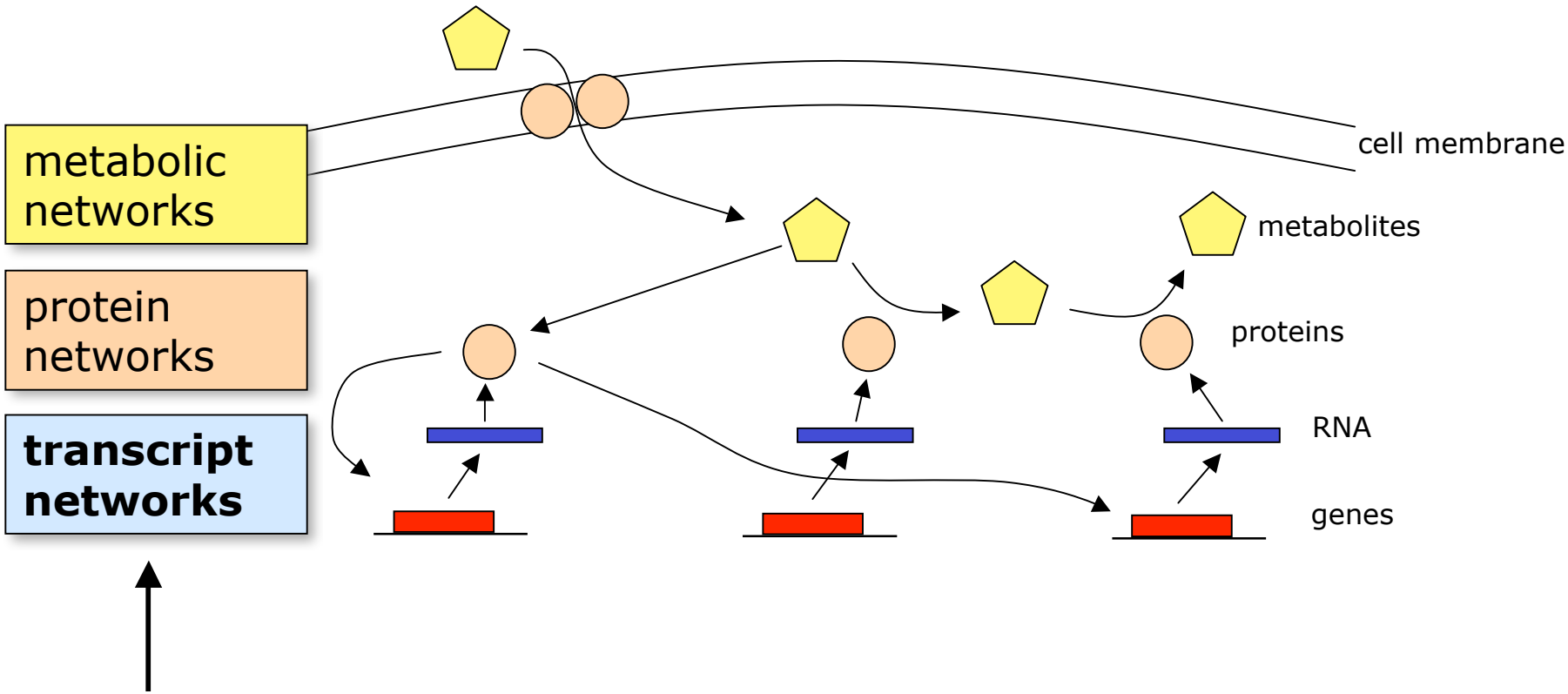
Gene network inference in diseases and drug discovery

Diego di Bernardo
Telethon Institute of Genetics and Medicine

19 Dicembre 2007, Napoli

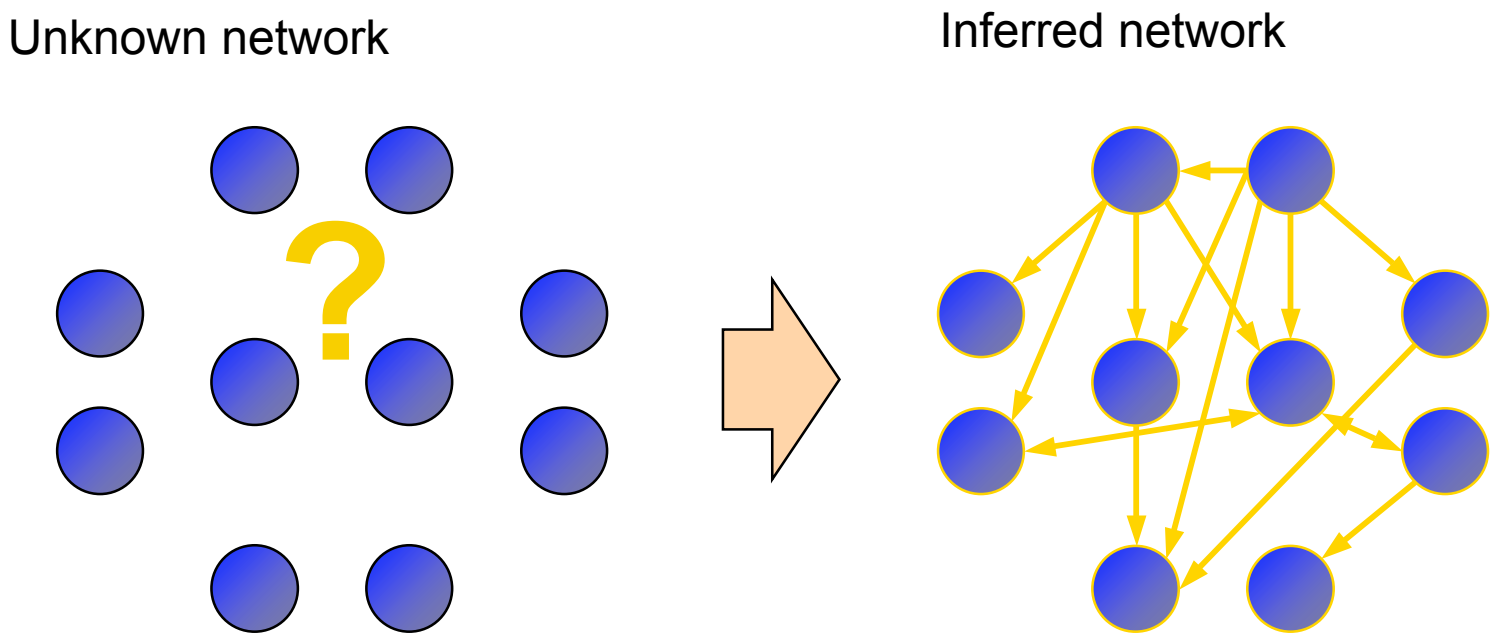
TIGEM
TELETHON INSTITUTE OF GENETICS AND MEDICINE

Gene Networks

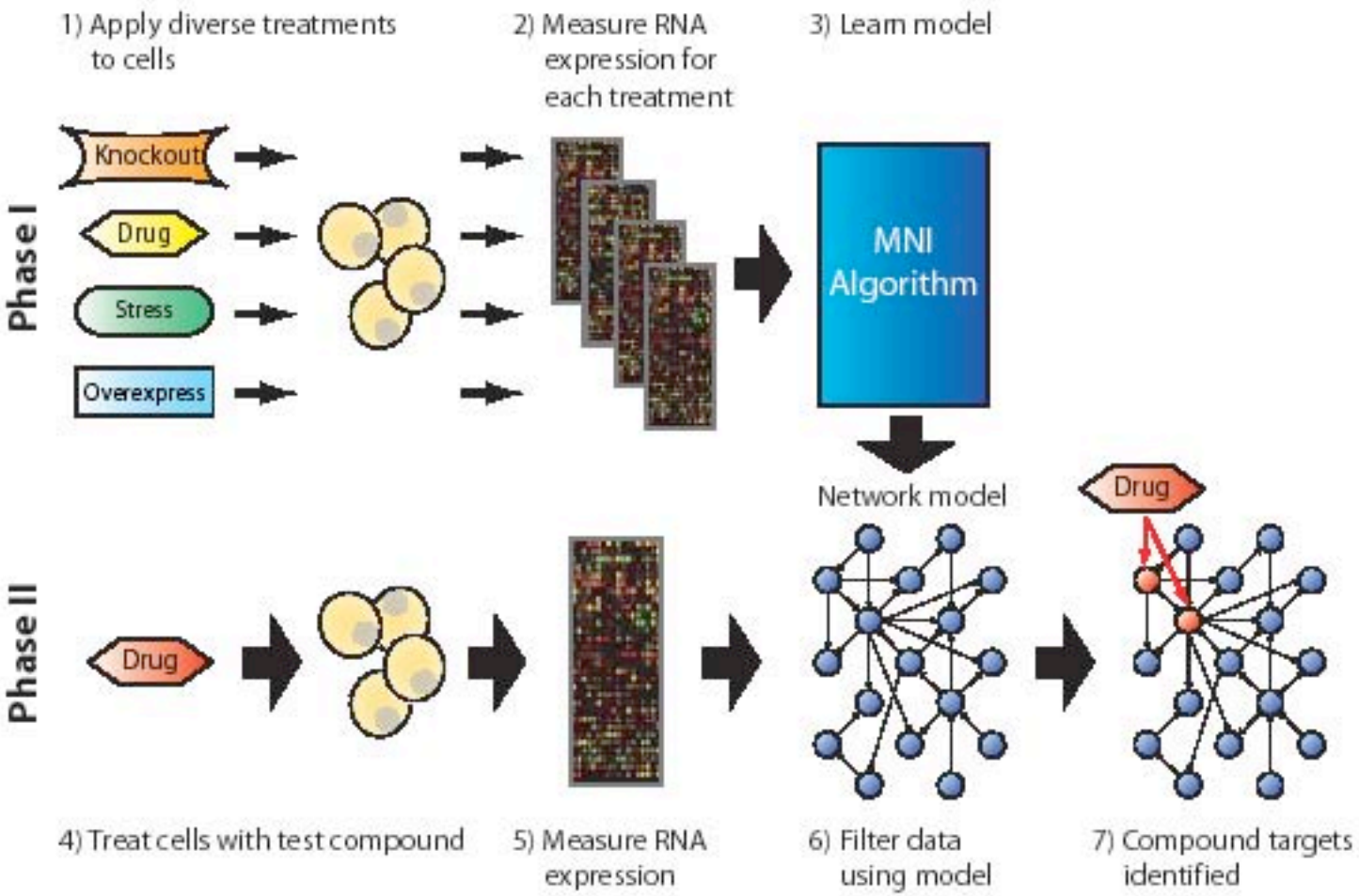


Our focus: methods to decode transcription regulation networks

Reverse engineering (or system id) gene networks:



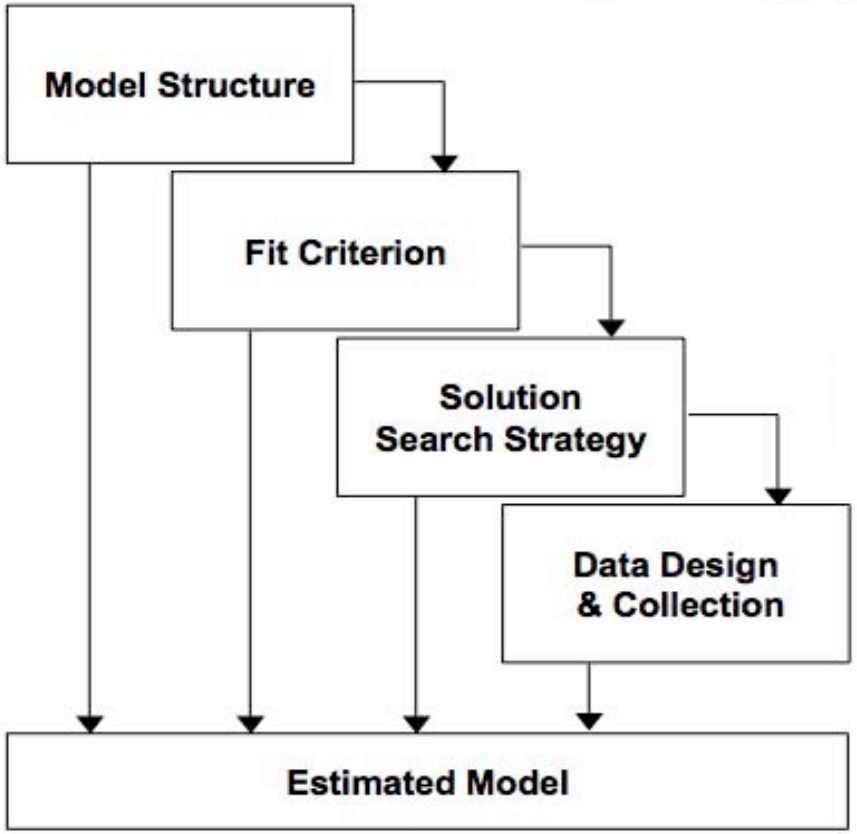
Aim of reverse engineering: gene function and drug MOA



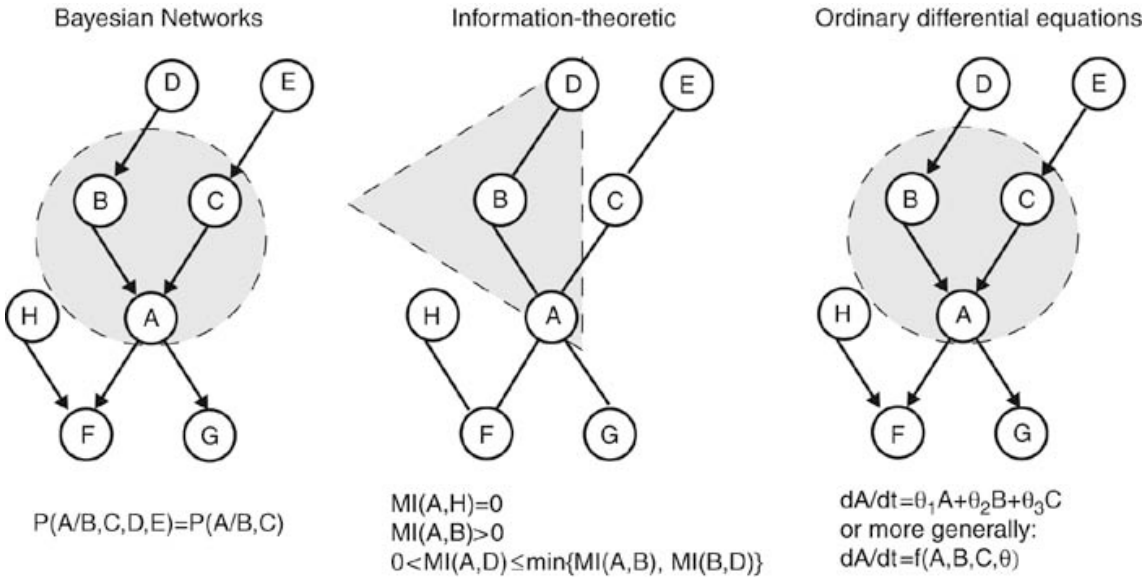
D di Bernardo et al, Nature Biotechnolgy, 2005; TS Gardner, D di Bernardo et al, Science, 2003;

Reverse-engineering strategy:

- Choose a model
- Choose a fit criterion (cost function) to measure the fit of the model to the data
- Define a strategy to find the parameters that best fit the data (i.e. that minimise cost function)
- Perform appropriate experiments to collect the experimental data:



Comparision of different models



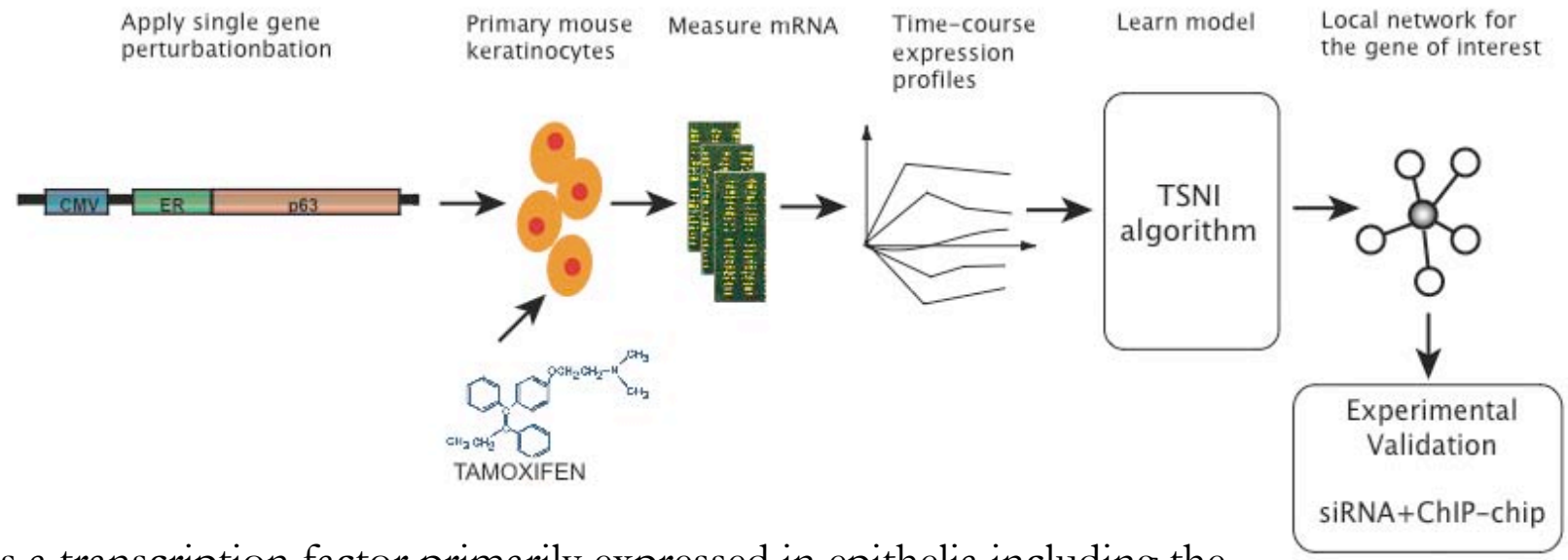
Bayesian network (Hartemink, A. (2005) Nature Biotechnology, 2005.)

Information-theoretic: (Basso et al., Nature Genetics, 2006)

ODEs (Gardner, di Bernardo et al, Science, 2003; di Bernardo et al, Nature Biotechnolgy, 2005)

Mukesh Bansal, Vincenzo Belcastro, Alberto Ambesi, Diego di Bernardo. How to infer gene networks from expression profiles. Molecular Systems Biology, 2007.

Understanding gene function in a genetic disease:

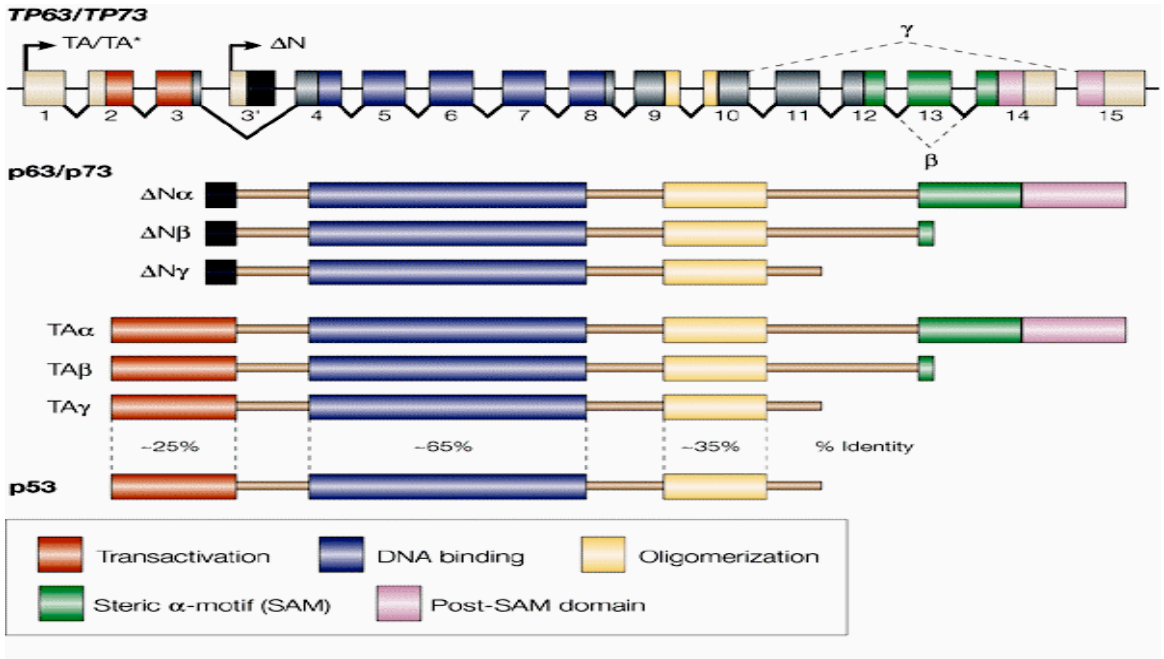


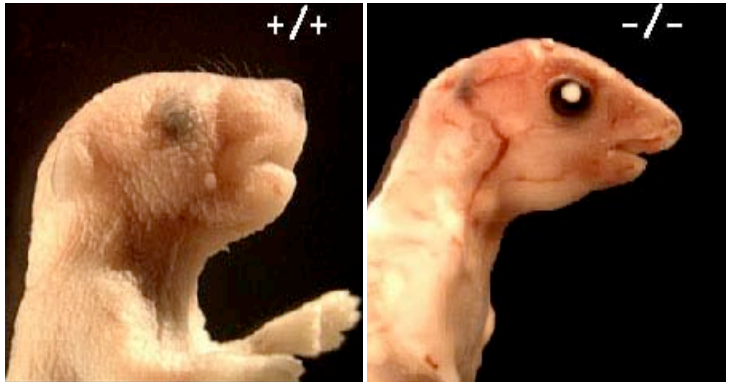
P63 is a transcription factor primarily expressed in epithelia including the proliferative compartments of the skin.

It plays an essential role in modulating cellular differentiation by unknown mechanisms.

Mutation in its DNA binding and SAM domains have been linked to five human malformation syndromes.

The transcription factor p63:

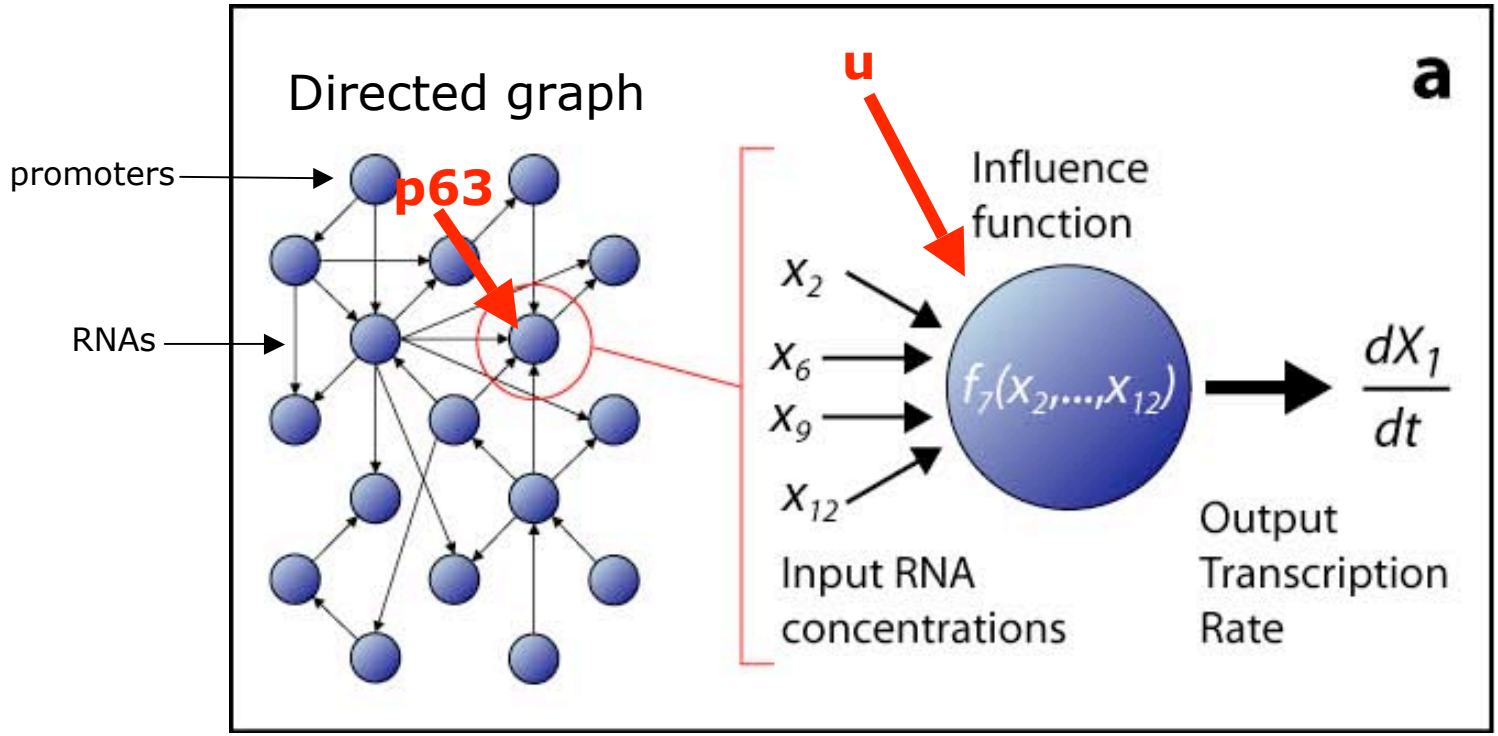




- P63 is a Transcription Factor (unknown targets)
- p63 gene has three different promoter that gives rise to two different transcripts (TA and ΔN);
- each transcript has three alternative splicing isoforms (α,β,γ)

- Lack of epidermis and hair follicles
- Defects in all squamous epithelia
- Limb truncations
- Craniofacial defects: lack of tooth primordia and eyelids
- Maxilla and mandible are truncated and secondary palate fails to close

Model structure: Ordinary differential equations



$$dX_1/dt = \mathbf{a_2}\, X_2 + \mathbf{a_6}\, X_6 + \mathbf{a_9}\, X_9 + \mathbf{a_{12}}\, X_{12} + \mathbf{b}_1 u$$

Model structure:

$$\begin{cases} x'_1(t) = a_{11}x_1(t) + a_{12}x_2(t) + \ldots + a_{1n}x_n(t) + b_1u(t) \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ x'_n(t) = a_{n1}x_1(t) + a_{n2}x_2(t) + \ldots + a_{nn}x_n(t) + b_nu(t) \end{cases}$$

Gene effect unknown

Model structure - Discrete model:

$$x_1(t_{k+1}) = a^d_{11}x_1(t_k) + a^d_{12}x_2(t_k) + \ldots + a^d_{1n}x_n(t_k) + b^d_1 u(t_k)$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$x_n(t_{k+1}) = a^d_{n1}x_1(t_k) + a^d_{n2}x_2(t_k) + \ldots + a^d_{nn}x_n(t_k) + b^d_n u(t_k)$$

Or in matrix format:

$$\mathbf{x}(t_{k+1}) = \mathbf{A}^d\mathbf{x}(t_k) + \mathbf{b}^d u(t_k)$$

For k=1..M

$$\mathbf{X}(t_{k+1}) = \mathbf{A}^d\mathbf{X}(t_k) + \mathbf{b}^d\mathbf{u}(t_k)$$

Where (NxM)=(NxN)(NxM)+(Nx1)(1xM)

Fit criterion (Least Squared Error):

$$\dot{X}(t_k) = AX(t_k) + BU(t_k) \quad k = 1 \ldots M$$

$$X(t_{k+1}) = A_d * X(t_k) + B_d * U(t_k)$$

$$X(t_{k+1}) = \begin{bmatrix} A_d & B_d \end{bmatrix} * \begin{bmatrix} X(t_k) \\ U(t_k) \end{bmatrix}$$

NxM = Nx(N+1) (N+1)xM

$$X(t_{k+1}) = H * Y(t_k)$$

$$B = (A_d + 1)^{-1} B_d$$

$$H = \begin{bmatrix} A_d & B_d \end{bmatrix}$$

$$Y(t_k) = \begin{bmatrix} X(t_k) \\ U(t_k) \end{bmatrix}$$

Solution search strategy:

- Usually N>>M (more unknown than data points; i.e. 1000 genes e 20 points)
- From the previous equation we know that a unique solution will exist only if $N \leq M-1$
- We need a trick to reduce number of unknowns by dimensional reduction:
  - Regression with Variable selection
  - Clustering
  - Regression with SVD

Solution search strategy: Regression using PCS/SVD

$$X(t_{k+1}) = H * Y(t_k)$$

$$H = \begin{bmatrix} A_d & B_d \end{bmatrix}$$

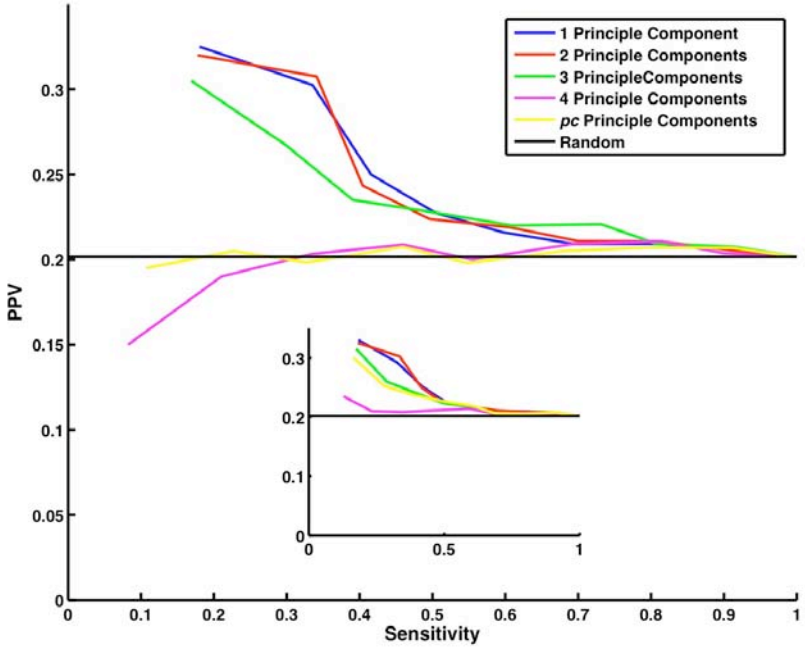
$$Y(t_k) = \begin{bmatrix} X(t_k) \\ U(t_k) \end{bmatrix}$$

- Singular Value Decomposition:
  - Y=UDV′ (N+1xM)=(N+1xM)(MxM)(MxM)

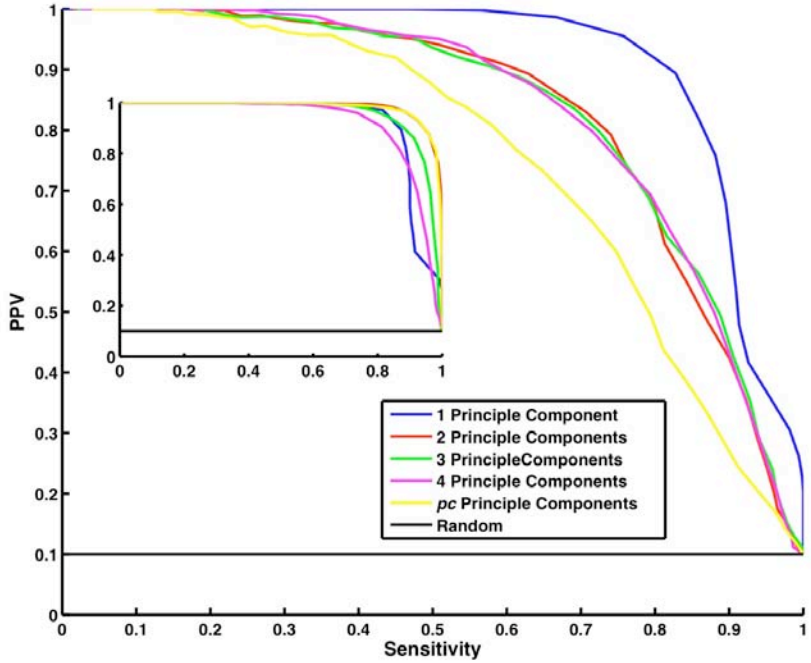
  - $Y=U_rD_rV_r'$ (N+1xM)=(N+1xK)(KxK)(KxM)
  - $X(t_{k+1})$ =HY= (NxN+1)(N+1xM) =
  - $X(t_{k+1})$ $=HU_rD_rV_r'=(HU_r)D_rV_r'=$**(NxK)(**KxK)(KxM)

Preliminary Results on noisy "*in silico*" data

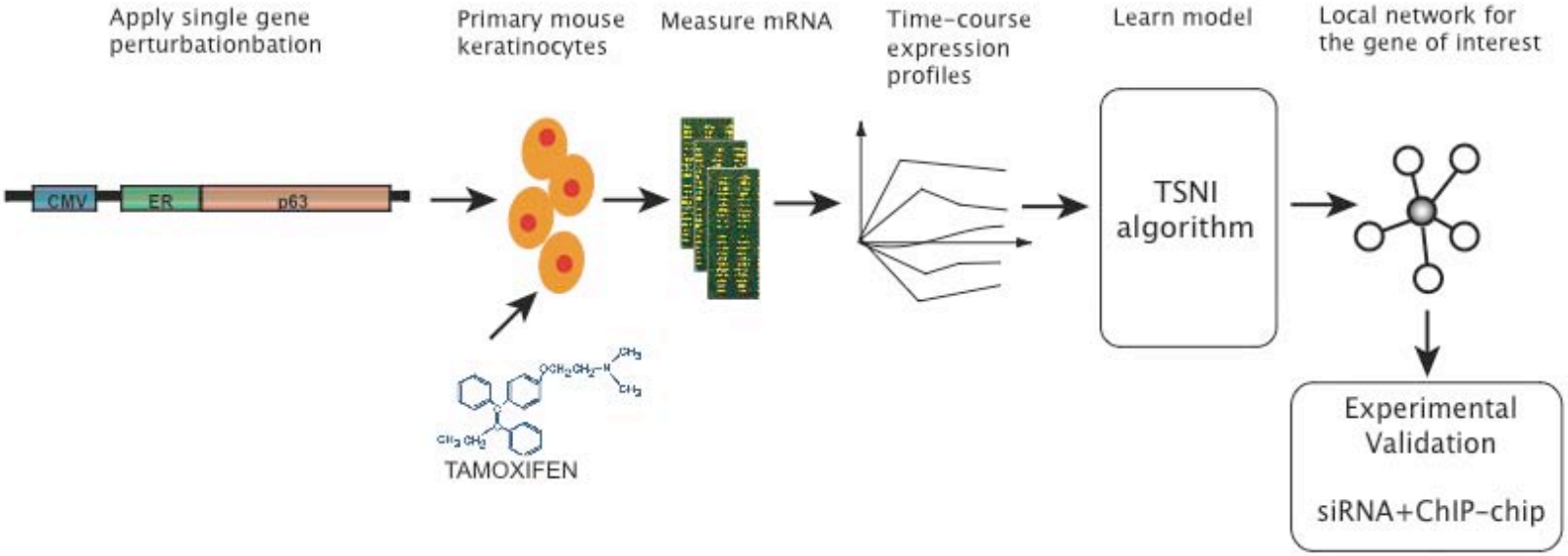
Network of 10 genes (**A**)



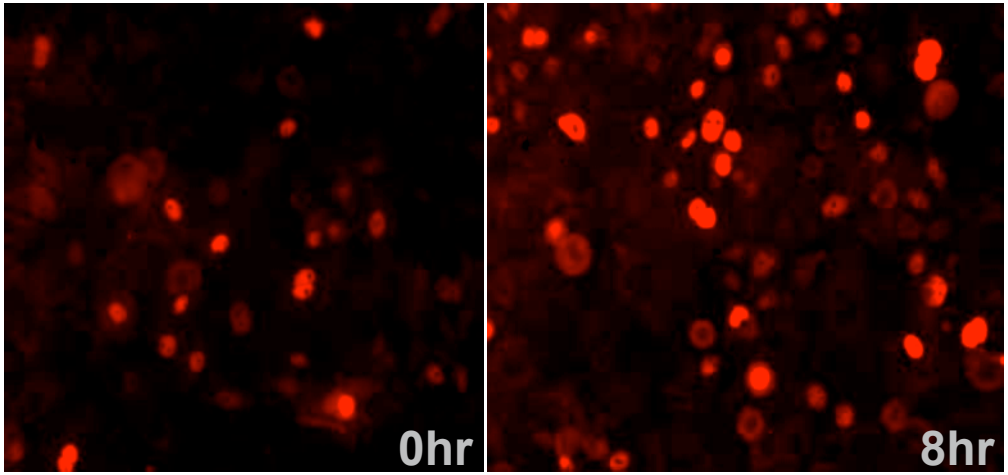
Target prediction in 1000 gene network (**B**)



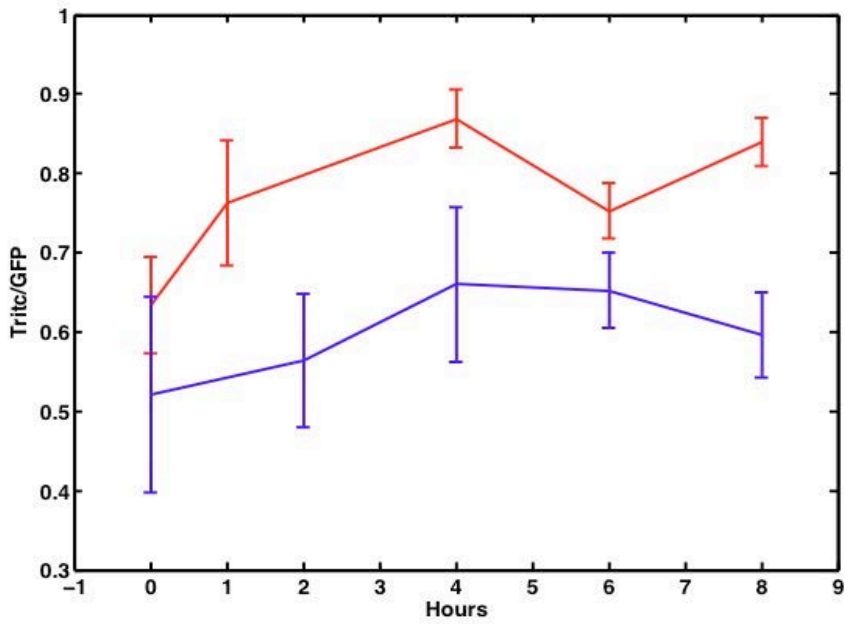
Application to p63:



Localization of ΔNp63 upon tamoxifen treatment by immunofluorescence



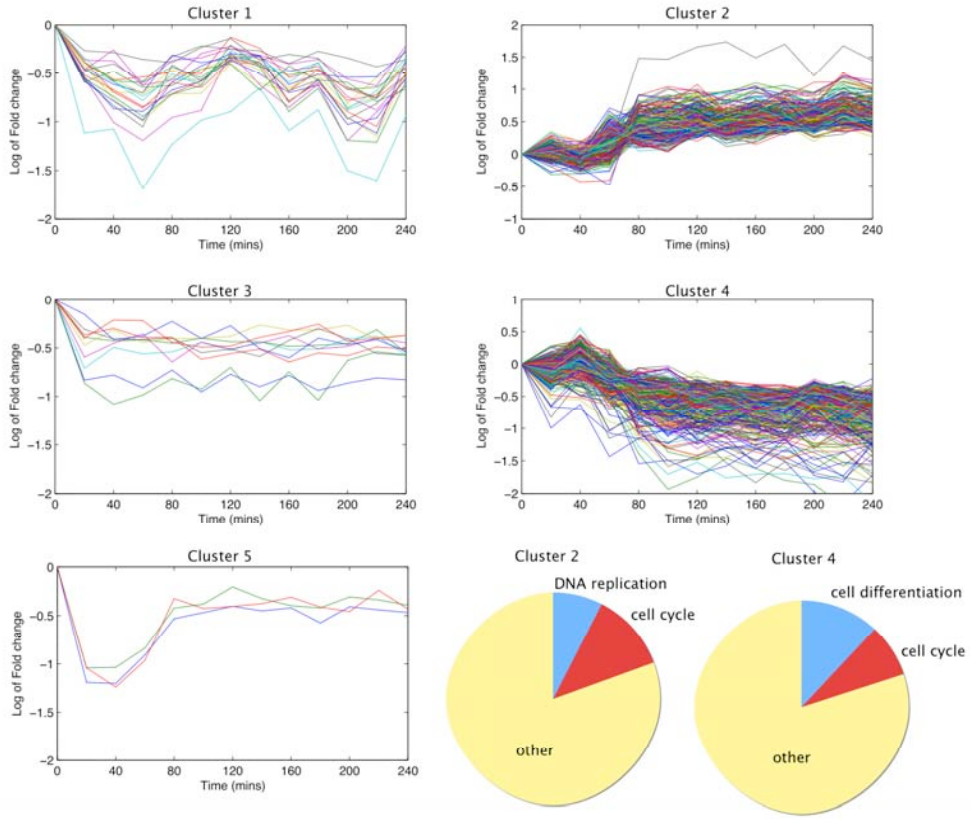
Murine keratinocytes are infected with Ginco ErDNp63 retrovirus. The cells are collected at the following time points from Tamoxifen induction: 1hr, 2hrs,4hrs, 8hrs.



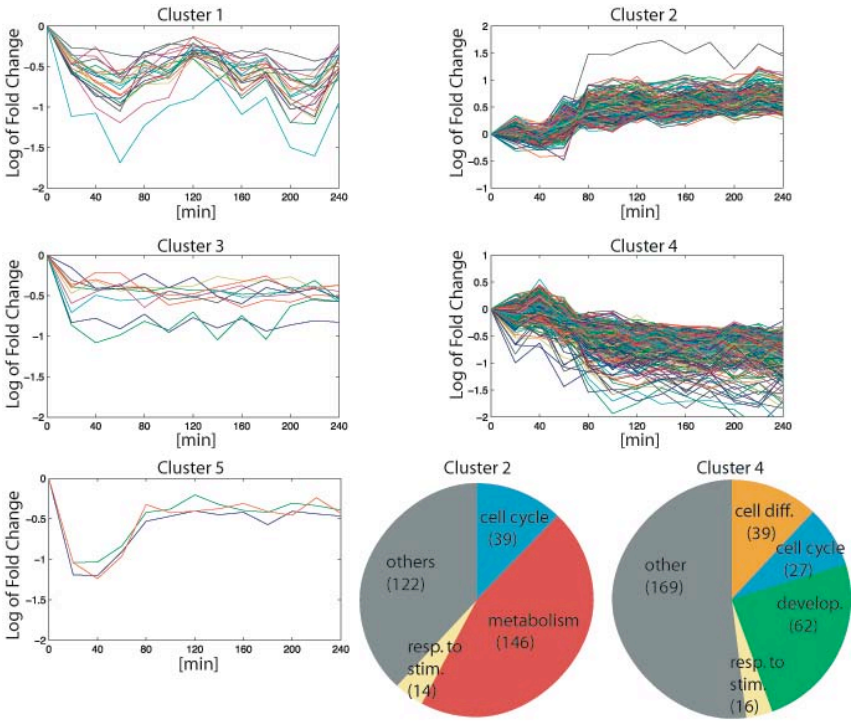
Immunofluorescence plot: red and blue lines represent two different immunofluorescence experiments.

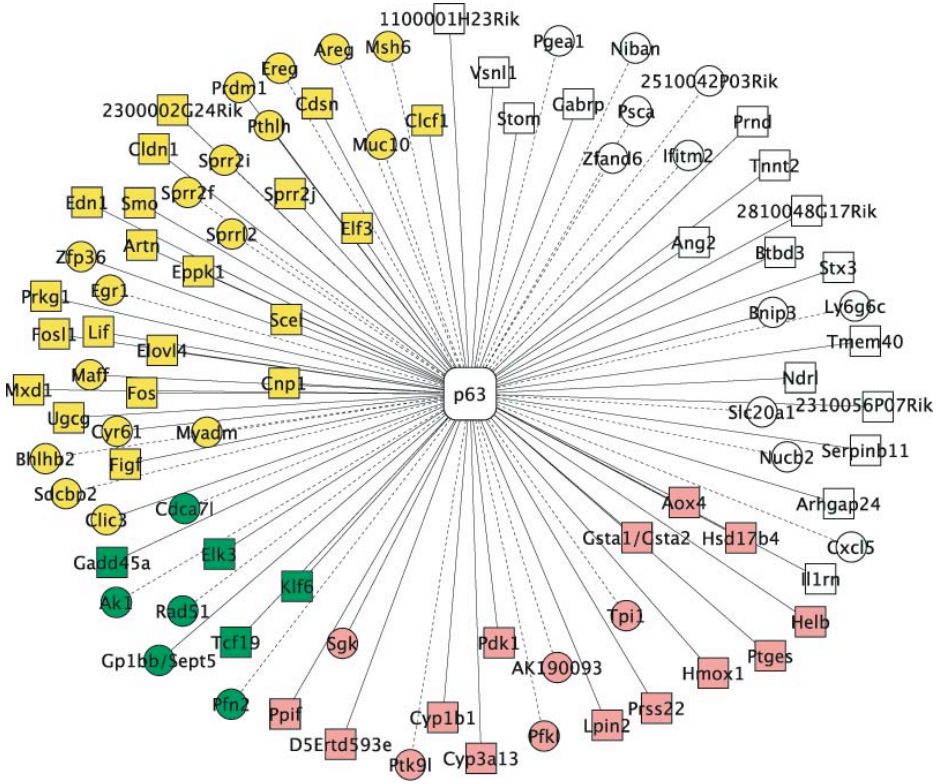
$$\dot{x}(t) = A \cdot x(t) + Bu(t)$$

Hierarchical Clustering of 786 expression profiles:



Application of TSNI on 786 expression profiles:





TSNI (B matrix)

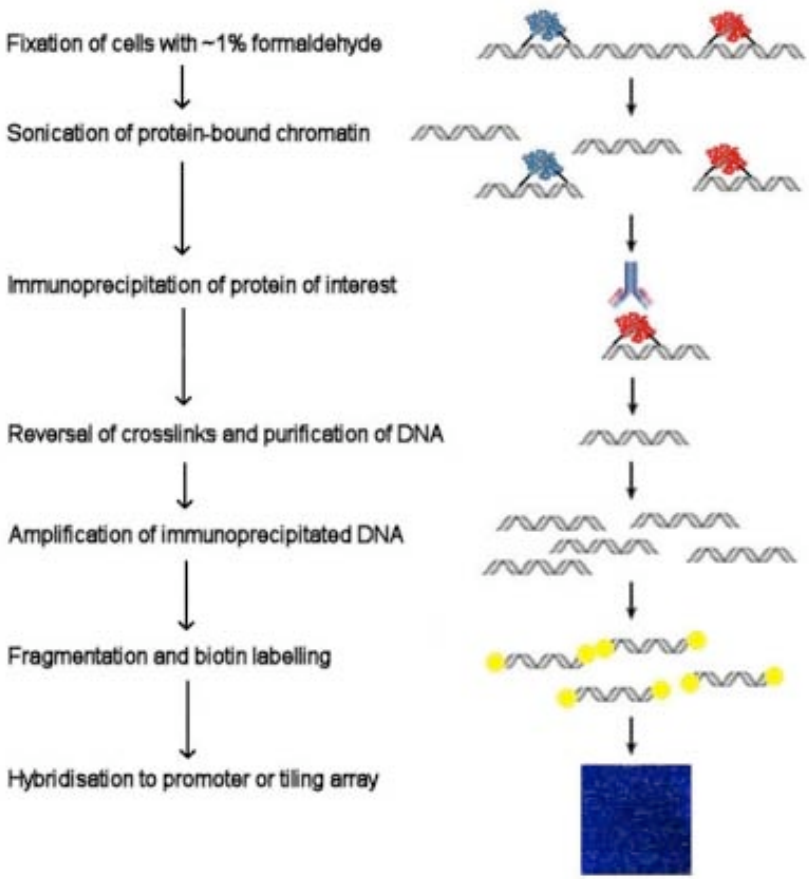
Validation of predicted targets: (1) ChIP-chip

• We selected the top 100 predicted targets and as control the bottom 200 predicted genes.

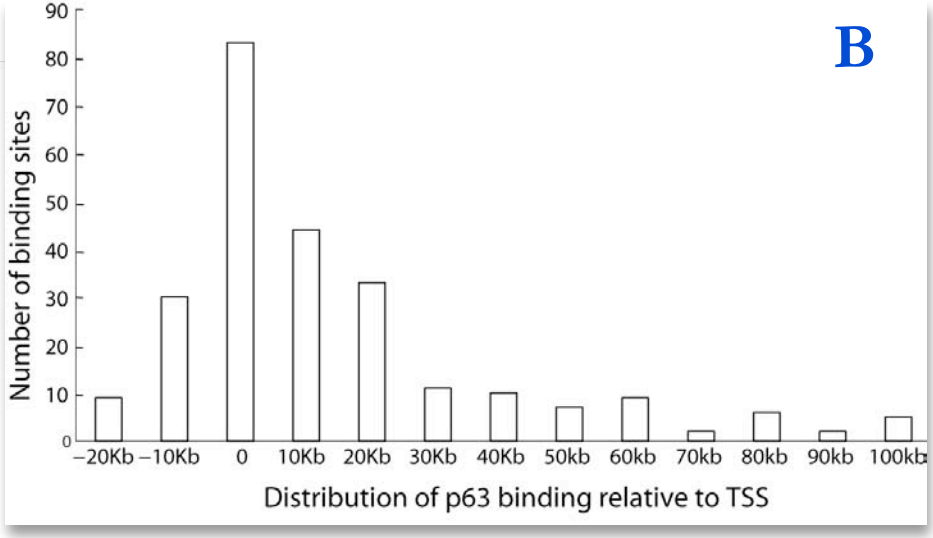
•For each gene we analysed 20 Kbp upstream+gene sequence via ChIP-chip (Agilent 200k custom-array)

•We found about 300 p63-binding sites bu ChIP-chip with a p-value<0.01

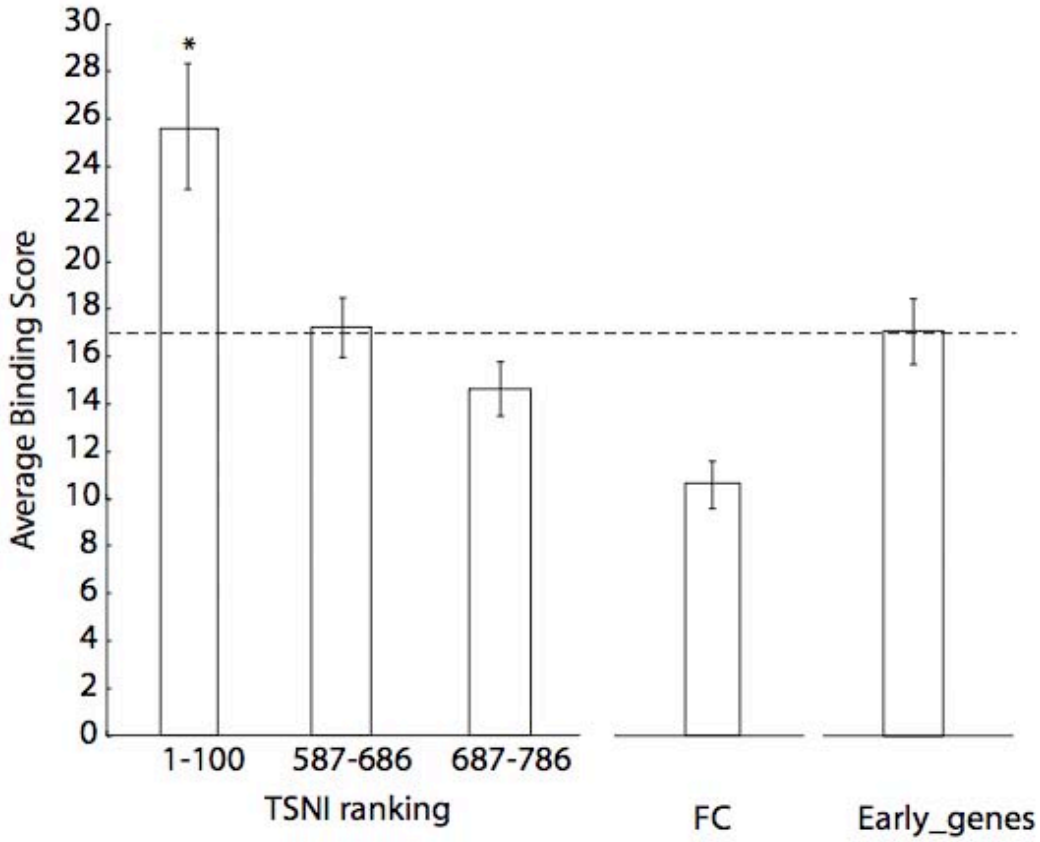
Validation of predicted targets: (1) ChIP-chip



Validation of predicted targets: (1) ChIP-chip

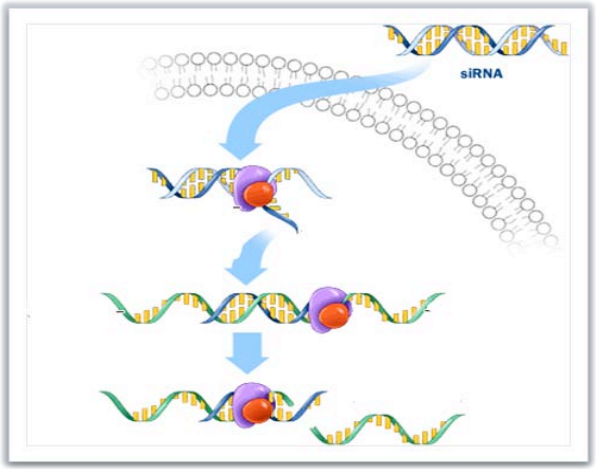


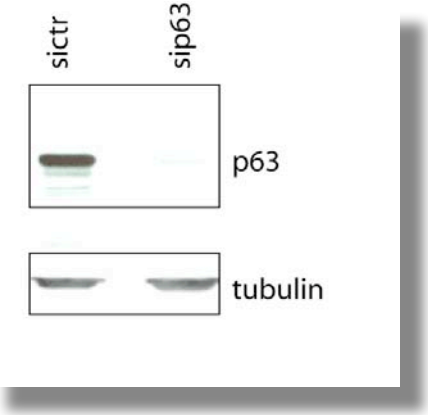
TSNI results are confirmed by ChIP-chip

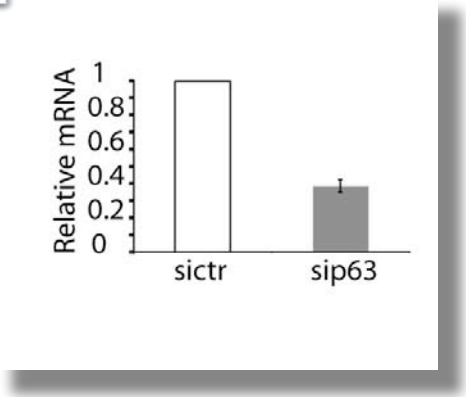


Knocking down p63 expression via siRNA:

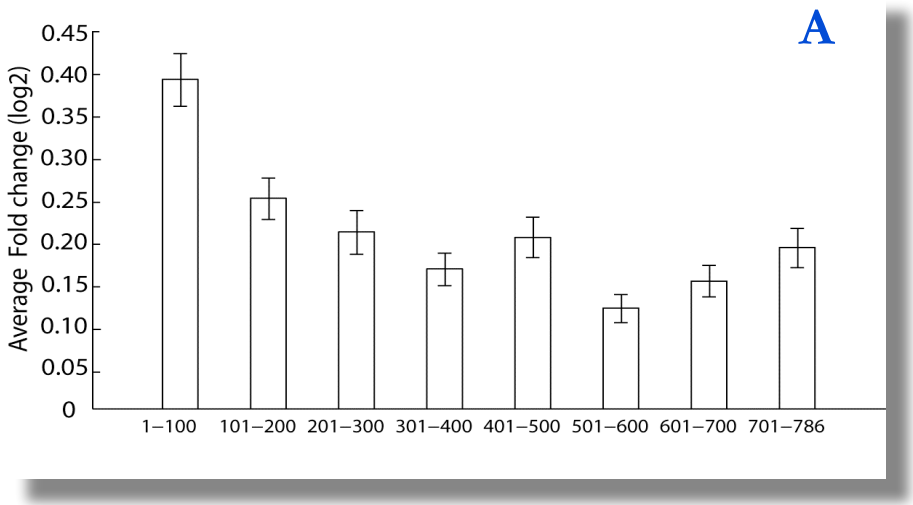
RNA interference against the p63 DNA binding domain

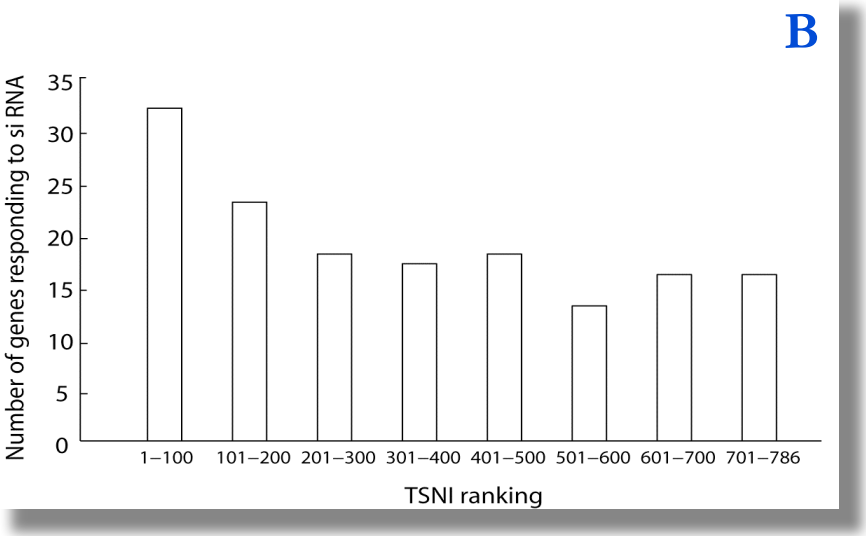




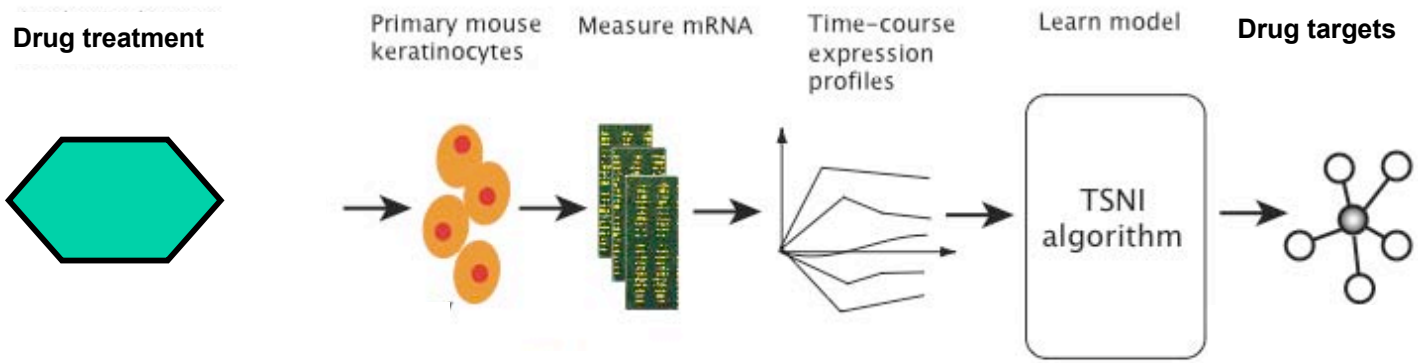


TSNI results are confirmed by siRNA



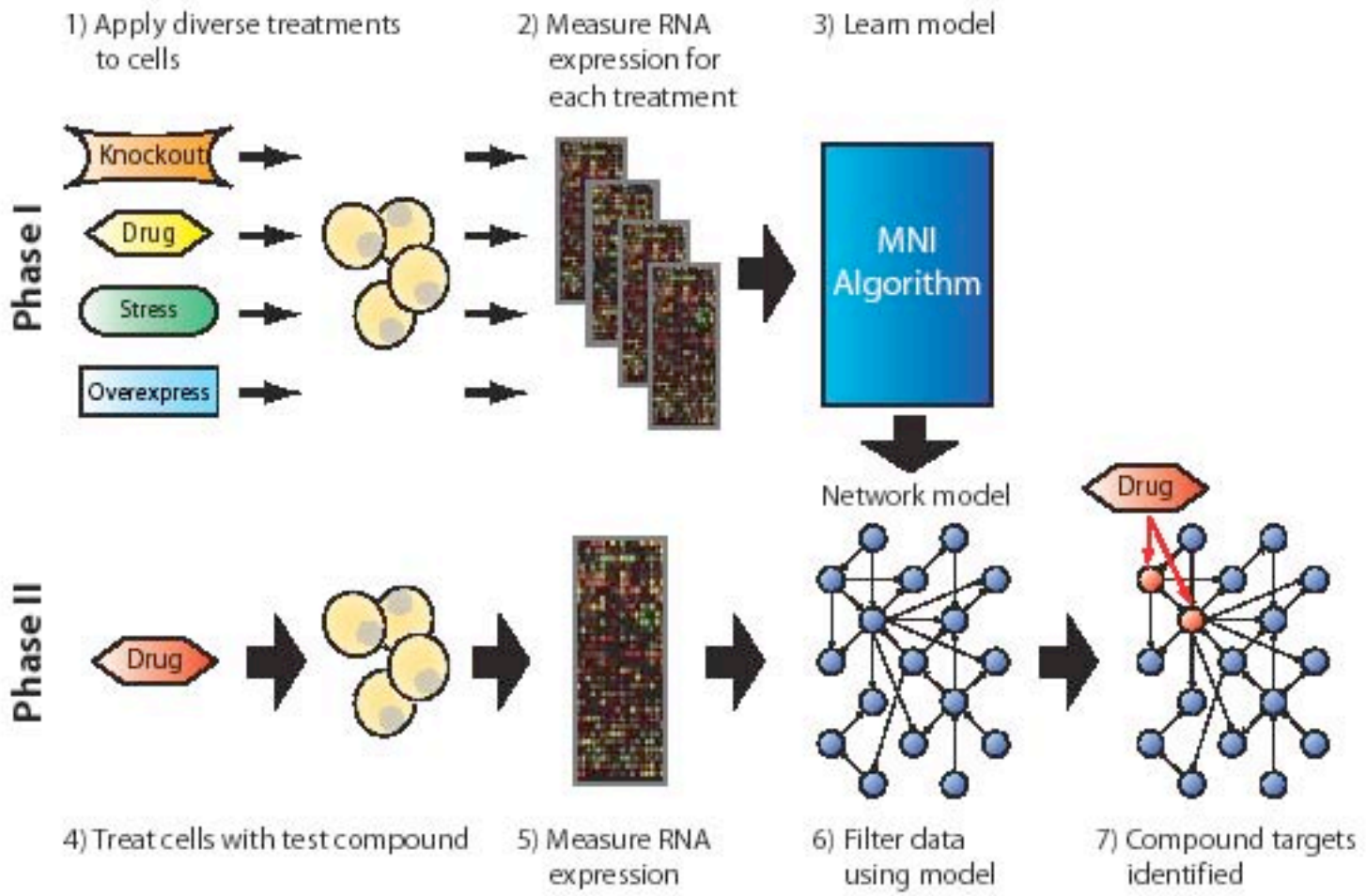


Application of TSNI to drug discovery :



A “steady-state” approach to drug target identification:

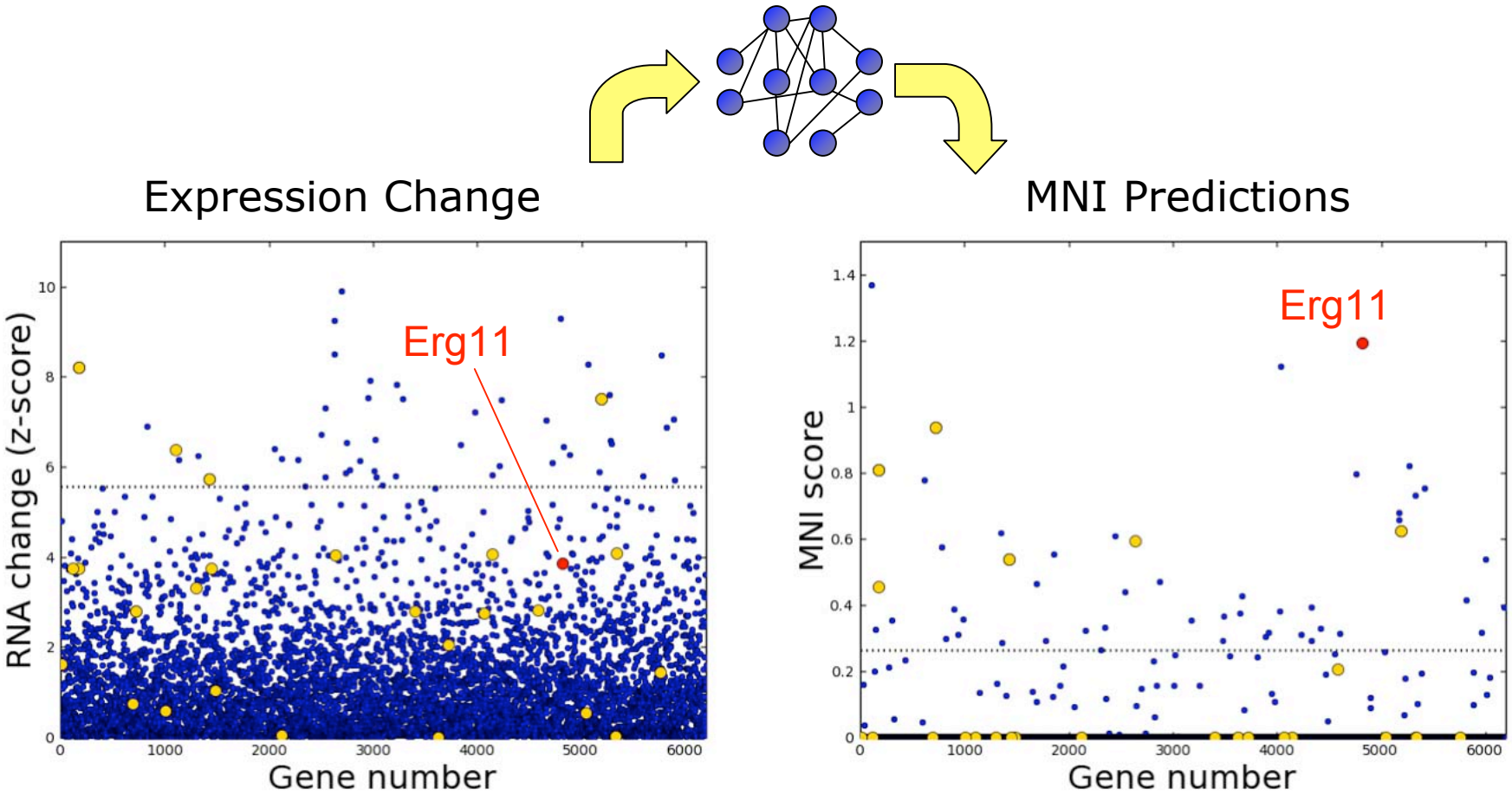
D di Bernardo et al
Nature Biotechnolgy, 2005



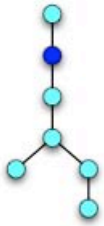
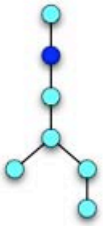
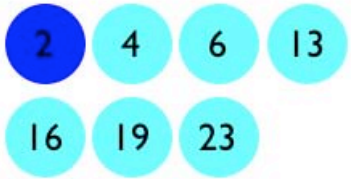
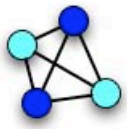
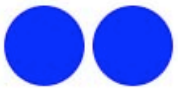
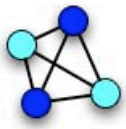
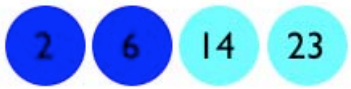
MNI identifies target of itraconazole in S. cerevisiae

Itraconazole treatment: a known target is ERG11

Filter through MNI-inferred network model



MNI results:

Compound	Known pathway	Known target	Predicted pathway	Ranked target genes (rank)
Itraconazole	ergosterol biosynthesis 	Erg11 	steroid metabolism 	***ERG11,*** *ERG24, ERG1, ERG25, CYB5, ERG27, ATF2* 
Hydroxyurea	DNA replication 	Rnr2, Rnr4 	DNA replication 	***RNR4, RNR2,*** *RNR1, RNR3* 

Pathway data from GO database

MNI identifies MOA for 8 out of 9 drugs:

Table 2 Pathways and associated genes targeted by drug compounds

Drug	Known pathway	Known target	Significant GO ontology (rank, *P*-value)	Ranked pathway genes (rank)
Terbinafine	Ergosterol biosynthesis[41]	Erg1	**Steroid metabolism** (1, 10^{-14})	*ERG7* (4), ***ERG1* (5)**, *ERG8* (11), *ERG26* (13), *UPC2* (17), *ERG28* (18), *ERG11* (20), *DAP1* (33), *HES1* (34), *ATF2* (36), *ERG5* (49)
Lovastatin	Ergosterol biosynthesis[42]	Hmg2, Hmg1	**Lipid metabolism** (1, 10^{-4})	*BST1* (1), *ERG1* (18), *YSR3* (23), ***HMG2* (30)**, *LCB5* (31), *ERG13* (36), *VRG4* (48)
Itraconazole	Ergosteroll biosynthesis[43]	Erg11	**Steroid metabolism** (1, 10^{-8})	***ERG11* (2)**, *ERG24* (4), *ERG1* (6), *ERG25* (13), *CYB5* (16), *ERG27* (19), *ATF2* (23)
Hydroxyurea	DNA replication[44]	Rnr2, Rnr4	Heteroduplex formation (1, 10^{-4})	*RAD51* (15), *RAD54* (47)
			DNA replication (2, 10^{-2})	***RNR4* (2)**, ***RNR2* (6)**, *RNR1* (14), *RNR3* (23)
Cycloheximide	Protein biosynthesis[45]	Ribosome	Nuclear mRNA splicing, via spliceosome (1, 10^{-4})	*SYF1* (3), *SMD3* (19), *HSH49* (42)
			–	***RPL26B* (32)**, ***RPS29A* (34)**
Tunicamycin	N-linked glycosylation[46]	Alg7	**Protein-ER targeting** (1, 10^{-3})	*SEC62* (1), *SIL1* (31), *SEC59*[a] (43)
Nikkomycin	Cell wall chitin biosynthesis[47]	Chs3	Protein amino acid alkylation(1, 10^{-3})	*SWD2* (3), *RMT2* (6)
		Drugs not in the original compendium data set		
3-aminotriazole	Histidine biosynthesis[48]	His3	**Organic acid metabolism** (1, 10^{-7})	*FRM2* (8), *BIO5* (9), *YAT2* (10), *ARO10* (18), *ARO9* (20), *CHA1* (21), *BIO3* (31), *ARG1* (33), *ARG4* (37), *HIS5*[†] (42), *LYS1* (47), *SAM2* (50)
	Oxygen and reactive oxygen species metabolism[30]	Cta1		
Dyclonine	Ergosterol biosynthesis[1]	Erg2	**Sterol biosynthesis** (1, 10^{-18})	*ERG3* (1), *ERG6* (2), *CYB5* (3), ***ERG2* (4)**, *ERG11* (6), *ERG28* (10), *ERG1* (12), *ERG5* (13), *ERG27* (18), *MVD1* (23), *ERG24* (30), *ERG26* (37)

Acknowledgements (http://dibernardo.tigem.it for more info):

Dr Alberto Ambesi

(Sequence analysis)

Giusy Della Gatta

(Experimental work)

Mukesh Bansal

(Computational work)

Velia Siciliano
Lucia Marucci
Giulia Cuccato

Vincenzo Belcastro

Francesco Iorio

Mario Lauria

…and Dr Caterina Missero (CEINGE, Napoli) for supervision of experiments; Prof. Tim Gardner and Prof. Jim Collins (Boston University, USA) for MNI/NIR related work.

Selection of Significant Genes in Time Series experiment:

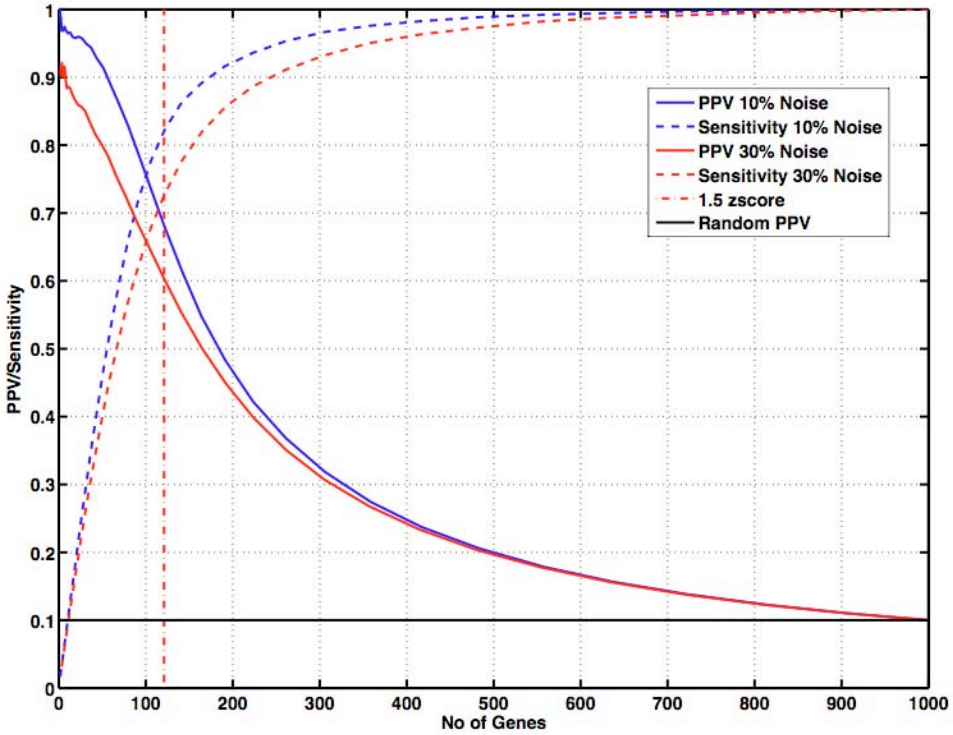
- First filter
  - Area under curve
- Second filter
  - Standard deviation is computed at each time point using smoothing and generalized cross validation algorithm

- Statistical test for significance (Chi-Square)

in silico with dual perturbation

$$dX_1/dt = a_2 X_2 + a_6 X_6 + a_9 X_9 + a_{12} X_{12} + b_{11}u_1 + b_{12}u_2$$

p63 Tamoxifen



Improvement of TSNI to identify the network (A)

- Integral Approach instead of differential approach

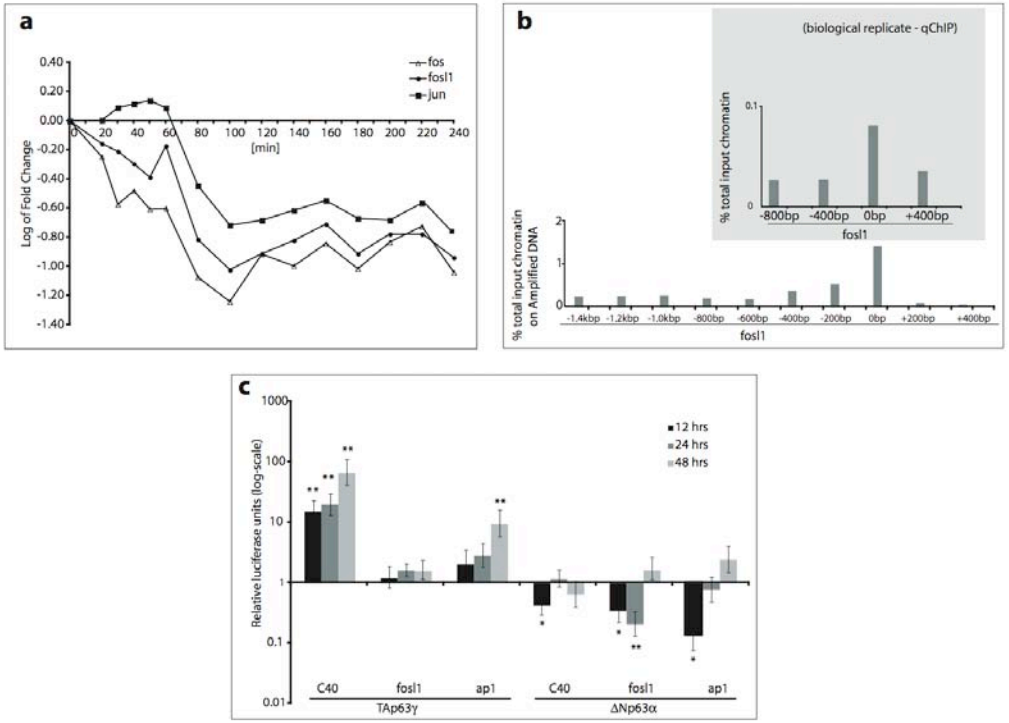
$$\int_0^{t_k} \dot{X}(t)dt = A\int_0^{t_k} X(t)dt + B\int_0^{t_k} U(t)dt \quad k = 1\ldots M$$

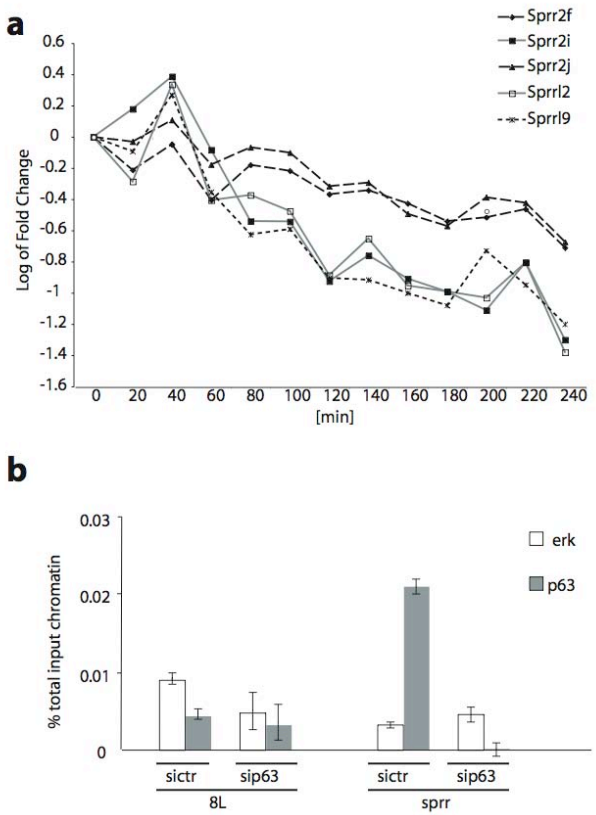
$$X(t_k) = A\int_0^{t_k} X(t)dt + B\int_0^{t_k} U(t)dt \quad k = 1\ldots M$$

- Bootstrapping

Mukesh Bansal, Diego di Bernardo
Inference of gene networks from temporal gene expression profiles. IET (Under Review)

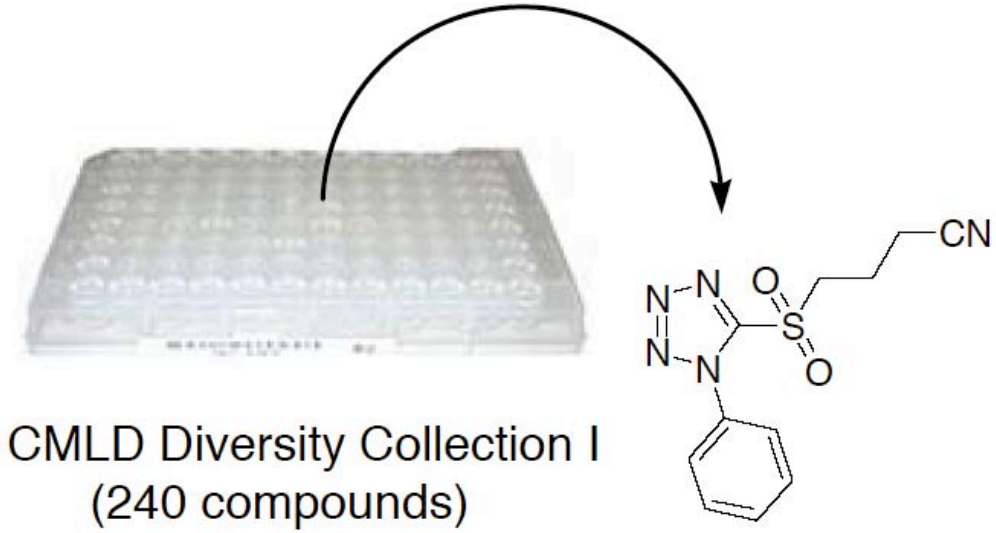
(1a) Time Series Network Identification Algorithm (NEW) :





P63 directly regulates at early times AP1 complex and markers of terminal keratinocyte differentiation *-Manuscript submitted-*

Identified novel anticancer compound via chemical screen



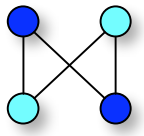
PTSB IC_{50} A549 5 μM

- PTSB inhibits growth in yeast and tumor cell lines

In collaboration with Schaus and Elliot laboratories
Dept. of Chemistry, Boston University
Center for Methodology and Library Development (CMLD), Boston U.

MNI identifies two enzyme targets of PTSB

Identifies thioredoxin (TRX2) and thioredoxin reductase (TRR1)

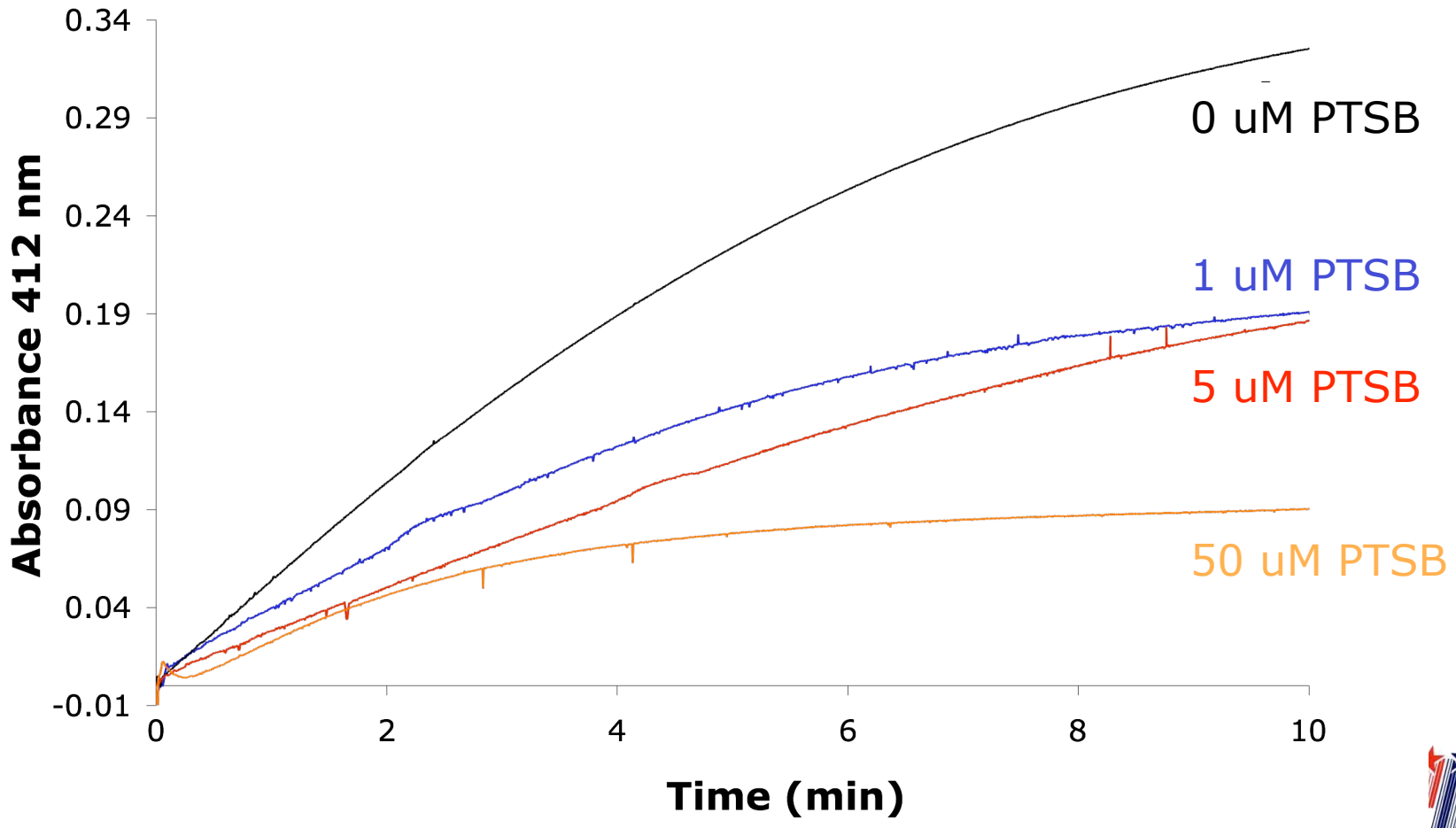
Compound	Known pathway	Known target	Predicted pathway	Ranked target genes (rank)
PTSB	unknown	unknown	cell redox homeostasis 	***TRR1, TRX2*** 32 36

TRR1/TRX2 activity inhibited in presence of PTSB

Assay:

Thioredoxin reduction of dithio(bis)nitrobenzoic acid (DTNB)

- Product of reaction = thiolate anion, measured via A412



MNI algorithm overview and application to S. cerevisiae

1) Learn Model 6000 genes → 512 microarrays

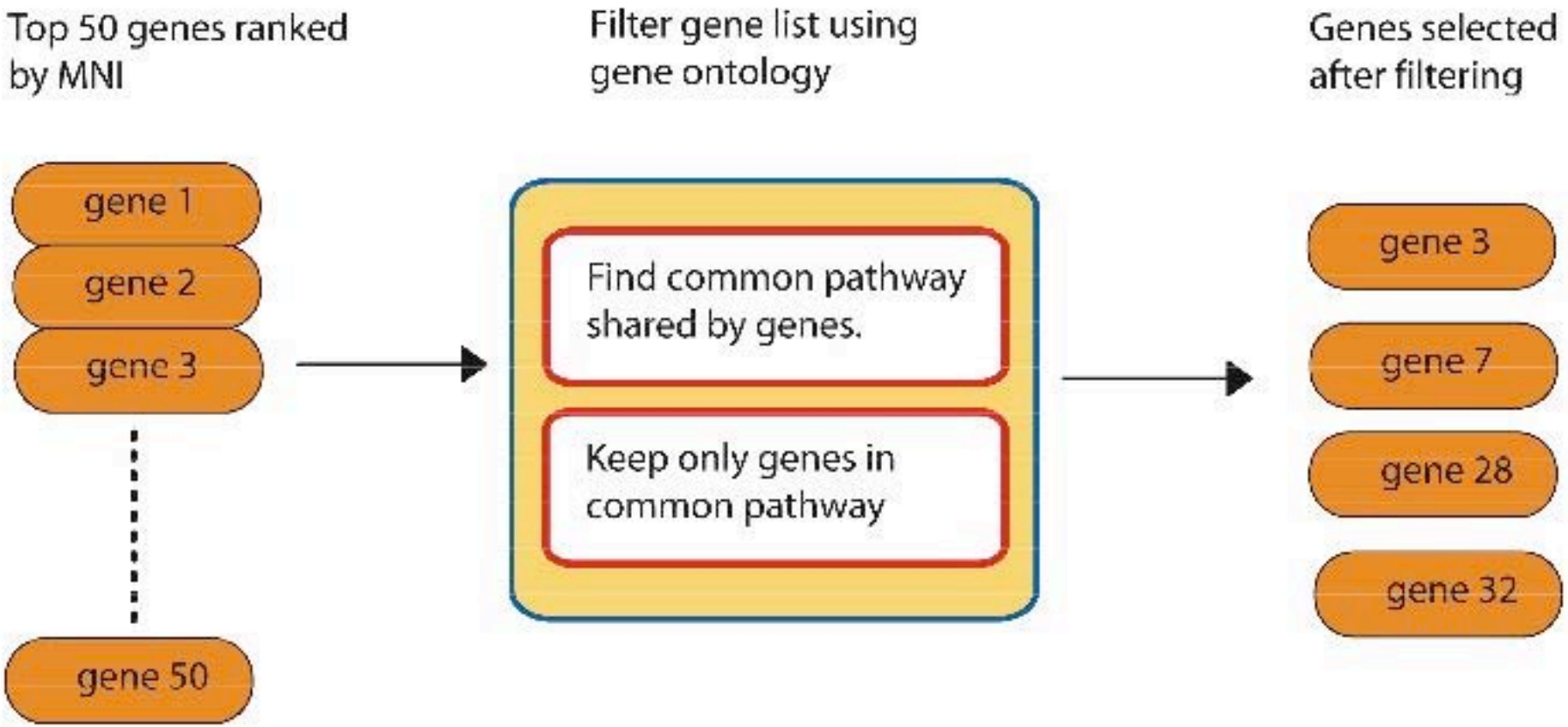
2) Measure profile in response to drug 15 compounds tested

3) Rank genes using model learned in step 1

4) Select top 50 genes, i.e. top 50 predicted drug targets

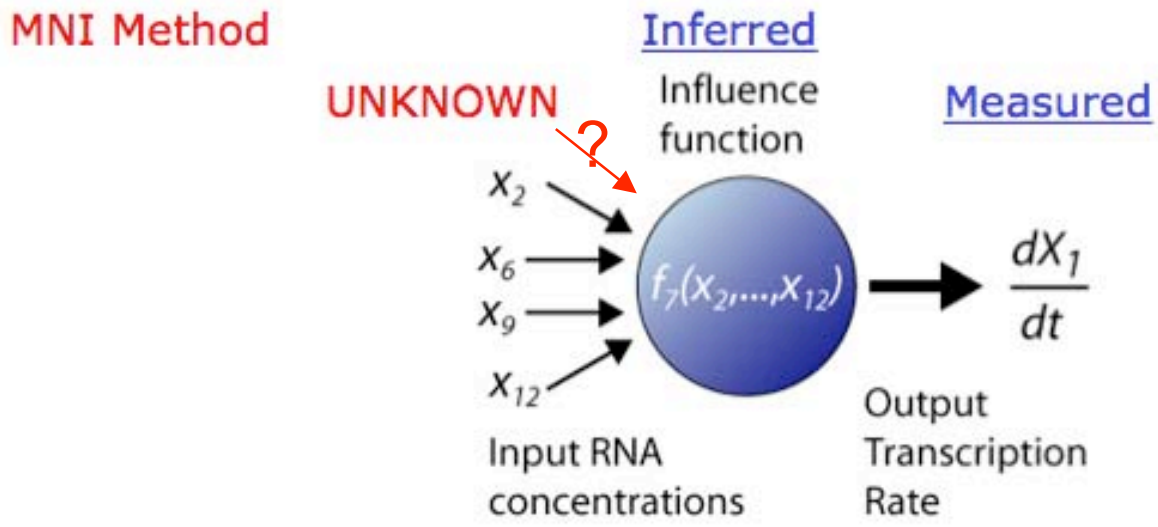
5) Find common pathway among top 50 predicted targets

Geno Ontology Filtering

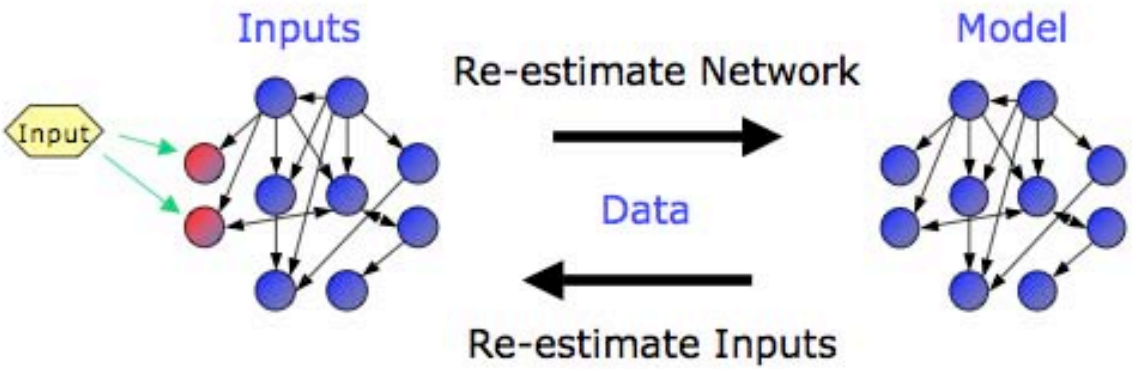


MNI: Microarray Network Identification

The MNI Algorithm: an unsupervised approach



MNI Algorithm: Recursively estimate inputs and model using data



Model structure:

$$\begin{cases} x'_1(t) = a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n + b_1u \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ x'_n(t) = a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nn}x_n + b_nu \end{cases}$$

Drug effect unknown

Or in matrix format:

DRUG EFFECT

$$x' = Ax + \underline{b}u$$

? ?

NETWORK

Fit criterion and search solution strategy – phase 1:

$$0 = a_{11}x_{11} + a_{12}x_{21} + \ldots + a_{1n}x_{n1} + \cancel{b_1 u}$$

.....

$$0 = a_{11}x_{1h} + a_{12}x_{2h} + \ldots + a_{1n}x_{nh} + \cancel{b_1 u}$$

Choose only the *h* experiments where gene 1 has not been perturbed and solve the eqs. for gene 1 with $a_{11} \neq 0$.

Repeat for all the genes and obtain matrix **A**

…we still did not say how to find the *h* experiments where gene *i* has not been perturbed, this is done simply by choosing those experiments where the gene has changed less…

Fit criterion and search solution strategy – phase II:

Say $\boldsymbol{x_d}$ the expression profile following the drug treatment, then the predicted targets of the drug are:

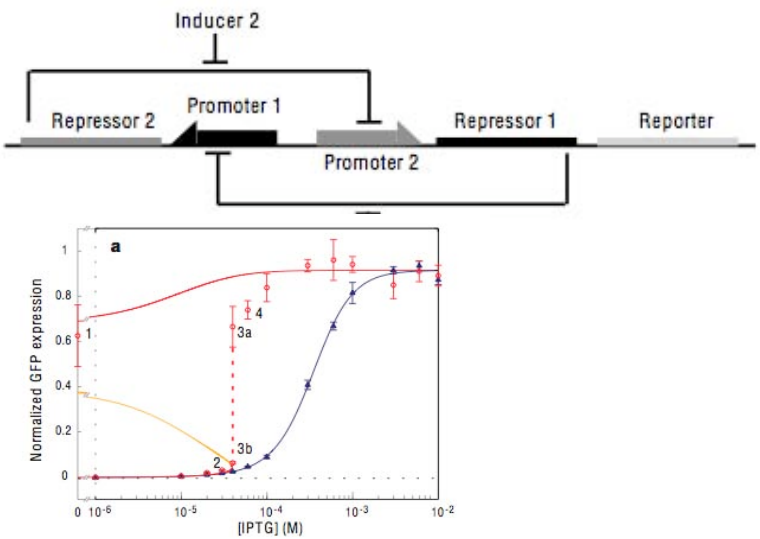
$$\mathbf{bu=-Ax_d}$$

Other projects going on in our lab:

- REVERSE ENGINEERING BASED ON MI AND BAYESIAN APPROACH
- IDENTIFICATION OF DRUG TARGETS
- SYNTHETIC BIOLOGY - COBIOS (coordinator) 120kE/yr for 3 yrs for my lab

Construction of a genetic toggle switch in *Escherichia coli*

Timothy S. Gardner*†, Charles R. Cantor* & James J. Collins*†



A synthetic oscillatory network of transcriptional regulators

Michael B. Elowitz & Stanislas Leibler