoxY
SU(H)
FoxF
unkn
SoxE
Smo
Ptc
Pks
SuTx
Dpt
FvMo1,2,3
CAPK
Decorin
Small Mic/CP
Abo NSM Diff.
Oral NSM Diff.
2
3
4
6
unkn mes/end rep
SoxB1
nβ-TCF
Hox11/13b
Eve
FoxA
Bra
Krl
Myc
Gcm
Bm1/2/4
Tgif
Hh
Dac
Endo16
Endo
Veg2 Endoderm
Wnt1
Wnt16
Wnt8
Lim1
Wnt5
(aboral)
Wnt4
Hnf1
Vegf3
Signal V2
Veg1 Endoderm
SoxB1
Emx
Unc4.1
Hox7
Msx
Sp5
Nk1
Wnt5 (aboral)
Veg1 Ectoderm

Reverse engineering of gene regulatory nets



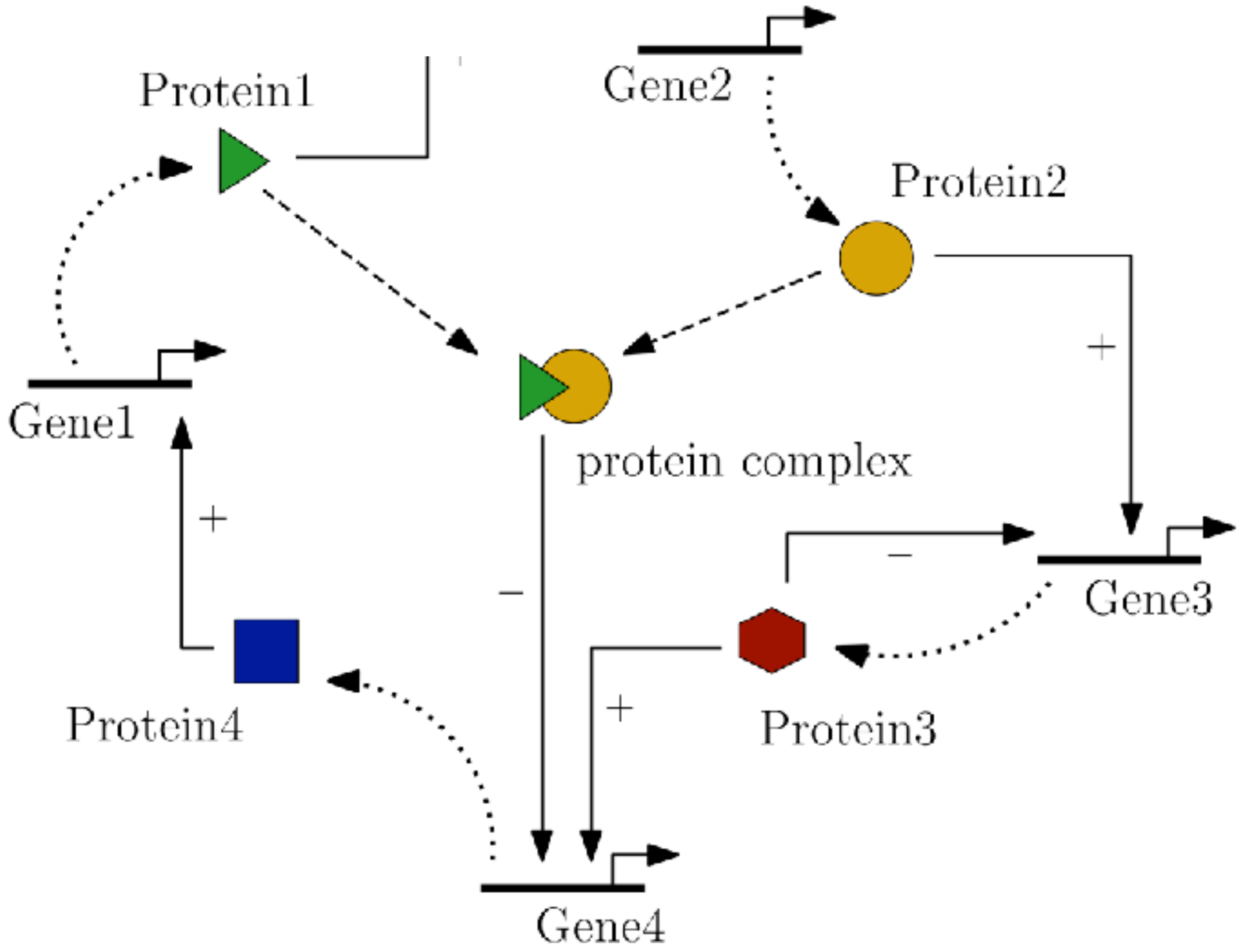
Living thing

Reverse Engineering

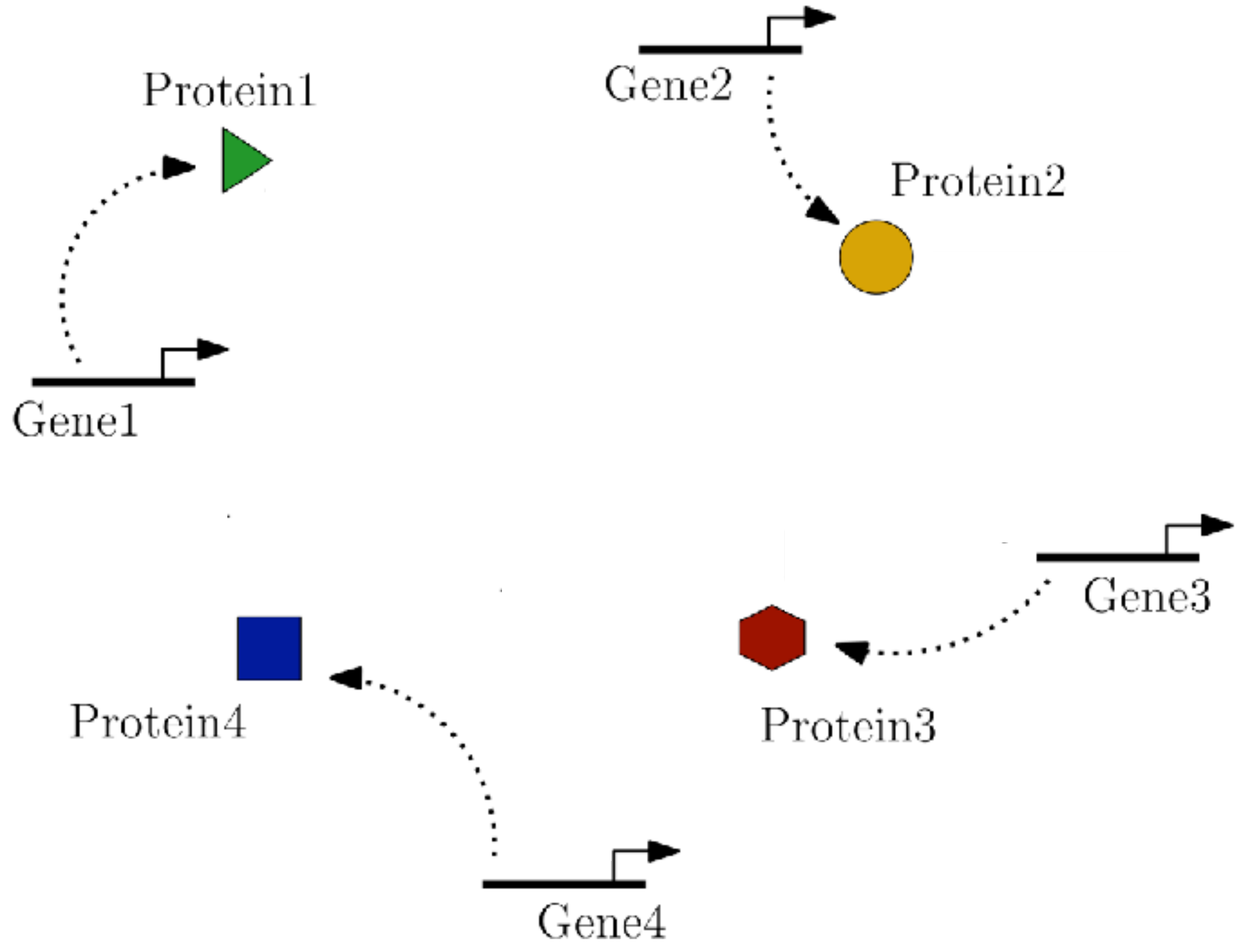


Gene regulatory Model

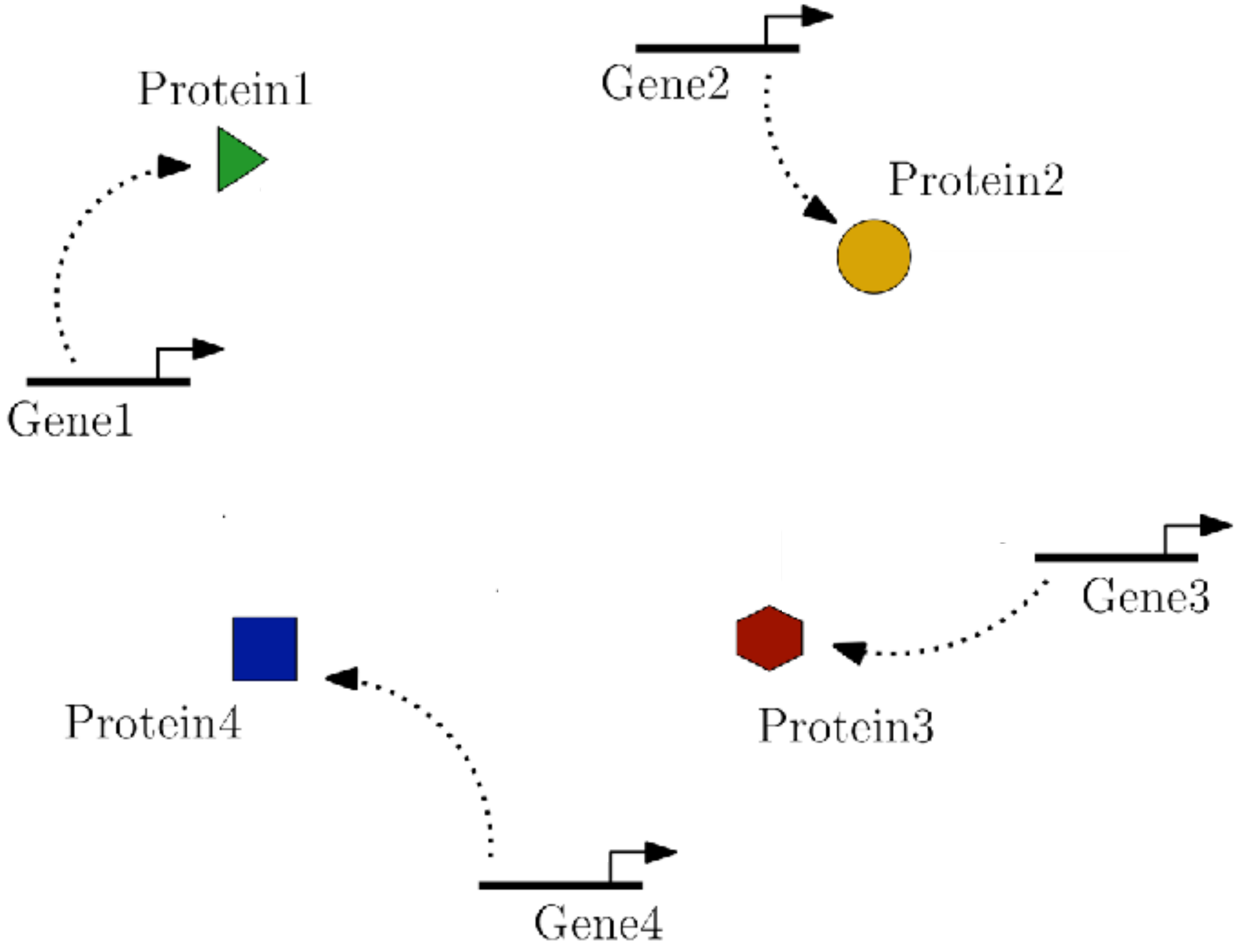
Reverse engineering of gene regulatory nets



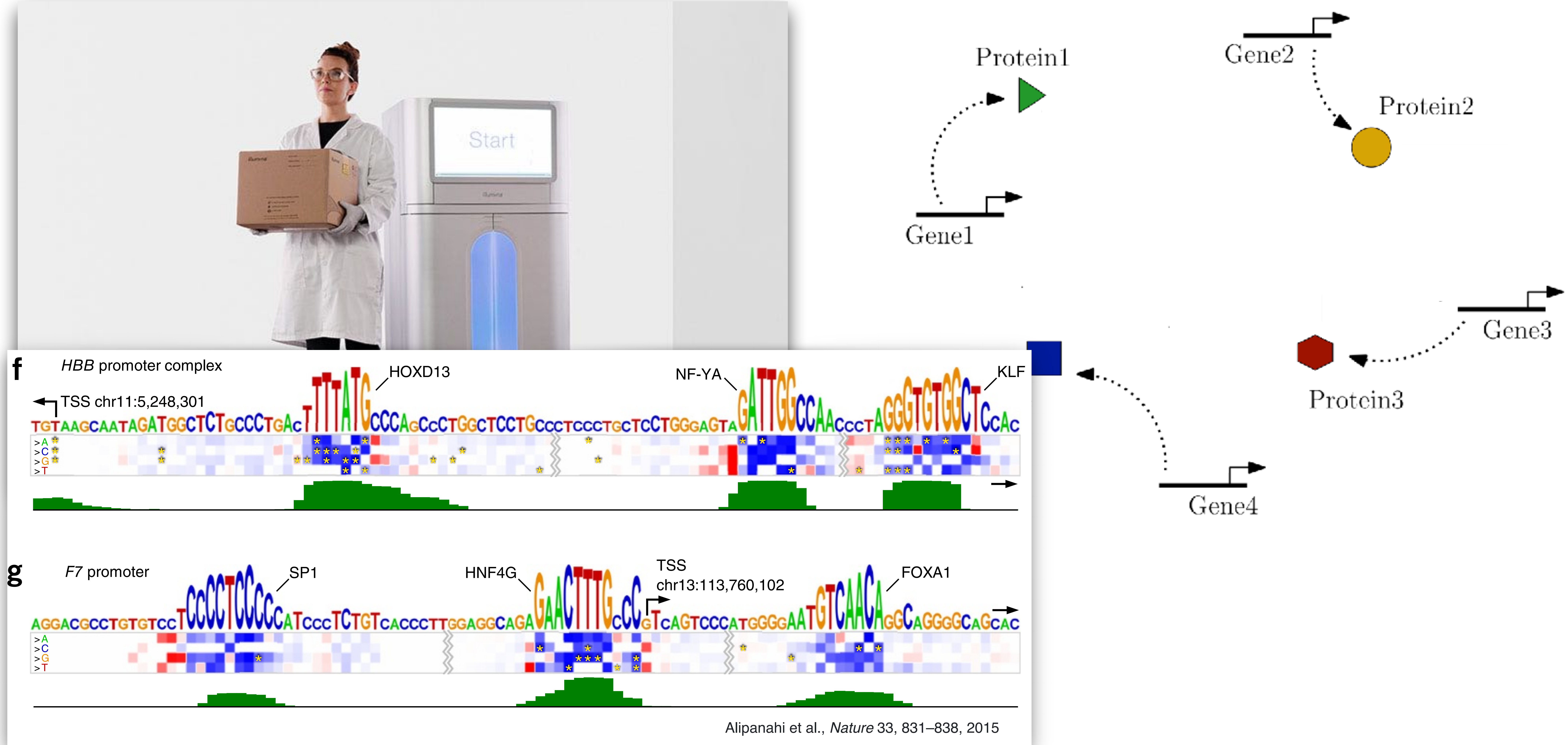
Reverse engineering of gene regulatory nets



Reverse engineering of gene regulatory nets

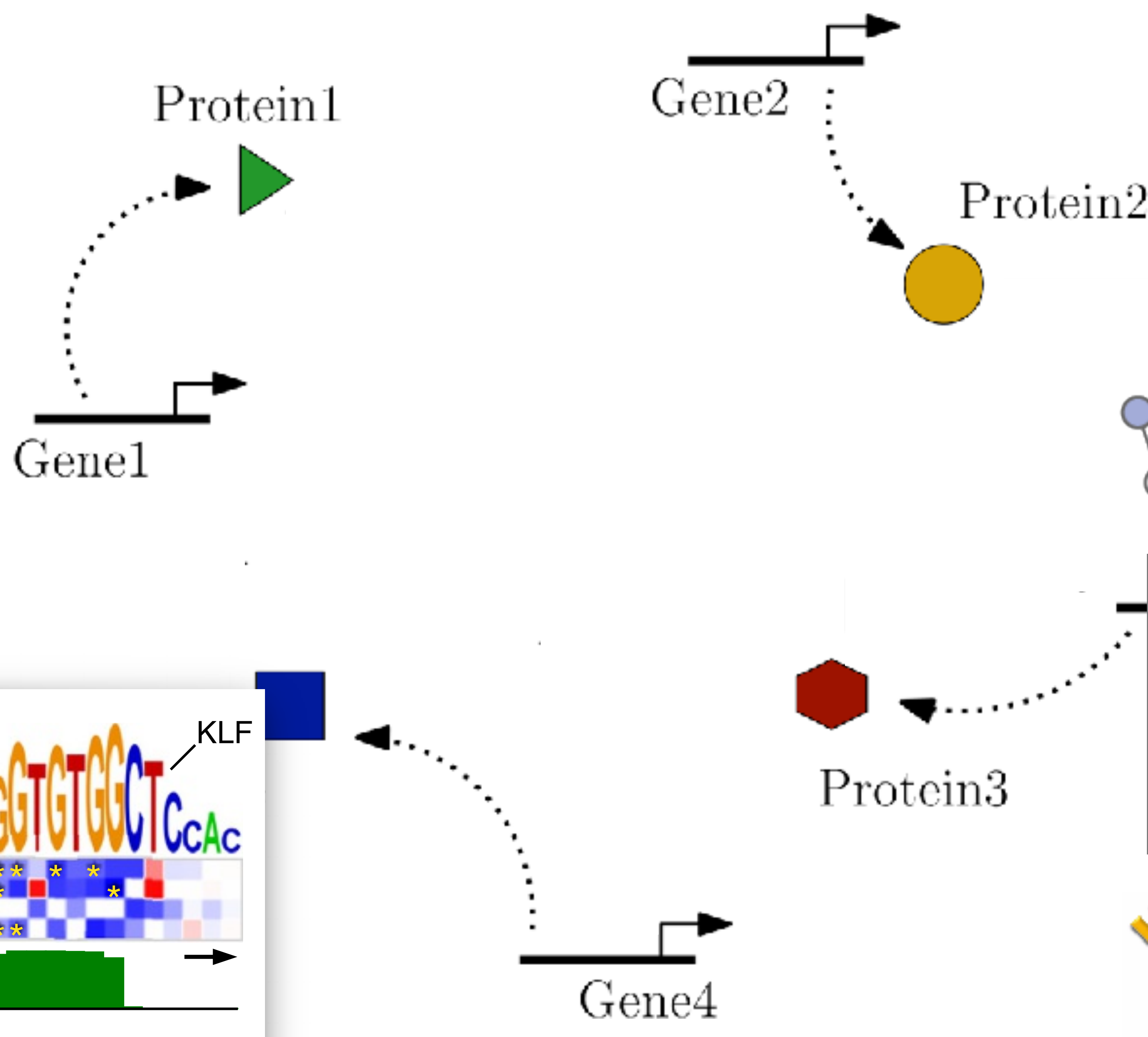


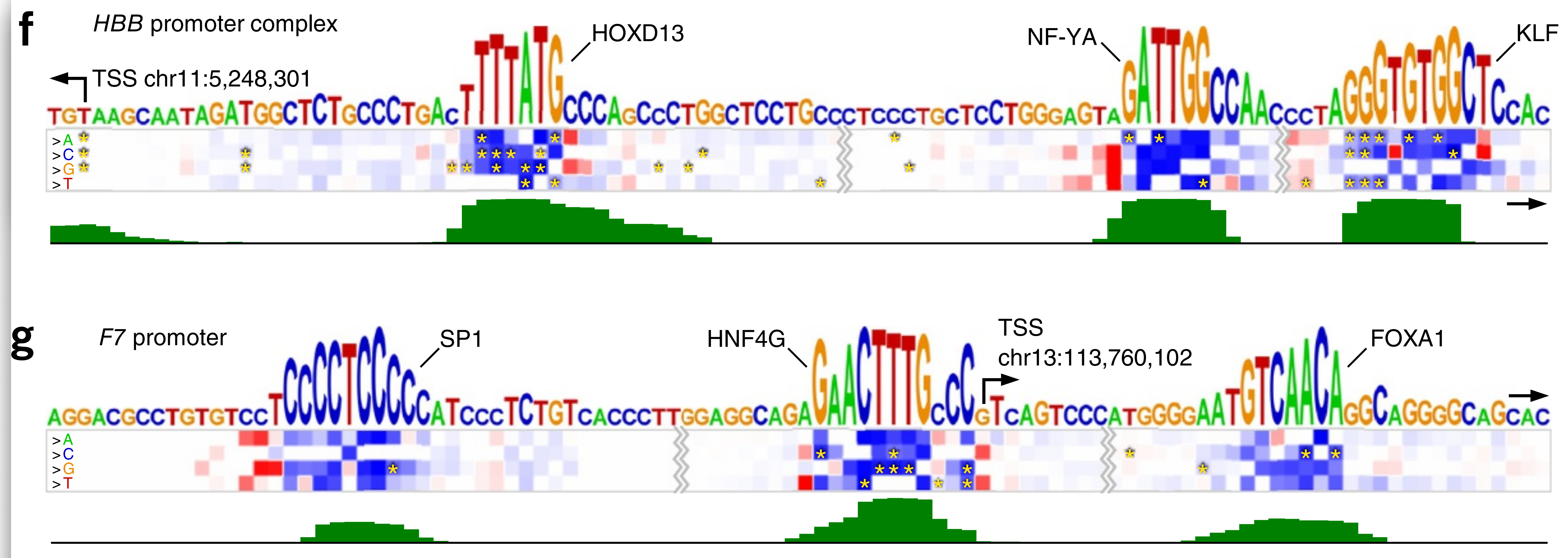
Reverse engineering of gene regulatory nets



Alipanahi et al., *Nature* 33, 831–838, 2015

Reverse engineering of gene regulatory nets





Alipanahi et al., *Nature* 33, 831–838, 2015

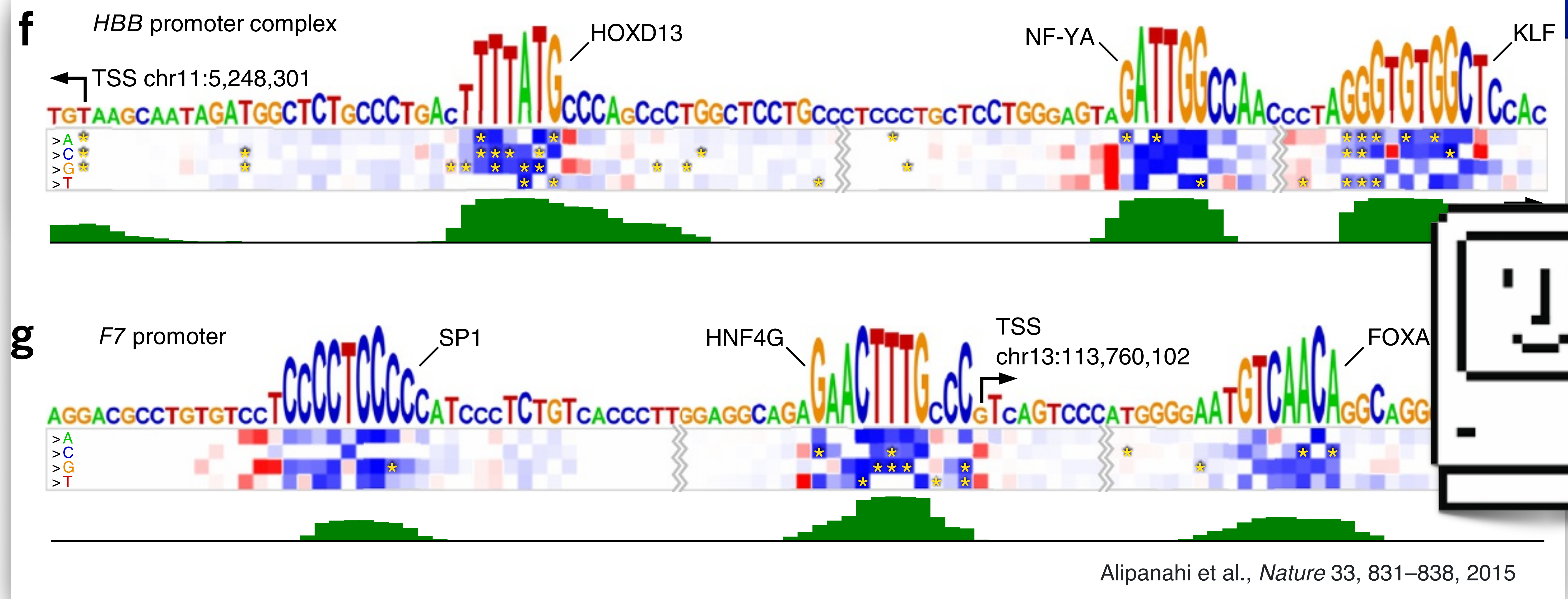
STRING

ENCODE

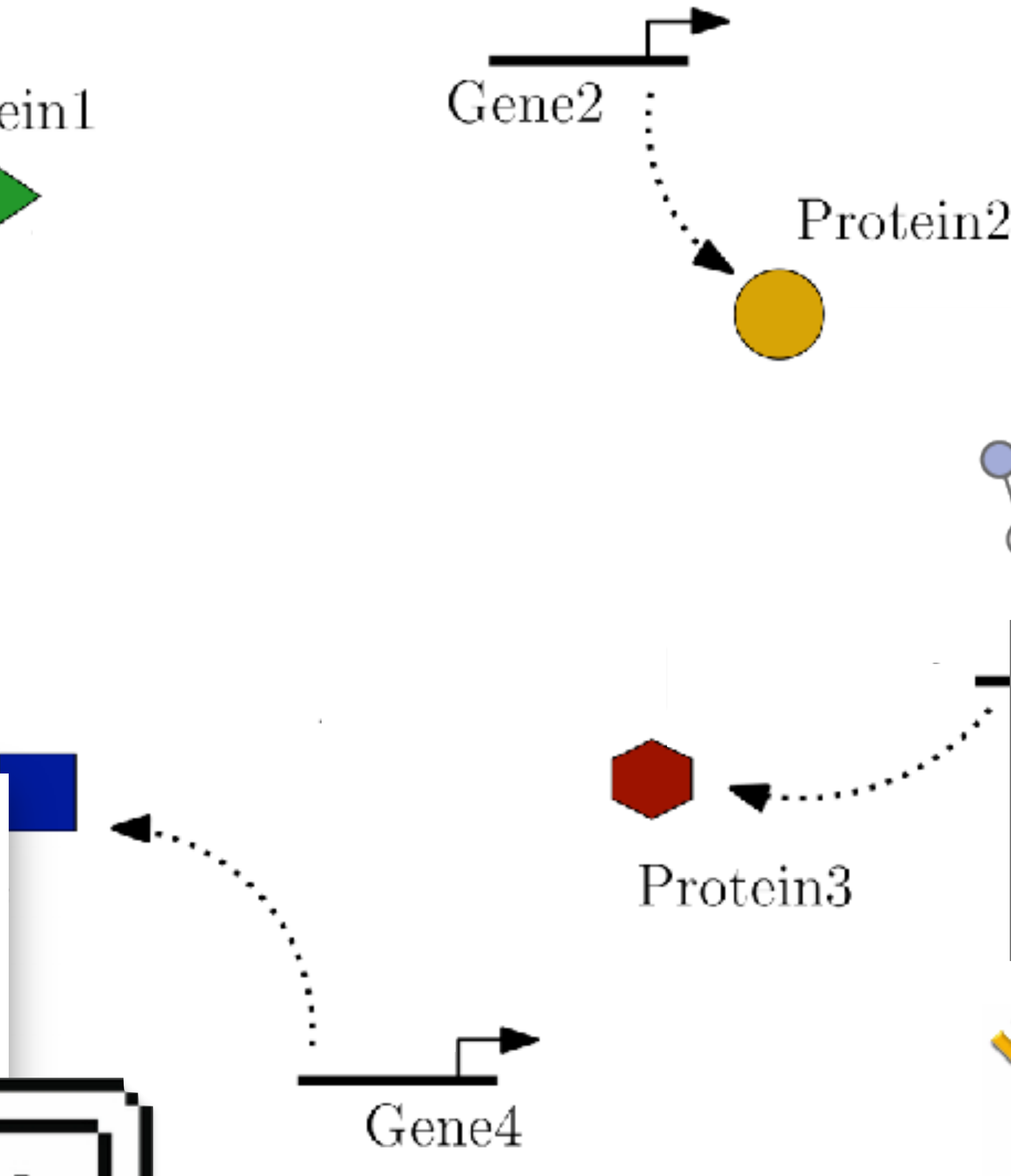
TRANSFAC

JASPAR

Reverse engineering of gene regulatory nets



Alipanahi et al., *Nature* 33, 831–838, 2015



STRING

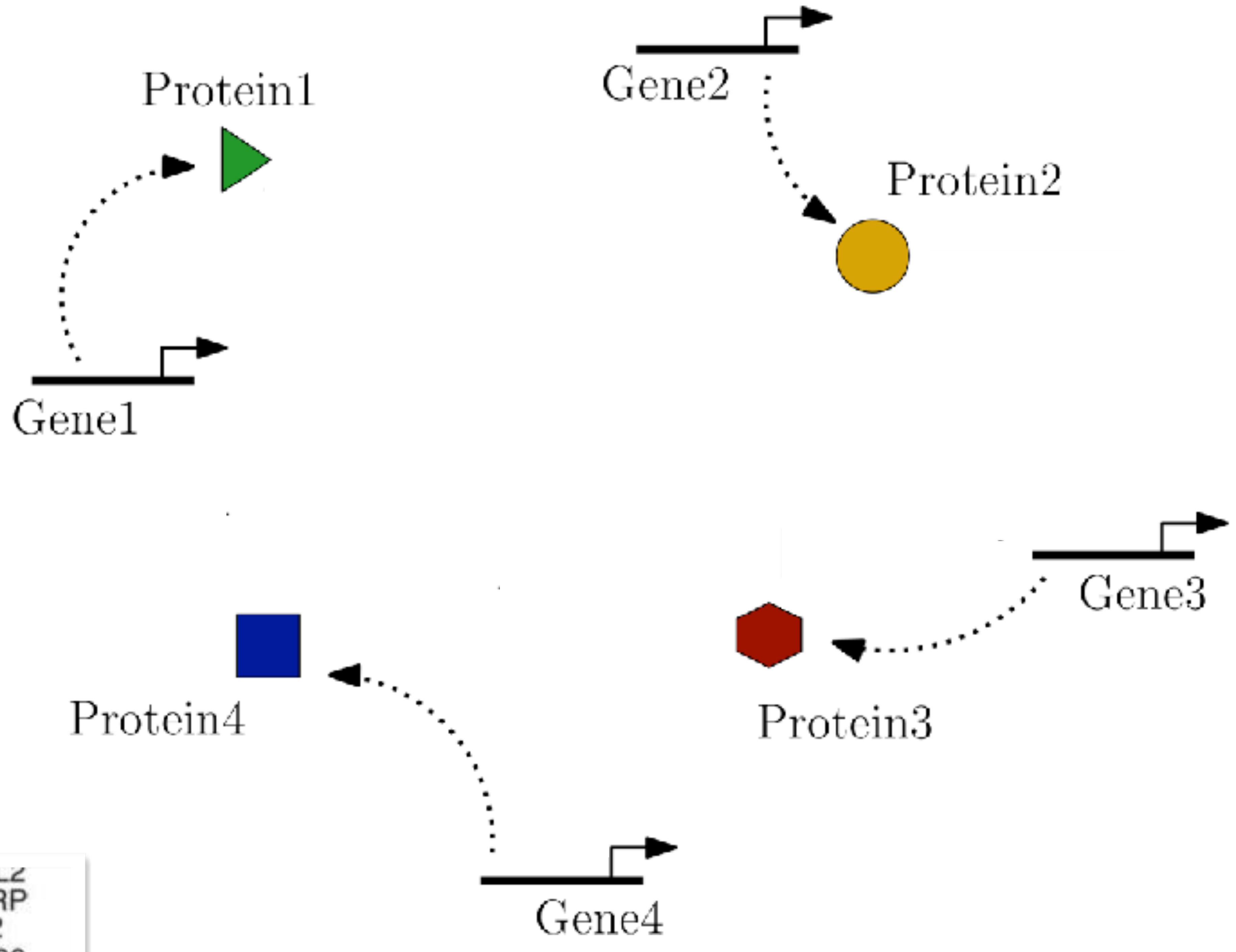
ENCODE

TRANSFAC

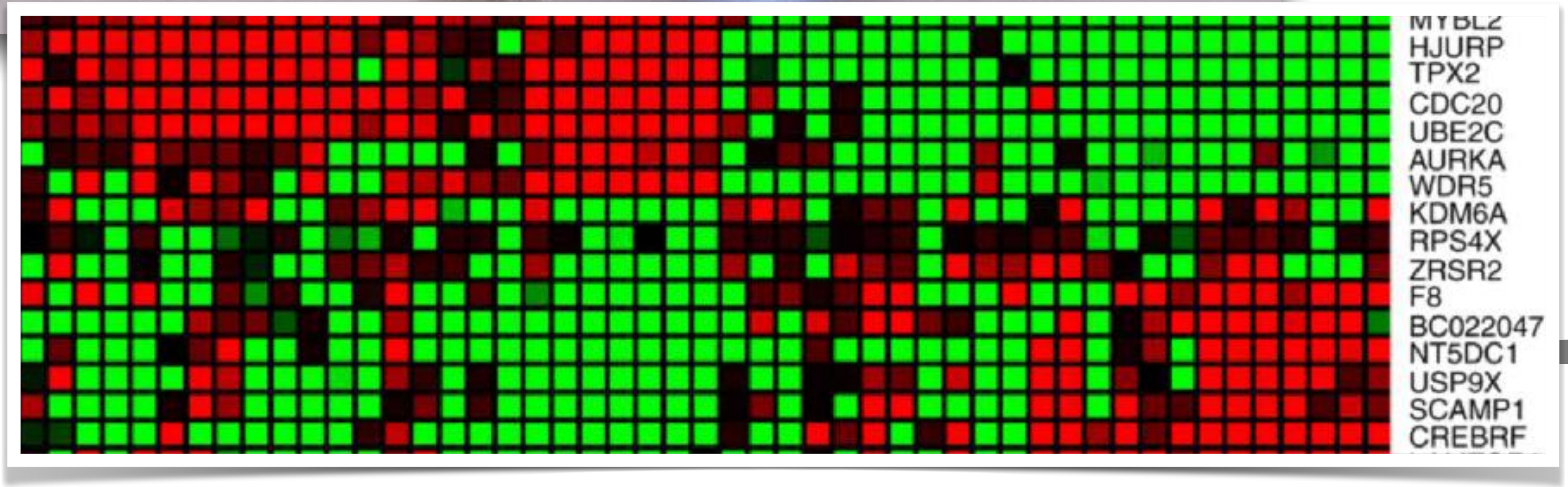
Promoter Analysis

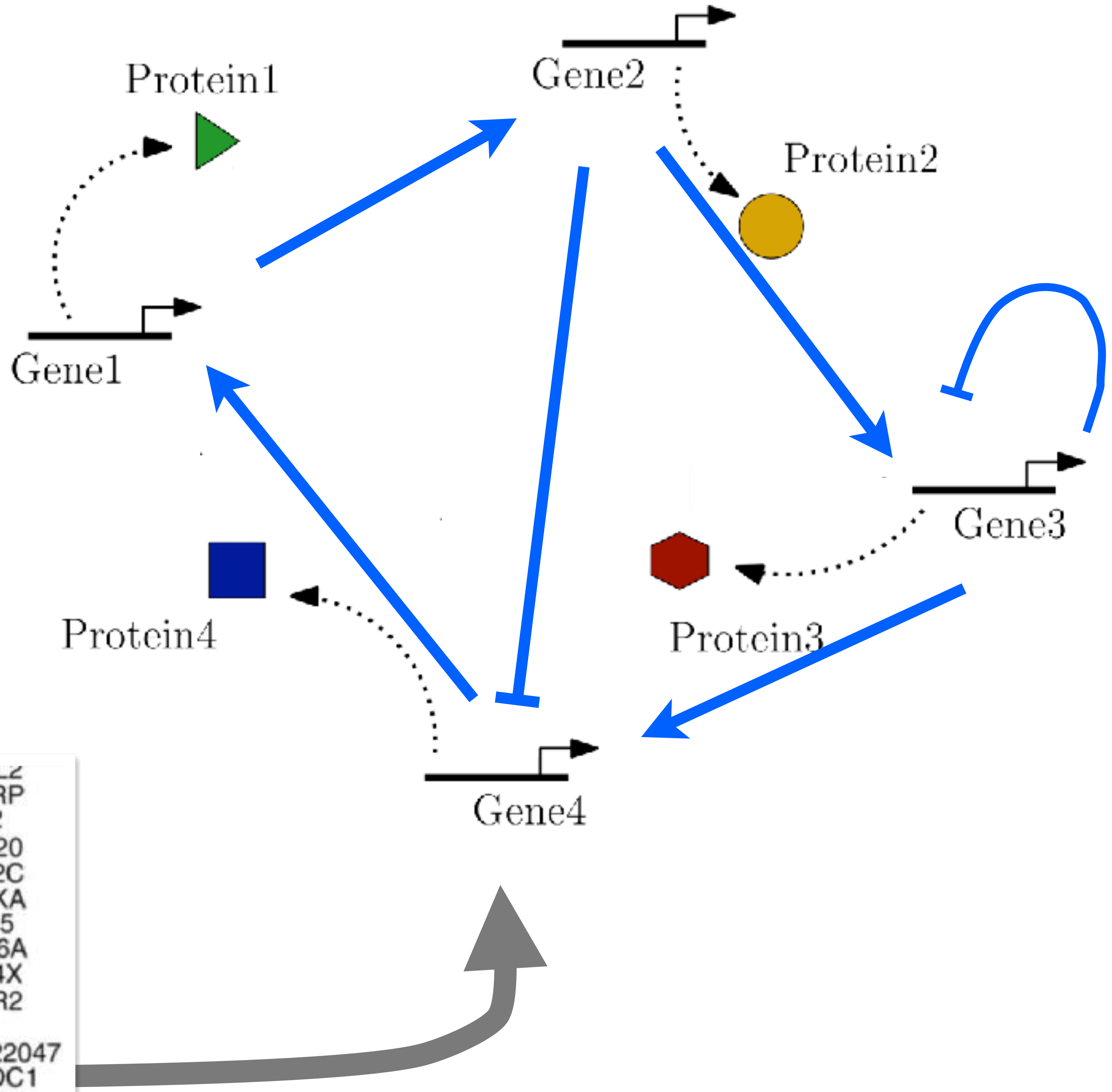
We'll try this in Lab session

Reverse engineering of gene regulatory nets



Reverse engineering of gene regulatory nets

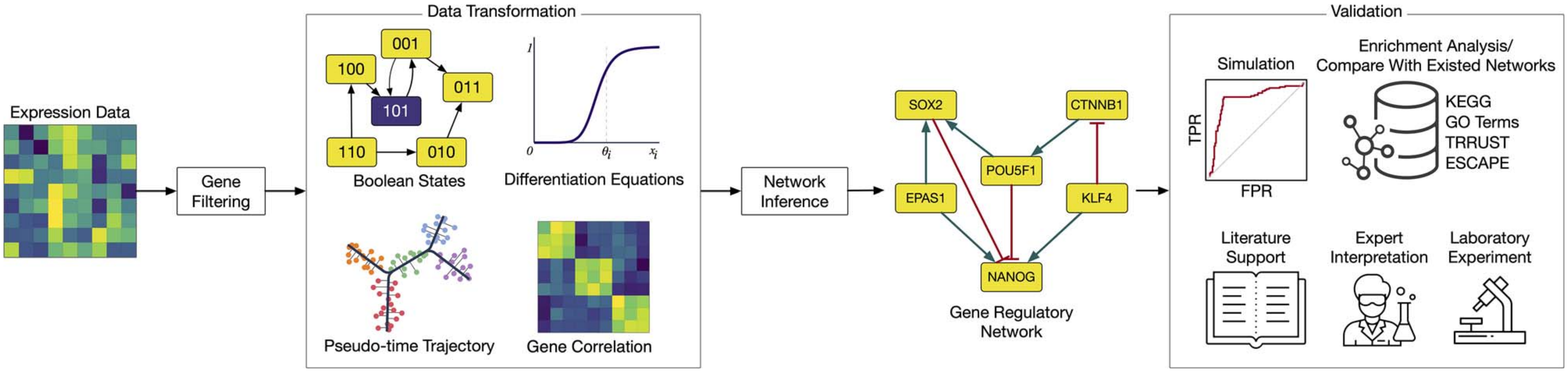




Reverse Engineering

Reverse engineering of gene regulatory nets

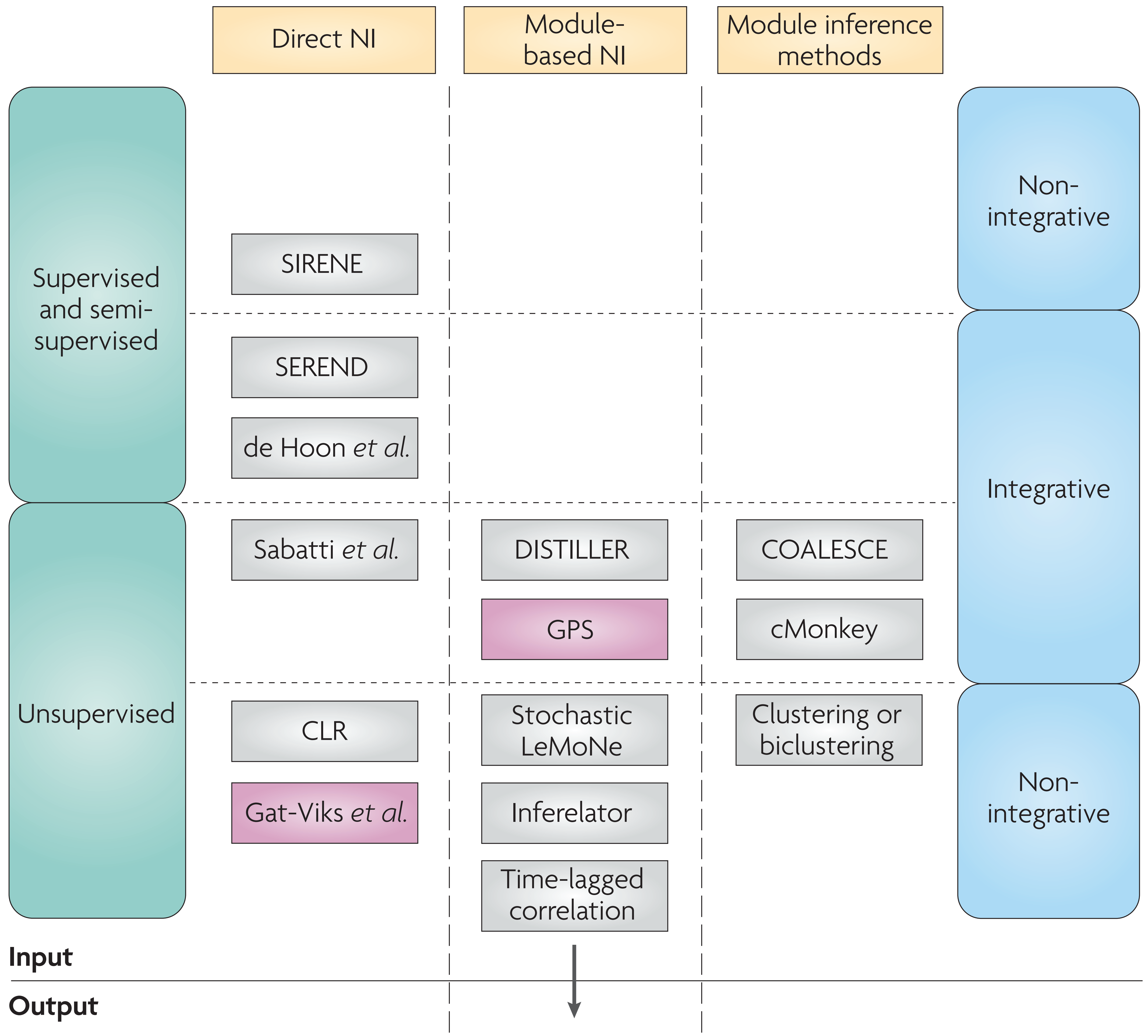
20 years of approaches



Nguyen et al., *Briefing in bioinformatics,* 2020

Reverse engineering of gene regulatory nets

20 years of approaches



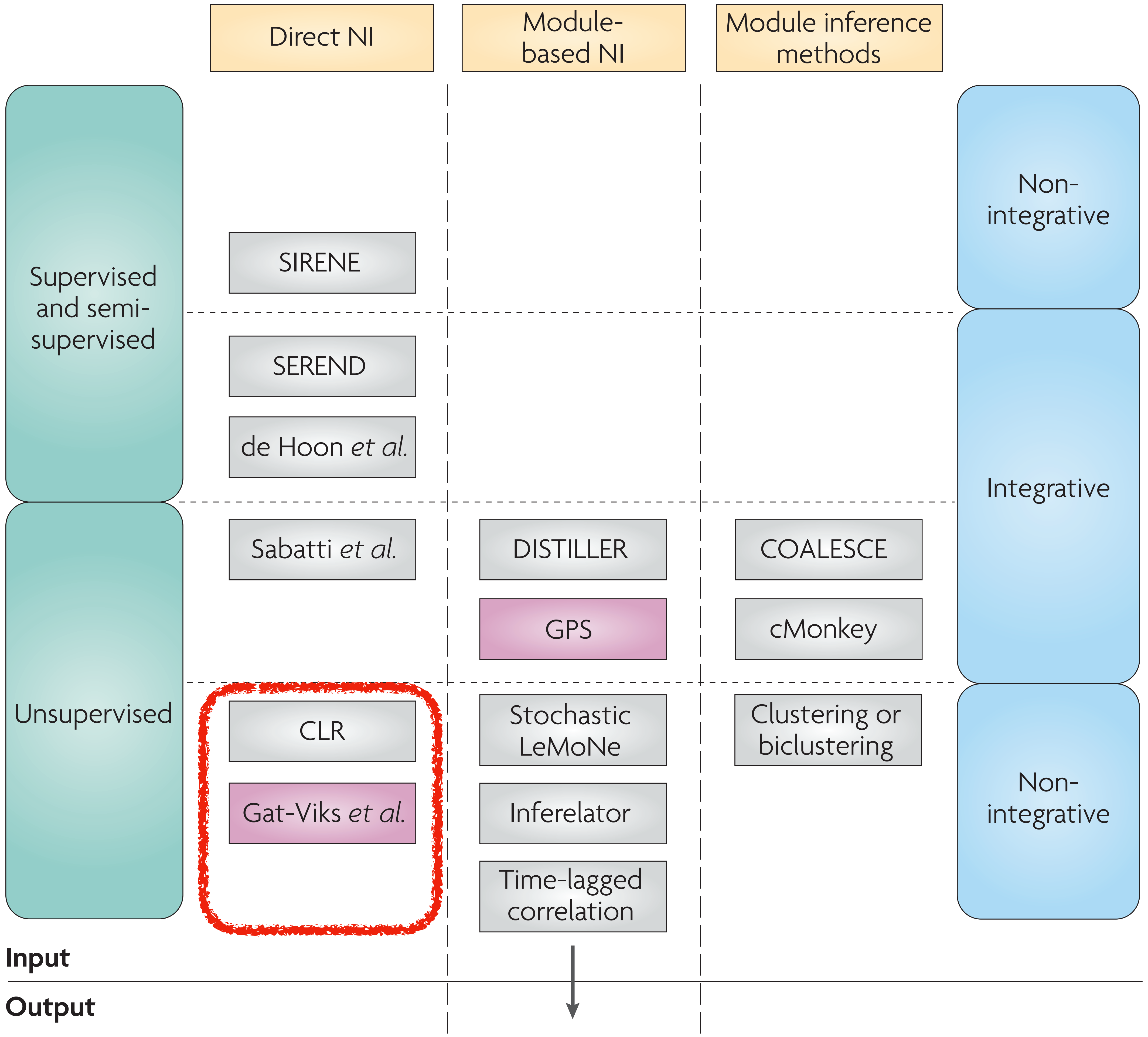
Smet and Marchal, *Nature Reviews Microbiology*, 2010

Reverse engineering of gene regulatory nets

20 years of approaches

Connect similar genes

- Correlation
- Mutual Information (ARACNE)

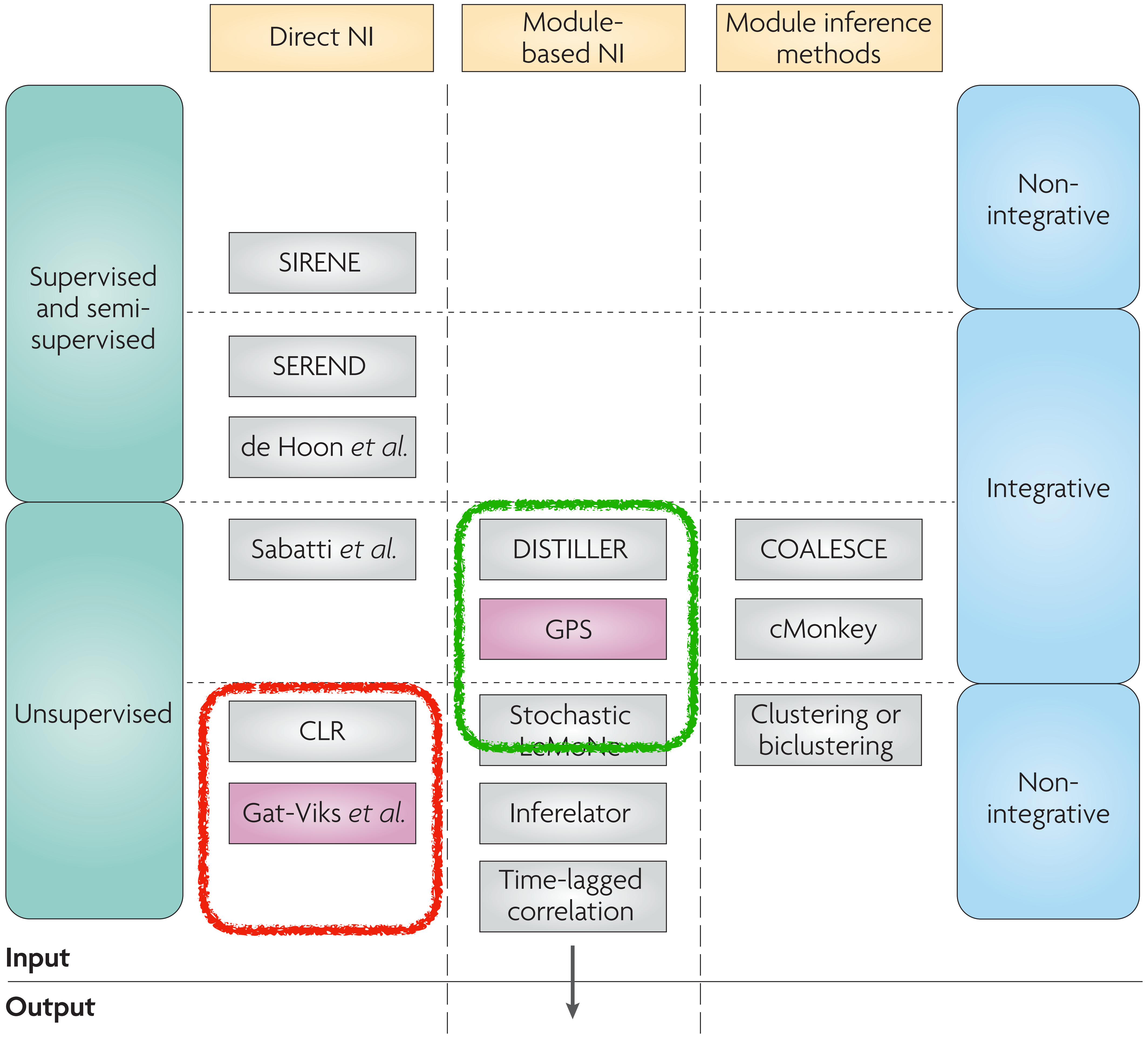


Smet and Marchal, *Nature Reviews Microbiology*, 2010

Reverse engineering of gene regulatory nets

20 years of approaches

- Connect similar genes
  - Correlation
  - Mutual Information (ARACNE)
- Model Based
  - Sparse regression (GENIE3)
  - Boolean/Bayesian networks



Direct NI

Module-based NI

Module inference methods

Supervised and semi-supervised

Unsupervised

Non-integrative

Integrative

Non-integrative

SIRENE

SEREND

de Hoon *et al.*

Sabatti *et al.*

DISTILLER

GPS

COALESCE

cMonkey

CLR

Gat-Viks *et al.*

Stochastic LeMoNe

Inferelator

Time-lagged correlation

Clustering or biclustering

Input

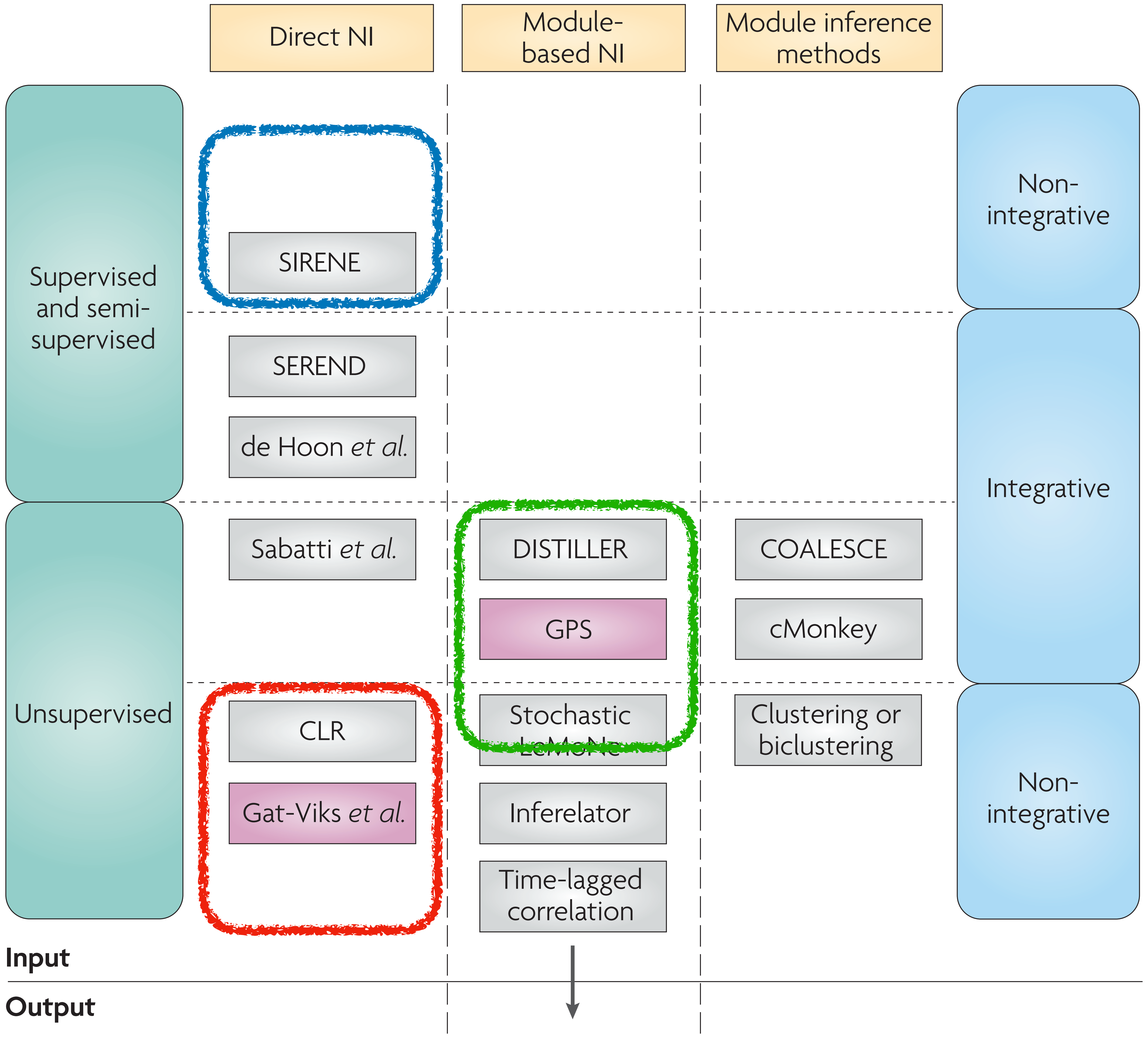
Output

Smet and Marchal, *Nature Reviews Microbiology*, 2010

Reverse engineering of gene regulatory nets

20 years of approaches

- Connect similar genes
  - Correlation
  - Mutual Information (ARACNE)
- Model Based
  - Sparse regression (GENIE3)
  - Boolean/Bayesian networks
- Supervised
  - SIRENE



Smet and Marchal, *Nature Reviews Microbiology*, 2010

Reverse engineering of gene regulatory nets

Relevance Networks

Reverse engineering of gene regulatory nets

Relevance Networks

- Choose a measure of association $F(X, Y)$

Reverse engineering of gene regulatory nets

Relevance Networks

- Choose a measure of association $F(X, Y)$
- Define a threshold value τ_F

Reverse engineering of gene regulatory nets

Relevance Networks

- Choose a measure of association $F(X, Y)$
- Define a threshold value τ_F
- For all pairs of genes (X, Y) compute their association $F(X, Y)$

Reverse engineering of gene regulatory nets

Relevance Networks

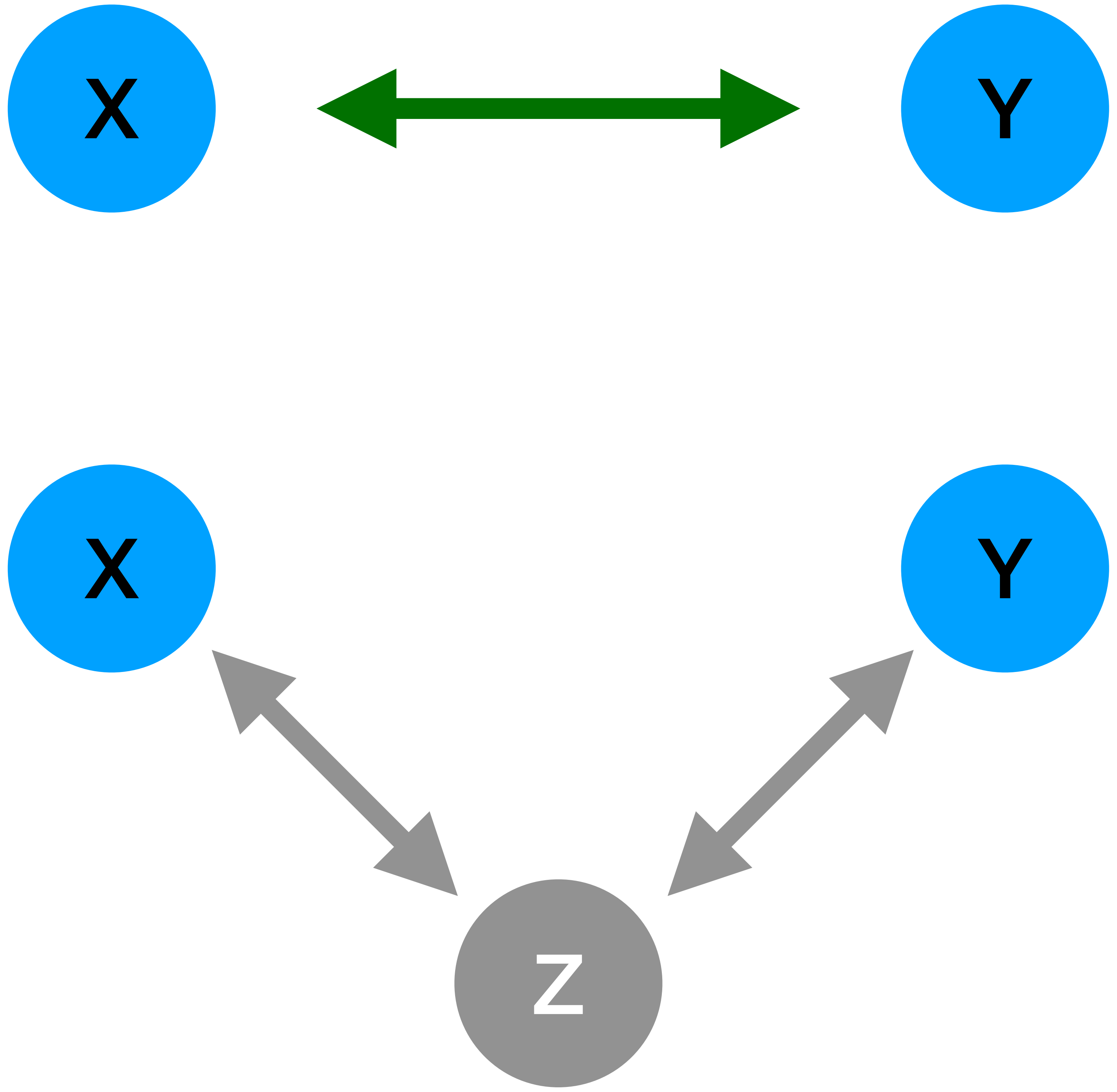
- Choose a measure of association $F(X, Y)$
- Define a threshold value τ_F
- For all pairs of genes (X, Y) compute their association $F(X, Y)$
- Retain associations such that $F(X, Y) > \tau_F$

Reverse engineering of gene regulatory nets

Relevance Networks

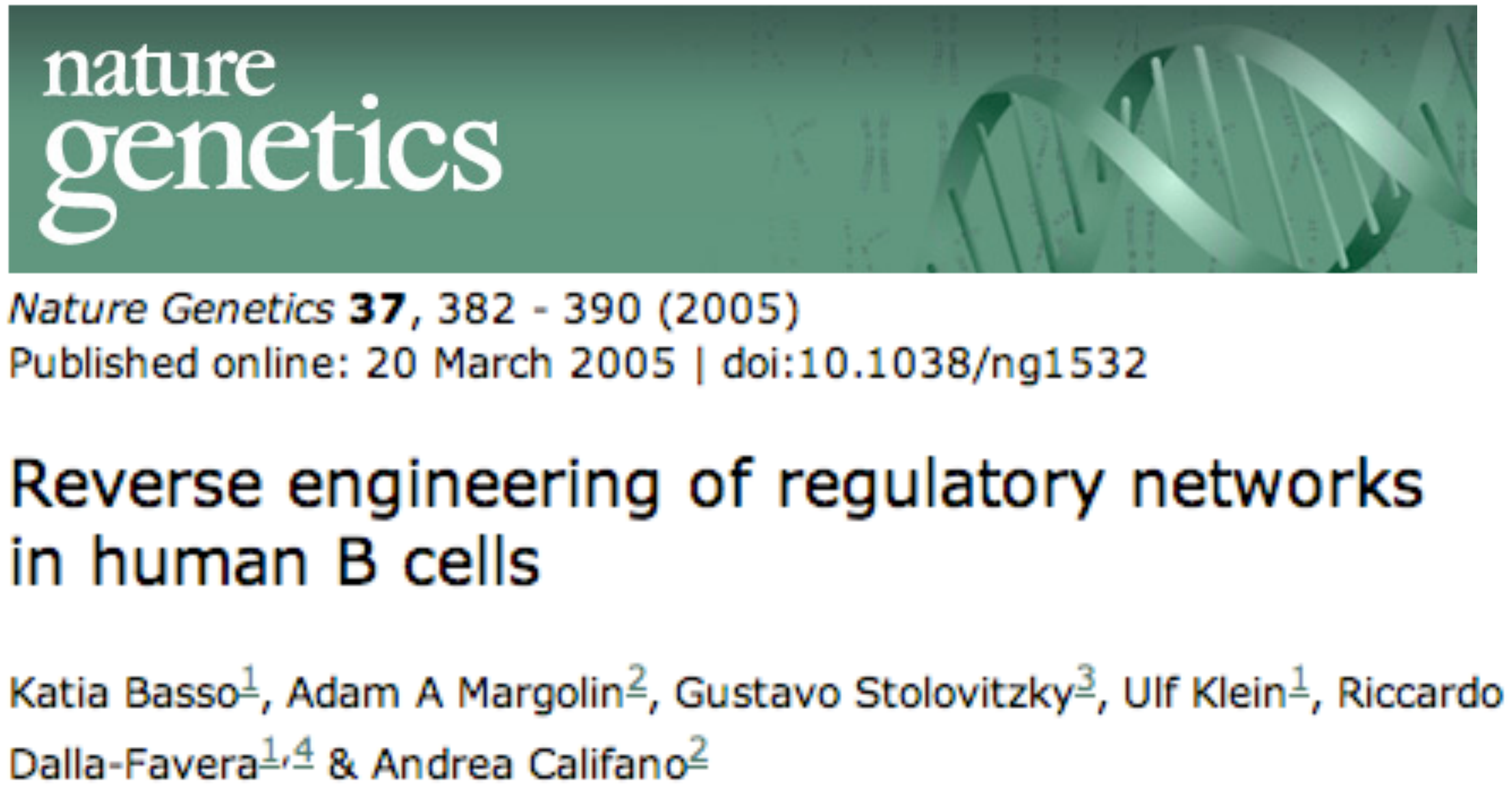
- Choose a measure of association $F(X, Y)$
- Define a threshold value τ_F
- For all pairs of genes (X, Y) compute their association $F(X, Y)$
- Retain associations such that $F(X, Y) > \tau_F$

No direction
Indirect association (confounding variables)



Reverse engineering of gene regulatory nets

ARACNE



nature genetics

Nature Genetics **37**, 382 - 390 (2005)
Published online: 20 March 2005 | doi:10.1038/ng1532

Reverse engineering of regulatory networks in human B cells

Katia Basso[1], Adam A Margolin[2], Gustavo Stolovitzky[3], Ulf Klein[1], Riccardo Dalla-Favera[1,4] & Andrea Califano[2]

ARACNE: An Algorithm for the Reconstruction of Gene Regulatory Networks in a Mammalian Cellular Context

Adam A Margolin, Ilya Nemenman, Katia Basso, Chris Wiggins, Gustavo Stolovitzky, Riccardo Dalla Favera & Andrea Califano

BMC Bioinformatics **7**, Article number: S7 (2006) | Cite this article

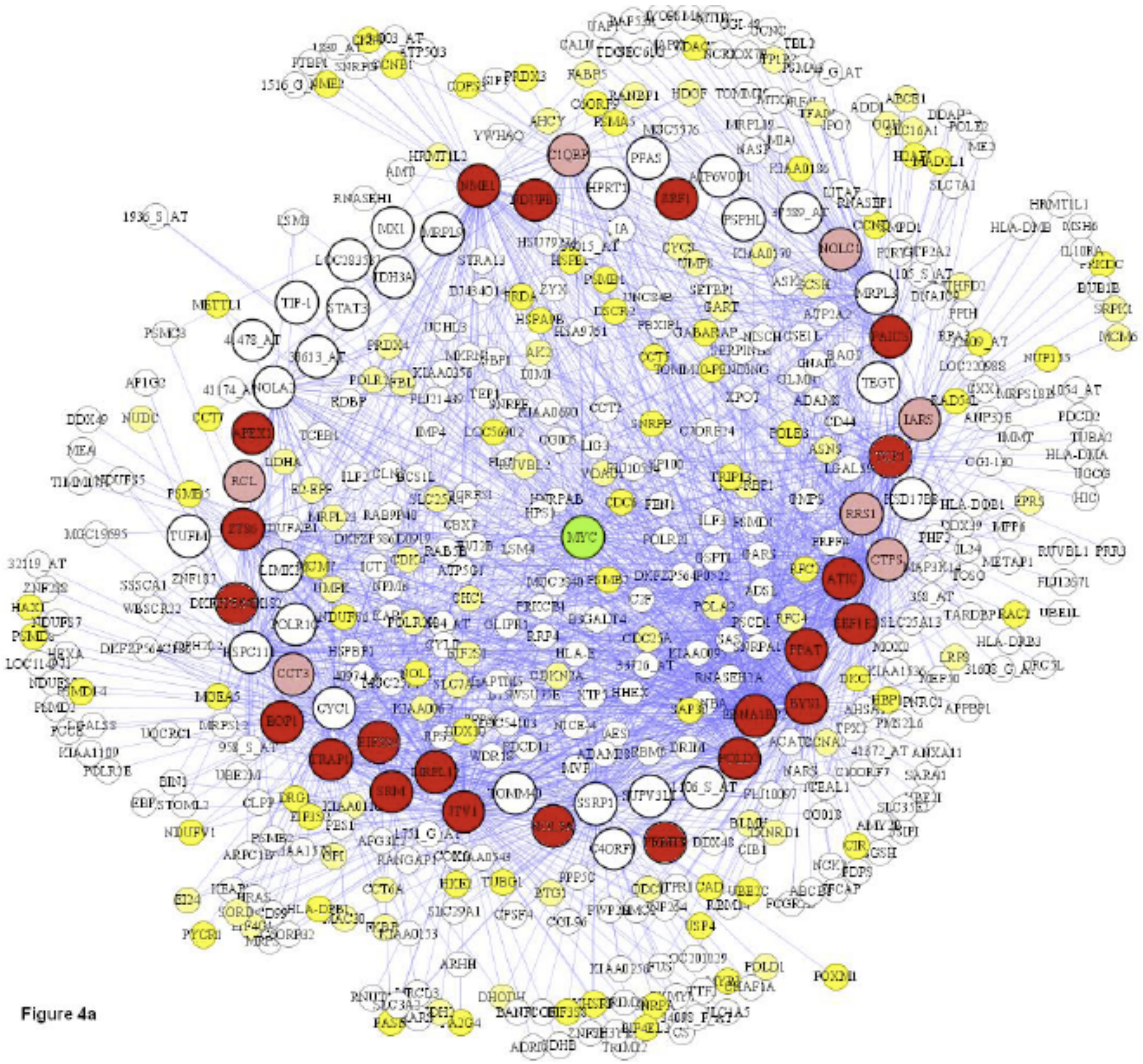
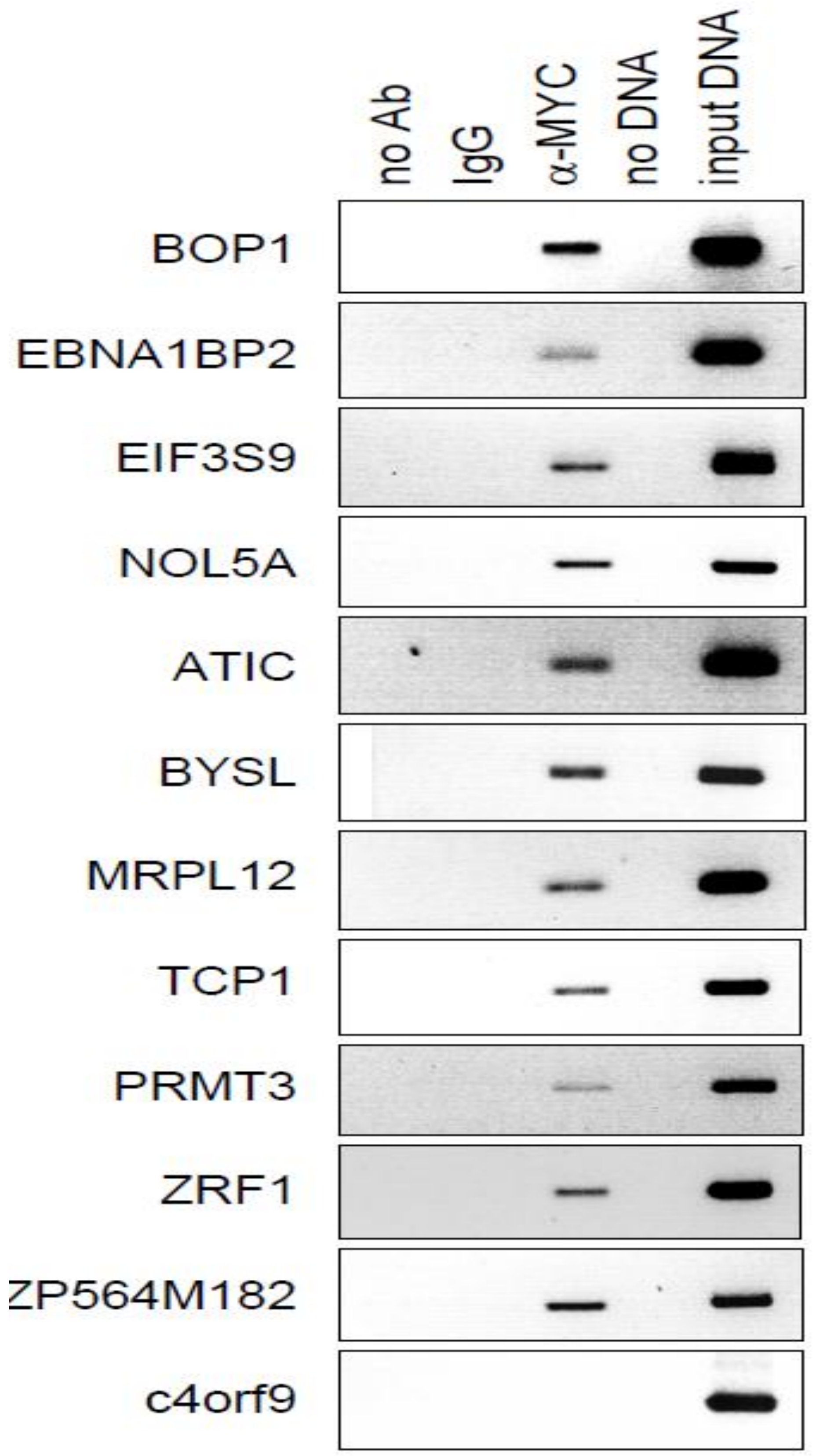


Figure 4a

Reverse engineering of gene regulatory nets

ARACNE



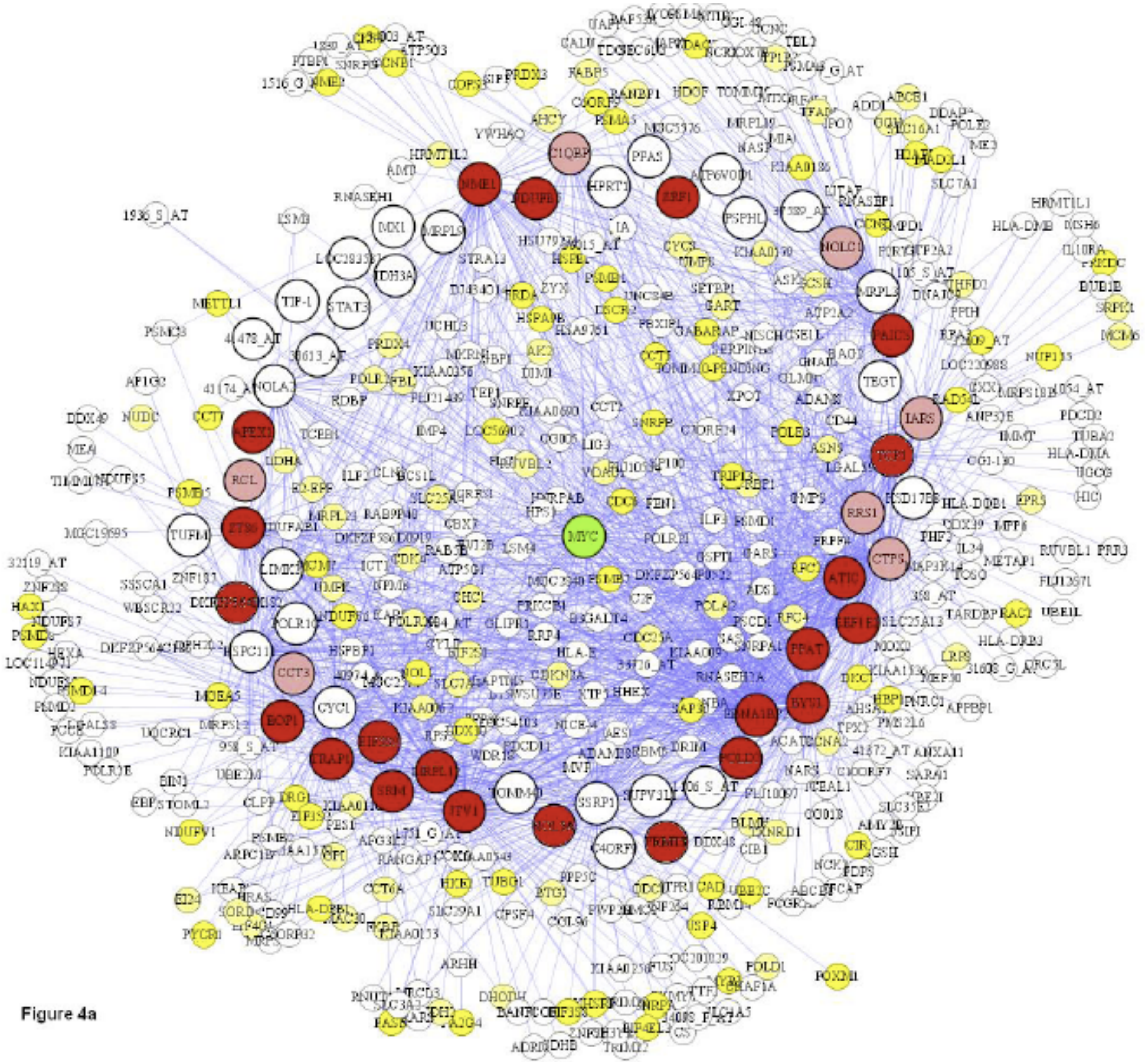


Figure 4a

Reverse engineering of gene regulatory nets

ARACNE

The measure of association $F(X, Y)$ is the Mutual Information between two genes $I(X, Y)$

Reverse engineering of gene regulatory nets

ARACNE

The measure of association $F(X, Y)$ is the Mutual Information between two genes $I(X, Y)$

The threshold value τ_I is computed with a permutation test

Reverse engineering of gene regulatory nets

ARACNE

The measure of association $F(X, Y)$ is the Mutual Information between two genes $I(X, Y)$

The threshold value τ_I is computed with a permutation test

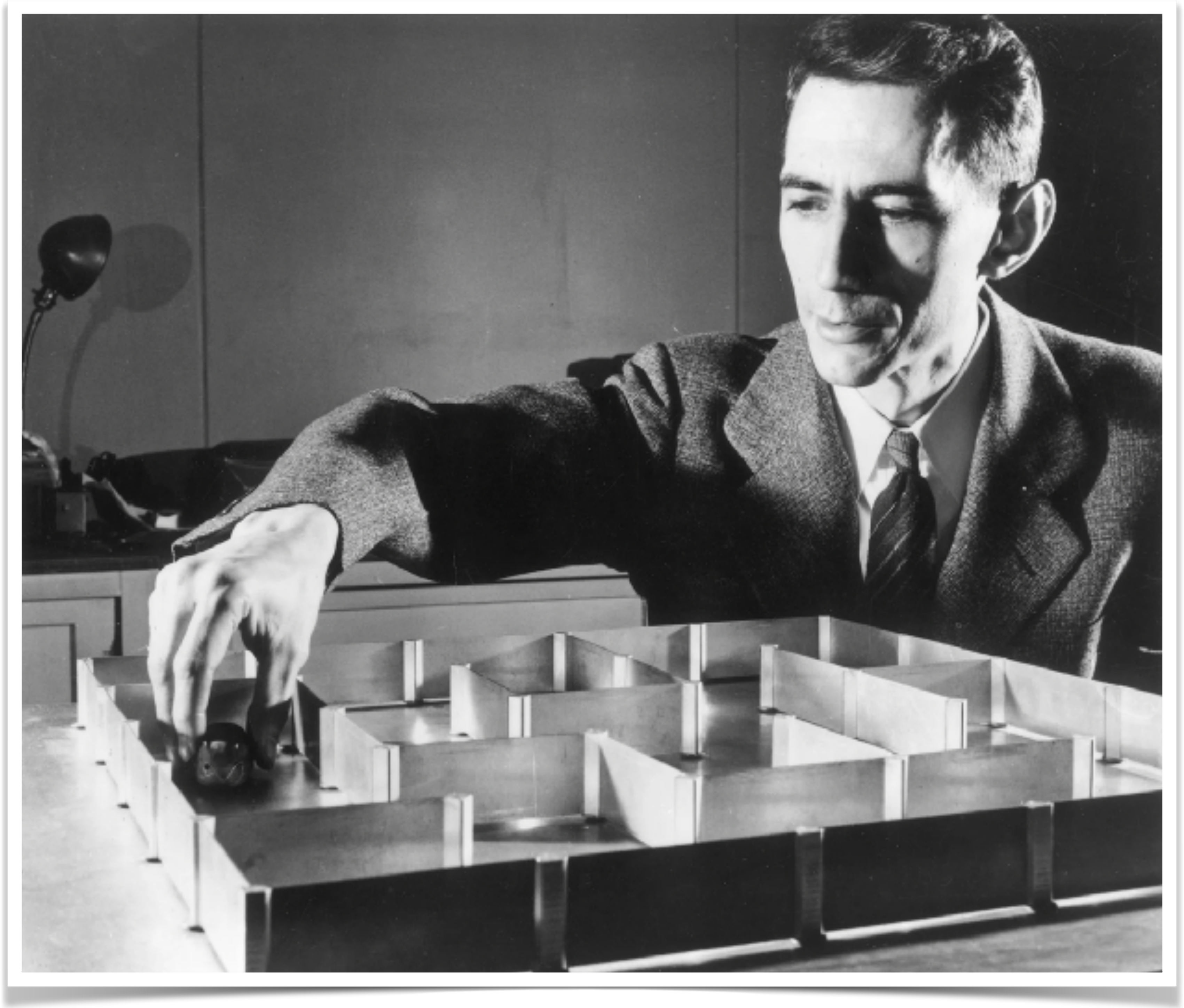
Indirect associations are removed through the Data Process Inequality

Reverse engineering of gene regulatory nets

ARACNE

Information Entropy H

BLA BLA BLA ...

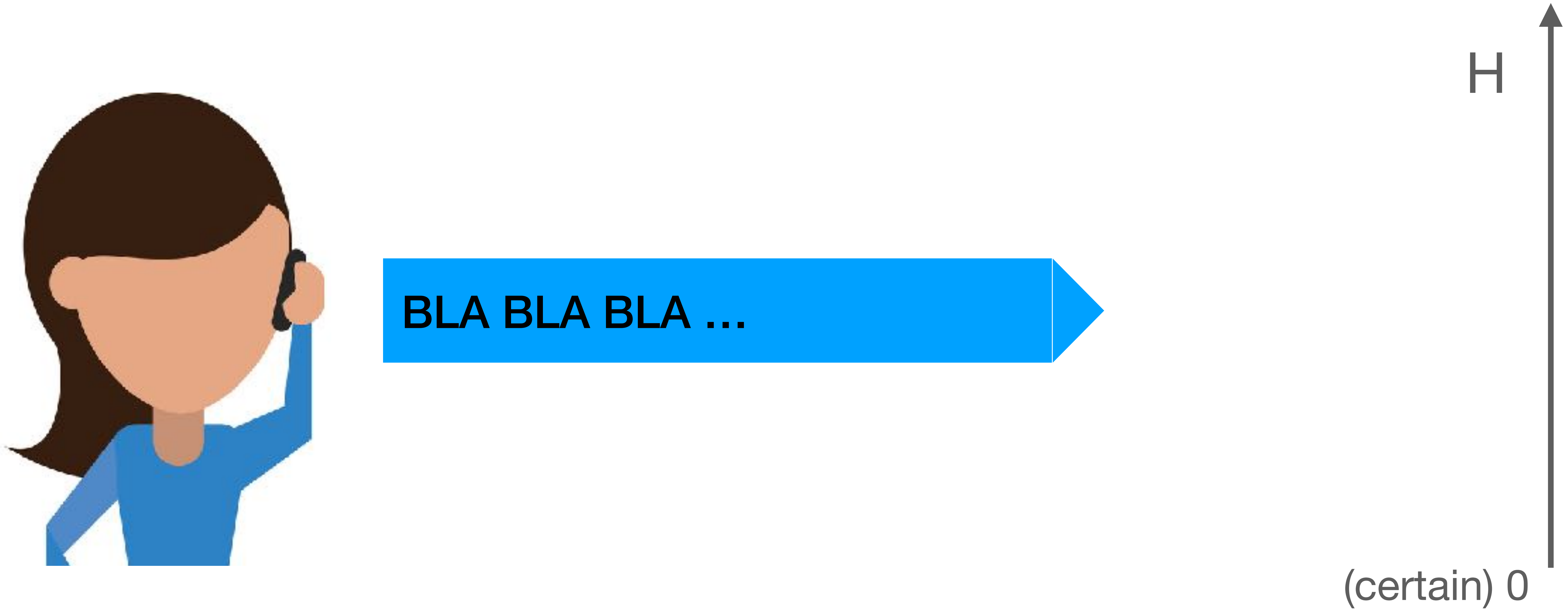


Claude Shannon
A mathematical theory of communication, 1948

Reverse engineering of gene regulatory nets

ARACNE

Information Entropy H



BLA BLA BLA ...

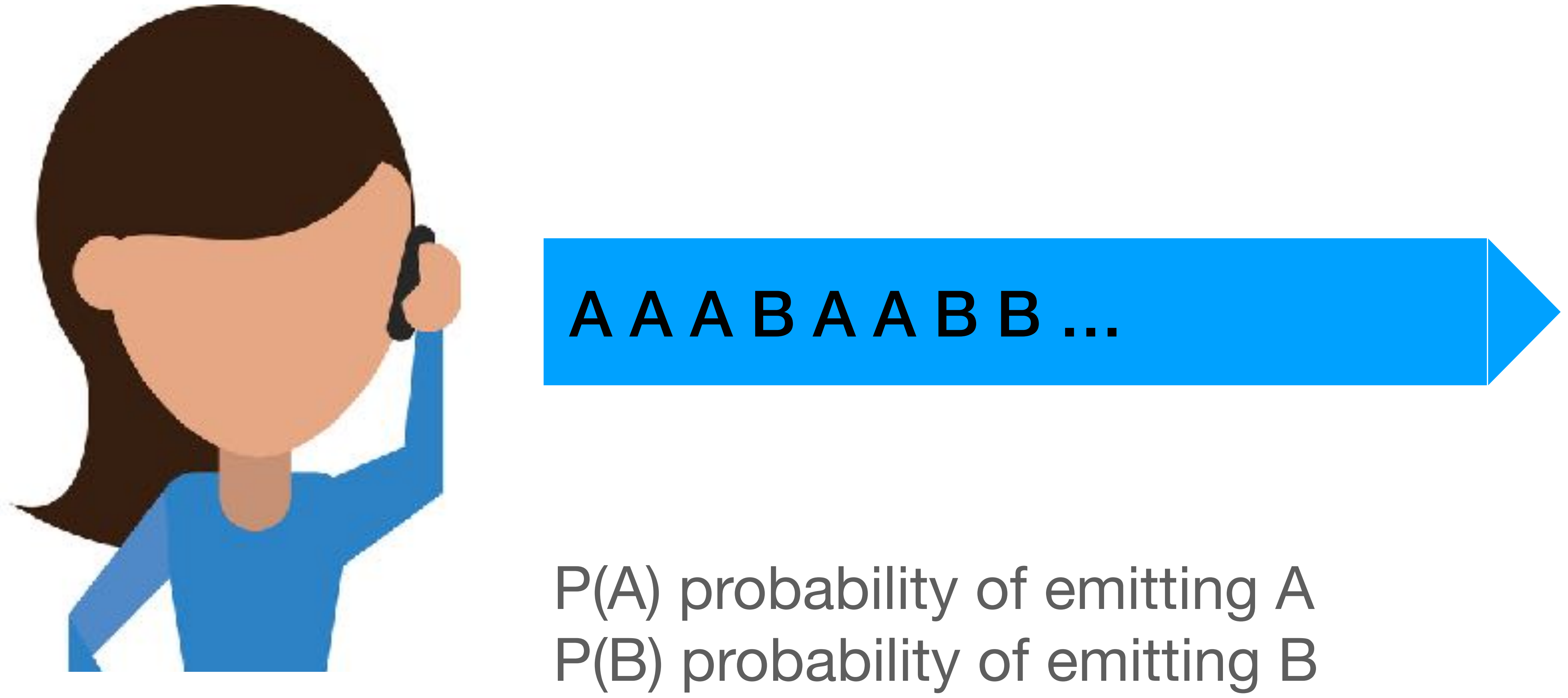
H

(certain) 0

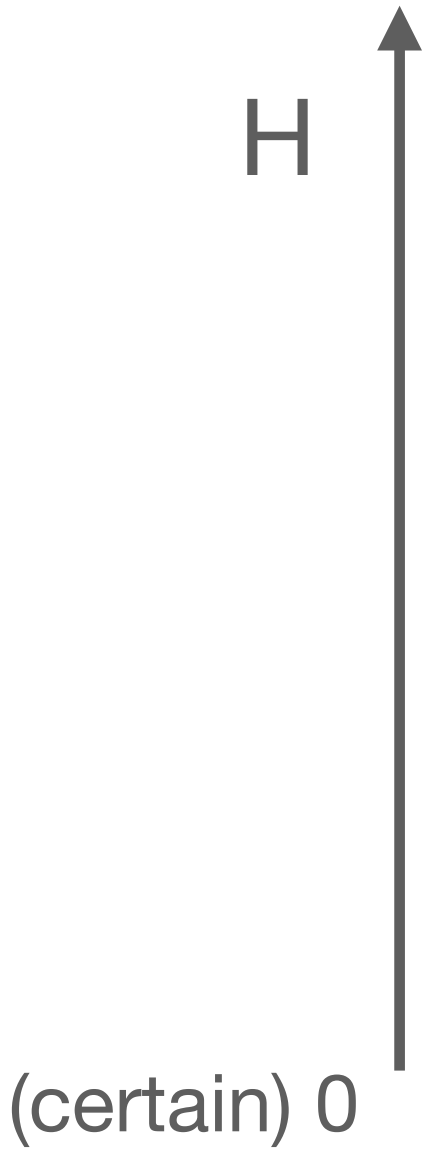
Reverse engineering of gene regulatory nets

ARACNE

Information Entropy H



P(A) probability of emitting A
P(B) probability of emitting B



$$H(Mary) = -P(A)log_2P(A) - P(B)log_2P(B)$$

Reverse engineering of gene regulatory nets

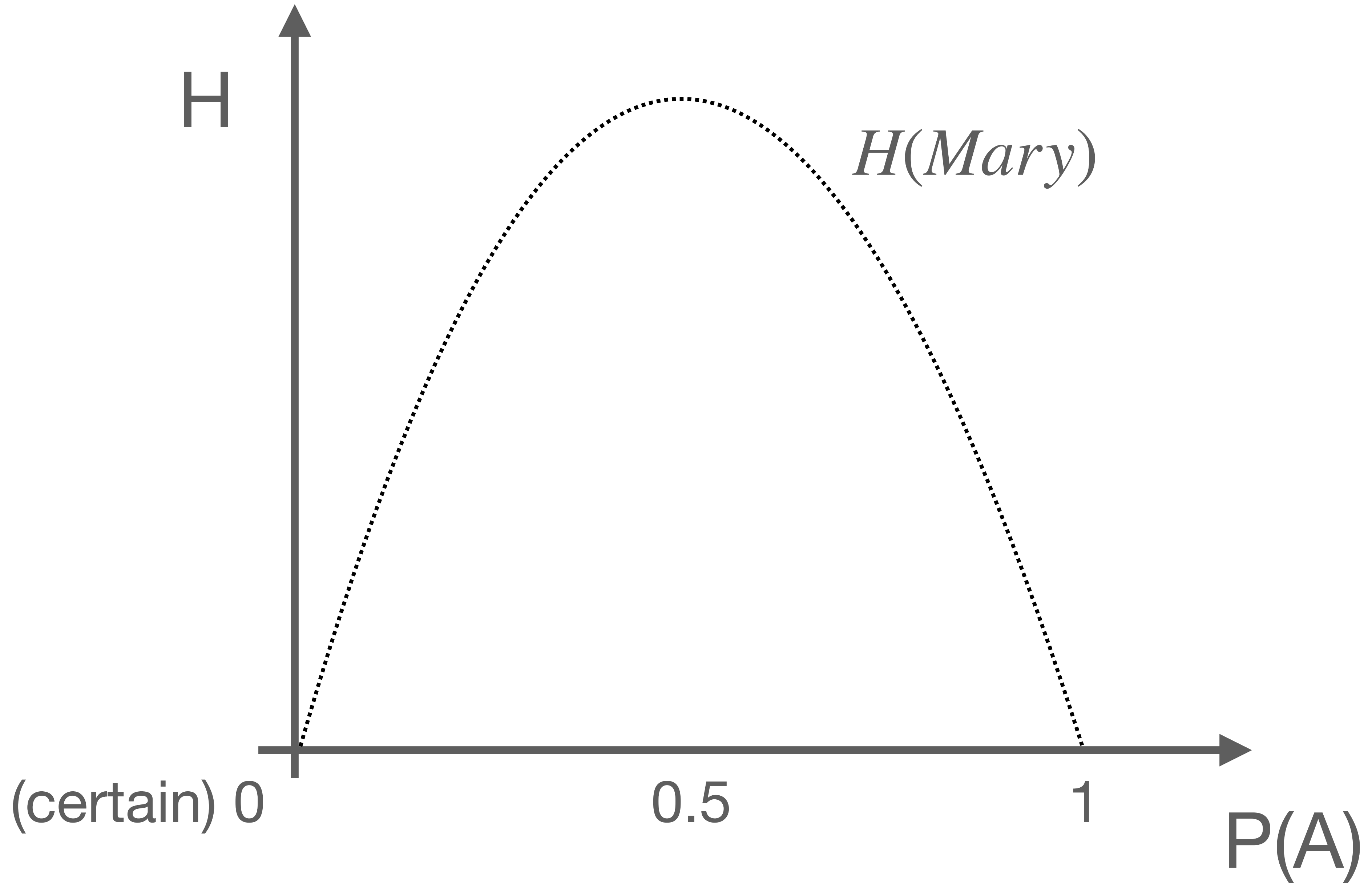
ARACNE

Information Entropy H

A A A B A A B B ...

P(A) probability of emitting A
P(B) probability of emitting B

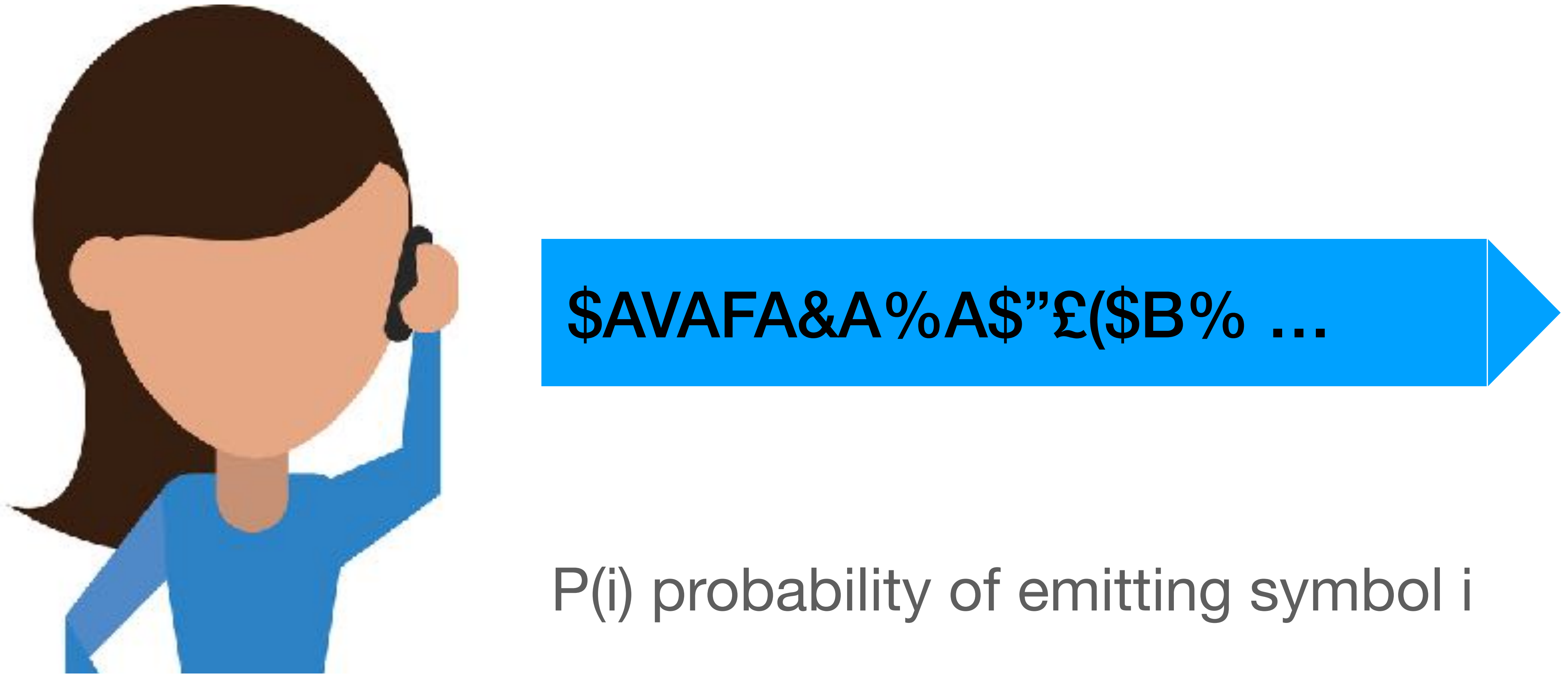
$$H(Mary) = -P(A)log_2P(A) - P(B)log_2P(B)$$



Reverse engineering of gene regulatory nets

ARACNE

Information Entropy H



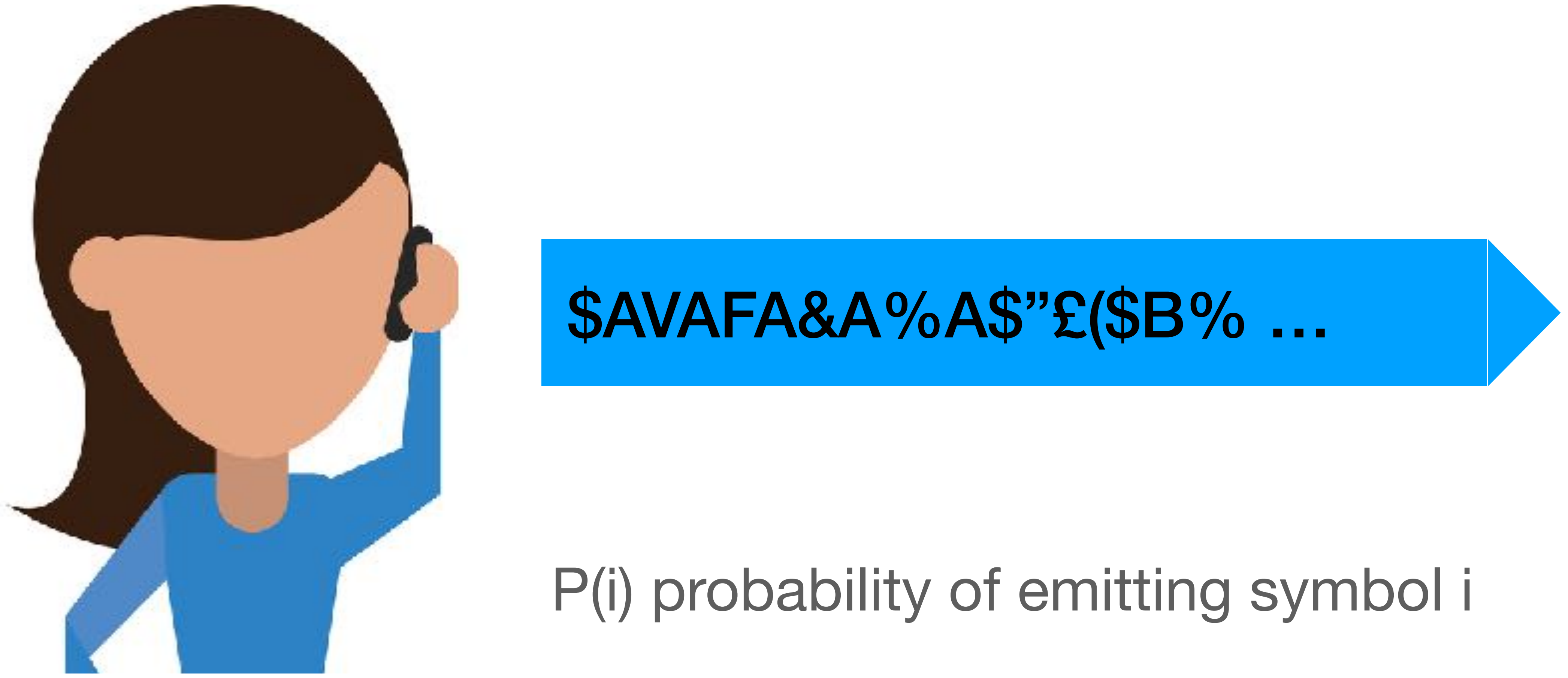
P(i) probability of emitting symbol i

$$H(Mary) = \sum_{i \in S} - P(i) log_2 P(i)$$

Reverse engineering of gene regulatory nets

ARACNE

Information Entropy H



$AVAFA&A%A$”£($B% …

P(i) probability of emitting symbol i

$$H(Mary) = \sum_{i \in S} - P(i) log_2 P(i)$$

- Measures (in bits) of the uncertainty associated with a random variable.
- How much information we learn on average from one instance of the random symbol i

Reverse engineering of gene regulatory nets

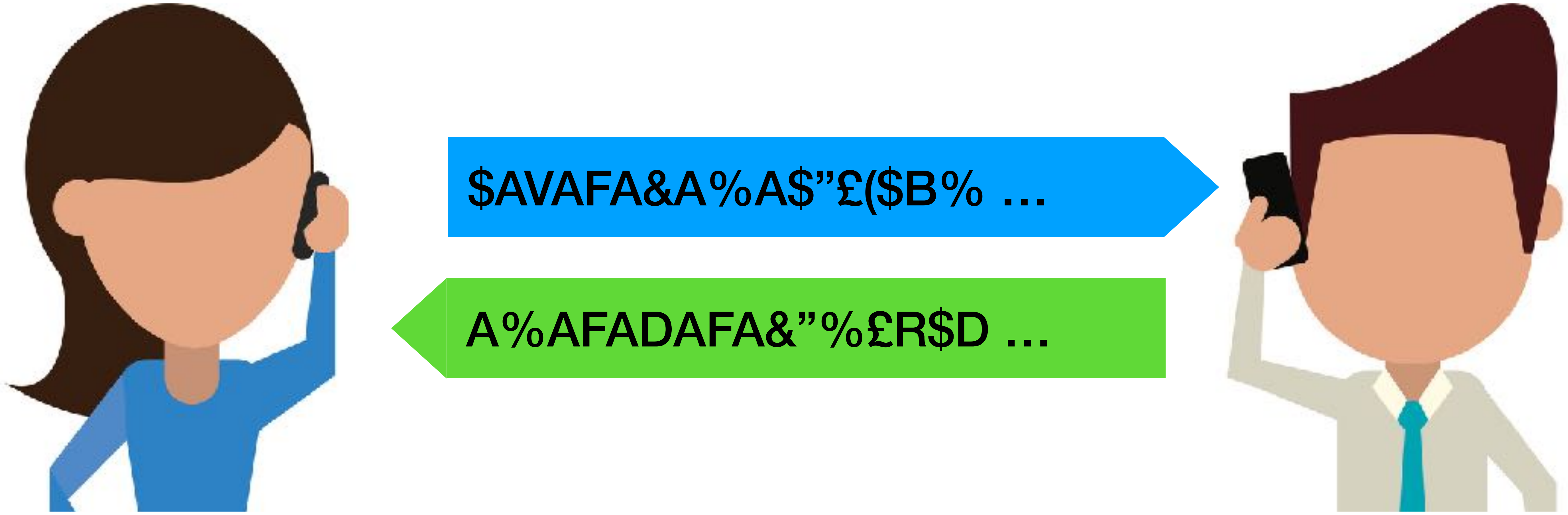
ARACNE

Mutual Information I

Reverse engineering of gene regulatory nets

ARACNE

Mutual Information I



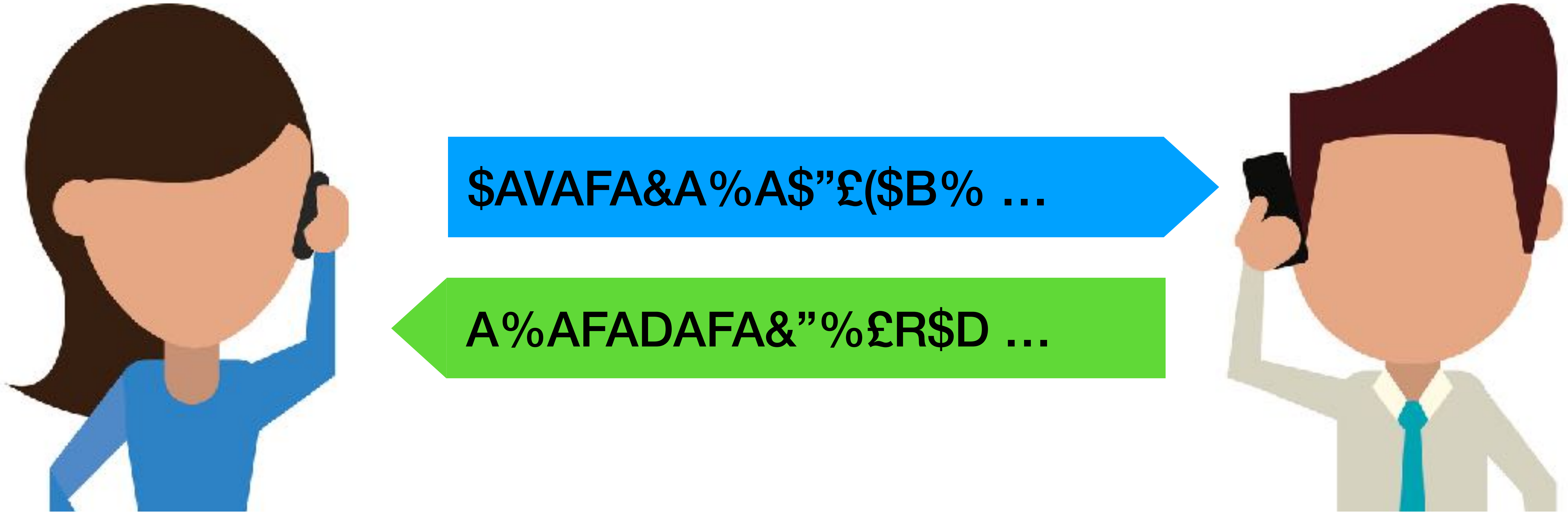
$AVAFA&A%A$”£($B% …

A%AFADAFA&”%£R$D …

Reverse engineering of gene regulatory nets

ARACNE

Mutual Information I



$AVAFA&A%A$”£($B% …

A%AFADAFA&”%£R$D …

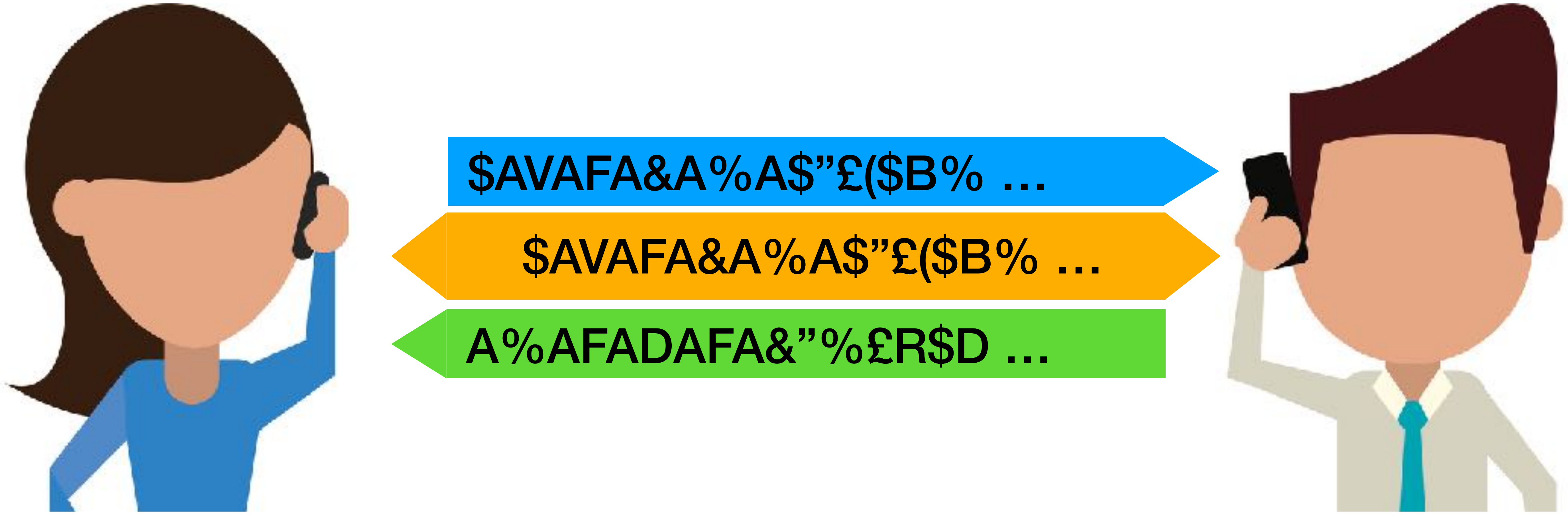
$$H(Mary) = \sum_{i \in S} - P(i) log_2 P(i)$$

$$H(Joe) = \sum_{j \in Z} - P(j) log_2 P(j)$$

Reverse engineering of gene regulatory nets

ARACNE

Mutual Information I



$AVAFA&A%A$"£($B% …

$AVAFA&A%A$"£($B% …

A%AFADAFA&"%£R$D …

$$H(Mary) = \sum_{i \in S} - P(i) log_2 P(i)$$

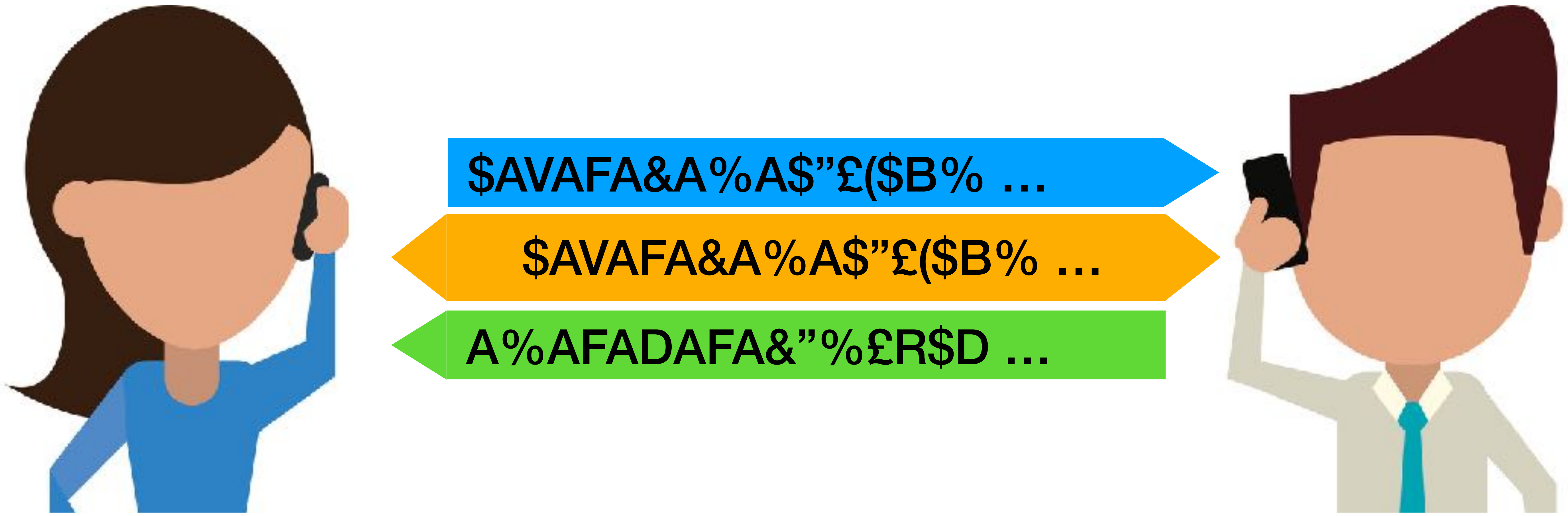
$$H(Joe) = \sum_{j \in Z} - P(j) log_2 P(j)$$

Reverse engineering of gene regulatory nets

ARACNE

Mutual Information I

$$H(Joe\,|\,Mary) = \sum_{i \in Z, j \in S} - P(i\,|\,j) log_2 P(i\,|\,j)$$



$AVAFA&A%A$"£($B% …

$AVAFA&A%A$"£($B% …

A%AFADAFA&"%£R$D …

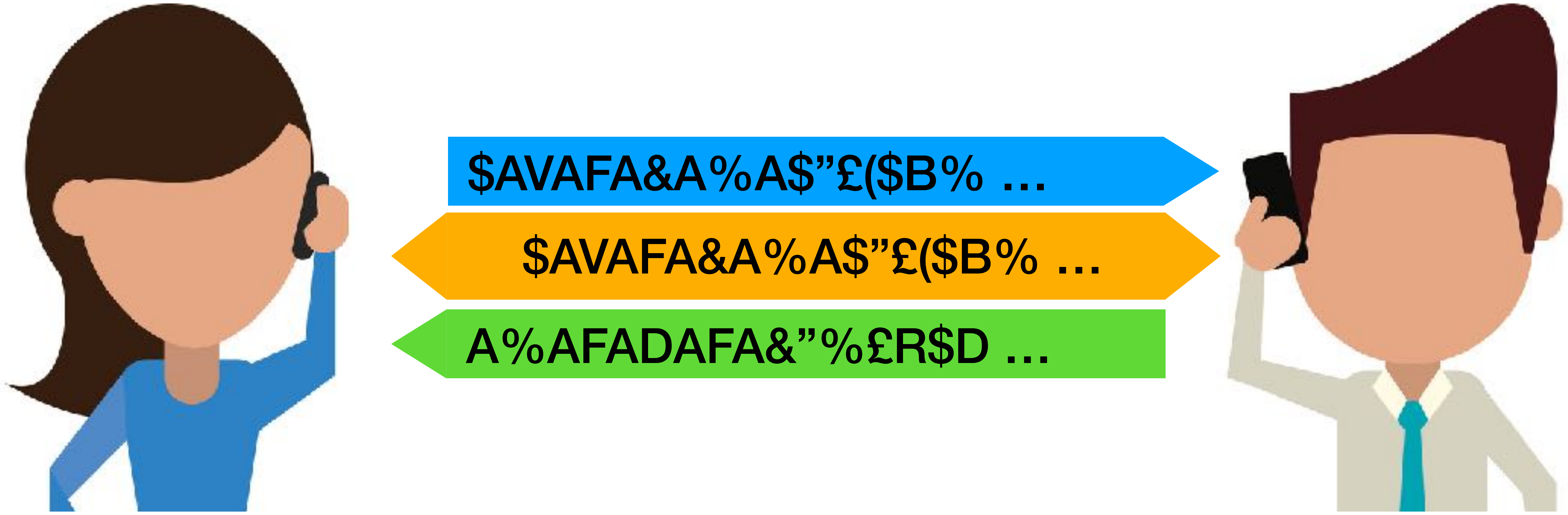
$$H(Mary) = \sum_{i \in S} - P(i) log_2 P(i)$$

$$H(Joe) = \sum_{j \in Z} - P(j) log_2 P(j)$$

Reverse engineering of gene regulatory nets

ARACNE

Mutual Information I



$$H(Joe\,|\,Mary) = \sum_{i\in Z, j\in S} -P(i\,|\,j)log_2P(i\,|\,j)$$

$$I(Joe, Mary) = H(Joe) - H(Joe\,|\,Mary)$$

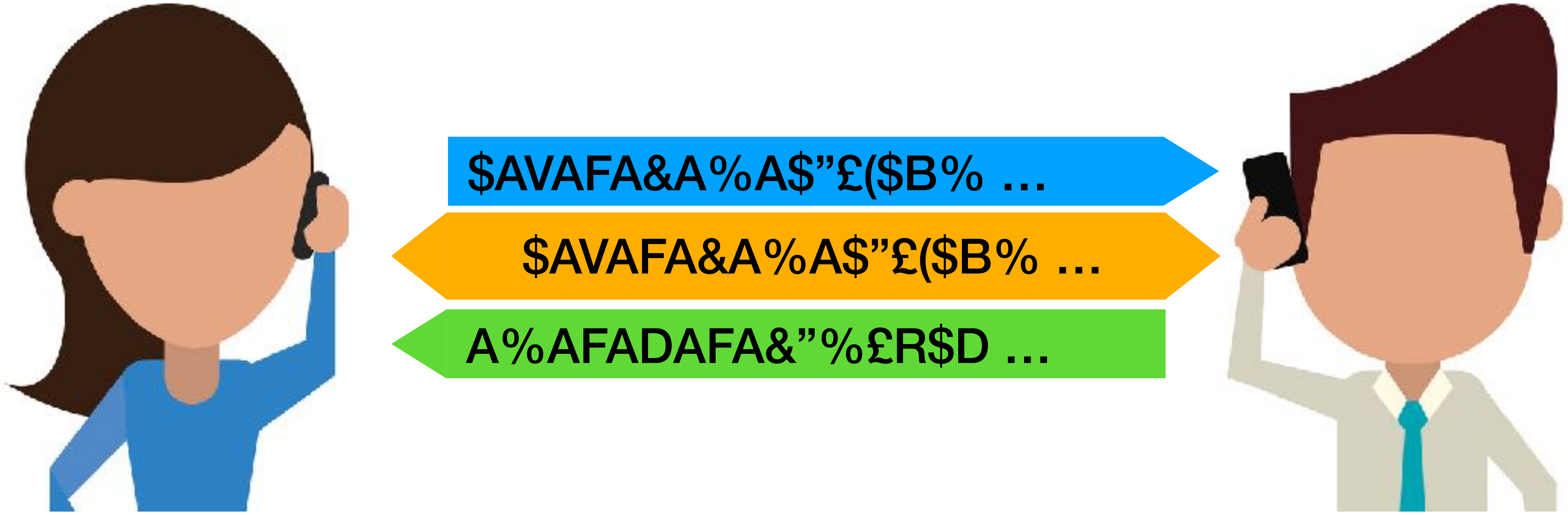
$$H(Mary) = \sum_{i\in S} -P(i)log_2P(i)$$

$$H(Joe) = \sum_{j\in Z} -P(j)log_2P(j)$$

Reverse engineering of gene regulatory nets

ARACNE

Mutual Information I



$AVAFA&A%A$”£($B% …

$AVAFA&A%A$”£($B% …

A%AFADAFA&”%£R$D …

$$H(Joe\,|\,Mary) = \sum_{i \in Z, j \in S} -P(i\,|\,j)log_2P(i\,|\,j)$$

$$\begin{aligned} I(Joe, Mary) &= H(Joe) - H(Joe\,|\,Mary) \\ &= H(Mary) - H(Mary\,|\,Joe) \end{aligned}$$

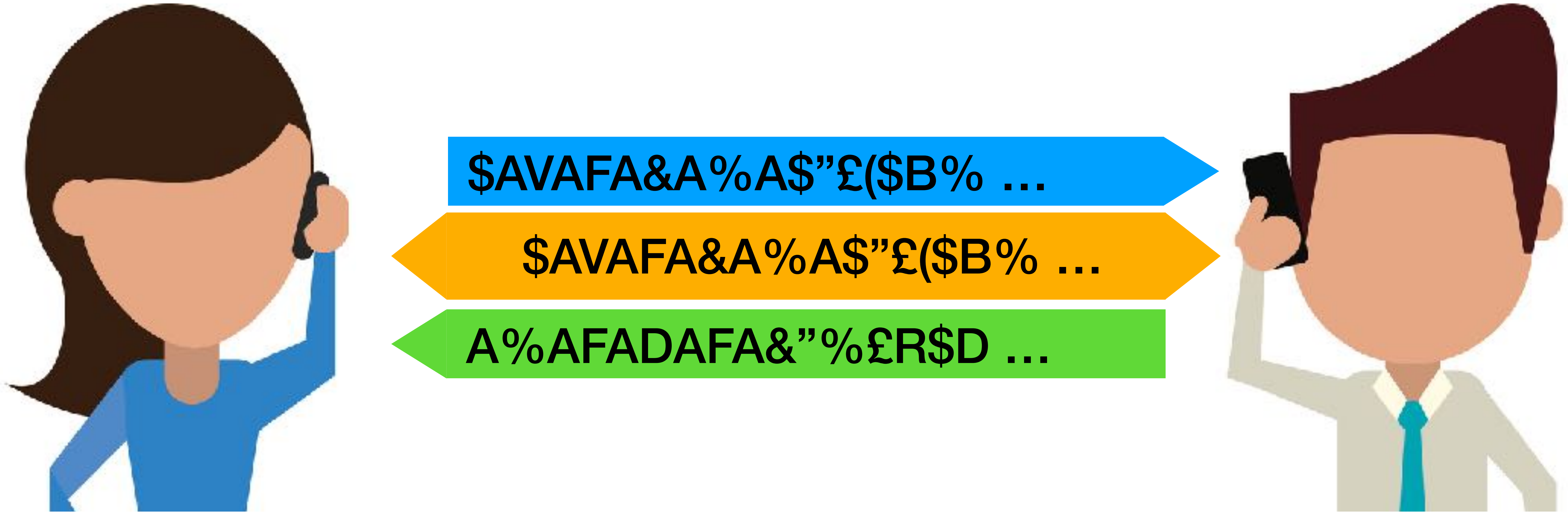
$$H(Mary) = \sum_{i \in S} -P(i)log_2P(i)$$

$$H(Joe) = \sum_{j \in Z} -P(j)log_2P(j)$$

Reverse engineering of gene regulatory nets

ARACNE

Mutual Information I



$AVAFA&A%A$"£($B% …

$AVAFA&A%A$"£($B% …

A%AFADAFA&"%£R$D …

$$H(Joe\,|\,Mary) = \sum_{i \in Z, j \in S} -P(i\,|\,j)log_2P(i\,|\,j)$$

$$\begin{aligned} I(Joe, Mary) &= H(Joe) - H(Joe\,|\,Mary) \\ &= H(Mary) - H(Mary\,|\,Joe) \\ &= I(Mary, Joe) \end{aligned}$$

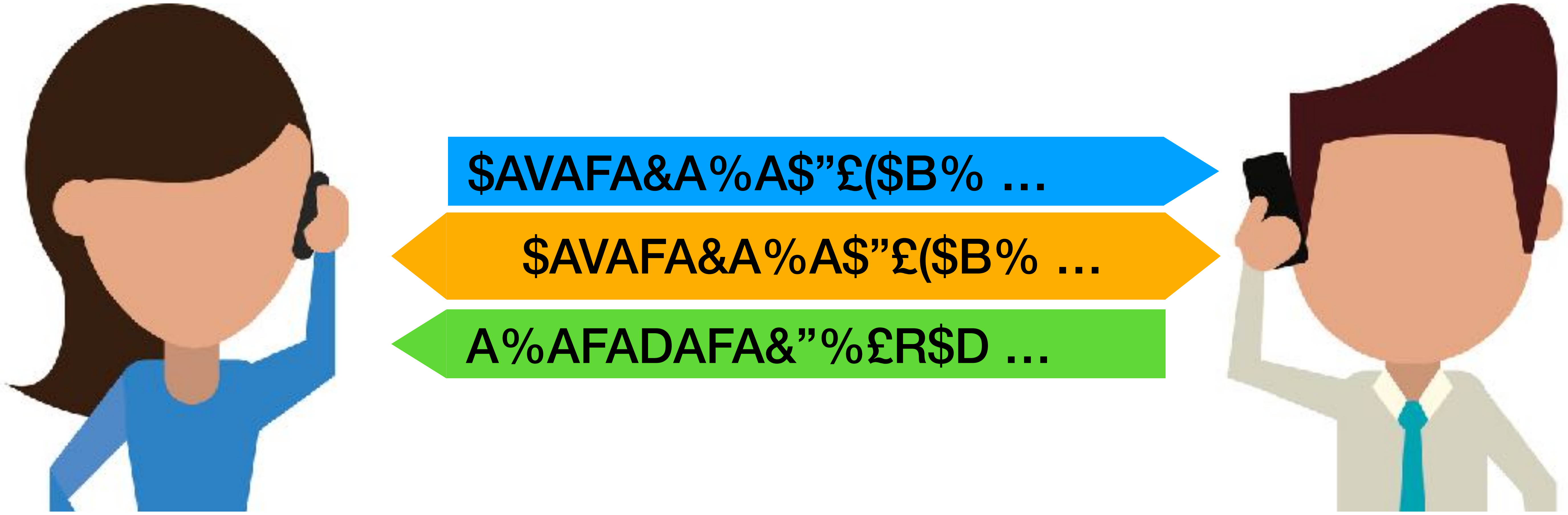
$$H(Mary) = \sum_{i \in S} -P(i)log_2P(i)$$

$$H(Joe) = \sum_{j \in Z} -P(j)log_2P(j)$$

Reverse engineering of gene regulatory nets

ARACNE

Mutual Information I



$AVAFA&A%A$”£($B% …

$AVAFA&A%A$”£($B% …

A%AFADAFA&”%£R$D …

$$H(Joe\,|\,Mary) = \sum_{i \in Z, j \in S} -P(i\,|\,j) log_2 P(i\,|\,j)$$

$$\begin{aligned} I(Joe, Mary) &= H(Joe) - H(Joe\,|\,Mary) \\ &= H(Mary) - H(Mary\,|\,Joe) \\ &= I(Mary, Joe) \\ &= \sum_{i \in S} \sum_{j \in Z} P(i, j) log_2 \frac{P(i, j)}{P(i)P(j)} \end{aligned}$$

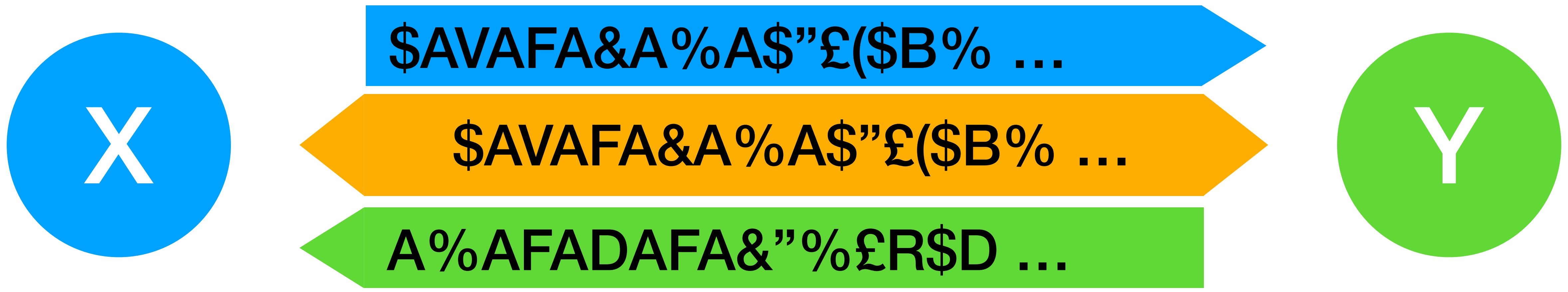
$$H(Mary) = \sum_{i \in S} -P(i) log_2 P(i)$$

$$H(Joe) = \sum_{j \in Z} -P(j) log_2 P(j)$$

Reverse engineering of gene regulatory nets

ARACNE

Mutual Information I between
two genes X and Y


X

$AVAFA&A%A$”£($B% …

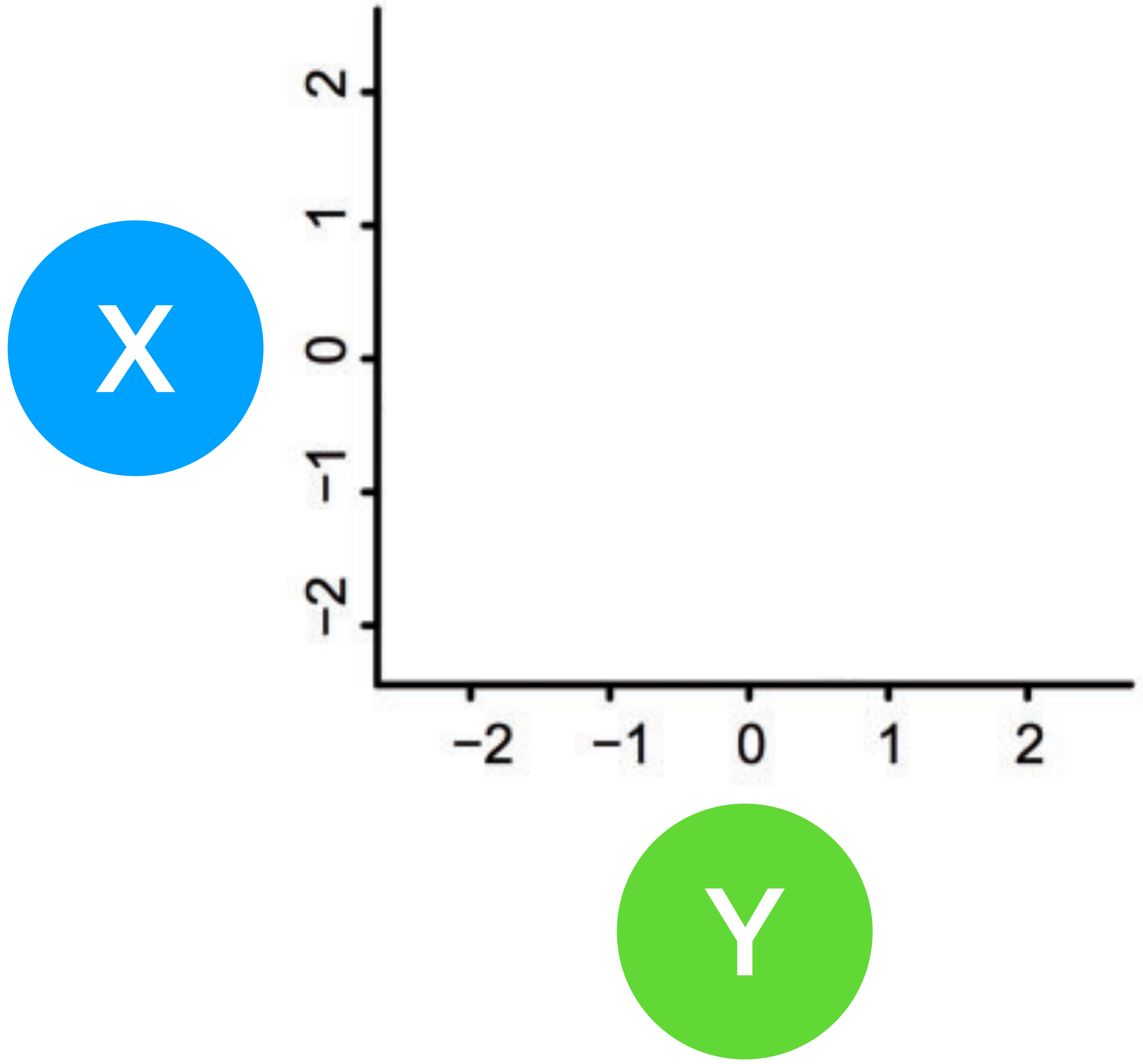
$AVAFA&A%A$”£($B% …

A%AFADAFA&”%£R$D …

Y

Reverse engineering of gene regulatory nets

ARACNE

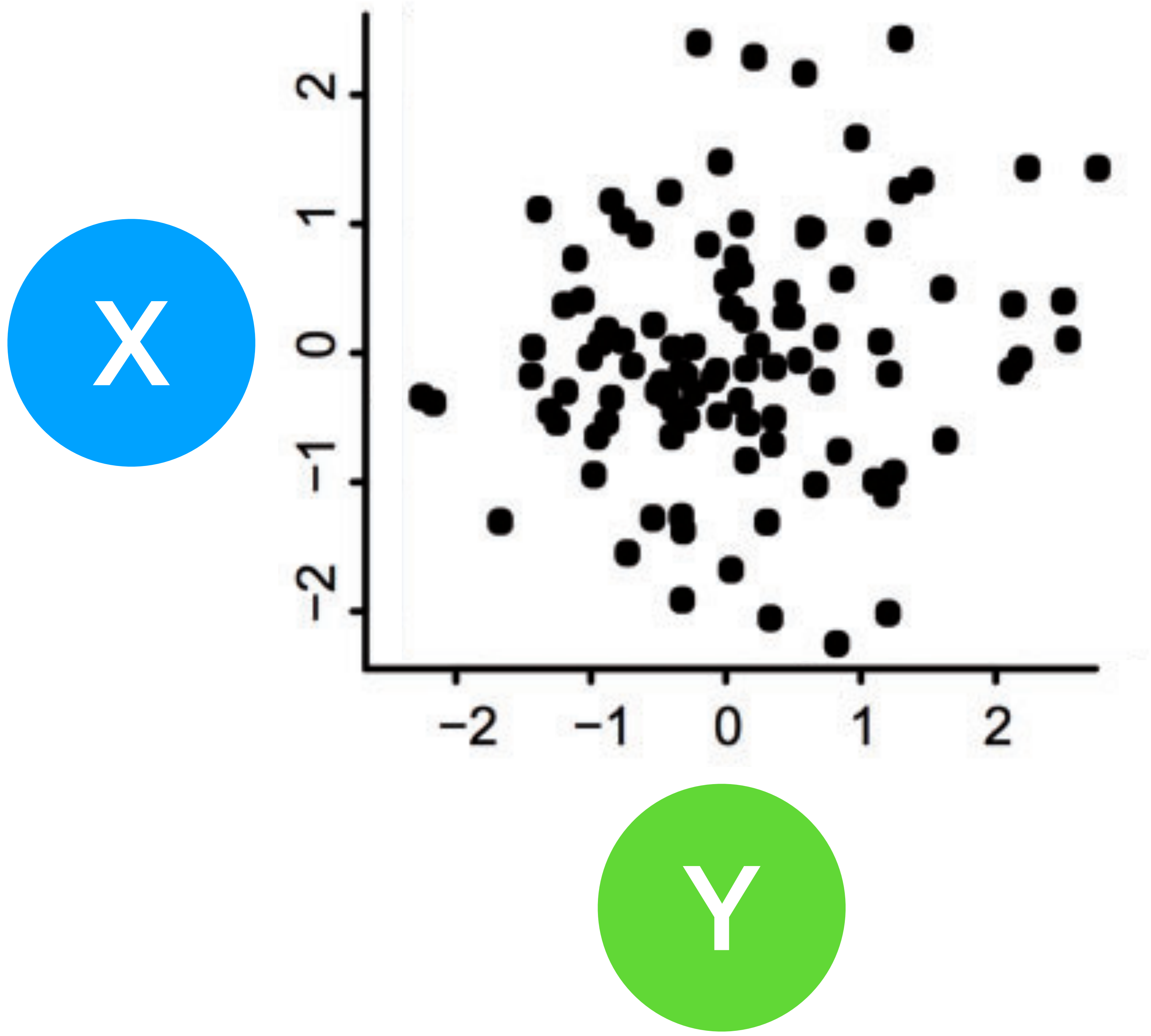


Measure the information that genes X and Y share

$$I(X,Y) = \int\int F_{X,Y}(x,y) log_2 \frac{F_{X,Y}(x,y)}{F_X(x)F_Y(y)}$$

Reverse engineering of gene regulatory nets

ARACNE

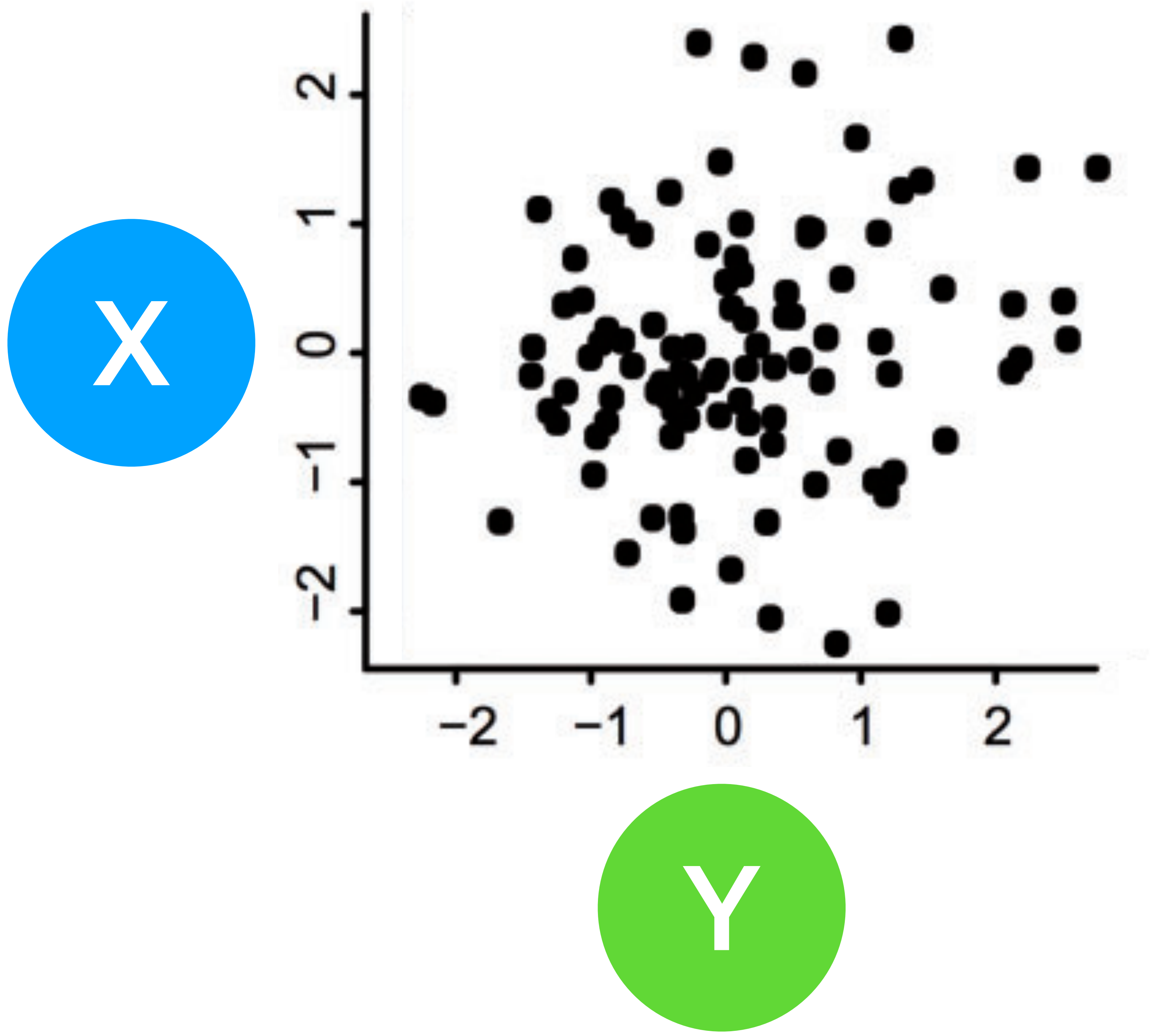


Measure the information that genes X and Y share

$$I(X,Y) = \int\int F_{X,Y}(x,y) log_2 \frac{F_{X,Y}(x,y)}{F_X(x)F_Y(y)}$$

Reverse engineering of gene regulatory nets

ARACNE



Measure the information that genes X and Y share

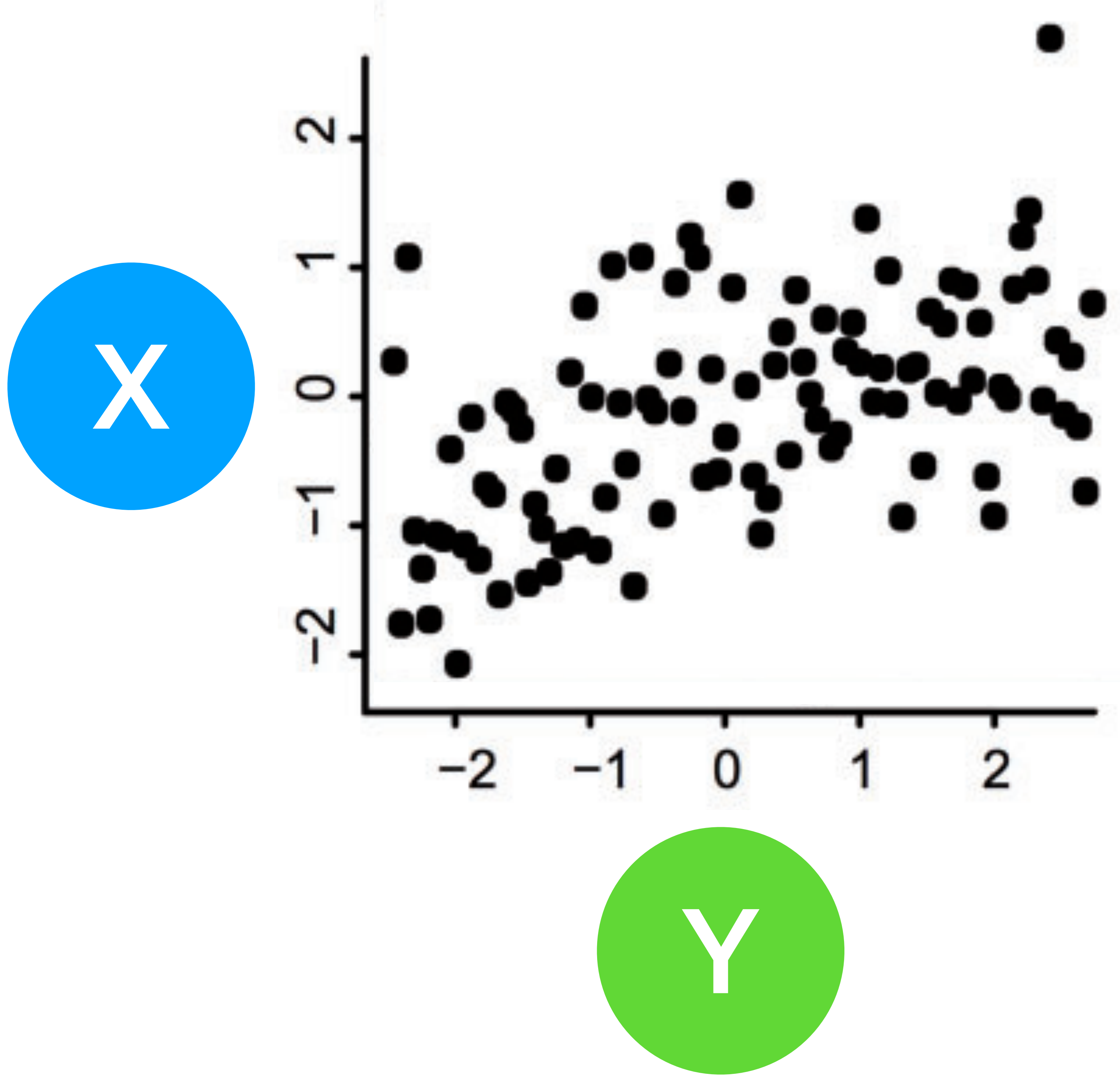
$$I(X,Y) = \int\int F_{X,Y}(x,y) log_2 \frac{F_{X,Y}(x,y)}{F_X(x)F_Y(y)}$$

Association	**Pearson**	**Mutual Information**
Independent	0.51	0.51

Siqueira Santos et al. Briefing in Bioinformatics, 2013

Reverse engineering of gene regulatory nets

ARACNE



Measure the information that genes X and Y share

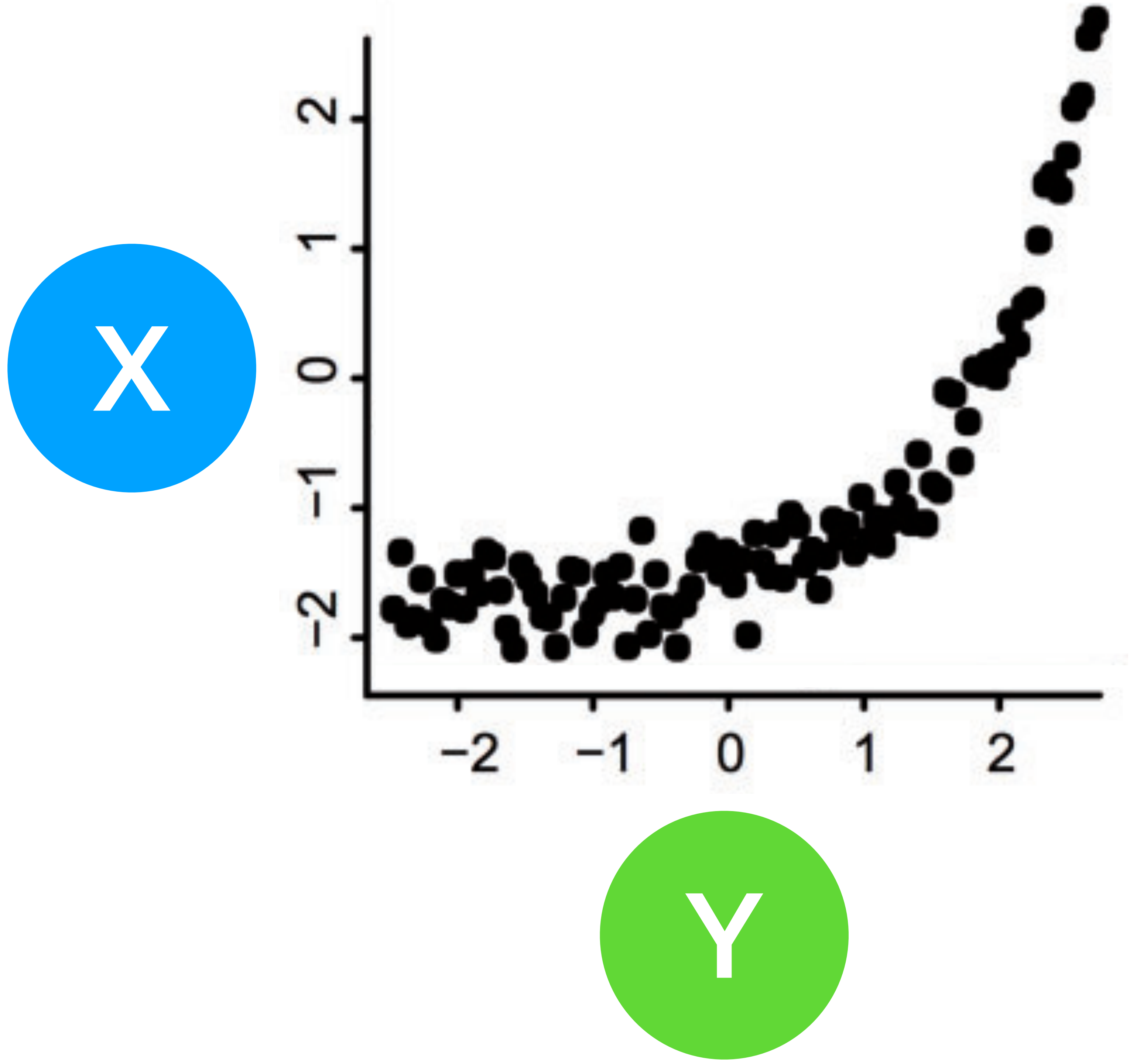
$$I(X,Y) = \int\int F_{X,Y}(x,y) log_2 \frac{F_{X,Y}(x,y)}{F_X(x)F_Y(y)}$$

Association	Pearson	Mutual Information
Independent	0.51	0.51
Linear	1.00	0.79

Siqueira Santos et al. Briefing in Bioinformatics, 2013

Reverse engineering of gene regulatory nets

ARACNE



Measure the information that genes X and Y share

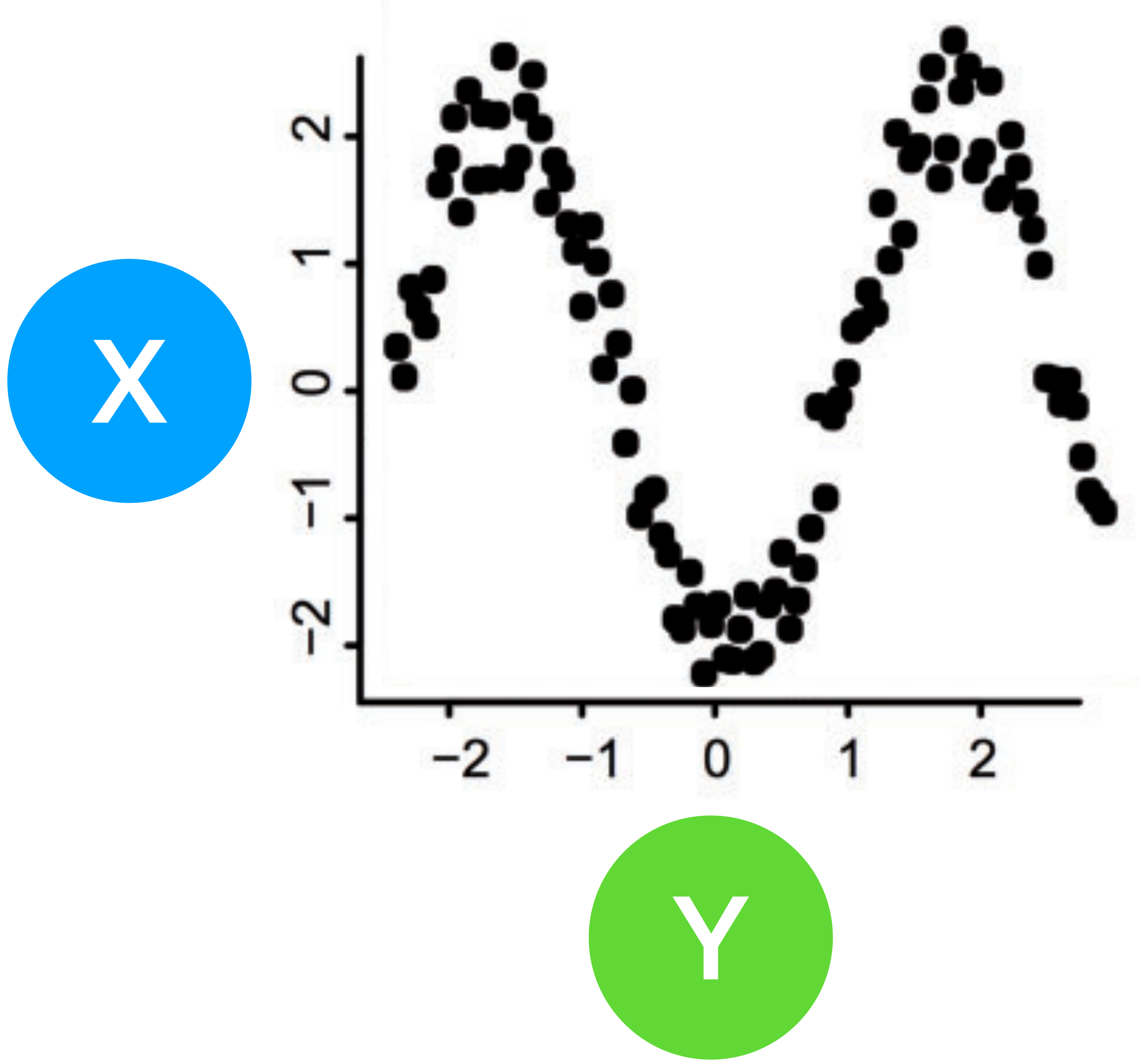
$$I(X,Y) = \int\int F_{X,Y}(x,y) log_2 \frac{F_{X,Y}(x,y)}{F_X(x)F_Y(y)}$$

Association	Pearson	Mutual Information
Independent	0.51	0.51
Linear	1.00	0.79
Exponential	0.99	0.90

Siqueira Santos et al. Briefing in Bioinformatics, 2013

Reverse engineering of gene regulatory nets

ARACNE



Measure the information that genes X and Y share

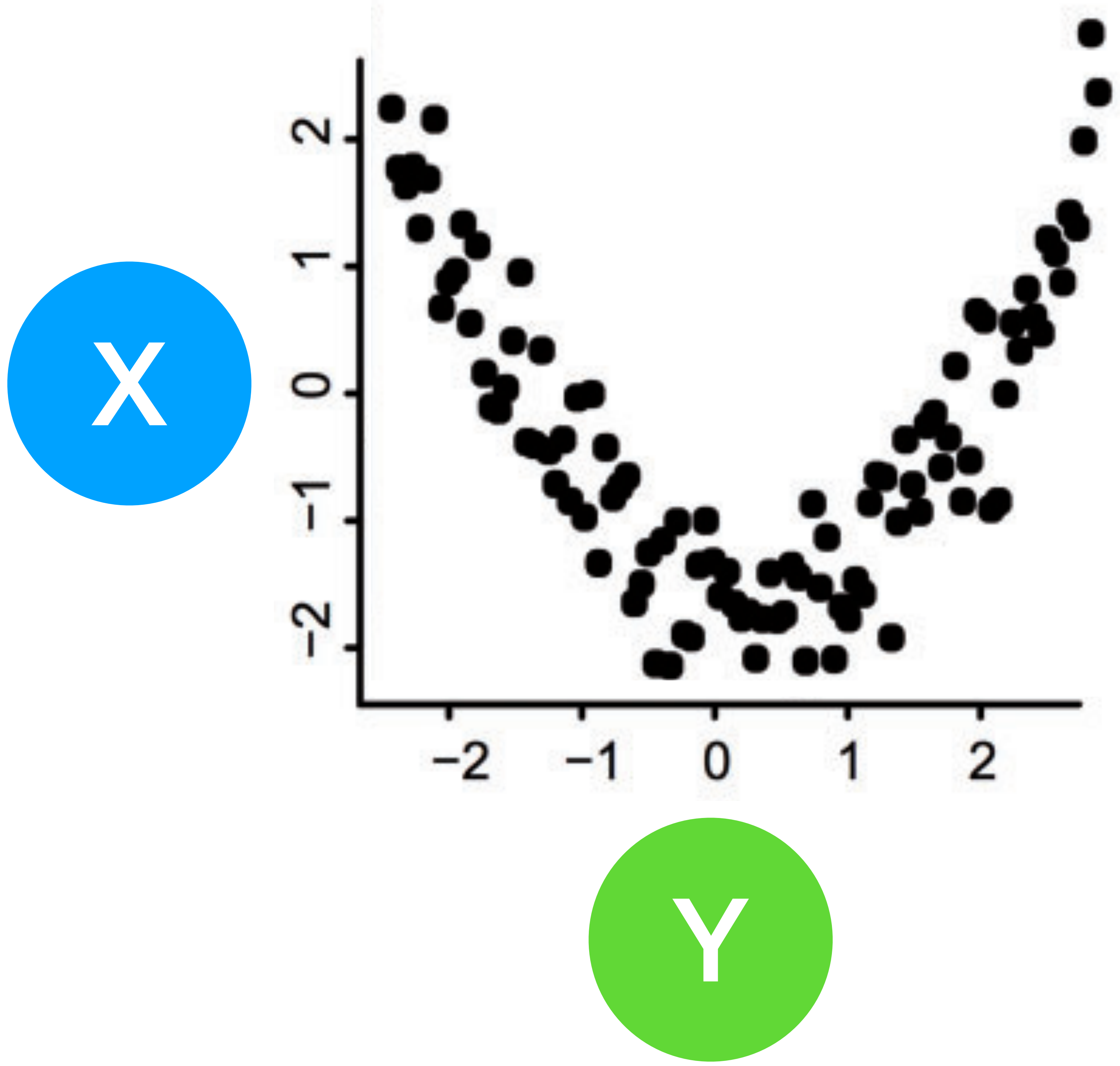
$$I(X,Y) = \int\int F_{X,Y}(x,y)\,log_2 \frac{F_{X,Y}(x,y)}{F_X(x)F_Y(y)}$$

Association	**Pearson**	**Mutual Information**
Independent	0.51	0.51
Linear	1.00	0.79
Exponential	0.99	0.90
Bistable	0.40	1.00

Siqueira Santos et al. Briefing in Bioinformatics, 2013

Reverse engineering of gene regulatory nets

ARACNE



Measure the information that genes X and Y share

$$I(X, Y) = \int\int F_{X,Y}(x, y) log_2 \frac{F_{X,Y}(x, y)}{F_X(x)F_Y(y)}$$

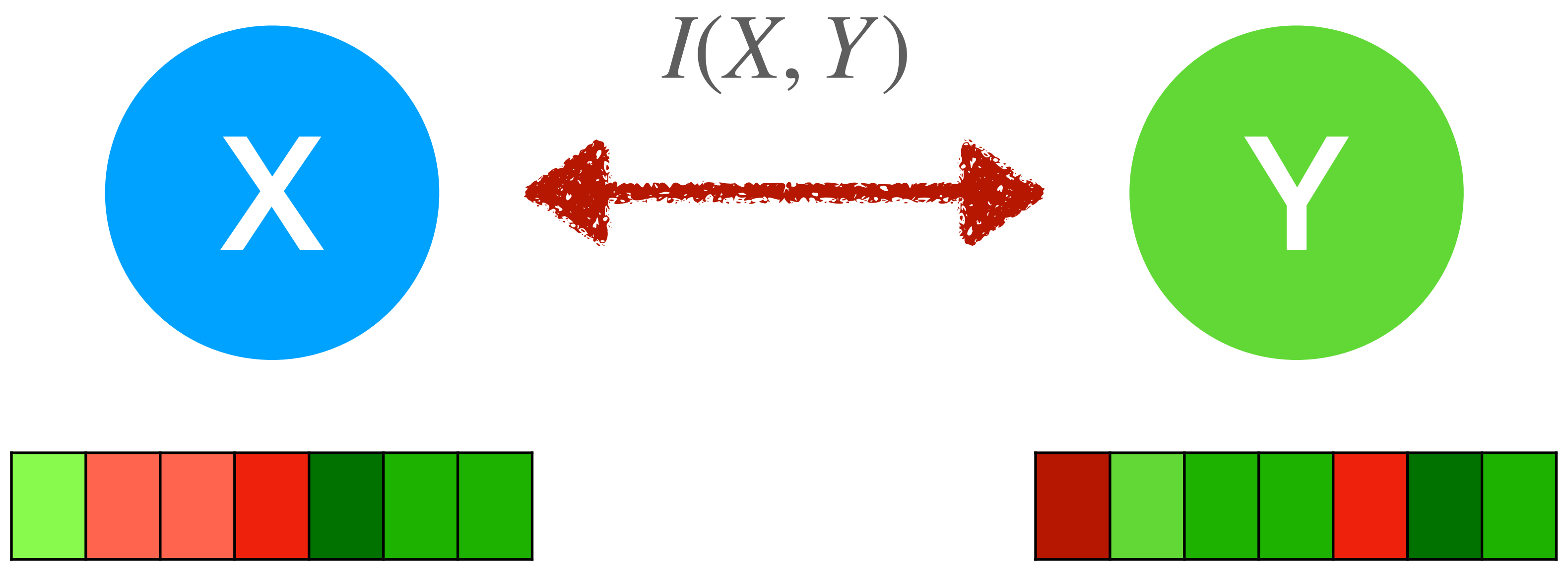
Association	**Pearson**	**Mutual Information**
Independent	0.51	0.51
Linear	1.00	0.79
Exponential	0.99	0.90
Bistable	0.40	1.00
Quadratic	0.21	1.00

Siqueira Santos et al. Briefing in Bioinformatics, 2013

Reverse engineering of gene regulatory nets

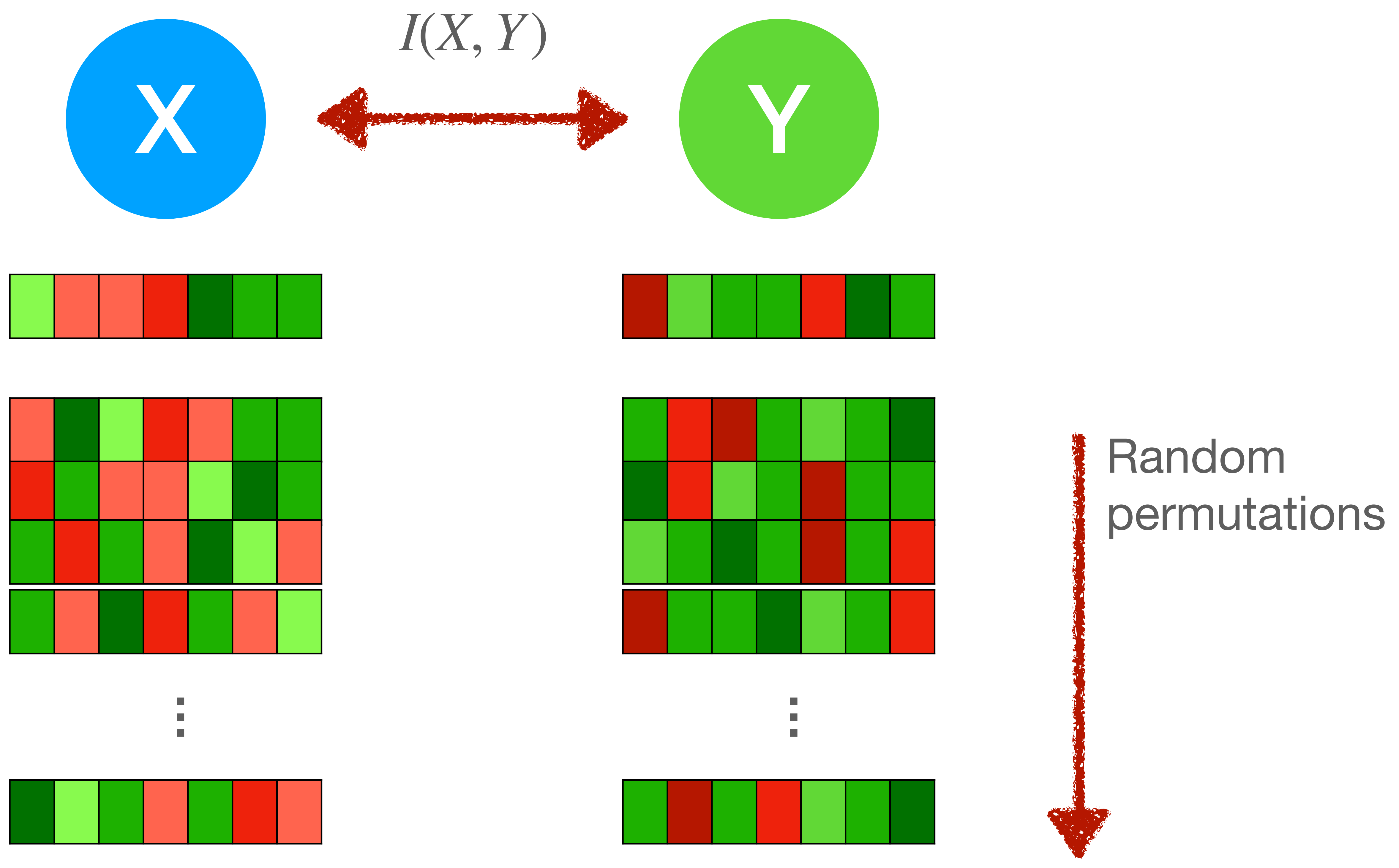
ARACNE

The threshold value τ_F is computed with a permutation test



Reverse engineering of gene regulatory nets
ARACNE

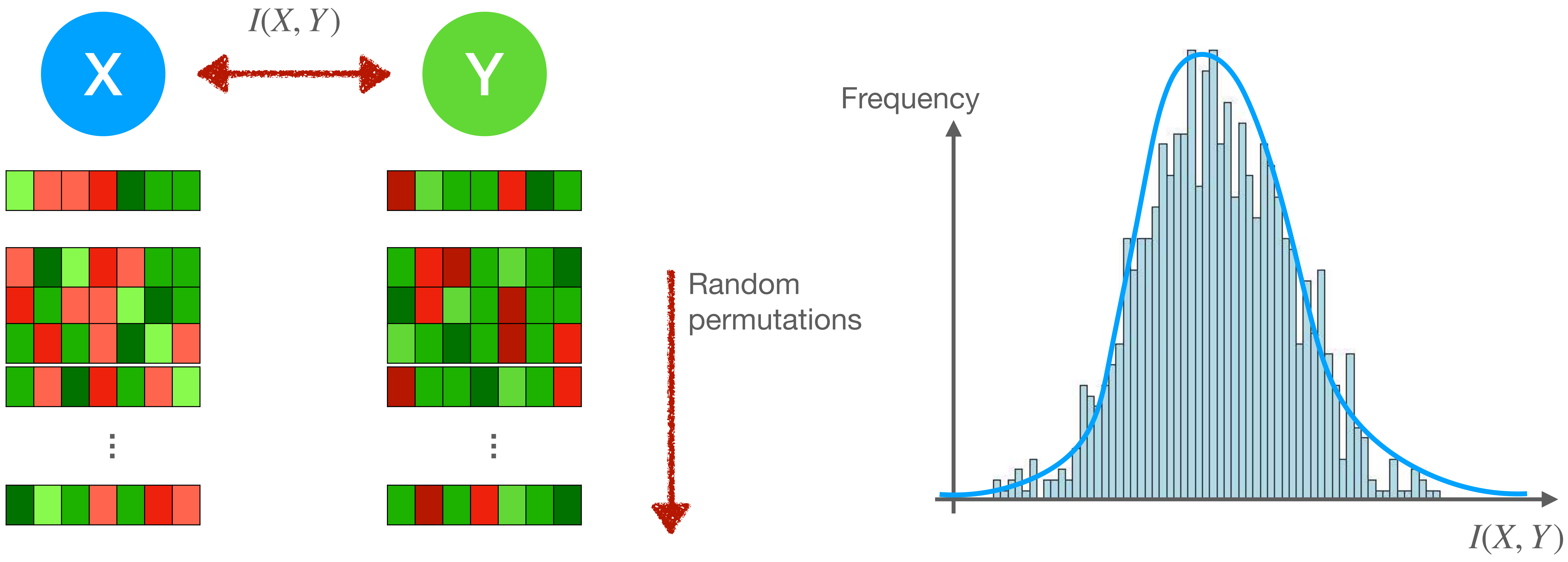
The threshold value τ_F is computed with a permutation test



Reverse engineering of gene regulatory nets

ARACNE

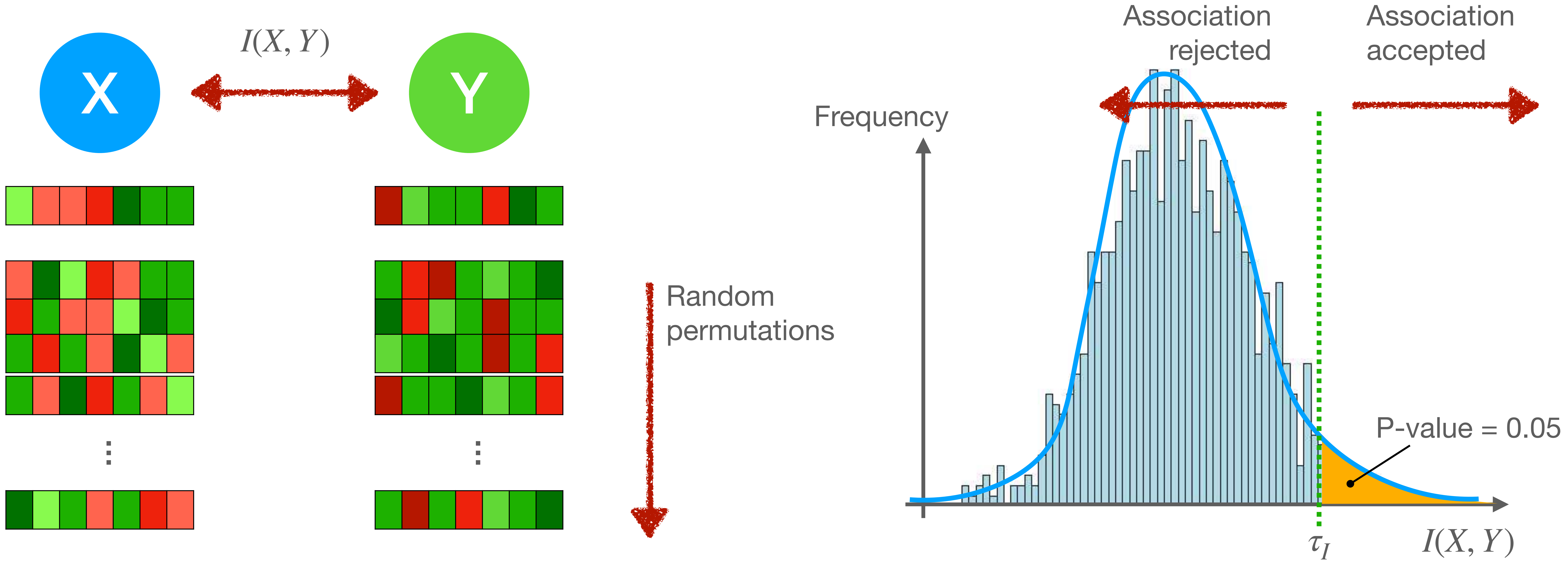
The threshold value τ_F is computed with a permutation test



Reverse engineering of gene regulatory nets

ARACNE

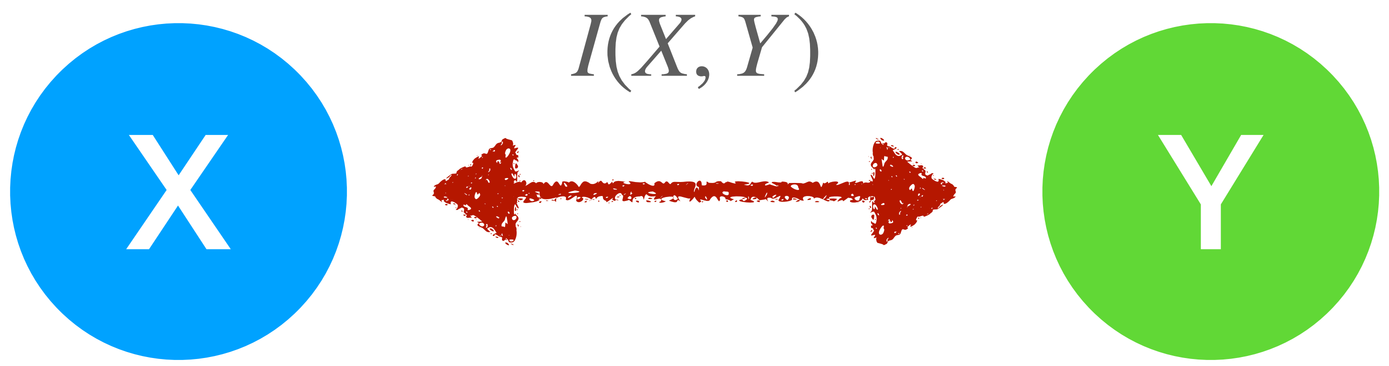
The threshold value τ_F is computed with a permutation test



Reverse engineering of gene regulatory nets

ARACNE

Indirect Associations



Reverse engineering of gene regulatory nets

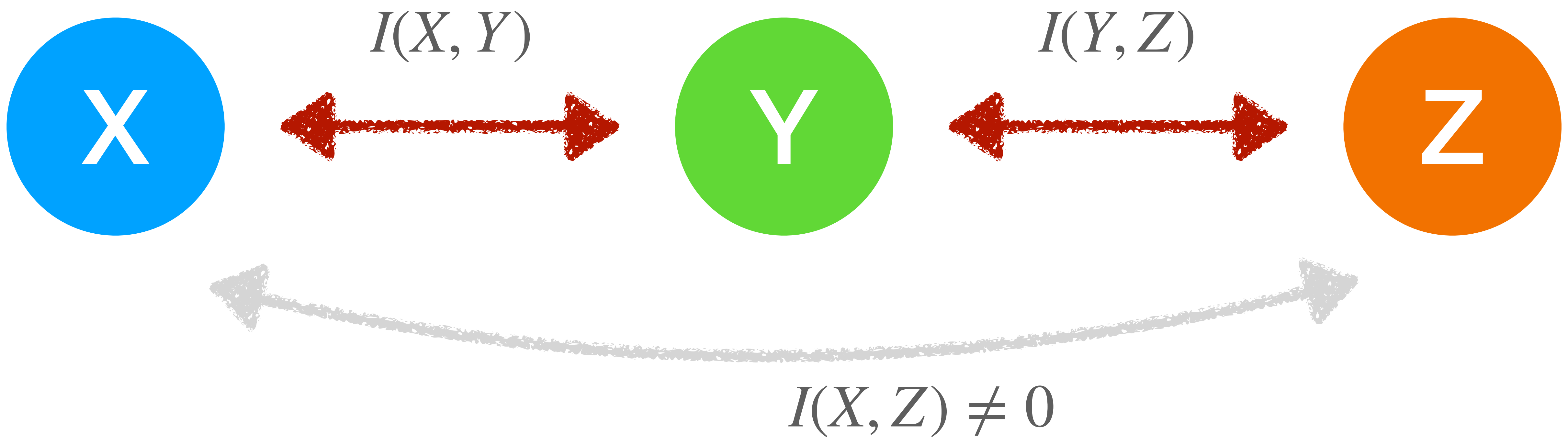
ARACNE

Indirect Associations

Reverse engineering of gene regulatory nets

ARACNE

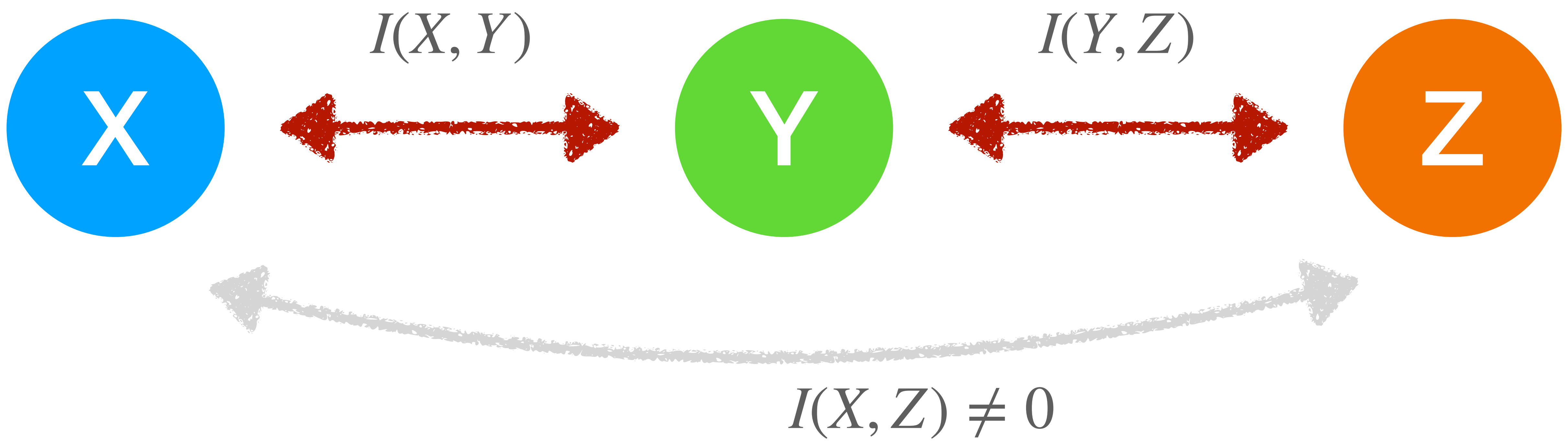
Indirect Associations



Reverse engineering of gene regulatory nets

ARACNE

Indirect Associations

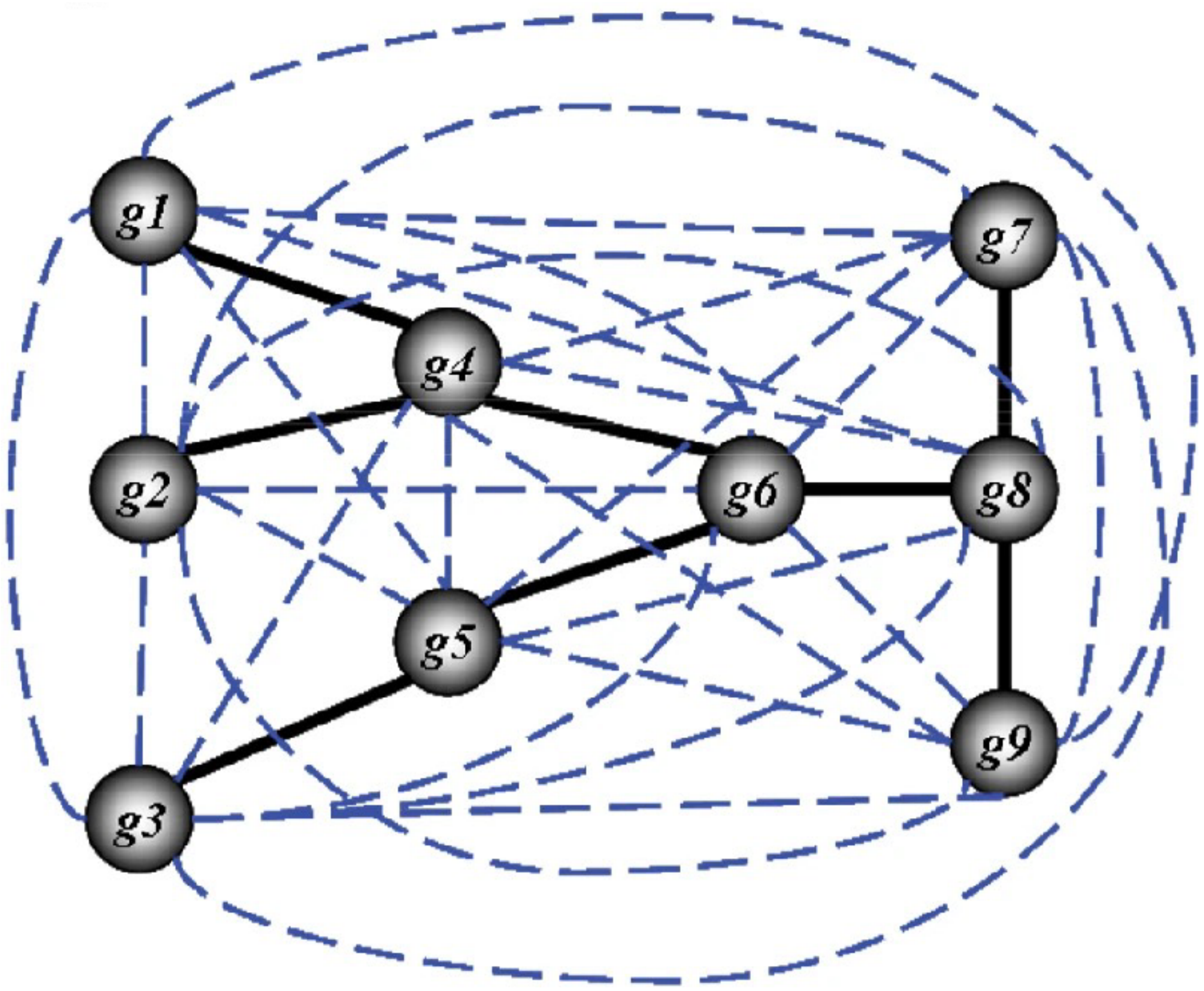


Data Process Inequality: links are removed if $I(X, Z) \leq min(I(X, Y), I(Y, Z))$

Reverse engineering of gene regulatory nets

ARACNE

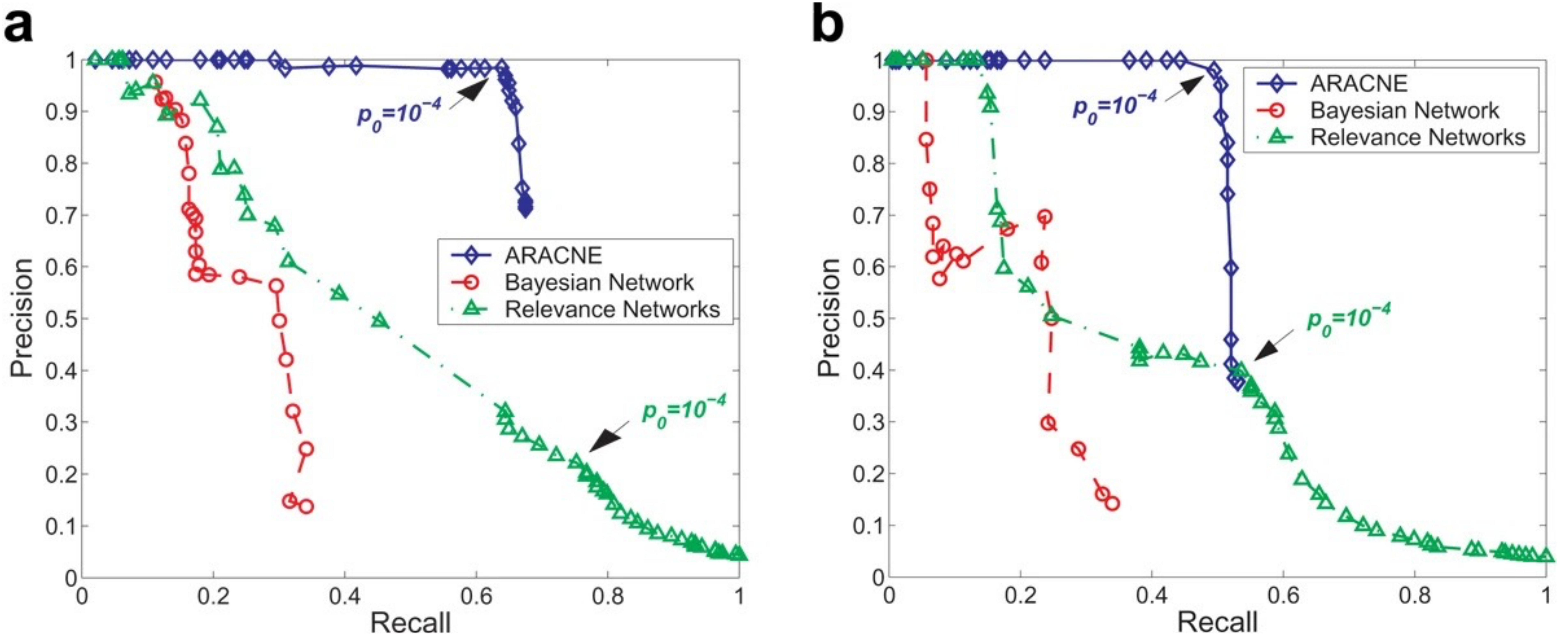
If Mutual Information can be measured without errors, ARACNE will reconstruct the network exactly by removing all false candidate interactions.



Margolin et al., *BMC Bioinformatics*, 2006

Reverse engineering of gene regulatory nets

ARACNE



Margolin et al., *BMC Bioinformatics*, 2006

Reverse engineering of gene regulatory nets

GENIE3

PLOS ONE

OPEN ACCESS PEER-REVIEWED

RESEARCH ARTICLE

Inferring Regulatory Networks from Expression Data Using Tree-Based Methods

Vân Anh Huynh-Thu, Alexandre Irrthum, Louis Wehenkel, Pierre Geurts

Published: September 28, 2010 • https://doi.org/10.1371/journal.pone.0012776

Reverse engineering of gene regulatory nets

GENIE3

PLOS ONE

OPEN ACCESS PEER-REVIEWED

RESEARCH ARTICLE

Inferring Regulatory Networks from Expression Data Using Tree-Based Methods

Vân Anh Huynh-Thu, Alexandre Irrthum, Louis Wehenkel, Pierre Geurts

Published: September 28, 2010 • https://doi.org/10.1371/journal.pone.0012776

DREAM CHALLENGES

powered by Sage Bionetworks

Overall winner of DREAM4 (2019) and DREAM5 (2010) challenges

Reverse engineering of gene regulatory nets

GENIE3

PLOS ONE

OPEN ACCESS PEER-REVIEWED

RESEARCH ARTICLE

Inferring Regulatory Networks from Expression Data Using Tree-Based Methods

Vân Anh Huynh-Thu, Alexandre Irrthum, Louis Wehenkel, Pierre Geurts

Published: September 28, 2010 • https://doi.org/10.1371/journal.pone.0012776

DREAM CHALLENGES

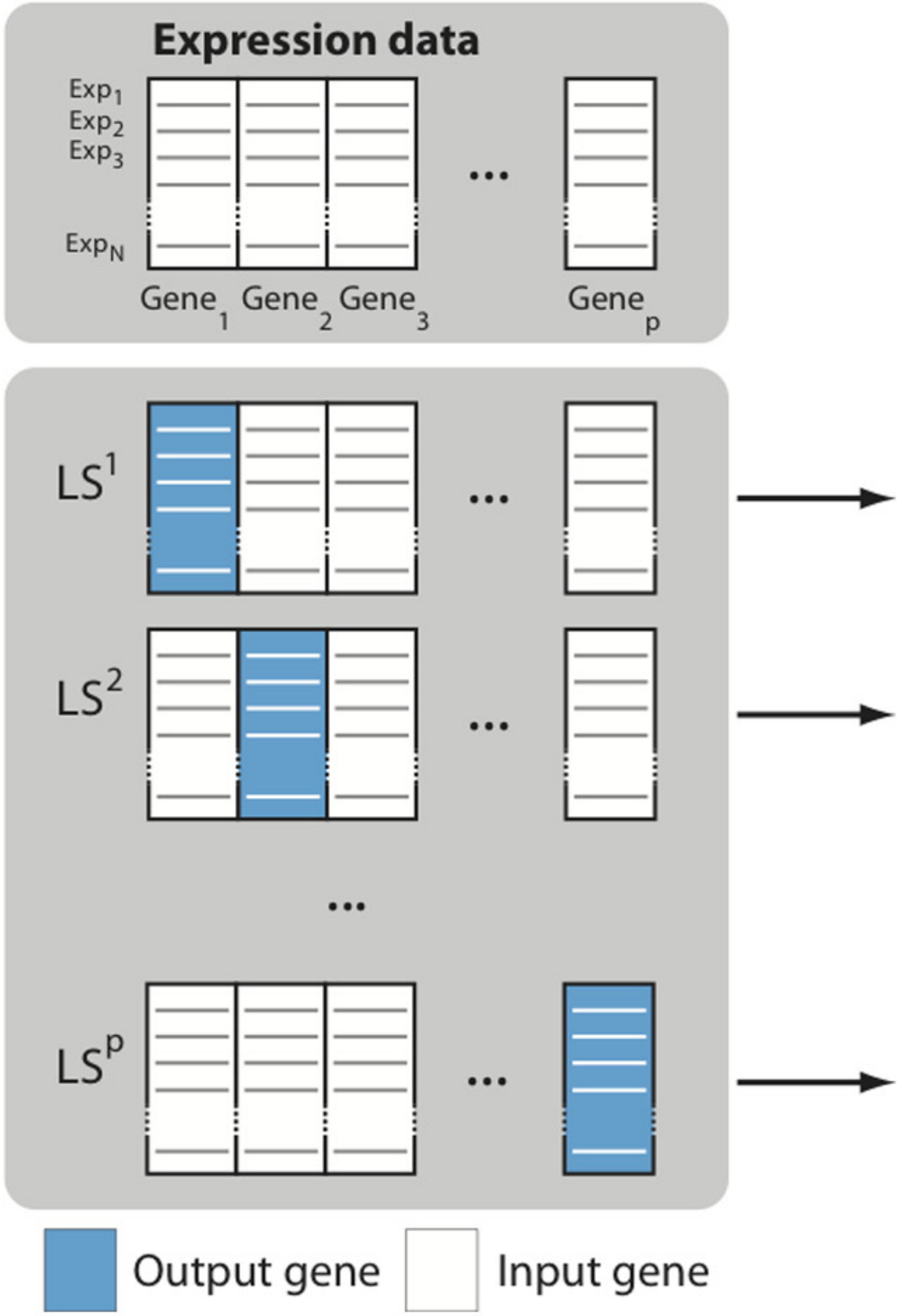
powered by Sage Bionetworks

Overall winner of DREAM4 (2019) and DREAM5 (2010) challenges

Model Based: learn a model that explains as well as possible the observed expression data and extract the network from this model

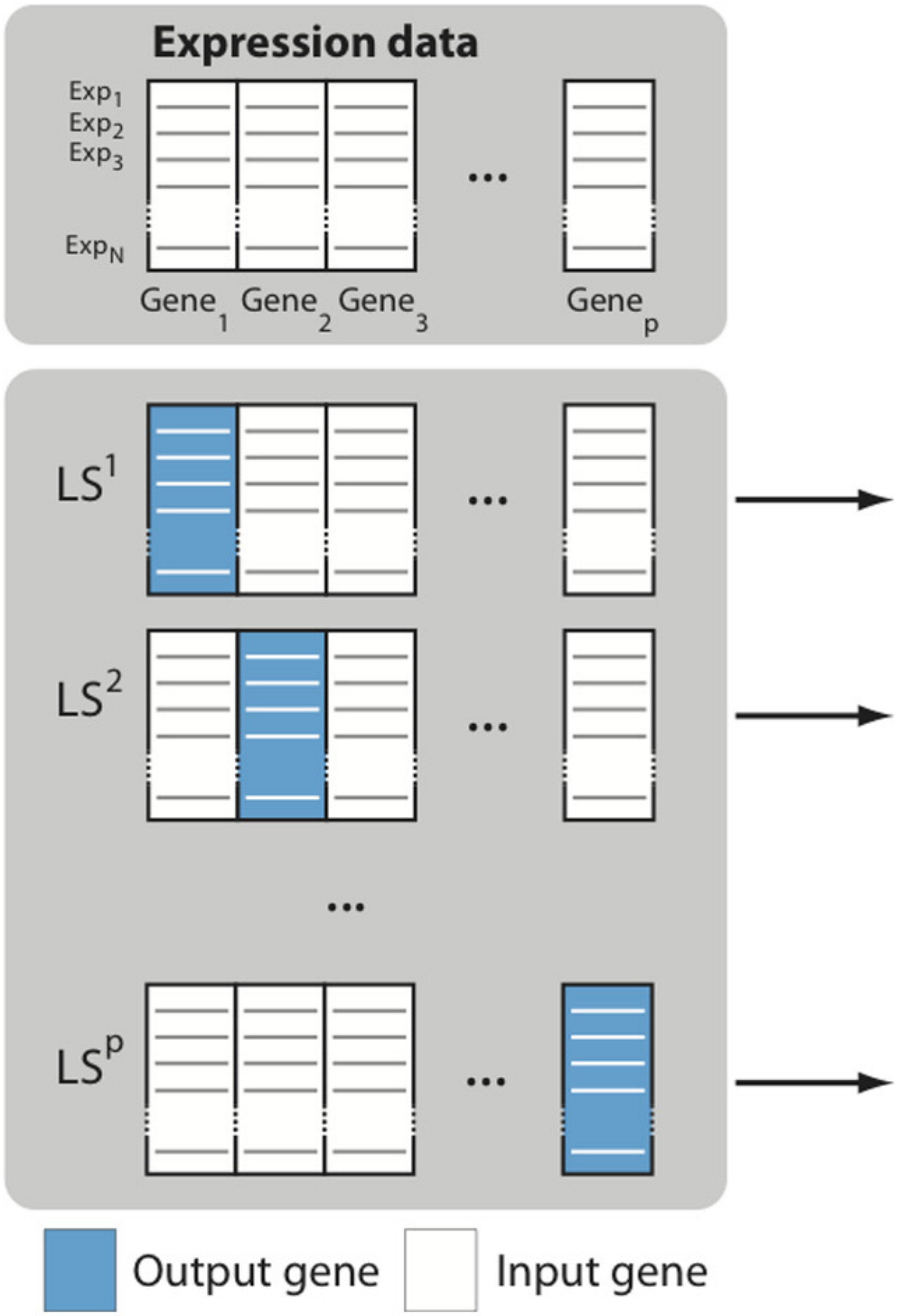
Reverse engineering of gene regulatory nets

GENIE3

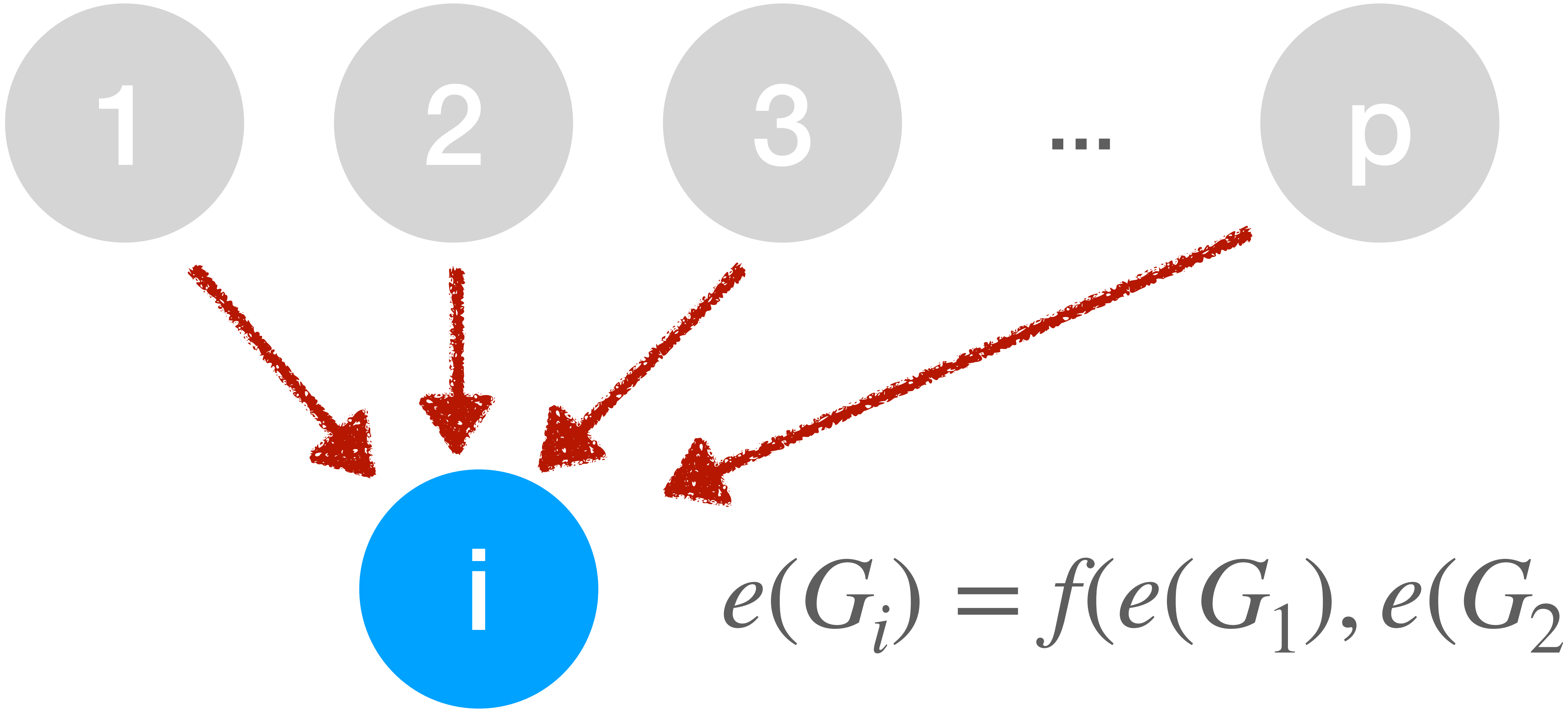


Reverse engineering of gene regulatory nets

GENIE3



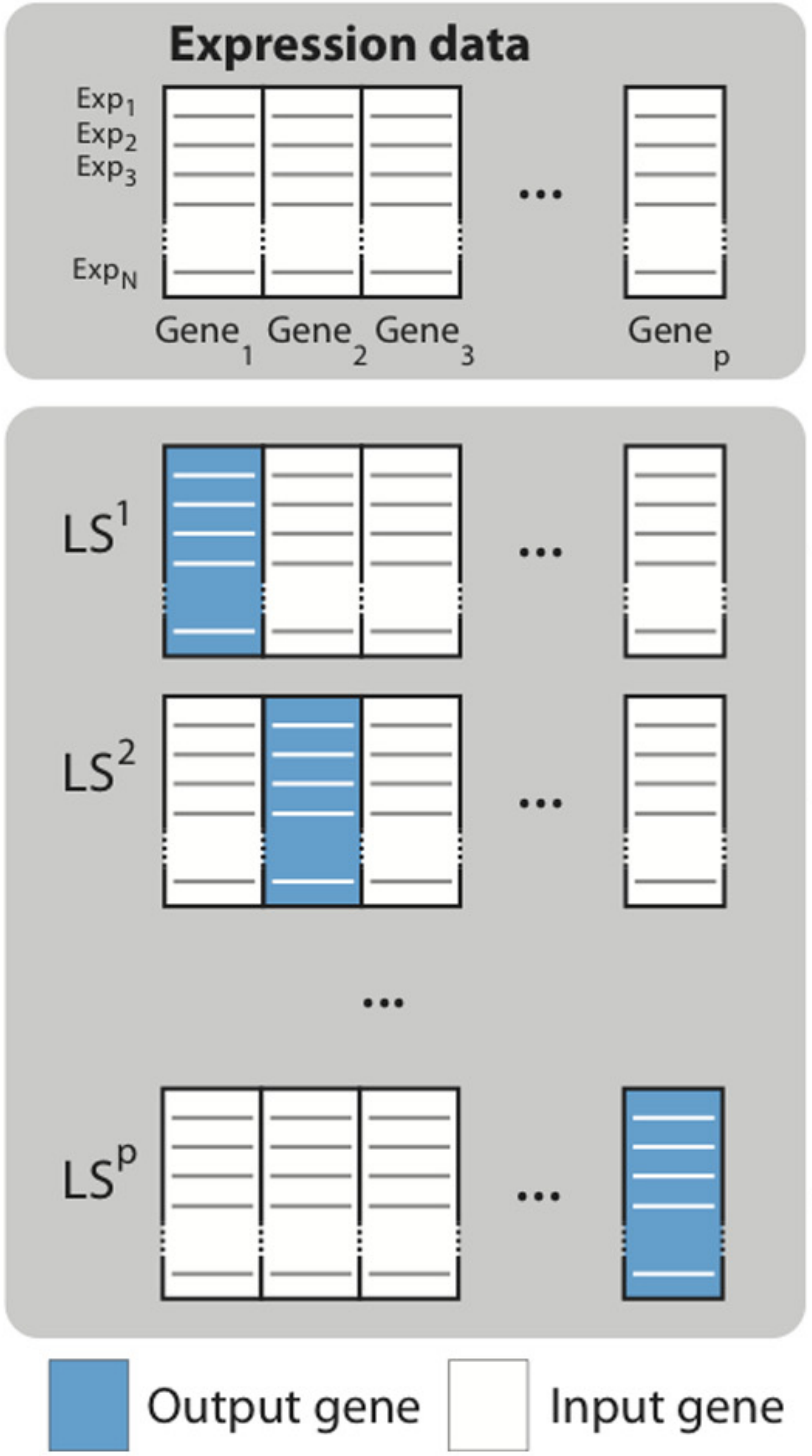
Decompose the problem into p sub-problems. Each sub-problem $i \in \{1,\ldots,p\}$ finds the most strong regulators of G_i



$$e(G_i) = f(e(G_1), e(G_2), \ldots, e(G_p))$$

Reverse engineering of gene regulatory nets

GENIE3



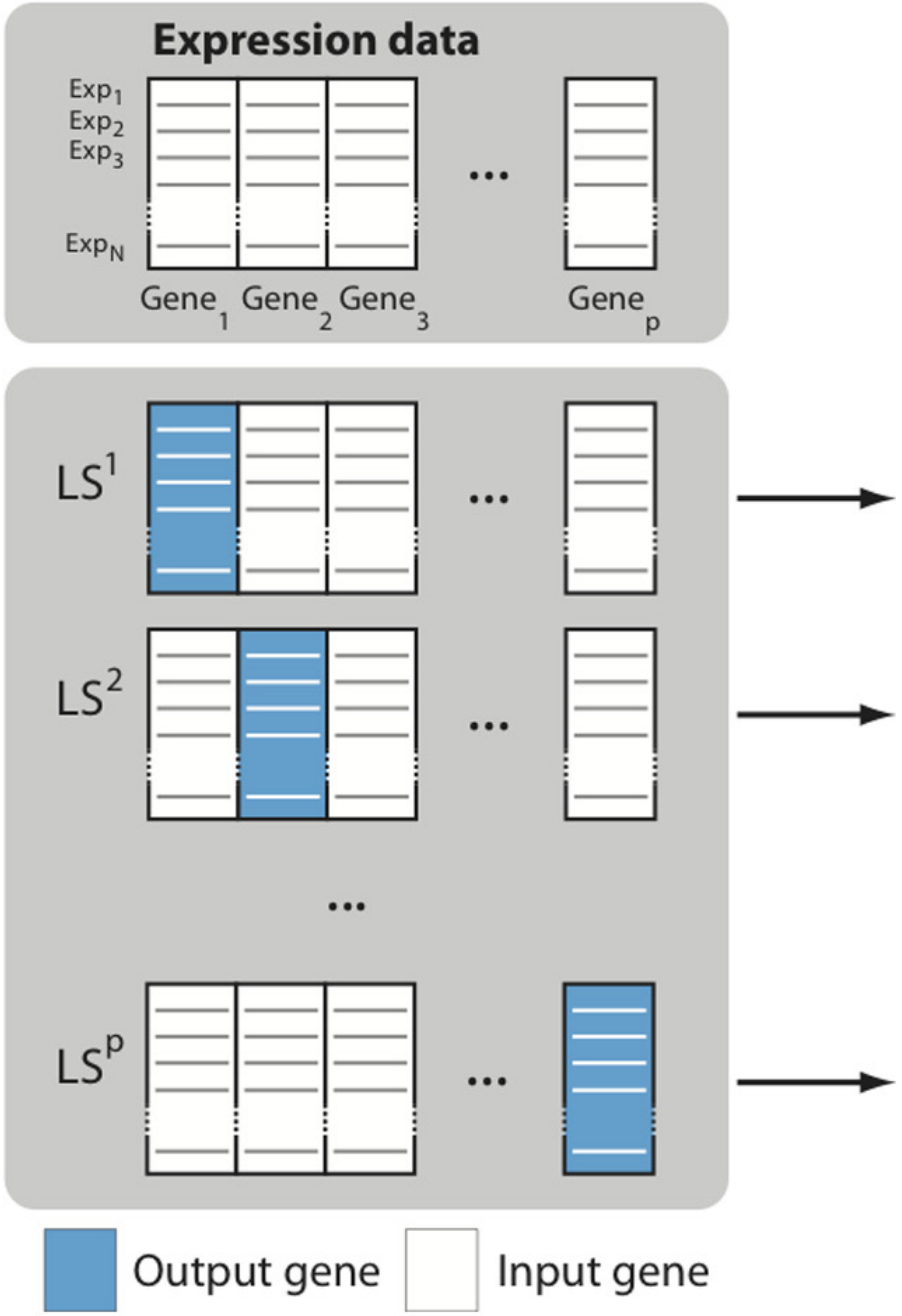
$$e(G_1) = f_1(e(G_2), \ldots, e(G_p))$$

$$e(G_2) = f_2(e(G_1), \ldots, e(G_p))$$

$$e(G_p) = f_2(e(G_1), \ldots, e(G_{p-1}))$$

Reverse engineering of gene regulatory nets

GENIE3



f_i can be learned in several ways

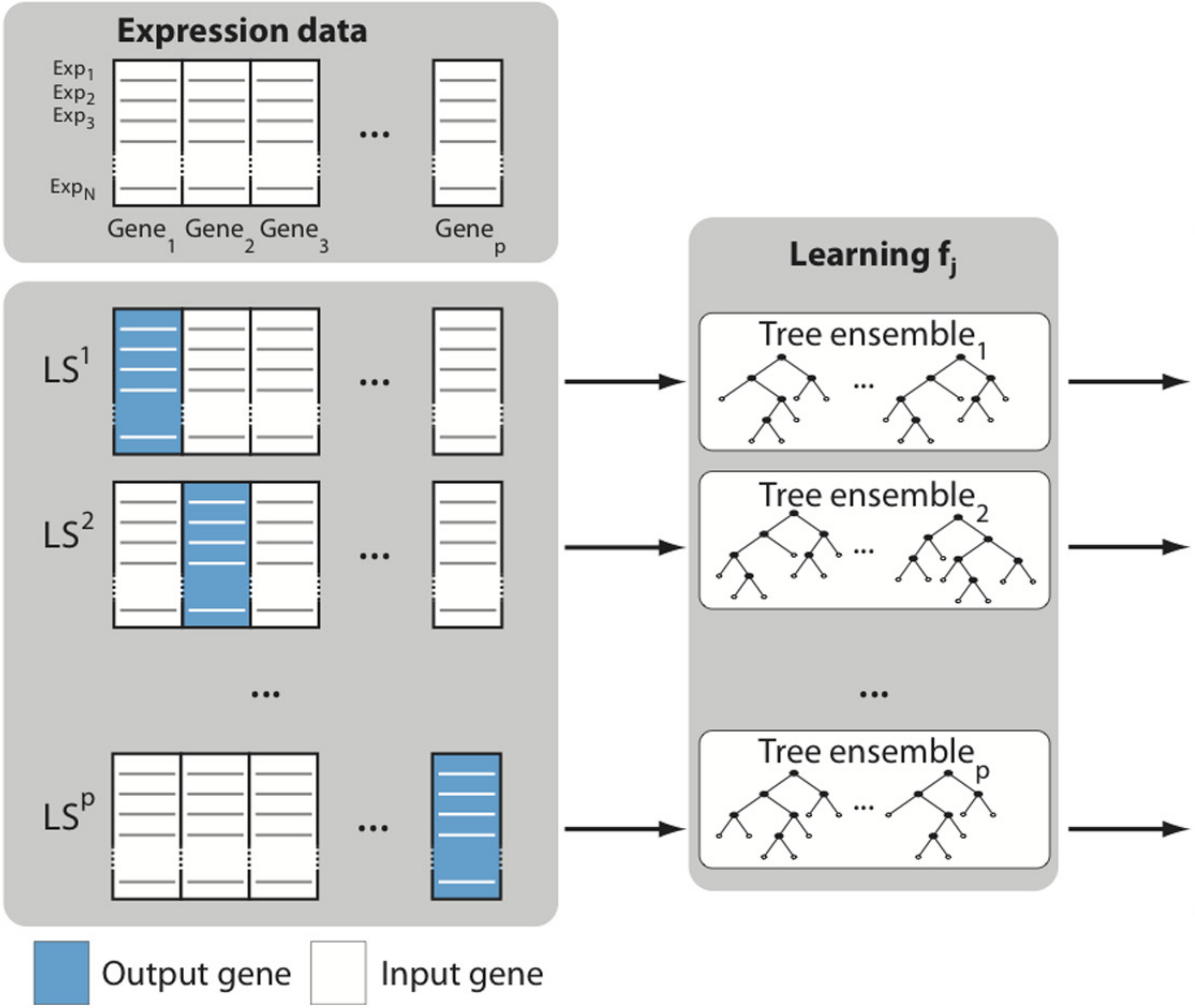
$$e(G_1) = f_1(e(G_2), \ldots, e(G_p))$$

$$e(G_2) = f_2(e(G_1), \ldots, e(G_p))$$

$$e(G_p) = f_2(e(G_1), \ldots, e(G_{p-1}))$$

Reverse engineering of gene regulatory nets

GENIE3

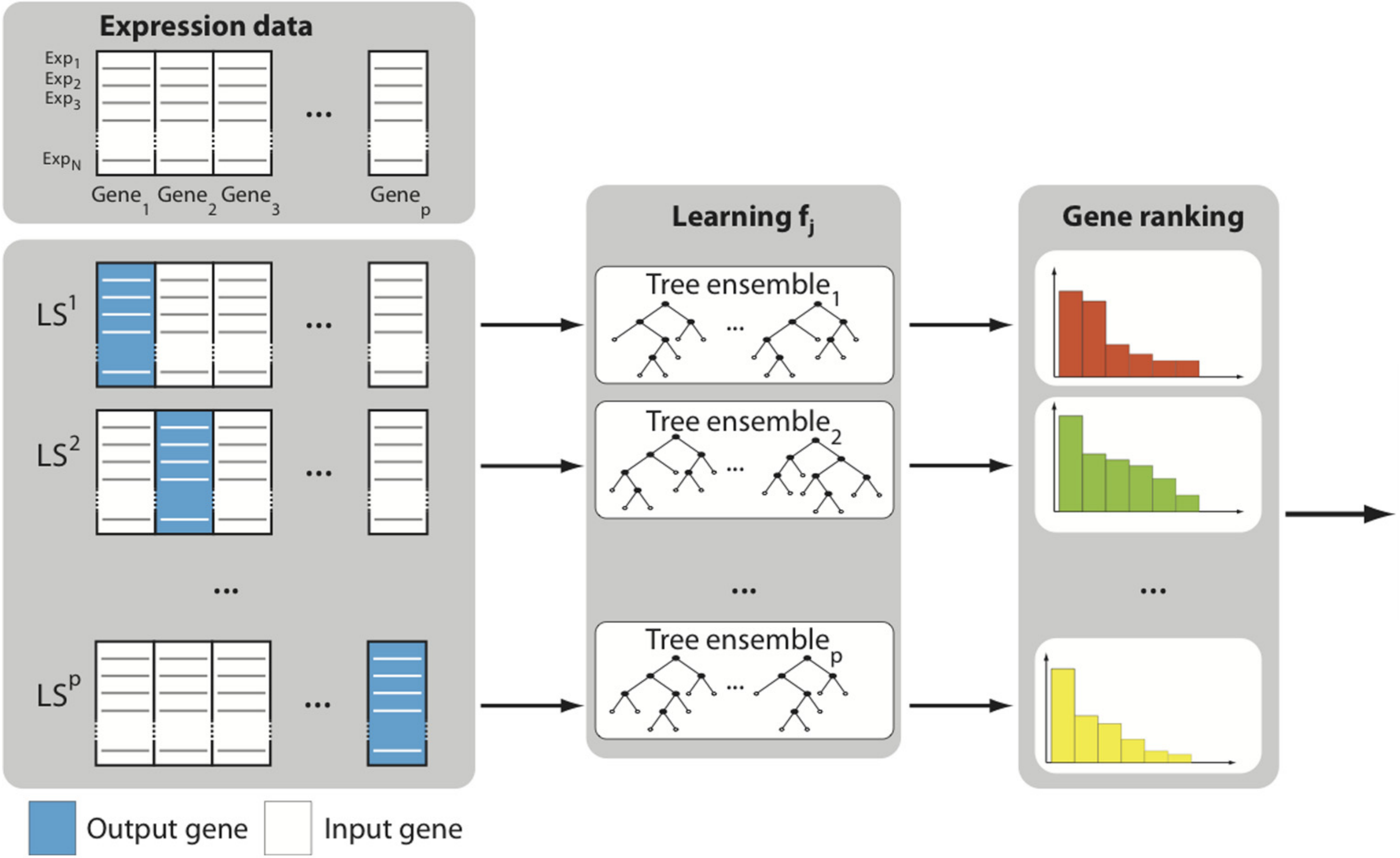


GENIE3 adopts a random forest approach which is:

- non-parametric
- can deal with interacting features
- work well with high dimensional datasets
- it is scalable

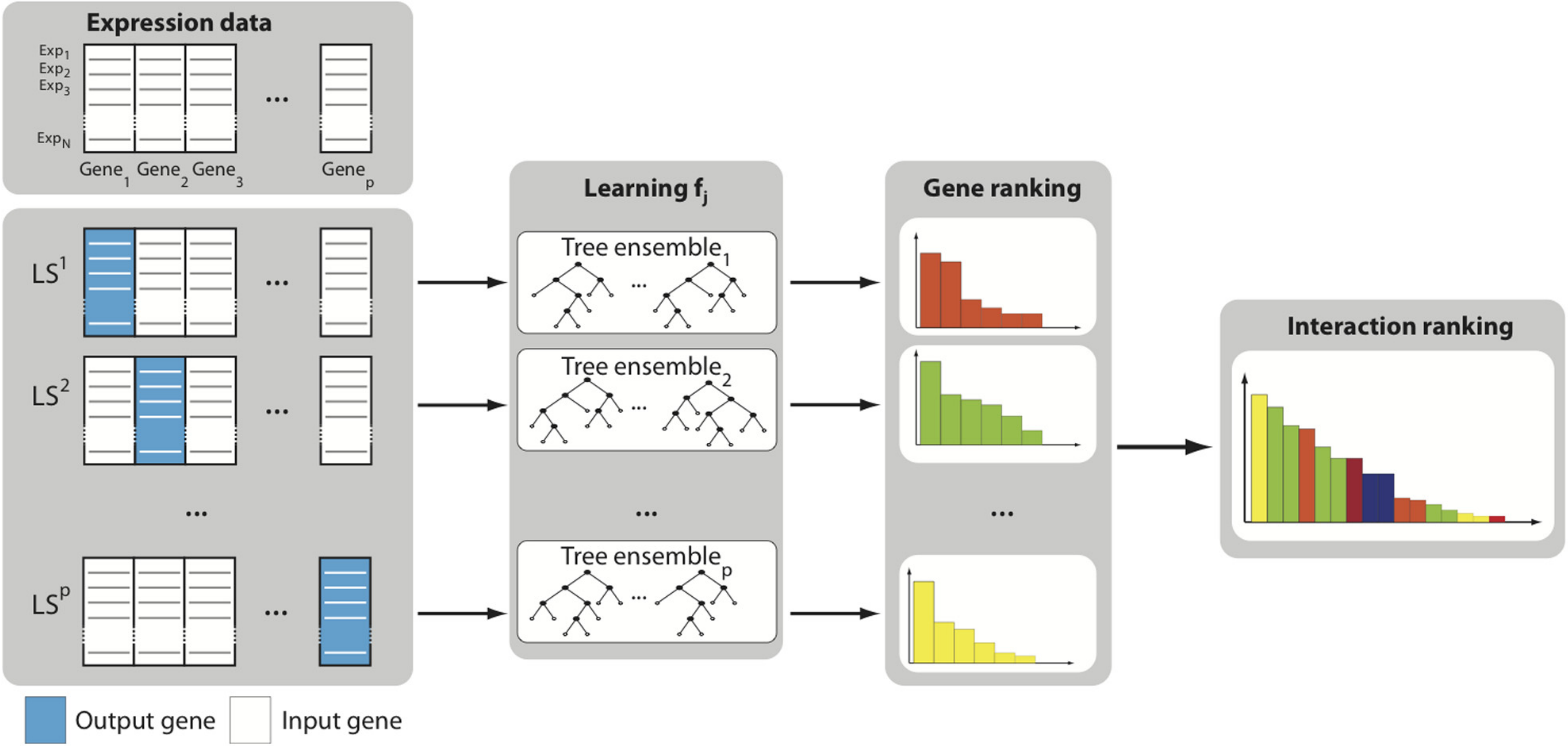
Reverse engineering of gene regulatory nets

GENIE3



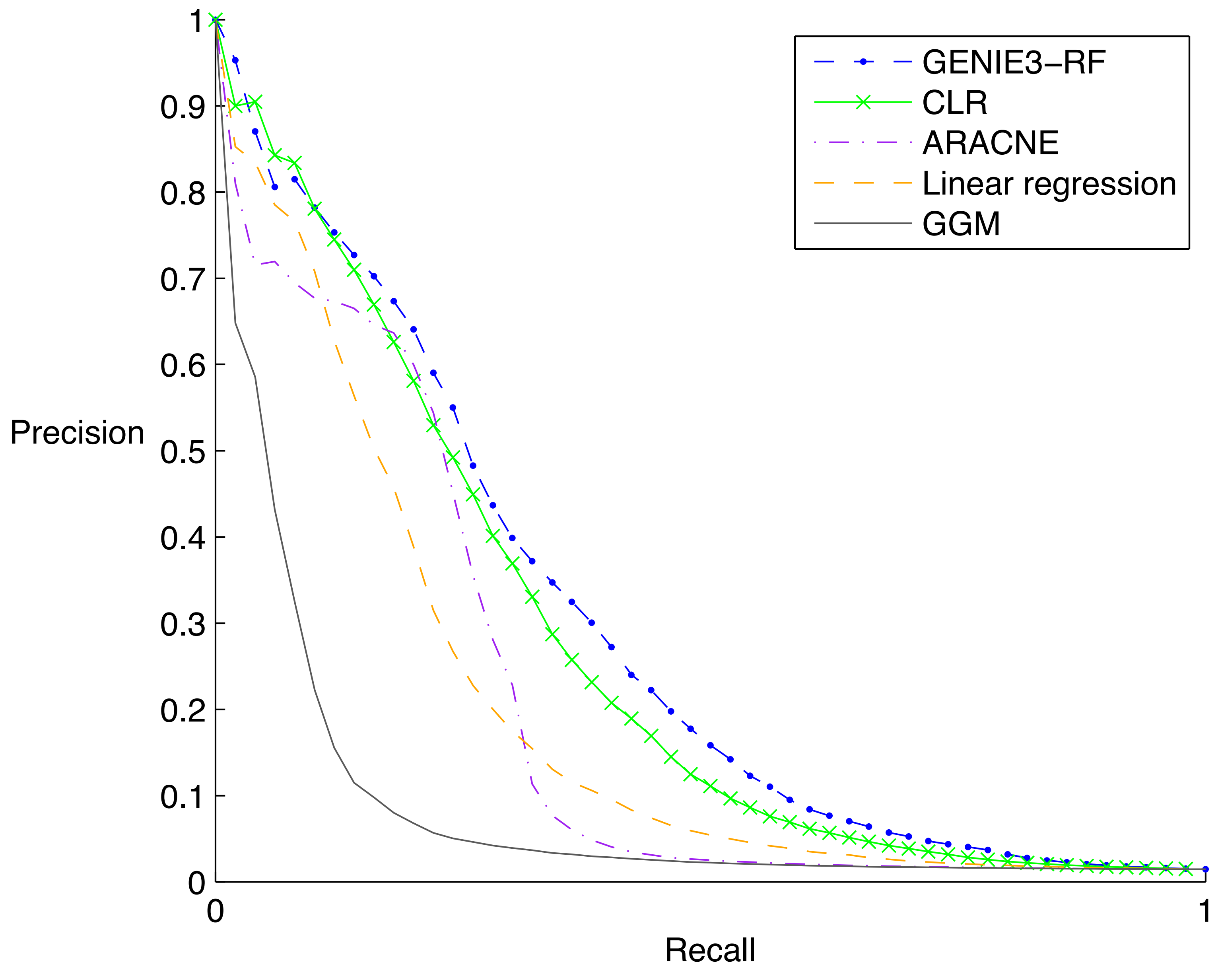
Reverse engineering of gene regulatory nets

GENIE3

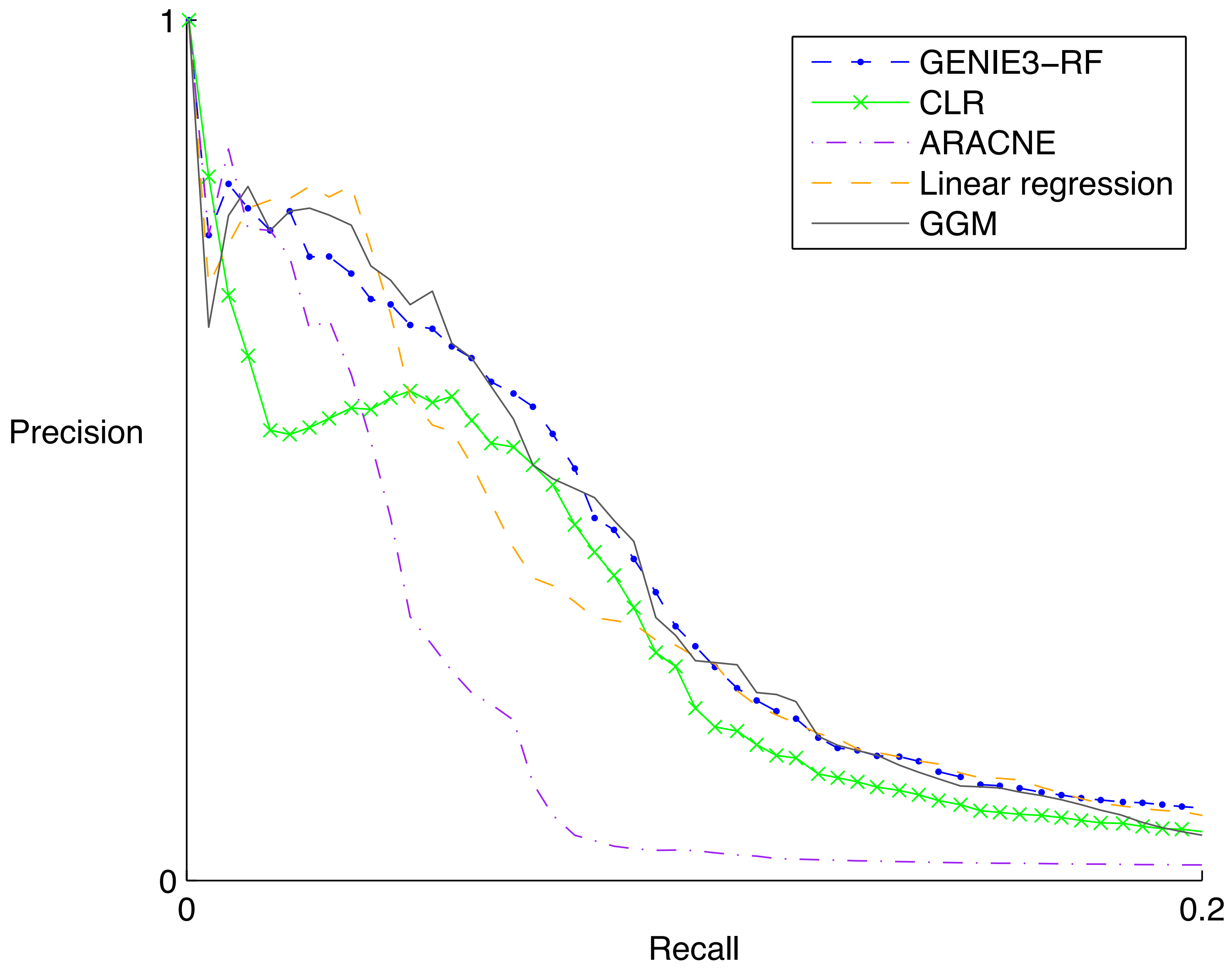


Reverse engineering of gene regulatory nets

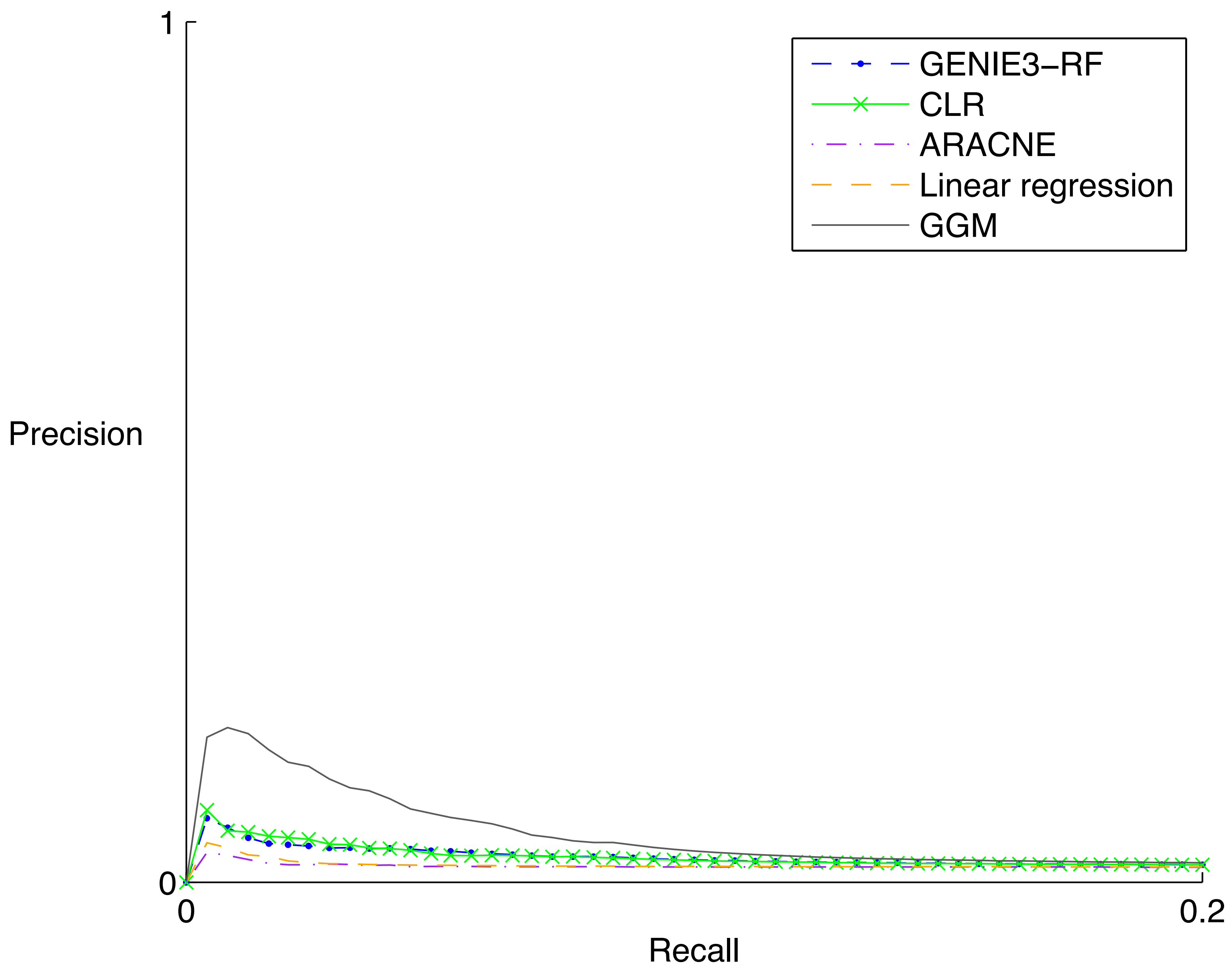
GENIE3



In silico



E. coli



S. Cerevisiae

Reverse engineering of gene regulatory nets

SIRENE

JOURNAL ARTICLE

SIRENE: supervised inference of regulatory networks

Fantine Mordelet ✉, Jean-Philippe Vert

Bioinformatics, Volume 24, Issue 16, 15 August 2008, Pages i76–i82,

Supervised Model Based: learn from known gene-gene interaction a SVM model that explains as well as possible the observed expression data and extract the network from this model

Reverse engineering of gene regulatory nets

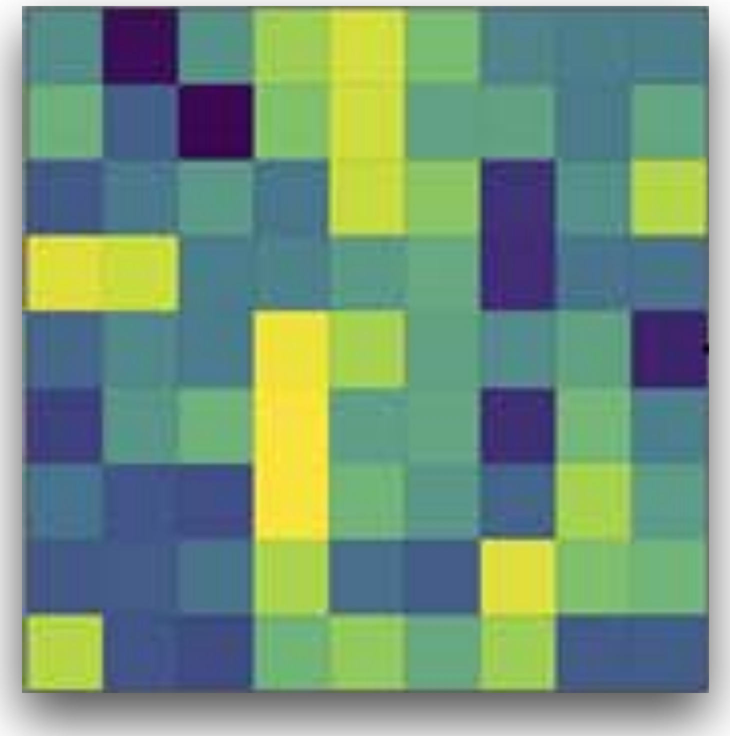
SIRENE

JOURNAL ARTICLE

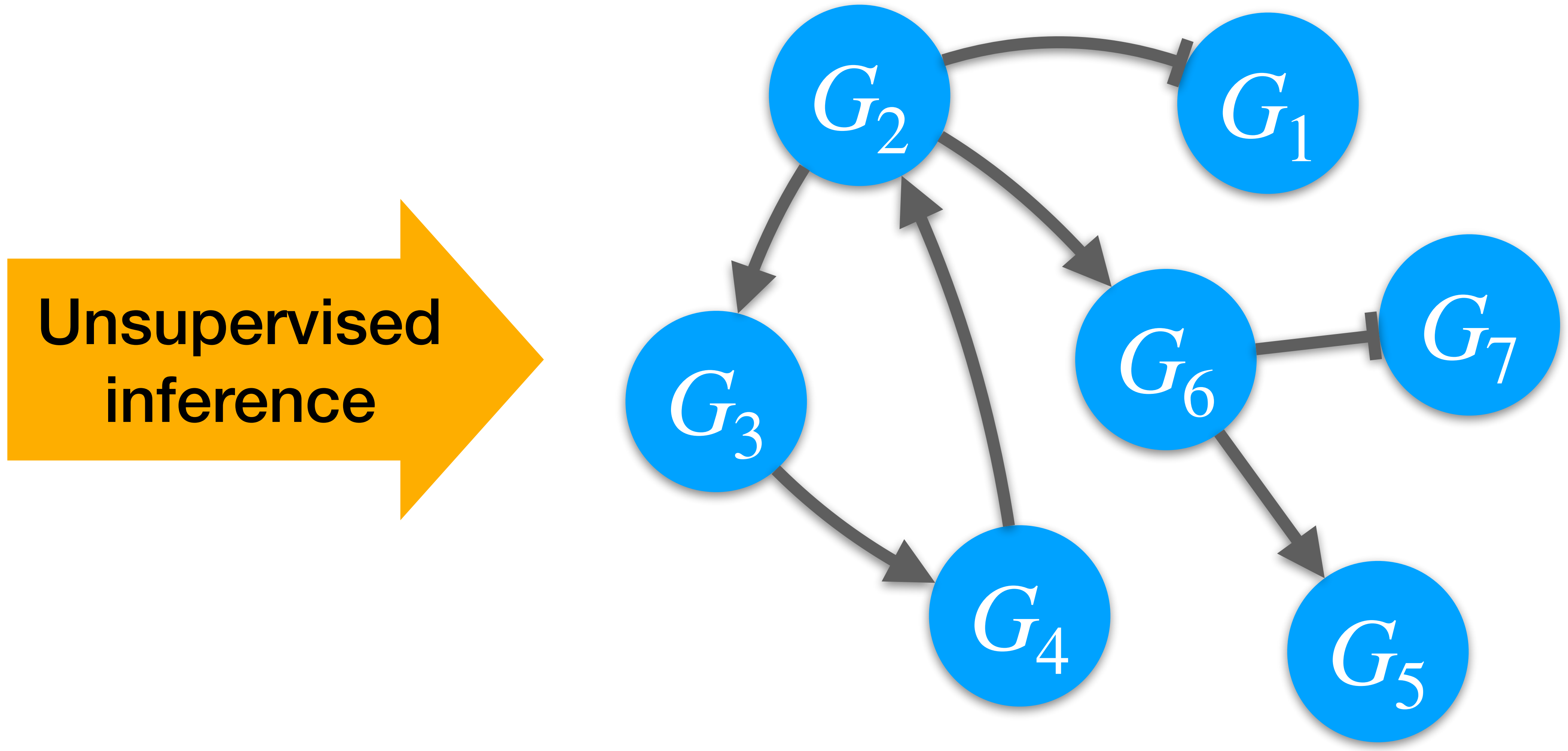
SIRENE: supervised inference of regulatory networks

Fantine Mordelet, Jean-Philippe Vert

Bioinformatics, Volume 24, Issue 16, 15 August 2008, Pages i76–i82,



Expression data



Unsupervised inference

Reverse engineering of gene regulatory nets

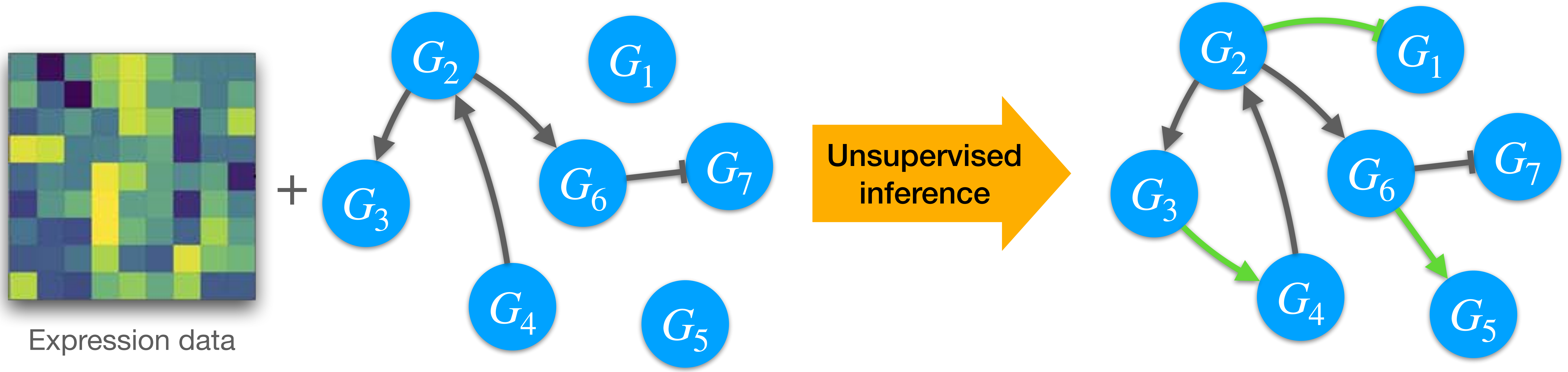
SIRENE

JOURNAL ARTICLE

SIRENE: supervised inference of regulatory networks

Fantine Mordelet, Jean-Philippe Vert

Bioinformatics, Volume 24, Issue 16, 15 August 2008, Pages i76–i82,



Unsupervised inference

Expression data

Reverse engineering of gene regulatory nets

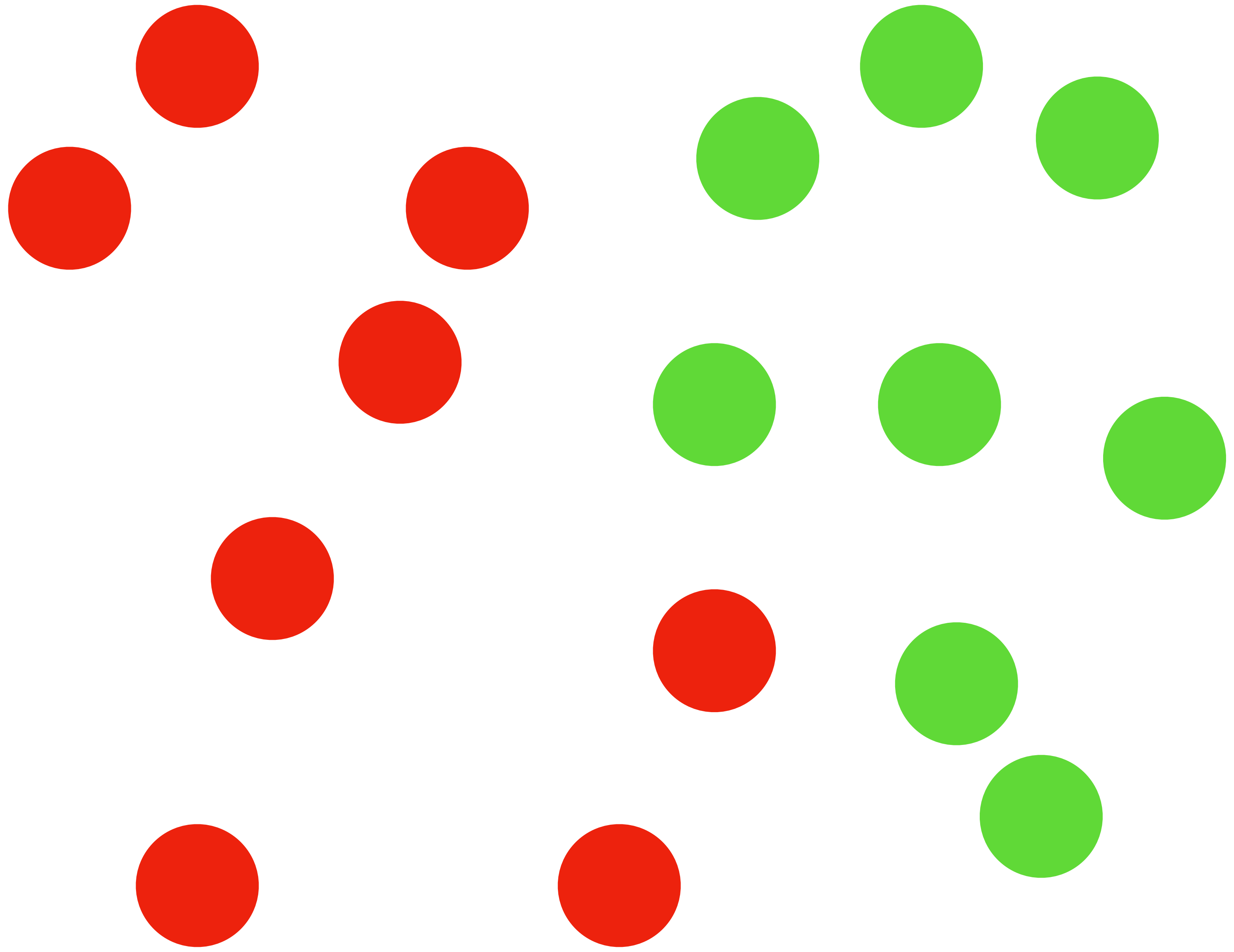
SIRENE

Machine Learning

Reverse engineering of gene regulatory nets

SIRENE

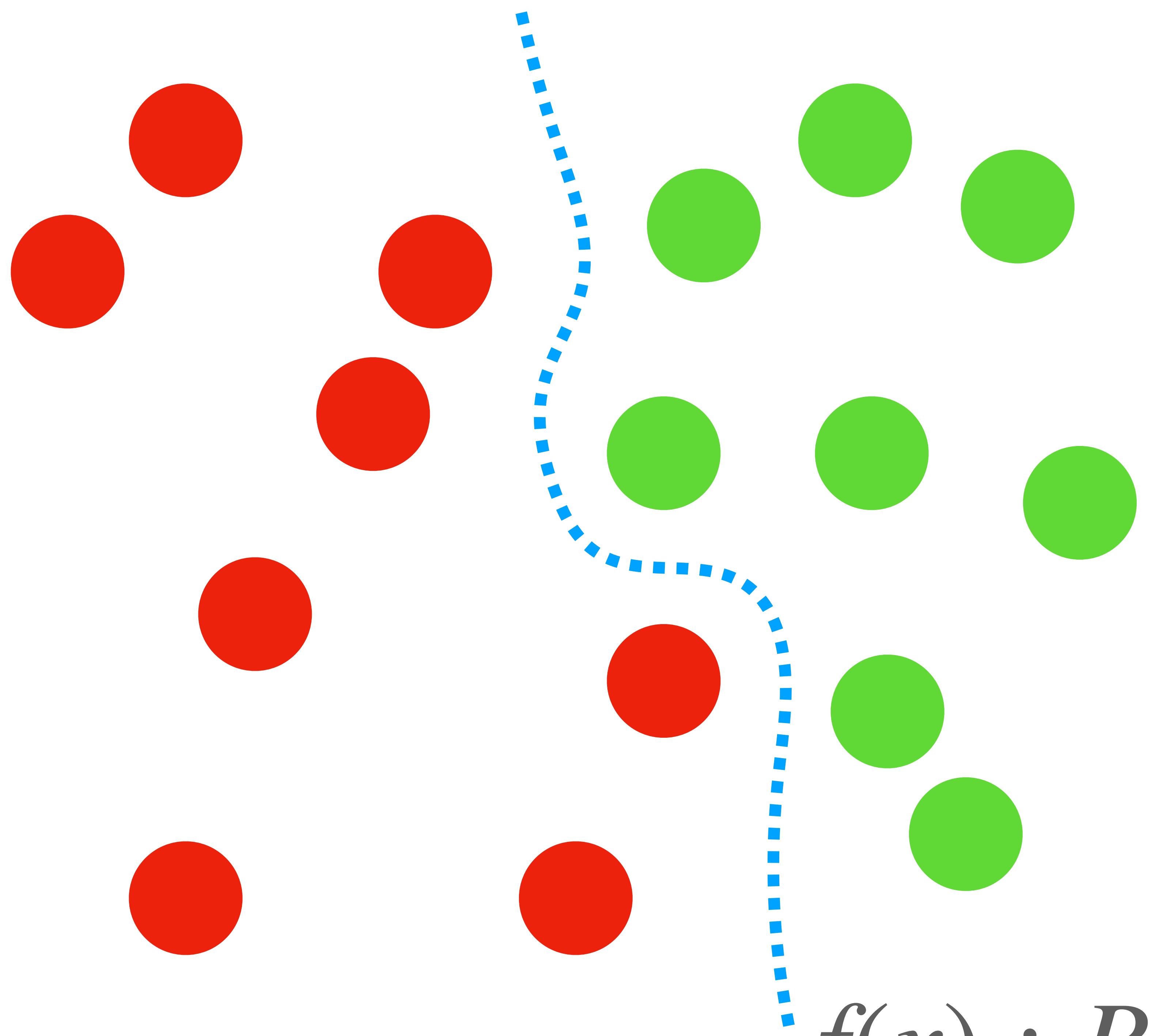
Machine Learning



Reverse engineering of gene regulatory nets

SIRENE

Machine Learning

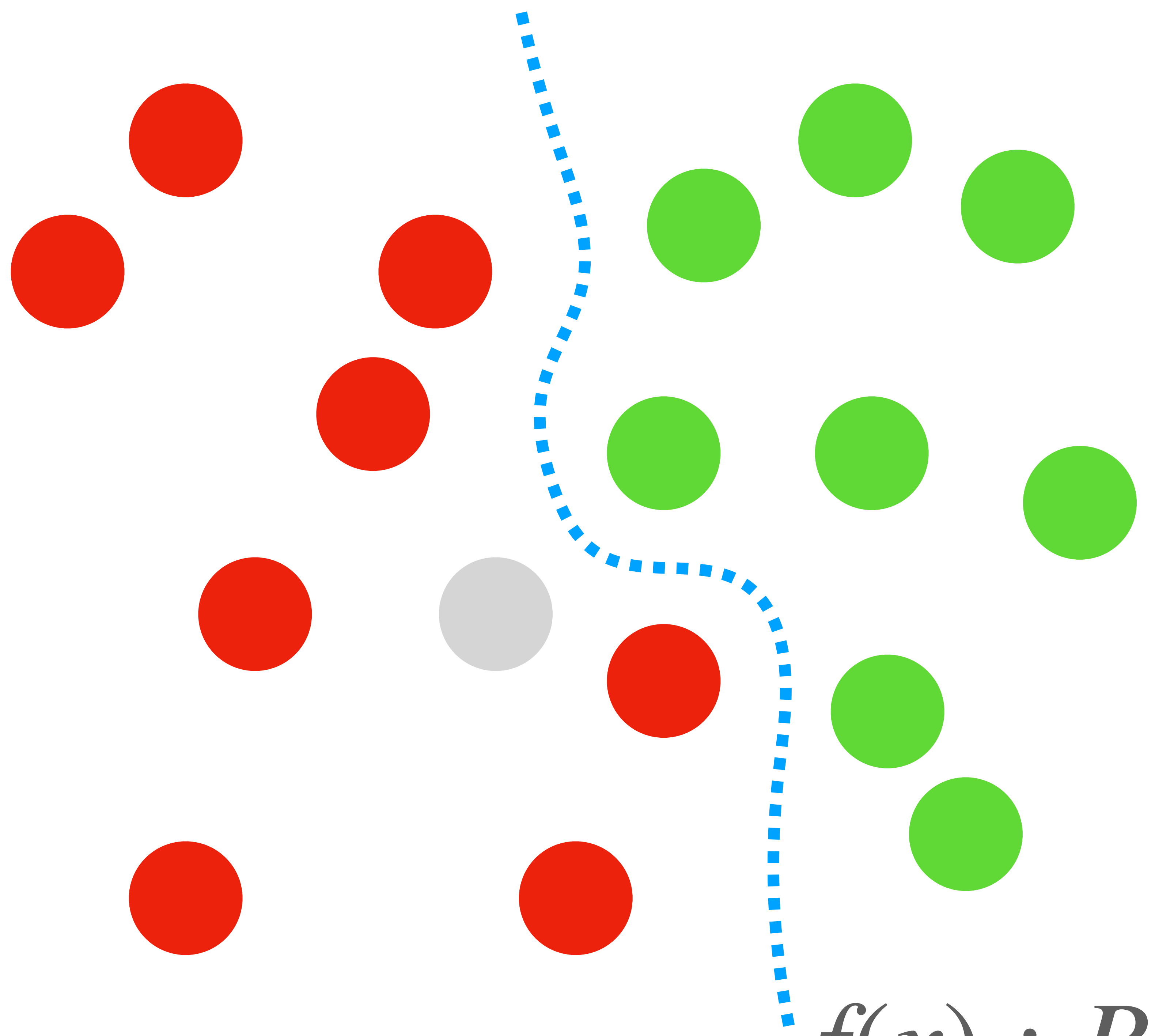


$f(x) : R^n \rightarrow [0,1]$

Reverse engineering of gene regulatory nets

SIRENE

Machine Learning

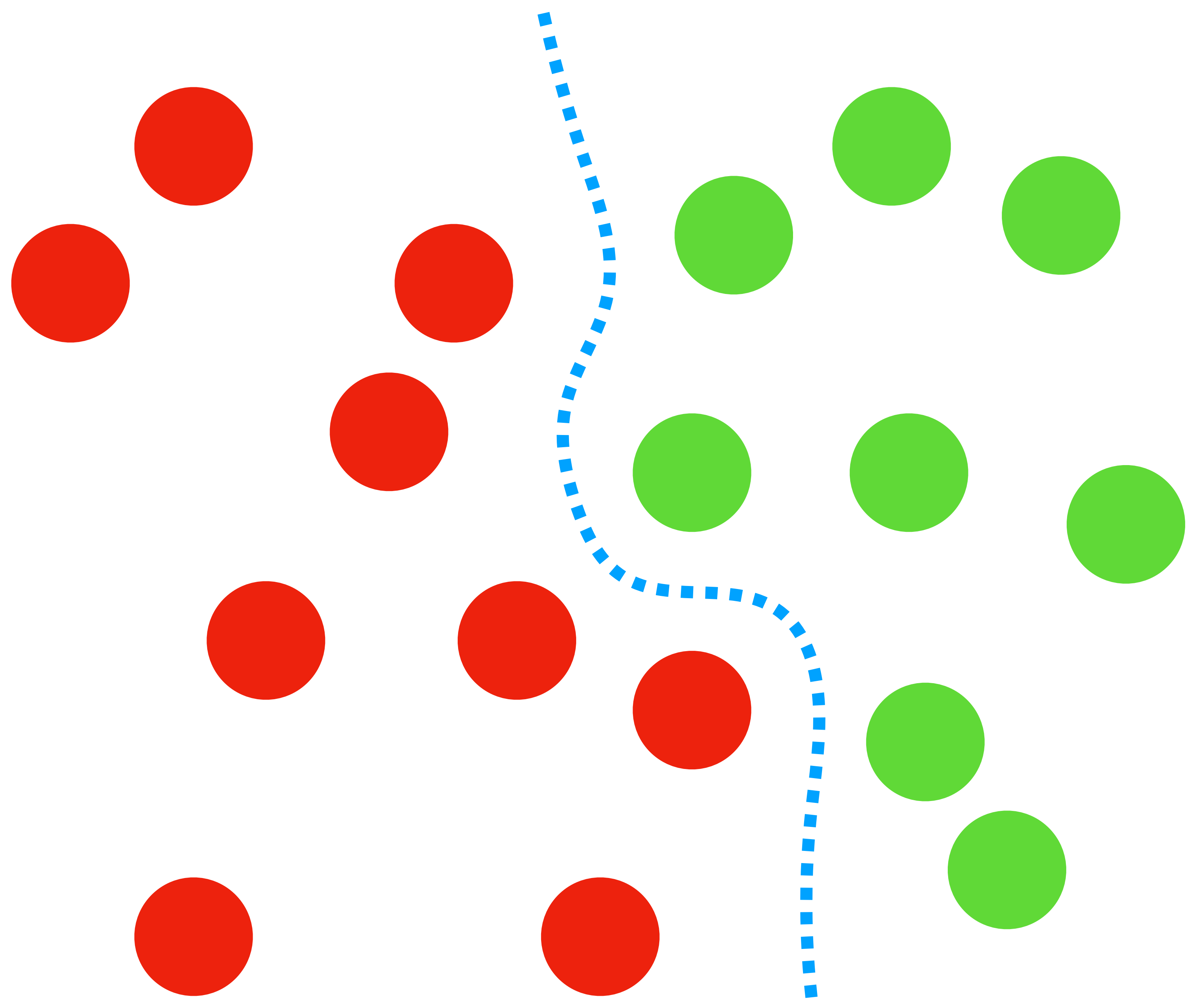


$f(x) : R^n \to [0,1]$

Reverse engineering of gene regulatory nets

SIRENE

Machine Learning

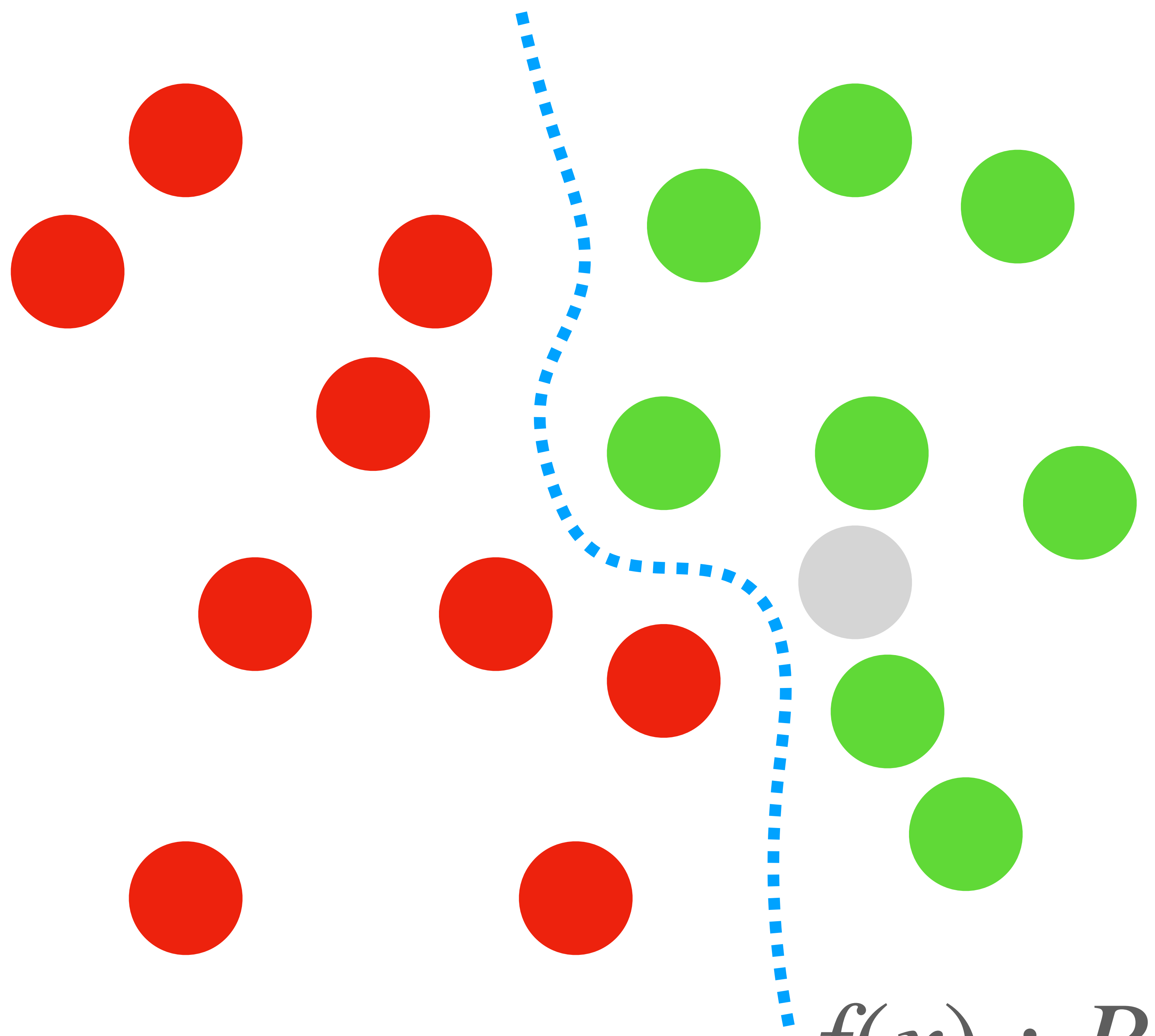


$f(x) : R^n \rightarrow [0,1]$

Reverse engineering of gene regulatory nets

SIRENE

Machine Learning

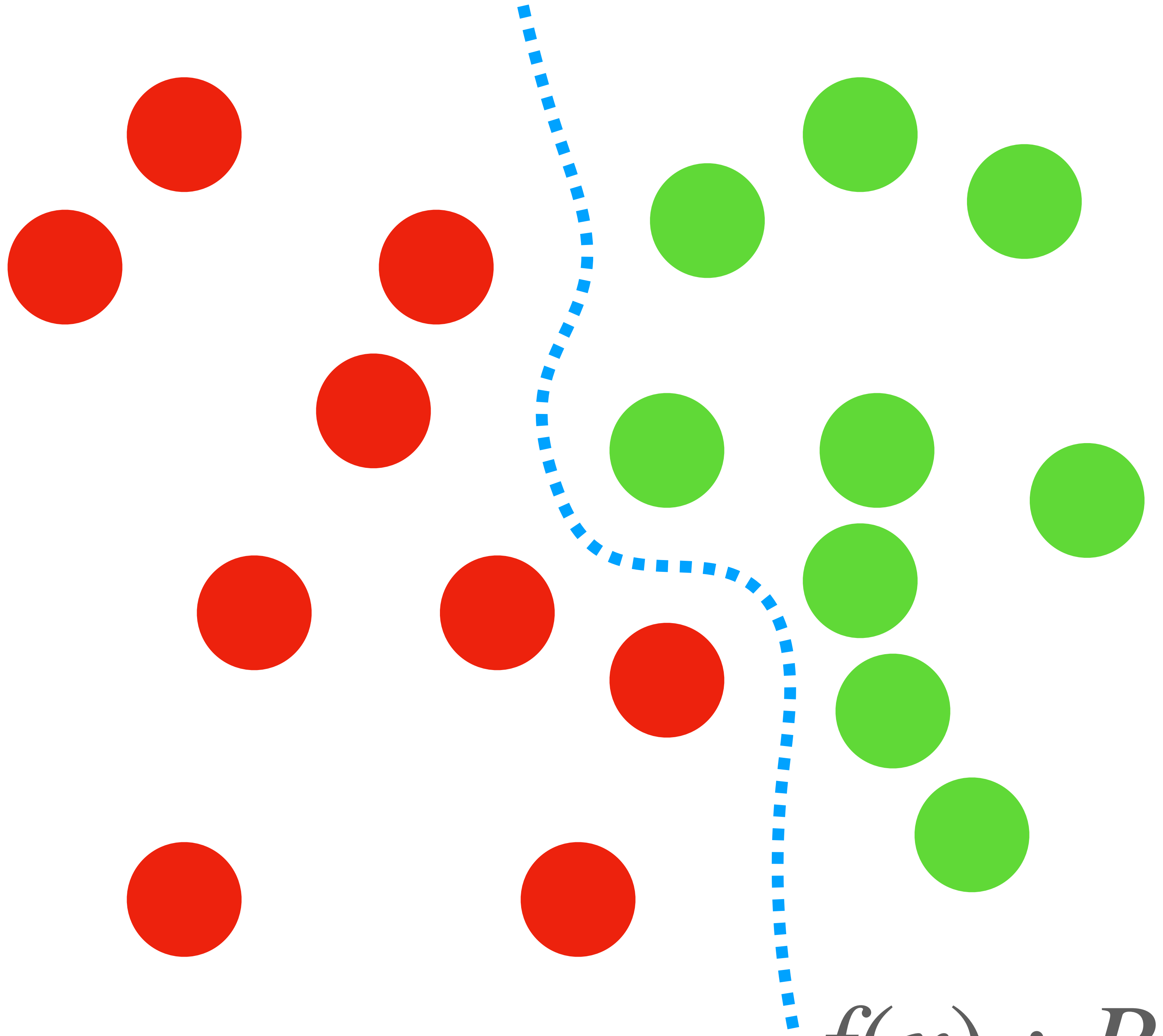


$f(x) : R^n \rightarrow [0,1]$

Reverse engineering of gene regulatory nets

SIRENE

Machine Learning

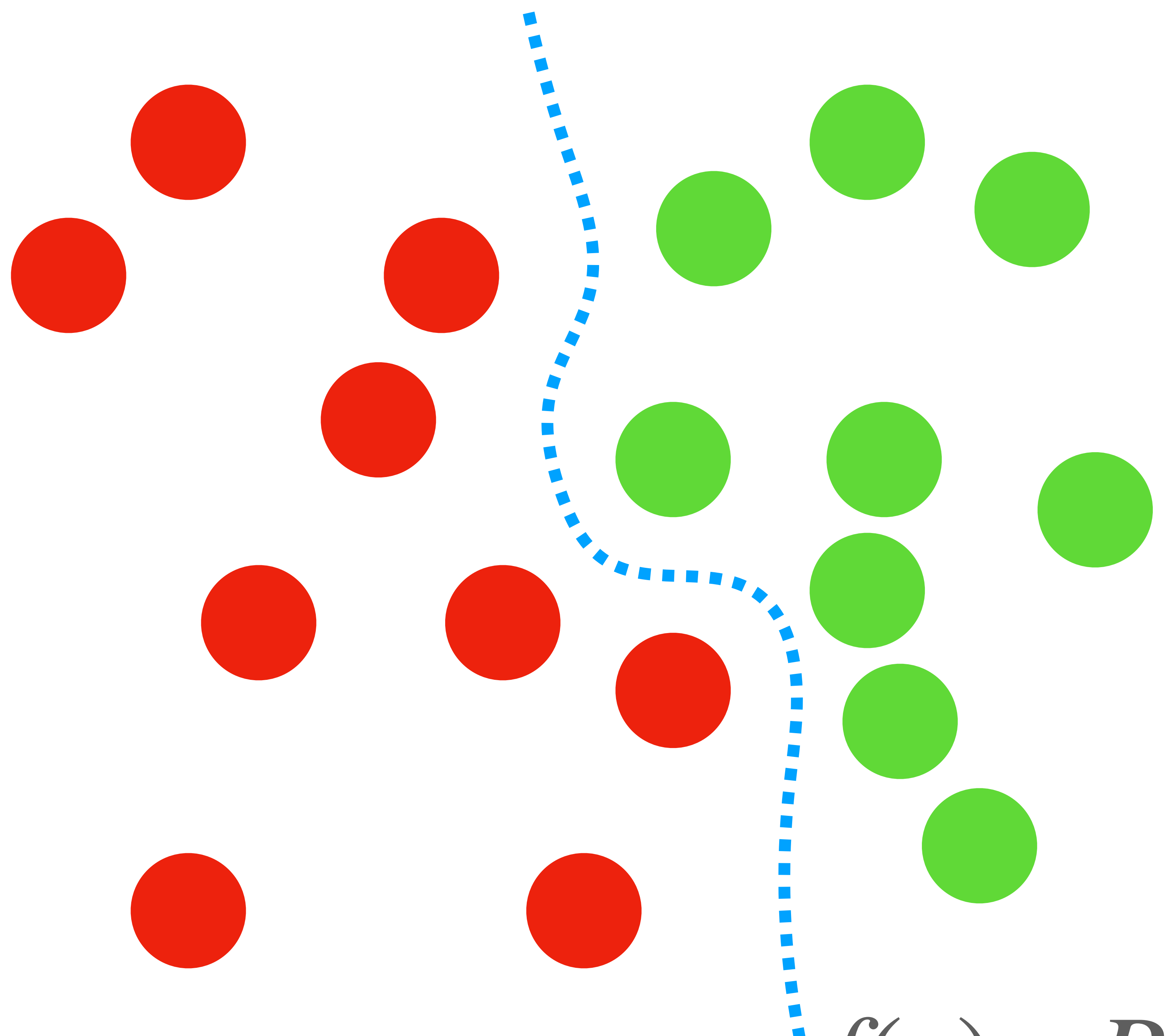


$$f(x) : R^n \rightarrow [0,1]$$

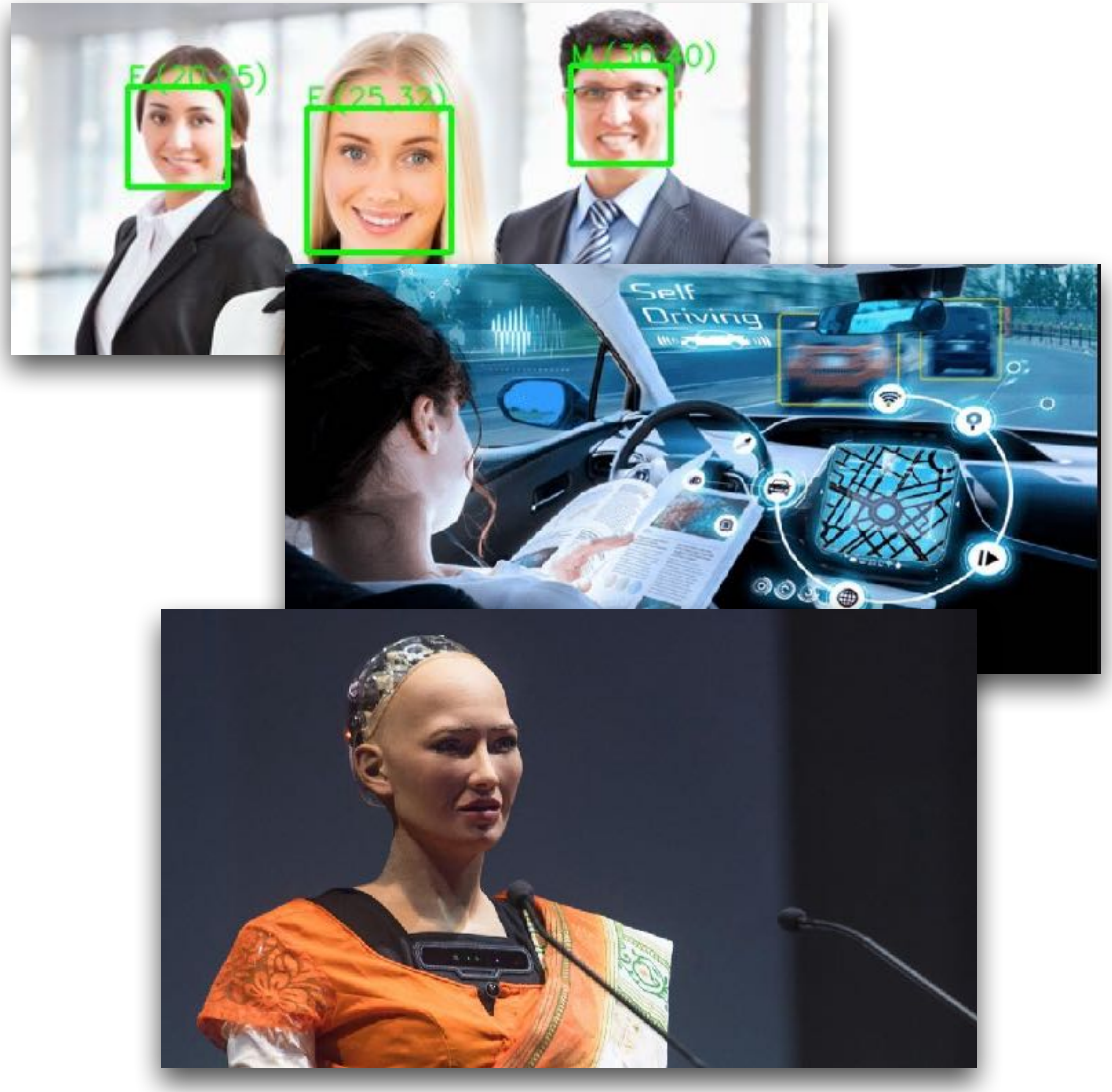
Reverse engineering of gene regulatory nets

SIRENE

Machine Learning



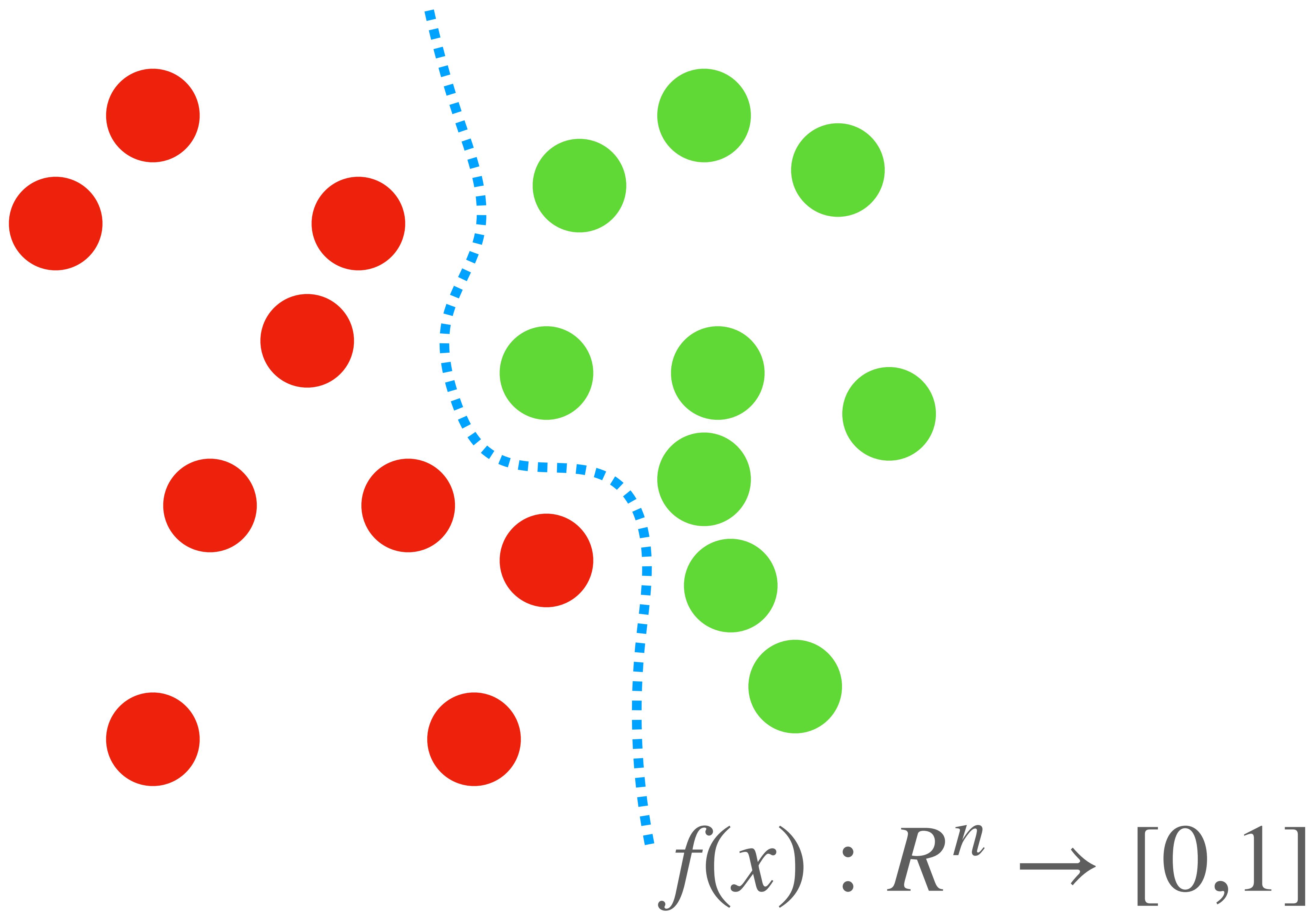
$$f(x) : R^n \rightarrow [0,1]$$



Reverse engineering of gene regulatory nets

SIRENE

Machine Learning


$f(x) : R^n \rightarrow [0,1]$

Many algorithms

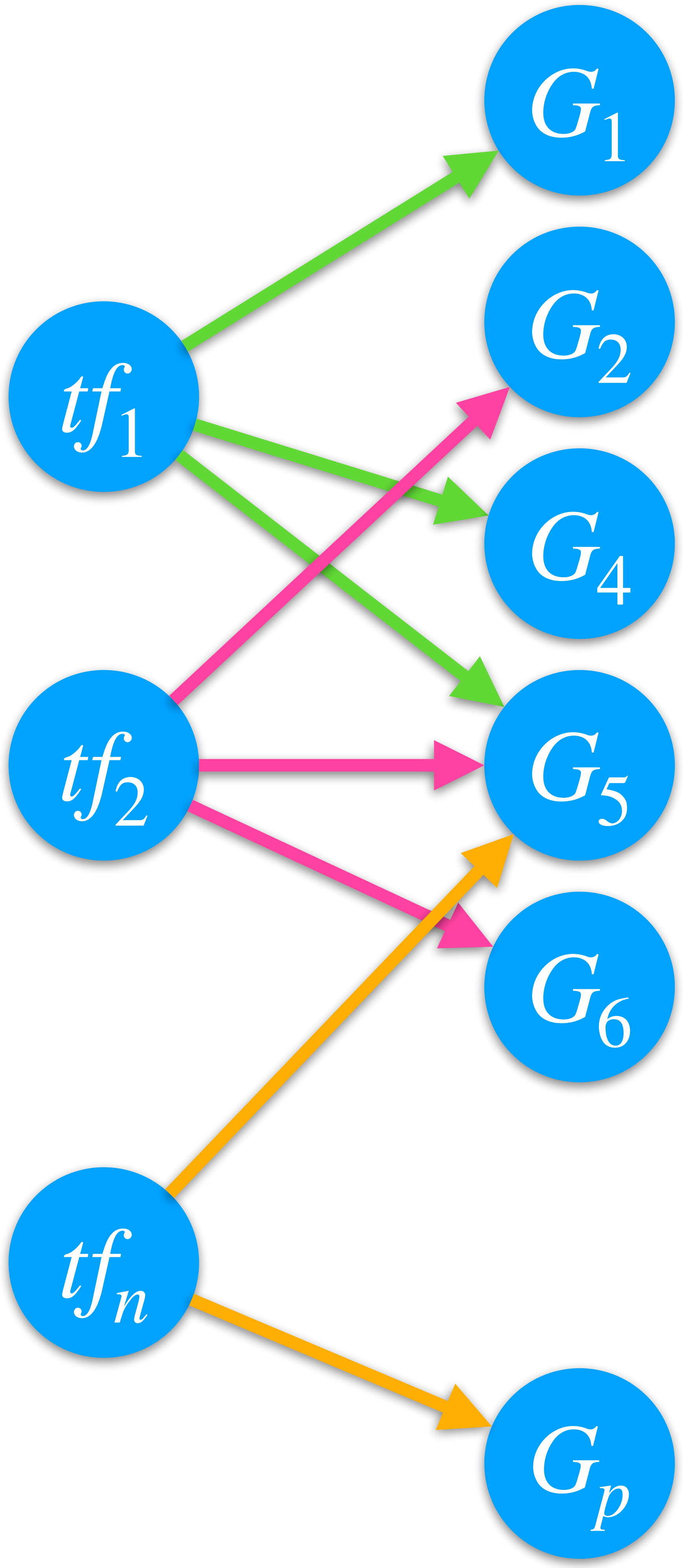
- kNN
- Naive Bayes
- Decision Tree
- Random Forest
- Support Vector Machine
- Deep Learning (Neural Networks)

Reverse engineering of gene regulatory nets

SIRENE

Local Model

for each Transcription Factor (tf_i), learn to discriminate the regulated vs non-regulated genes



Reverse engineering of gene regulatory nets

SIRENE

Local Model (One-class approach)

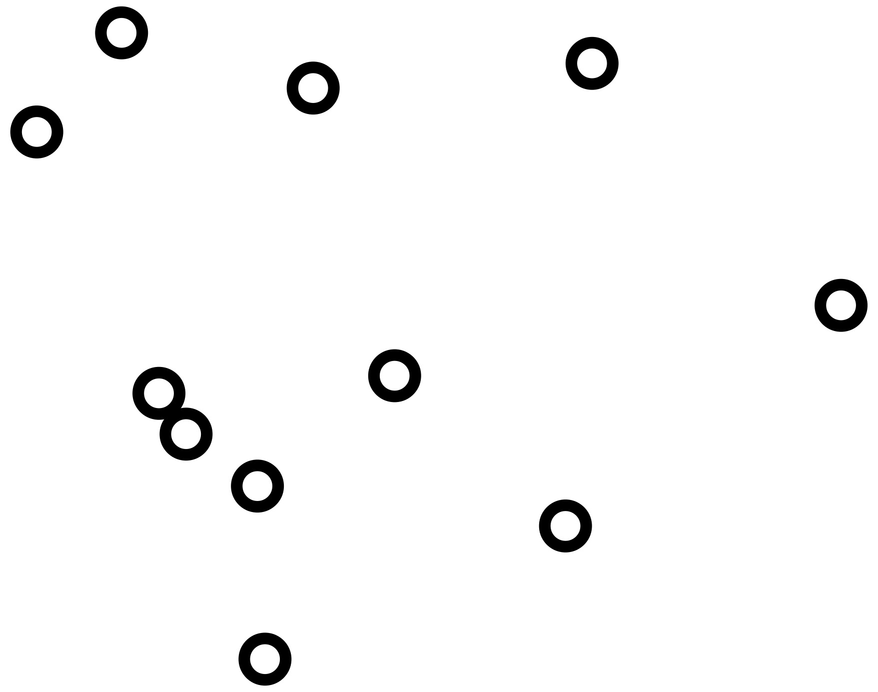
for each Transcription Factor (tf_i), learn a score $S_{tf_i}(e(G_i))$ function to assess the similarity with the set of known positive expression profiles. Then classify the unknown genes decreasing score.

Reverse engineering of gene regulatory nets

SIRENE

Local Model (One-class approach)

for each Transcription Factor (tf_i), learn a score $S_{tf_i}(e(G_i))$ function to assess the similarity with the set of known positive expression profiles. Then classify the unknown genes decreasing score.



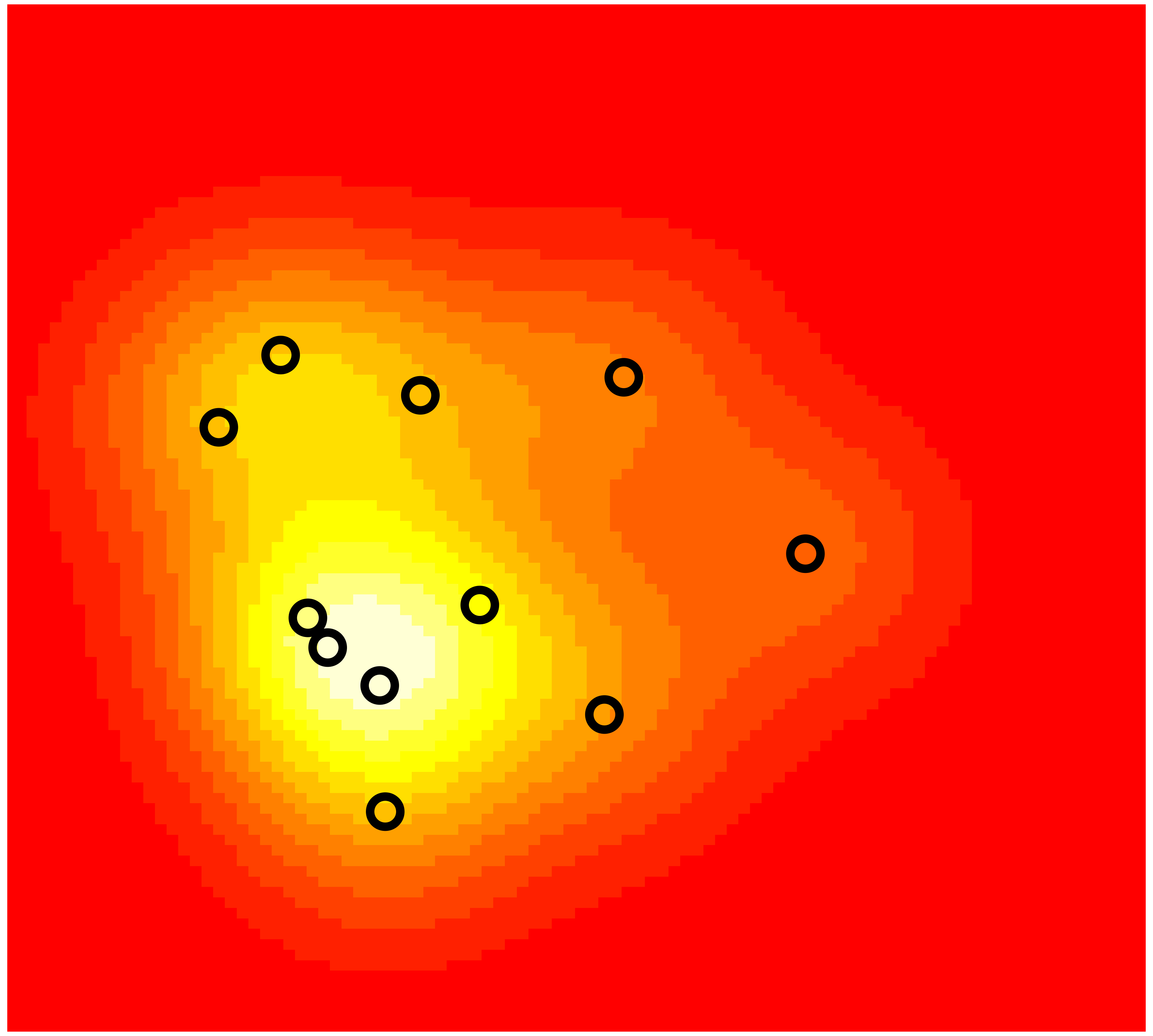
Known Positives
regulated by tf_i

Reverse engineering of gene regulatory nets

SIRENE

Local Model (One-class approach)

for each Transcription Factor (tf_i), learn a score $S_{tf_i}(e(G_i))$ function to assess the similarity with the set of known positive expression profiles. Then classify the unknown genes decreasing score.



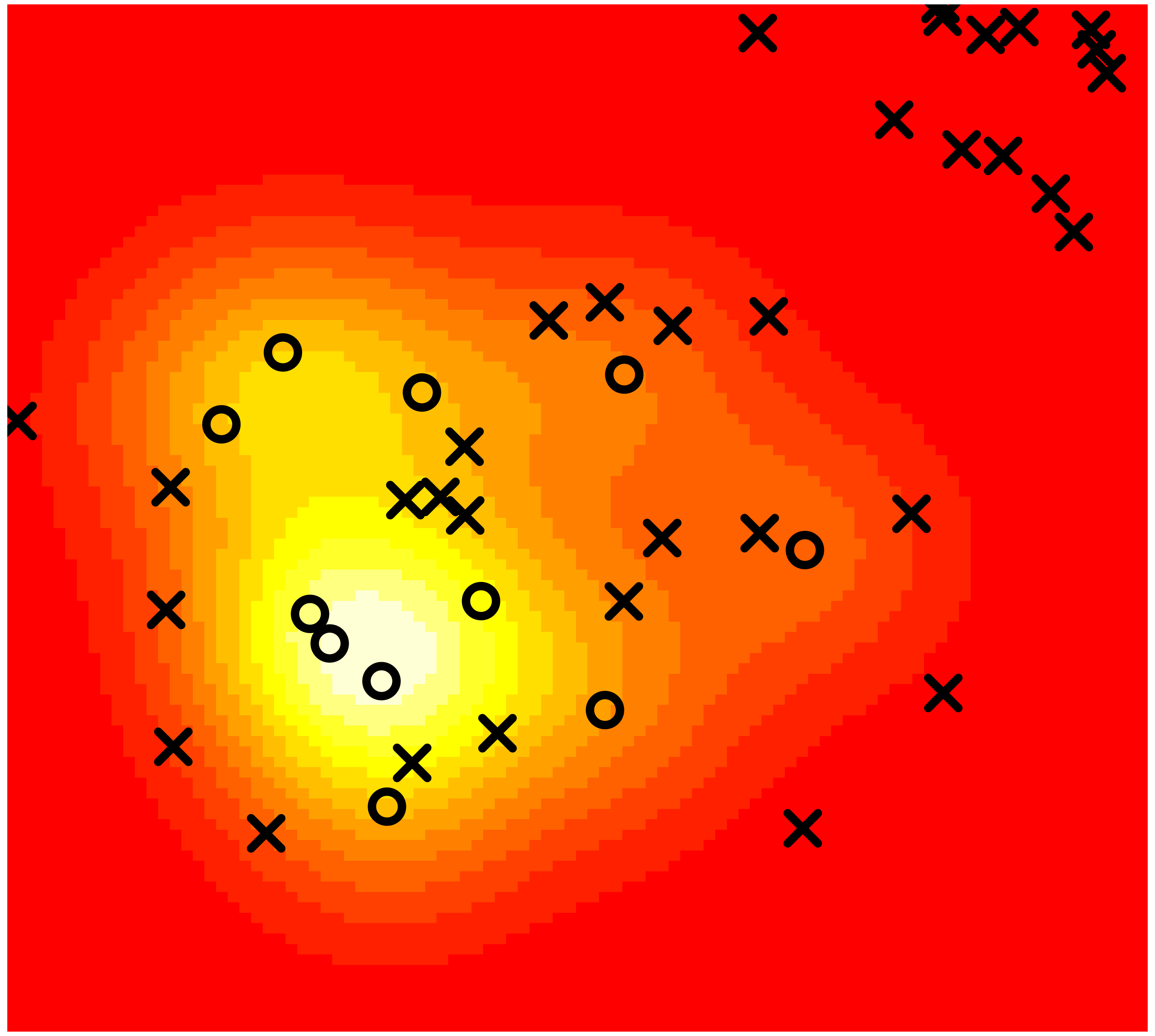
Learned score function $S_{tf_i}()$

Reverse engineering of gene regulatory nets

SIRENE

Local Model (One-class approach)

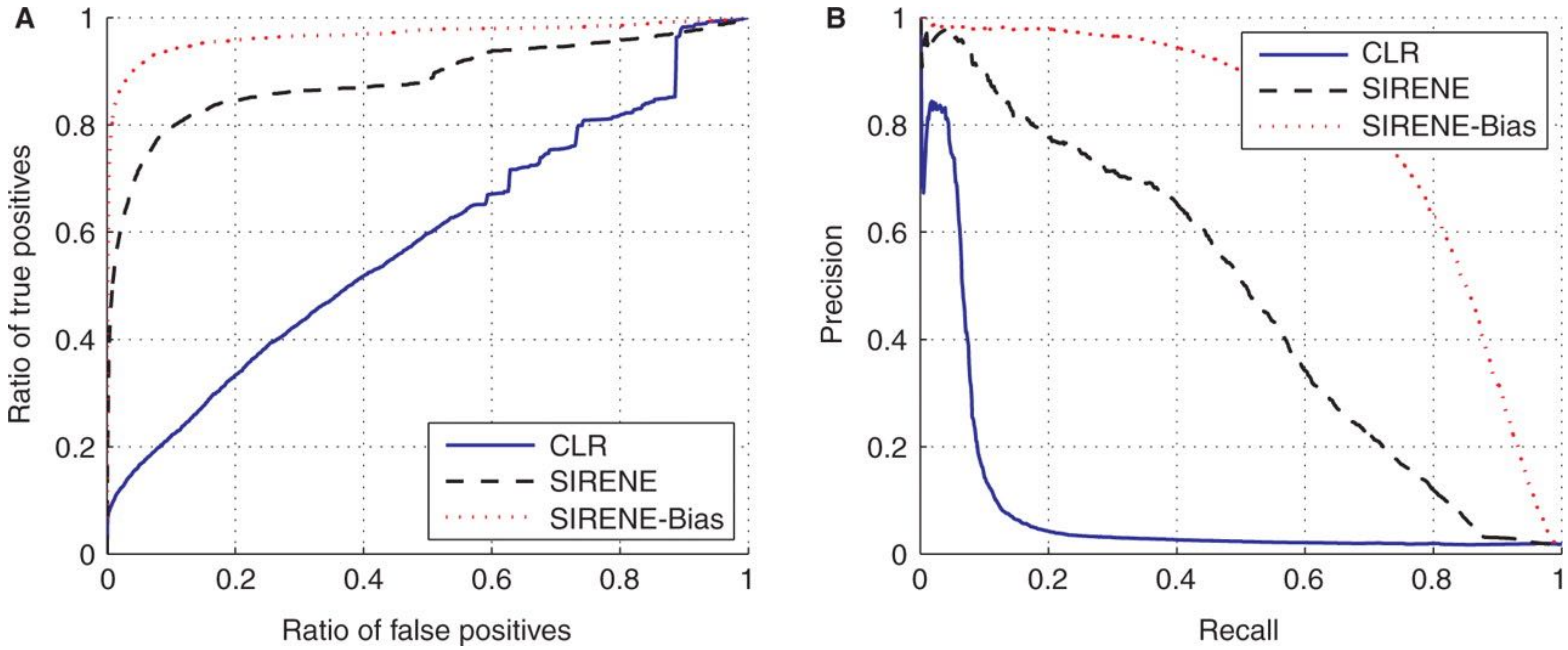
for each Transcription Factor (tf_i), learn a score $S_{tf_i}(e(G_i))$ function to assess the similarity with the set of known positive expression profiles. Then classify the unknown genes decreasing score.



Classification of
Unknown genes

Reverse engineering of gene regulatory nets

SIRENE

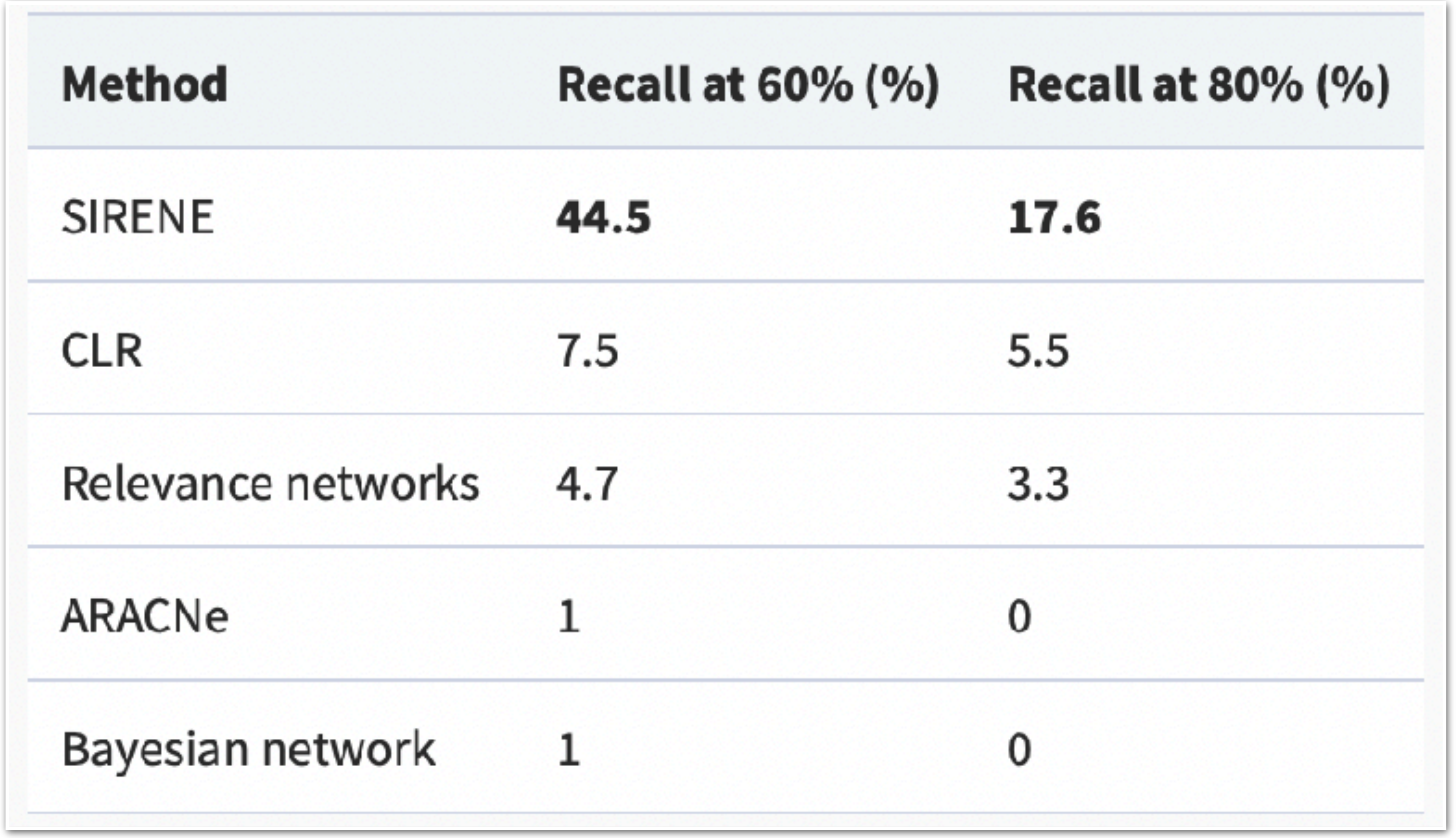


Mordelet and Vert, Bioinformatics, 2008

E.coli benchmark used by Faith et al., (2007)

Reverse engineering of gene regulatory nets

SIRENE



Method	Recall at 60% (%)	Recall at 80% (%)
SIRENE	**44.5**	**17.6**
CLR	7.5	5.5
Relevance networks	4.7	3.3
ARACNe	1	0
Bayesian network	1	0

Mordelet and Vert, Bioinformatics, 2008

E.coli benchmark used by Faith et al., (2007)

Reverse engineering of gene regulatory nets

SIRENE

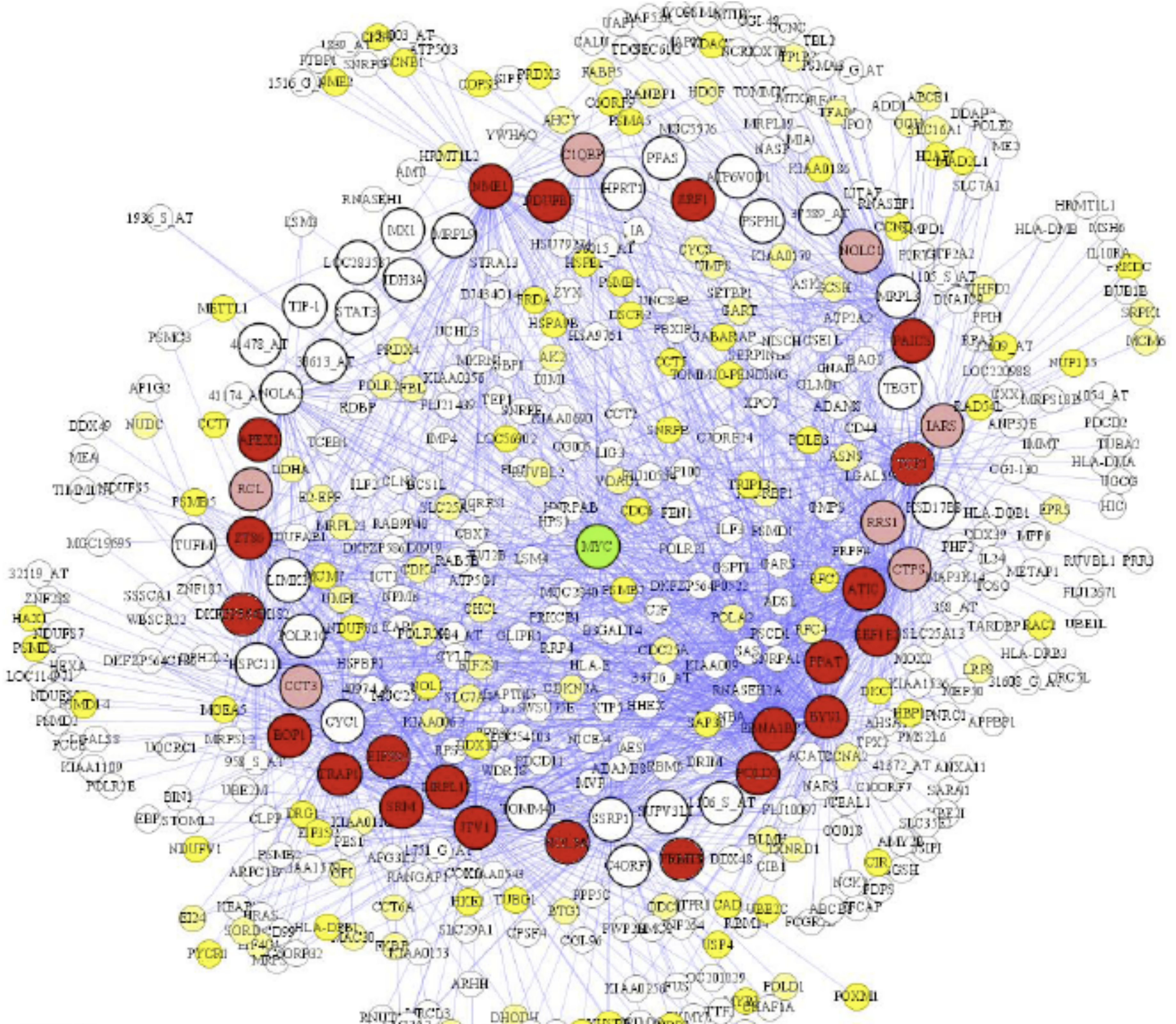
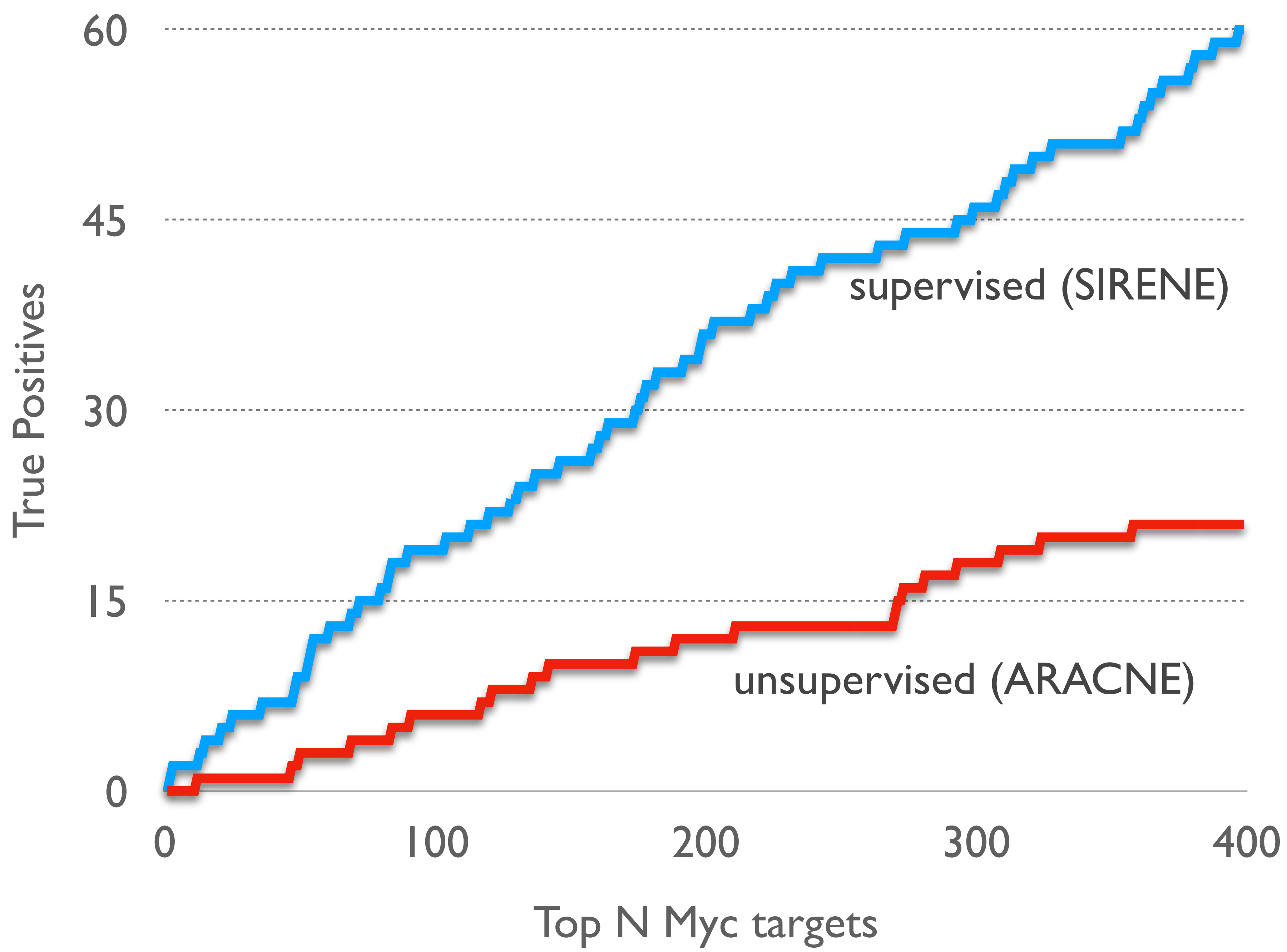
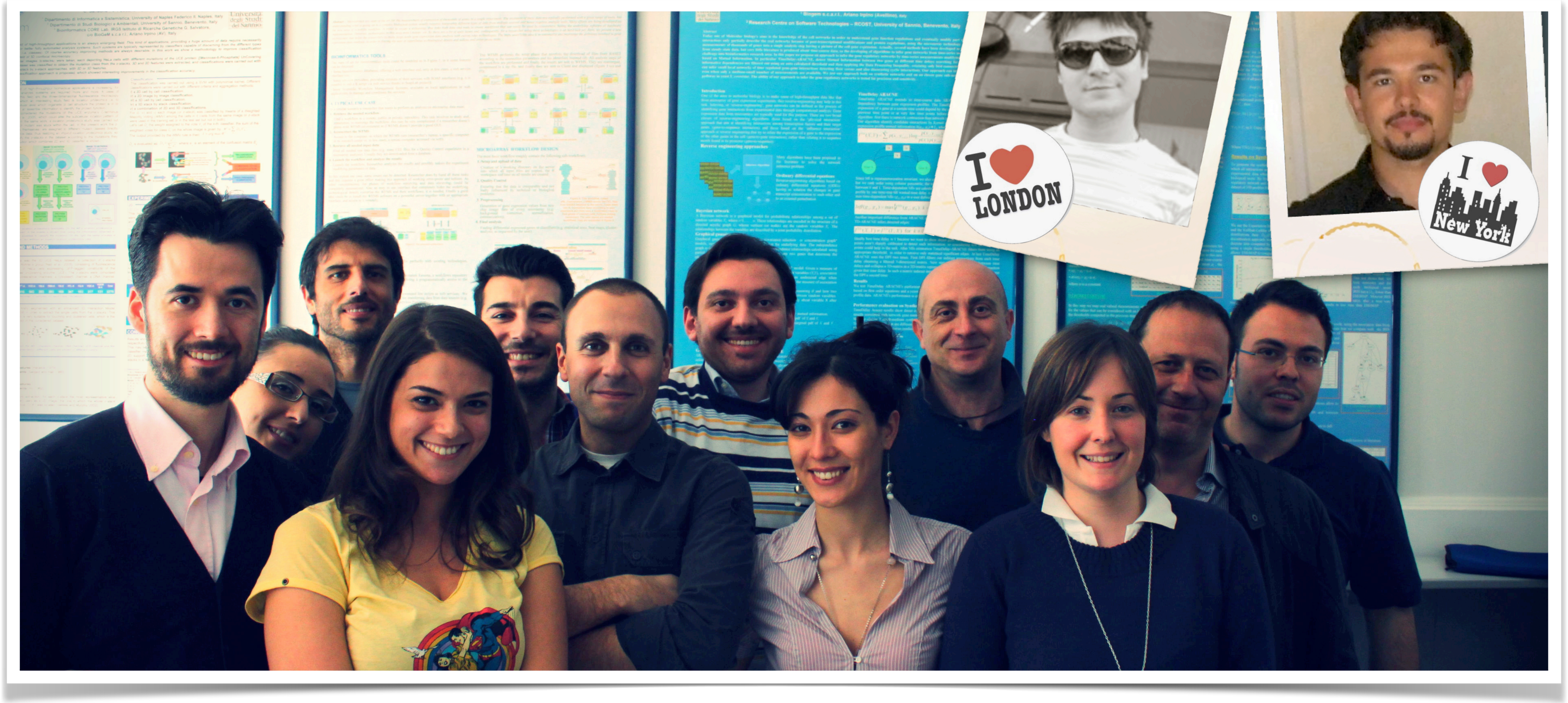


Figure 4a



Our bioinformatics group

Michele Ceccarelli
Luigi Cerulo
Francesco Napolitano
Francesca P. Caruso
Teresa M.R. Noviello
Rossella Tufano
Maria Claudia Costa
Luigi Ferraro
Antonio De Falco



2008 Bioinformatics @Unisannio @Biogem

2022 Bioinformatics @Unina @Unisannio @Biogem

Our bioinformatics group

...see you at lab session

2008 Bioinformatics @Unisannio @Biogem

2022 Bioinformatics @Unina @Unisannio @Biogem