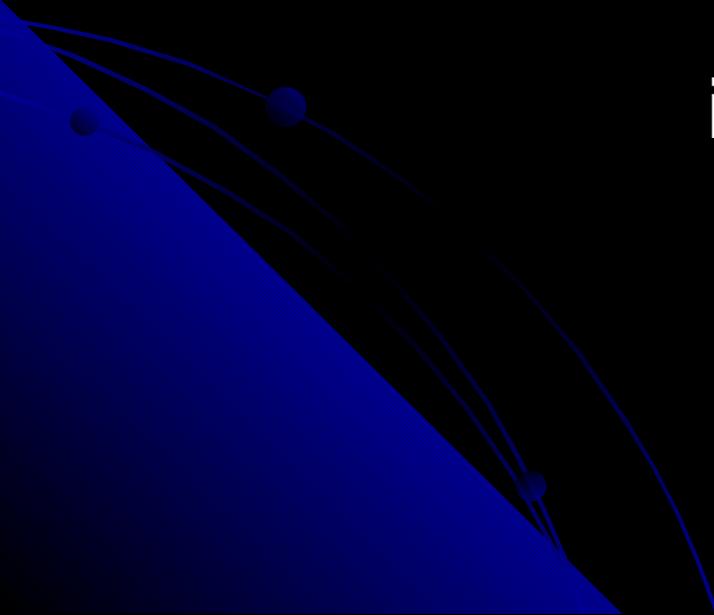
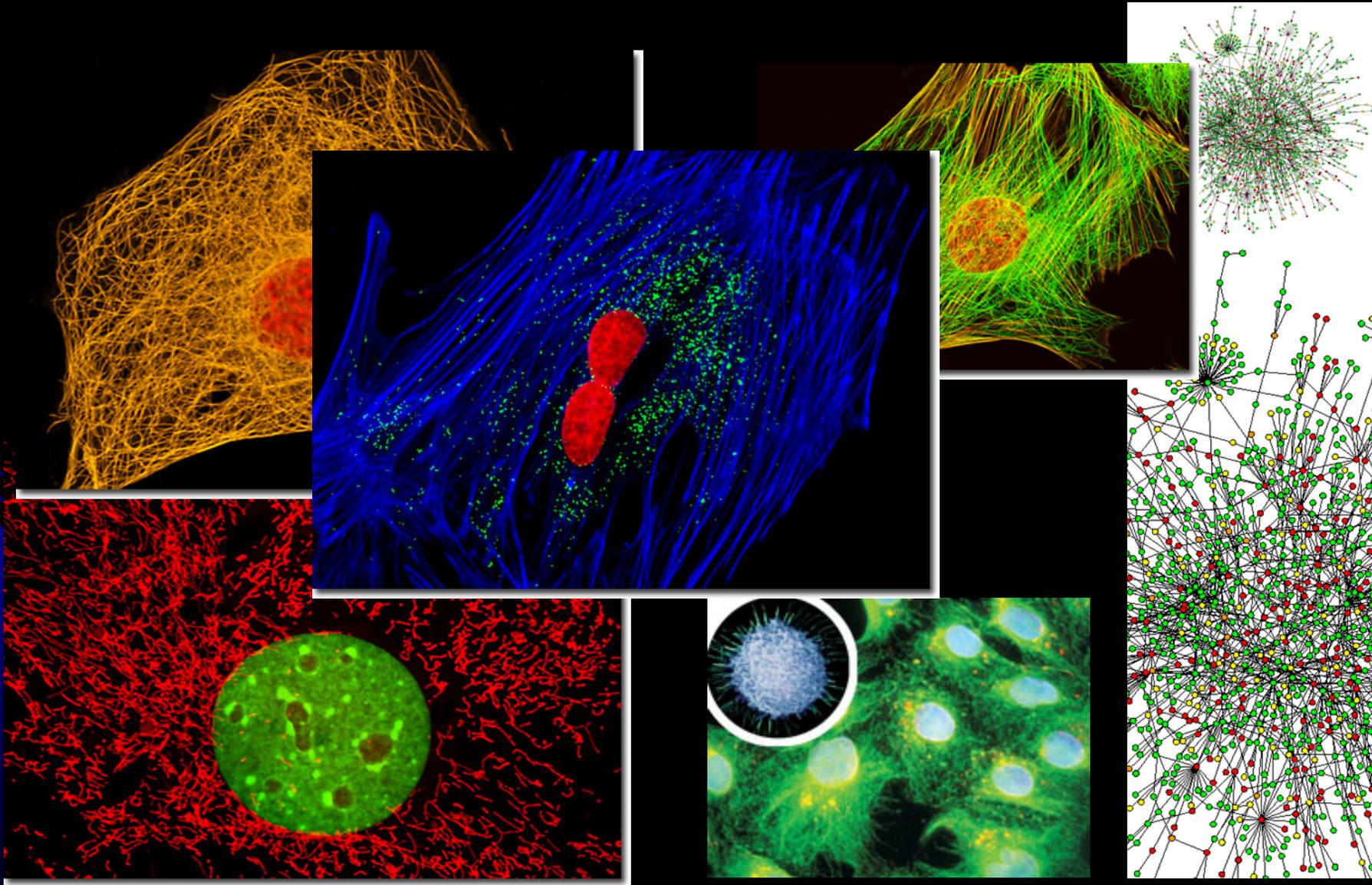


# Modeling and Inference of Transcriptional Regulatory Networks

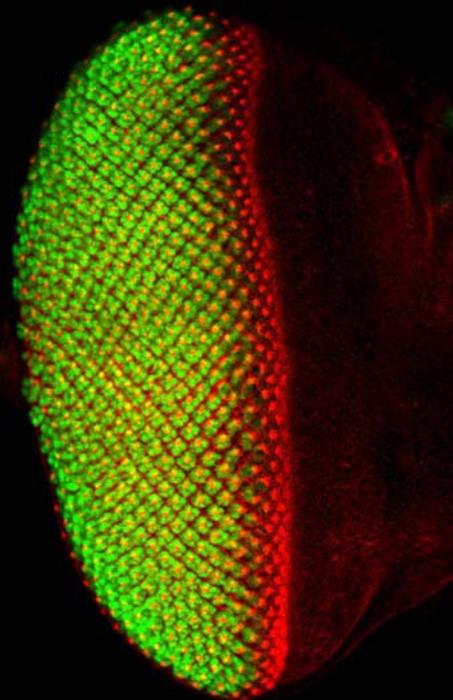
ilya shmulevich



# cells are systems of interacting molecules



organisms are systems of  
interacting cells



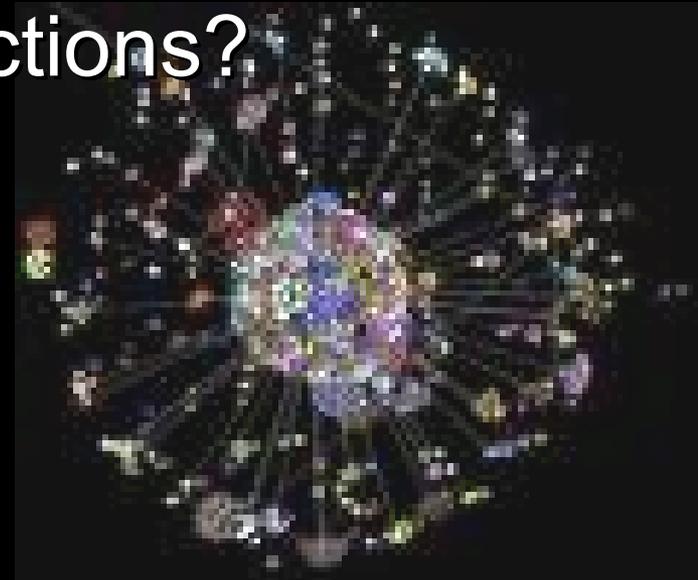
# societies are systems of interacting organisms



# living systems

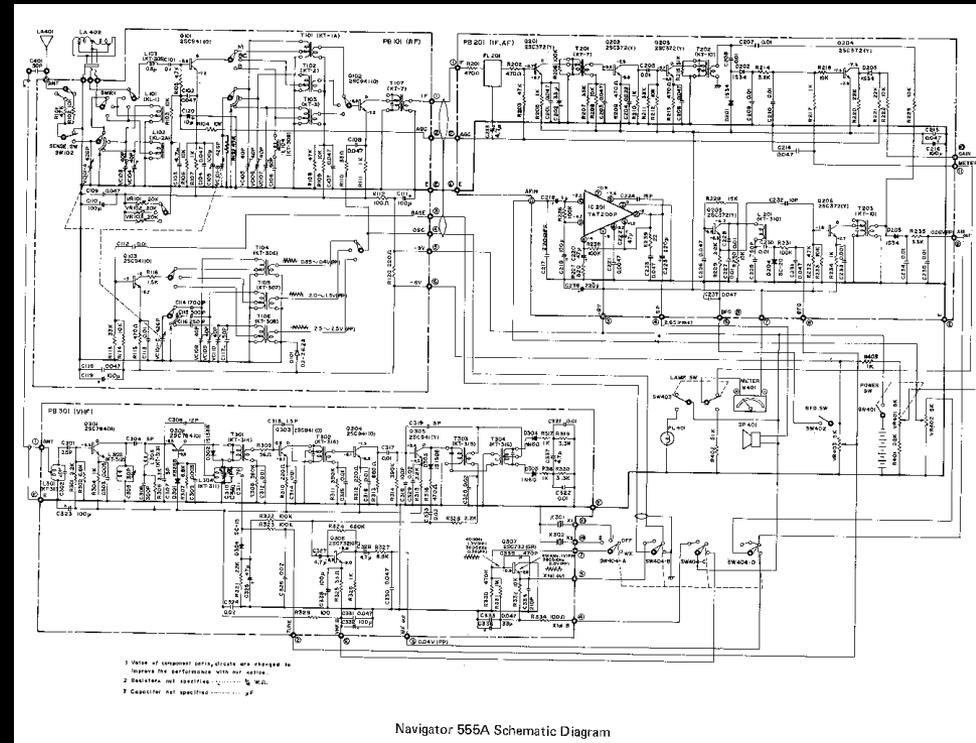
self-organized complex dynamical  
systems of interacting parts

how can we understand the emergent  
macroscopic properties of the system  
from its parts and their interactions?



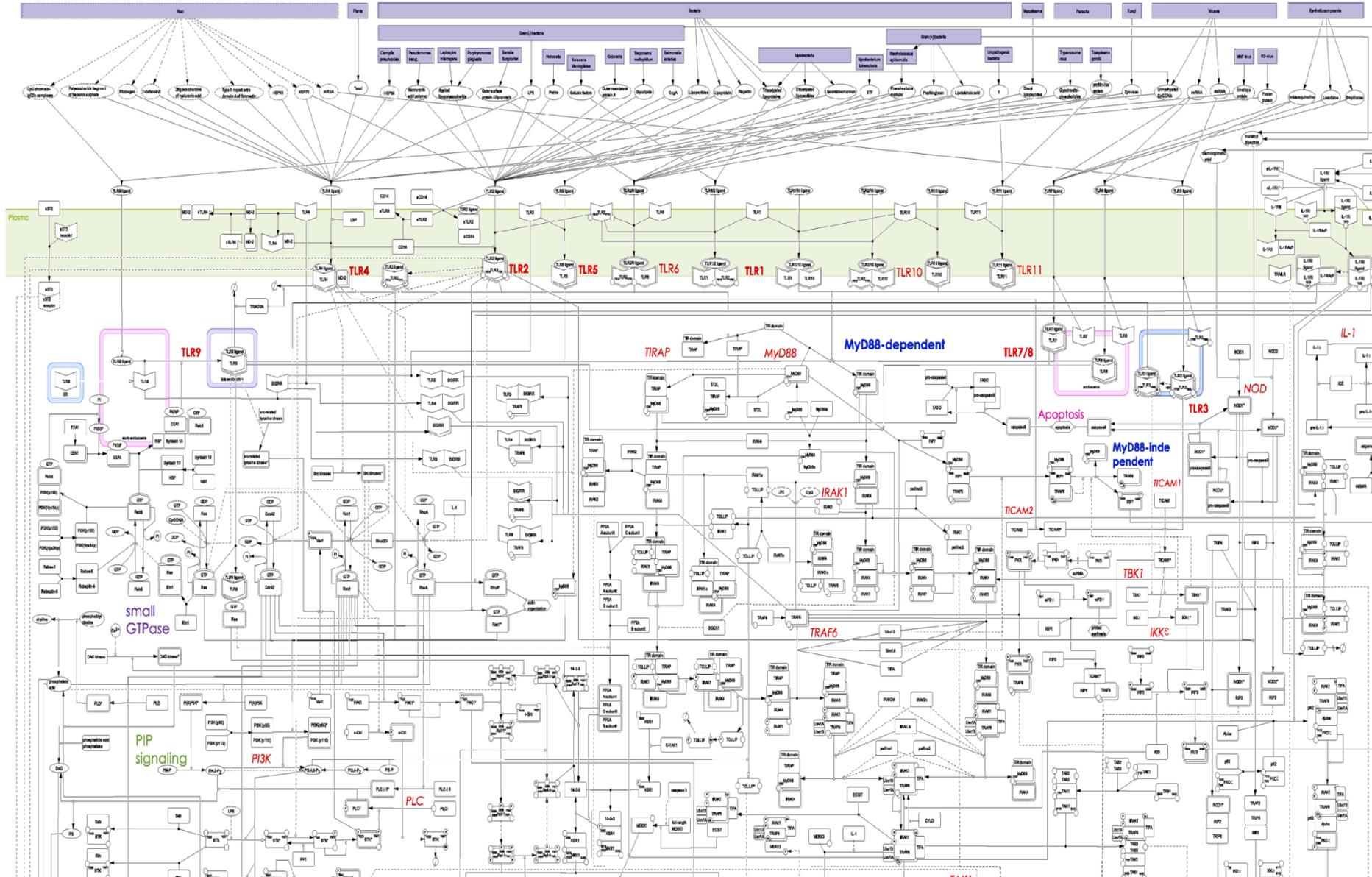
# genetic networks

- **Complex regulatory networks among genes and their products control cell behaviors such as:**
  - cell cycle
  - apoptosis
  - cell differentiation
  - communication between cells in tissues
- **A paramount problem is to understand the dynamical interactions among these genes, transcription factors, and signaling cascades, which govern the integrated behavior of the cell.**

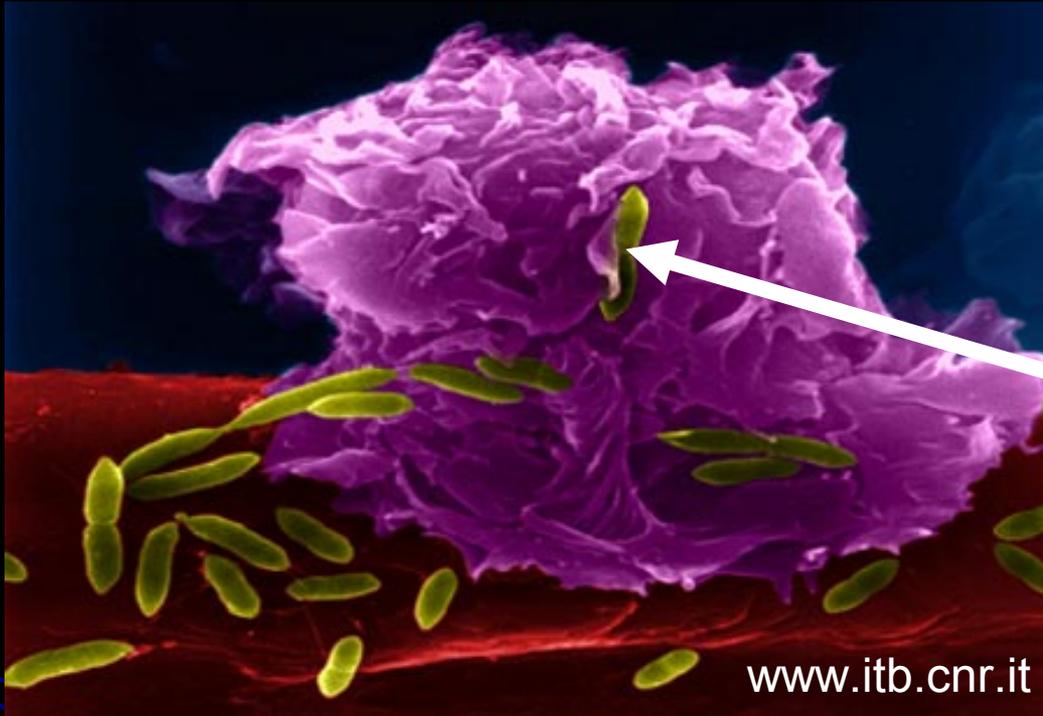


**Analogy: circuit diagram**

# Map of the TLR signaling pathway in the macrophage



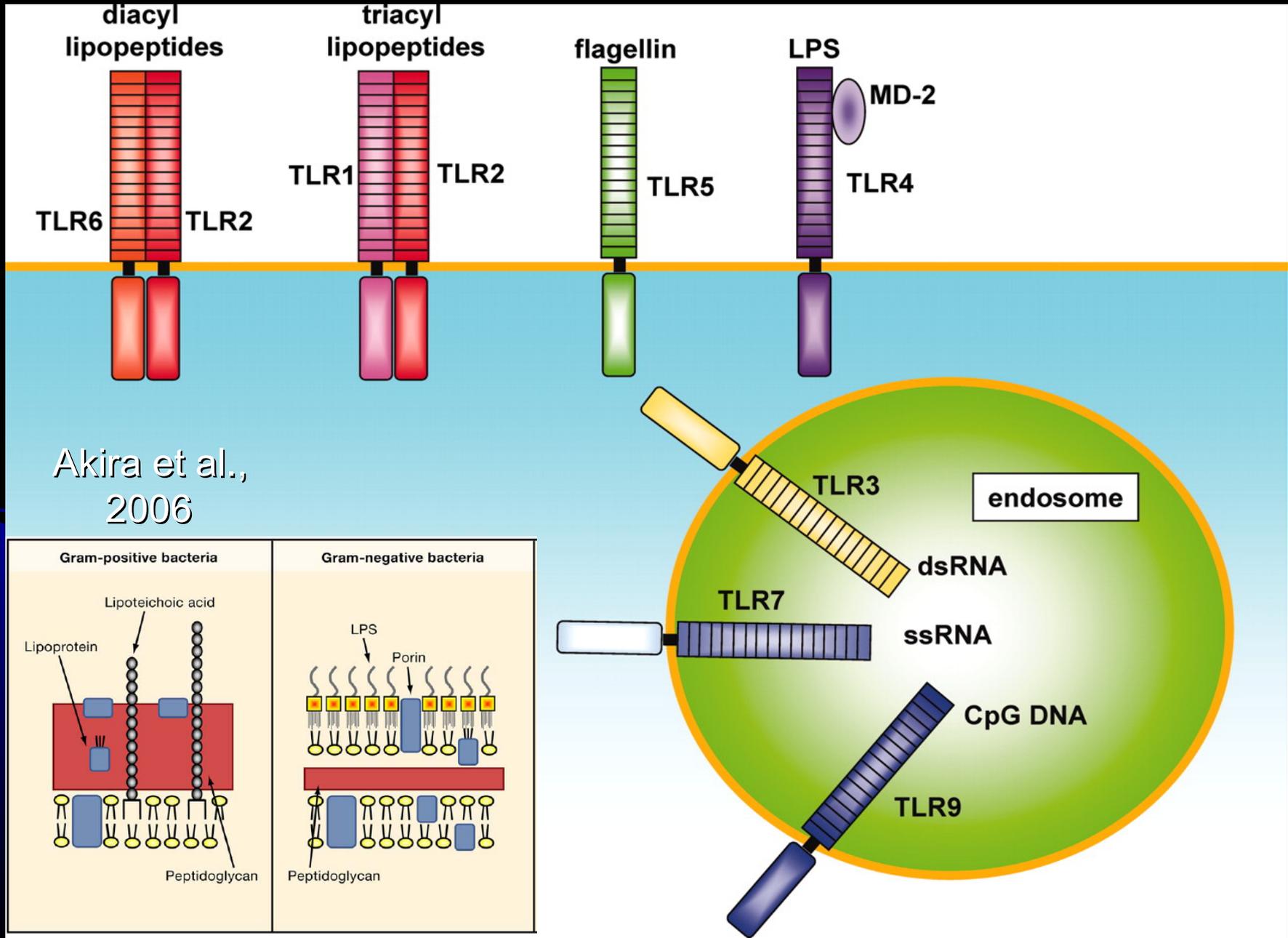
# Macrophages are key immune cells



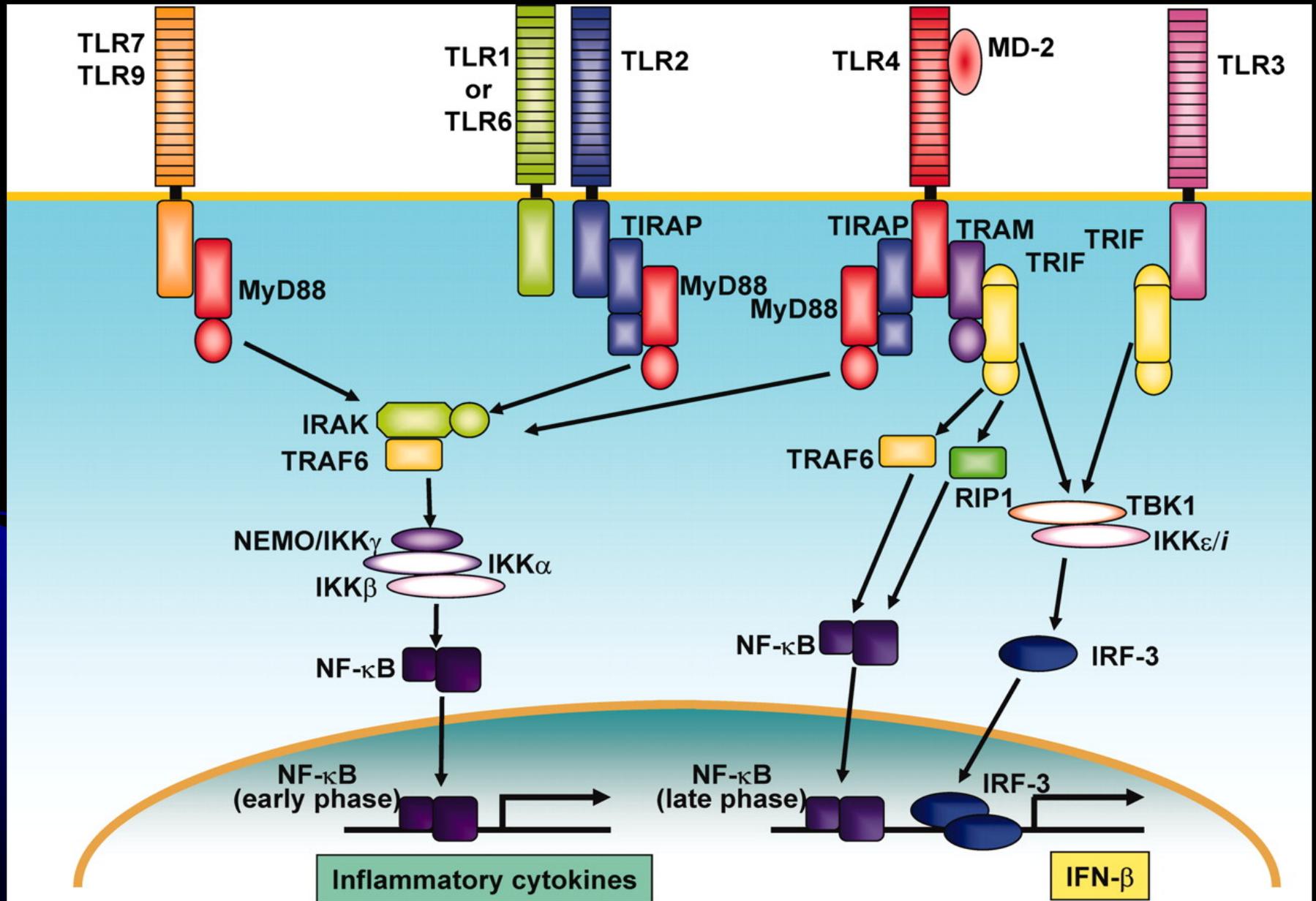
bacterium  
being engulfed

- Phagocytosis
- Antigen presentation
- Secretion of proinflammatory cytokines
- Wound healing

# TLRs and their associated PAMPs



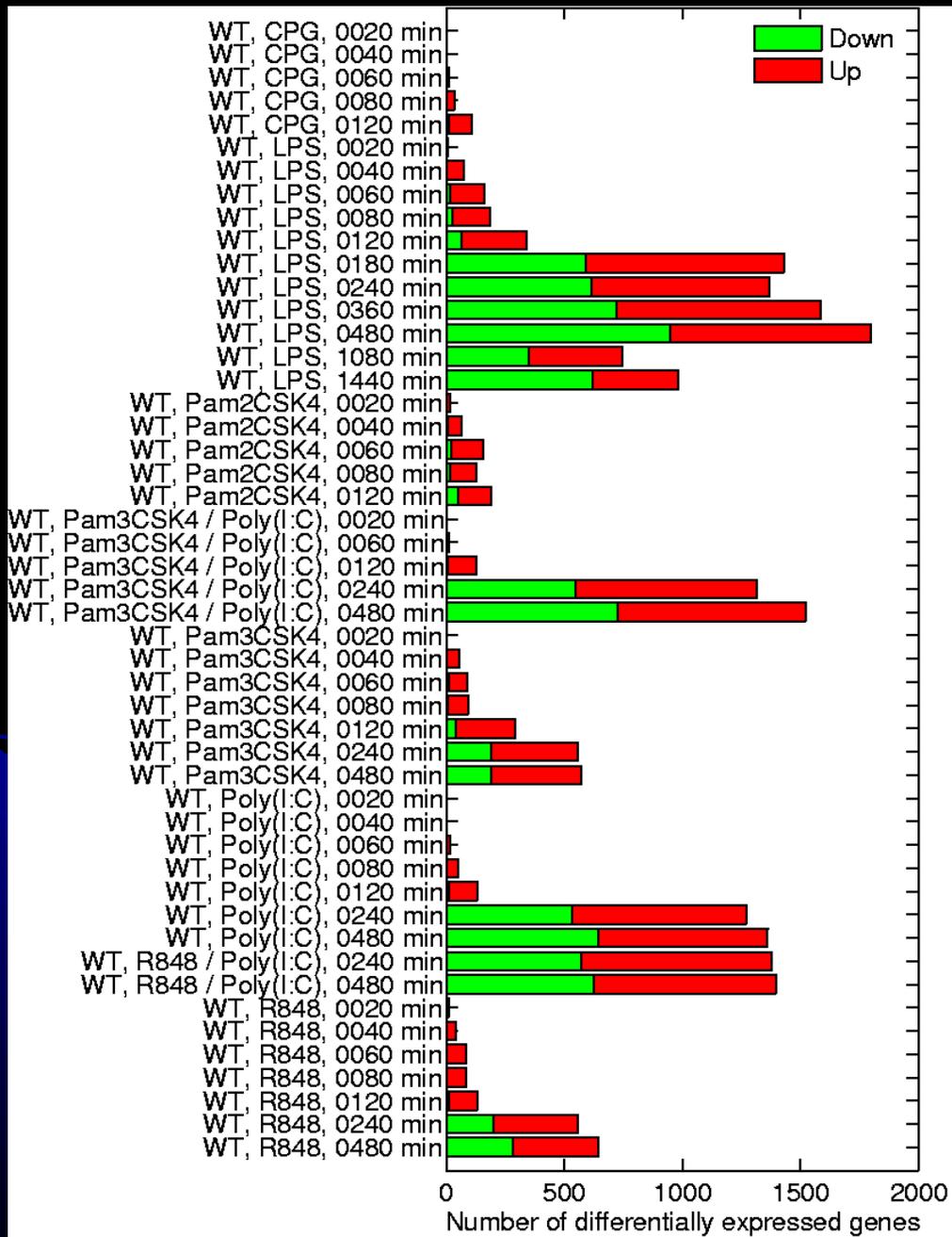
# TLR signaling pathway





# The ingredients

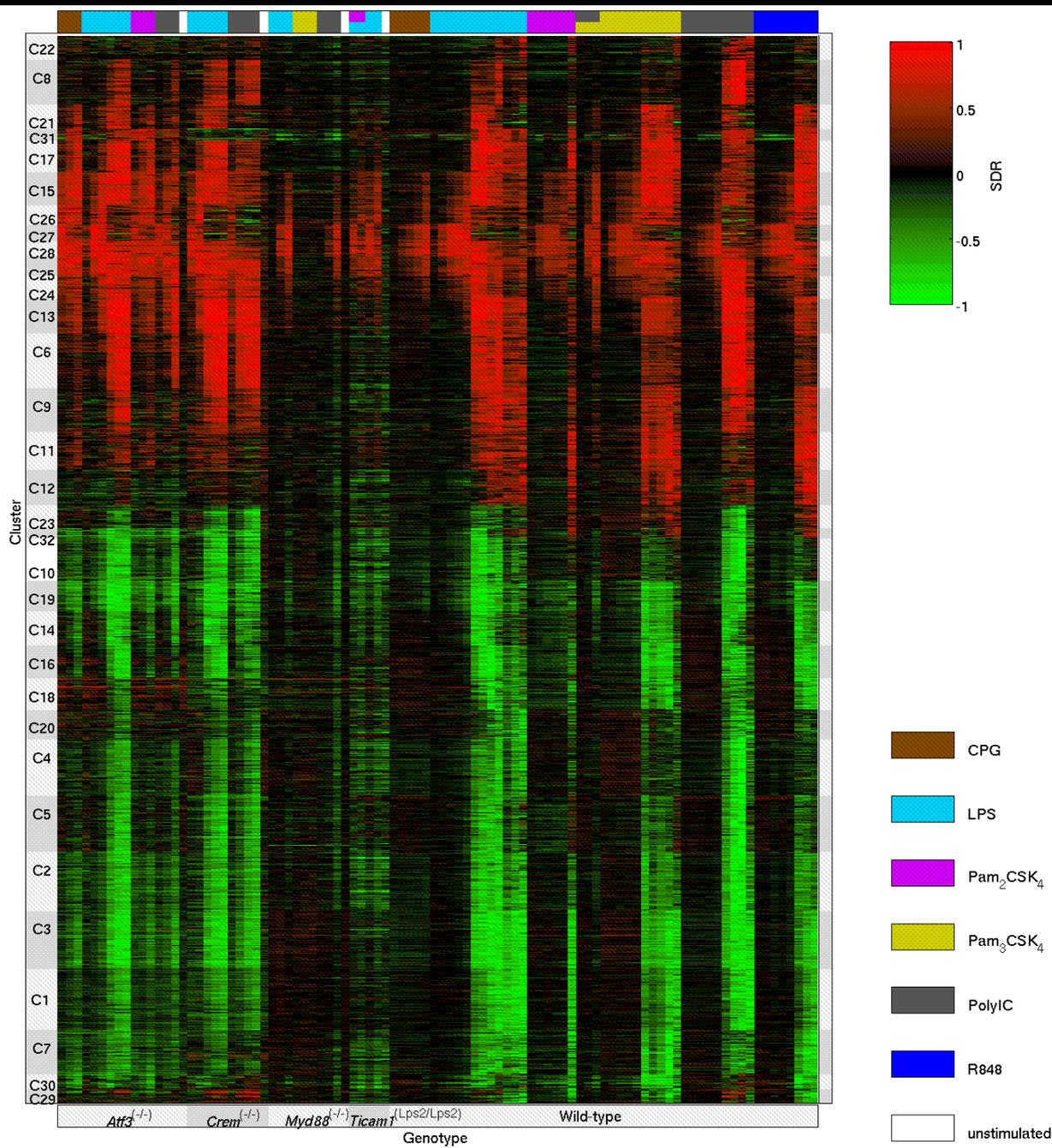
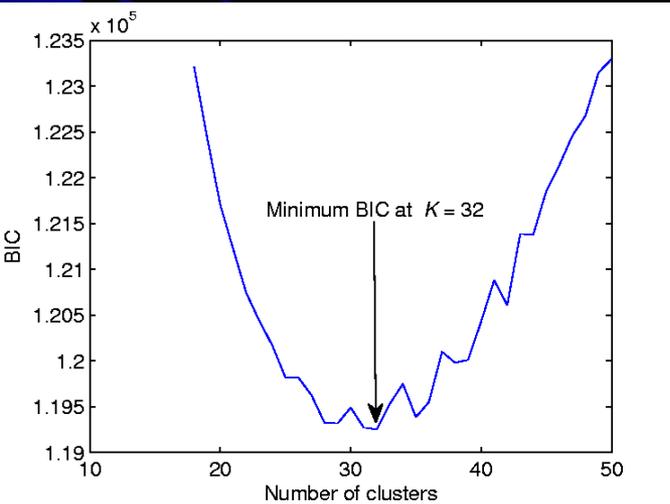
- 130 microarray experiments (253 arrays)
  - Seven mouse strains (WT,  $Ahr^{(-/-)}$ ,  $Atf3^{(-/-)}$ ,  $Crem^{(-/-)}$ ,  $Cebpd^{(-/-)}$ ,  $Myd88^{(-/-)}$ ,  $Ticam1^{(Lps2/Lps2)}$ )
  - Combinations of six stimuli (LPS, Pam<sub>3</sub>CSK<sub>4</sub>, Pam<sub>2</sub>CSK<sub>4</sub>, poly I:C, CpG, R848, T091317)
  - Time courses out to 8 hours (24 for LPS)
- Mouse genome promoters (UCSC)
- TRANSFAC Professional 10.3
- Curated list of ~1800 human TFs



Number of differentially expressed genes vs. elapsed time (by stimulus)

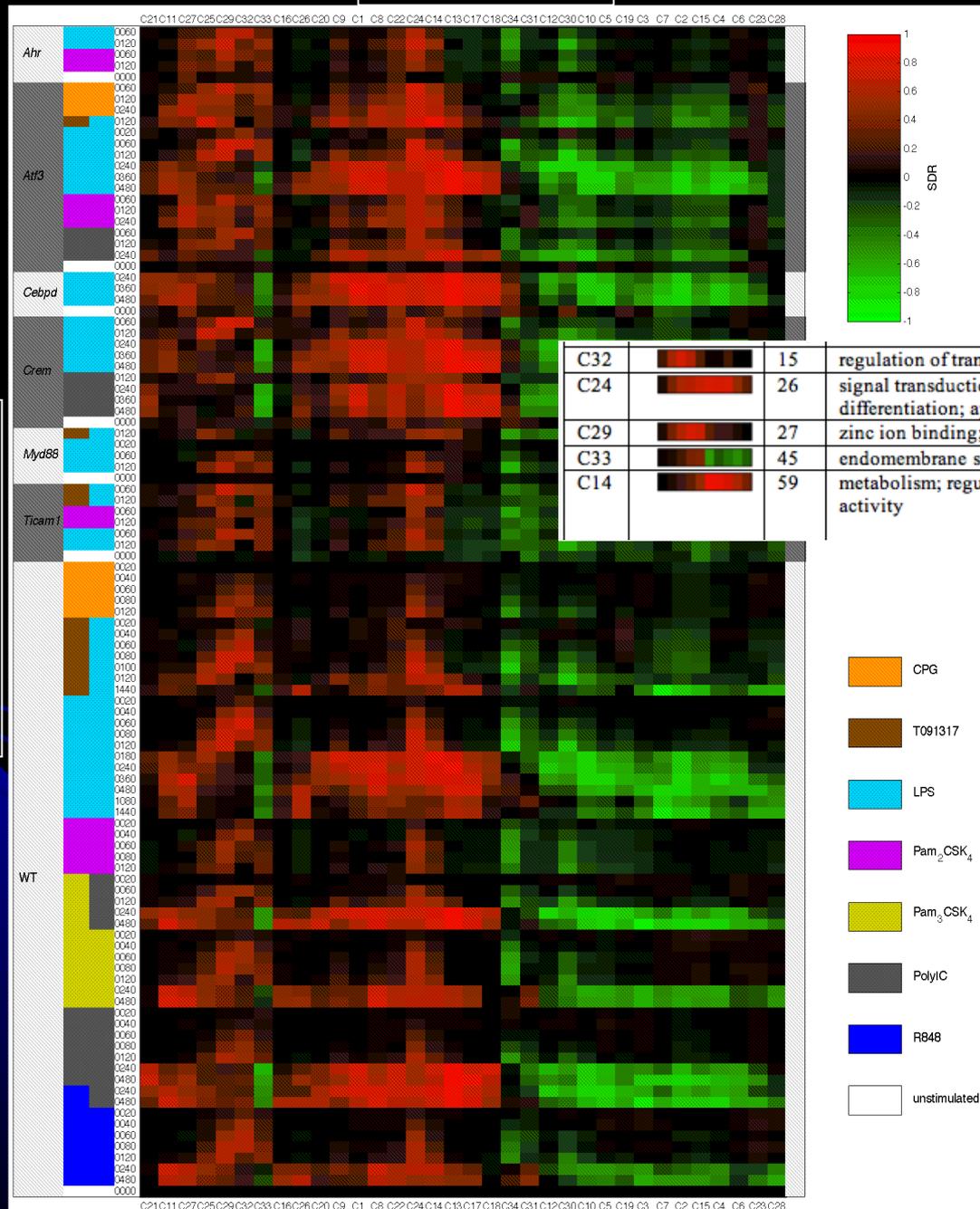
Total number: 2,562

# clustering of gene expression profiles



34 Clusters

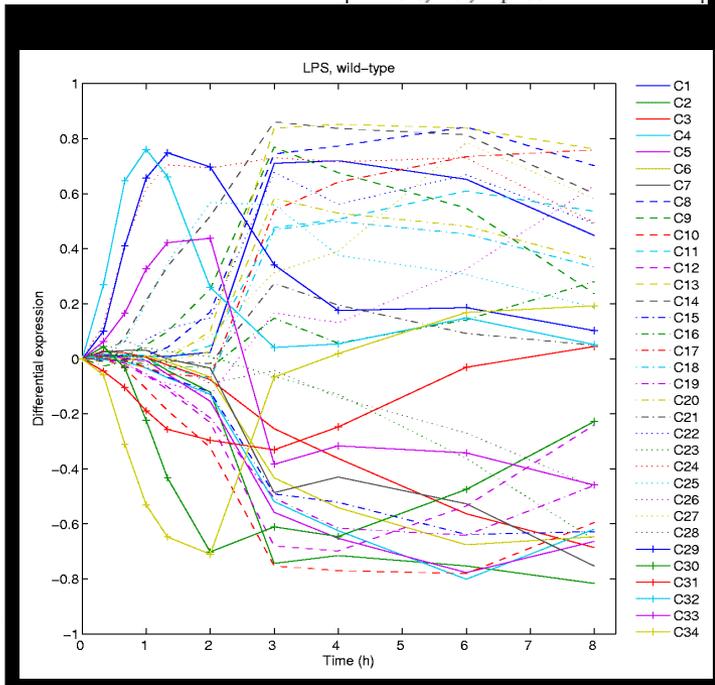
130 Experiments



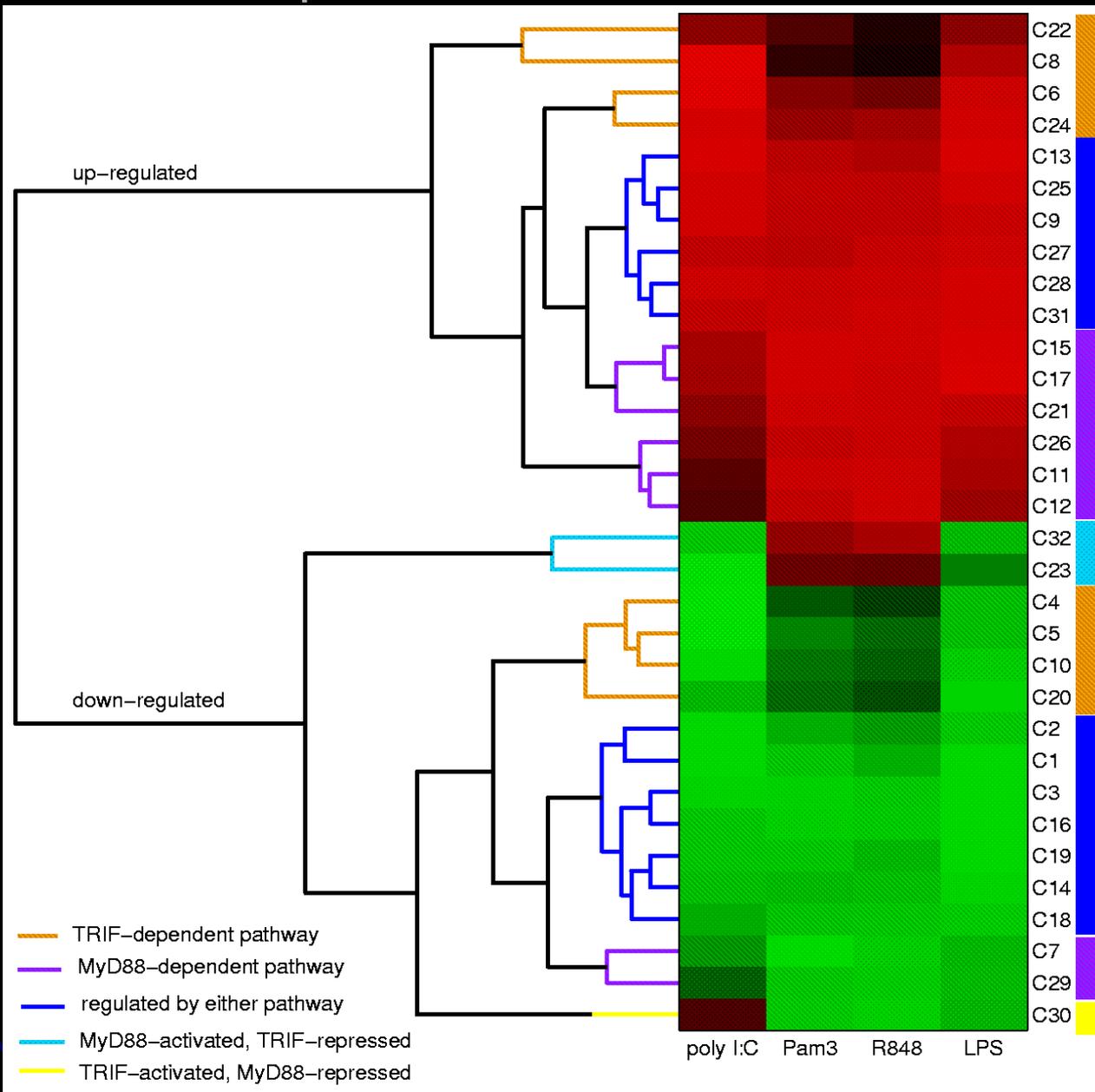
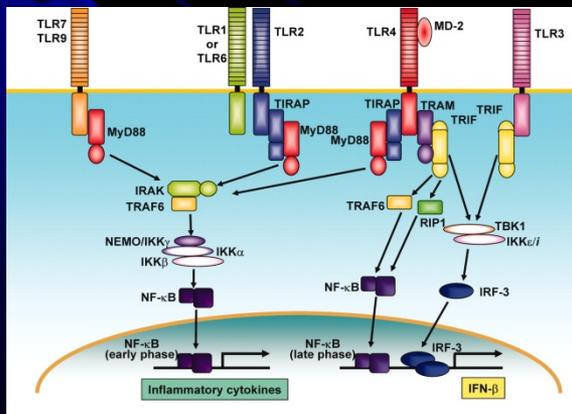
Cluster summarized expression profiles

C32		15	regulation of transcription; response to stress	Btg2, Ewsr1, Fos, Hipk1, Tgif
C24		26	signal transduction; regulation of transcription; cell differentiation; apoptosis; cytokine activity	Atf3, Ccrn41, Etv3, Irf1, Junb, Klf6, Nfkbiz, Prdm1
C29		27	zinc ion binding; regulation of transcription	Egr1, Egr2, Egr3, Maff
C33		45	endomembrane system	Id2, Phf19
C14		59	metabolism; regulation of transcription; cytokine activity	Arid5a, Bcl3, Hivep1, Hivep2, Klf7, Lass6, Nfil3, Nfkb1, Nfkb2, Nfkbie, Rel, Zfp263

- CPG
- T091317
- LPS
- Pam<sub>2</sub>CSK<sub>4</sub>
- Pam<sub>3</sub>CSK<sub>4</sub>
- PolyIC
- R84B
- unstimulated

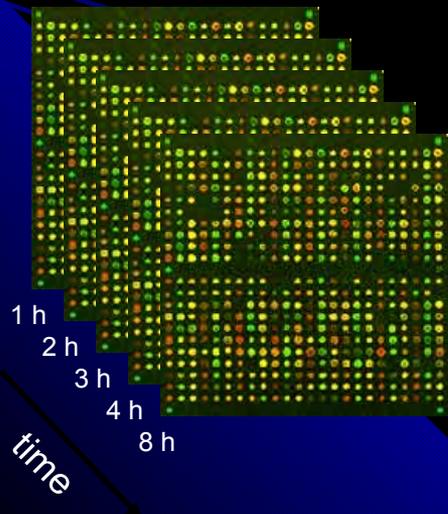


# Stimulus-specific cluster responses

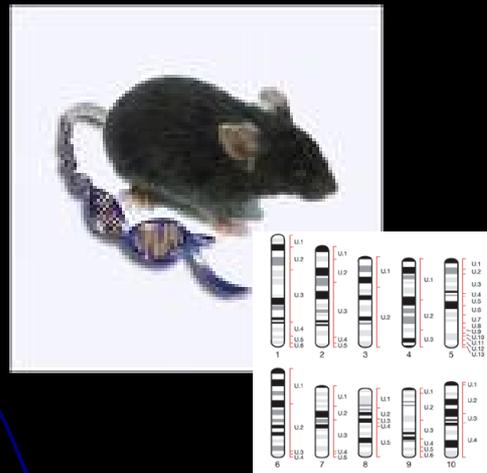


In conjunction with transcription factor binding site prediction, we can use the *timing* of expression to identify induced transcription factors that are associated with downstream groups of genes that they regulate.

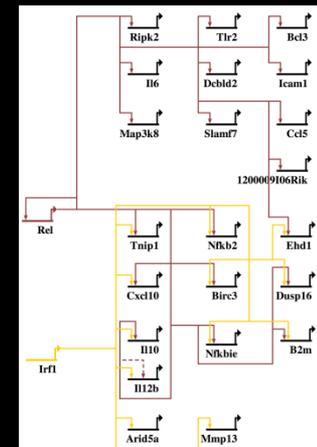
Expression *dynamics*



Genomic sequence data

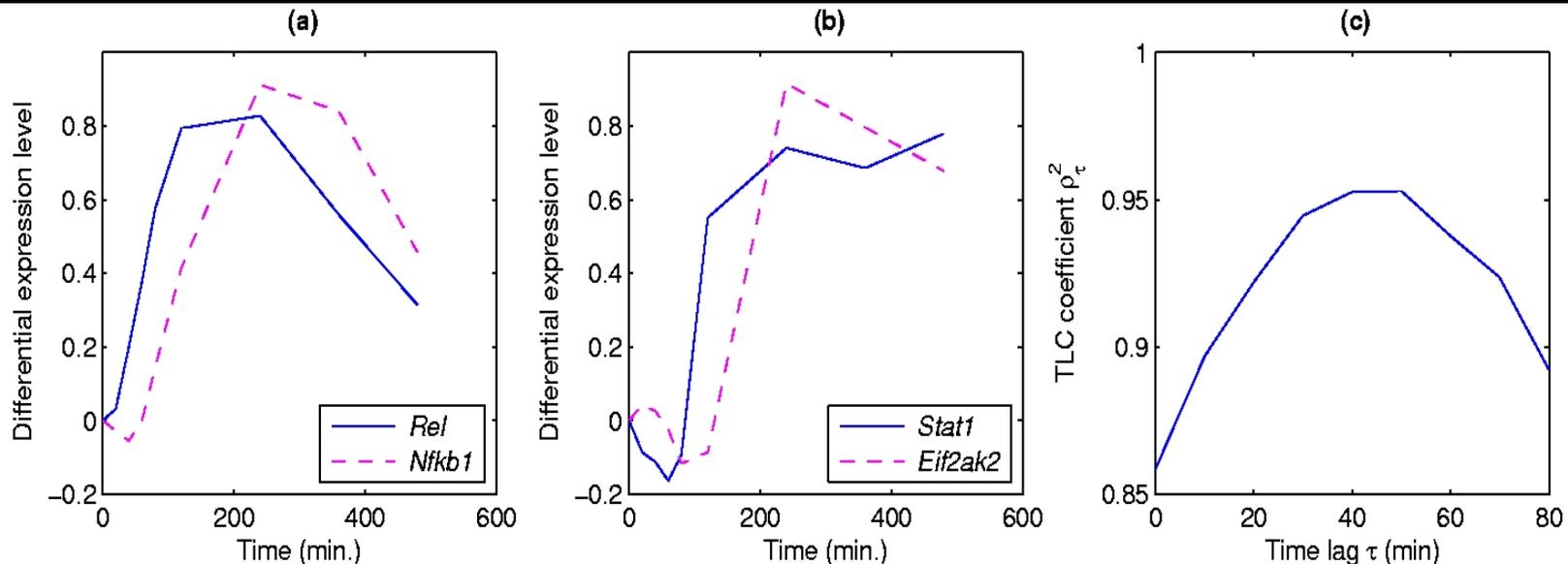


Transcriptional network



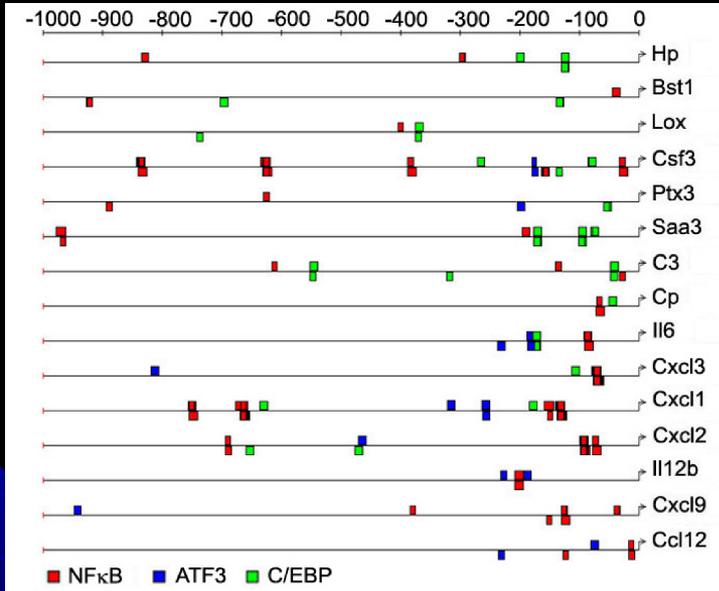
# Components of the gene-gene transcriptional time delay

- mRNA half-life of TF gene
- Translation and folding of TF protein
- Diffusion of TF back into nucleus
- Turnover rate (half-life) for TF protein
- Transcriptional elongation of target gene

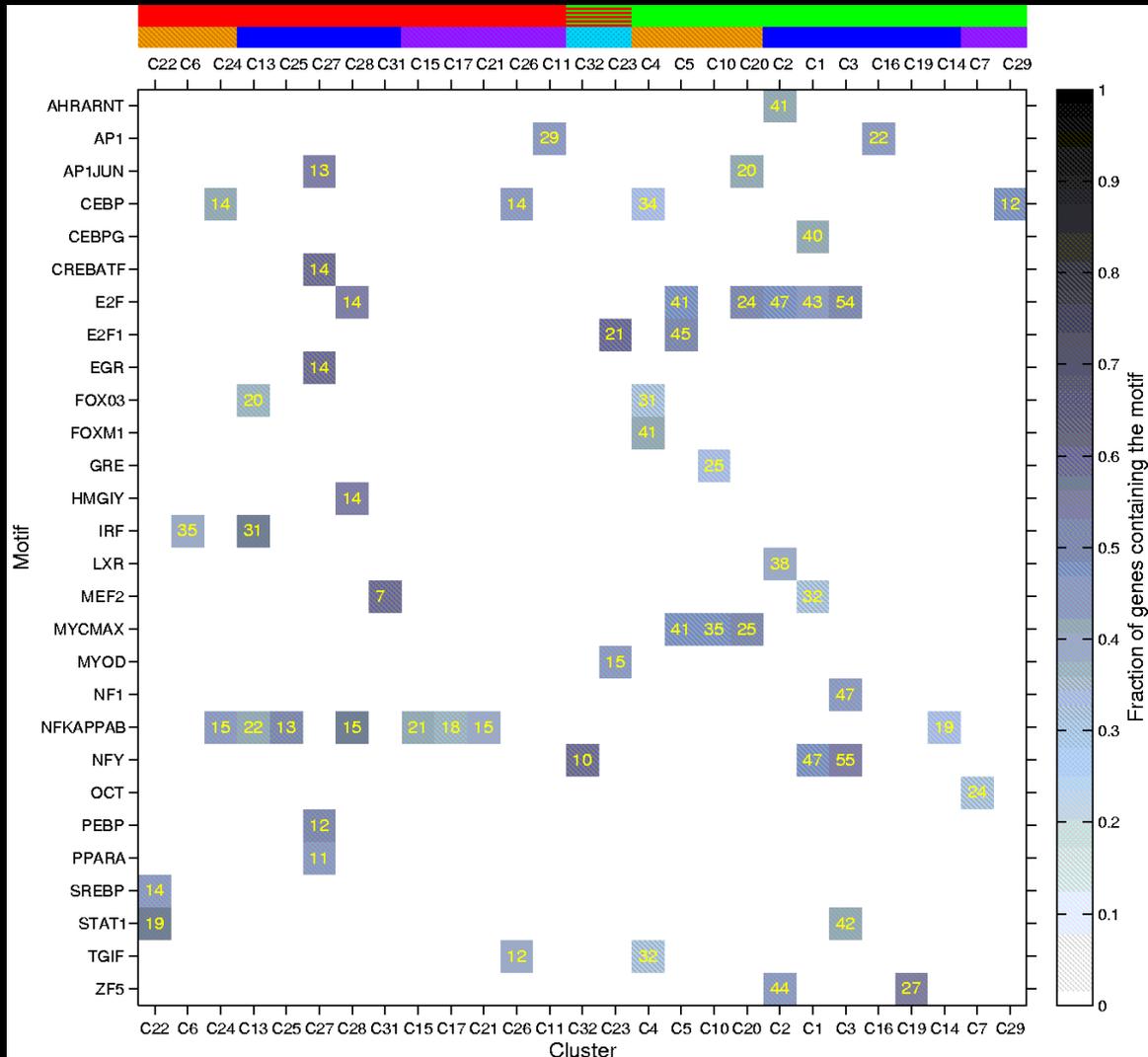


# Scan for TF binding motif matches

Scanning:



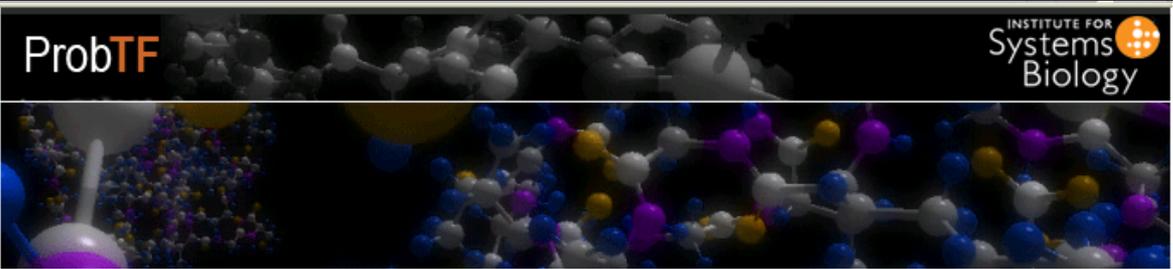
Motif enrichment matrix:



TF recognition site motif:



# Probabilistic Framework for Transcription Factor Binding Site Prediction



[About](#)   [FAQ](#)   [Contact](#)   [Acknowledgements](#)   [Changes](#)

This is a web server that enables the analysis of DNA sequences using **mouse-specific** position weight matrices from the [TRANSFAC™](#) database. Help on using this server can be found by clicking on the linked features within the page and using the [FAQ](#).

Upload [sequence in FASTA](#) format

Limit: 5K base pairs

---

Upload [evidence scores](#) [Optional]

The number of evidence scores MUST be the same length as the number of basepairs in the uploaded sequence file

---

Select the [order of background model](#) to use

0  
  1  
  2  
  3

Select transcription factor matrices to scan with

Up to 10 may be selected.  
[Hold down the Ctrl key to select multiple factors]

Ahr

Ap1

Ap2a

Arnt

Atf2

Bach1

Bach2

Cart1

Cebpa

Cebpb

Chx10

Creb1

Ddit3

E2f

E2f1

Press  to submit information, or  to reset fields.

www.probtff.org

motif model  $\theta^{(\pi_i)}$  at location  $a_i$   
 round model

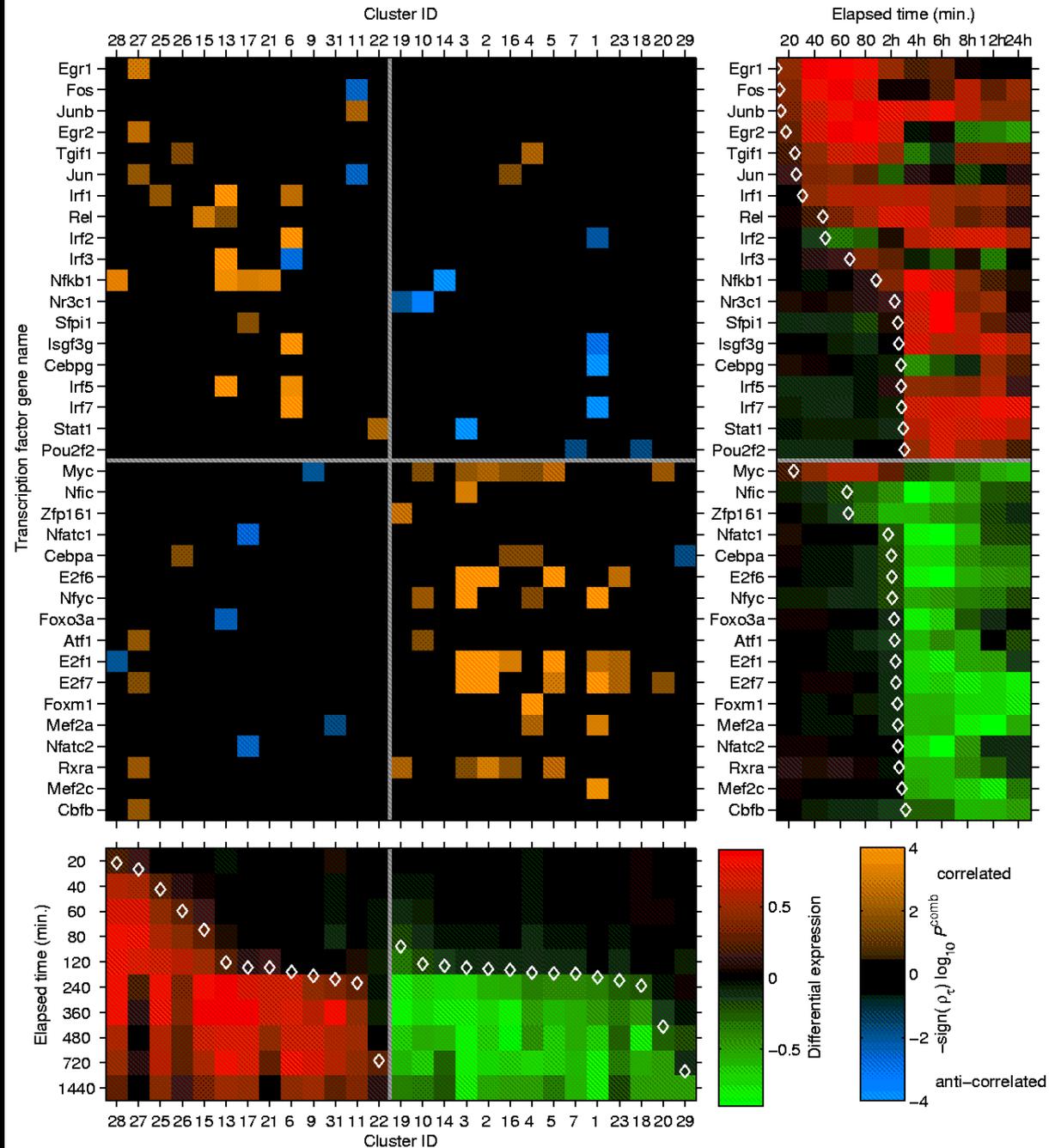
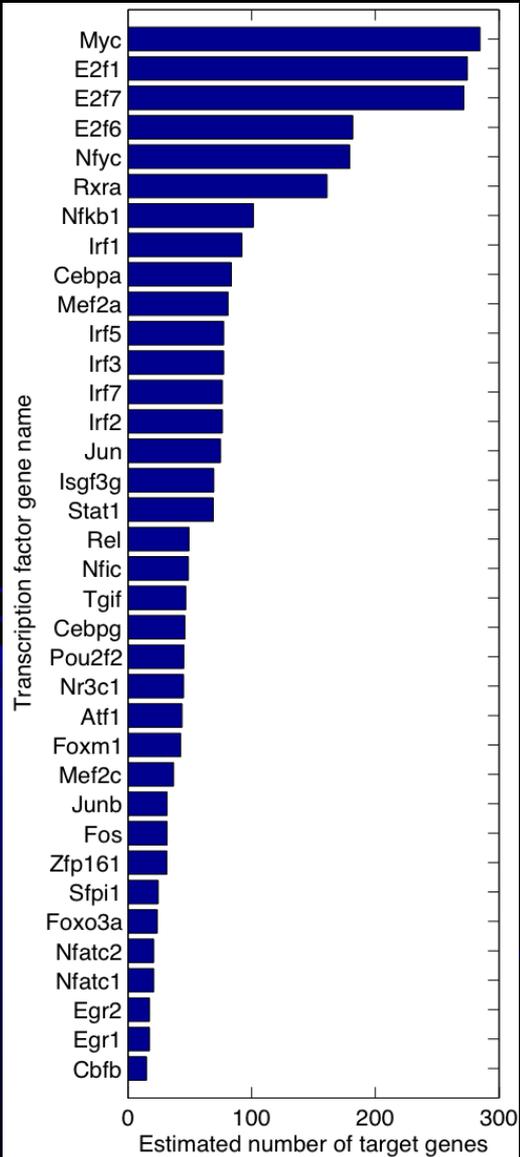
## Data Fusion

$$P(S, \mathcal{D} | A, \pi, \Theta, \phi) = P(S | A, \pi, \Theta, \phi) P(\mathcal{D} | A, \pi)$$

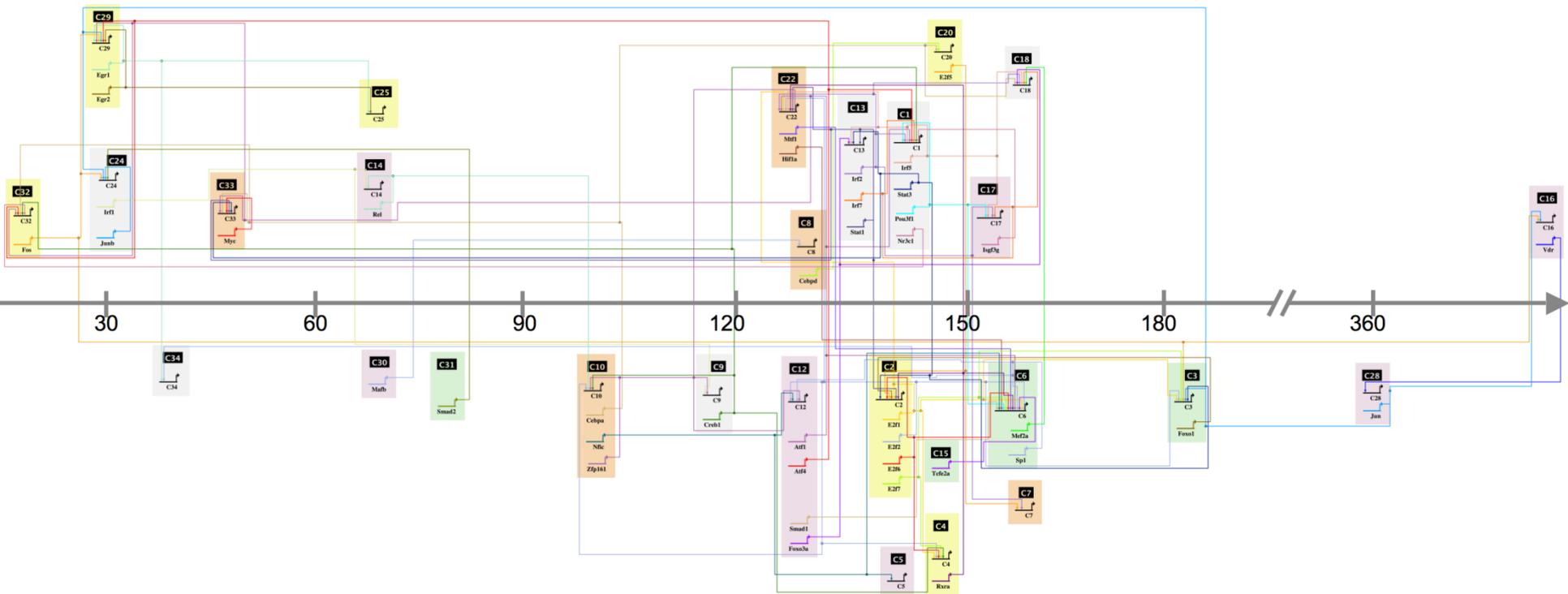
$$P(S, \mathcal{D}) = \frac{P(S, \mathcal{D} | A, \pi) P(A, \pi)}{P(S, \mathcal{D})}$$

$$= \frac{P(S | A, \pi) P(\mathcal{D} | A, \pi) P(A, \pi)}{P(S, \mathcal{D})}$$

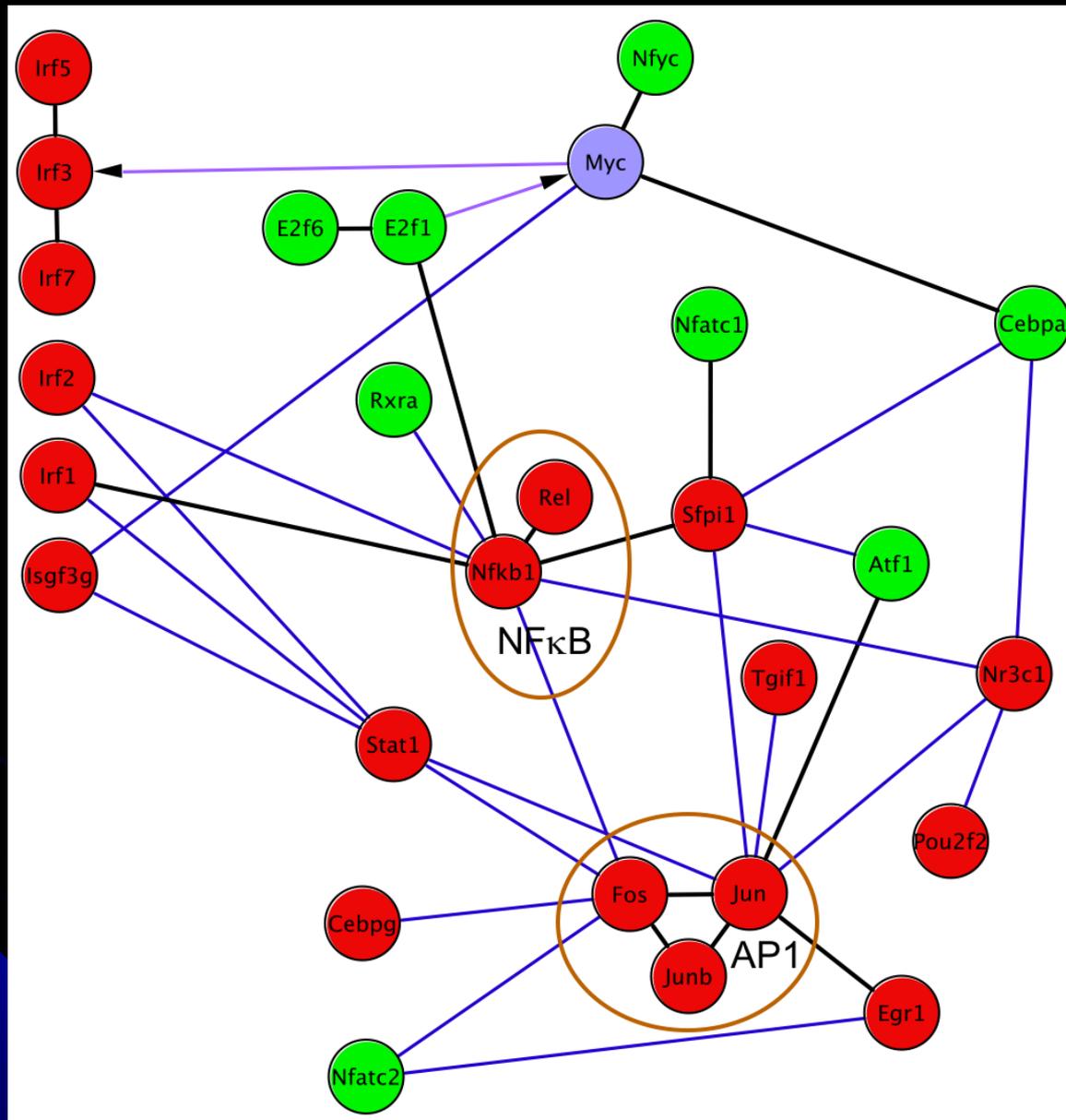
# Map network



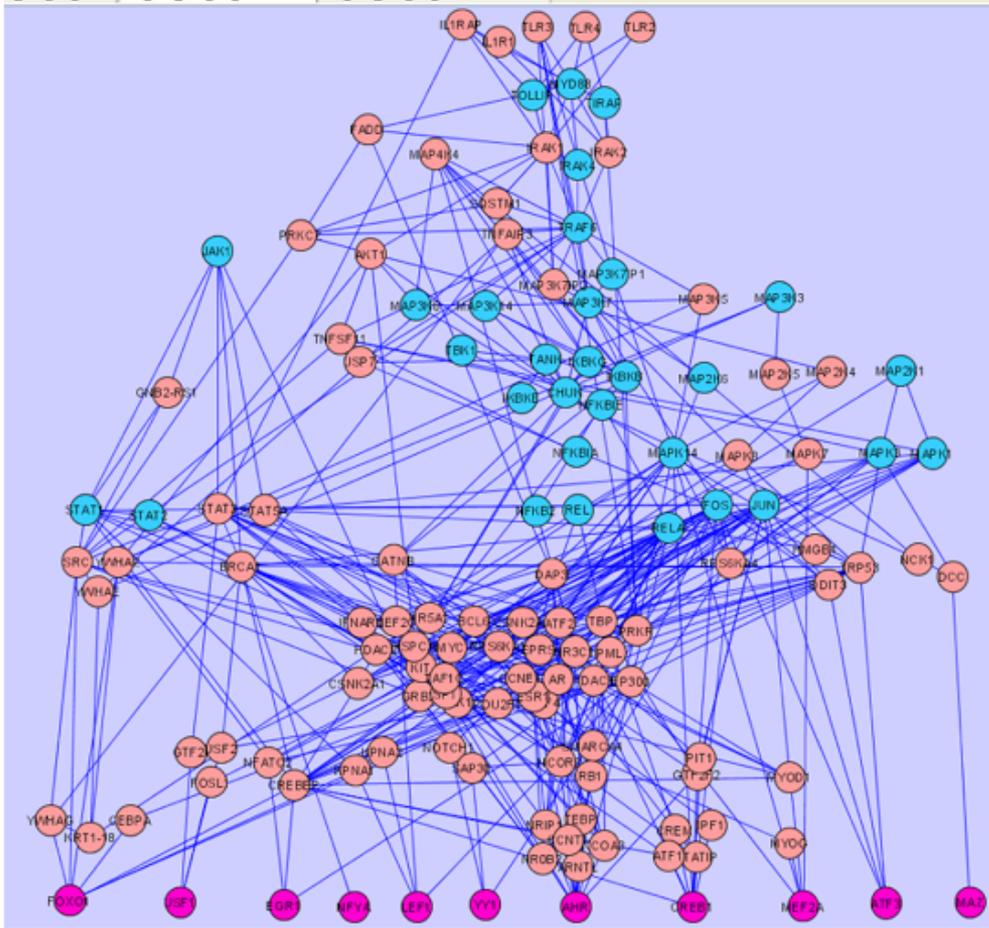
# TF-to-cluster association network



# TF interactions



# TLR signaling pathway and downstream predicted transcription factors: using protein-protein interactions



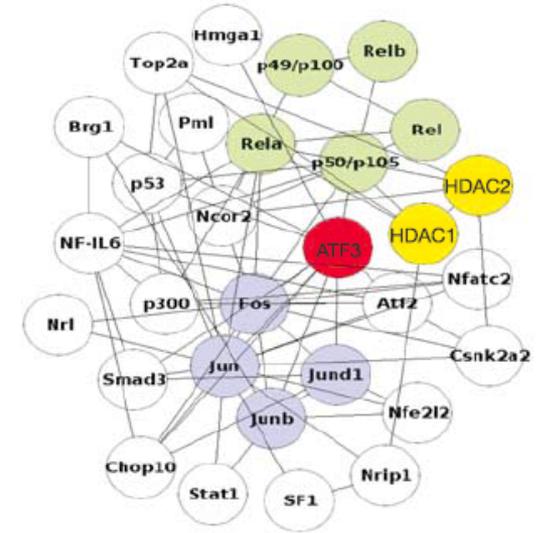
Known  
TLR  
Signaling  
pathway

(Known TFs)

Link to  
Upstream prot.

Link to  
Known TFs

Predicted TFs



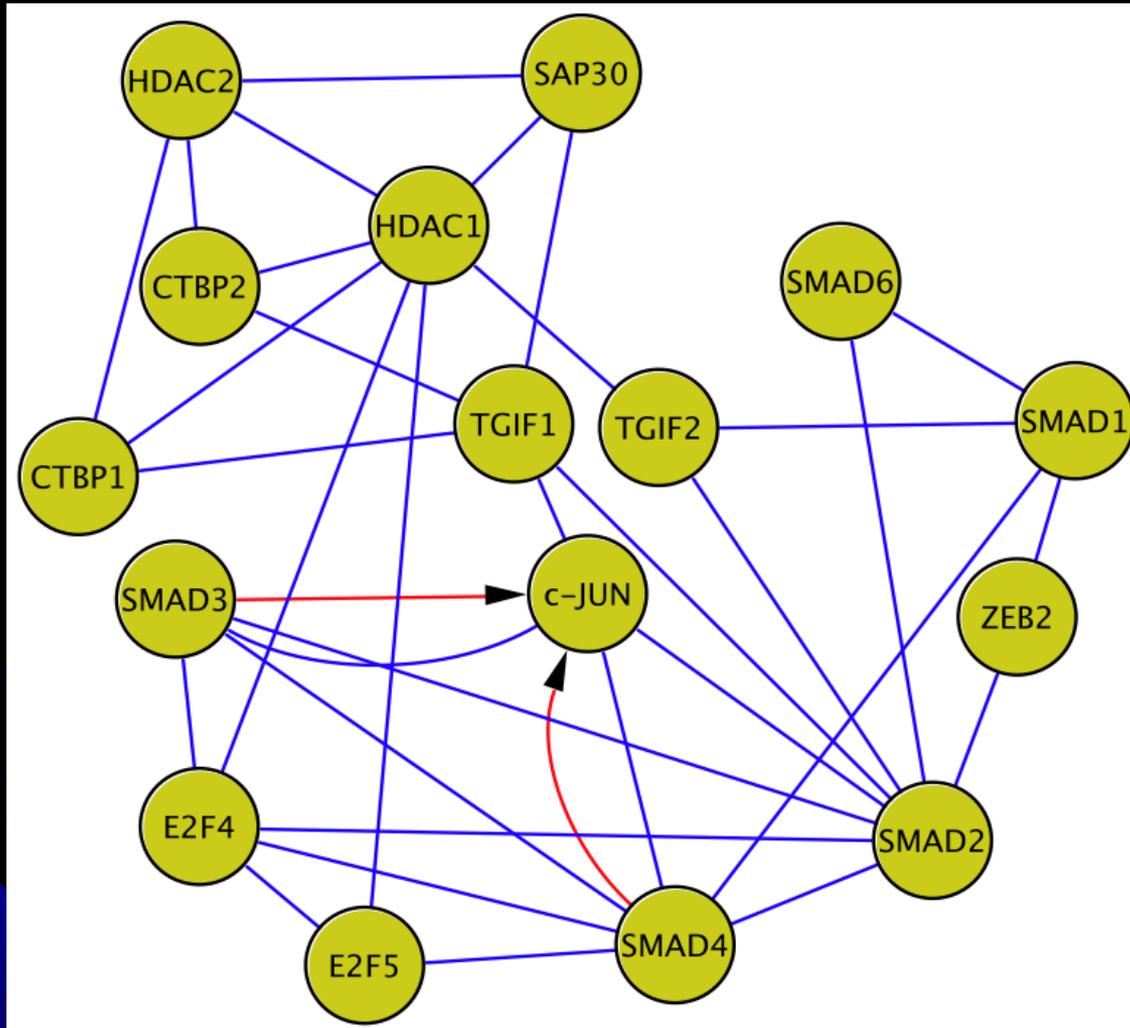
ATF3 (red) is predicted to interact with a number of TFs, including members of the AP1 and NF- $\kappa$ B TF complexes.

# ChIP-on-chip validation

**Table 2.** Validation of transcription factor-to-cluster associations using ChIP-on-chip

TF	Matrix	Stim.	Clust	Time Points	In Clust	On Chip	Bound	P-Value
NFκB/p50	NFKB_Q6	LPS	C13	1 h, 2 h	64	23	18	$1.1 \times 10^{-3}$
NFκB/p50	NFKB_Q6	LPS	C17	1 h, 2 h	58	20	11	$2.5 \times 10^{-1}$
NFκB/p50	NFKAPPAB_01	LPS	C28	1 h, 2 h	28	21	20	$1.1 \times 10^{-6}$
IRF1	IRF_Q6_01	LPS	C13	1 h, 2 h, 4 h	64	23	18	$2.3 \times 10^{-3}$
IRF1	IRF_Q6_01	LPS	C25	1 h, 2 h, 4 h	37	22	18	$8.8 \times 10^{-4}$

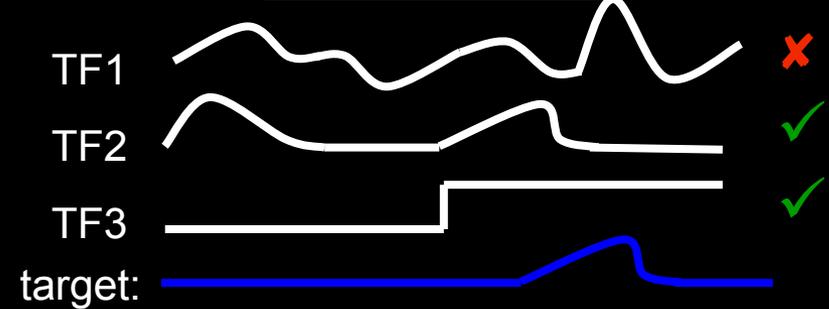
# TGIF enhanceosome



## Constrain Possible Regulators

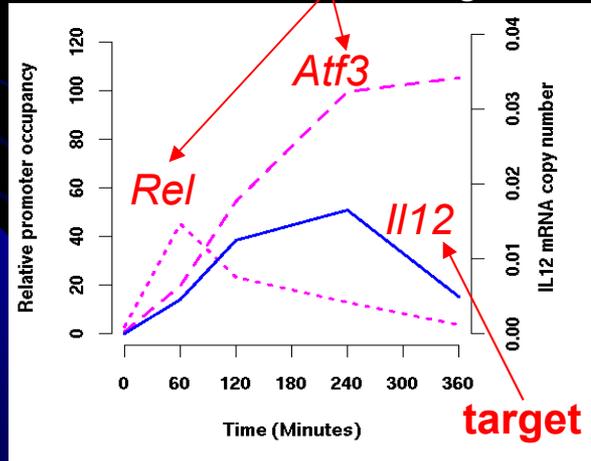
- Gene expression (timing, etc.)
- Promoter sequence analysis
- Immunoprecipitation data

## Kinetic Selection



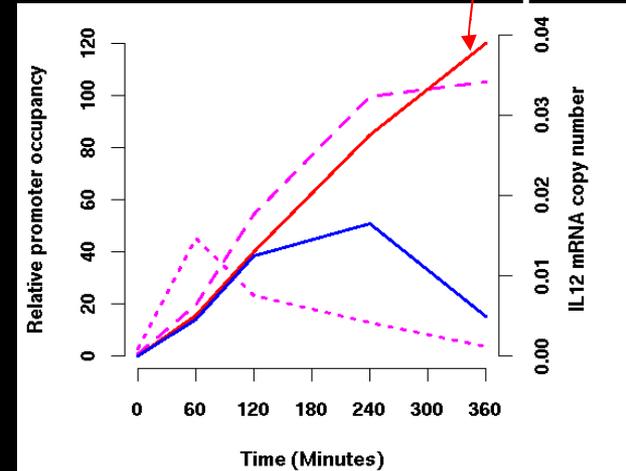
## Parameter Fitting

one coefficient for each regulator:



## Prediction

*Atf3*<sup>(-/-)</sup>  
prediction



(Gilchrist *et al.* Nature. 2006; 441(7090):173.)

V. Thorsson

$$\tau \frac{d(IL12)}{dt} = -IL12 + g(\beta)$$

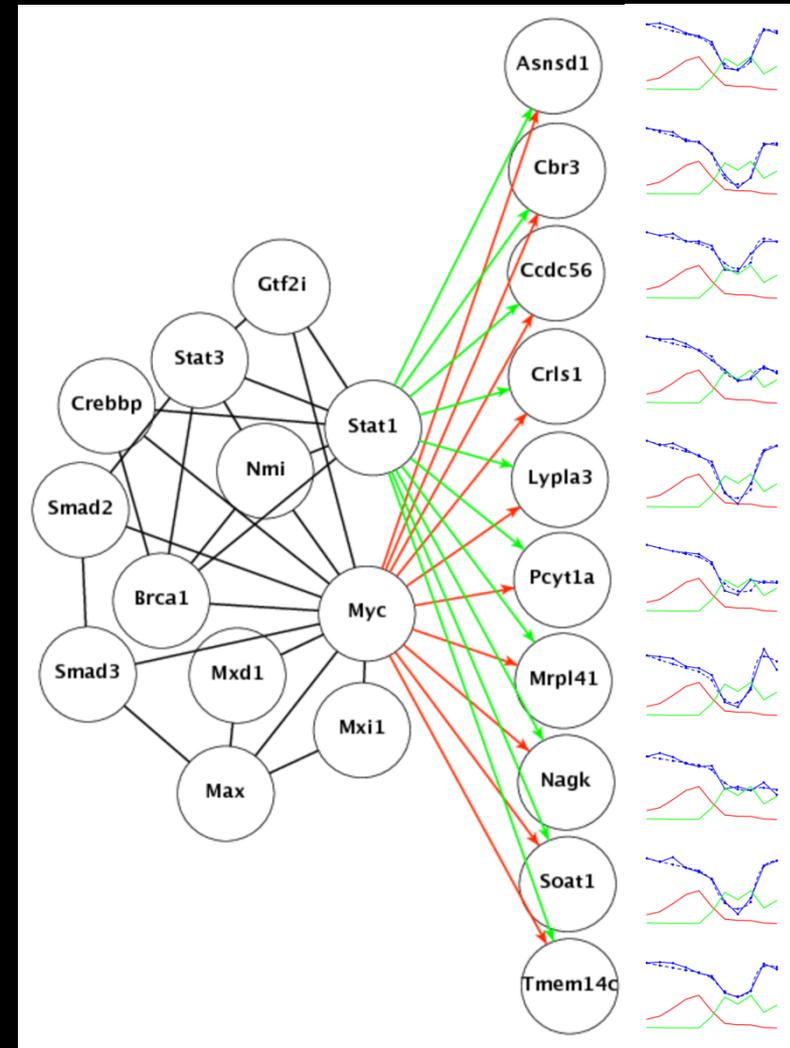
# Network Predictions

## Macrophage

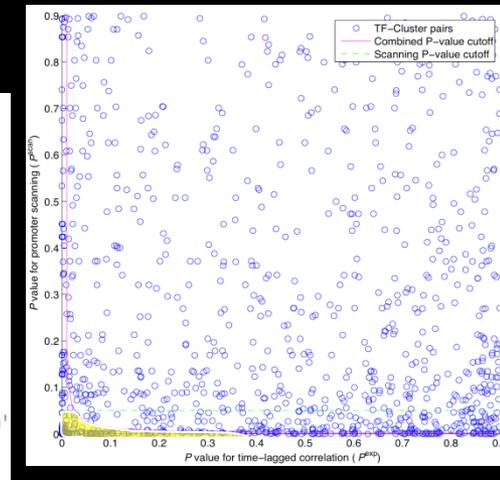
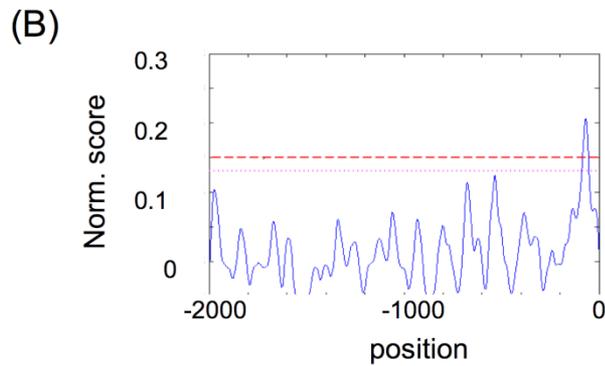
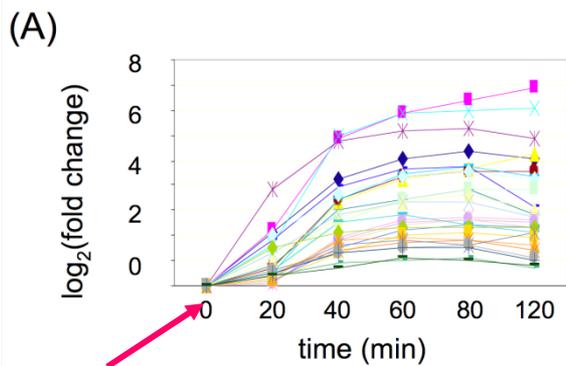
Predicted network contains co-regulated gene groups with shared:

- Binding sites predictions, for single motifs or pairs of motifs
- Kinetics
- Predicted activators and repressors
- Predicted co-factors
- Function (GO categories)

Predictions can be tested with directed binding and/or functional assays

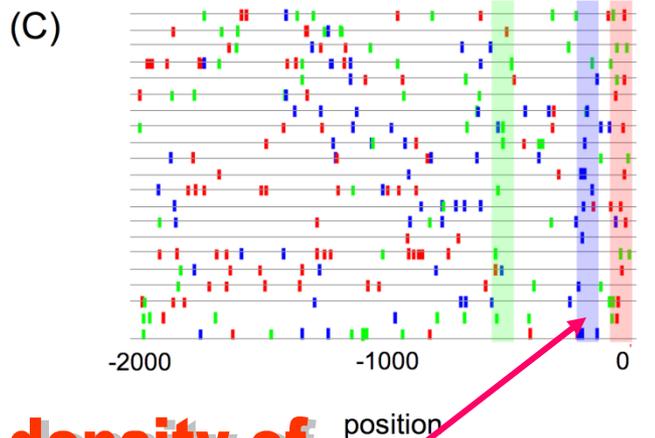






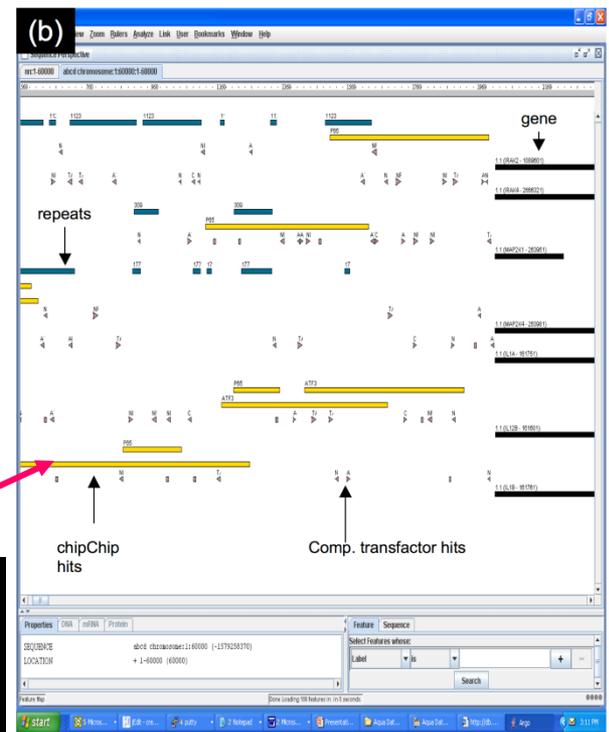
**Microarray time-course data**

**data**



**Combined density of predicted TF binding sites**

**ChIP-chip data**



# Innate Immune Database (IIDB)

- Genomic annotations and *cis*-regulatory element predictions for immune-related genes.
- Web-based software tool for querying and visualization.
- Display expression time-course data (and clusters)
- Graphical visualization of genomic annotations combining many different data types

[v](#)
[i](#)
[d](#)
[e](#)
[o](#)
[f](#)
[T](#)
[F](#)
[B](#)
[S](#)
[e](#)
[n](#)
[s](#)
[i](#)
[n](#)
[g](#)
[i](#)
[n](#)
[g](#)
[a](#)
[n](#)
[n](#)
[o](#)
[t](#)
[a](#)
[n](#)
[n](#)
[o](#)
[t](#)
[s](#)

The screenshot displays the Innate Immune Database (IIDB) homepage. At the top, it identifies the database as part of the Institute for Systems Biology. The main navigation bar includes links for 'IIDB Home', 'About IIDB', 'How to Use IIDB', 'IIDB Tutorial', 'Site Map', and 'Questions/Contact'. The central area is divided into three columns: 'Your Favorite Gene' (with a search box and 'List of Annotated Genes'), 'Computationally Predicted Co-regulated Genes' (with sub-sections for 'ISB Co-regulated Gene Clusters' and 'LPS Responsive Gene Clusters'), and 'Advanced Analyses' (with options like 'Search for TFBS', 'Search Genes for Shared TFBS', 'Create Gene Groups by GO-Annotation', 'Get a Sequence File', and 'Explore ChIP-chip Data'). A search bar is located at the top right. A red watermark 'db.systemsbiology.net/IIDB' is overlaid across the middle of the page, with red arrows pointing to the search bar, the navigation bar, and the 'Your Favorite Gene' section.

# Cytoscape

**Network Metadata for Yeast Network (galFiltered.gml)**

Data Label	Value
Title	Yeast Network (galFiltered.gml)
Identifier	N/A
Source	http://chianti.ucsd.edu/idekerlab/
Type	Protein-Protein Interaction
Format	Cytoscape-XGMML
Date	2006-05-31 15:02:11

**Node Attribute Browser**

ID	GO Aliases	GO Biological Process	GO Cellular Component	GO Molecular Function
YNL145W	[MFA2]	[G-protein coupled...	[extracellular region]	[binding, mating p...
YJL159W	[CCW7, HSP150, ...]	[cell organization a...	[cell, cell part, cell ...]	[structural constitu...
YJL157C	[FAR1]	[G-protein coupled...	[cell]	[binding, mating p...
YER111C	[ART1, SWI4]	[G1/S transition of ...]	[cell]	[intracellular membrane-bound organ...
YAL040C	[CLN3, DAF1, FUN...	[G1/S transition of ...]	[cell]	[intracellular organelle]
YJL194W	[CDC6]	[DNA metabolism, ...]	[cell]	[intracellular part]
YMR043W	[FUN80, MCM1]	[DNA metabolism, ...]	[cell]	[mating protection]

Visually Integrate  
gene expression,  
protein state, protein  
interactions, and  
protein class (ontology)

Analysis plug-in  
modules

Implemented in Java

(networks, attributes, network metadata, etc.)

<http://www.cytoscape.org/>

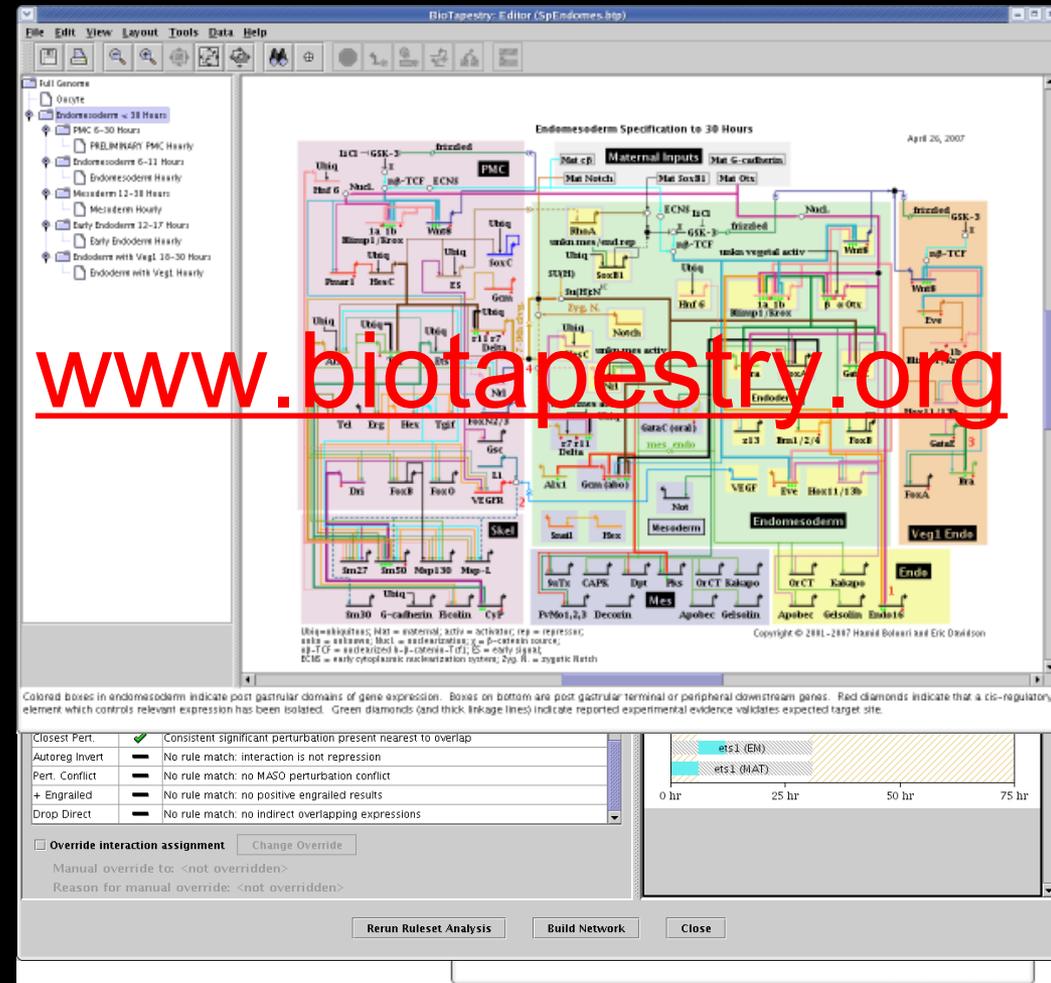
# BioTapestry

Graphical application for building & visualising gene regulatory networks

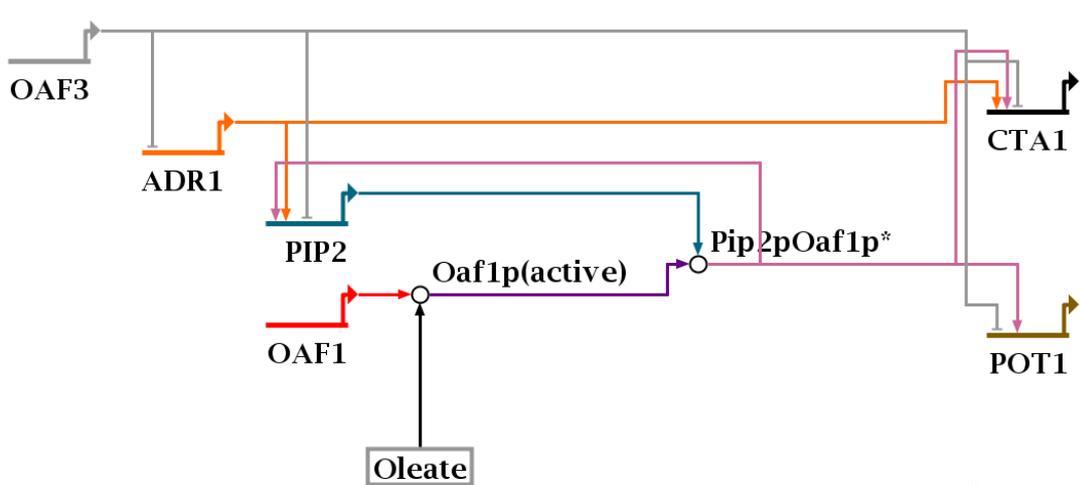
Hierarchical network model for spatially and temporally complex network activation programs

View network activity over time, based on time-course expression

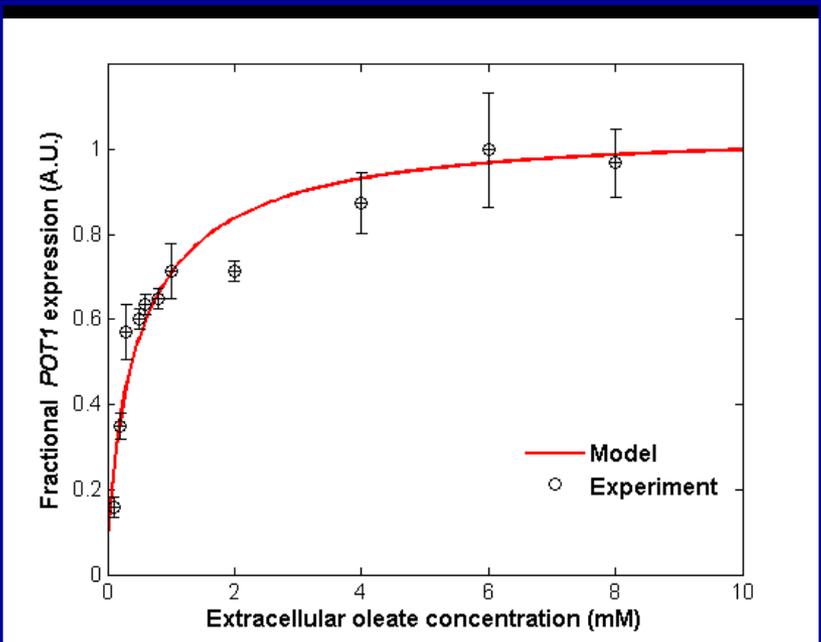
- Build networks from high-throughput data using worksheet feature (under development)



[www.biotapestry.org](http://www.biotapestry.org)



# modeling the oleate-response transcriptional network

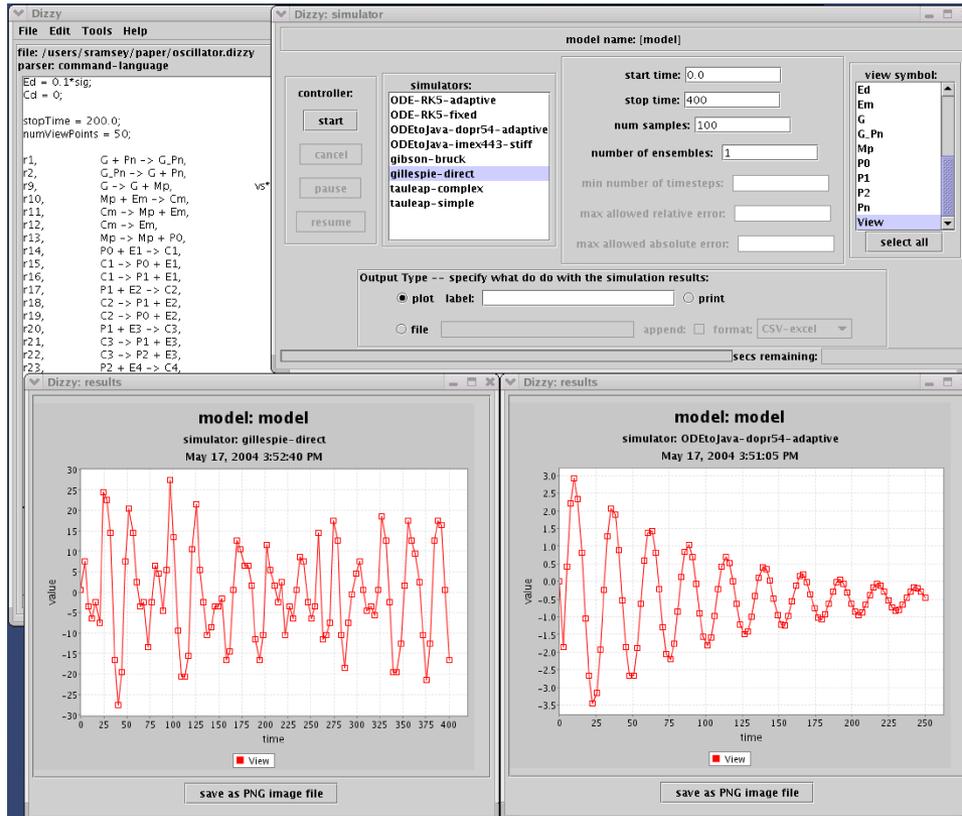
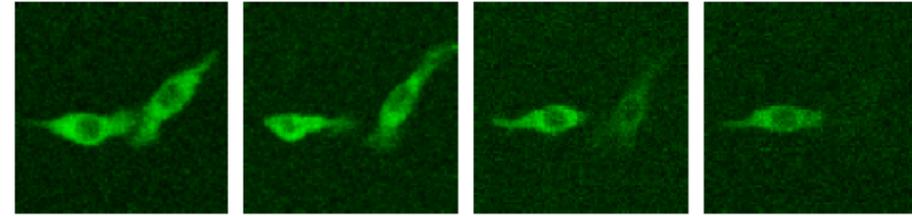


Steady-state dose-response

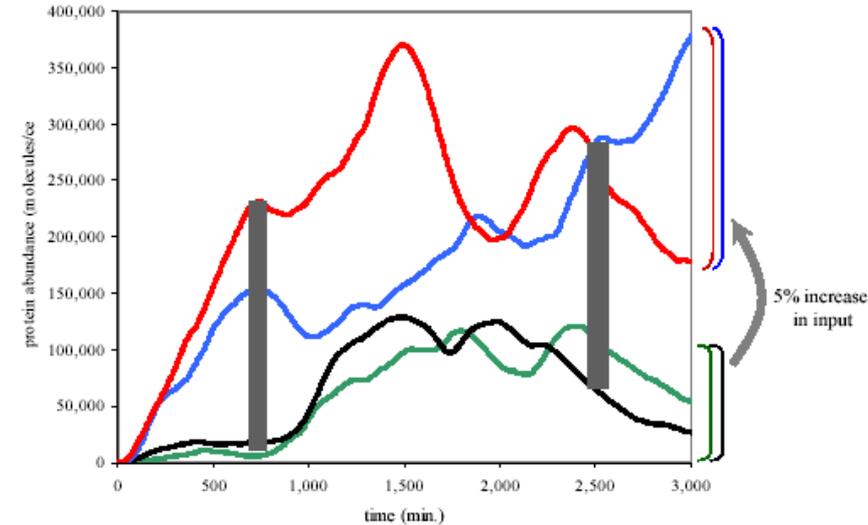
$\pi_o(t) = \frac{Fp_o(t)O_{ic}}{K_{M,o} + O_{ic}}$	Activation of Oaf1p by intracellular oleate ( $\pi_o$ denotes activated Oaf1p)
$h(t) = \left( \frac{Q}{2K_{D,h}} \left( 1 + \frac{Q}{K_{D,h}} (\pi_o + p_p) - \sqrt{ \left( 1 + \frac{Q}{K_{D,h}} (\pi_o + p_p) \right)^2 - 4 \left( \frac{Q}{K_{D,h}} \right) \pi_o p_p } \right) \right)$	Heterodimerization of activated Oaf1p with Pip2p ( $h(t)$ represents the concentration of Oaf1p-Pip2p heterodimer)
$S = \frac{O_{ic}}{K_{M,s} + O_{ic}}$	Activation of transcriptional response of ADR1, OAF1, and OAF3 by oleate.
$\frac{dr_\gamma(t)}{dt} = k_{i,\gamma} f_{r,\gamma}(t - \tau_b) - k_{d,\gamma} r_\gamma(t), \quad \forall \gamma \in \{a,c,o,v,p,f\}$	Gene expression
$\frac{dp_\gamma(t)}{dt} = k_{i,p,\gamma} r_\gamma(t - \tau_p) - k_{d,p,\gamma} p_\gamma(t), \quad \forall \gamma \in \{a,c,o,v,p,f\}$	Protein expression
$f_{r,a} = \left( \frac{1}{r_{x,a}} \right) [r_{g,a} + S(r_{o,a} - r_{g,a})]$	Fractional activity of ADR1 gene
$f_{r,c}(t) = \left( \frac{1}{r_{x,c}} \right) \left[ r_{g,c} + \left\{ w_p \left( \frac{h(t)}{K_{M,h} + h(t)} \right) + w_a S \left( \frac{p_a(t)}{K_{M,a} + p_a(t)} \right) \right\} (r_{o,c} - r_{g,c}) \right]$	Fractional activity of CTA1 gene
$f_{r,o} = \left( \frac{1}{r_{x,o}} \right) [r_{g,o} + S(r_{o,o} - r_{g,o})]$	Fractional activity of OAF1 gene
$f_{r,y} = \left( \frac{1}{r_{x,y}} \right) [r_{g,y} + S(r_{o,y} - r_{g,y})]$	Fractional activity of OAF3 gene
$f_{r,p}(t) = \left( \frac{1}{r_{x,p}} \right) \left[ r_{g,p} + \left\{ w_p \left( \frac{h(t)}{K_{M,h} + h(t)} \right) + w_a S \left( \frac{p_p(t)}{K_{M,a} + p_p(t)} \right) \right\} (r_{o,p} - r_{g,p}) \right]$	Fractional activity of PIP2 gene
$f_{r,f}(t) = \left( \frac{1}{r_{x,f}} \right) \left[ r_{g,f} + \left( \frac{h(t)}{K_{M,h} + h(t)} \right) (r_{o,p} - r_{g,p}) \right]$	Fractional activity of POT1 gene

# Network Simulation

## Stochastic Simulation (Dizzy)



The screenshot shows the Dizzy simulator interface. The top window is the 'Dizzy: simulator' control panel, which includes a list of chemical reactions (e.g.,  $G + Pn \rightarrow G, Pn$ ), simulation parameters (start time, stop time, number of samples), and a list of available simulators (ODE-RK5-adaptive, Gillespie-direct, etc.). The bottom window displays two 'Dizzy: results' plots. The left plot, titled 'model: model', shows a time series of a variable fluctuating between -30 and 30. The right plot, titled 'model: model', shows a time series of a variable fluctuating between -3.5 and 3.0. Both plots include a 'View' button and a 'save as PNG image file' button.



Individual macrophage cells show different levels of I $\kappa$ B $\alpha$ :GFP after stimulation by dsRNA (TLR3 ligand)

# Acknowledgments



Steve Ramsey

Vesteinn Thorsson

Alistair Rust

Bin Li

John Boyle

Sandy Klemm

Harri Lähdesmäki

Martin Korb

Matti Nykter

Sarah Killcoyne

Chris Cavnor

Bill Longabaugh

Antti Niemistö

Ricardo Vencio

**Aderem Group**

## Support

NIH/NIGMS R21 GM070600

NIH/NIGMS R01 GM072855

NIH/NIAID U54 AI54253

NIH/NIGMS P50 MO-76547