Reconstructing networks from **experimental** and **natural** genetic **perturbations**

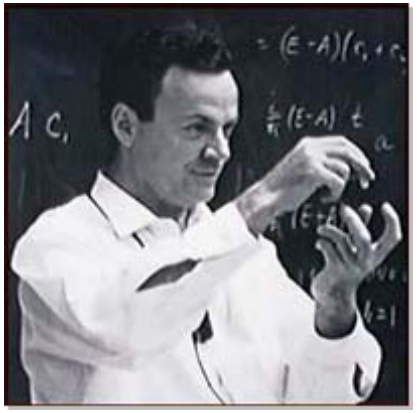
Phenotypes, pathways and **cancer**

Florian Markowetz
markowetzlab.org

LICSB 2010
Warwick University

CANCER RESEARCH UK Cambridge Research Institute

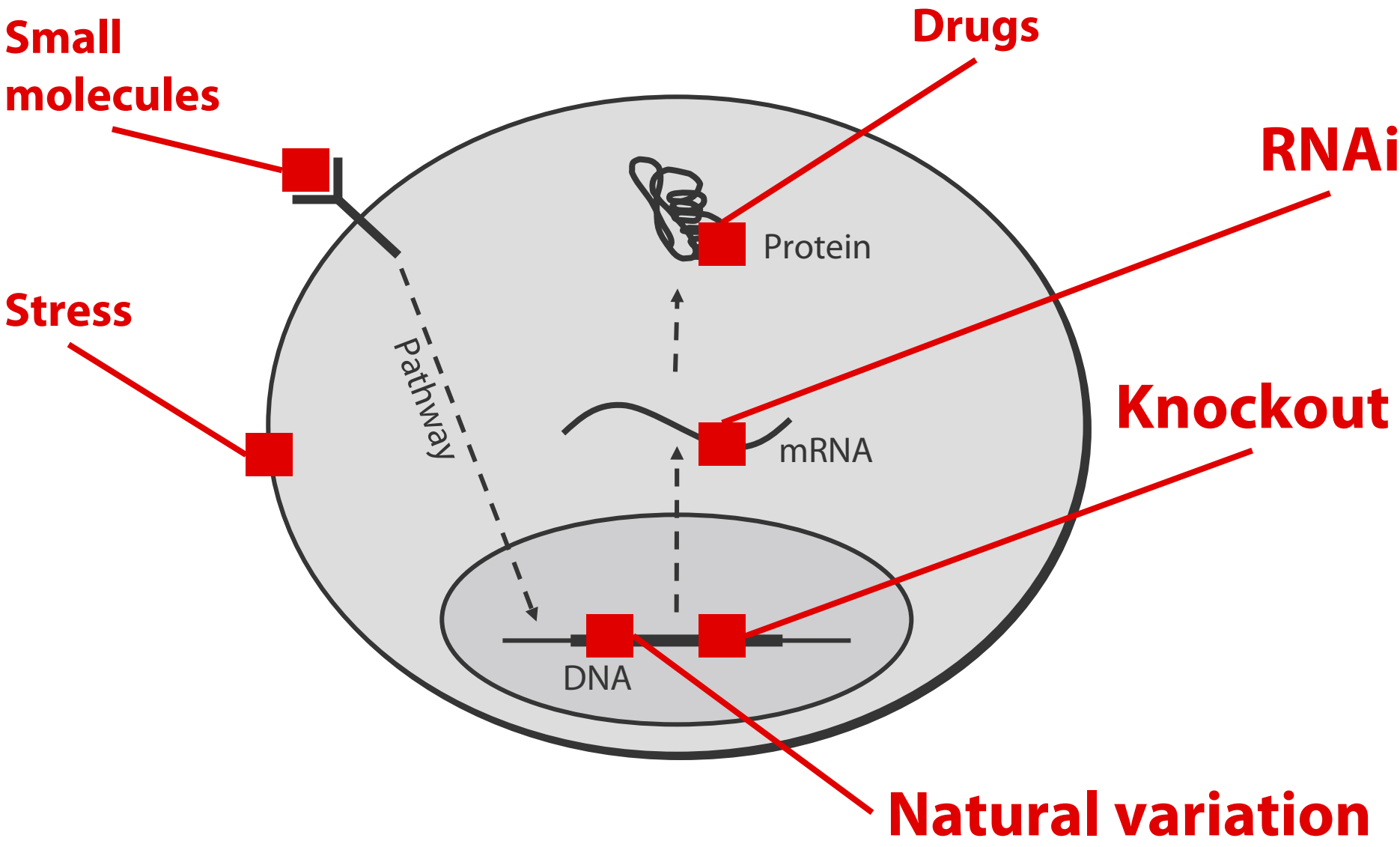
How to understand a complex system?



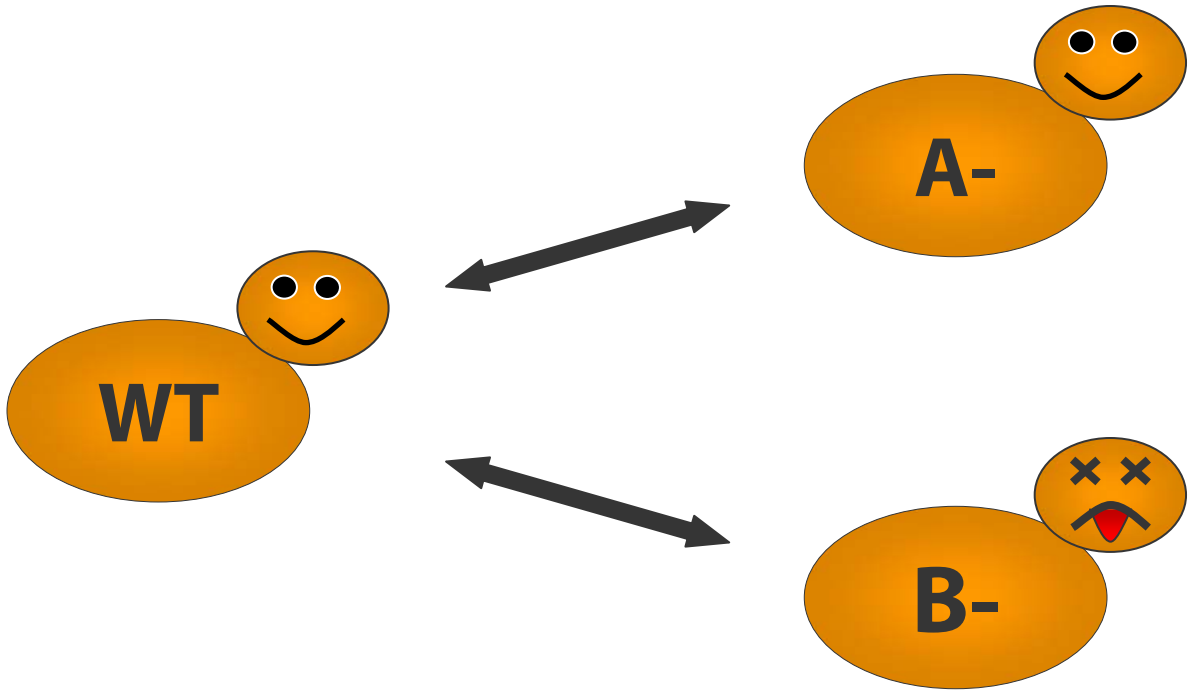
Richard Feynman:
"What I cannot **create**, I do not understand."

Functional Genomics:
"What I cannot **break**, I do not understand."

Breaking the system



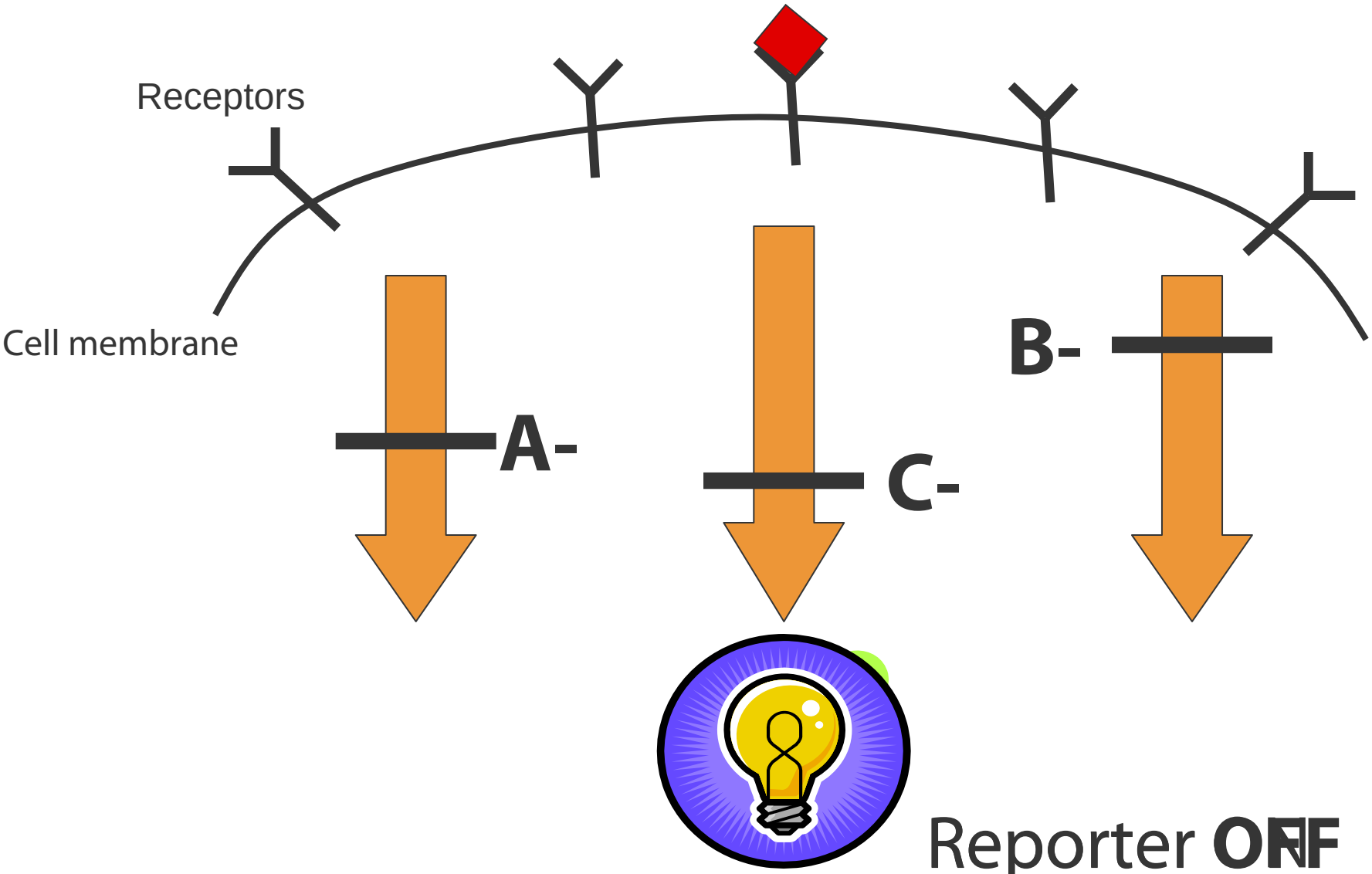
Phenotype: **viability** versus cell death



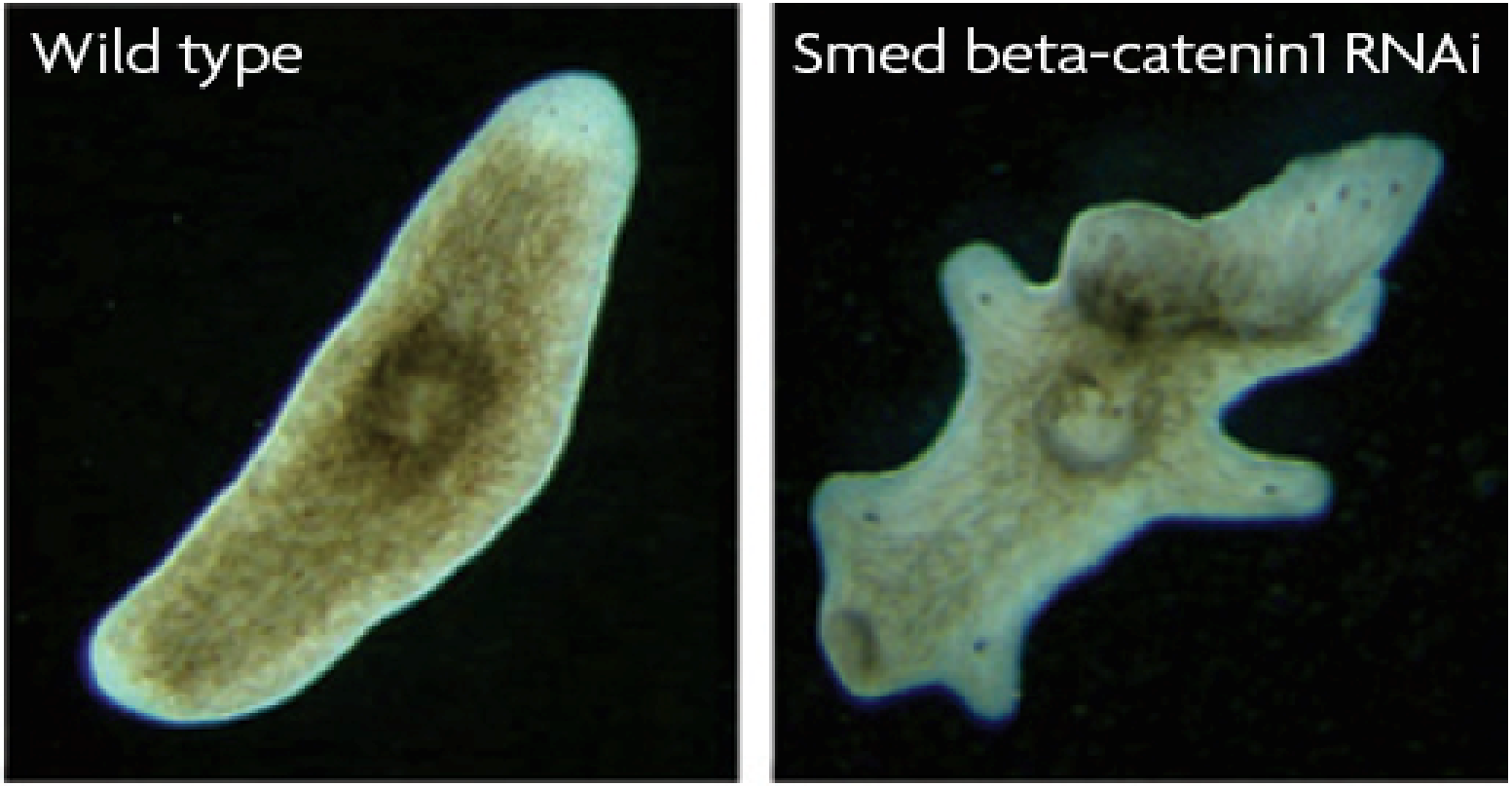
Interpretation:
- non-essential gene
- redundancy

Interpretation:
an **essential** gene
for the organism

Phenotype: **pathway activity**



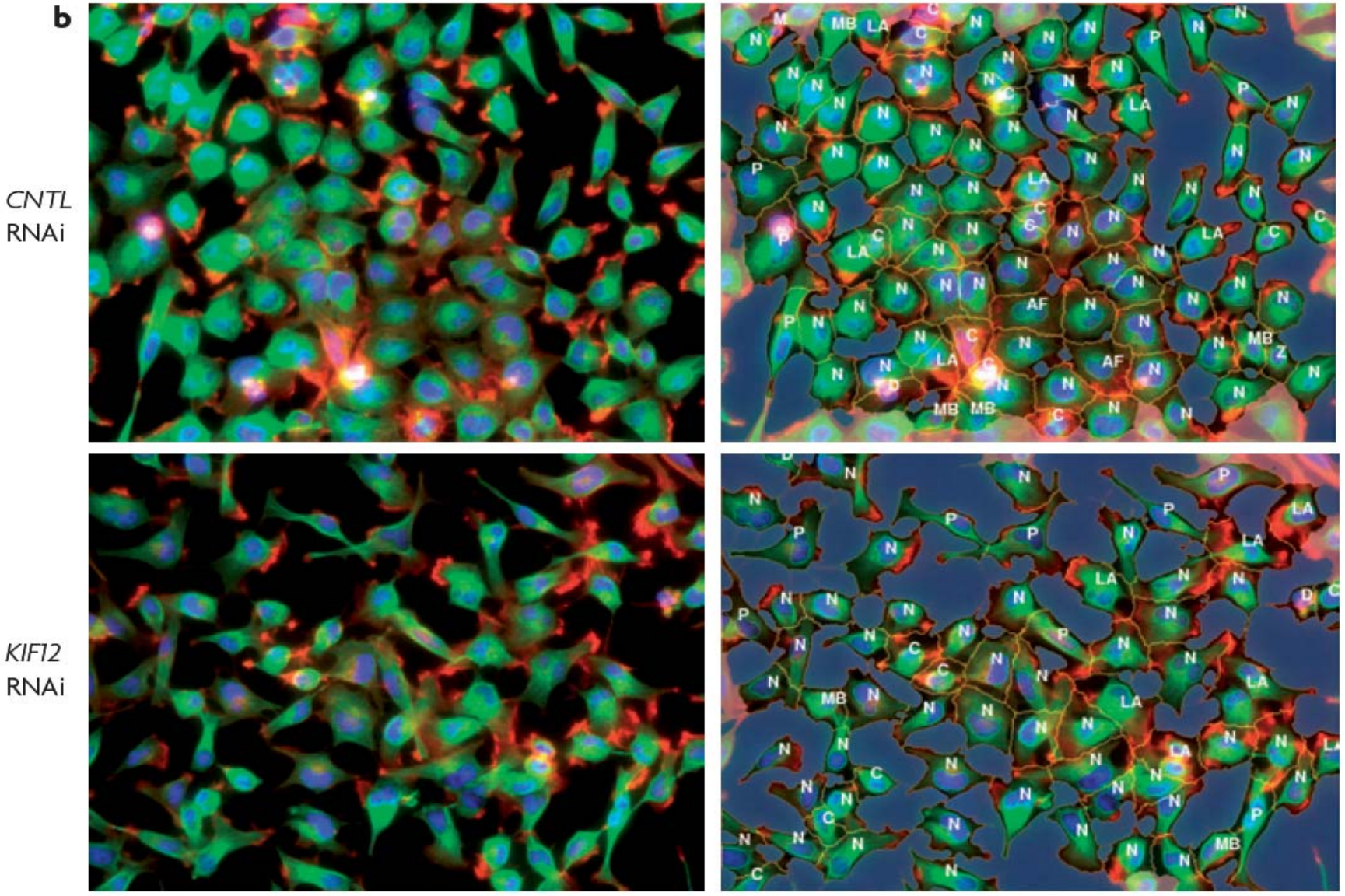
Phenotype: **organism** morphology



Whole organism: planaria

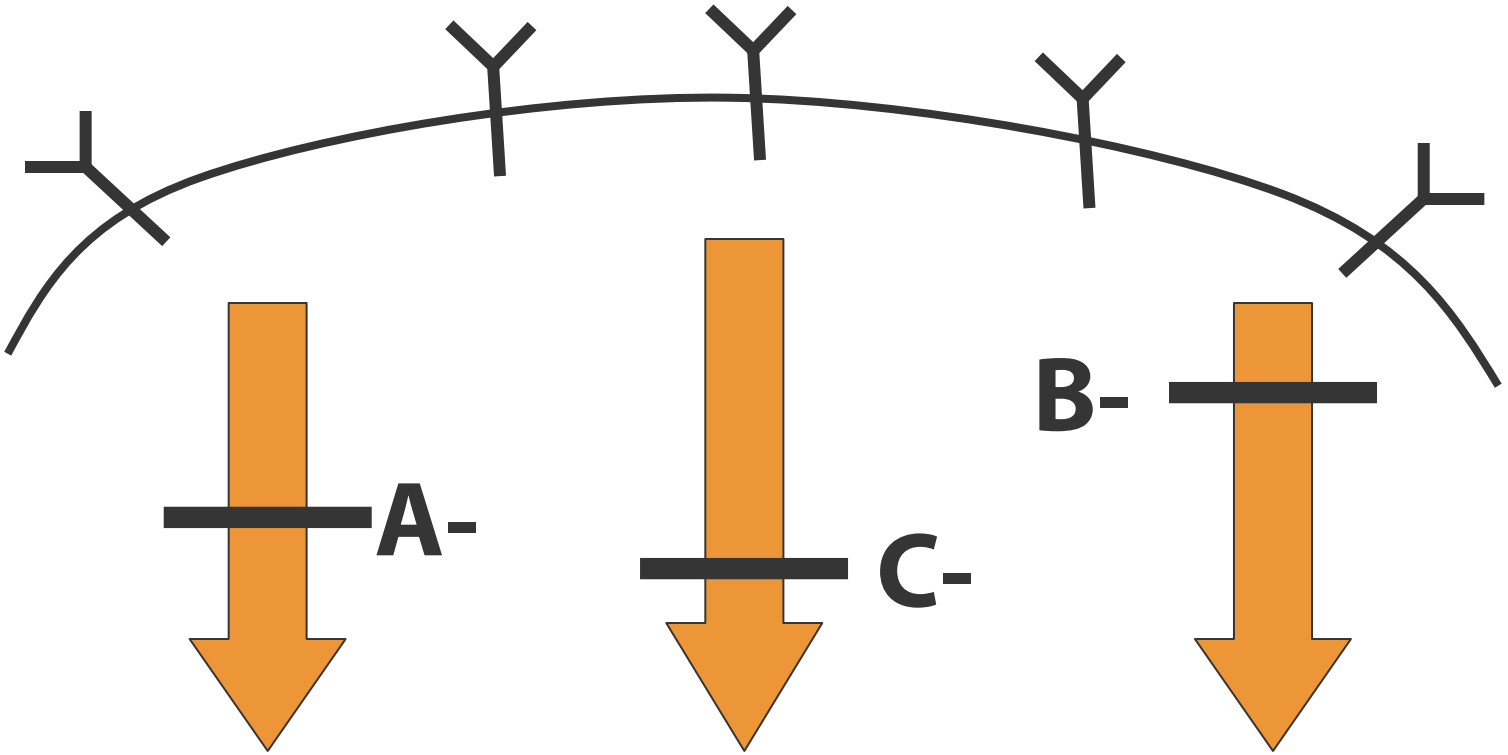
Boutros and Ahringer, Nat Rev 2008

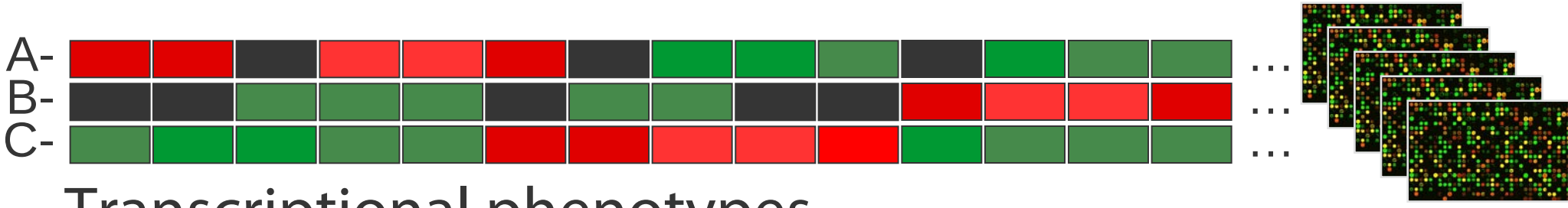
Phenotype: **cell** morphology



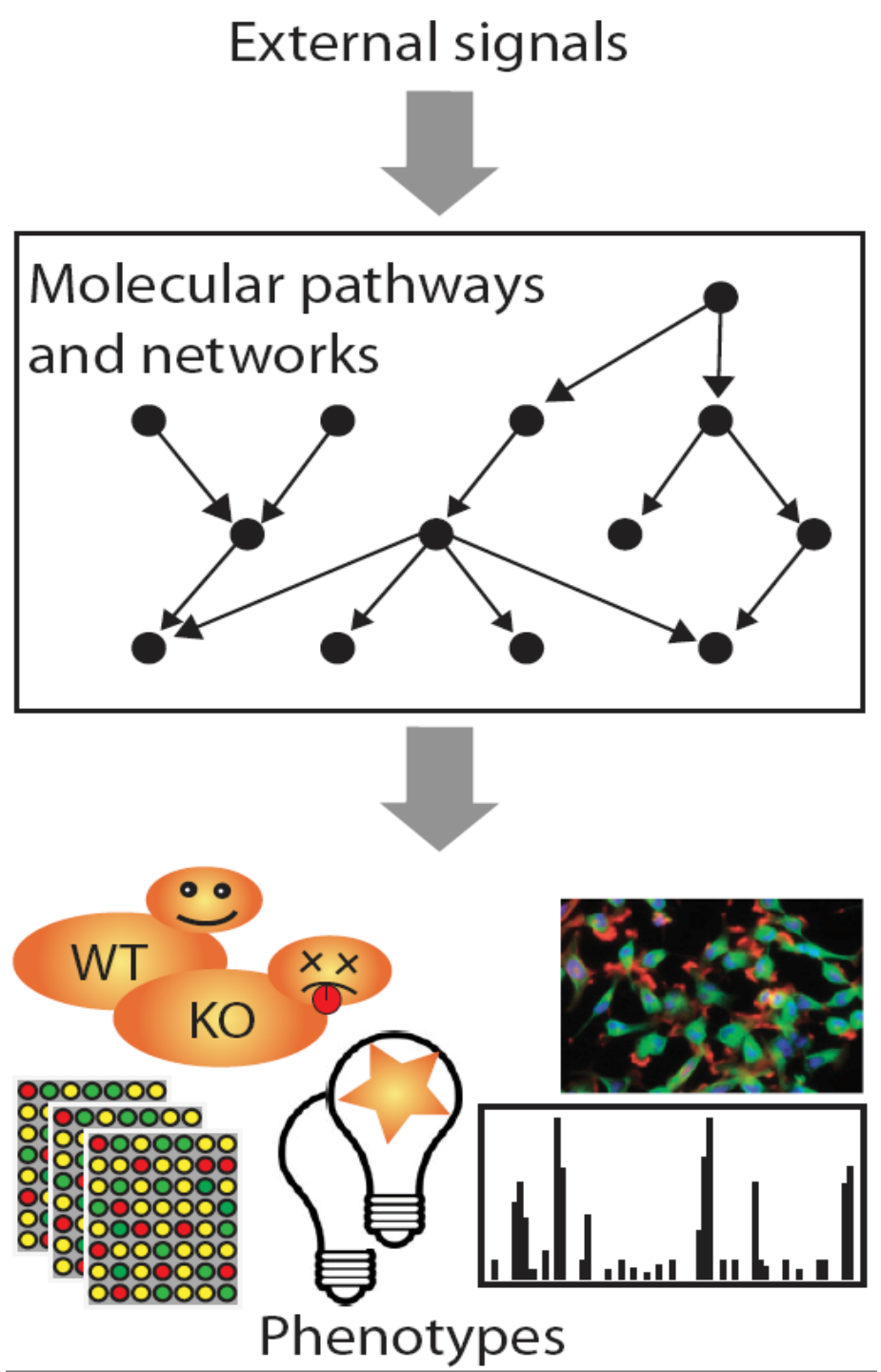
Boutros and Ahringer, Nat Rev 2008

Phenotype: **global gene expression**

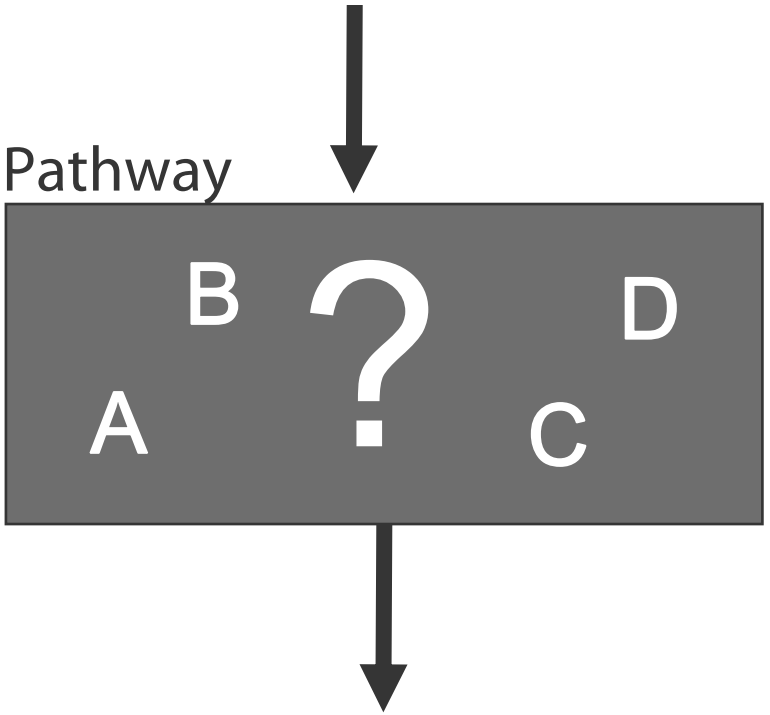




Transcriptional phenotypes


External signals
Molecular pathways
and networks
WT
KO
Phenotypes

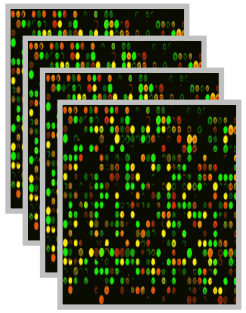
Network reconstruction from phenotypes



"Classical" approach

Corr
models:
Nets
GGMs
Mutual Information

no visible changes

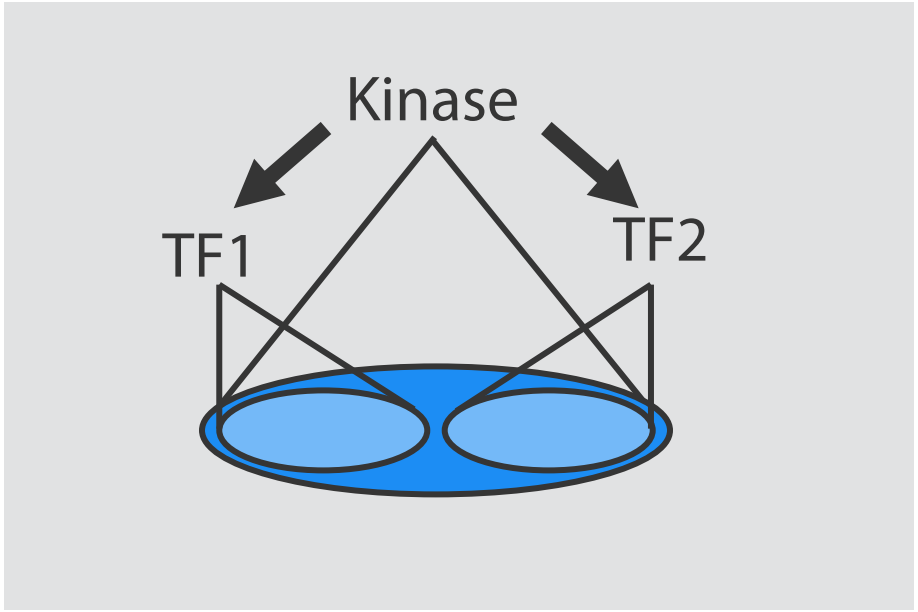


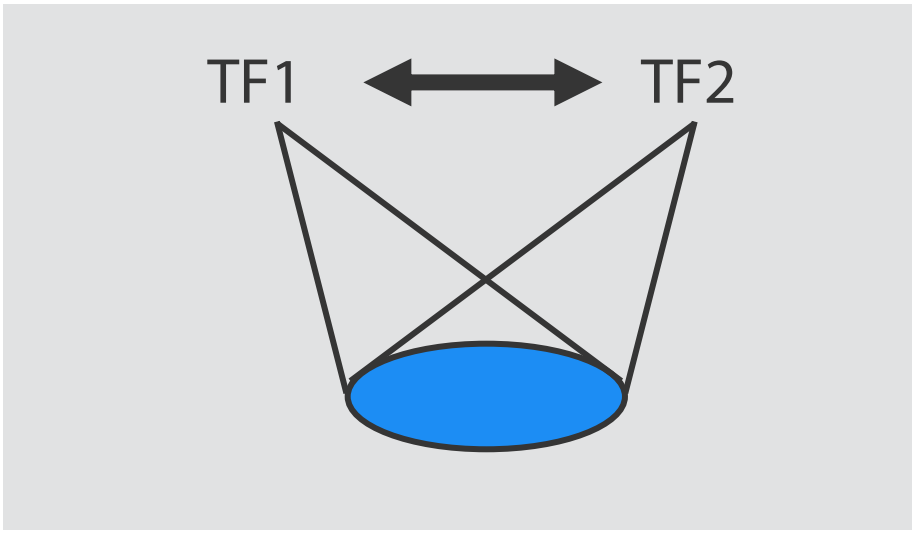
1. Correlation and clustering
2. Exploit **additional structure** in the data

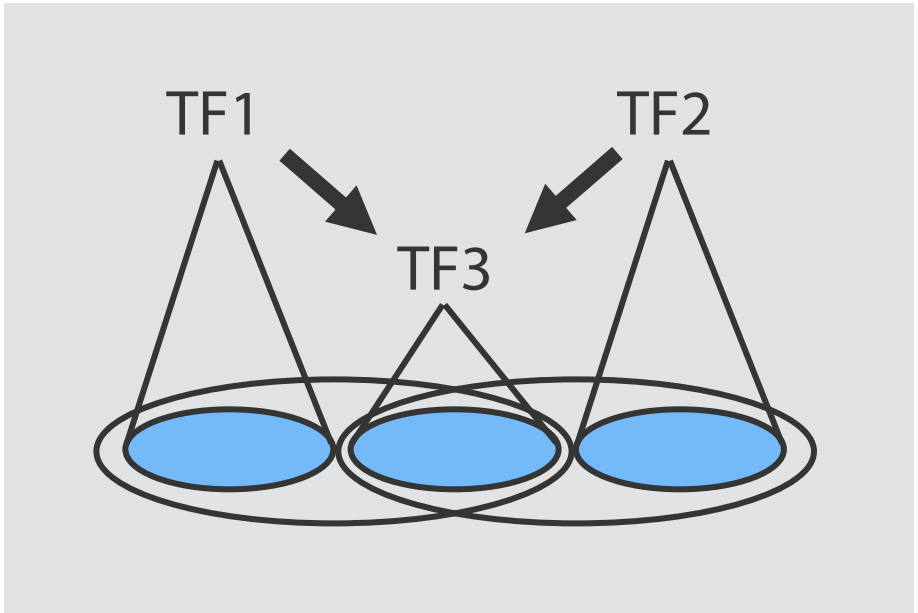
=> subset relations

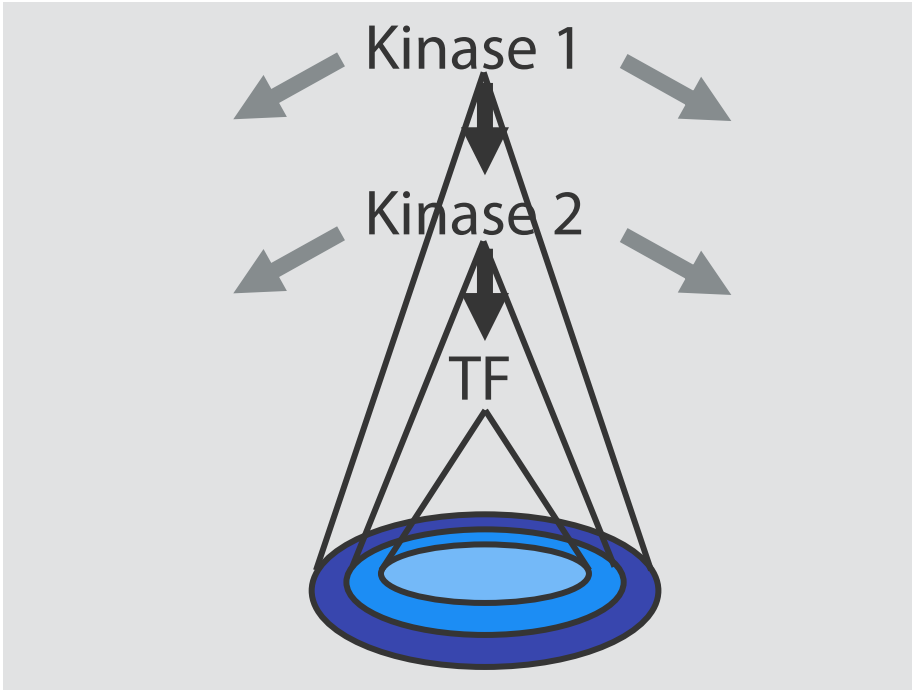
Nested Effects Models

Markowetz et al 2005, 2007
Tresch and Markowetz 2008



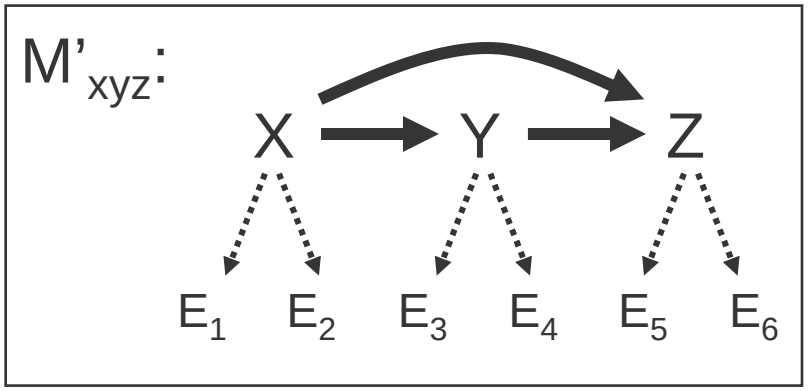


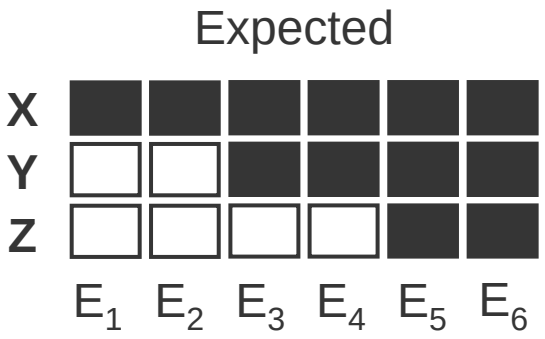


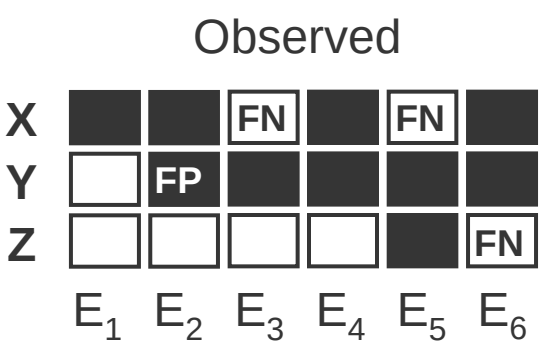


Nested Effects Models

Markowetz et al 2005, 2007
Tresch and Markowetz 2008







Pathway genes: X, Y, Z
- core topology
- to be reconstructed

  = Model M

Effect reporters: $E_1, \ldots, E_6$
- states are observed

  = Data D
- positions in pathway unknown

  = Parameters θ

Marginal likelihood

Posterior: $P(M \mid D) = 1/Z \cdot P(D \mid M) \cdot P(M)$

Marginal likelihood

$$P(D \mid M) = \int P(D \mid M, \Theta) P(\Theta \mid M) d\Theta$$

$$= \frac{1}{n^m} \prod_i \sum_j \prod_k P(e_{ik} \mid M, \theta_i = j)$$

Uniform prior over positions

Product over **all** effect reporters

Average over **possible positions** in the pathway

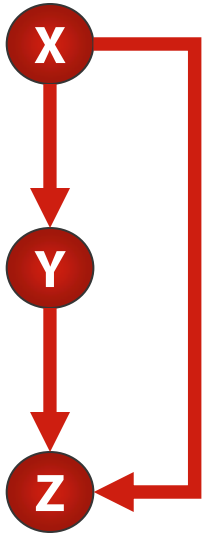
Product over **replicate** observation

Distribution of **single** effect reporter with **known** position

Markowetz et al, 2005

NEM model space

Subset relation =
reflexive and **transitive**
= quasi-order



NEM model space
= transitively closed directed graphs

n	a(n)
0	1
1	1
2	4
3	29
4	355
5	6942
6	209527
7	9535241
8	642779354
9	63260289423
10	8977053873043
11	1816846038736192
12	519355571065774021
13	207881393656668953041
14	115617051977054267807460
15	88736269118586244492485121
16	93411113411710039565210494095
17	134137950093337880672321868725846
18	261492535743634374805066126901117203

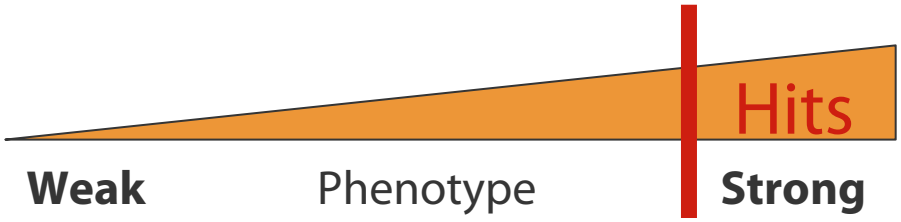
Exhaustive enumeration not feasible

http://www.research.att.com/~njas/sequences/
Online encyclopedia of integer sequences

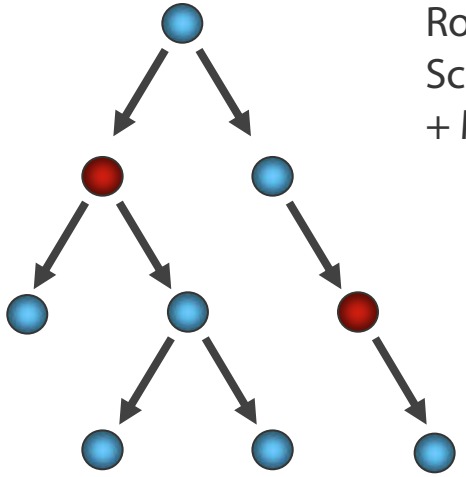
Anatomy of the NFκB pathway

Roland Schwarz
+ MPI IB Berlin

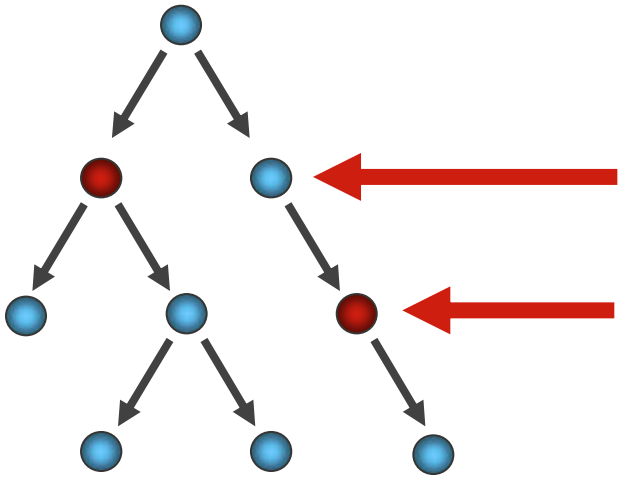
Step 1



?



Step 2



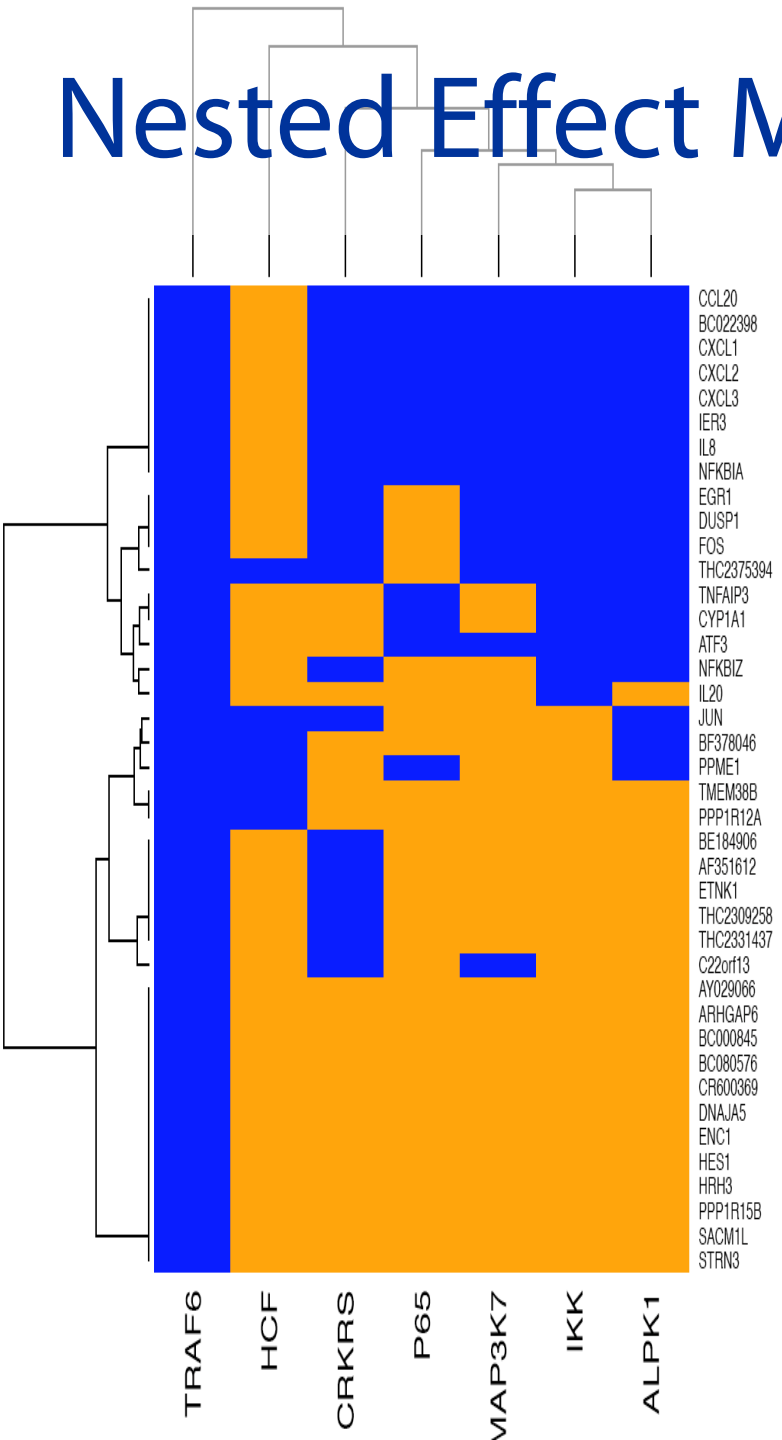
Knock-down

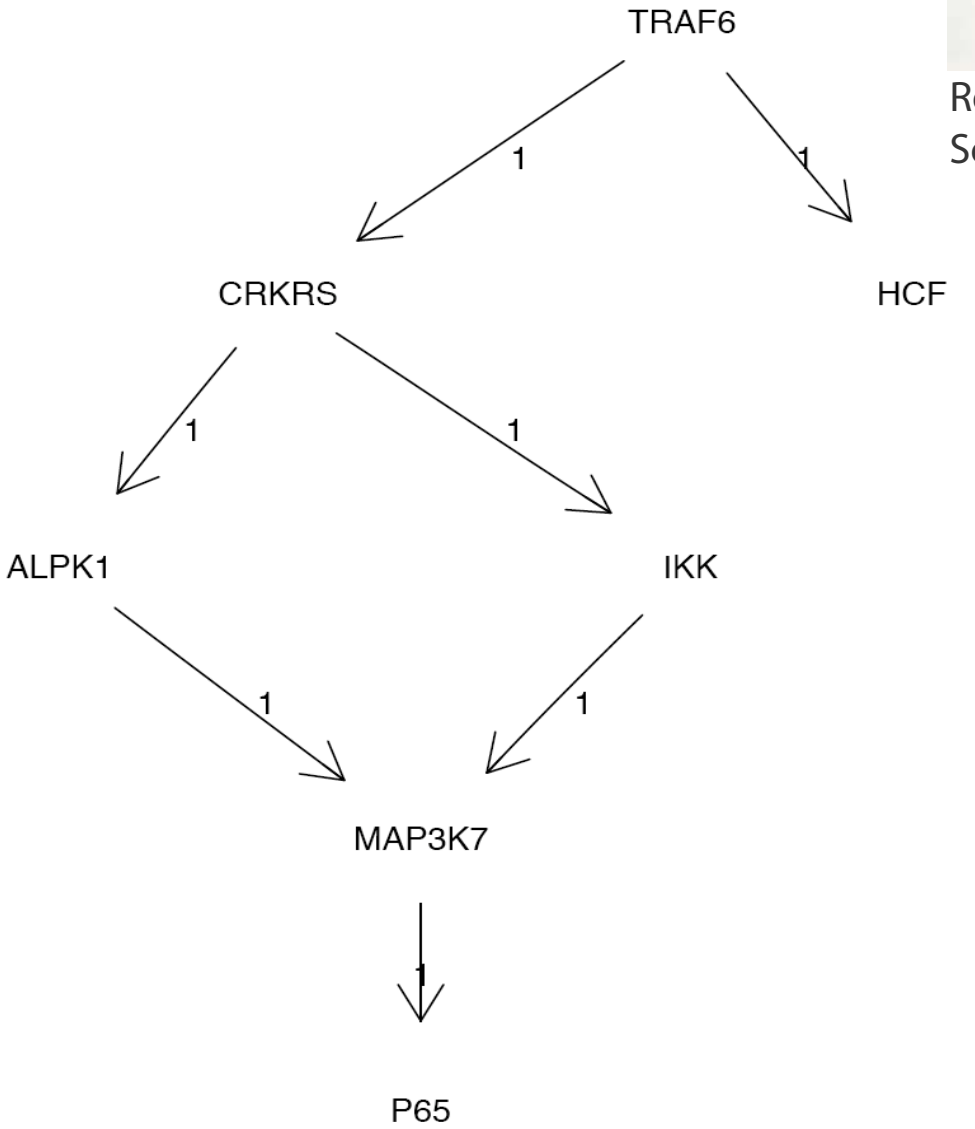
Known pathway members

New RNAi Hits

Compare expression phenotypes by **NEMs**

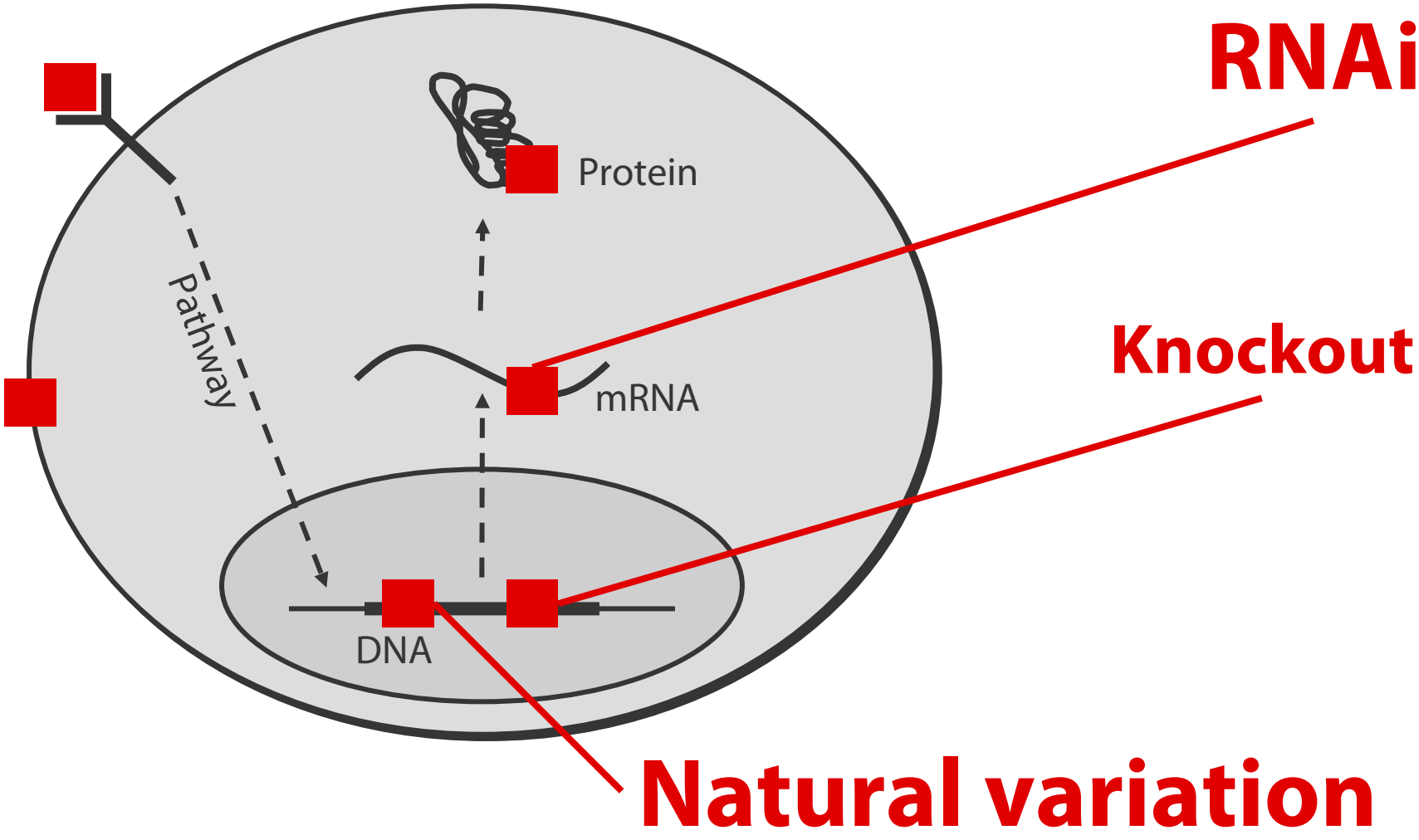
Nested Effect Models for NFκB

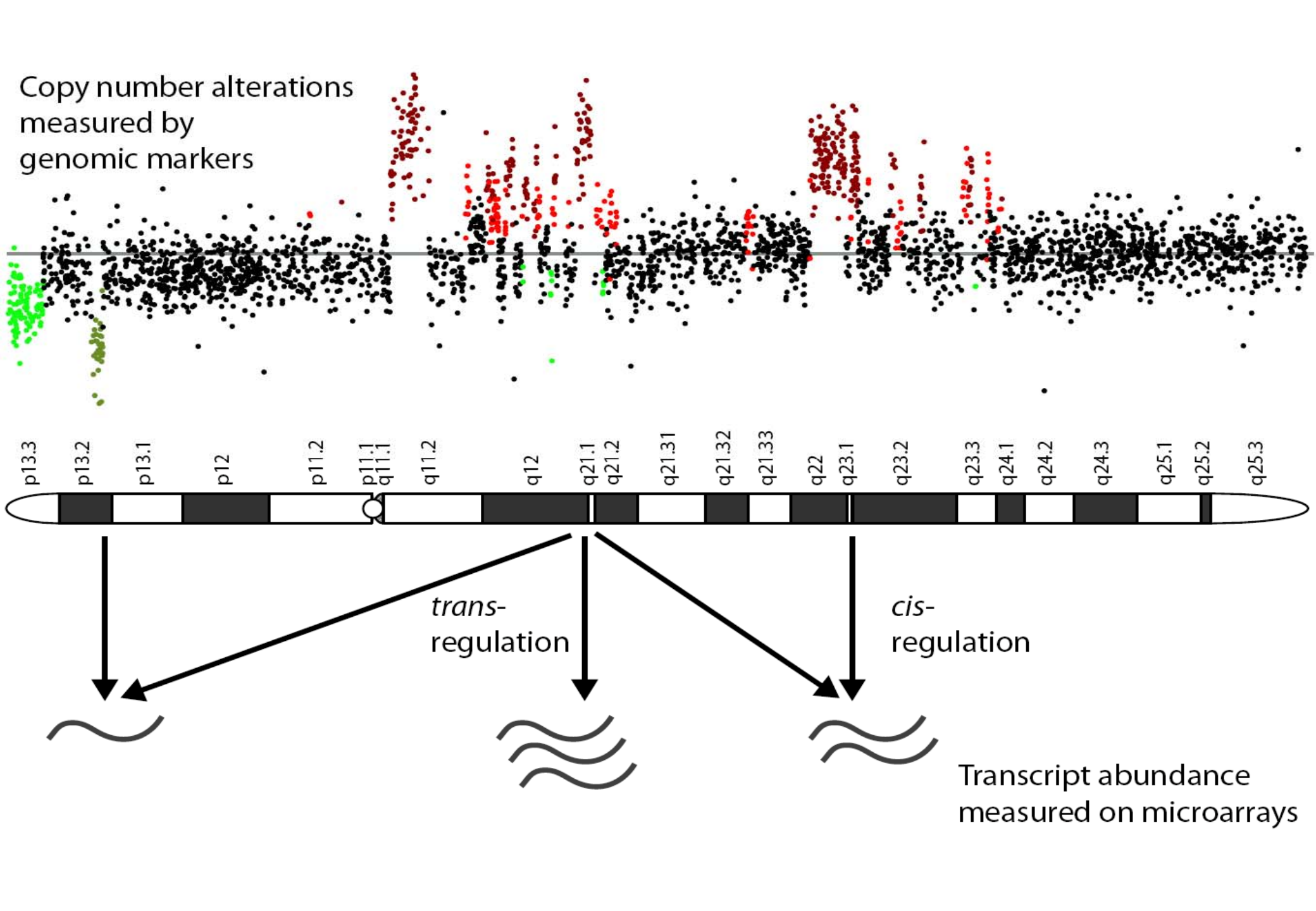




Roland Schwarz

Natural experiments




Copy number alterations
measured by
genomic markers
p13.3
p13.2
p13.1
p12
p11.2
p11.1
q11.1
q11.2
q12
q21.1
q21.2
q21.31
q21.32
q21.33
q22
q23.1
q23.2
q23.3
q24.1
q24.2
q24.3
q25.1
q25.2
q25.3
trans-
regulation
cis-
regulation
Transcript abundance
measured on microarrays

The **METABRIC** project

With data from **>1000 tumours**

1. describe the **genomic landscape** of breast cancer
   - **chromosomal alterations**, allelic ratios, breakpoints, genomic instability, mutations in oncogenes, **gene expression**, …
2. correlate **molecular** profiles with **clinical** outcome
   - to find **predictive** markers for eg. survival
   - to define molecular **subsets** of tumours with unique clinical phenotypes

Impact of CNA on expression

Yinyin Yuan

Global: Which transcriptional changes are copy-number dependent?

Local: for each copy-number dependent gene, which particular genomic loci have most influence on its expression?

Gene-expression

$$y_i = y_i$$ Baseline

$$y_i = x_i\beta_i + \epsilon_i$$ *cis*-effect

Copy-number

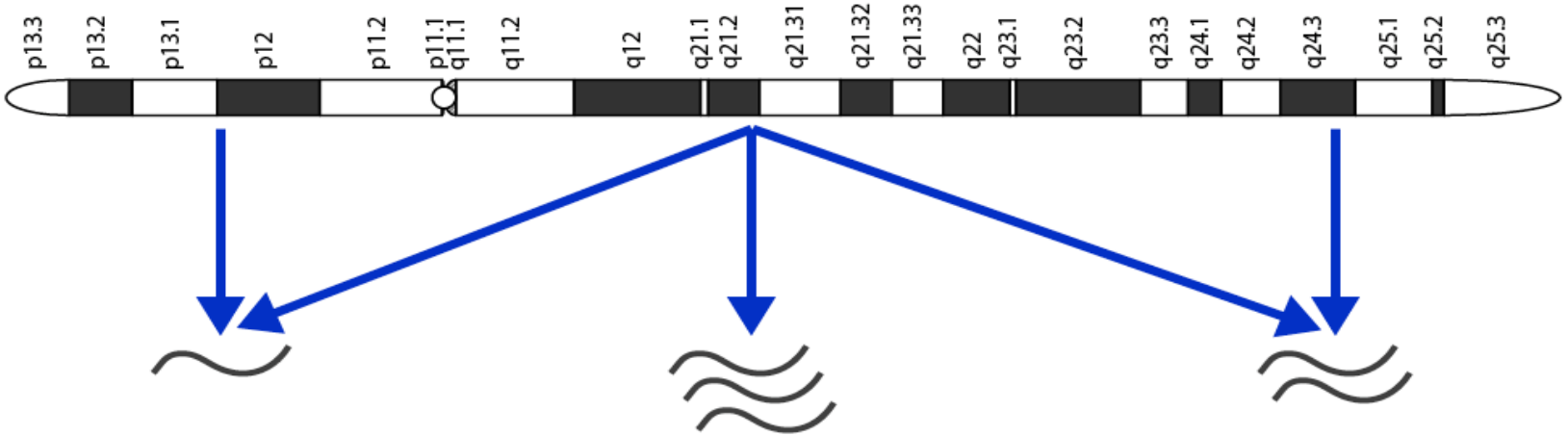
$$y_i = x_i\beta_i + \sum_{j, j \neq i} x_j\beta_j + \epsilon_i'$$ cis- and trans-effects

$$\text{score}_i = -\ln\left(\frac{\sigma^2_{with}}{\sigma^2_{without}}\right)$$

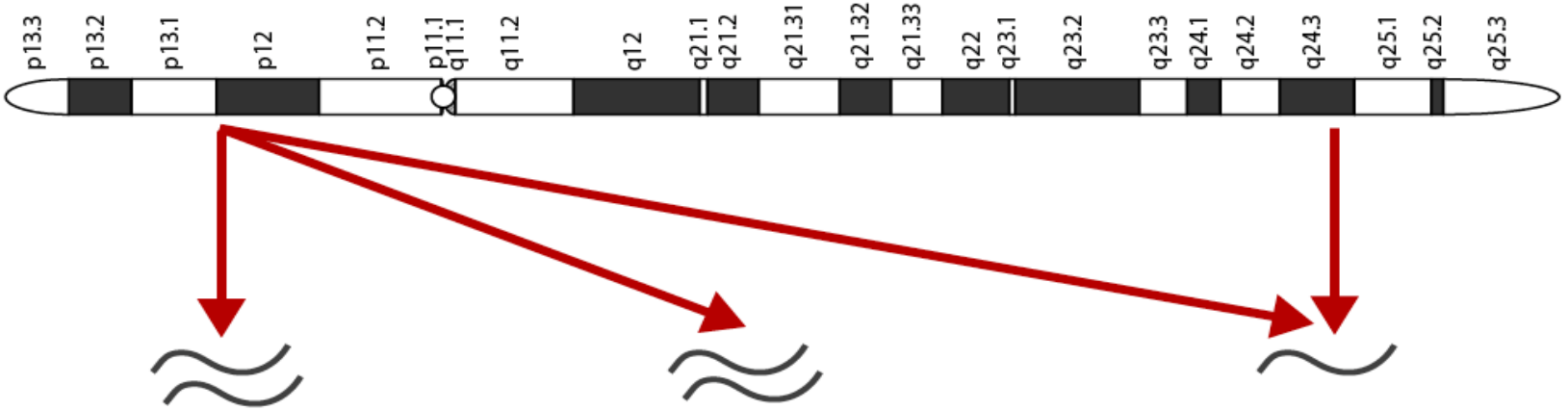
Differential regulation

Yinyin Yuan

Subtype A, eg ER+ breast cancer



Subtype B, eg ER- breast cancer



Differential regulation

Yinyin Yuan

Gene Expression

Copy-number

$$Y_1 = X_1 B^r + X_1 B^d + \epsilon_1$$

$$Y_2 = X_2 B^r + \epsilon_2$$

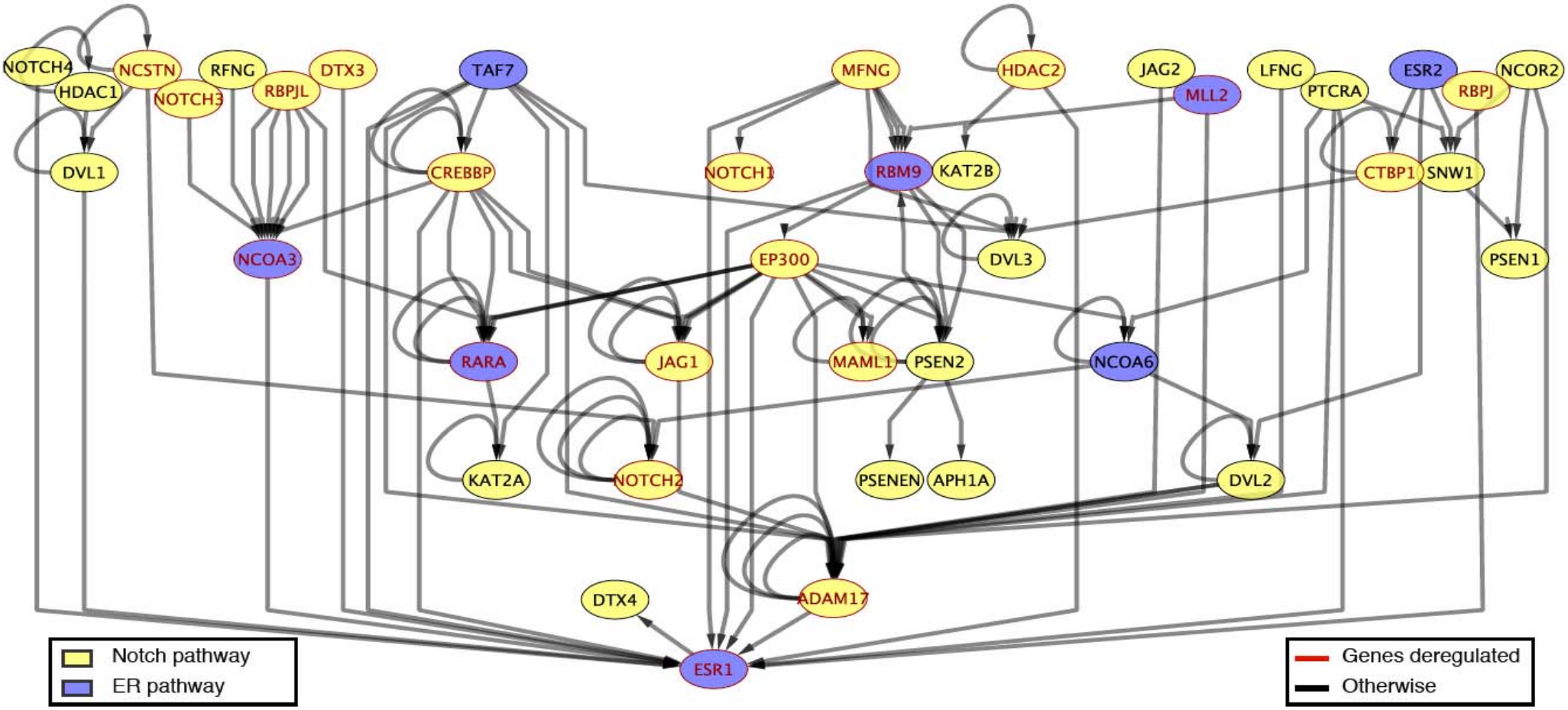
Differential "network"

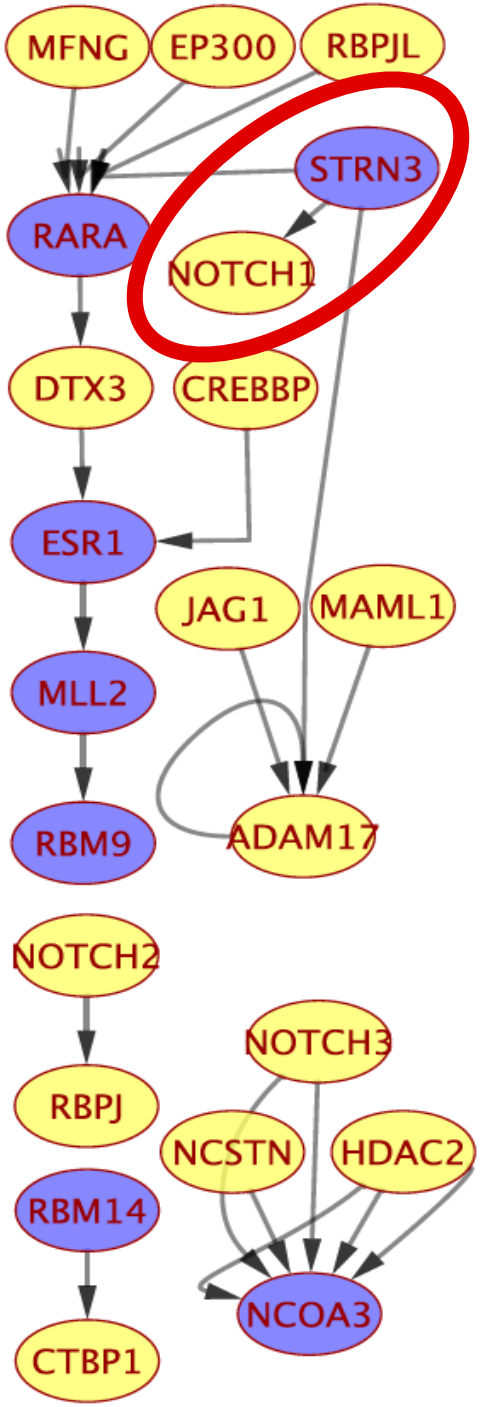
Reference "network"

$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} X_1 & X_1 \\ X_2 & 0 \end{bmatrix} \begin{bmatrix} B^r \\ B^d \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \end{bmatrix}$$

solved by Lasso

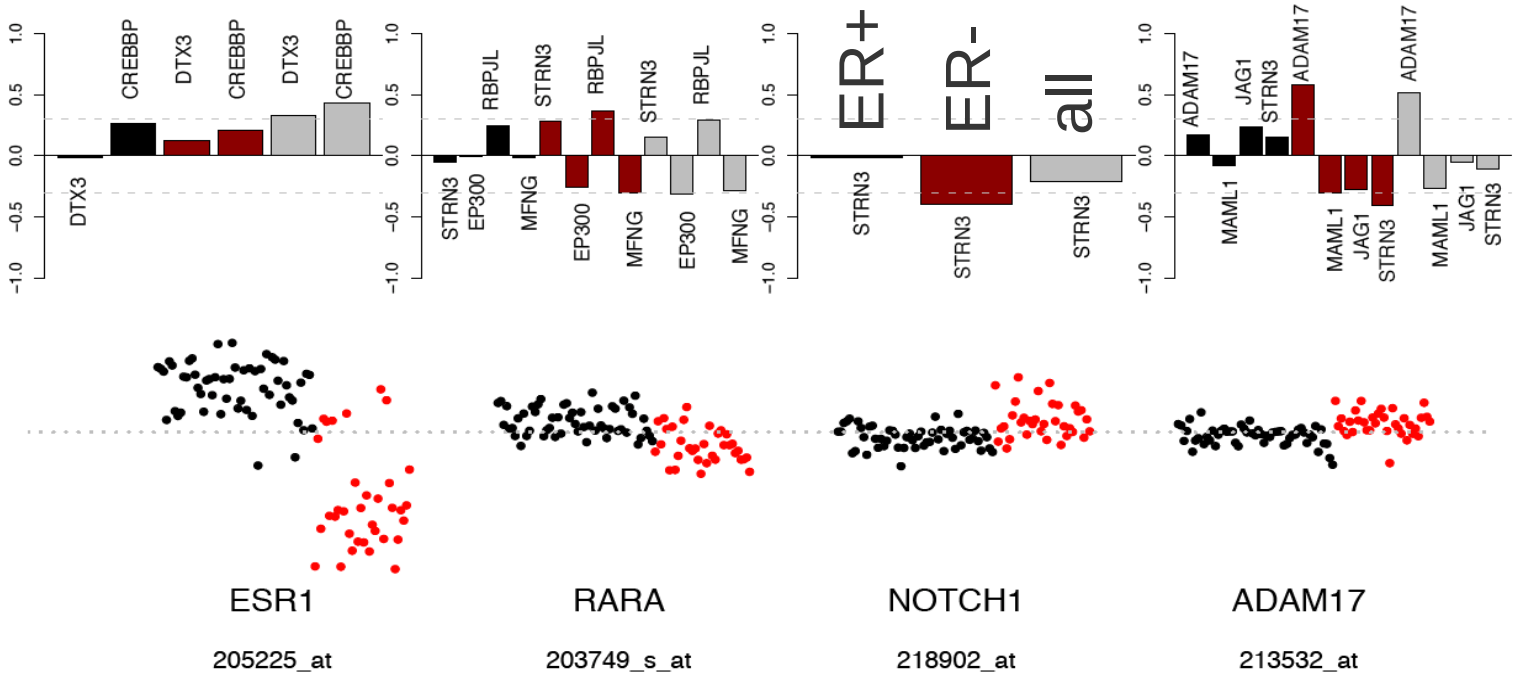
Reference network (ER+/-)





Differential network (ER-)

Copy-number changes
at regulating loci



Gene expression

Summary

1. Gene perturbation screens

- Nested Effects Models reconstruct pathways from **nested structure** of downstream effects
- Application in NFkB

2. Breast cancer genomics

- Metabric: the genomic landscape of breast cancer
- Copy-number alterations => Gene expression
- Regression models to identify **differential regulation** in cancer sub-types

Future plans

- **Nested Effects Models**
  - **dynamic models**: infer pathway from phenotypes observed over time
  - **re-wiring** of network over time
- **Breast cancer genomics**
  - **Stratification** into disease sub-populations
  - **Predict clinical outcome** by CNV and Expression and others

Xin Wang

Yinyin Yuan

OPEN ACCESS Freely available online

PLoS COMPUTATIONAL BIOLOGY

Education

PLoS Comp Bio, 2010 Feb 26;6(2):e1000655

How to Understand the Cell by Breaking It: Network Analysis of Gene Perturbation Screens

Florian Markowetz*

Cancer Research UK Cambridge Research Institute, Cambridge, United Kingdom

Introduction

Functional genomics has demonstrated considerable success in inferring the inner working of a cell through analysis of its response to various perturbations. In recent years several technological advances have pushed gene perturbation screens to the forefront of functional genomics. Most importantly, modern technologies make it possible to probe gene function on a genome-wide scale in many model organisms and human. For example, large collections of knock-out mutants play a prominent role in the study of *Saccharomyces cerevisiae* [1], and RNA interference (RNAi) and survival of cancer cell lines are also the least studied [5].

A goal becoming more and more prominent in both experimental as well as computational research is to leverage gene perturbation screens to the identification of molecular interactions, cellular pathways, and regulatory mechanisms. Research focus is shifting from understanding the phenotypes of single proteins to understanding how proteins fulfill their function, what other proteins they interact with, and where they act in a pathway. Novel ideas on how to use perturbation screens to uncover cellular wiring diagrams can lead to a better understanding of how cellular networks are deregulated in diseases like cancer. This knowledge is indispensable for finding new drug targets to attack the drivers of a disease and not only the symptoms.

This review surveys the current state-of-

activity of reporter constructs, e.g., a luciferase, downstream of a pathway of interest [9]. Low-dimensional phenotyping screens can identify candidate genes on a genome-wide scale and are often used as a first step for follow-up analysis. We will discuss methods to functionally interpret hits from low-dimensional phenotyping screens and to place them in the context of cellular networks in the first part of this review.

The second part will be devoted to high-dimensional phenotyping screens, which evaluate a large number of cellular features at the same time. Observing system-wide changes promises key insights into cellular mechanisms and pathways that can not be supplied by low-dimensional screens. For example, high-dimensional phenotypes can include changes in cell morphology [10–13], or growth rates under a wide range of conditions [14], or

Acknowledgements

- **NFkB:** Meyer lab at MPI for Infection Biology Berlin, especially Andre Maurer and Cindy Rechner
- **Breast cancer:** Carlos Caldas, Christina Curtis, and the rest of METABRIC
- **My group at CRI:**

Florian Roland Xin Yinyin Henrik

Reconstructing networks from **experimental** and **natural** genetic **perturbations**

Thank you !

Florian Markowetz
markowetzlab.org

CANCER RESEARCH UK Cambridge Research Institute