

# Reverse engineering of gene regulatory networks

applications in cancer datasets and beyond

**Luigi Cerulo**

Dep. of Science and Technology  
University of Sannio



biogem

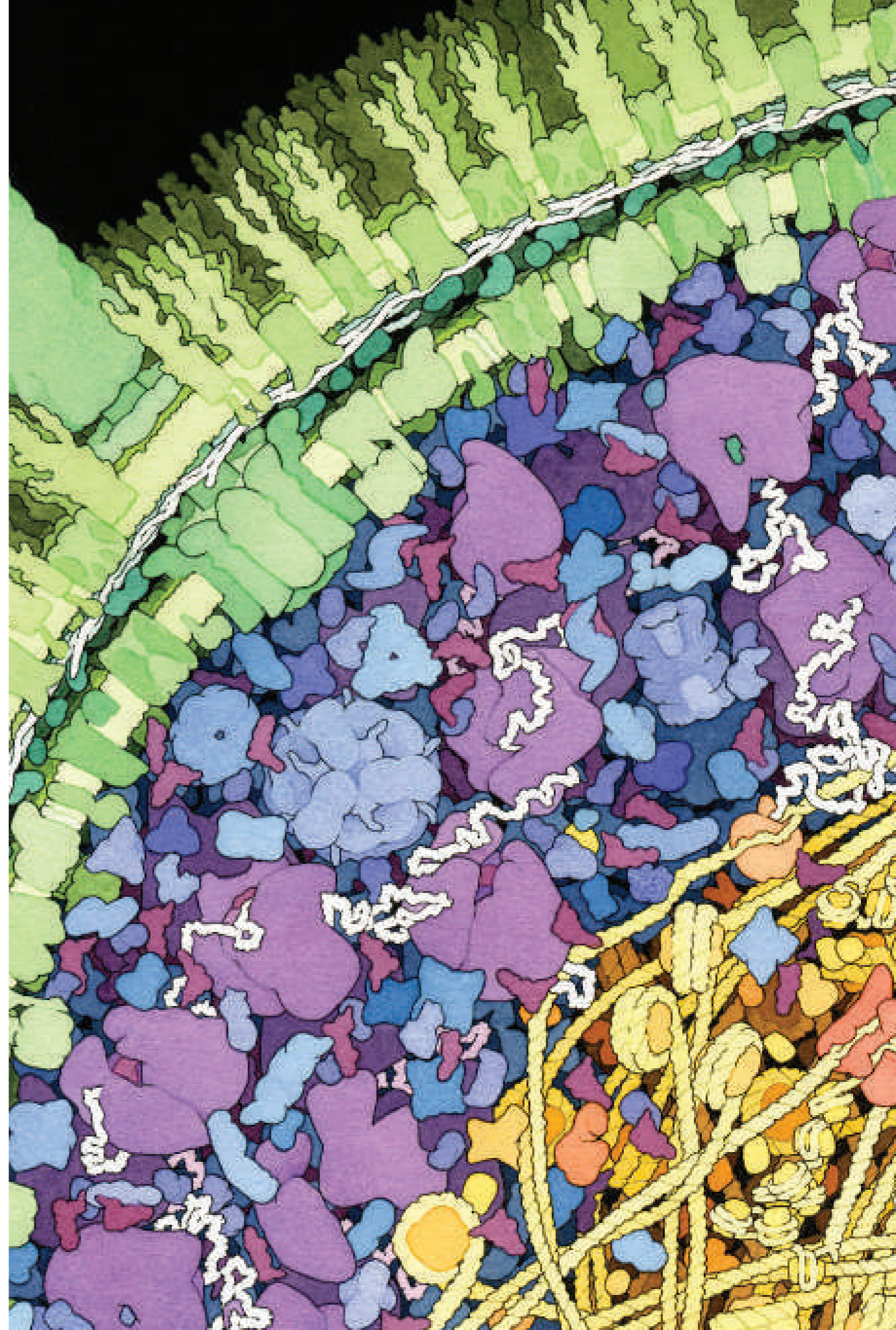
Nov, 2016

# System Biology

---

- A cell can be considered a **complex system** with a huge number of interconnected components
- It is necessary a systemic approach to understand how the cell is organized and how genes and proteins interact  
(Ludwig von Bertalanffy, 1934)

Nature, 2007



# How to understand a complex system?

---

- Forward Engineering

“What I cannot **create/simulate**, I do not understand”



# How to understand a complex system?

---

- Forward Engineering

“What I cannot **create/simulate**, I do not understand”



# How to understand a complex system?

---

- Reverse Engineering

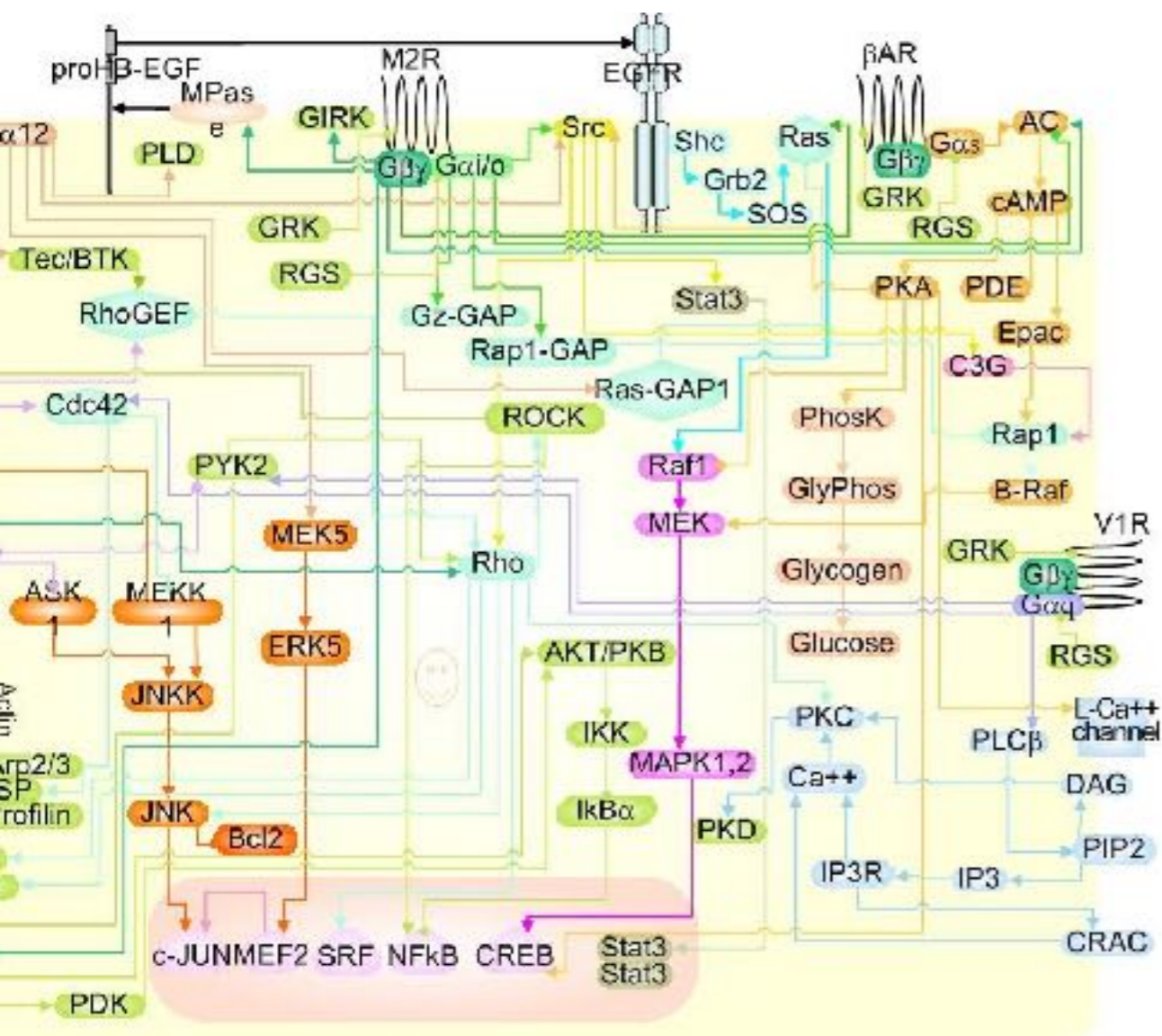
“What I cannot **break**, I do not understand”



# How to understand a complex system?

- Reverse Engineering

“What I cannot **break**, I do not understand”

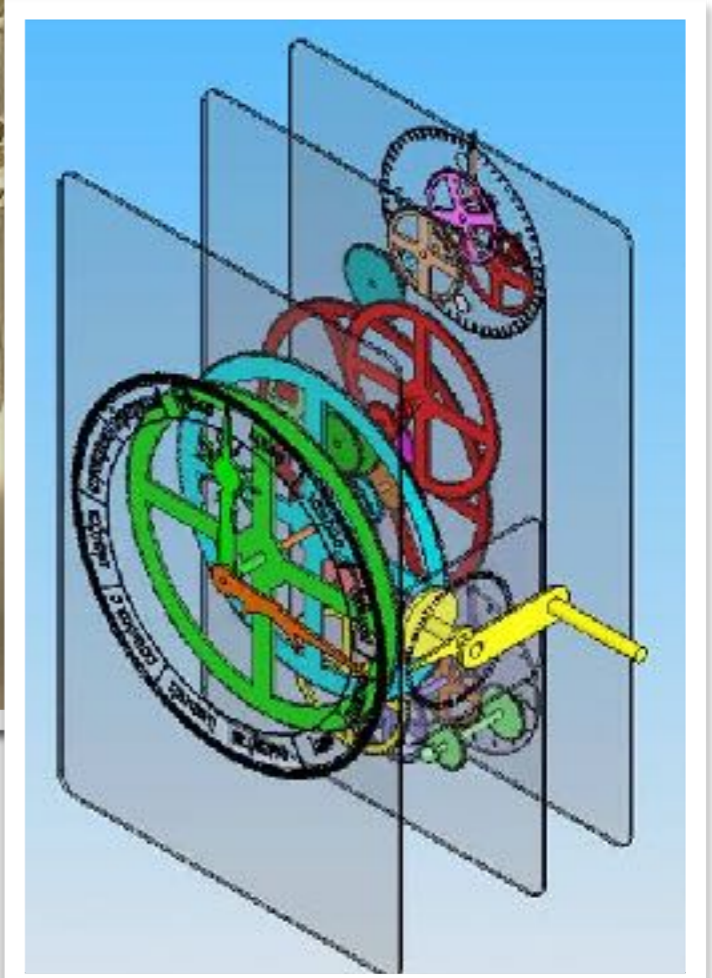
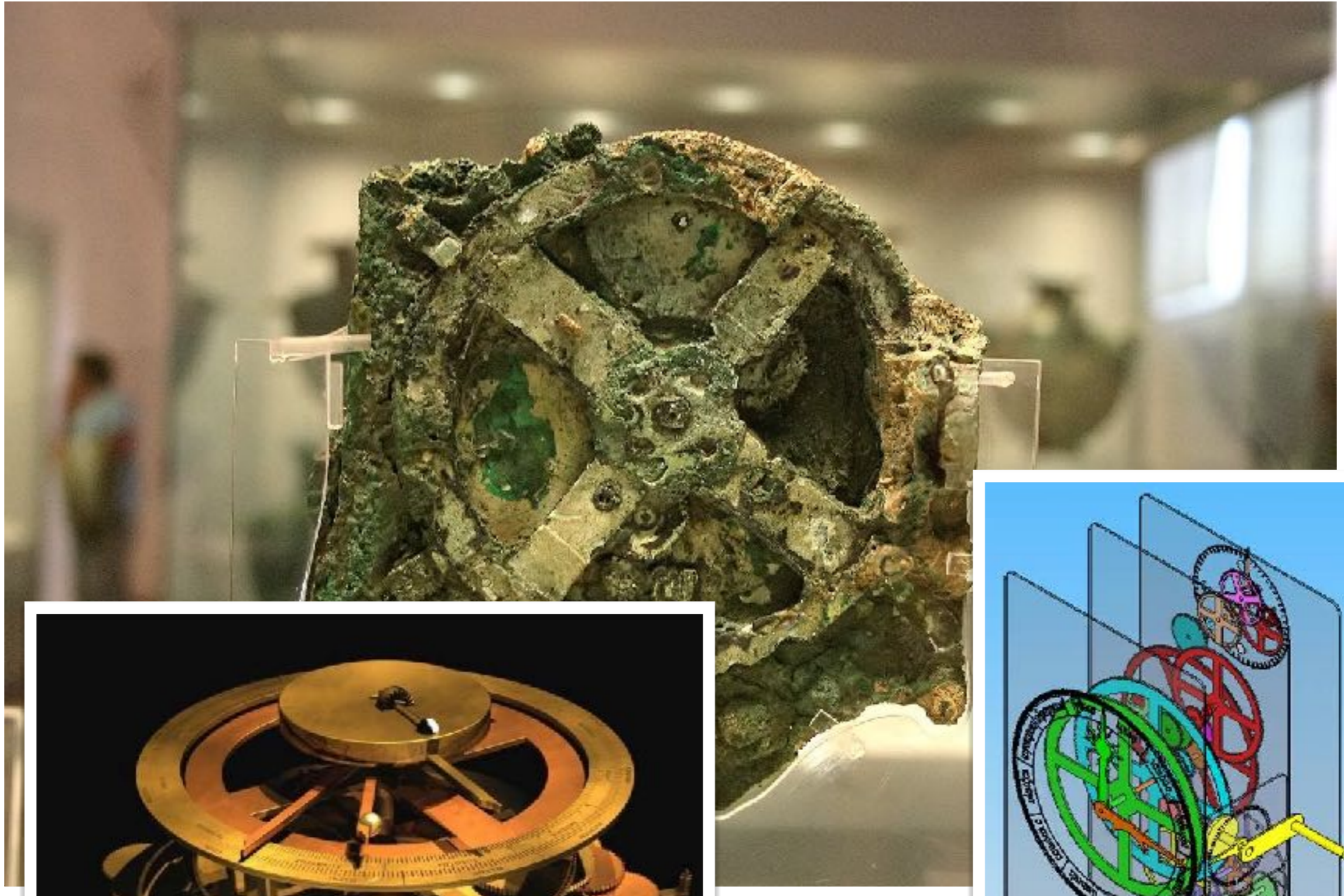






The Antikythera mechanism (National Museum Athen, probably 65 a.C.)

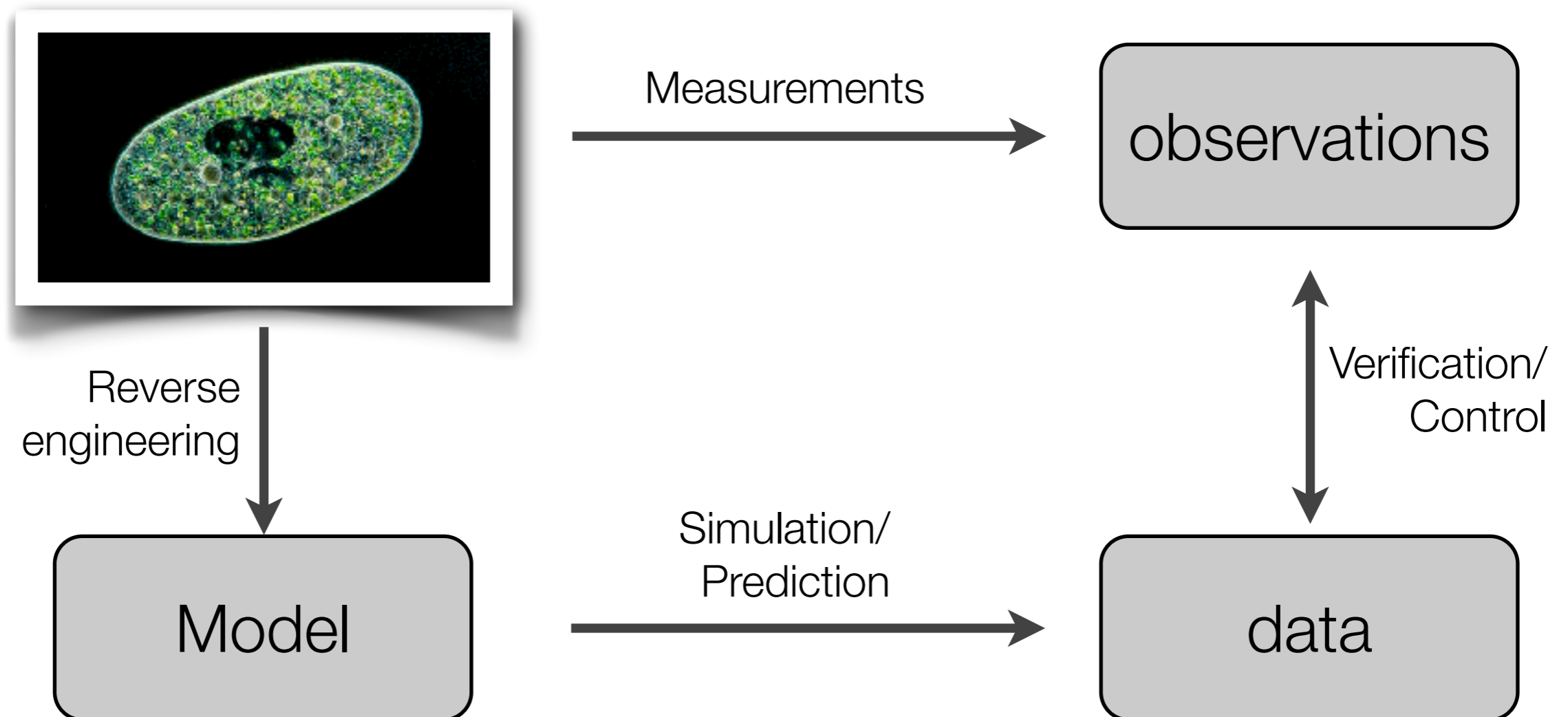




# Reverse engineering

---

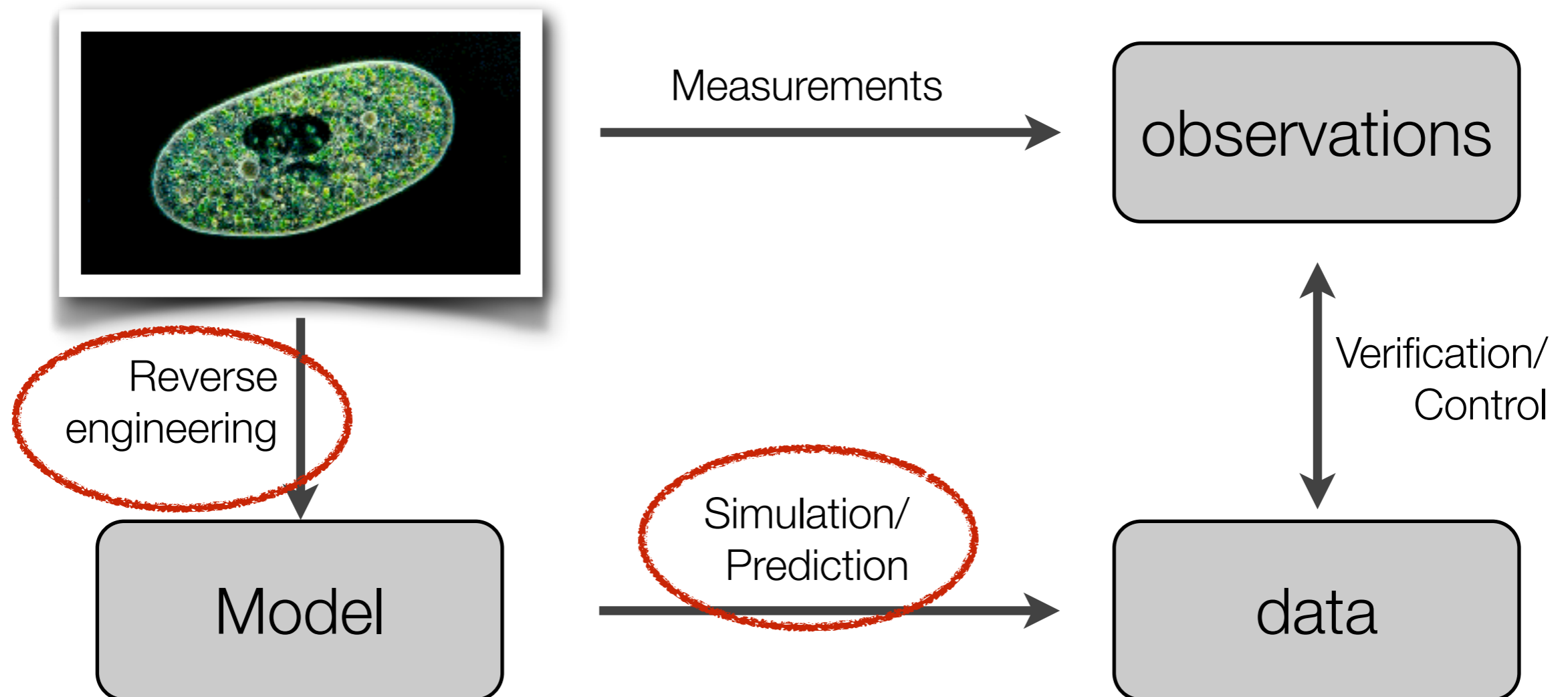
- The Reverse Engineering of a biological system consists of building a mathematical model that is able to describe/simulate (part of) its behavior



# Reverse engineering

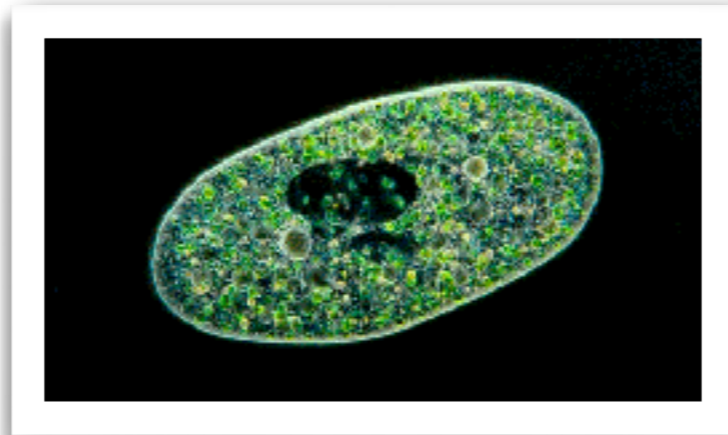
---

- The Reverse Engineering of a biological system consists of building a mathematical model that is able to describe/simulate (part of) its behavior



# Reverse engineering

---



Hypothesis/  
Model

# Reverse engineering

---



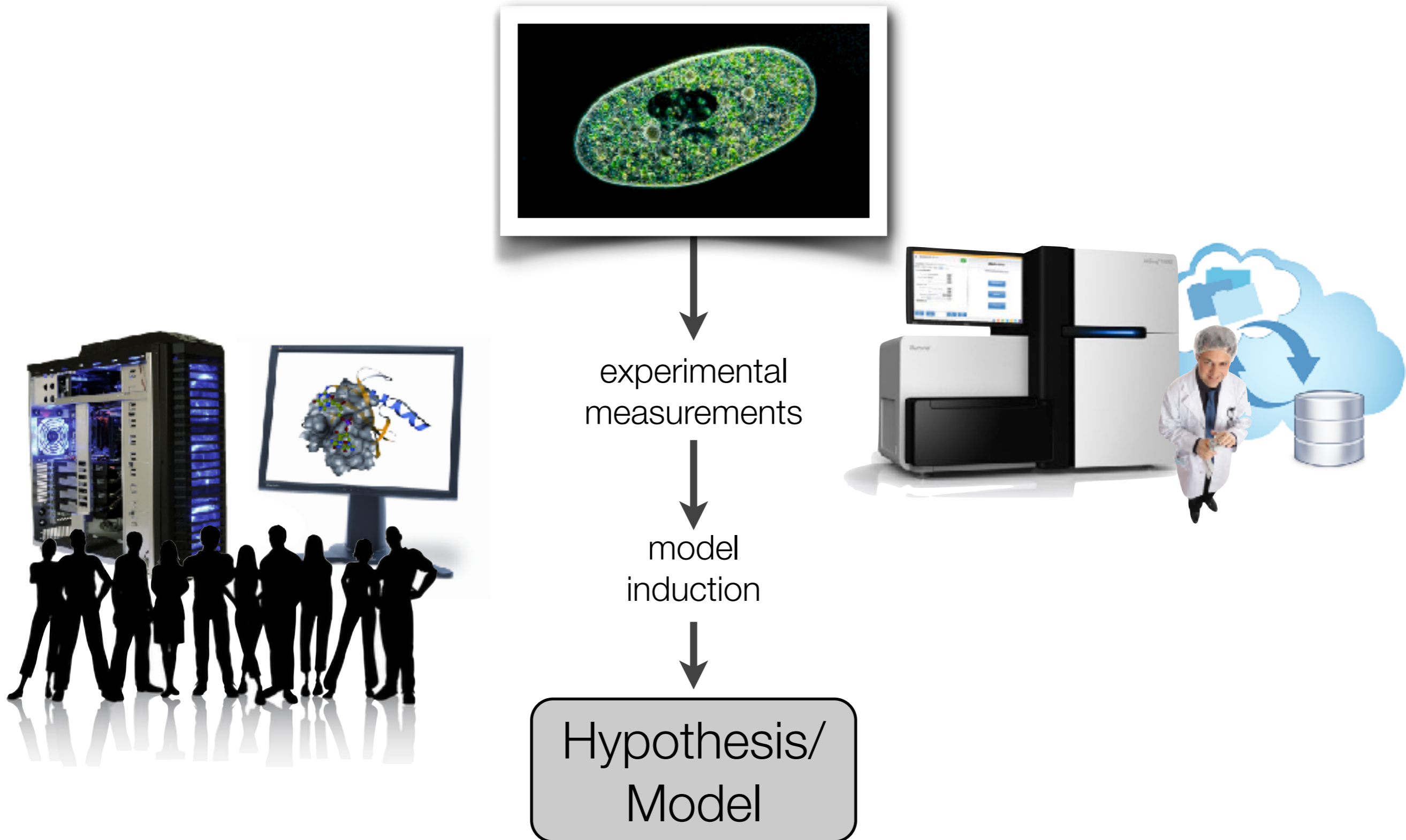
experimental  
measurements

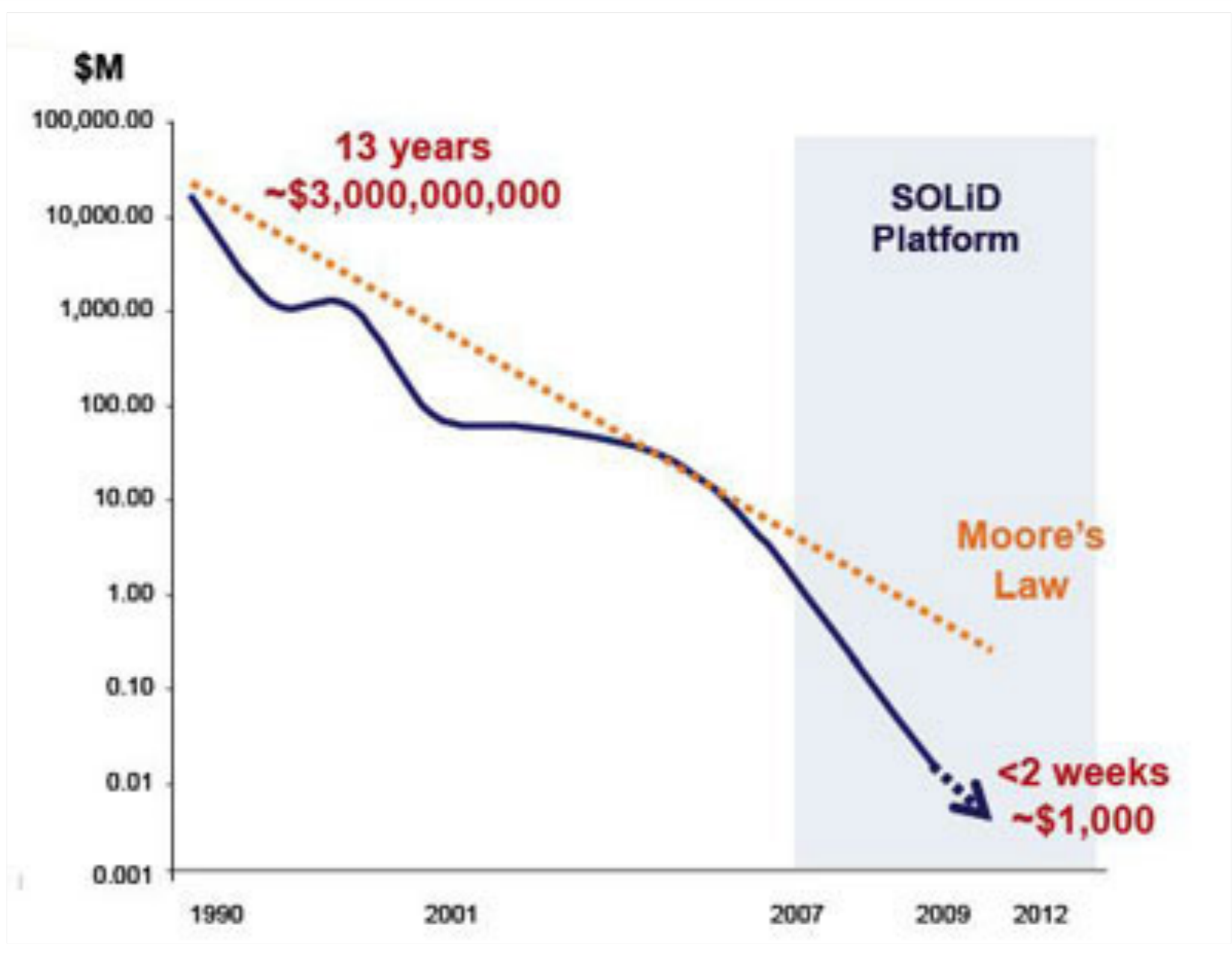


Hypothesis/  
Model

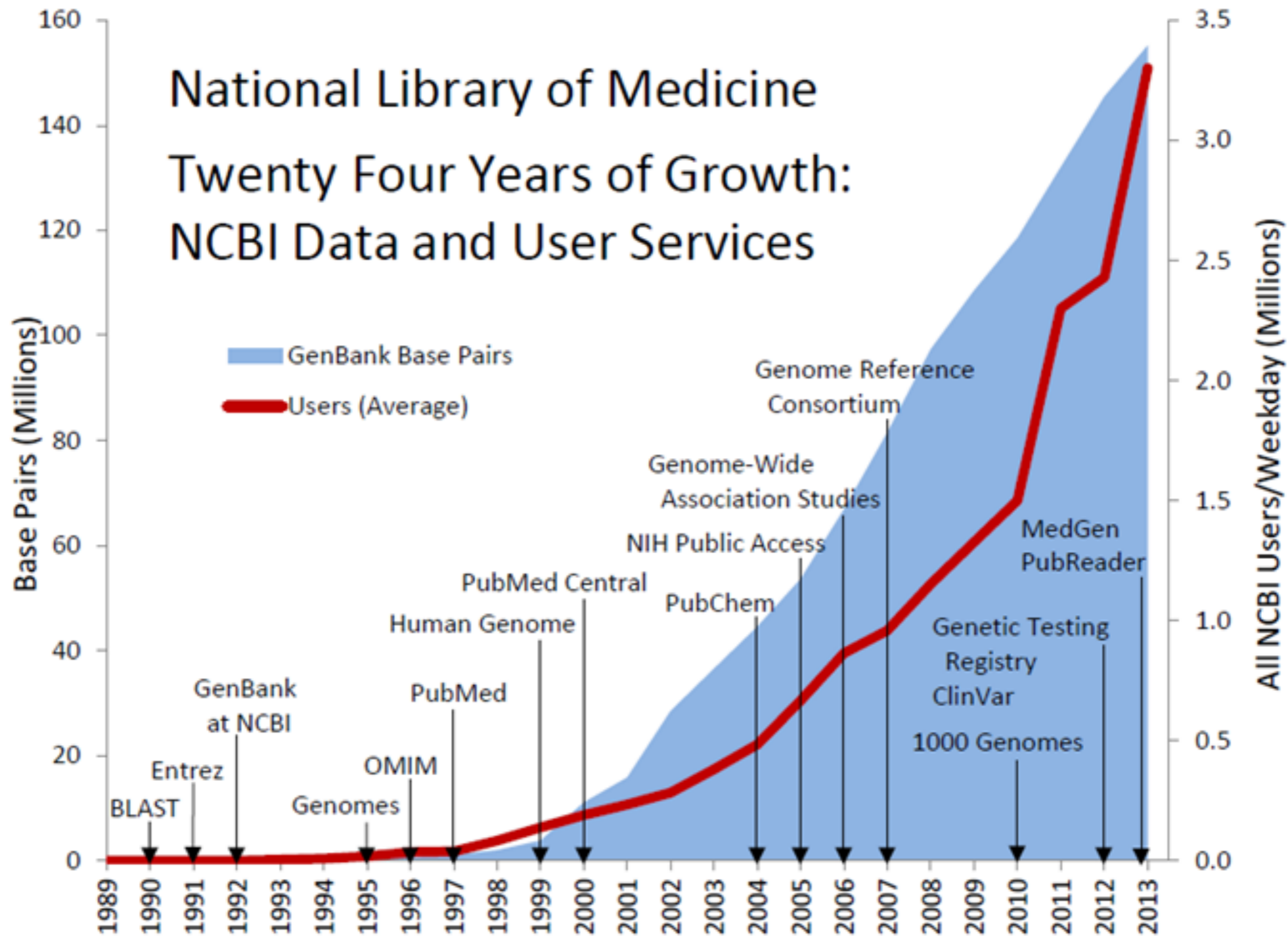
# Reverse engineering

---



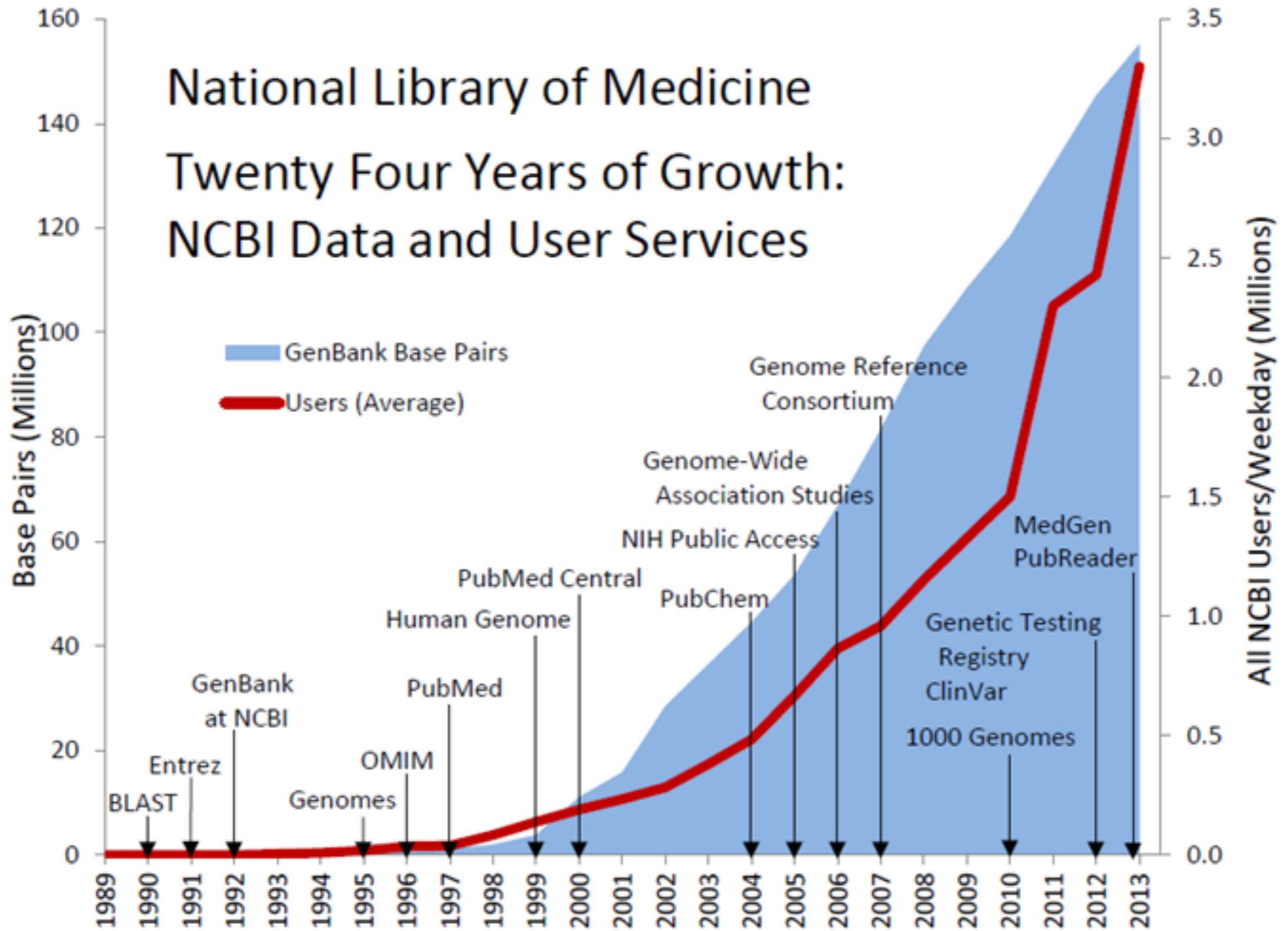


# National Library of Medicine Twenty Four Years of Growth: NCBI Data and User Services





# National Library of Medicine Twenty Four Years of Growth: NCBI Data and User Services



fits on 4 floppy disk



fits on 1 cd rom



If you torture the data long enough, it will confess anything

---



*“If you don't reveal some insights soon, I'm going to be forced to slice, dice, and drill!”*

# A cell has different control layers

---

- Key actors

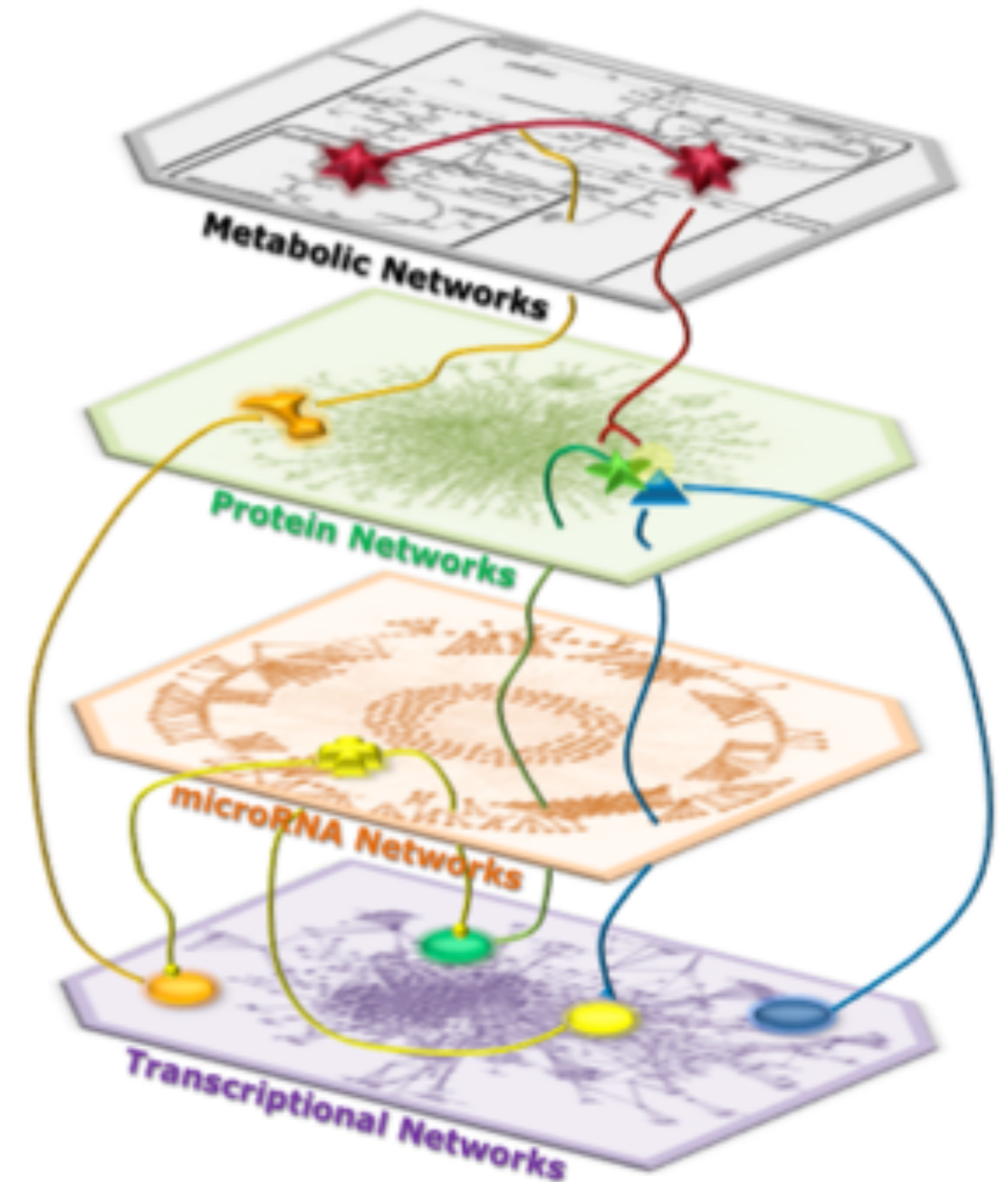
Transcription Factors

ncRNAs

Kinases

Enzymes

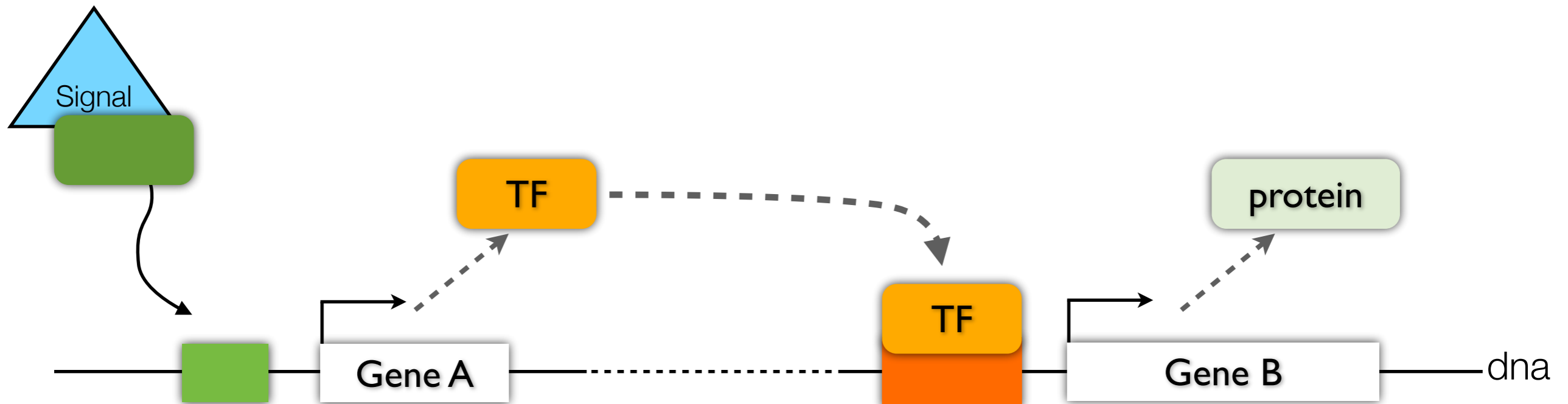
Metabolites



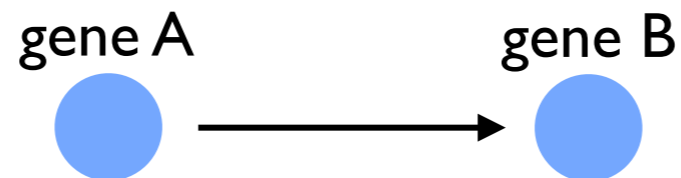
# Gene Space Layer

## gene interactions

---

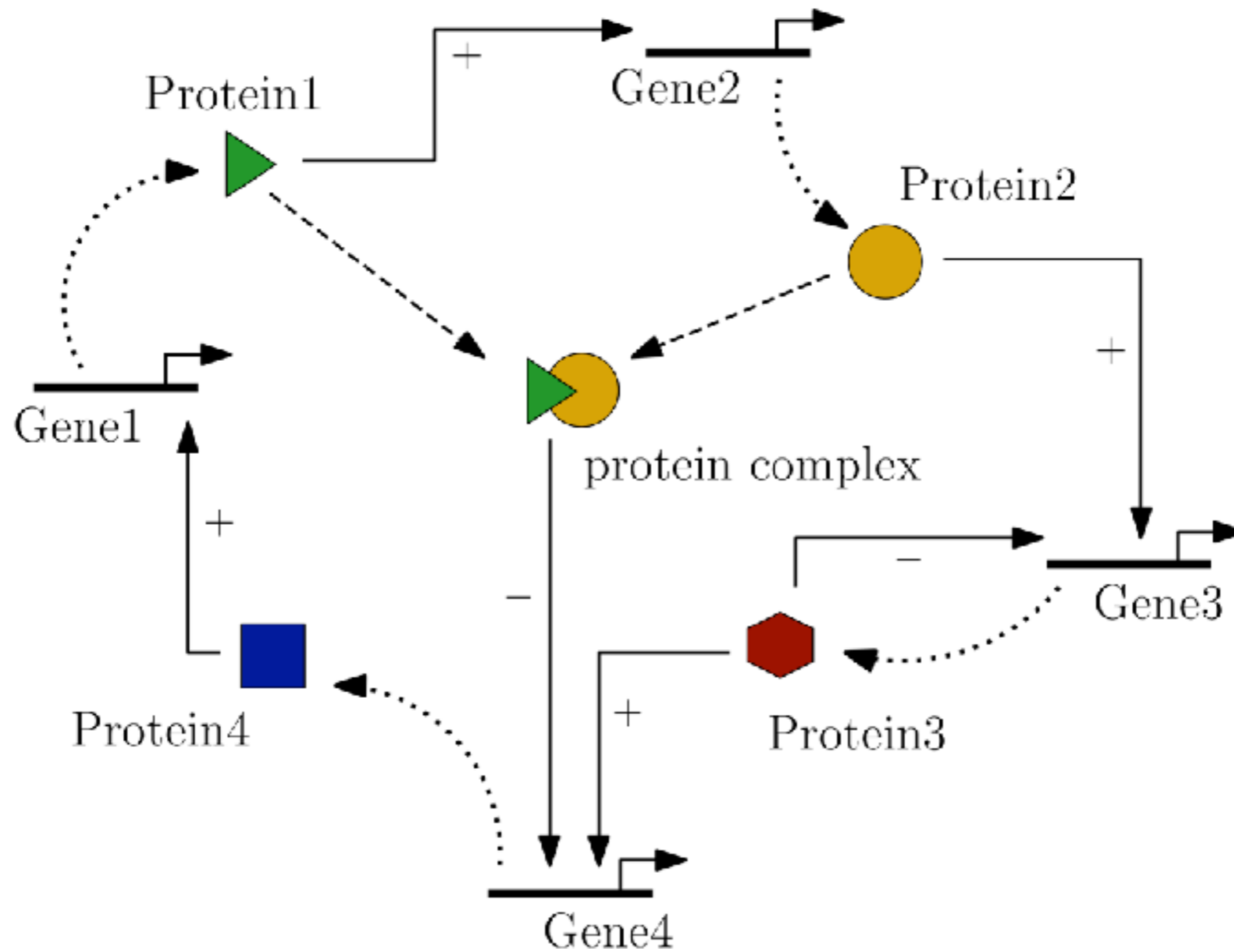


model

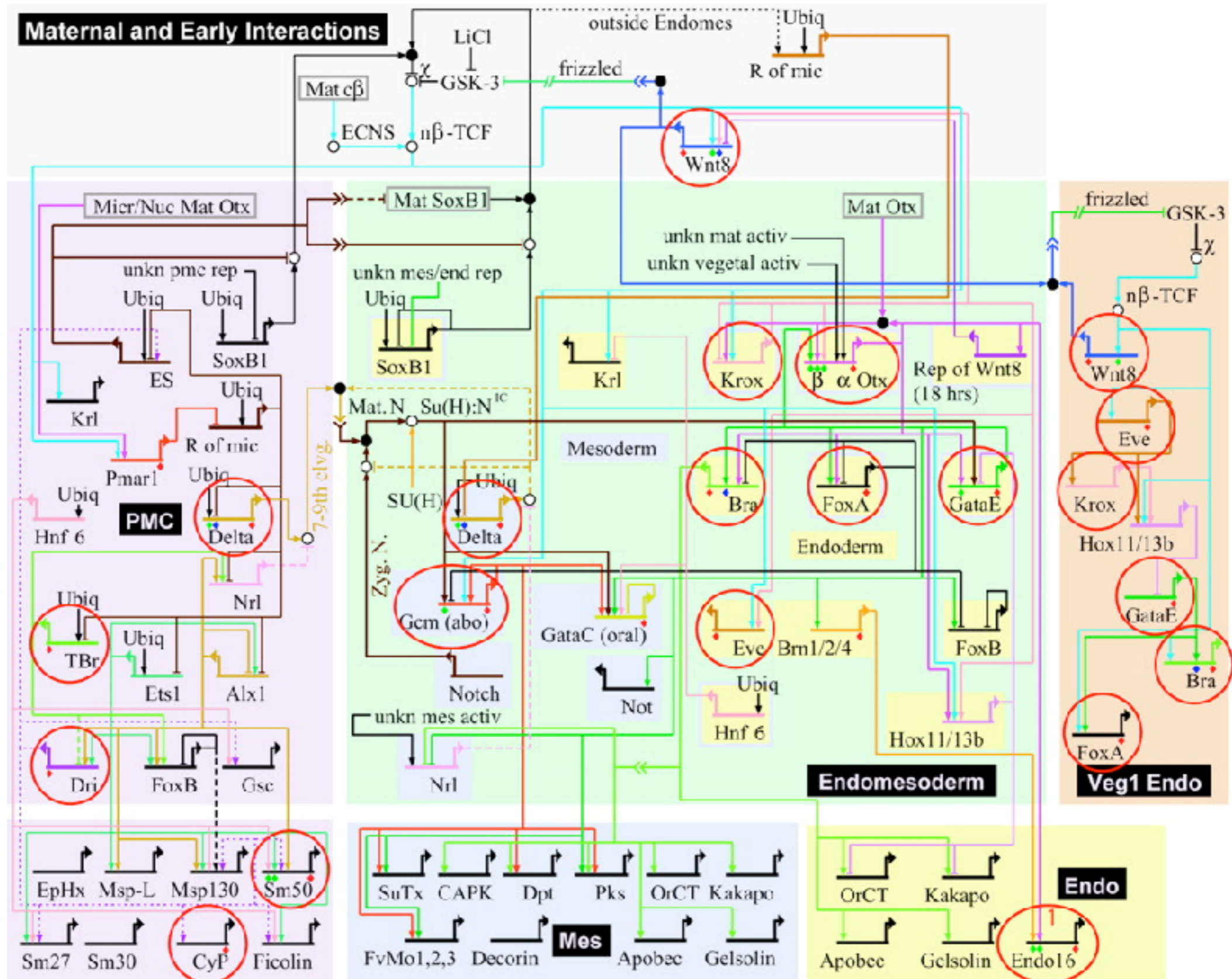


# Gene regulatory network

---



# Maternal and Early Interactions

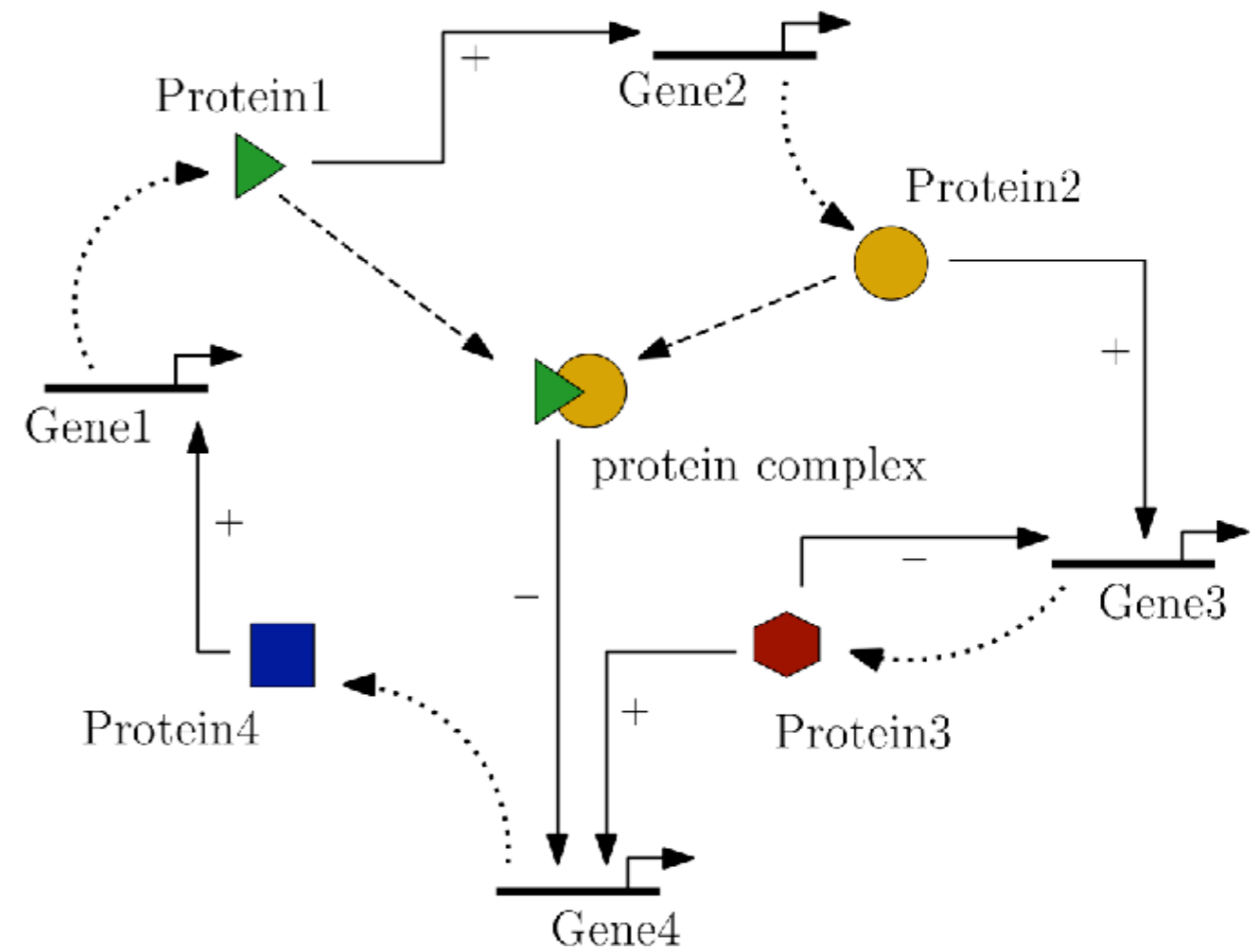






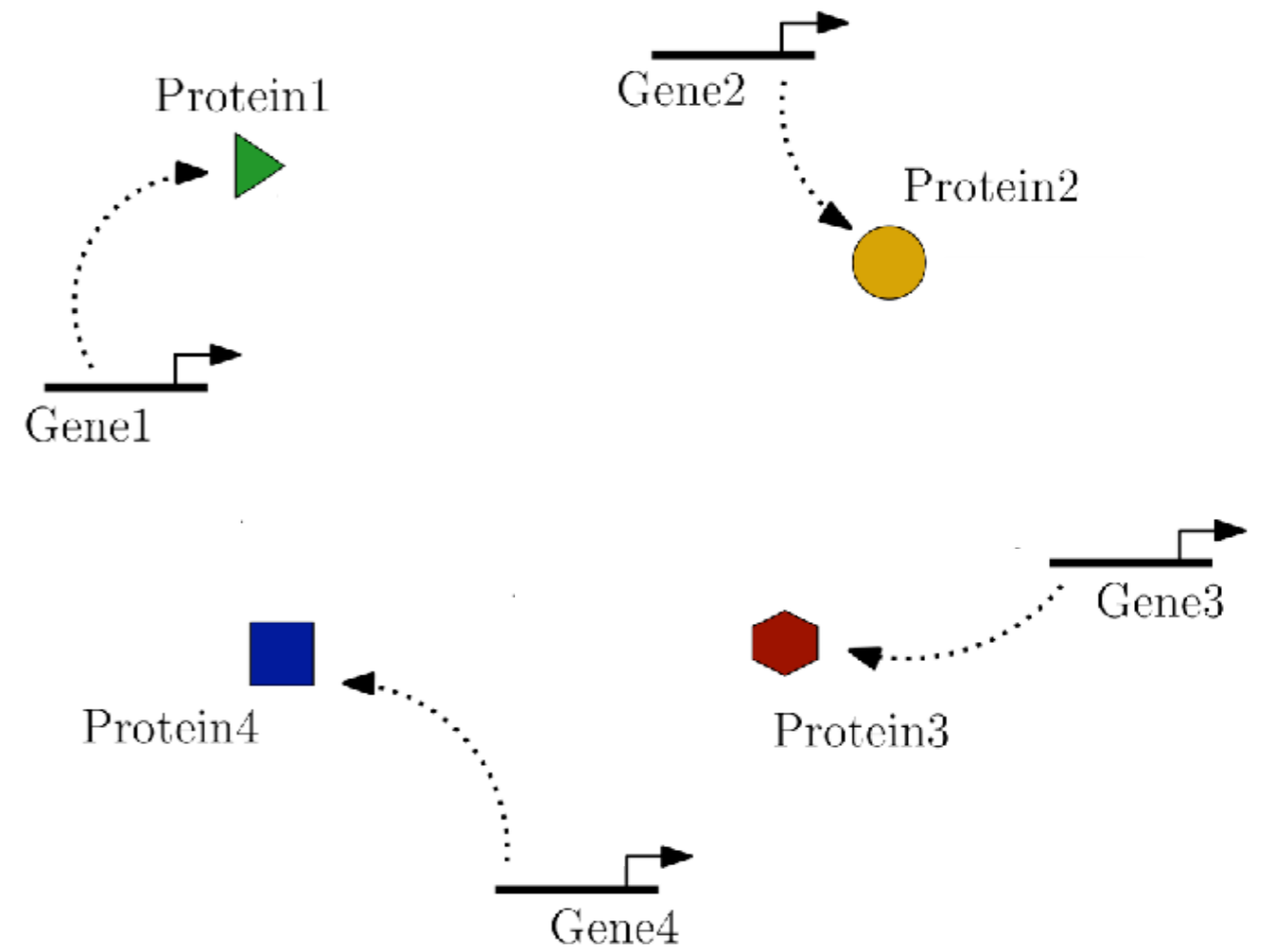
# Reverse engineering of gene regulatory networks

---



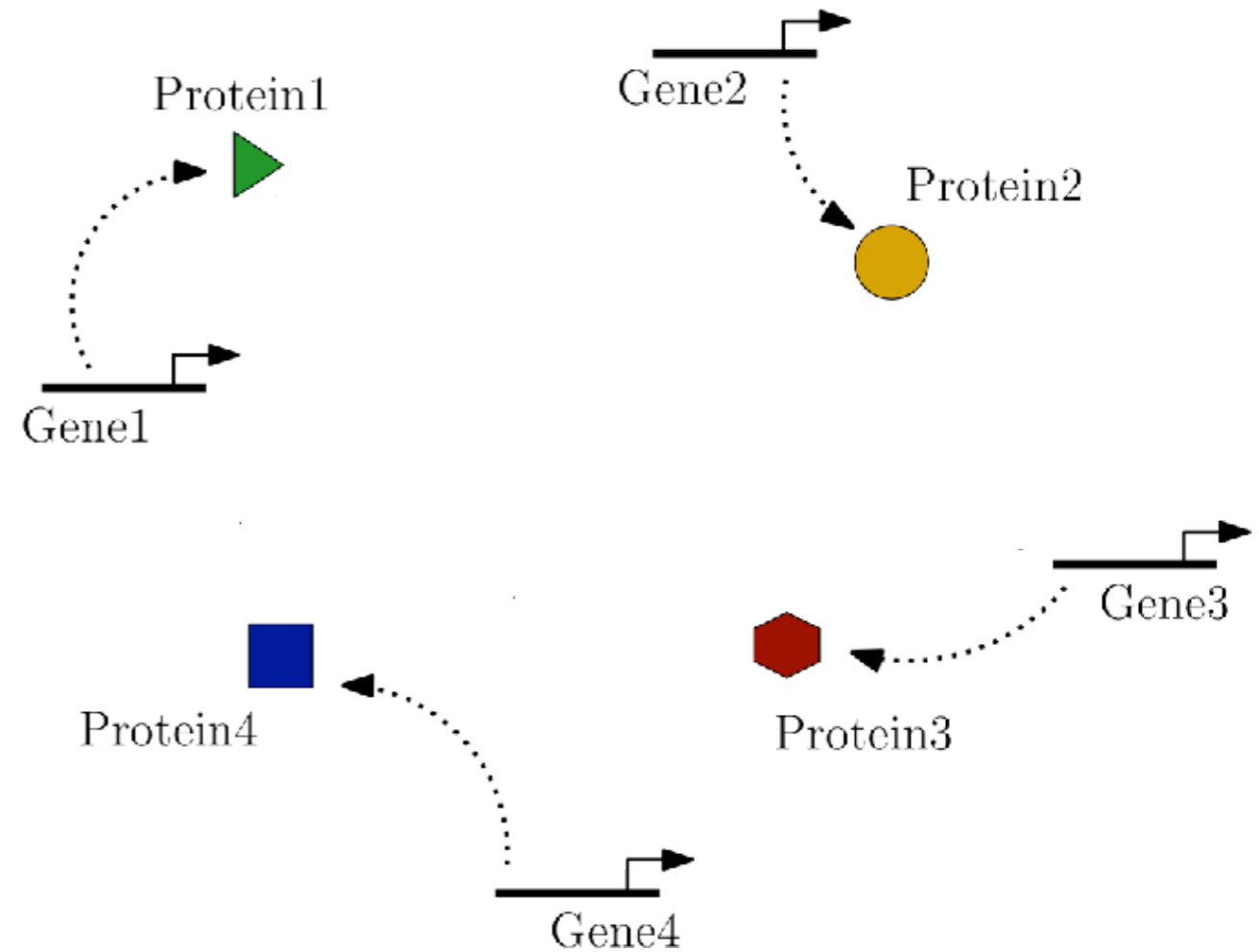
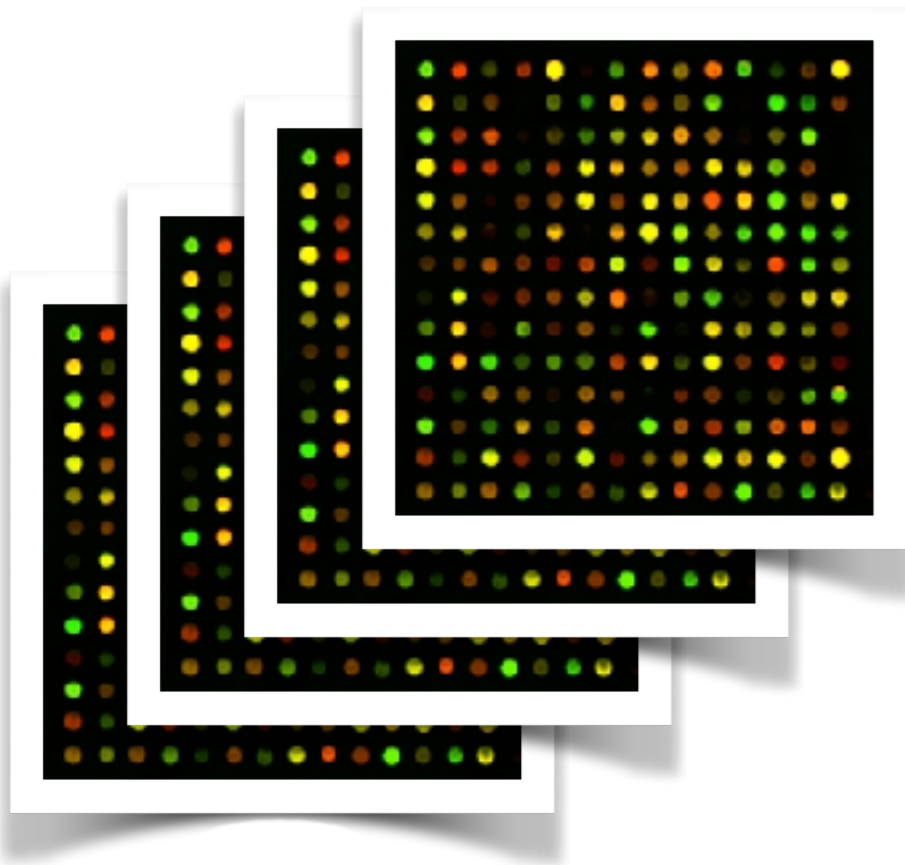
# What can we measure?

---



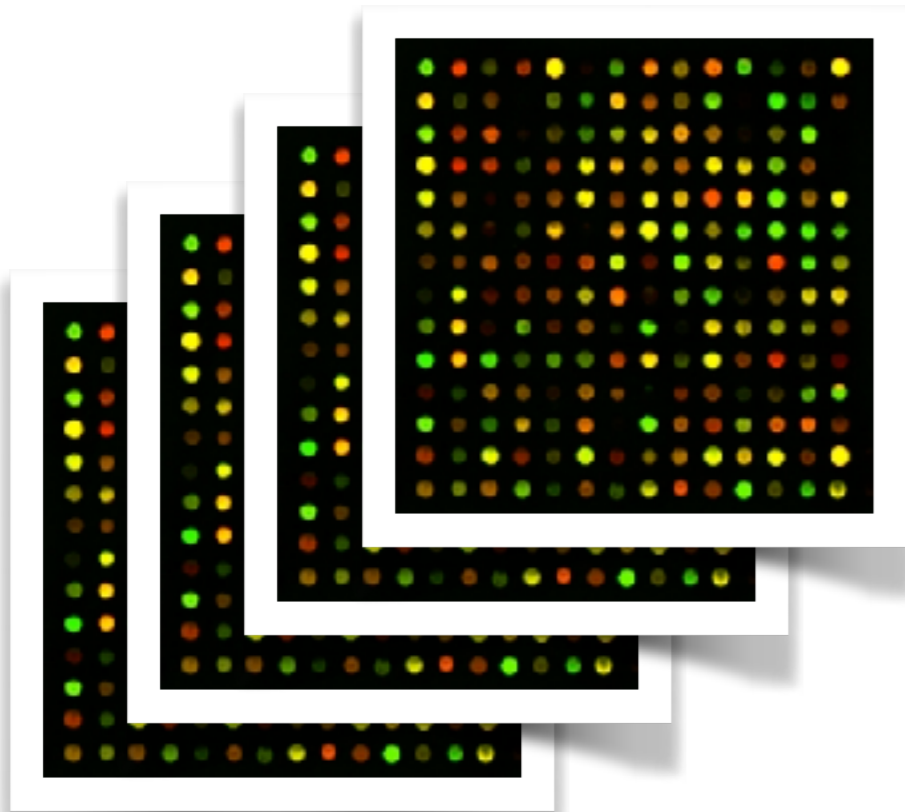
# Measure of gene activity

e.g. microarray experiments, RNA-seq

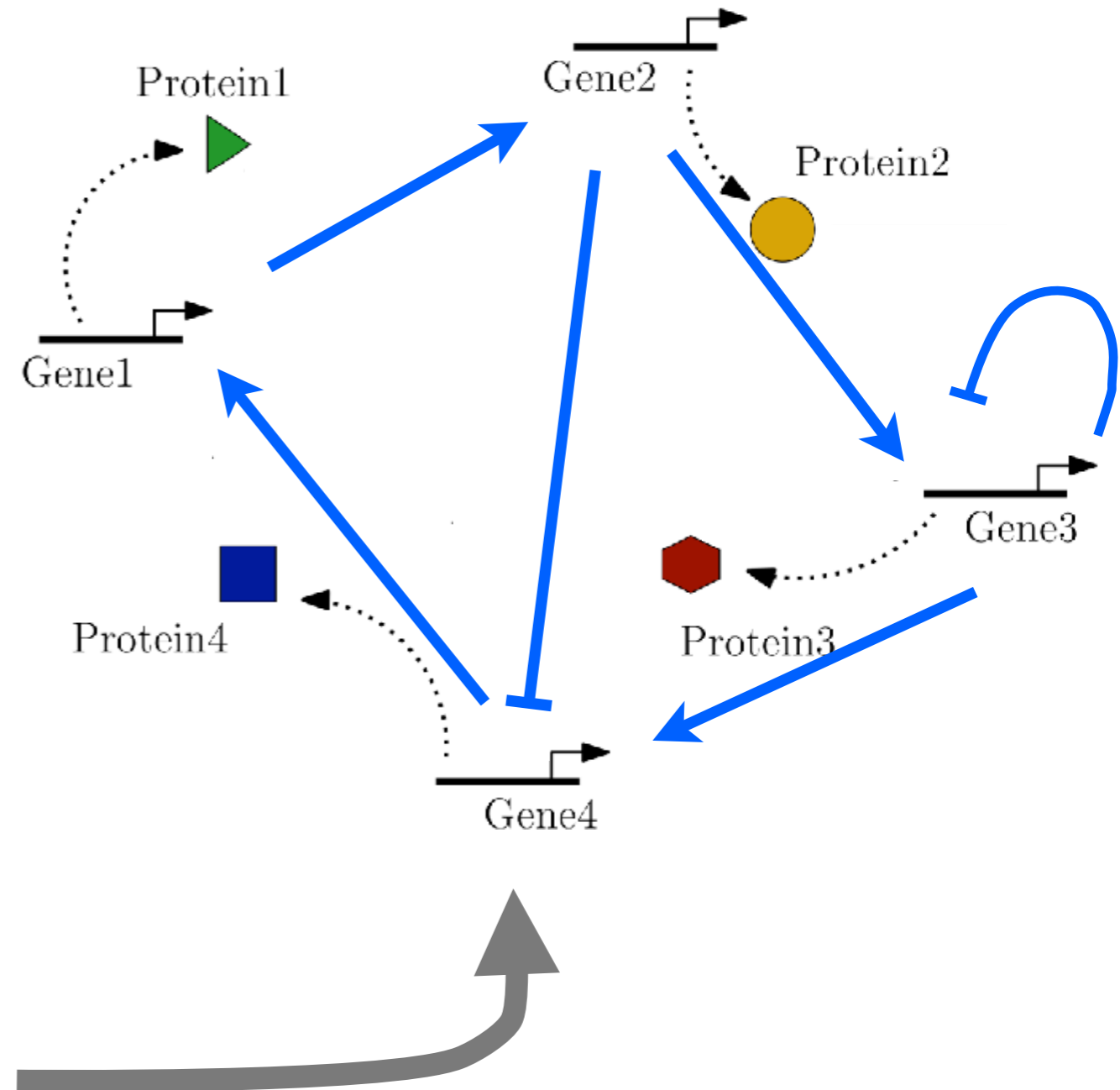


	T1	T2	T3	T4	T5	T6
gene1	20	25	30	32	23	22
gene2	20	23	21	20	22	21
gene3	1	1	2	10	15	13
gene4	0	30	34	37	40	45

# The aim: infer the regulatory relationships



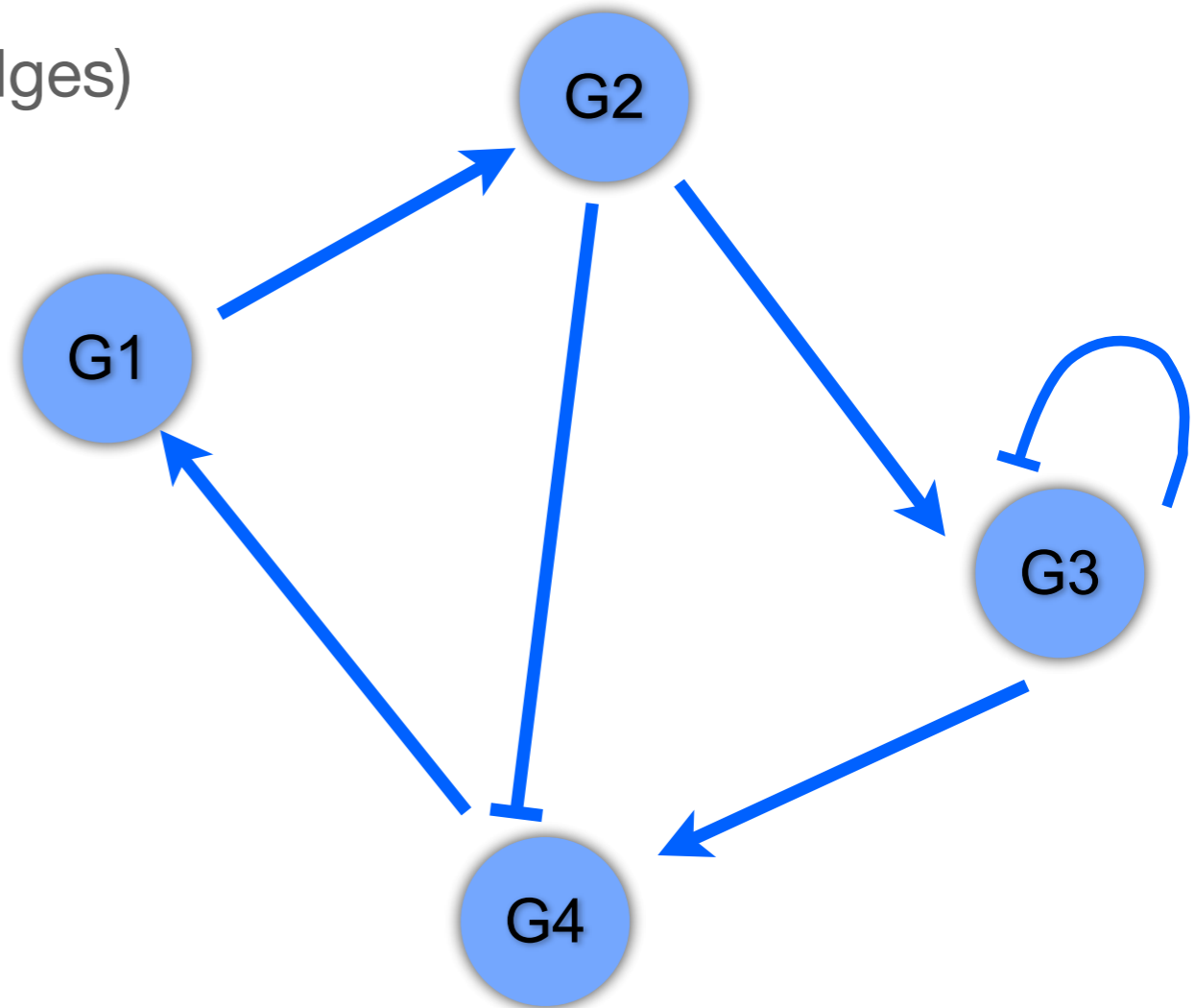
	T1	T2	T3	T4	T5	T6
gene1	20	25	30	32	23	22
gene2	20	23	21	20	22	21
gene3	1	1	2	10	15	13
gene4	0	30	34	37	40	45



# Gene Regulatory Network (GRN)

---

- A gene regulatory network can be represented as a graph  $G = (\text{Vertices}, \text{Edges})$
- Vertices = Genes
  - G1, G2, G3, ...
- Edges = Interactions
  - Activation  $\longrightarrow$
  - Inhibition  $\dashv$



# Dynamic System

---

- A systemic approach is to view the problem as a dynamic system described by a linear system of equations
- Suppose we have a set of genes  $\{G_1, G_2, \dots, G_m\}$  and their expression levels at different time steps  $\{t_1, t_2, \dots, t_n\}$
- The expression level of  $G_j$  at time  $t_i$  is denoted by  $e_j(t_i)$

# Linear dependence hypothesis

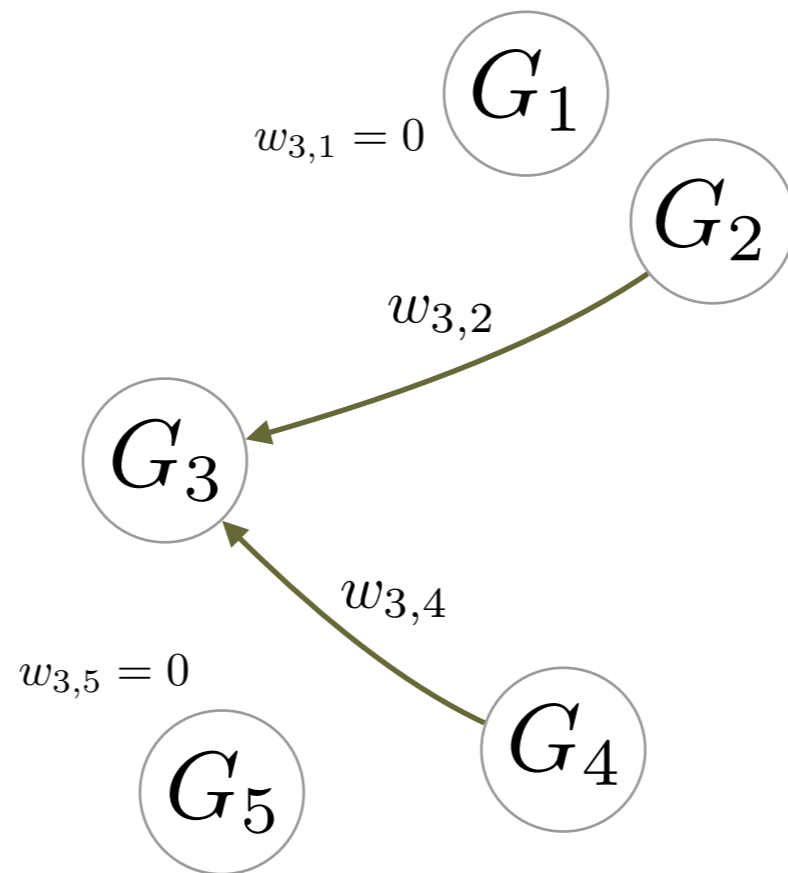
---

- We suppose that the expression level of a gene at time  $t$  depends on a linear combination of expression levels of other genes at the previous time (first order Markov Chain)

$$e_k(t_q) = \sum_{j=1}^m w_{k,j} e_j(t_{q-1}) + \beta_k$$

# Example

---



$$e_3(t_q) = w_{3,2}e_2(t_{q-1}) + w_{3,4}e_4(t_{q-1}) + \beta_3$$



# System of linear equations

**m** genes  
**n** time steps

---

- For each time step **t** [2, 3, 4, ..., **n**] it is possible to define **m** linear equations with **m** unknown variables

# System of linear equations

**m** genes  
**n** time steps

- For each time step **t** [2, 3, 4, ..., **n**] it is possible to define **m** linear equations with **m****x****m** unknown variables

$$\begin{aligned} e_1(t_2) &= w_{1,1}e_1(t_1) + w_{1,2}e_2(t_1) + \cdots + w_{1,m}e_m(t_1) + \beta_1 \\ e_2(t_2) &= w_{2,1}e_1(t_1) + w_{2,2}e_2(t_1) + \cdots + w_{2,m}e_m(t_1) + \beta_2 \\ &\vdots \\ e_m(t_2) &= w_{m,1}e_1(t_1) + w_{m,2}e_2(t_1) + \cdots + w_{m,m}e_m(t_1) + \beta_m \end{aligned}$$

m equations at  
time  $t_2$

# System of linear equations

**m** genes  
**n** time steps

- For each time step **t** [2, 3, 4, ..., **n**] it is possible to define **m** linear equations with **m****x****m** unknown variables

$$\begin{aligned} e_1(t_2) &= w_{1,1}e_1(t_1) + w_{1,2}e_2(t_1) + \cdots + w_{1,m}e_m(t_1) + \beta_1 \\ e_2(t_2) &= w_{2,1}e_1(t_1) + w_{2,2}e_2(t_1) + \cdots + w_{2,m}e_m(t_1) + \beta_2 \\ &\vdots \\ e_m(t_2) &= w_{m,1}e_1(t_1) + w_{m,2}e_2(t_1) + \cdots + w_{m,m}e_m(t_1) + \beta_m \\ e_1(t_3) &= w_{1,1}e_1(t_2) + w_{1,2}e_2(t_2) + \cdots + w_{1,m}e_m(t_2) + \beta_1 \\ e_2(t_3) &= w_{2,1}e_1(t_2) + w_{2,2}e_2(t_2) + \cdots + w_{2,m}e_m(t_2) + \beta_2 \\ &\vdots \\ e_m(t_3) &= w_{m,1}e_1(t_2) + w_{m,2}e_2(t_2) + \cdots + w_{m,m}e_m(t_2) + \beta_m \end{aligned}$$

m equations at  
time  $t_2$

m equations  
at time  $t_3$

# System of linear equations

**m** genes  
**n** time steps

- For each time step **t** [2, 3, 4, ..., **n**] it is possible to define **m** linear equations with **m****x****m** unknown variables

$$\begin{aligned} e_1(t_2) &= w_{1,1}e_1(t_1) + w_{1,2}e_2(t_1) + \cdots + w_{1,m}e_m(t_1) + \beta_1 \\ e_2(t_2) &= w_{2,1}e_1(t_1) + w_{2,2}e_2(t_1) + \cdots + w_{2,m}e_m(t_1) + \beta_2 \\ &\vdots \\ e_m(t_2) &= w_{m,1}e_1(t_1) + w_{m,2}e_2(t_1) + \cdots + w_{m,m}e_m(t_1) + \beta_m \end{aligned}$$

m equations at  
time  $t_2$

$$\begin{aligned} e_1(t_3) &= w_{1,1}e_1(t_2) + w_{1,2}e_2(t_2) + \cdots + w_{1,m}e_m(t_2) + \beta_1 \\ e_2(t_3) &= w_{2,1}e_1(t_2) + w_{2,2}e_2(t_2) + \cdots + w_{2,m}e_m(t_2) + \beta_2 \\ &\vdots \\ e_m(t_3) &= w_{m,1}e_1(t_2) + w_{m,2}e_2(t_2) + \cdots + w_{m,m}e_m(t_2) + \beta_m \end{aligned}$$

m equations  
at time  $t_3$

$$\begin{aligned} &\vdots \\ e_1(t_n) &= w_{1,1}e_1(t_{n-1}) + w_{1,2}e_2(t_{n-1}) + \cdots + w_{1,m}e_m(t_{n-1}) + \beta_1 \\ e_2(t_n) &= w_{2,1}e_1(t_{n-1}) + w_{2,2}e_2(t_{n-1}) + \cdots + w_{2,m}e_m(t_{n-1}) + \beta_2 \\ &\vdots \\ e_m(t_n) &= w_{m,1}e_1(t_{n-1}) + w_{m,2}e_2(t_{n-1}) + \cdots + w_{m,m}e_m(t_{n-1}) + \beta_m \end{aligned}$$

m equations  
at time  $t_n$

# System of linear equations

**m** genes  
**n** time steps

- For each time step **t** [2, 3, 4, ..., **n**] it is possible to define **m** linear equations with **m****x****m** unknown variables

In summary we have **m****x****m** unknown vars and **m****x**(**n**-1) independent linear equations

$$\begin{aligned} e_1(t_2) &= w_{1,1}e_1(t_1) + w_{1,2}e_2(t_1) + \cdots + w_{1,m}e_m(t_1) + \beta_1 \\ e_2(t_2) &= w_{2,1}e_1(t_1) + w_{2,2}e_2(t_1) + \cdots + w_{2,m}e_m(t_1) + \beta_2 \\ &\vdots \\ e_m(t_2) &= w_{m,1}e_1(t_1) + w_{m,2}e_2(t_1) + \cdots + w_{m,m}e_m(t_1) + \beta_m \end{aligned}$$

m equations at  
time  $t_2$

$$\begin{aligned} e_1(t_3) &= w_{1,1}e_1(t_2) + w_{1,2}e_2(t_2) + \cdots + w_{1,m}e_m(t_2) + \beta_1 \\ e_2(t_3) &= w_{2,1}e_1(t_2) + w_{2,2}e_2(t_2) + \cdots + w_{2,m}e_m(t_2) + \beta_2 \\ &\vdots \\ e_m(t_3) &= w_{m,1}e_1(t_2) + w_{m,2}e_2(t_2) + \cdots + w_{m,m}e_m(t_2) + \beta_m \end{aligned}$$

m equations  
at time  $t_3$

$$\begin{aligned} e_1(t_n) &= w_{1,1}e_1(t_{n-1}) + w_{1,2}e_2(t_{n-1}) + \cdots + w_{1,m}e_m(t_{n-1}) + \beta_1 \\ e_2(t_n) &= w_{2,1}e_1(t_{n-1}) + w_{2,2}e_2(t_{n-1}) + \cdots + w_{2,m}e_m(t_{n-1}) + \beta_2 \\ &\vdots \\ e_m(t_n) &= w_{m,1}e_1(t_{n-1}) + w_{m,2}e_2(t_{n-1}) + \cdots + w_{m,m}e_m(t_{n-1}) + \beta_m \end{aligned}$$

m equations  
at time  $t_n$

# Solution of the system

**m** genes  
**n** time steps

---

- Form linear algebra to resolve a system of **m** unknown variables we need at least **m = m(n-1)** independent linear equations
  - **m = (n-1)** unique solution or no solution
  - **m > (n-1)** undetermined system (infinite solutions)
  - **m < (n-1)** overdetermined system (it could not have solutions or they could be estimated with linear regression models).

# Solution of the system

**m** genes  
**n** time steps

---

- Form linear algebra to resolve a system of **m****x****m** unknown variables we need at least **m****x****m** = **m(n-1)** independent linear equations
  - **m** = **(n-1)** unique solution or no solution
  - **m** > **(n-1)** undetermined system (infinite solutions)
  - **m** < **(n-1)** overdetermined system (it could not have solutions or they could be estimated with linear regression models).

In biological contexts **m** >> **(n-1)**

number of genes **m** ~ 20000

number of samples **n** ~ 100

# The Big issue

---

- From a computational point of view the reconstruction of gene regulatory networks is an **undetermined problem** as the large number of possible solutions is typically high in contrast to the number of available independent data points!

Many approaches based on Heuristics

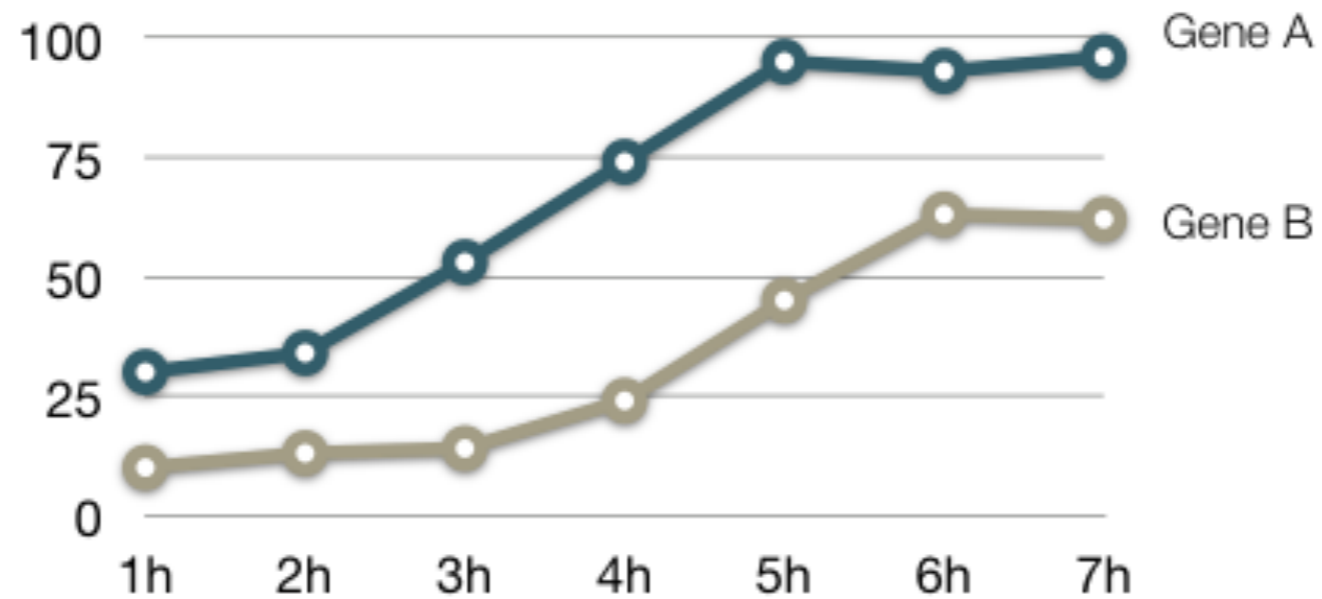
- Clustering
- Correlation methods (Pearson, Mutual Information,...)
- Boolean Networks
- Bayesian Networks
- ...



# A very simple example

---

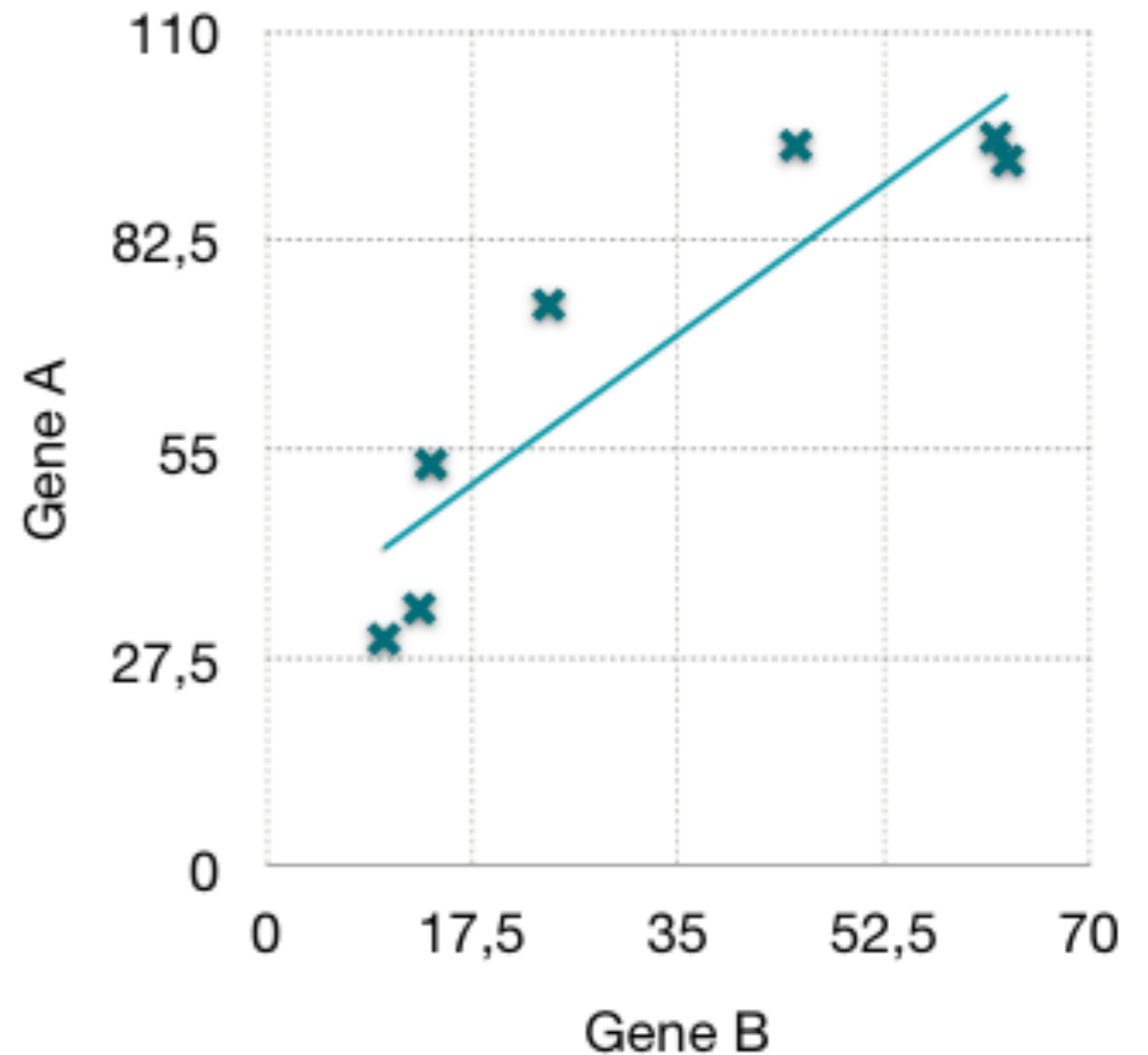
- A researcher would like to study the interaction between two genes A e B in mouse in the first embryonic developmental stages
- Measures the expression level of such genes from the first to the seventh hour



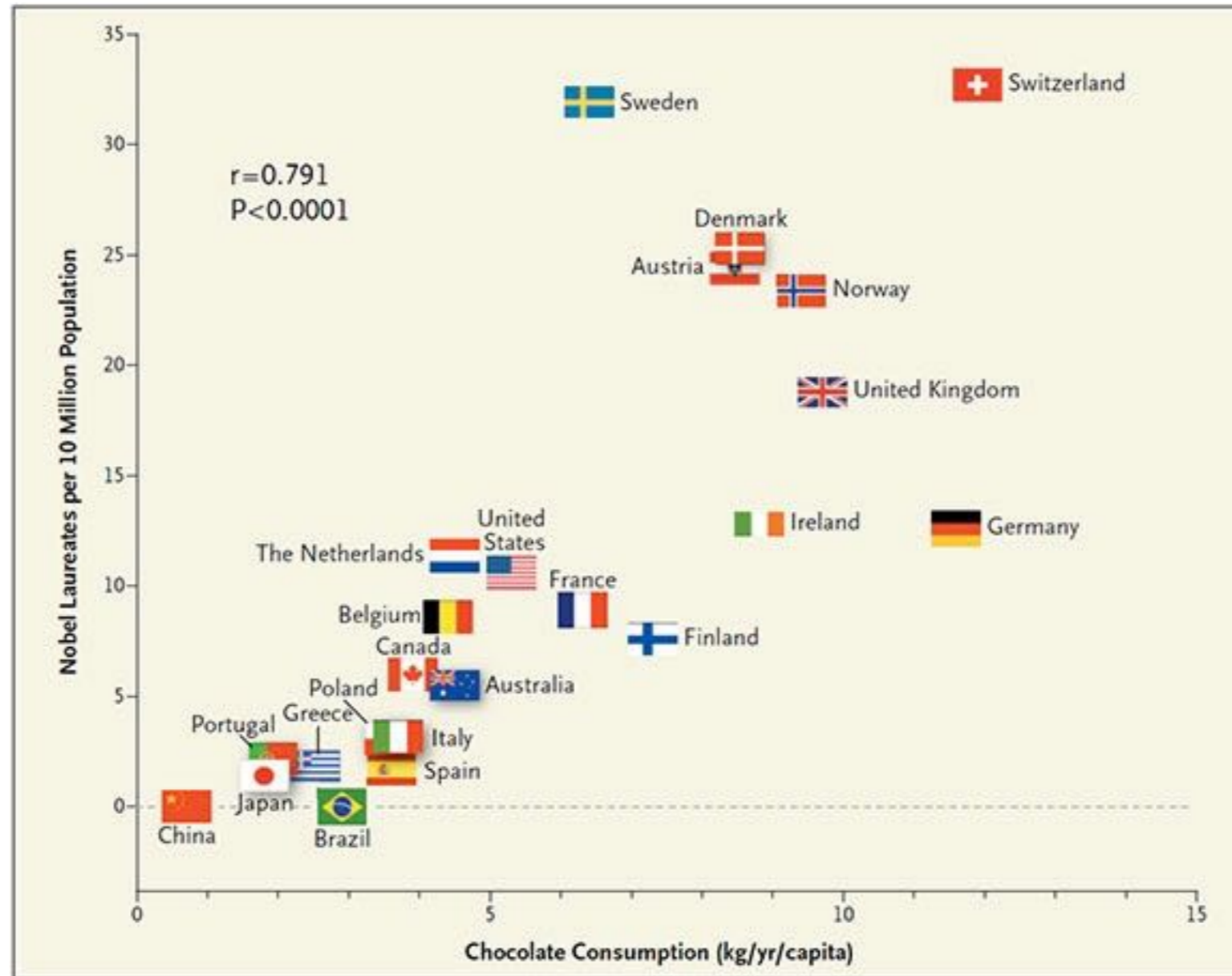
# Statistical Correlation (e.g. Pearson)

---

- It is the most widely used mathematical tool by biologists
- It assumes a linear relationship
- The number of points should be at least 3



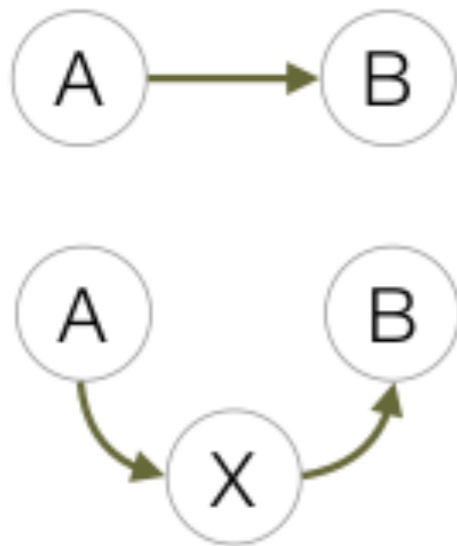
# Correlation does not imply cause-effect relationships



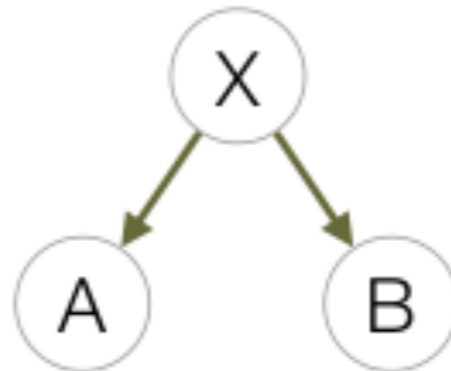
# Three possible cases

---

A regulates B or vice versa. Such a relationship could be direct or indirect (i.e. mediated by other genes not measured)



A and B are co-regulated by another gene X

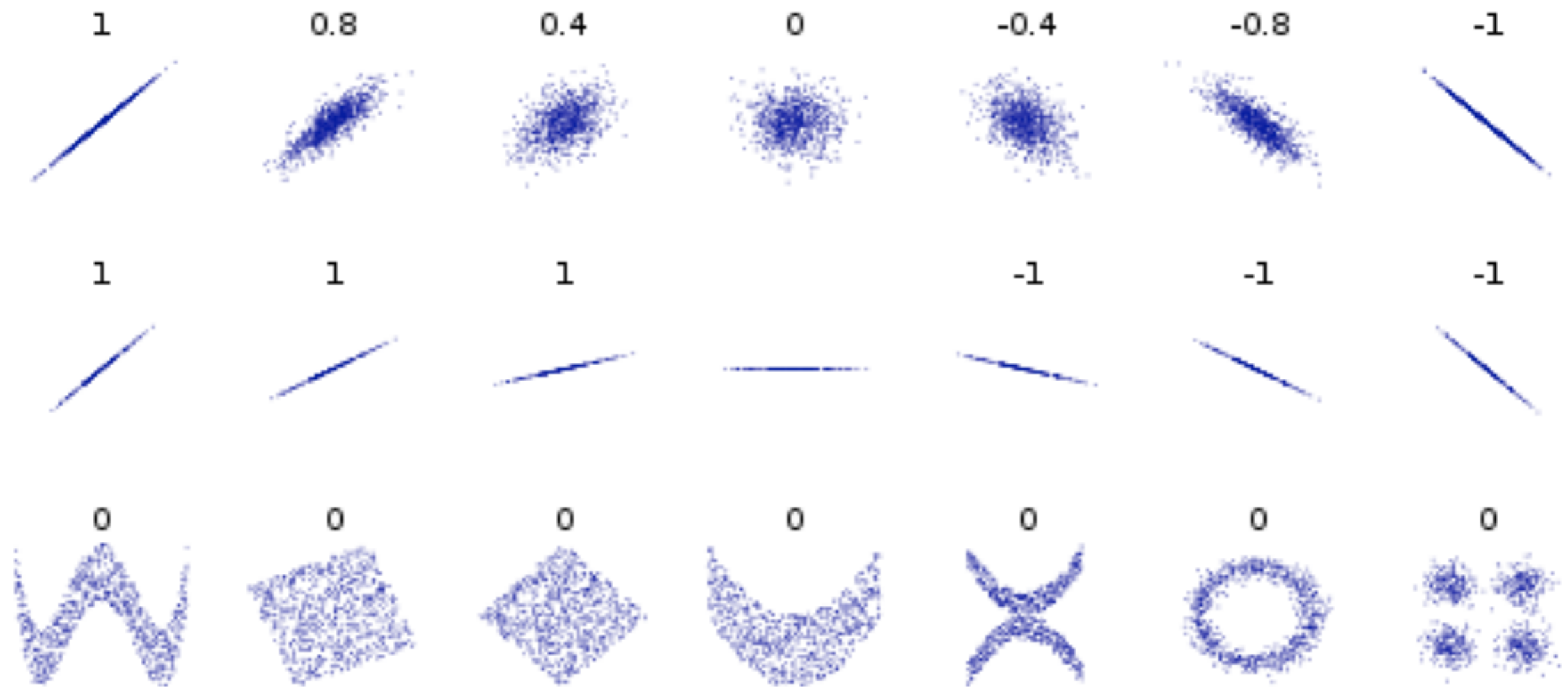


A and B do not interact and do not share any regulation mechanism

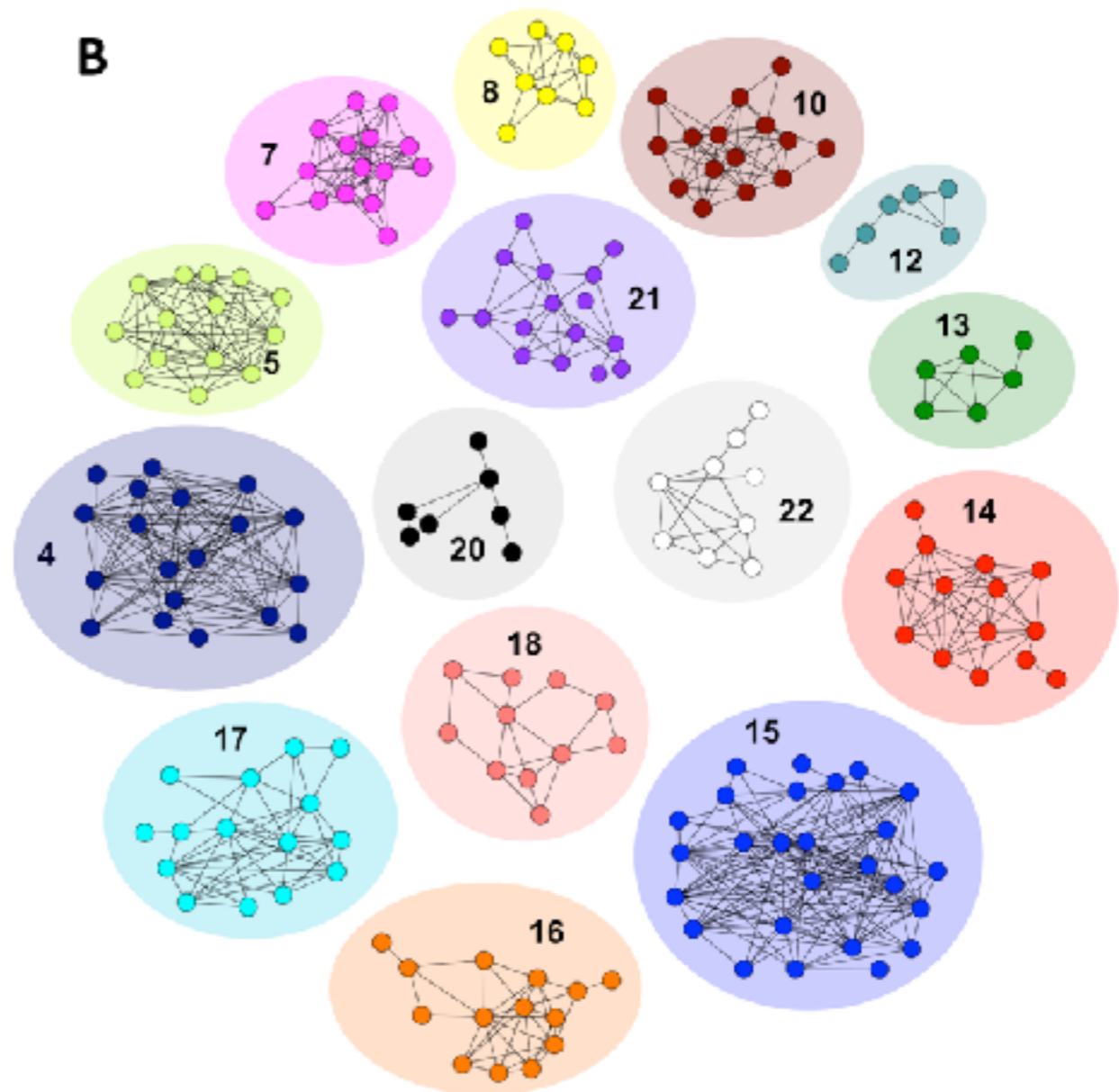
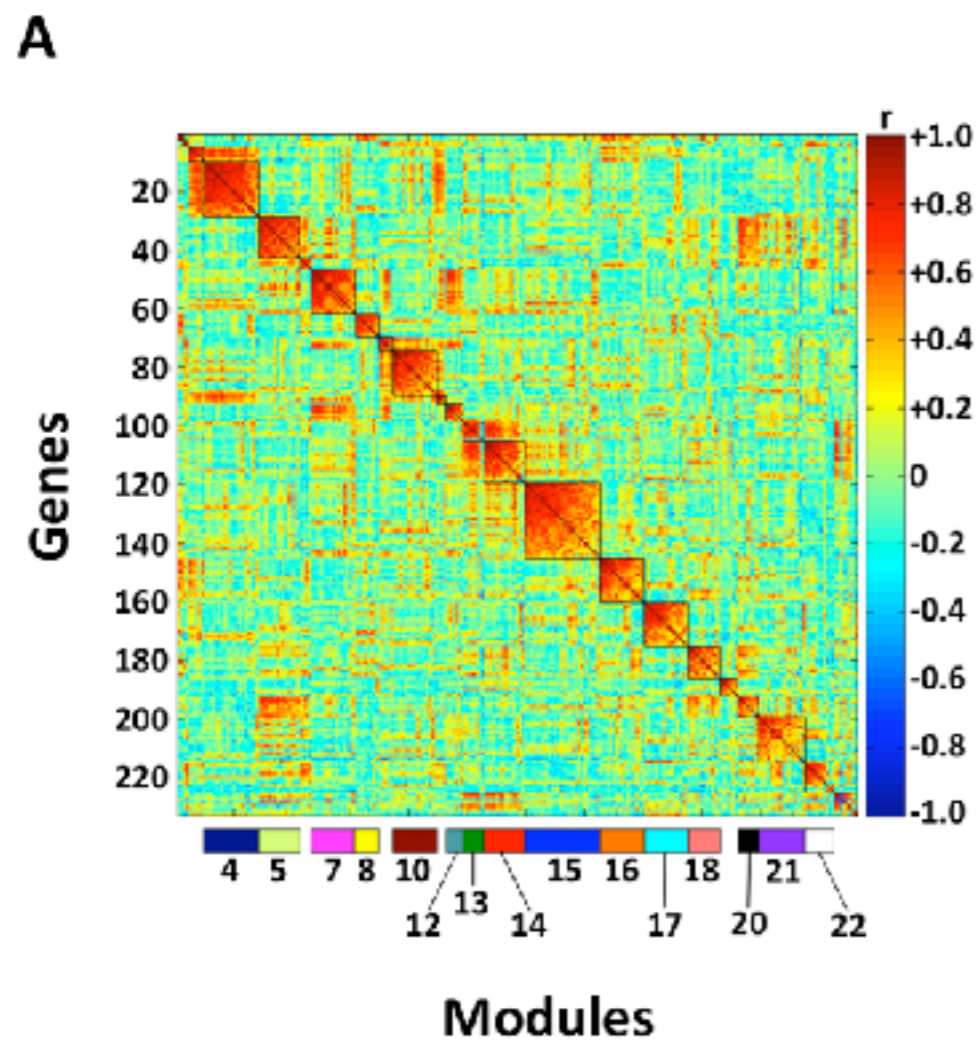


# Pearson correlation could fail with non-linear relationships

---



# Clustering



# Reverse Engineering of Gene Regulatory Networks

## Approaches

---

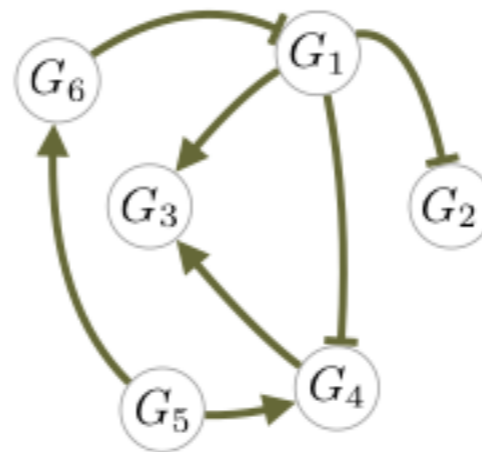
- Un-supervised

- Supervised

# Reverse Engineering of Gene Regulatory Networks Approaches

---

- Un-supervised



- Supervised



# Reverse Engineering of Gene Regulatory Networks

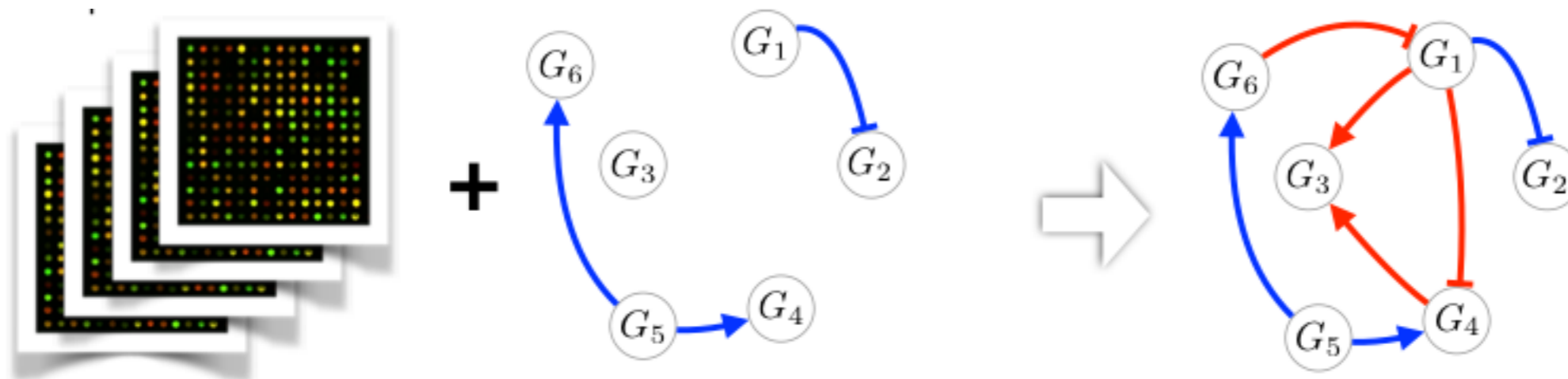
## Approaches

---

- Un-supervised



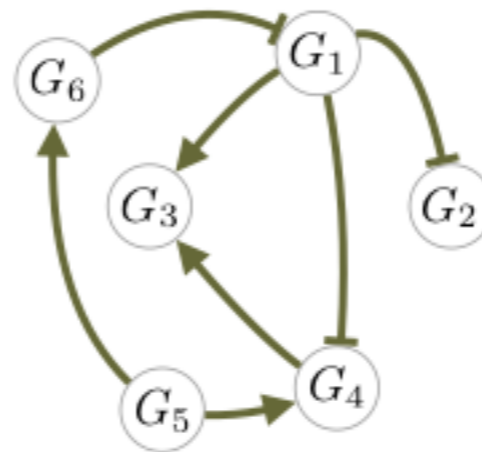
- Supervised



# Reverse Engineering of Gene Regulatory Networks

## Approaches

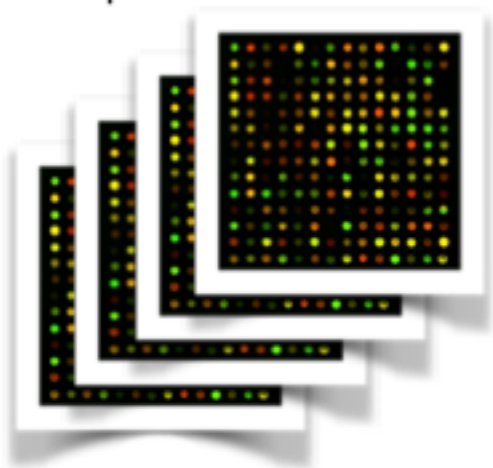
- Un-supervised



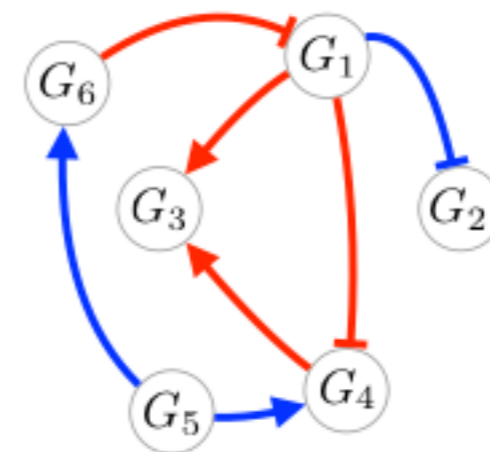
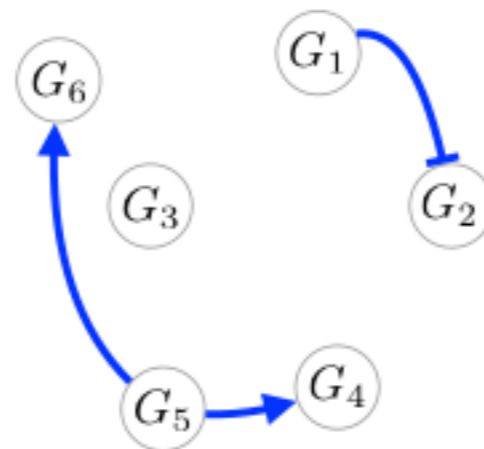
ARACNE  
CLR  
GENIE3  
TD-Aracne  
FormalM  
TNI  
BANJO

...

- Supervised



+



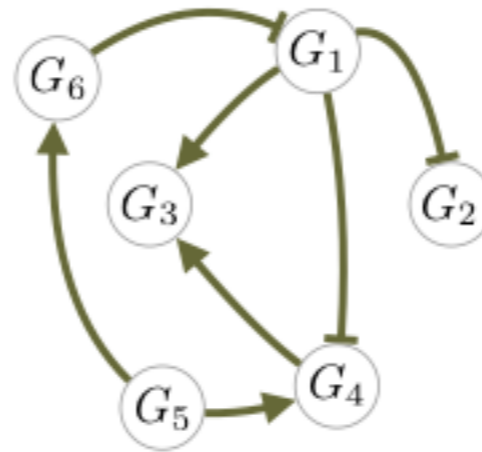
SIRENE  
POS-ONLY

...

# Reverse Engineering of Gene Regulatory Networks

## Approaches

- Un-supervised

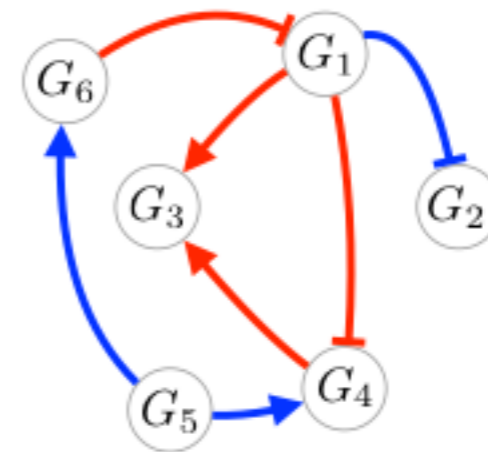
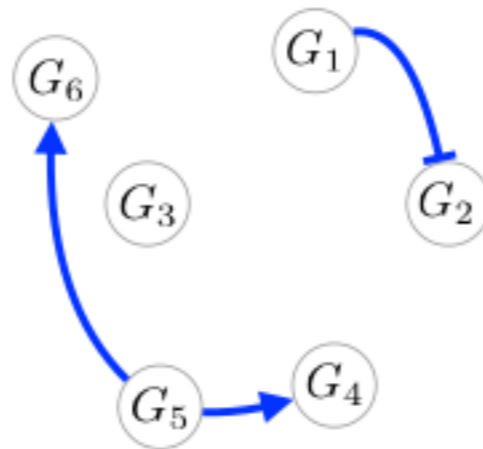


ARACNE  
CLR  
GENIE3  
TD-Aracne  
FormalM  
TNI  
BANJO  
...

- Supervised



+



SIRENE  
POS-ONLY  
...

# ARACNE



*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<sup>1</sup>, Adam A Margolin<sup>2</sup>, Gustavo Stolovitzky<sup>3</sup>, Ulf Klein<sup>1</sup>, Riccardo Dalla-Favera<sup>1,4</sup> & Andrea Califano<sup>2</sup>

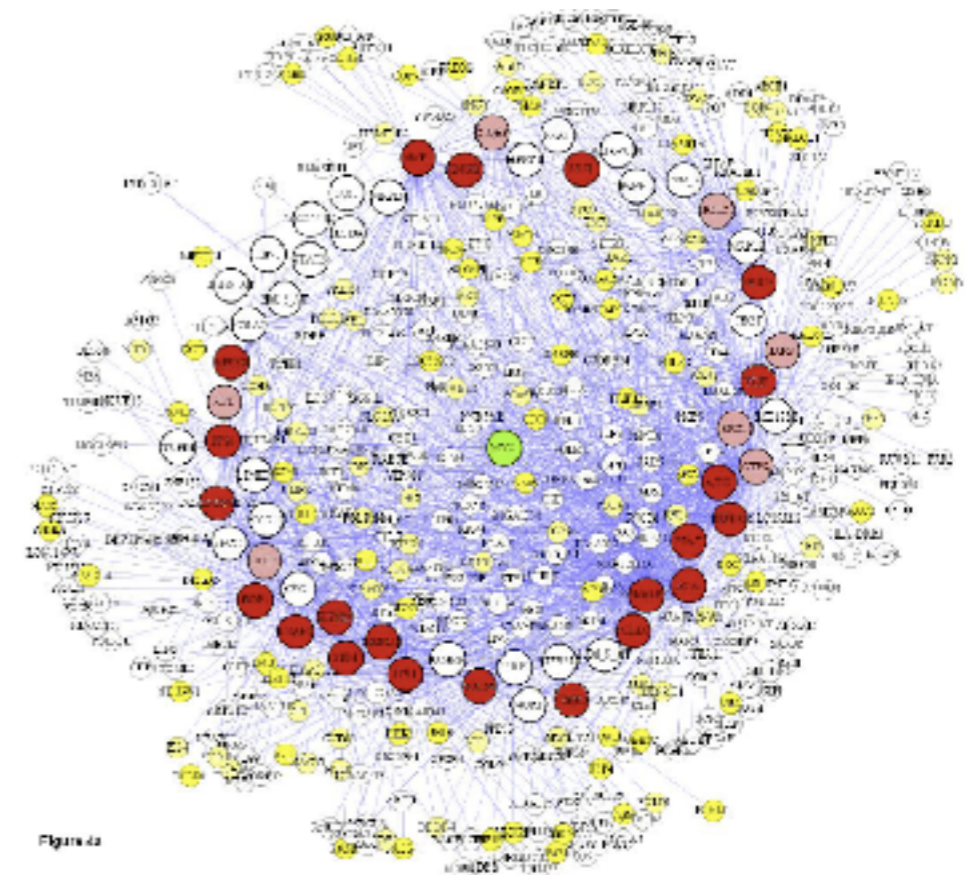


Figure 4a

# Novel Myc targets validated with ChIP-chip

(accuracy >90%)

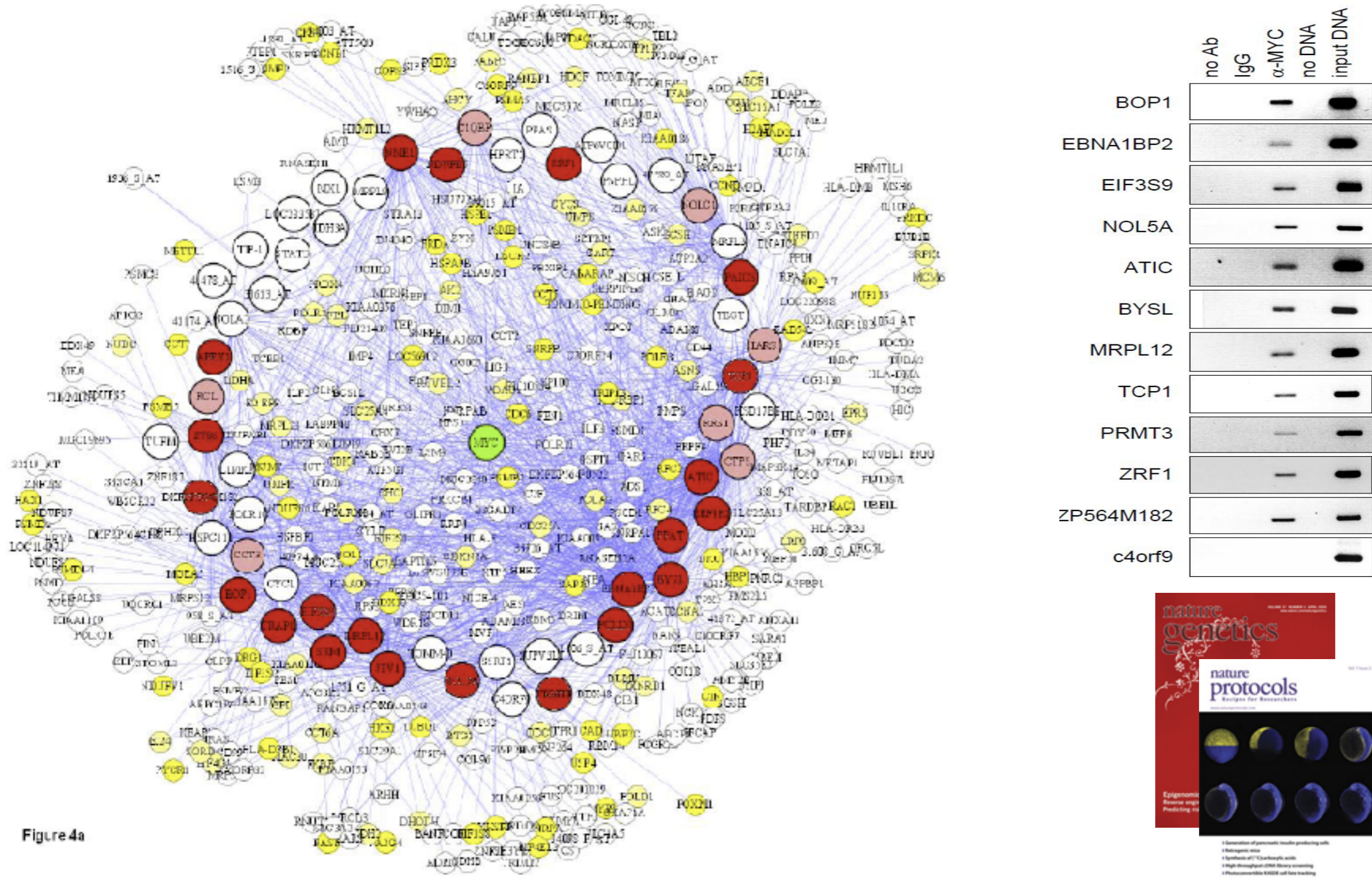


Figure 4a



# Information Entropy H

---



Mary

blah, blah, blah, blah, blah, blah, blah,  
blah, blah, blah, blah, blah, blah, blah,  
blah, ...

$H[Mary]$

>0 bits

↑ Uncertainty

0 Certainty

# Information Entropy

---



Mary

+ + - - - + - + - - +

$p(+)$ , probability of emitting +  
 $p(-)$ , probability of emitting -

$$\text{Entropy}(S) \equiv -p_{\oplus} \log_2 p_{\oplus} - p_{\ominus} \log_2 p_{\ominus}$$

# Information Entropy

---

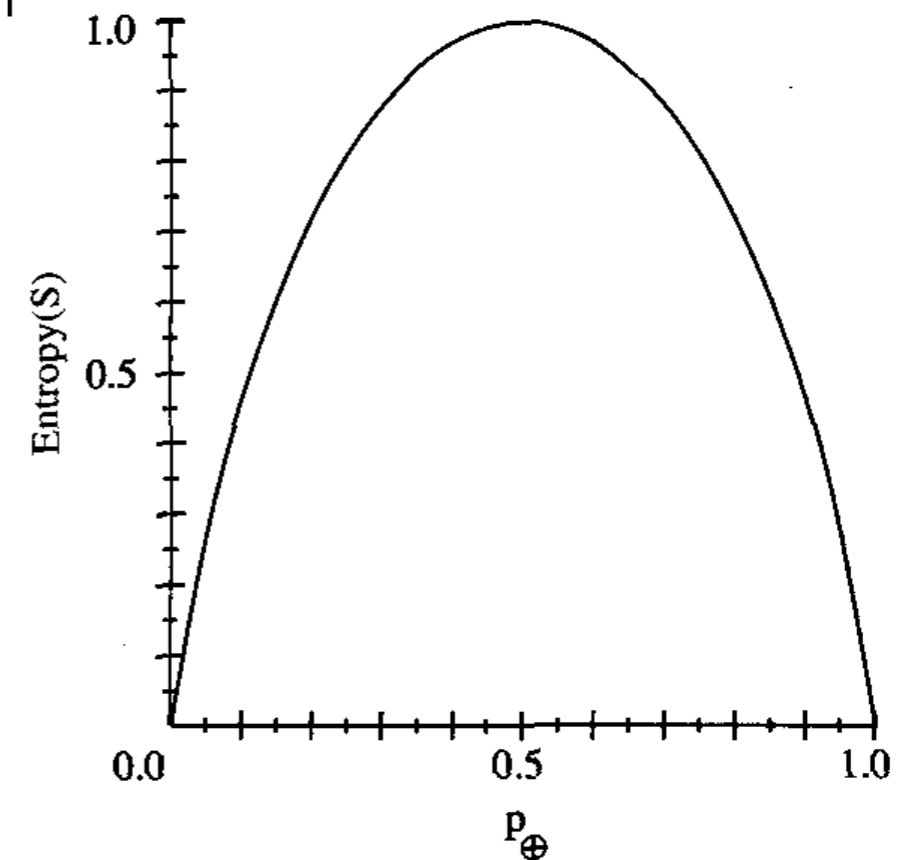


Mary

+ + - - - + - + - - +

$p(+)$ , probability of emitting +  
 $p(-)$ , probability of emitting -

$$\text{Entropy}(S) \equiv -p_{\oplus} \log_2 p_{\oplus} - p_{\ominus} \log_2 p_{\ominus}$$





# Information Entropy

---



Mary

! & ! % \$ & "\$ !£/ !"£ /!"£2  
78 3786" !( £/ &!" £

$p_i$ , probability of the symbol  $i$

$$\text{Entropy}(S) \equiv \sum_{i=1}^c -p_i \log_2 p_i$$

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

# Mutual Information

---



Mary

blah, blah, blah, ...  $H[Mary]$



Joe

blah, blah, blah, ...  $H[Joe]$

# Mutual Information

---



Mary

blah, blah, blah, ...  $H[Mary]$



Joe

blah, blah, blah, ...  $H[Joe]$

# Mutual Information

---



Mary

blah, blah, blah, ...  $H[Mary]$

$H[Joe|Mary]$



Joe

blah, blah, blah, ...  $H[Joe]$

# Mutual Information

---



Mary

blah, blah, blah, ...  $H[Mary]$

$$I[Joe; Mary] = H[Joe] - H[Joe|Mary]$$



Joe

blah, blah, blah, ...  $H[Joe]$

# Mutual Information

---



Mary

blah, blah, blah, ...  $H[Mary]$

$$\begin{aligned} I[Joe; Mary] &= H[Joe] - H[Joe|Mary] \\ &= H[Mary] - H[Mary|Joe] \\ &= I[Mary; Joe] \end{aligned}$$



Joe

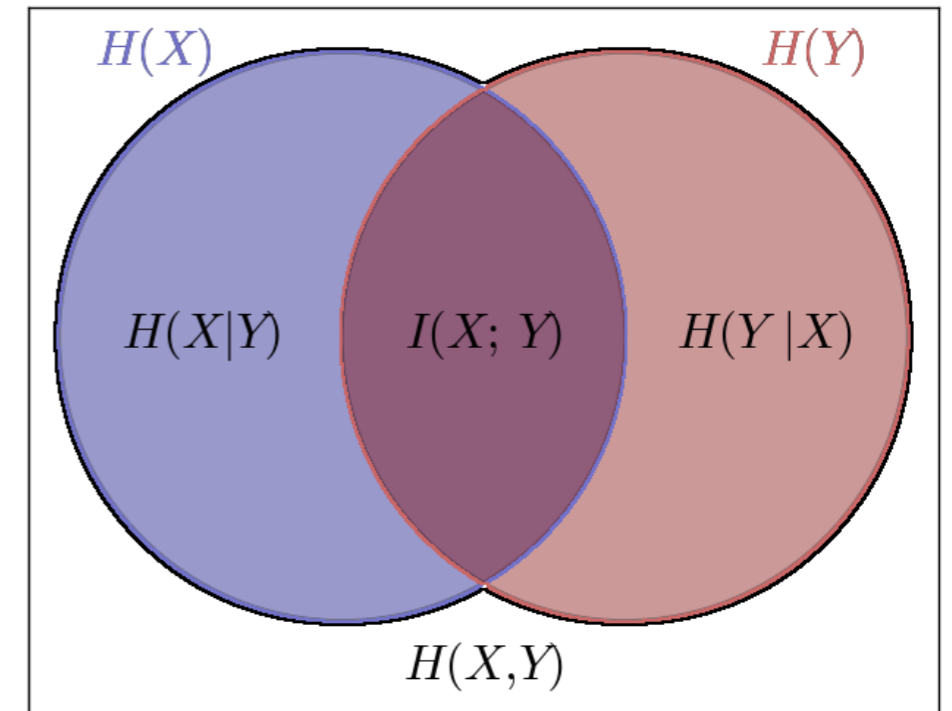
blah, blah, blah, ...  $H[Joe]$

# Mutual Information

---

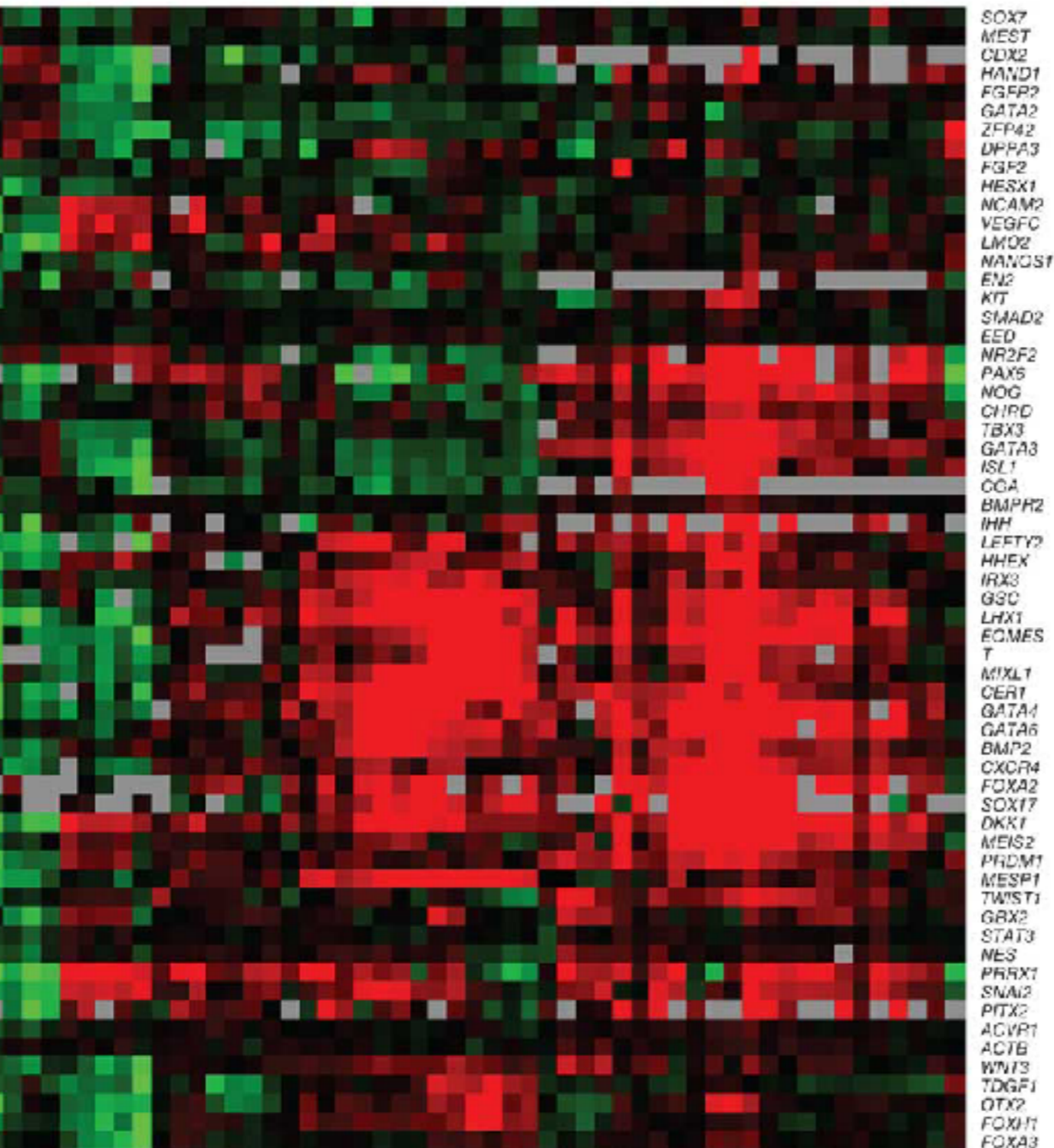
- Measures the information that X and Y share.
- $I(X;Y) = H(X) + H(Y) - H(X,Y)$
- $I(X;Y) = 0$  if X and Y are independent

$$I(X;Y) = \sum_{y \in Y} \sum_{x \in X} p(x,y) \log \left( \frac{p(x,y)}{p_1(x) p_2(y)} \right)$$



# ARACNE approach

---



- A gene  $X$  is modeled as a continuous random variable with a p.d.f  $f(x)$
- The mutual information between two continuous random variable  $X$  and  $Y$  with a join p.d.f  $f(x,y)$  is given by:

$$I(X;Y) = \int \int f(x,y) \log \frac{f(x,y)}{f(x)f(y)} dx dy.$$

- We need a way to estimate the p.d.f from sample data.



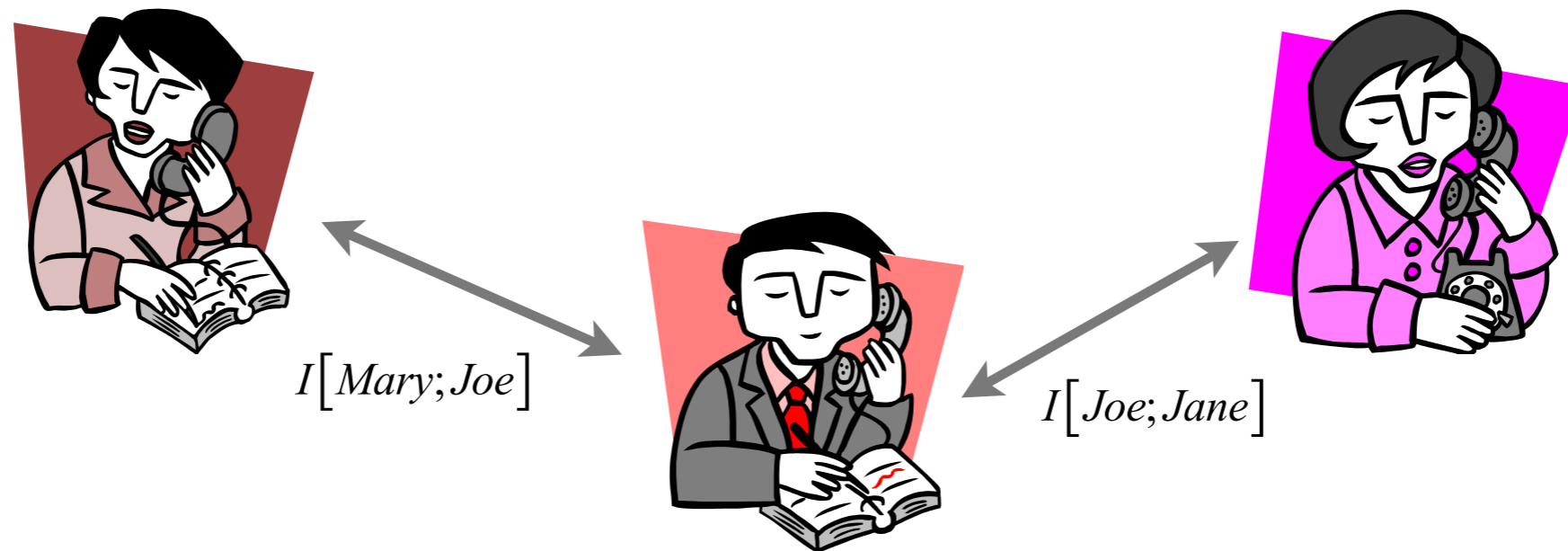
# p.d.f estimators

---

- k-Nearest Neighbor Estimator;
- Kernel Density Estimator.
- Maximum Likelihood Estimator;
- Miller-Madow Corrected Estimator;
- Bayesian estimators:
  - Krichevsky and Tromov Estimator;
  - Holste Estimator;
  - Schürmann and Grassberger Estimator;
  - Minimax Estimator;

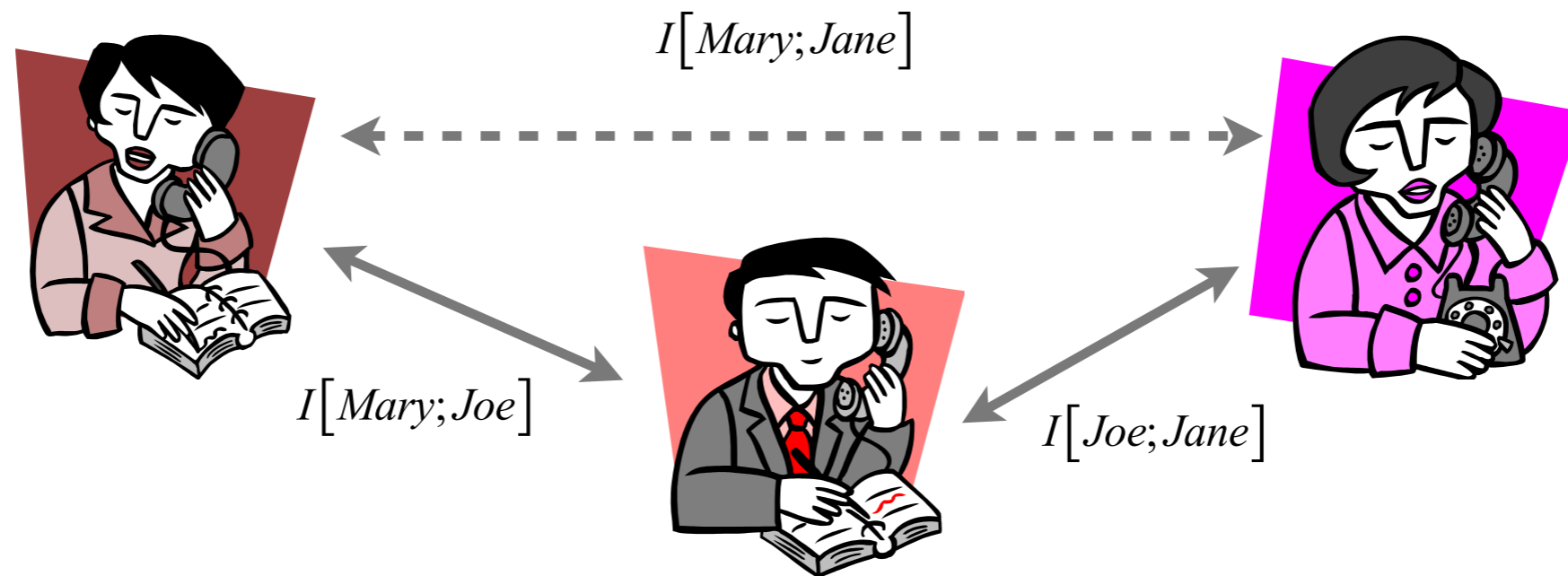
# The indirect relationship problem of information theoretic approaches

---



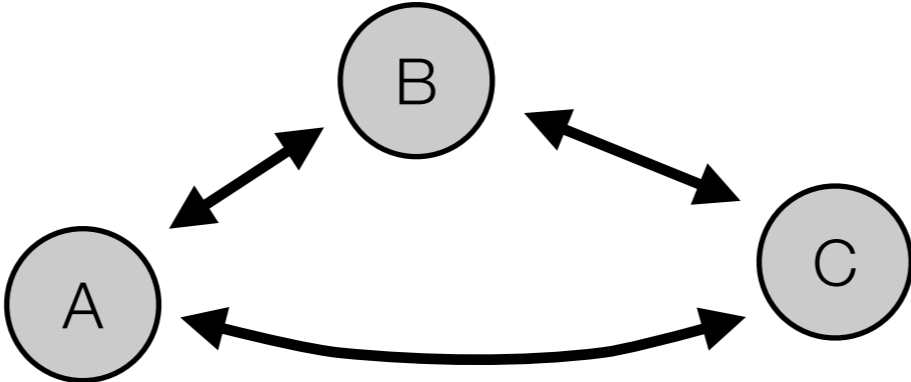
# The indirect relationship problem of information theoretic approaches

---



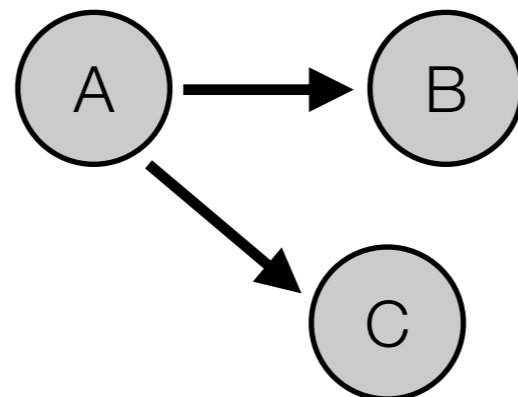
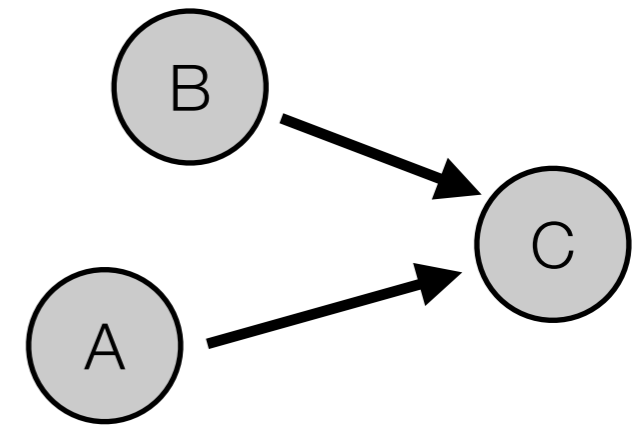
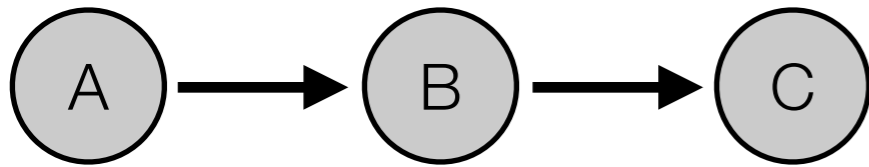
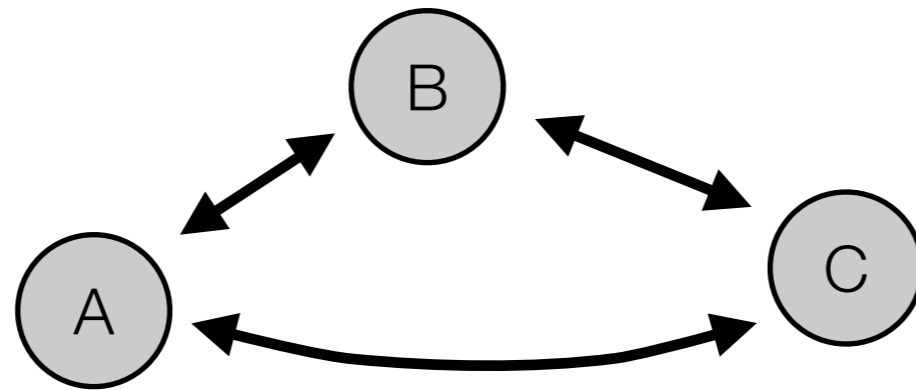
Which is the correct network?

---



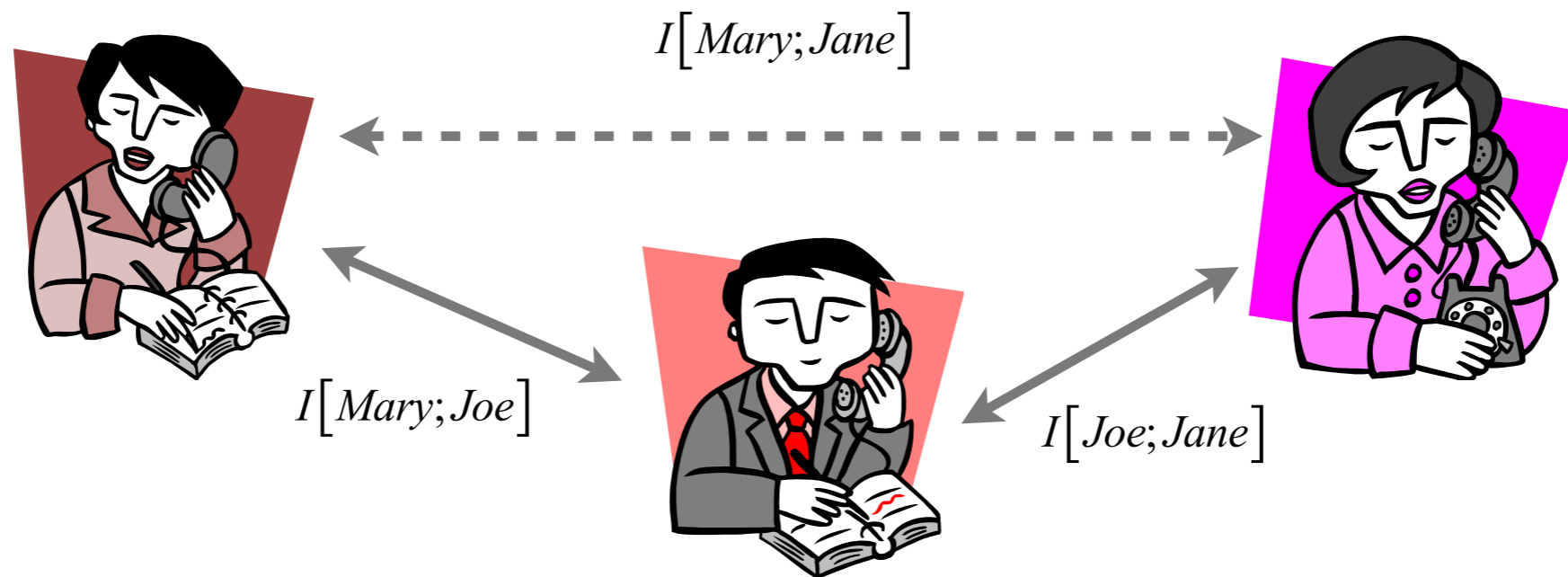
# Which is the correct network?

---



# Information theory

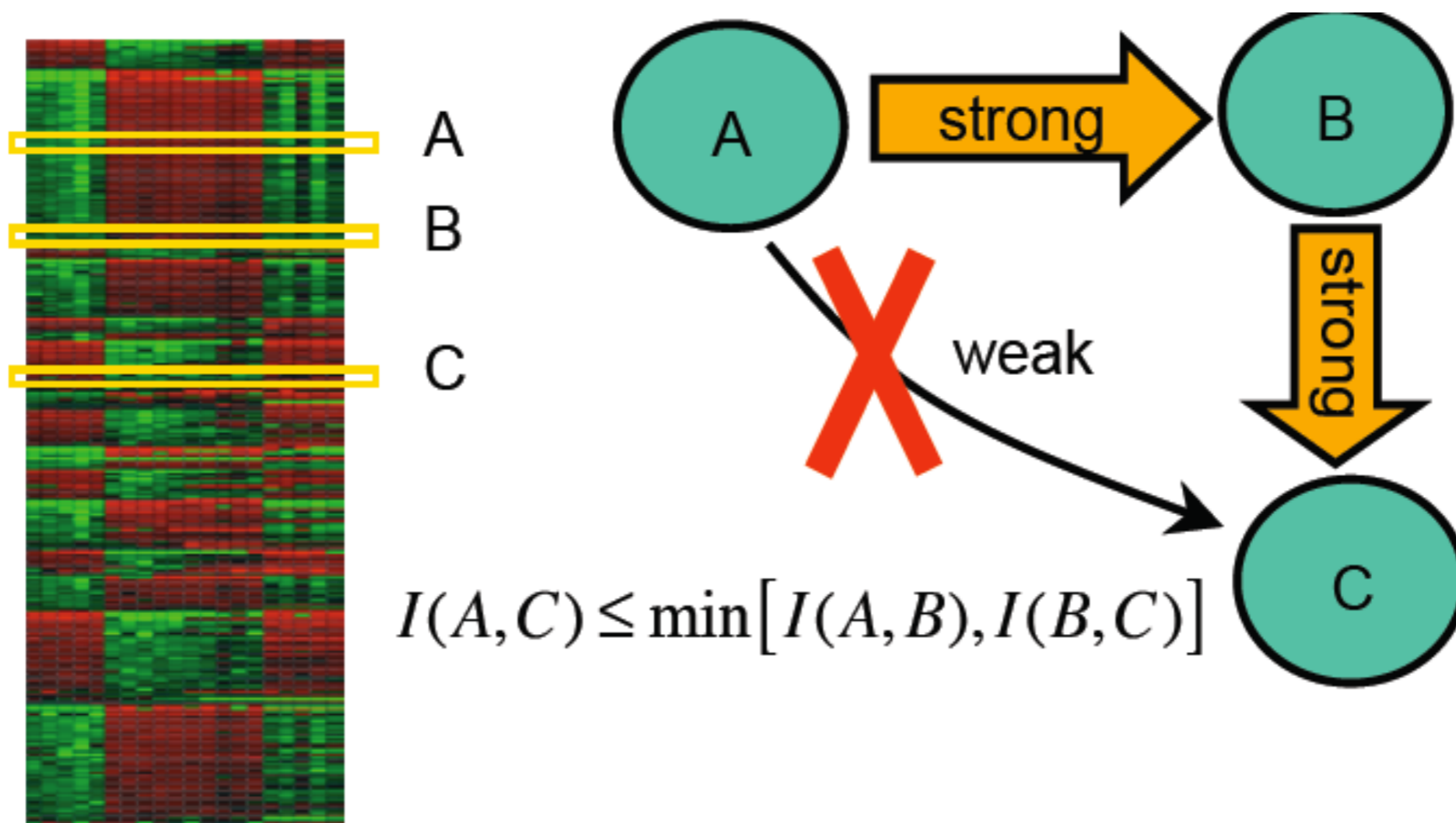
---



$$I[Mary; Jane] \leq \min\{I[Mary; Joe], I[Joe; Jane]\}$$

# ARACNE DPI approach

---



# TD-ARACNE

Zuppoli et al. BMC Bioinformatics 2010, 11:151  
http://www.biomedcentral.com/1471-2105/11/151

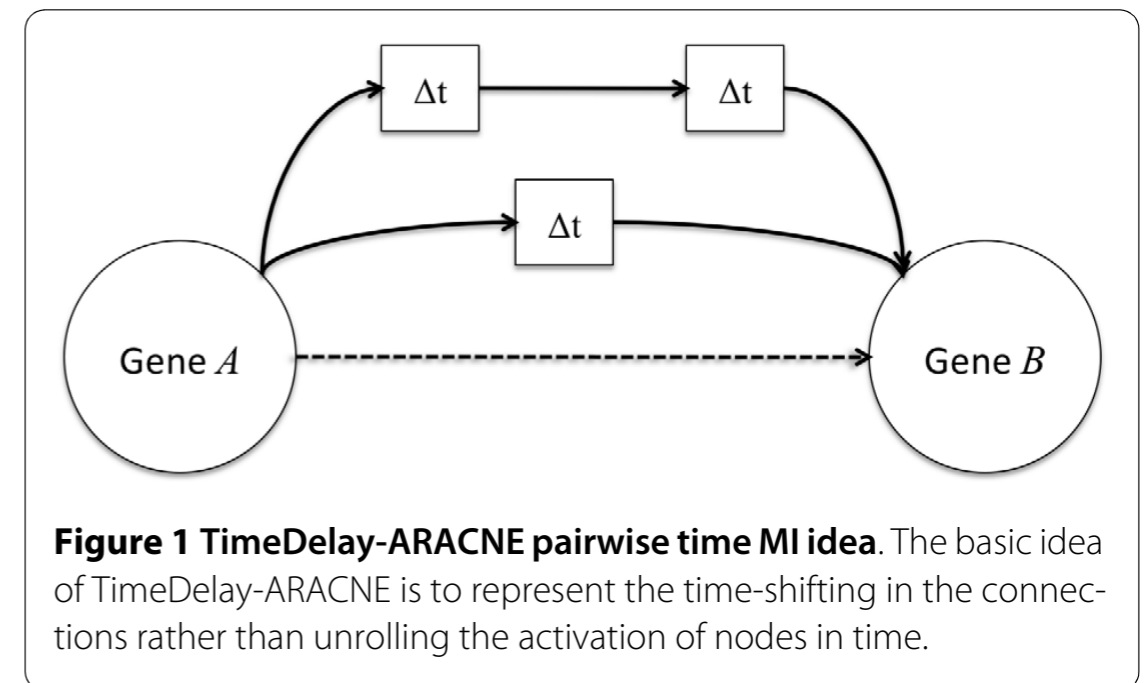
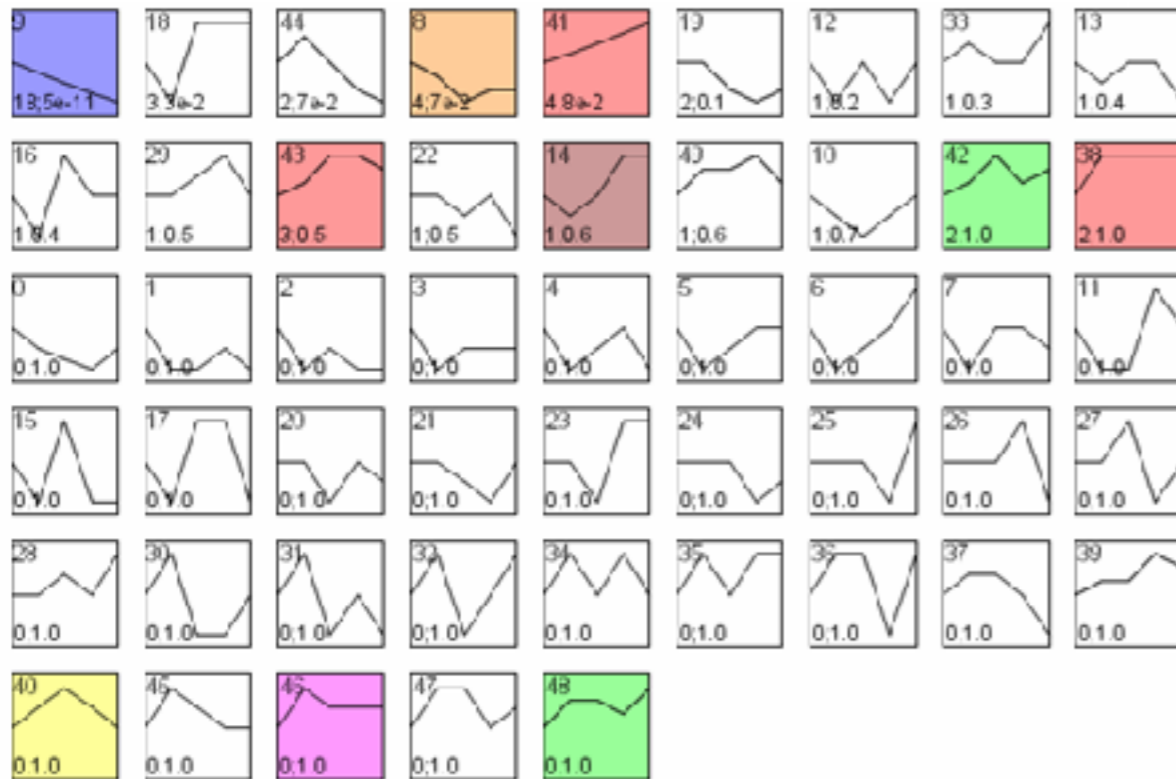


METHODOLOGY ARTICLE

Open Access

## TimeDelay-ARACNE: Reverse engineering of gene networks from time-course data by an information theoretic approach

Pietro Zuppoli<sup>1,2</sup>, Sandro Morganello<sup>1,2</sup> and Michele Ceccarelli<sup>\*1,2</sup>



**Figure 1 TimeDelay-ARACNE pairwise time MI idea.** The basic idea of TimeDelay-ARACNE is to represent the time-shifting in the connections rather than unrolling the activation of nodes in time.

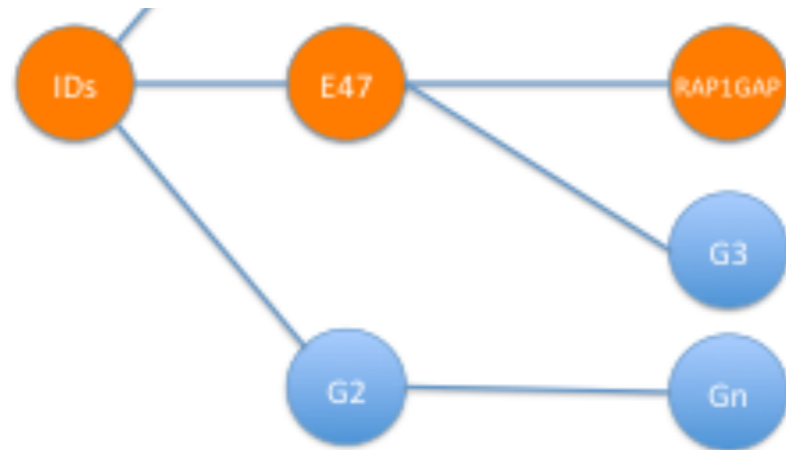
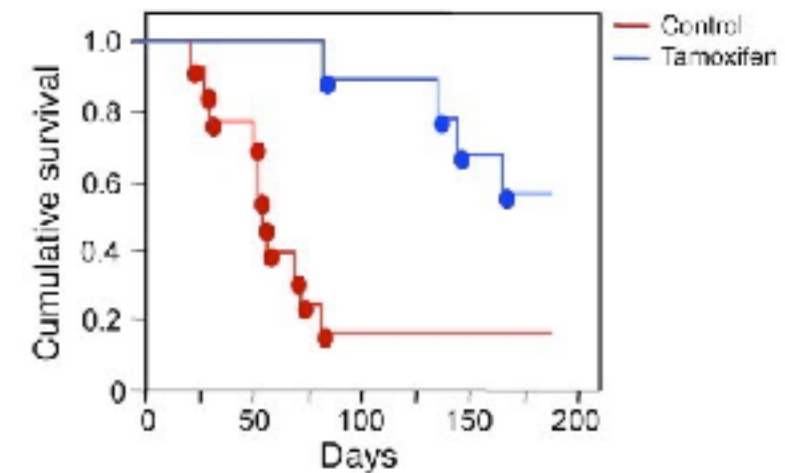


# Gene regulatory network from time series with inhibition of ID proteins

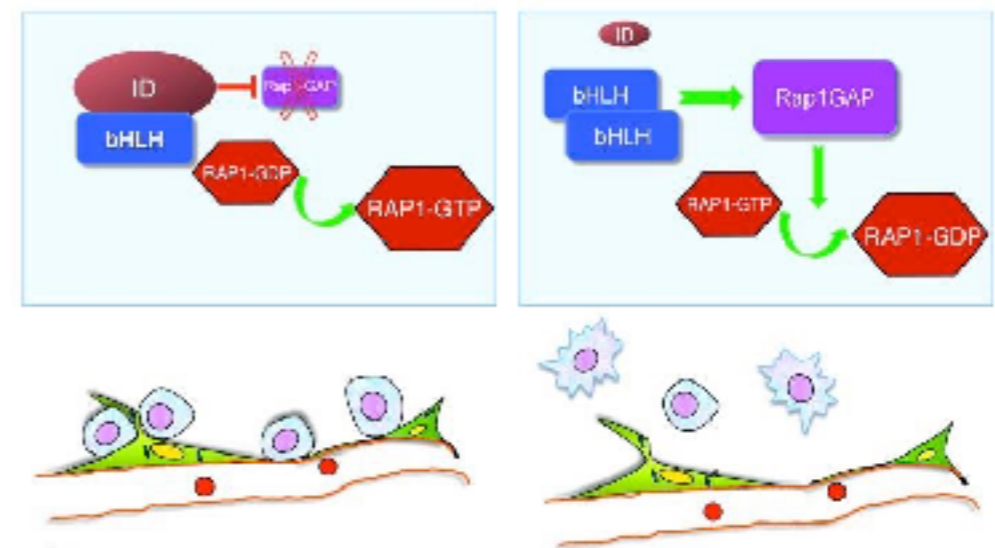


## Mesenchymal high-grade glioma is maintained by the ID-RAP1 axis

Francesco Nola,<sup>1</sup> Xudong Zhao,<sup>1</sup> Devendra Singh,<sup>1</sup> Ryan Sullivan,<sup>1</sup> Angelica Castano,<sup>1</sup> Antonio Verrico,<sup>1</sup> Pietro Zoppoli,<sup>1</sup> Dinorah Friedmann-Morvinski,<sup>2</sup> Erik Sulman,<sup>3</sup> Lindy Barrett,<sup>4</sup> Yuan Zhuang,<sup>5</sup> Inder Verma,<sup>2</sup> Robert Bonczra,<sup>4</sup> Ken Aldape,<sup>3</sup> Antonio Iavarone,<sup>1,6,7</sup> and Anna Lasorella<sup>1,7,8</sup>



Predicted network

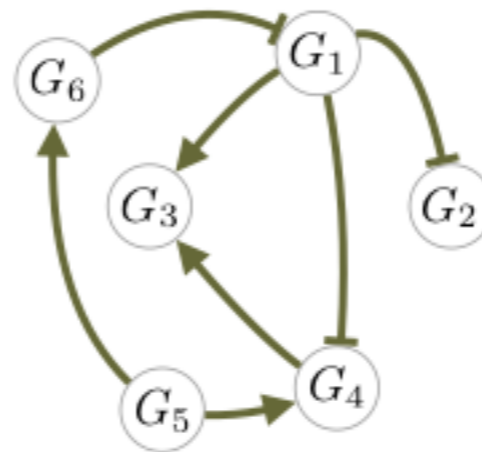


Biological mechanism

# Reverse Engineering of Gene Regulatory Networks

## Approaches

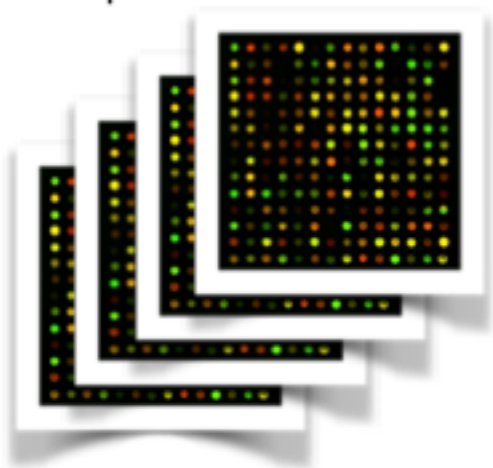
- Un-supervised



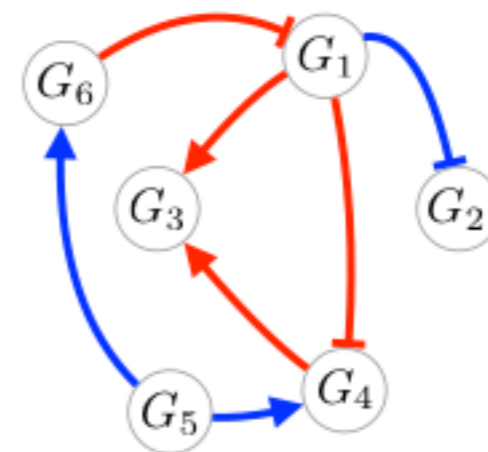
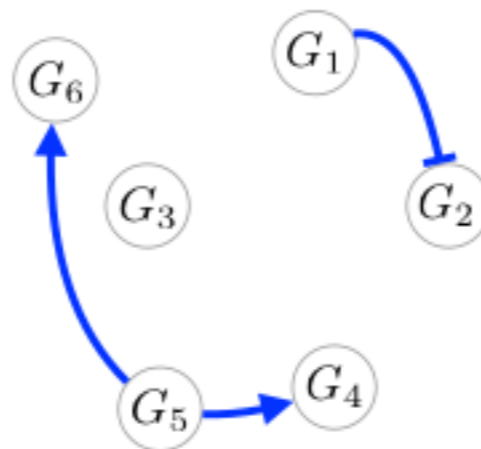
ARACNE  
CLR  
GENIE3  
TD-Aracne  
FormalM  
TNI  
BANJO

...

- Supervised



+



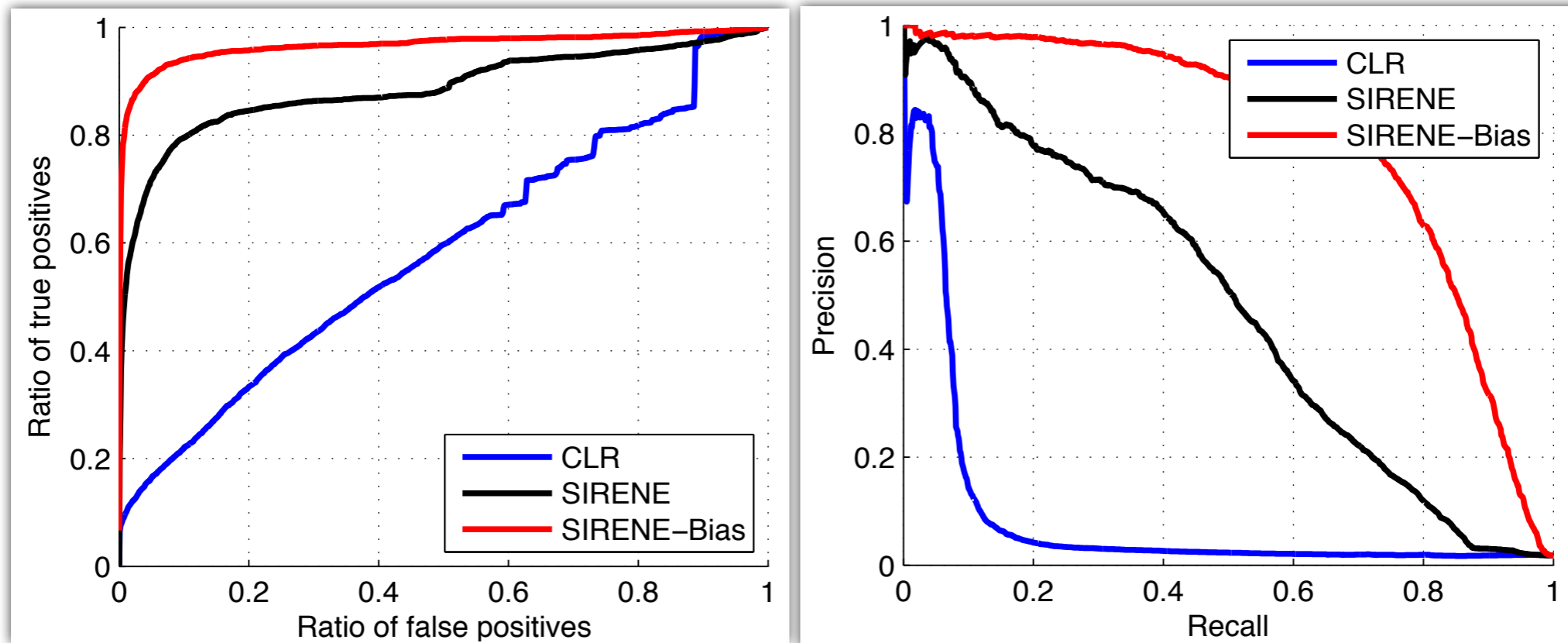
SIRENE  
POS-ONLY

...

## SIRENE: supervised inference of regulatory networks

Fantine Mordélet<sup>1,2,3,4,\*</sup> and Jean-Philippe Vert<sup>1,2,3</sup>

<sup>1</sup>Ecole des Mines de Paris, ParisTech, 35 rue Saint-Honoré, Fontainebleau F-77300, <sup>2</sup>Institut Curie, Paris F-75248, <sup>3</sup>INSERM, U900, Paris F-75248 and <sup>4</sup>CREST, INSEE, 3 av. Pierre Larousse, Malakoff, F-92240 France



| Method           | Recall at 60% of Precision | Recall at 80% of Precision |
|------------------|----------------------------|----------------------------|
| SIRENE           | 44,5%                      | 17,6%                      |
| ARACNe           | 1%                         | 0%                         |
| Bayesian network | 1%                         | 0%                         |

# Prediction of Myc Targets

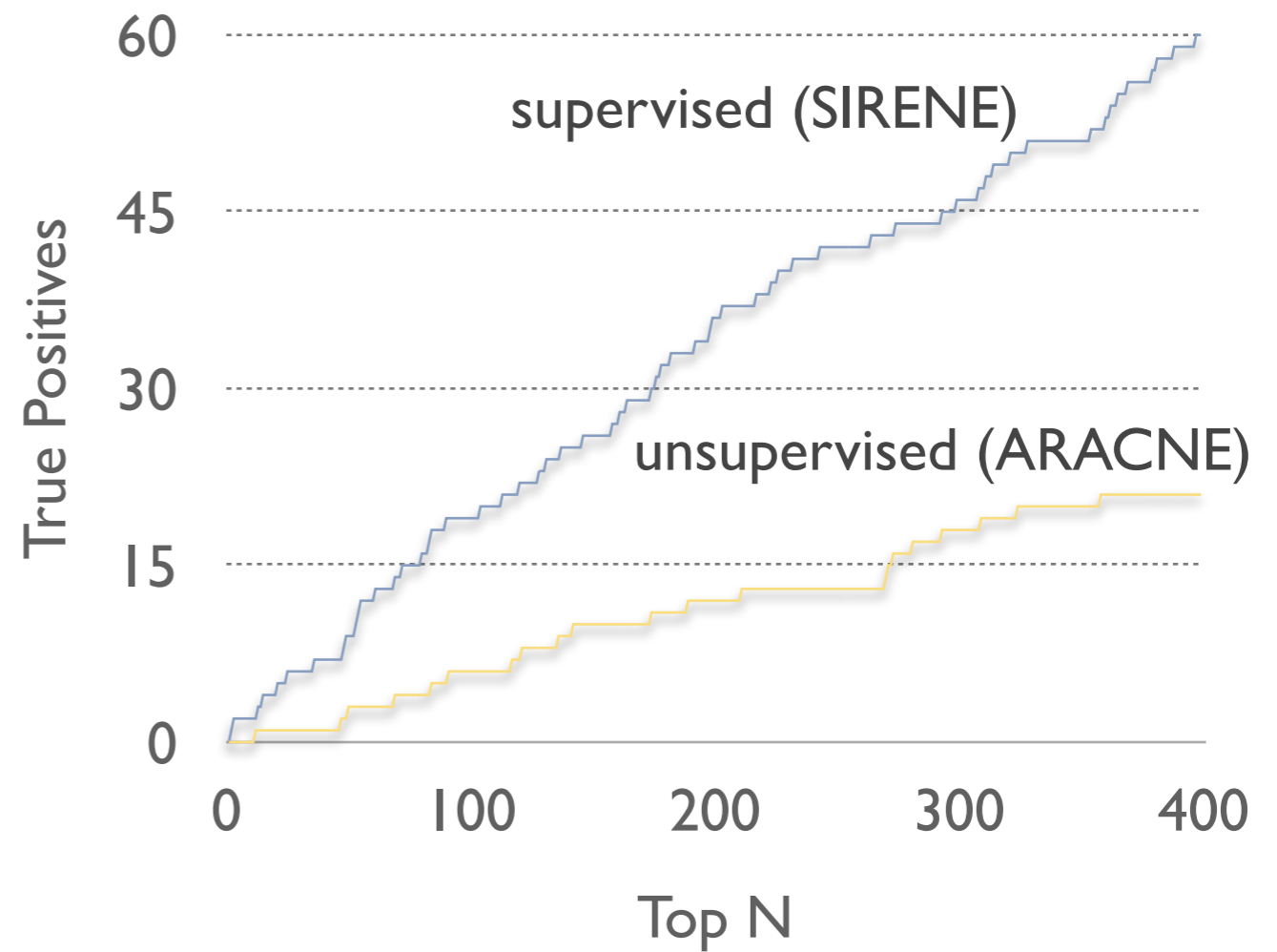
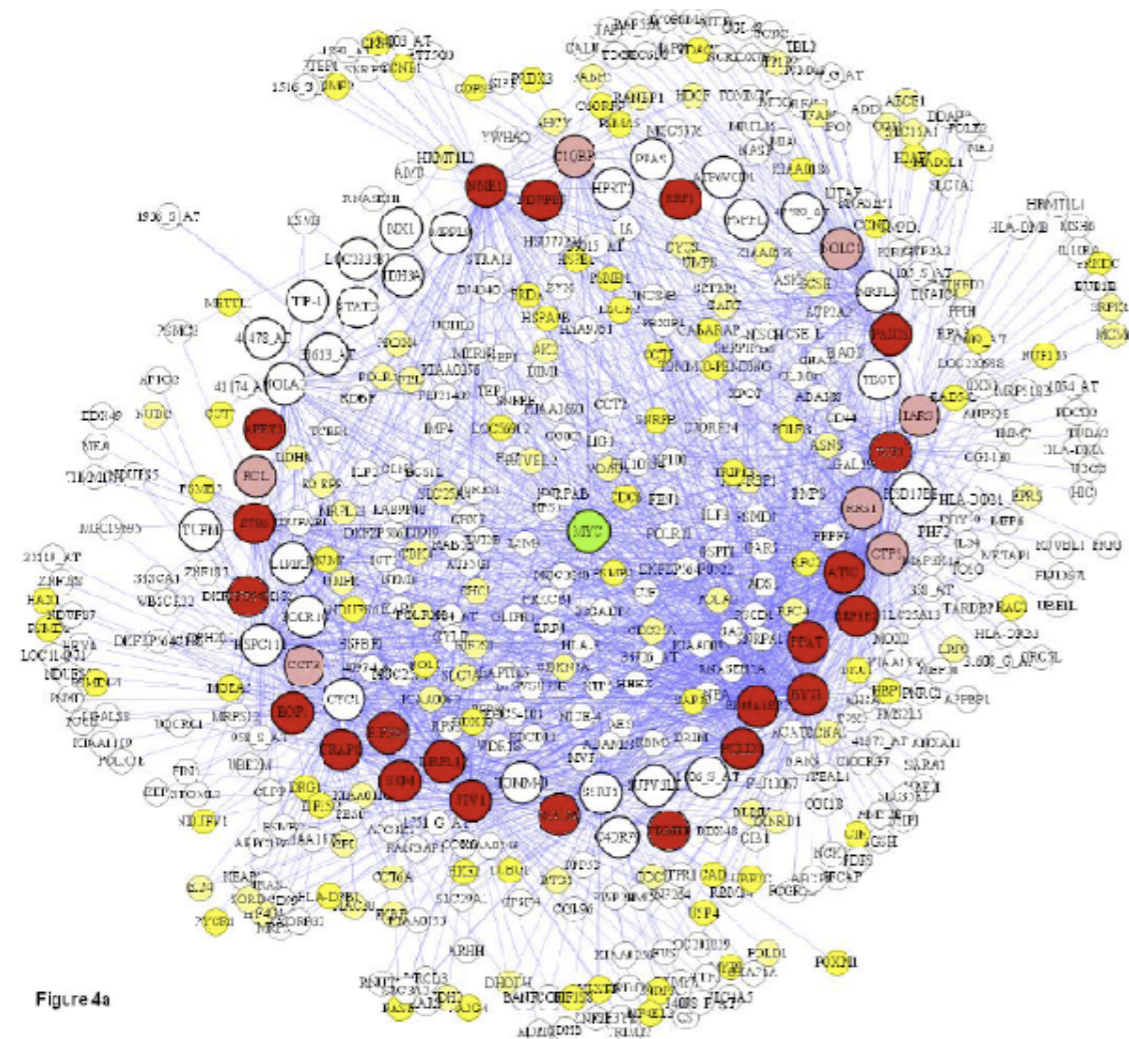
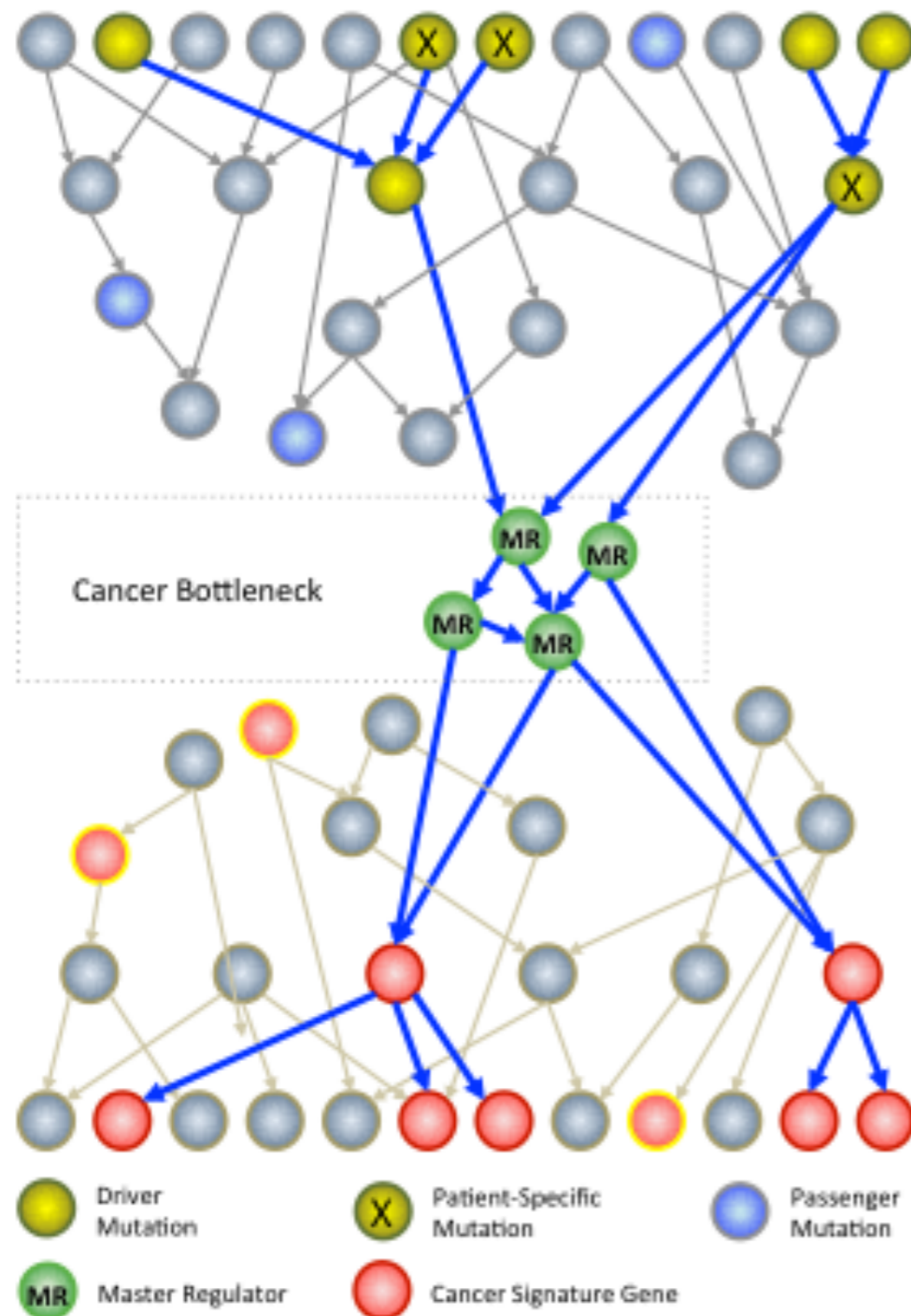


Figure 4a

# Master Regulators in Cancer biology



## Breast Cancer

- MR: SPDEF, ER $\alpha$ , FOXA1, GATA3 and PTTG1
- Fletcher et al. Nature Communications, 2013

## Glioblastoma

- MR: CEBP $\beta/\delta$  and Stat3
- Carro et al. Nature 2010

## DLBCL

- MR: NF $\kappa$ B
- Compagno et al, Nature 2009

# Master regulator analysis

---

- Scale-free networks  
(Hubs and Workers)

Social networks

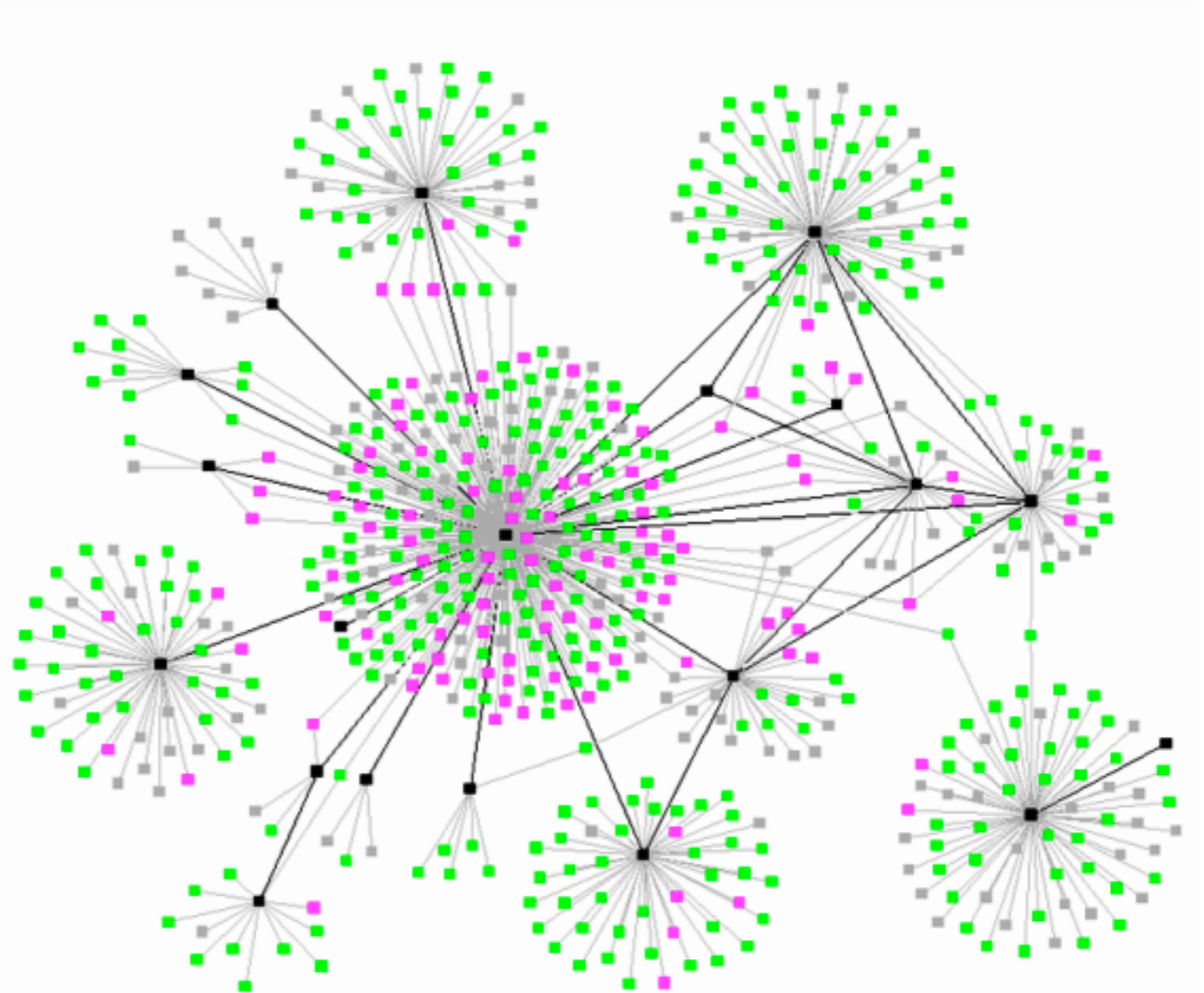
Computer Networks

Flight connections networks

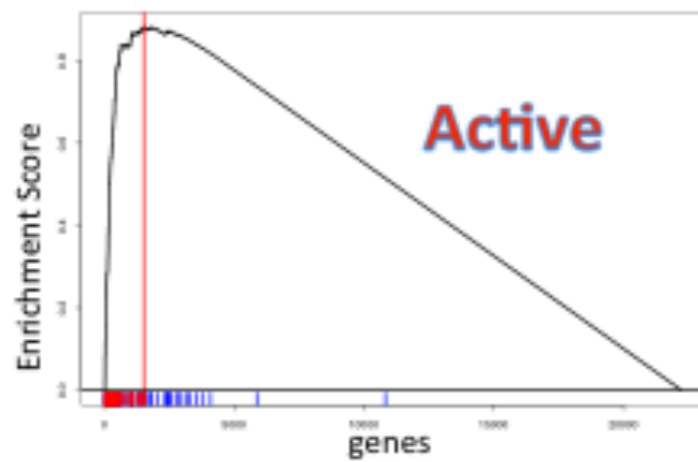
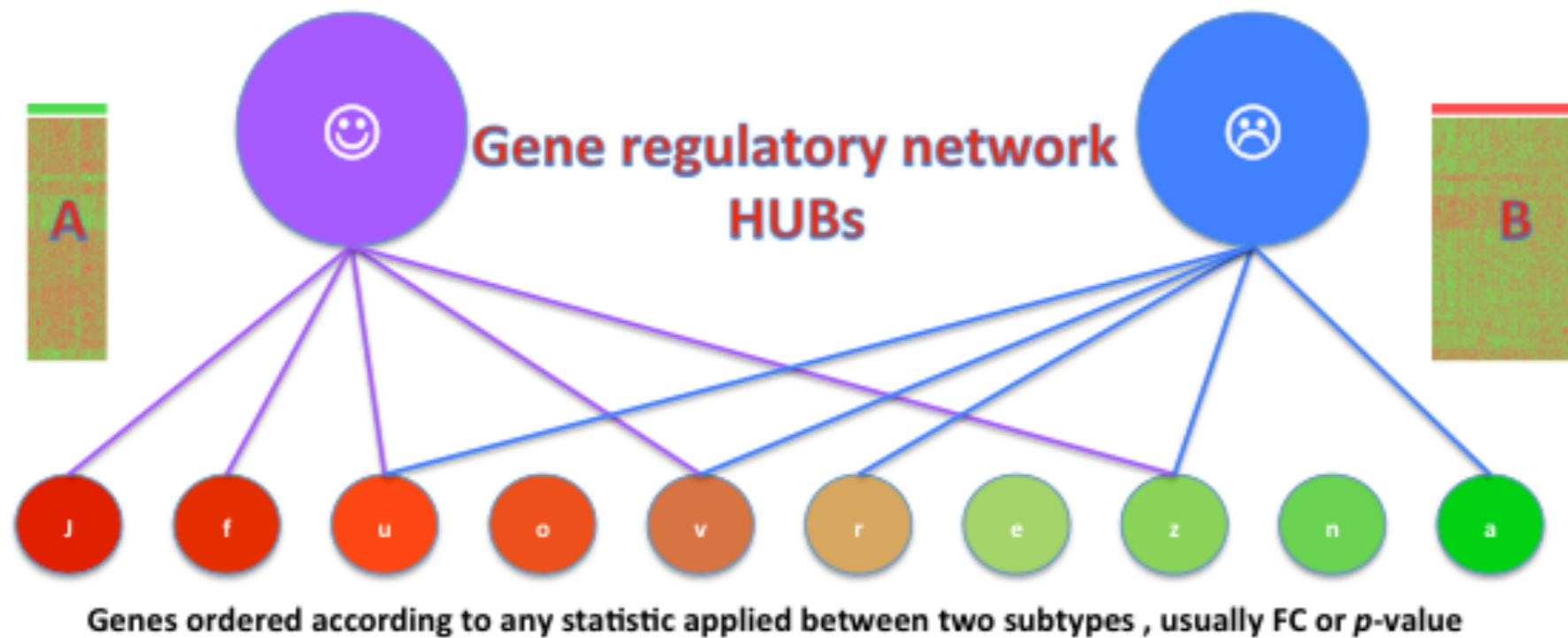
Financial Networks

co-authorship networks

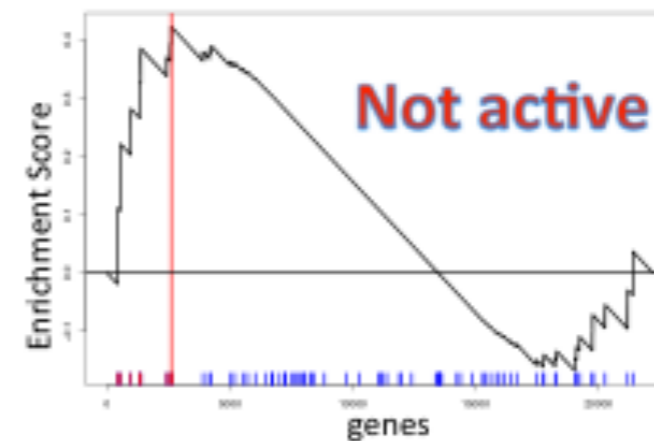
Biological networks



# Master regulator analysis



**GSEA using  
Regulon  
as Gene Set**





Research

Highly accessed

Open Access

## Systems biology analysis reveals NFAT5 as a novel biomarker and master regulator of inflammatory breast cancer

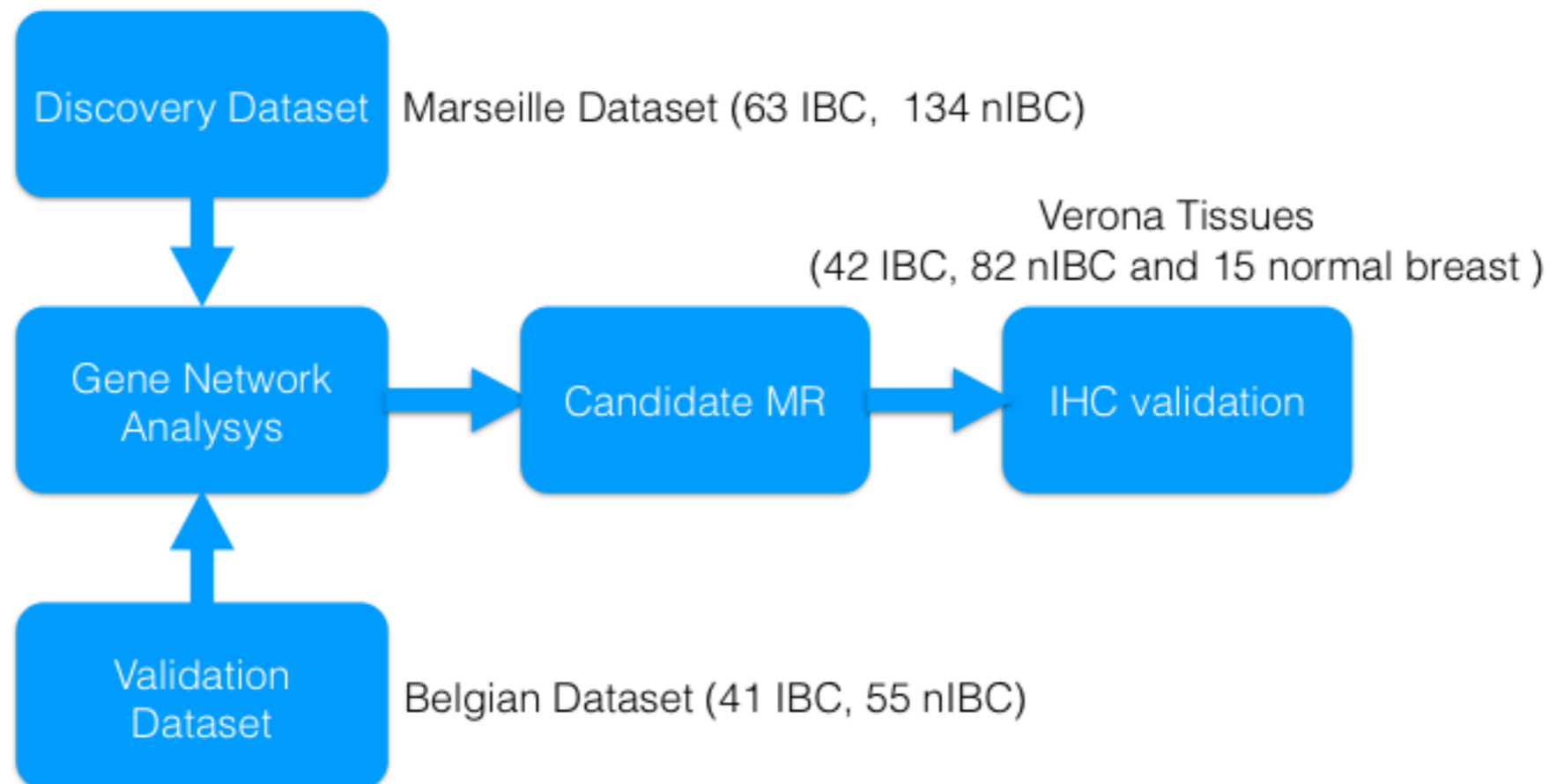
Andrea Remo<sup>†</sup>, Ines Simeone<sup>†</sup>, Massimo Pancione<sup>†</sup>, Pietro Parcesepe, Pascal Finetti, Luigi Cerulo, Halima Bensmail, Daniel Birnbaum, Steven J Van Laere, Vittorio Colantuoni, Franco Bonetti, François Bertucci, Erminia Manfrin<sup>\*</sup> and Michele Ceccarelli<sup>\*</sup>

\* Corresponding authors: Erminia Manfrin [erminia.manfrin@univr.it](mailto:erminia.manfrin@univr.it) - Michele Ceccarelli [m.ceccarelli@gmail.com](mailto:m.ceccarelli@gmail.com)

† Equal contributors

Journal of Translational Medicine 2015, 13:138

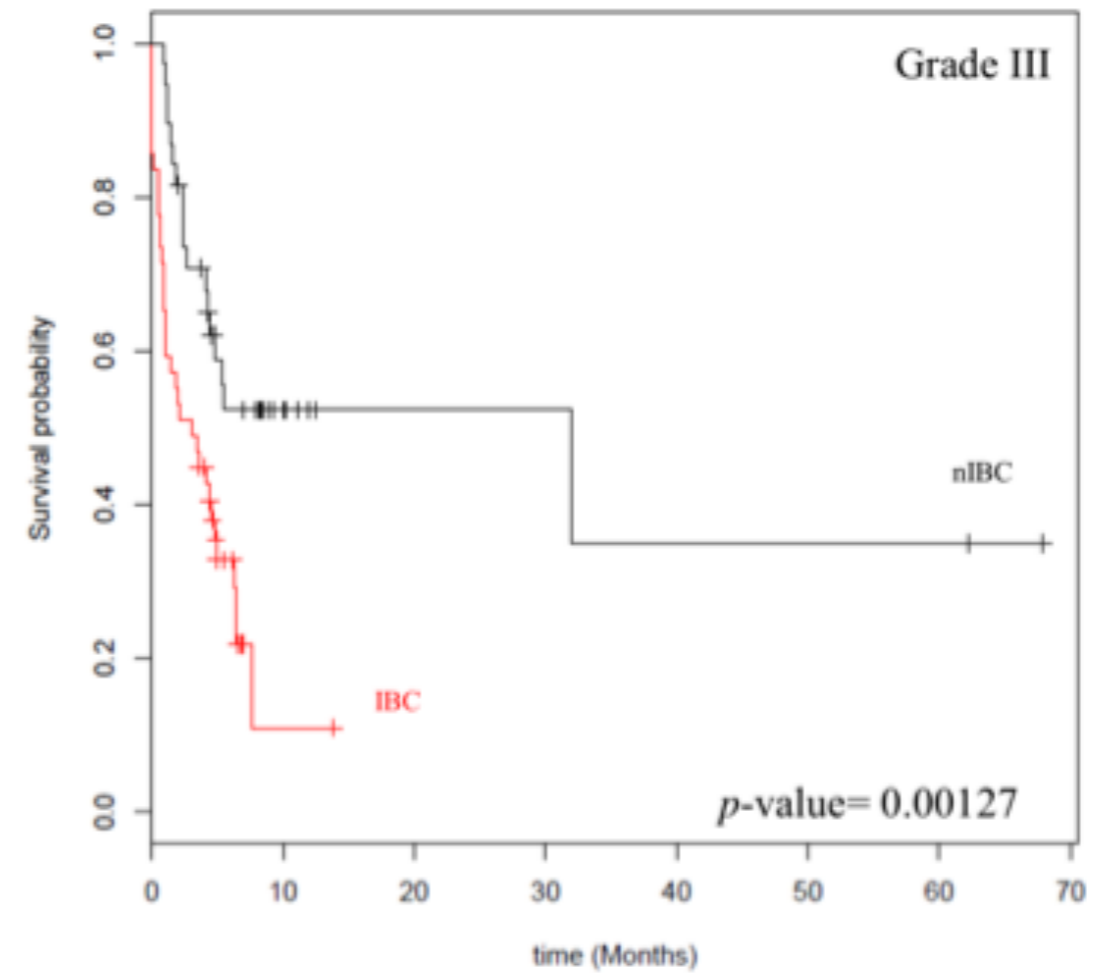
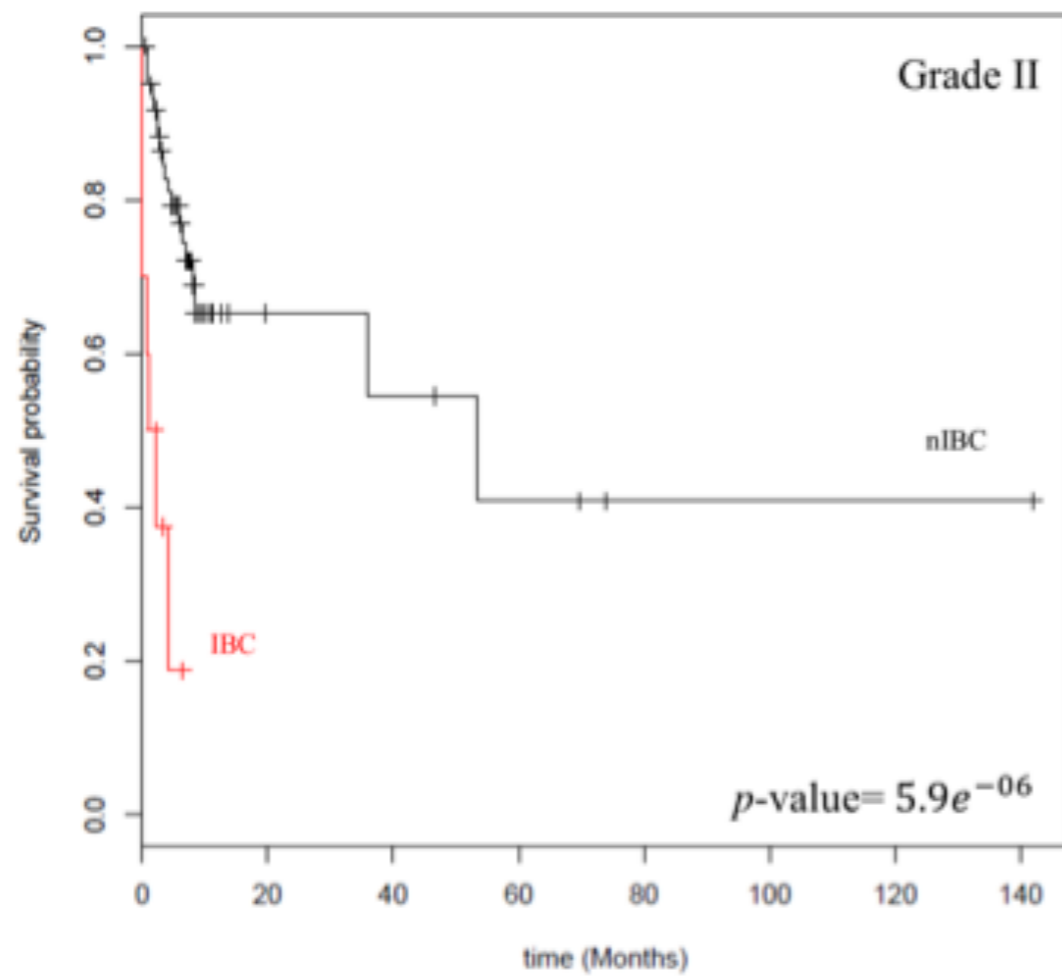
doi:10.1186/s12967-015-0492-2





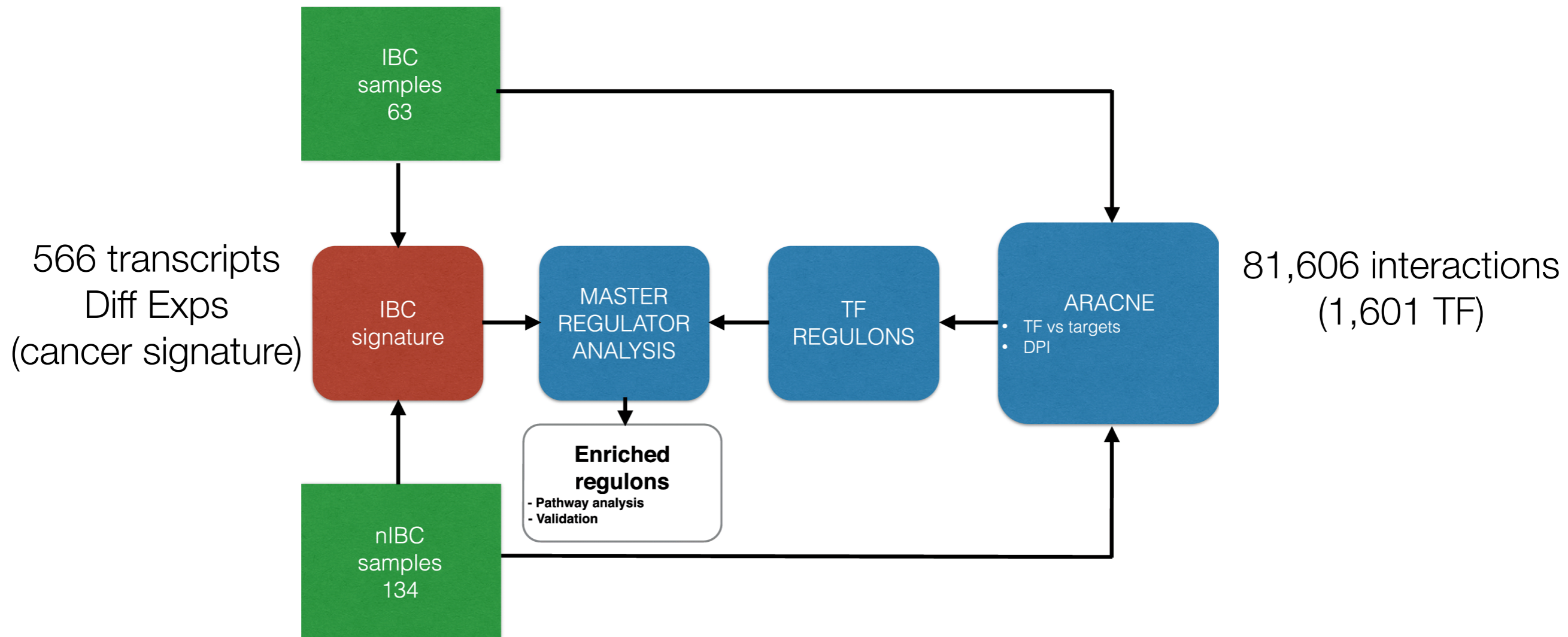
# Survival analysis

---



# Search for master regulators in Inflammatory Breast Cancer - bioinformatics pipeline

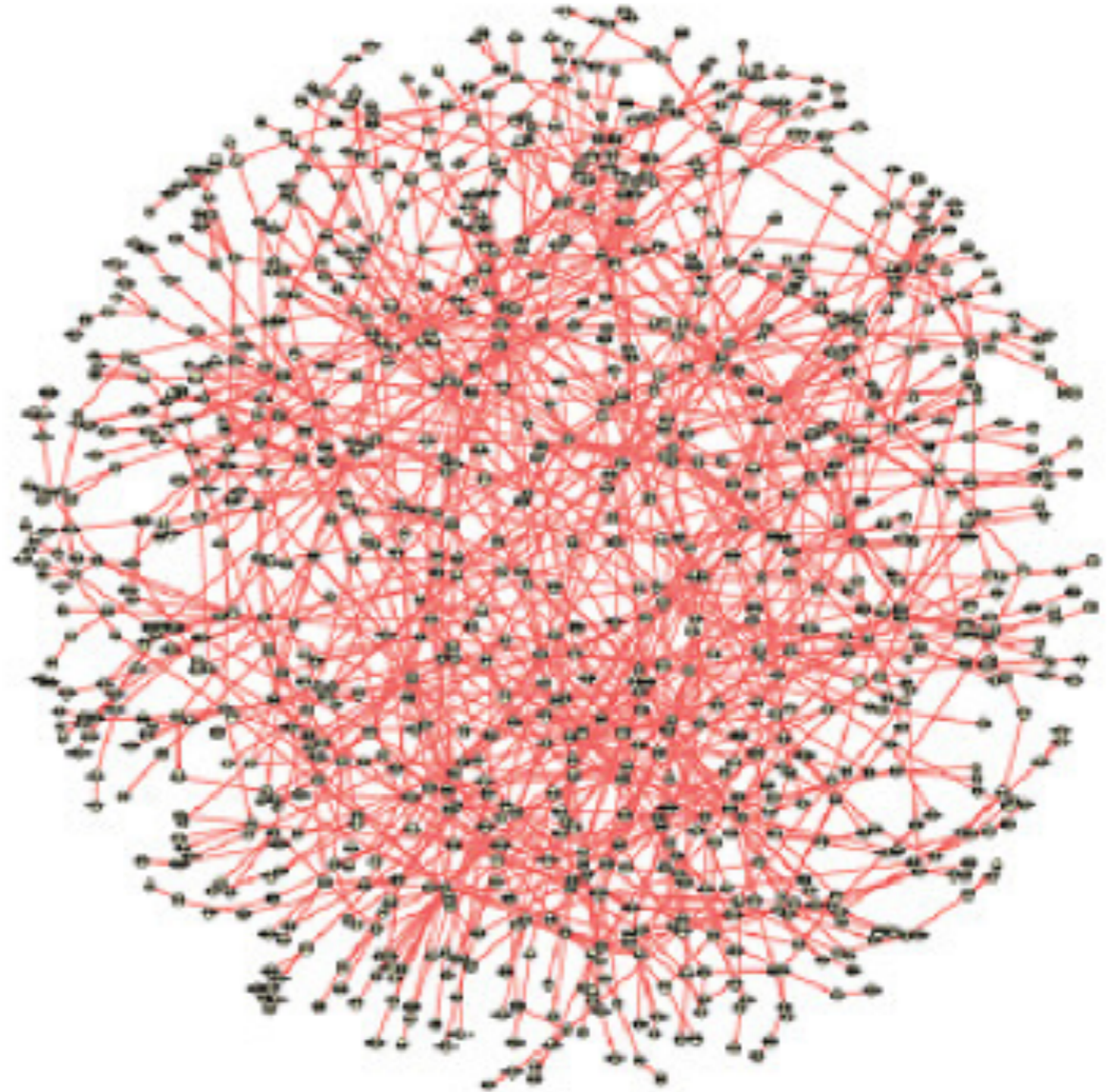
- Discovery dataset ~47000 transcripts



# Predicted network with ARACNE

---

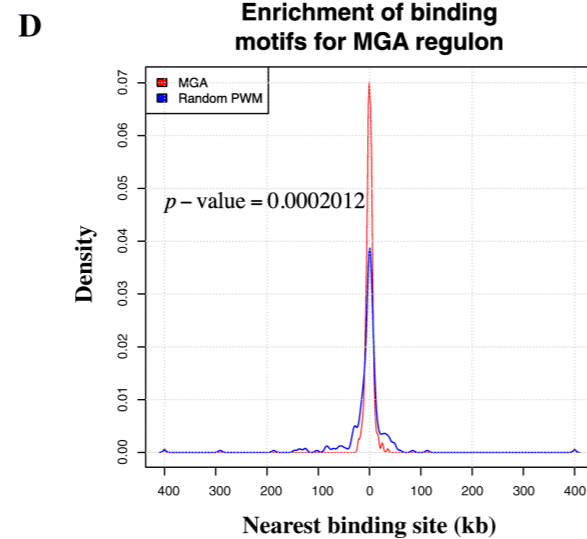
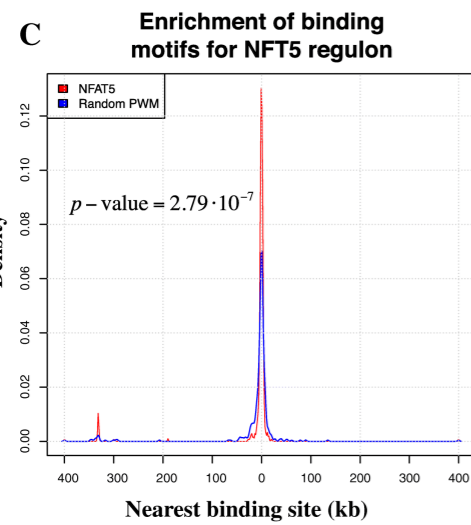
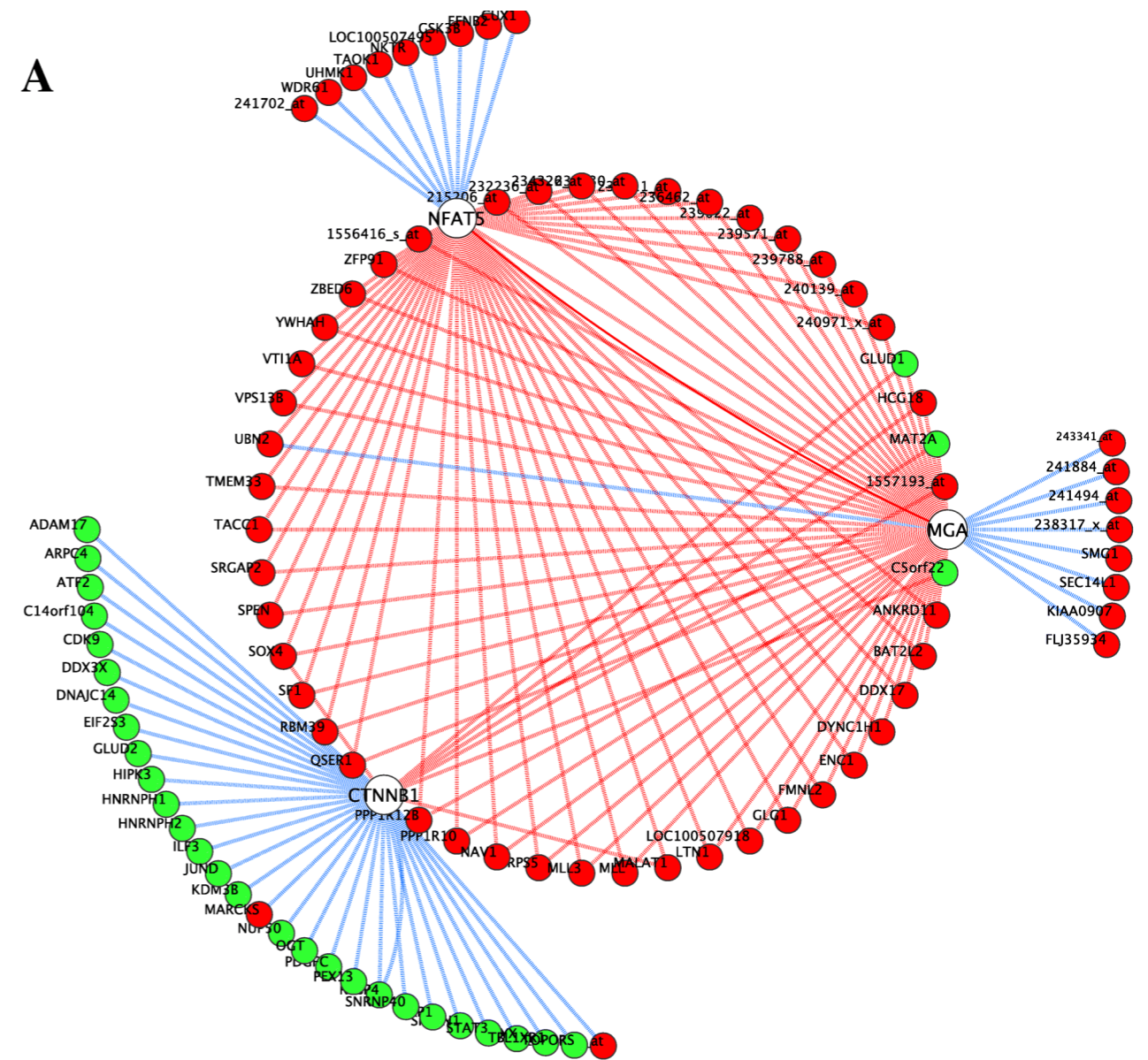
- 81,606 interactions for 1,601 Transcription factors



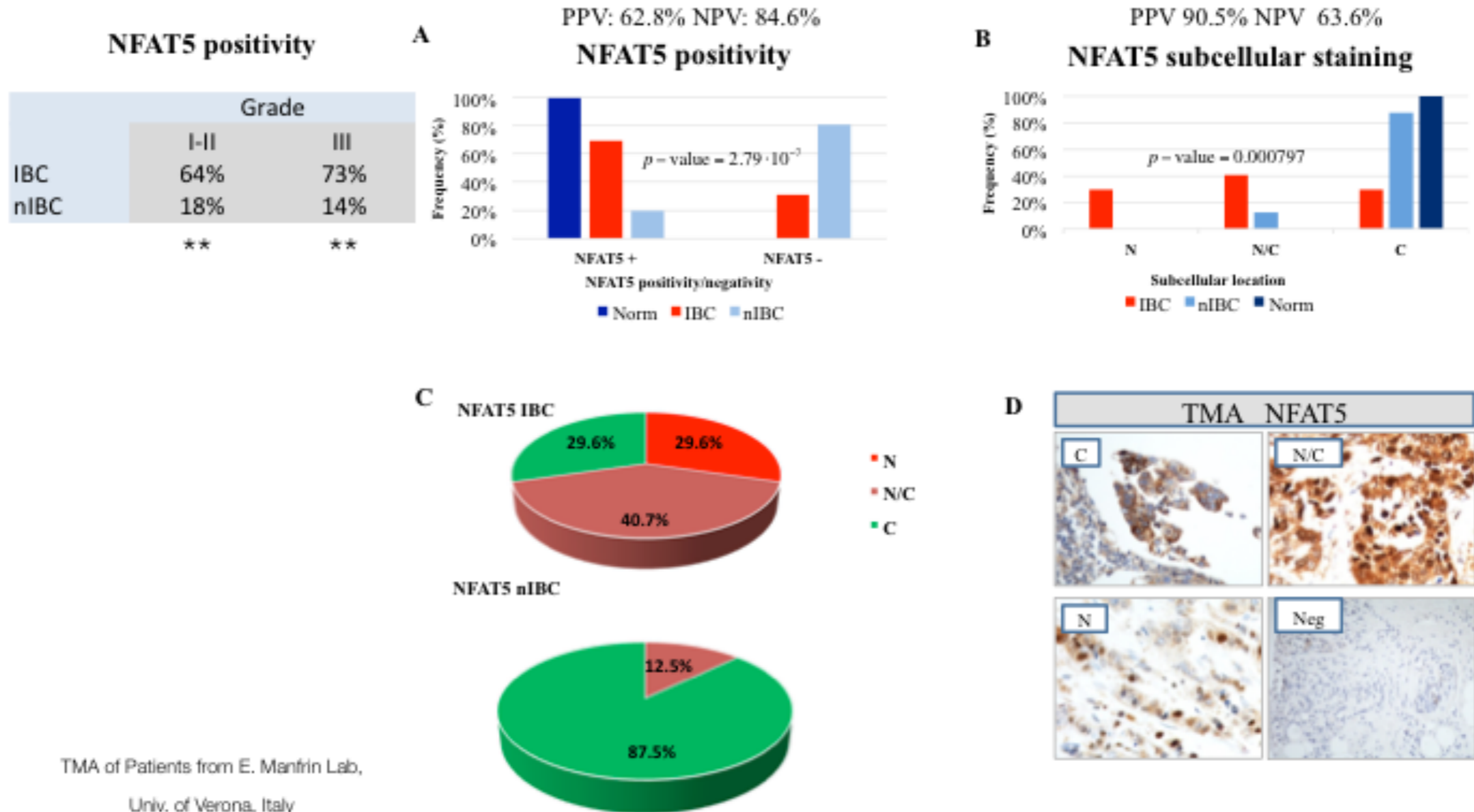
# Most enriched master regulators

- NFAT5 (Nuclear Factor of Activated T-Cells 5)  
pvalue  $10^{-29}$
- MGA (MAX Gene Associated)  
pvalue  $10^{-35}$
- CTNNB1 (Catenin Beta-1)  
pvalue  $10^{-33}$

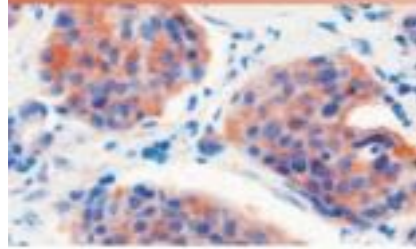
A



# NFAT5 is a novel master regulator and biological marker of IBC invasiveness

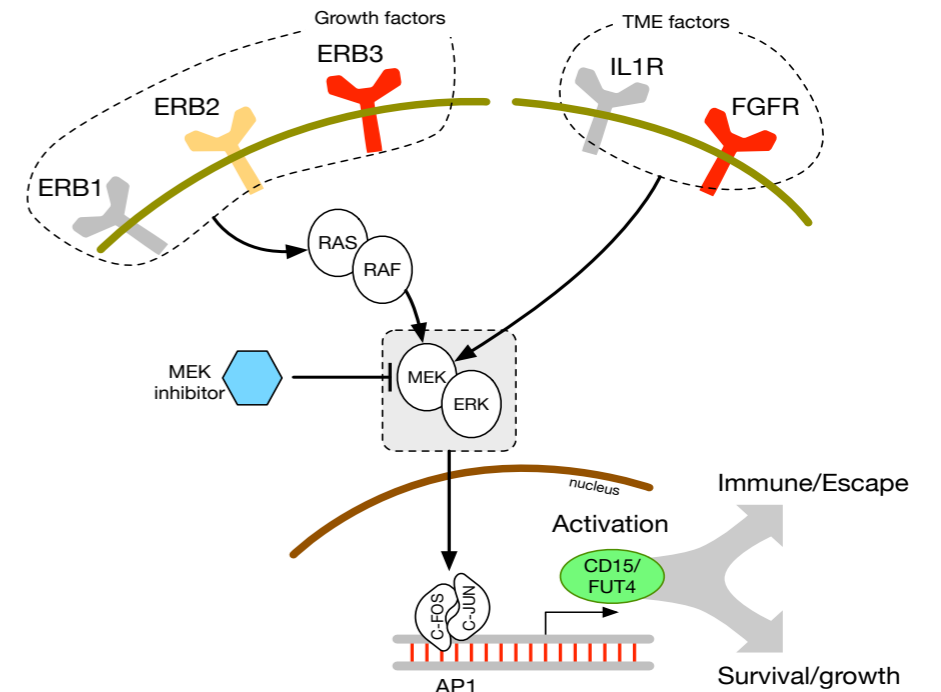
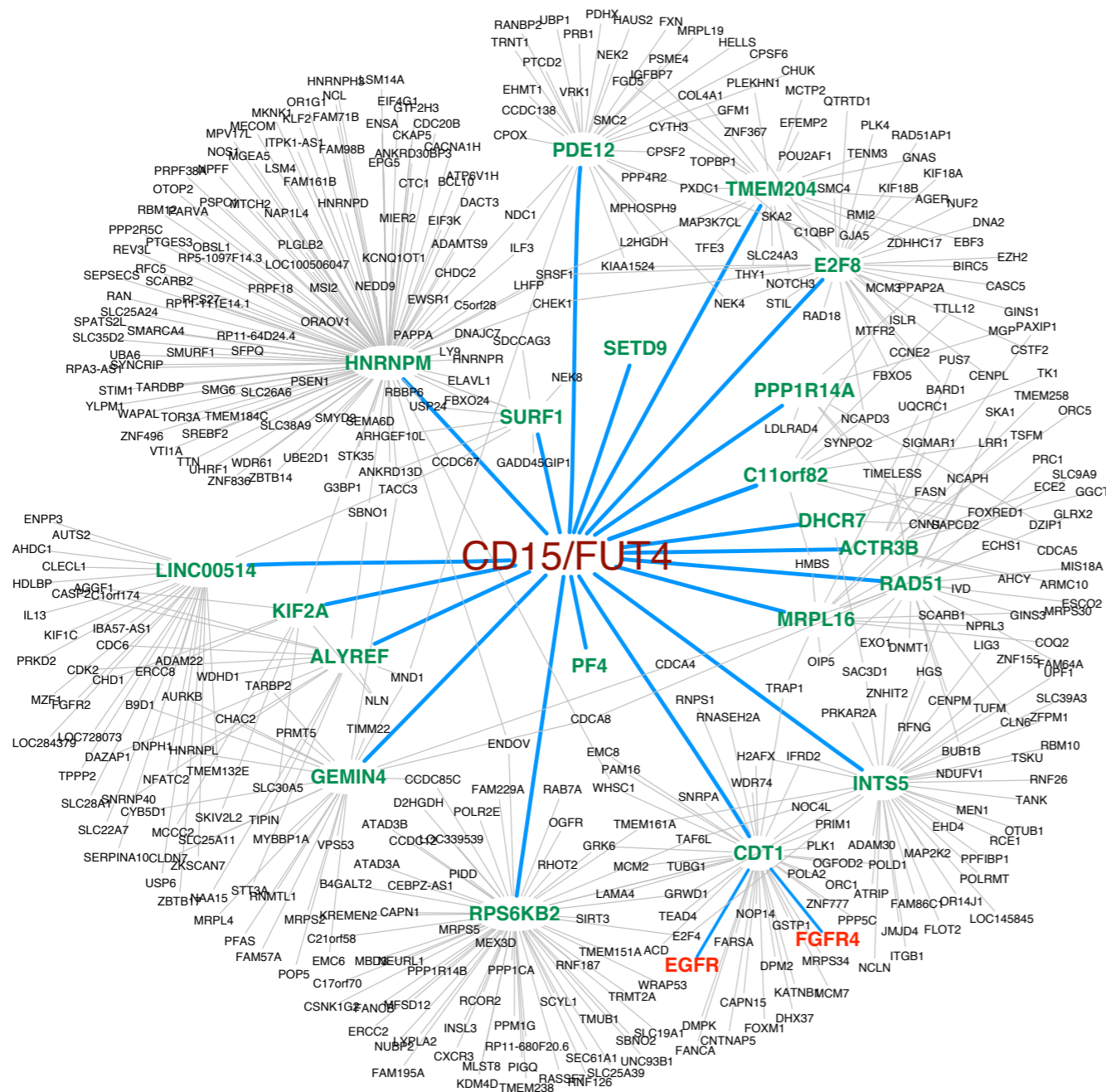


TMA of Patients from E. Manfrin Lab,  
Univ. of Verona, Italy  
42 IBC, 82 nIBC and 15 normal breast

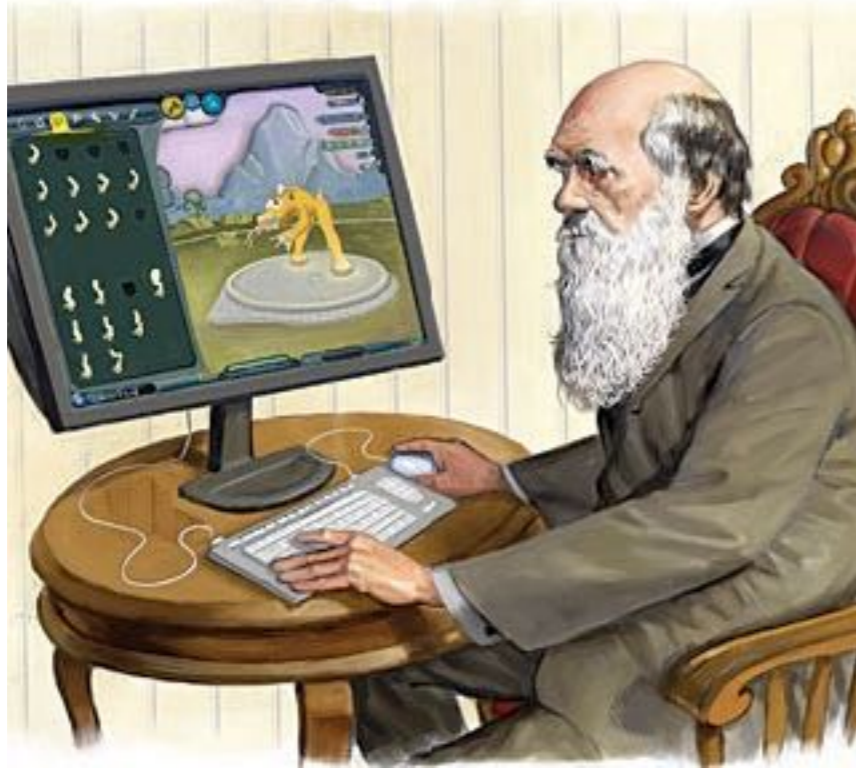


# Cancer-related CD15/FUT4 overexpression decreases benefit to agents targeting EGFR or VEGF acting as a novel RAF-MEK-ERK kinase downstream regulator in metastatic colorectal cancer

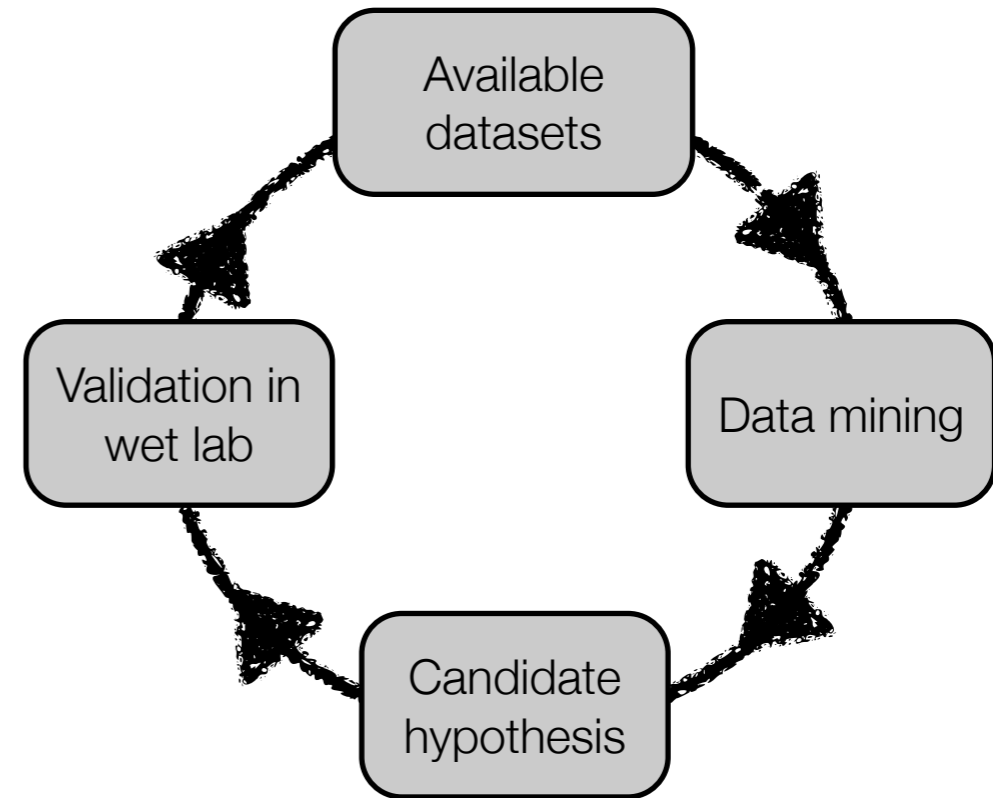
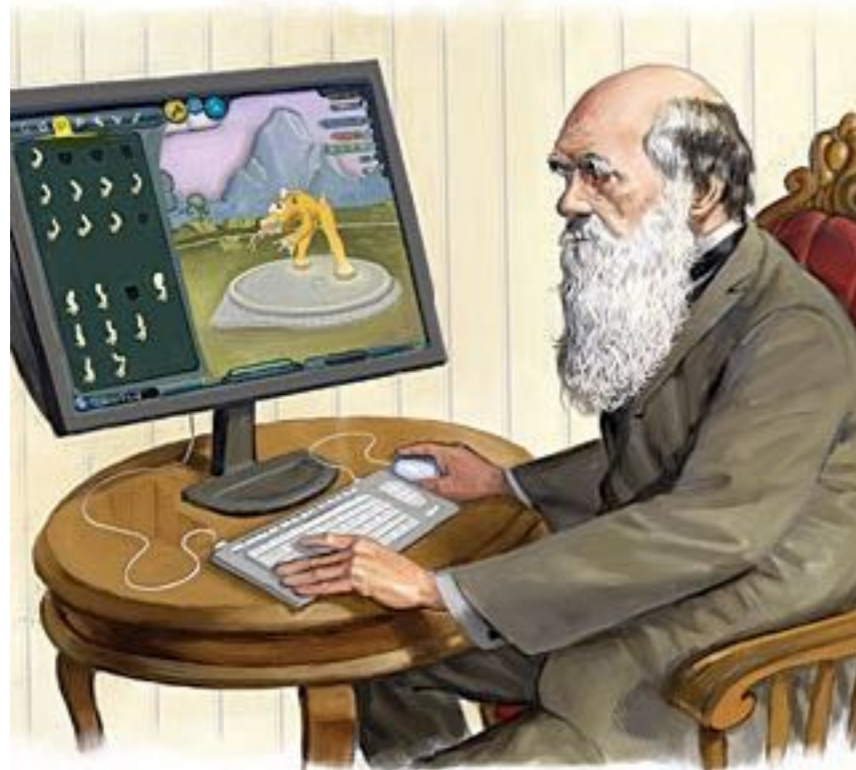
Guido Giordano<sup>1</sup>, Antonio Febraro<sup>1</sup>, Eugenio Tomaselli<sup>2</sup>, Maria Lucia Sarnicola<sup>3</sup>, Pietro Parcesepe<sup>4</sup>, Domenico Parente<sup>2</sup>, Nicola Forte<sup>2</sup>, Alessio Fabozzi<sup>1</sup>, Andrea Remo<sup>5</sup>, Andrea Bonetti<sup>5</sup>, Erminia Manfrin<sup>4</sup>, Somayehsadat Ghasemi<sup>4</sup>, Michele Ceccarelli<sup>6,7</sup>, Luigi Cerulo<sup>6,7</sup>, Flavia Bazzoni<sup>4</sup> and Massimo Pancione<sup>7\*</sup>



... in conclusion

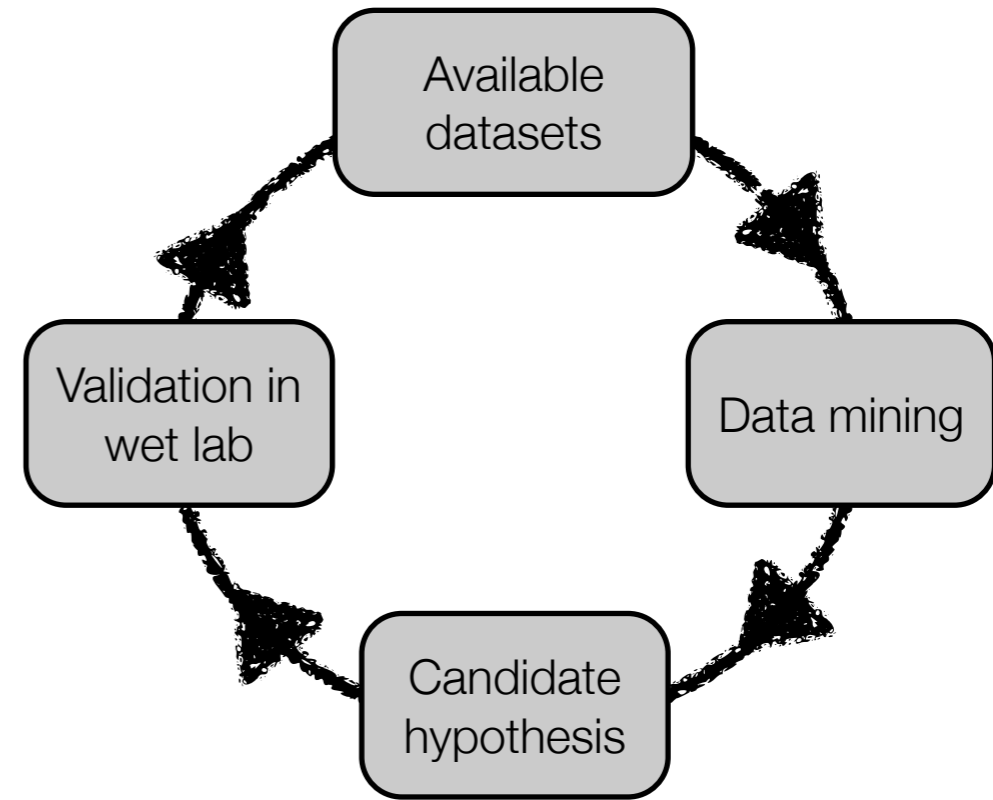
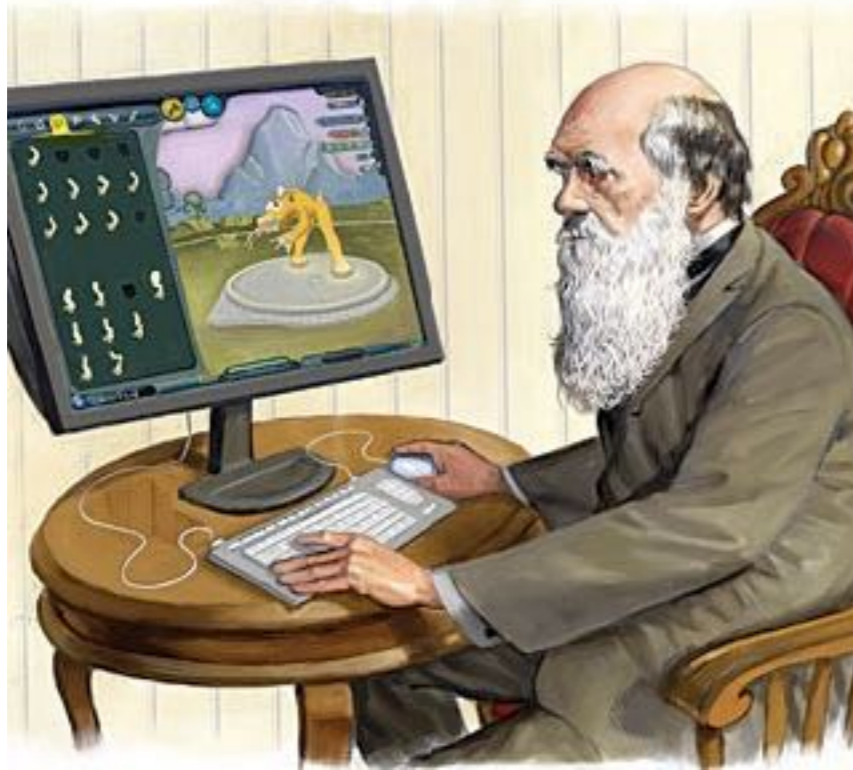


... in conclusion

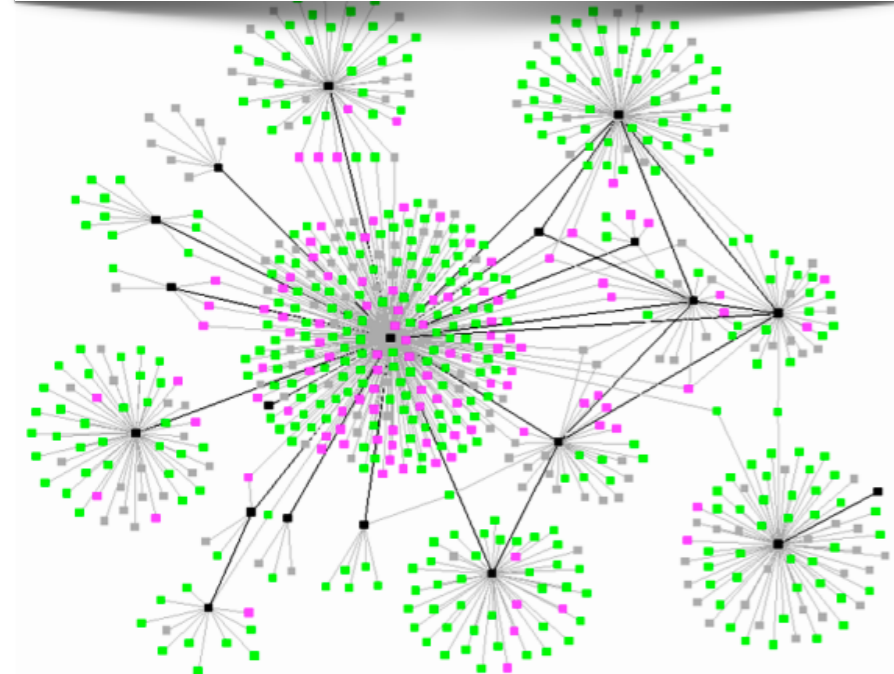




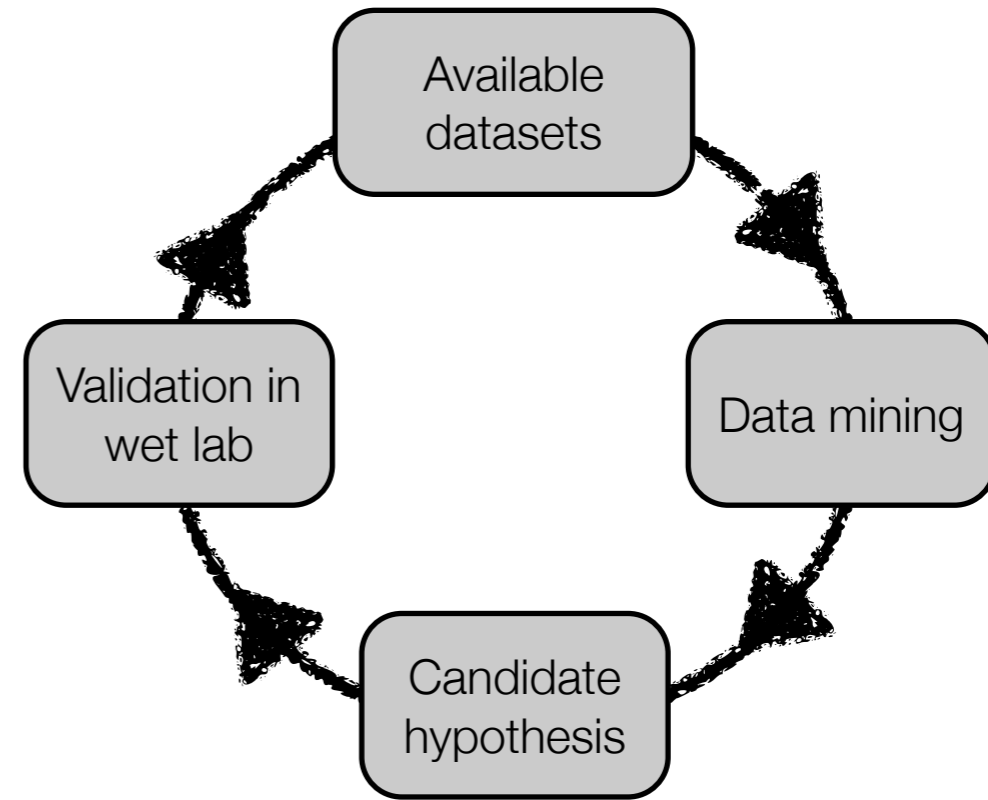
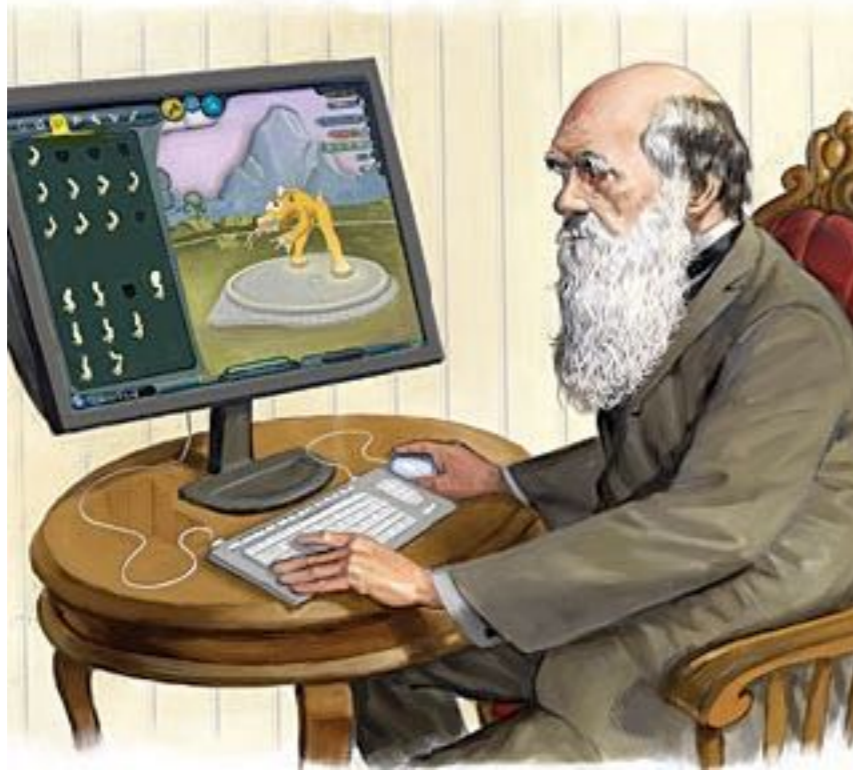
... in conclusion



Apply system biology approaches  
on wide public datasets

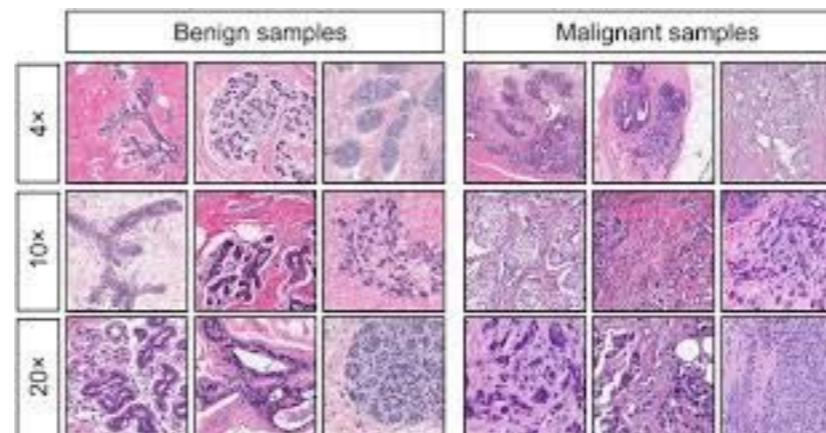
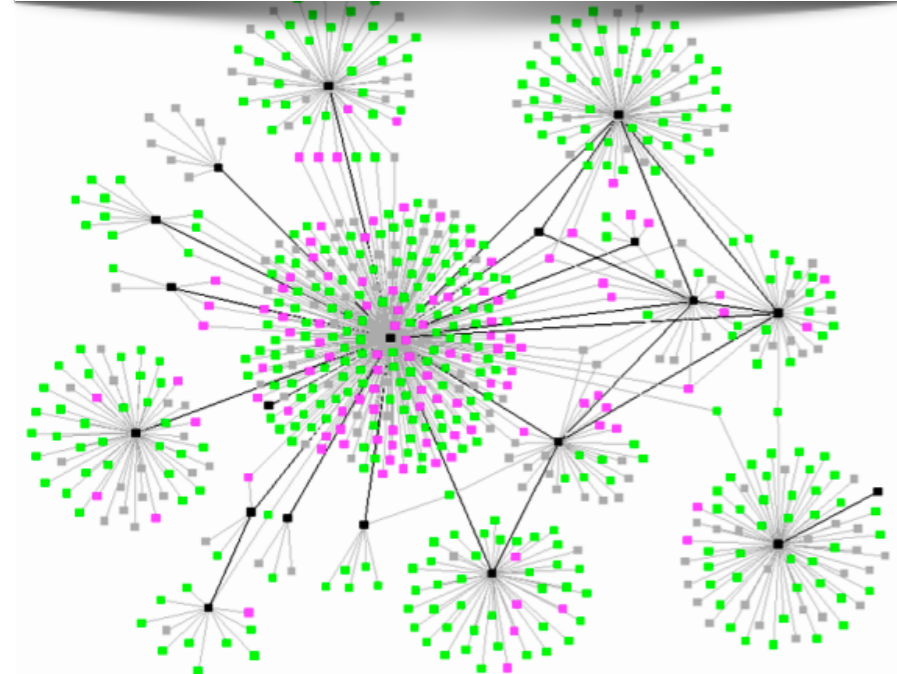


... in conclusion

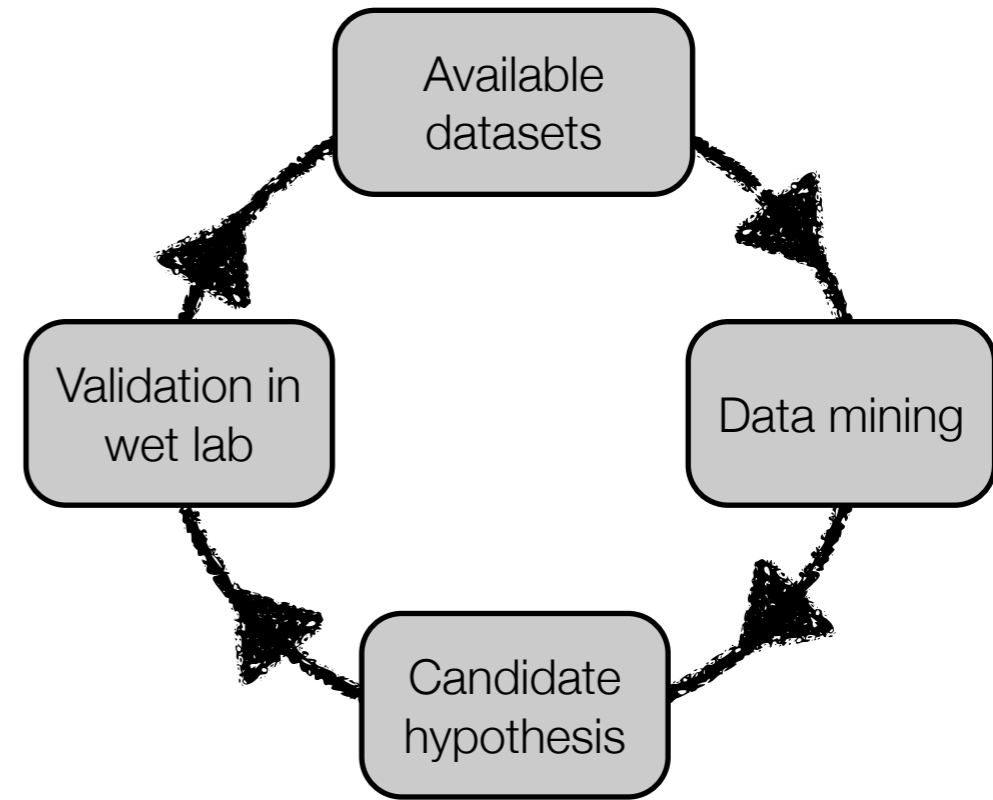
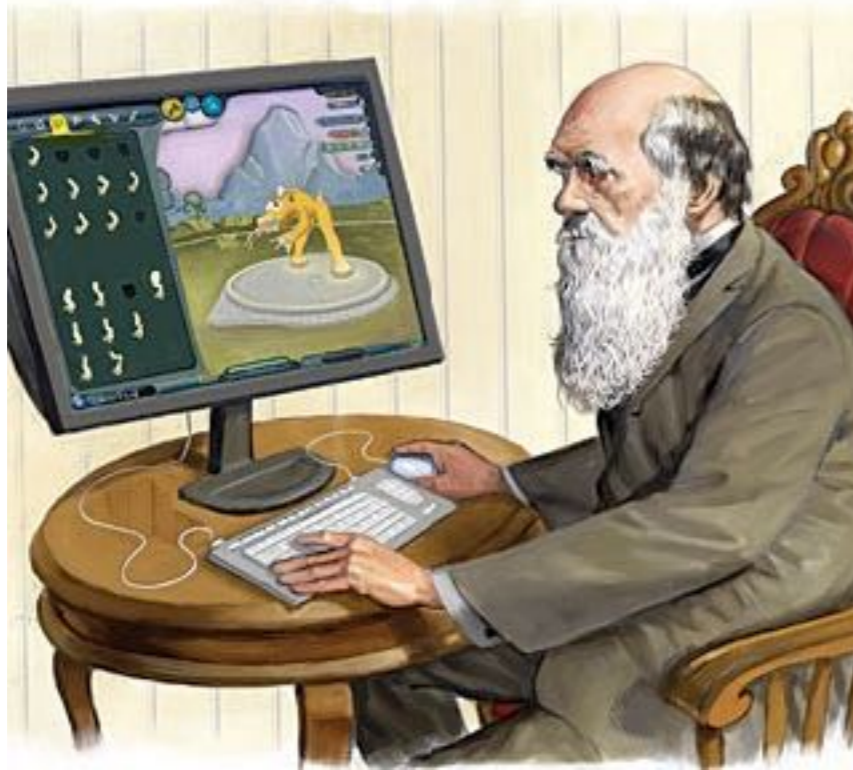


Apply system biology approaches on wide public datasets

Integrate results with the lab research hypothesis context



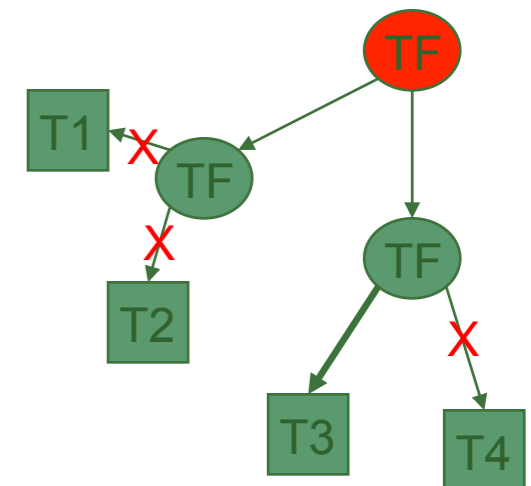
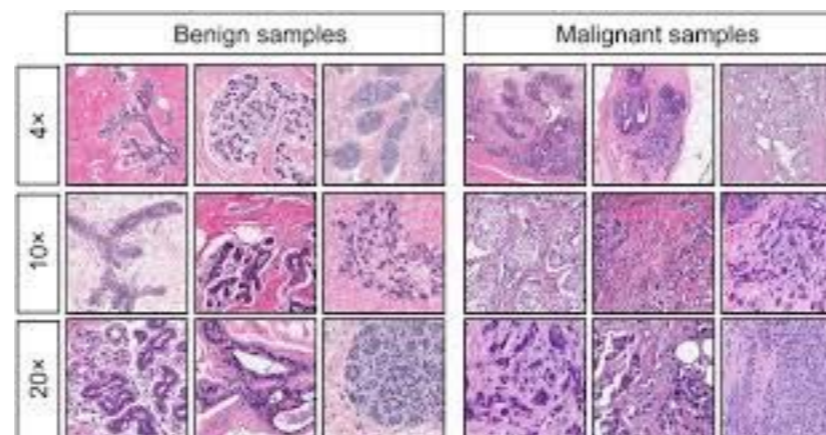
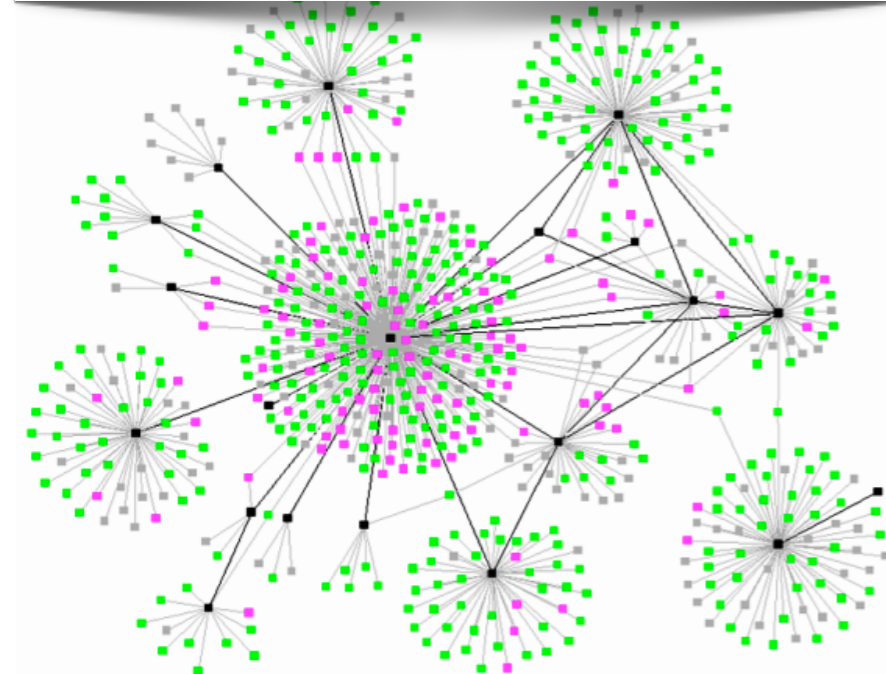
... in conclusion



Apply system biology approaches on wide public datasets

Integrate results with the lab research hypothesis context

Generate new hypothesis to be validated in wet lab



# Bioinformatics group @unisannio

---

Michele Ceccarelli  
Stefano M. Pagnotta  
Luigi Cerulo  
Fulvio D'Angelo  
Luciano Garofano  
Giovanna M. Ventola  
Ines Simeone  
Teresa M. Noviello



[www.bioinformatics-sannio.org](http://www.bioinformatics-sannio.org)

- Collaborations

University of Sannio, DST (M. Pancione, V. Colantuoni)

University of Sannio, DING (A. Santone)

University of Naples Federico II (G. Falco)

Stazione Zoologia "A. Dorhn" (S. D'Aniello, A. Spagnuolo)

University of Verona (A. Remo, E. Manfrin, P. Parcepese, F. Bonetti)

Qatar Computing Research Institute (M. Ceccarelli, I. Simeone)

Columbia University (A. Iavarone, P. Zoppoli, L. Garofano)