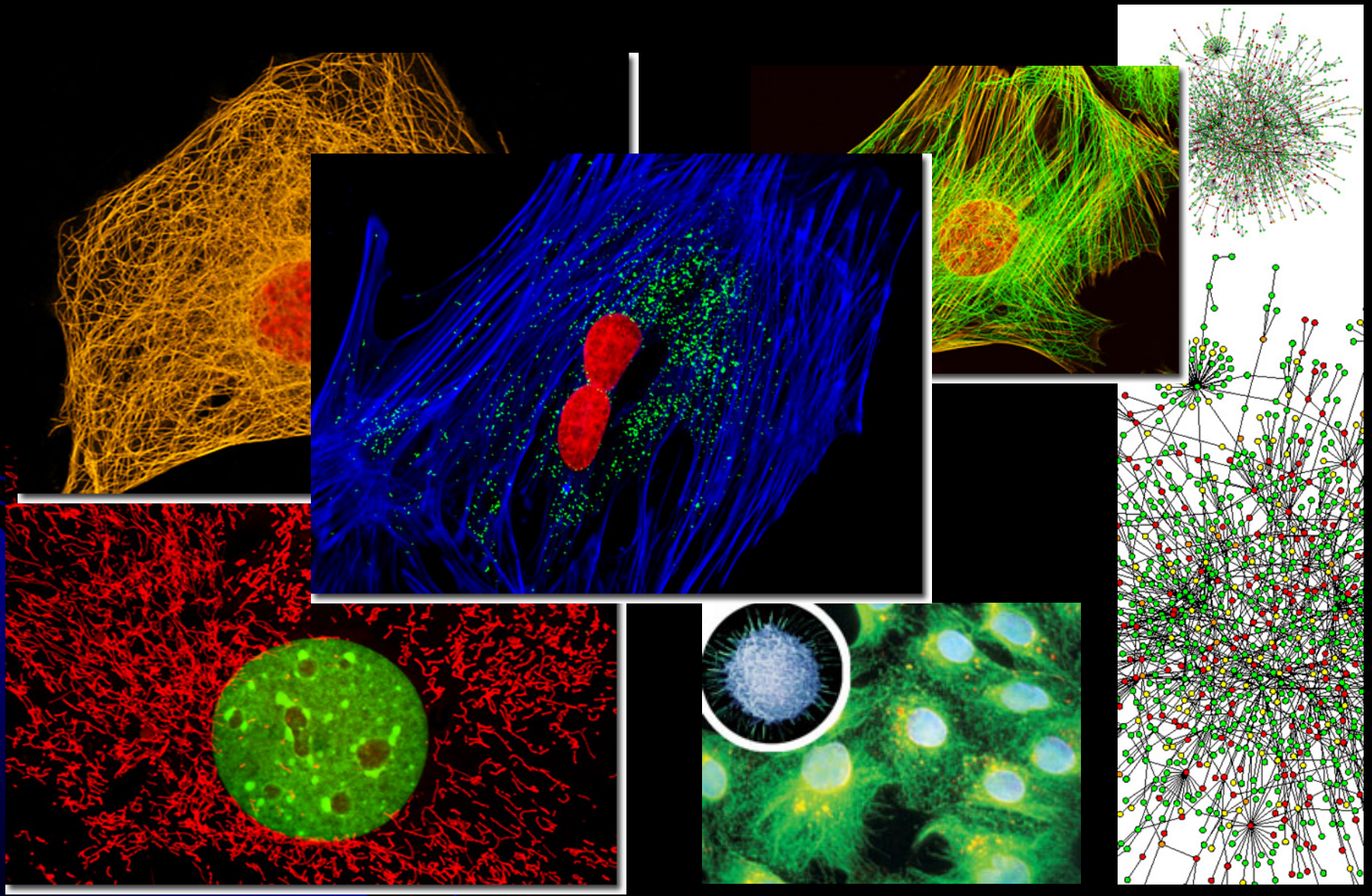
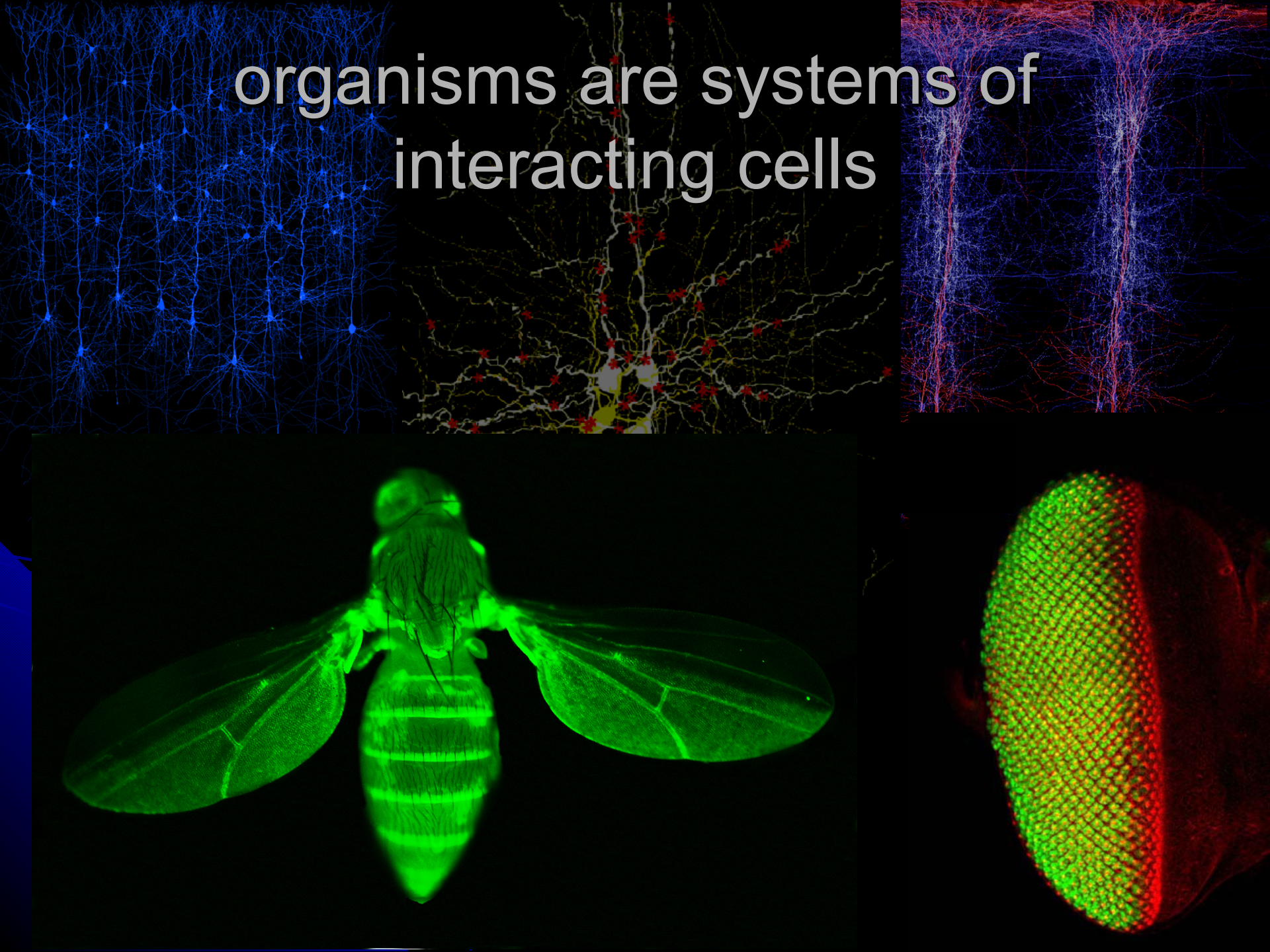
Modeling and Inference of Transcriptional Regulatory Networks

ilya shmulevich

cells are systems of interacting molecules



organisms are systems of interacting cells

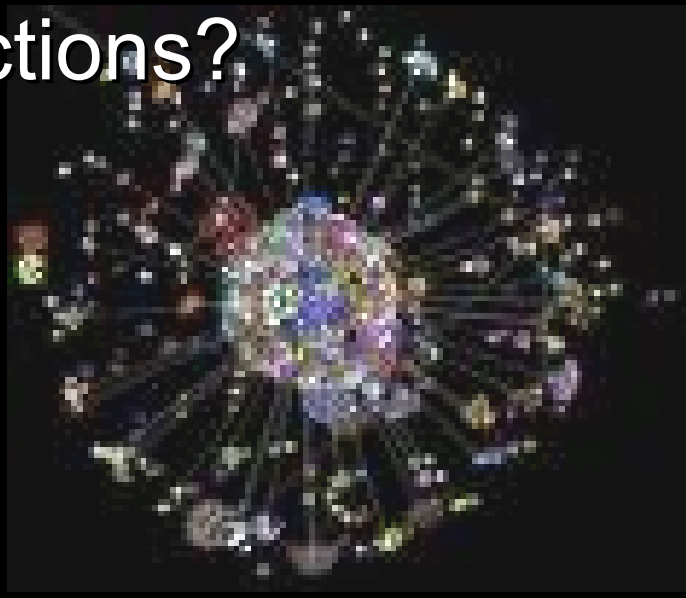


societies are systems of interacting organisms

living systems

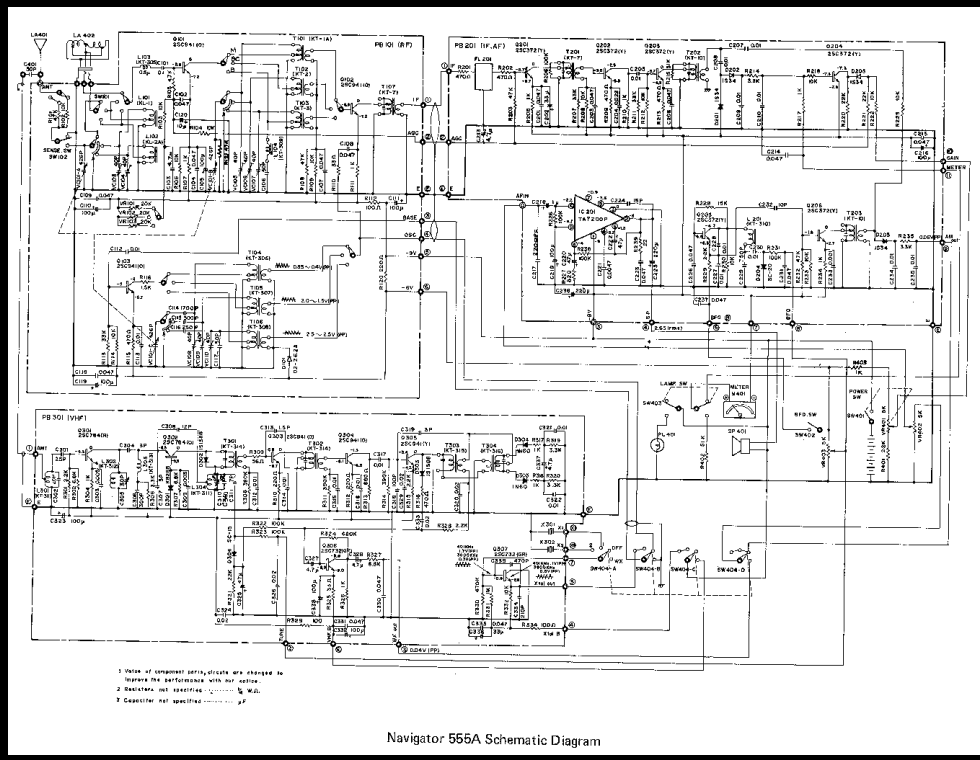
self-organized complex dynamical systems of interacting parts

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



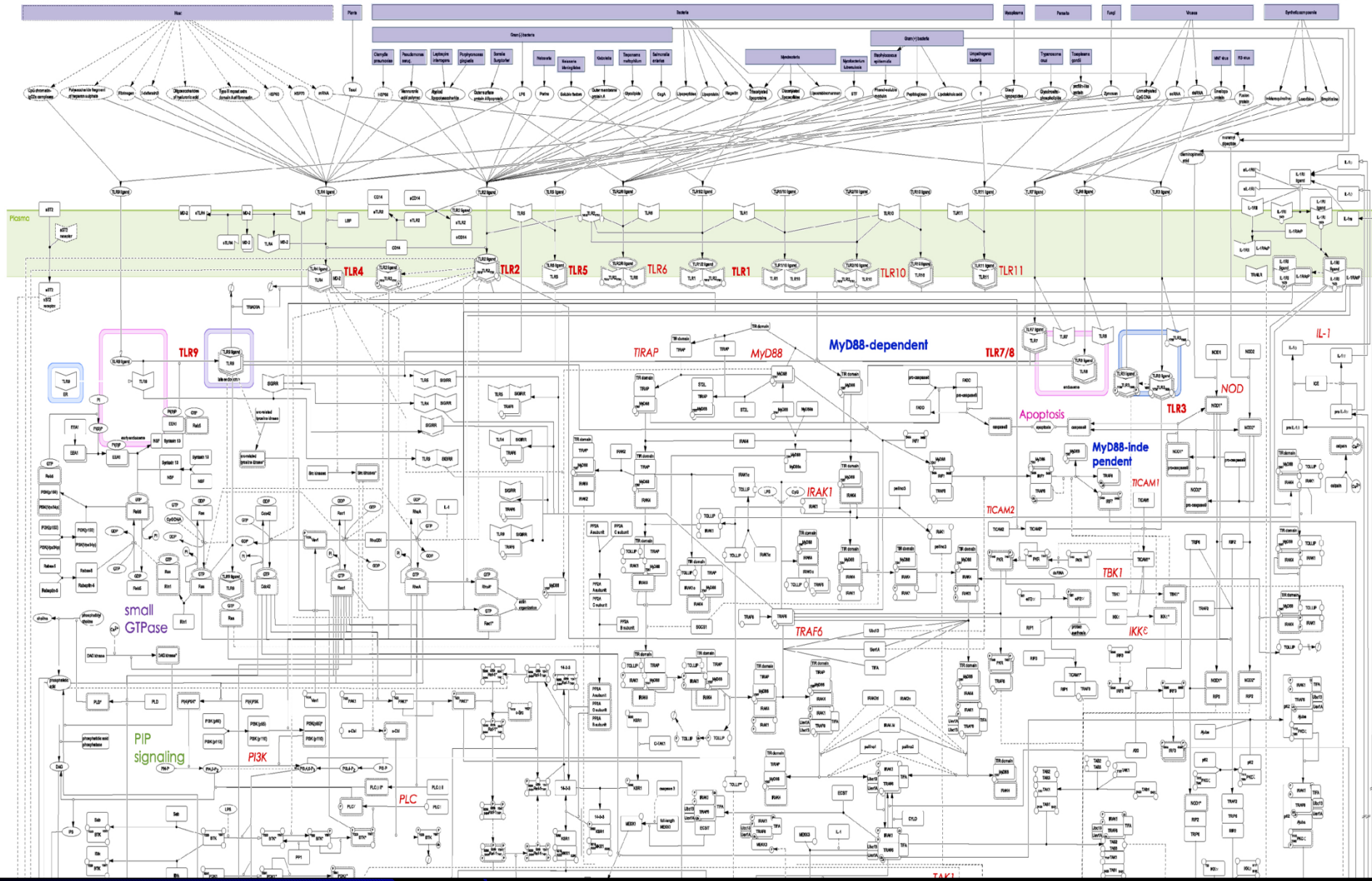
genetic networks

- 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.



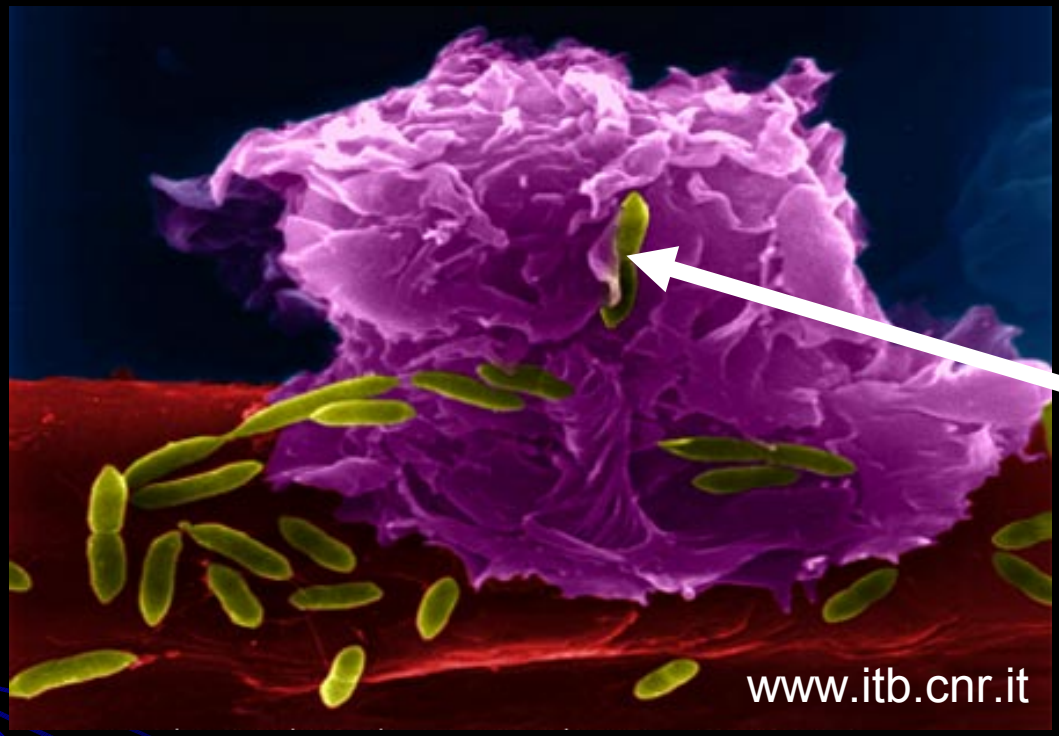
Analogy: circuit diagram

Map of the TLR signaling pathway in the macrophage



Mol Syst Biol. 2:2006.0015

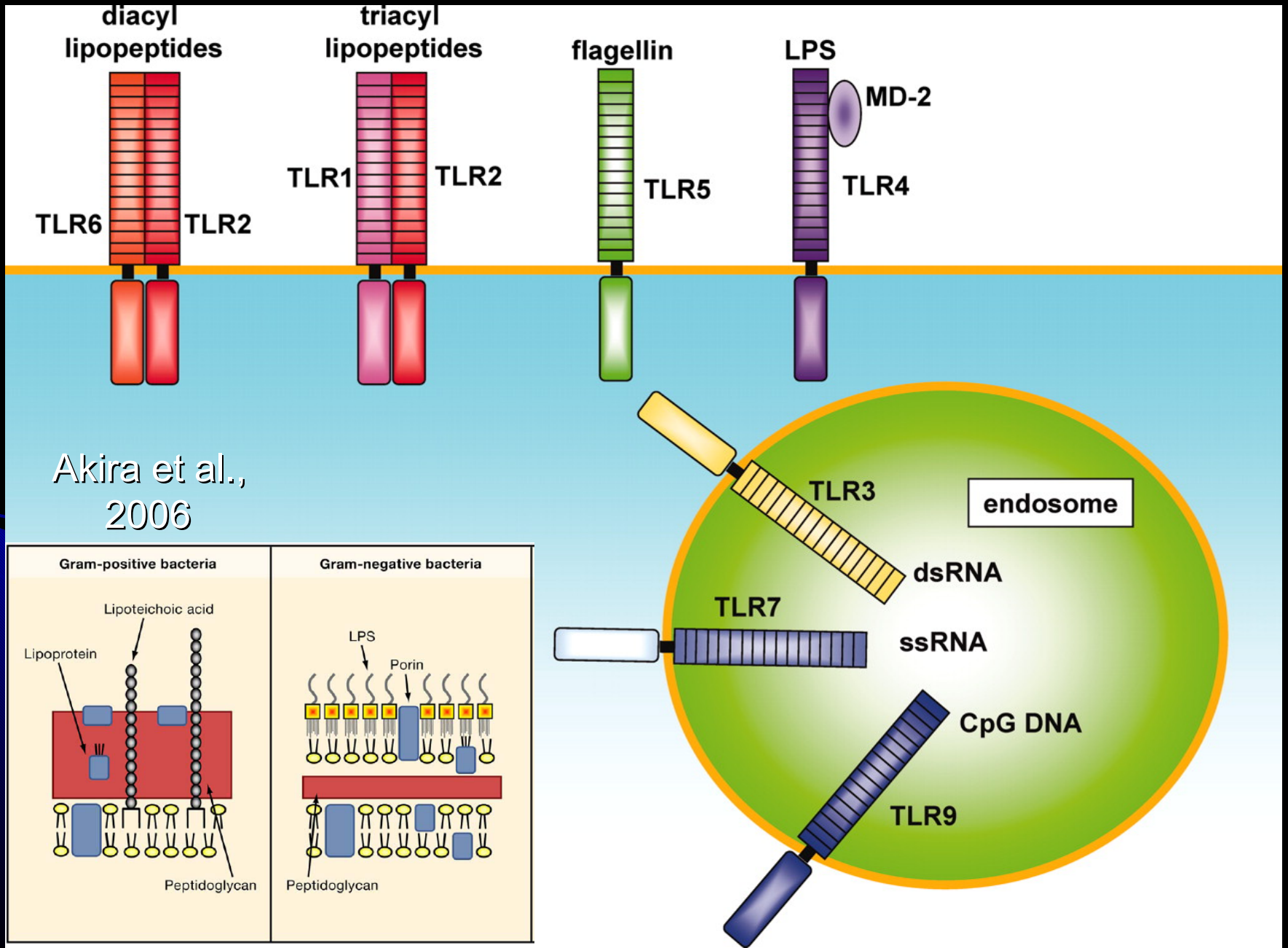
Macrophages are key immune cells



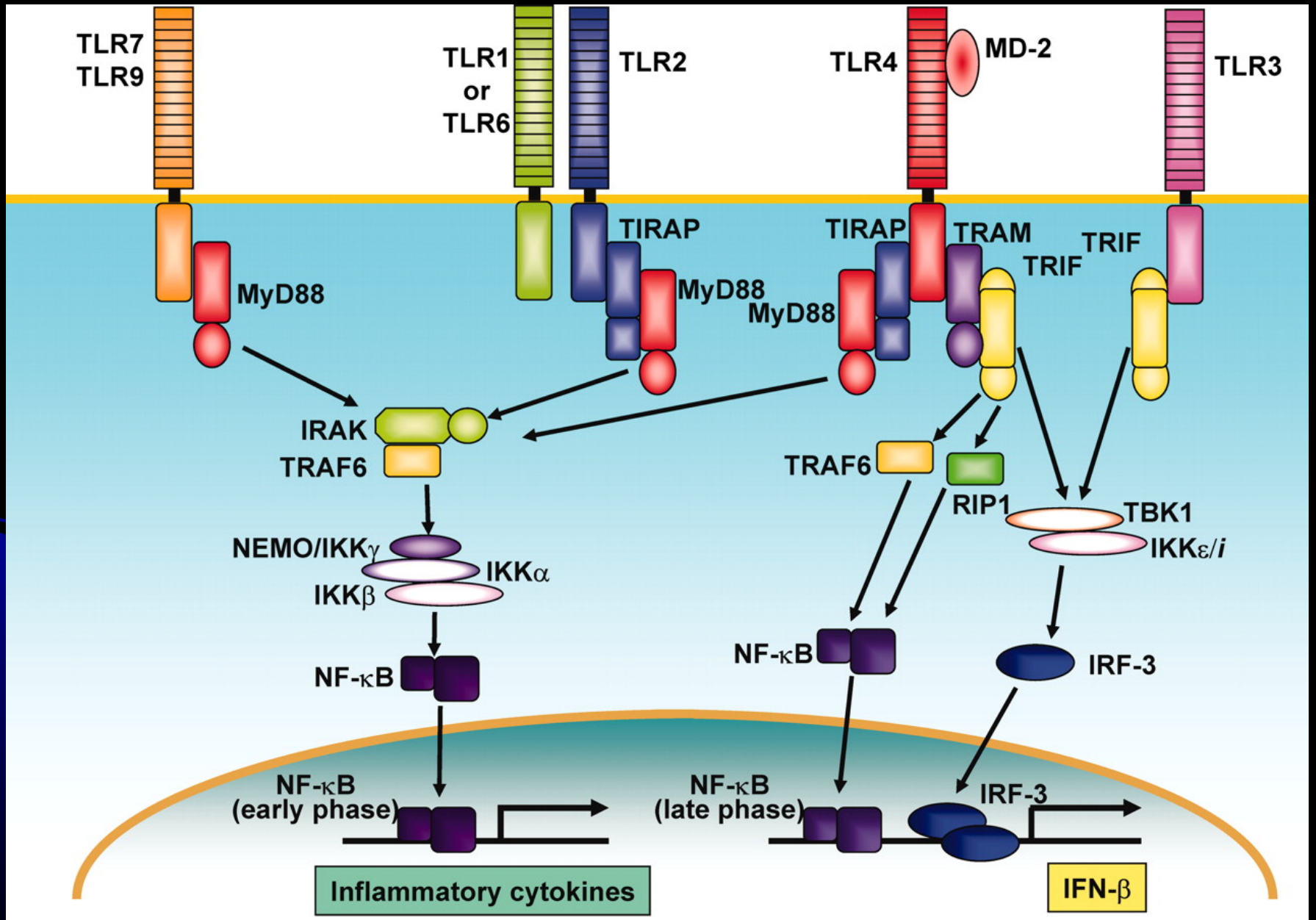
bacterium
being engulfed

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

TLRs and their associated PAMPs



TLR signaling pathway

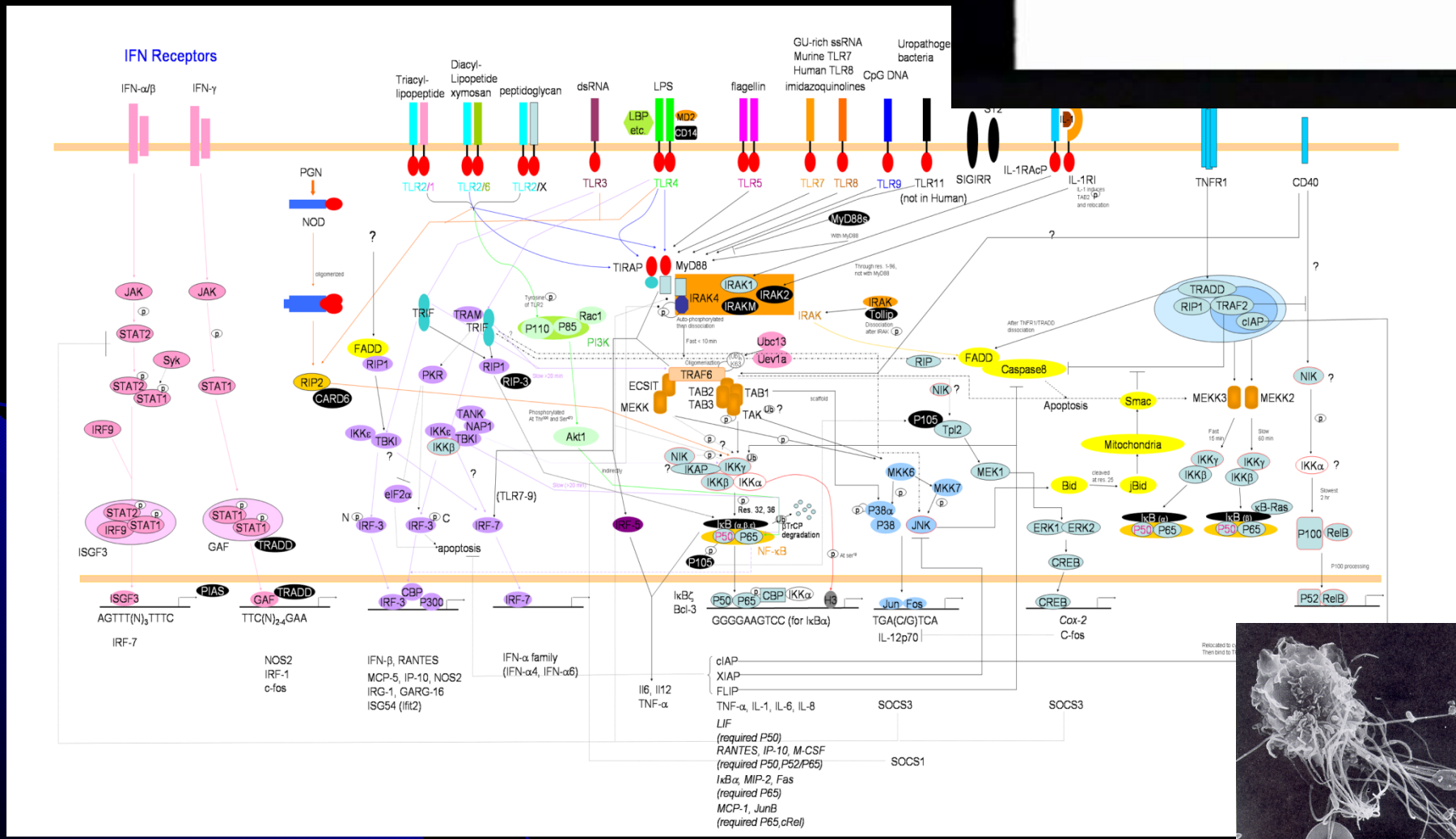


Map of the TLR signaling pathway in the macrophage

A few perspectives on

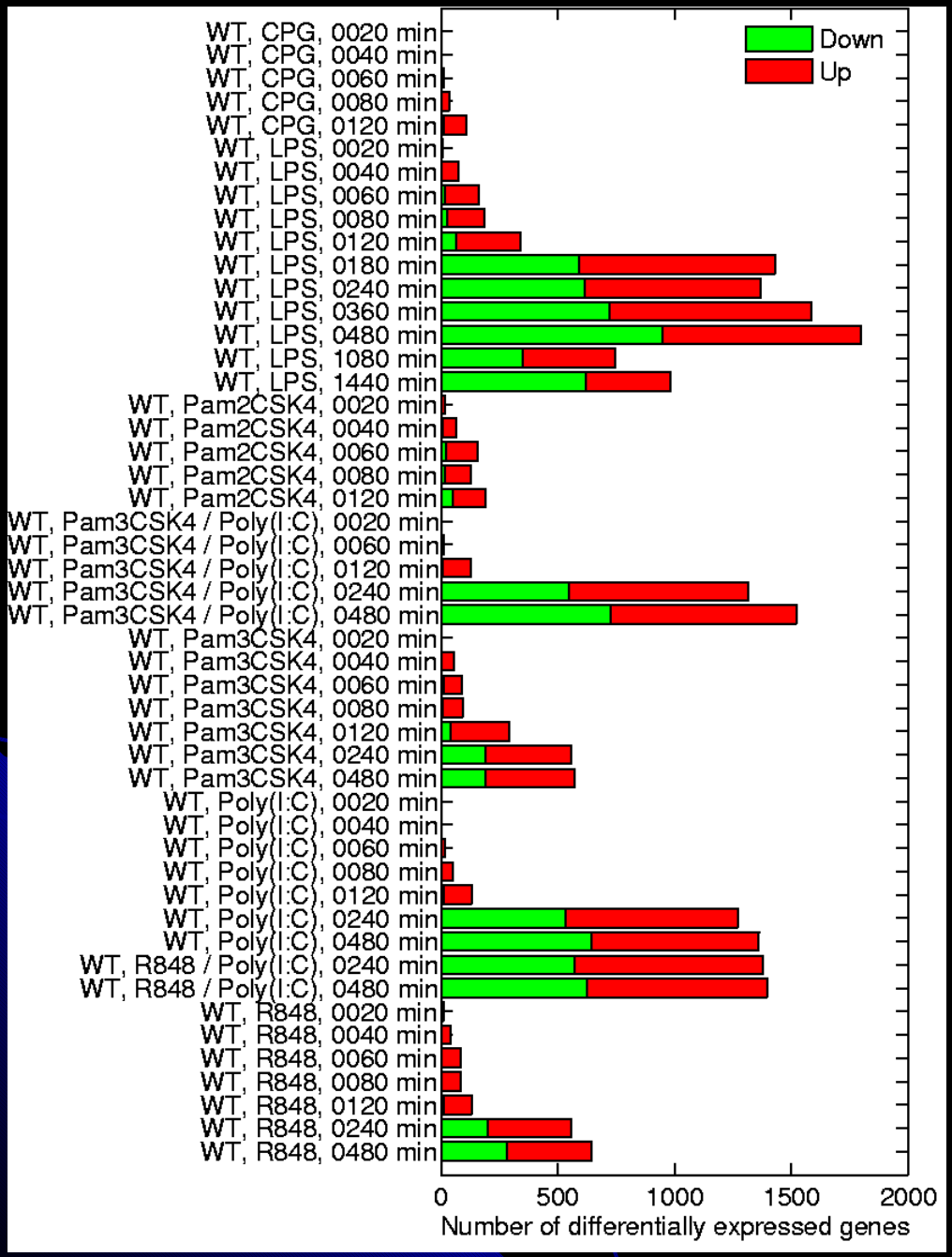
CELL MOTILITY

Division of Experimental Medicine
Brigham & Women's Hospital



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_3CSK_4, Pam_2CSK_4, 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

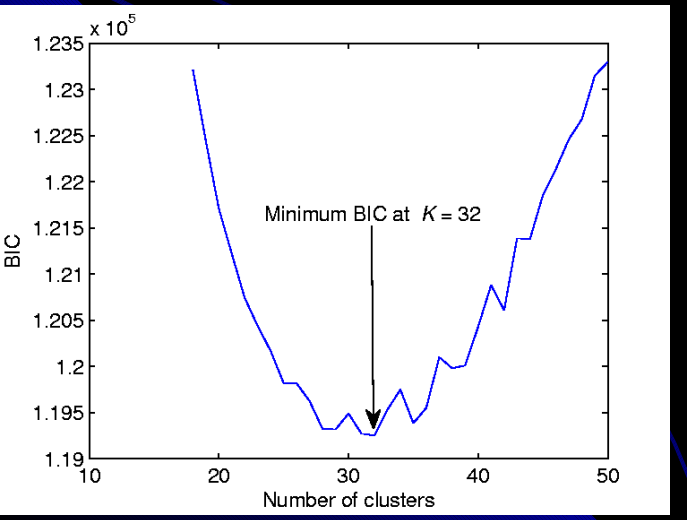


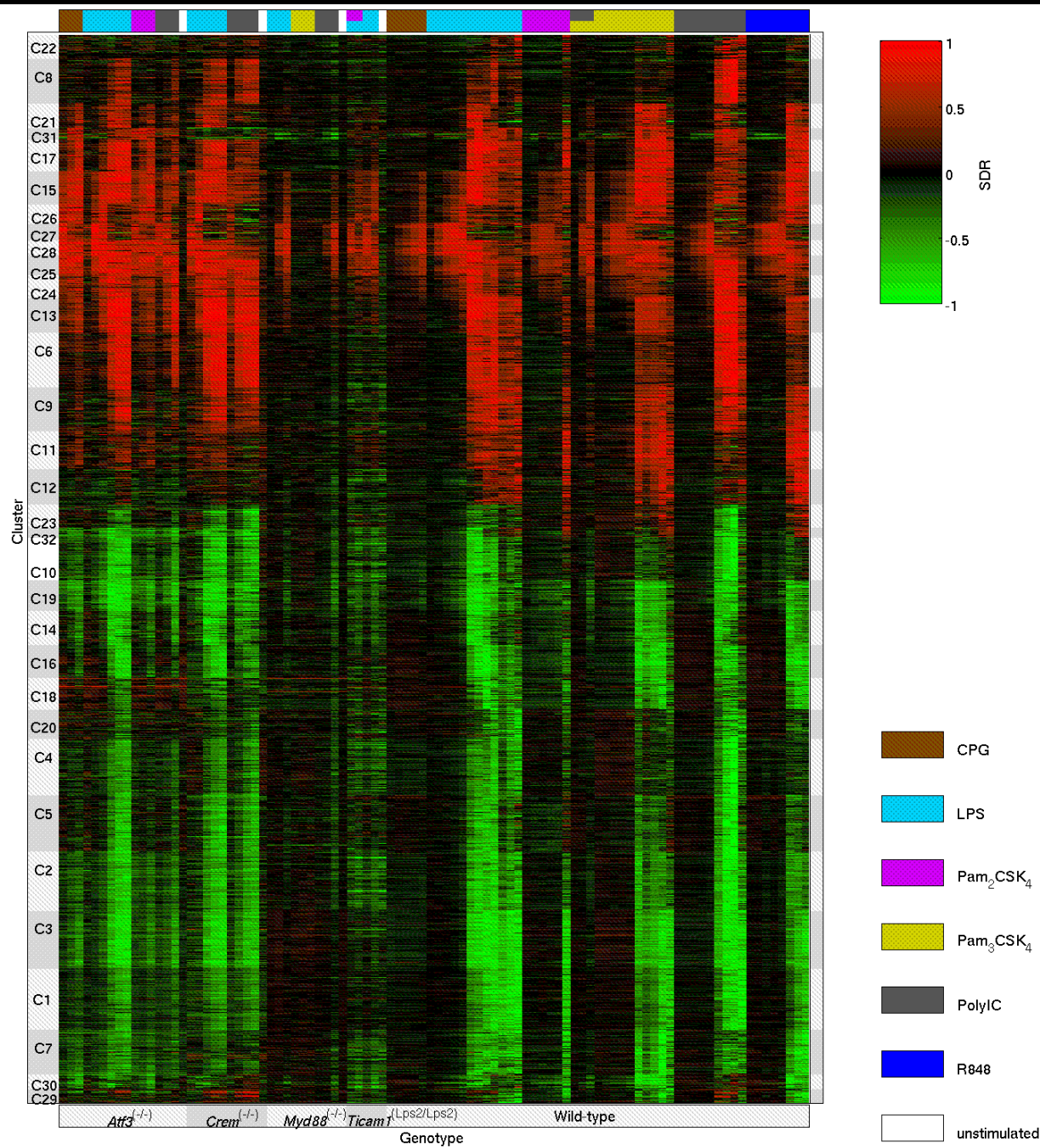
Number of differentially expressed genes vs. elapsed time (by stimulus)

Total number: 2,562

S. Ramsey, PLoS Comp Bio, 4,3, 2008.

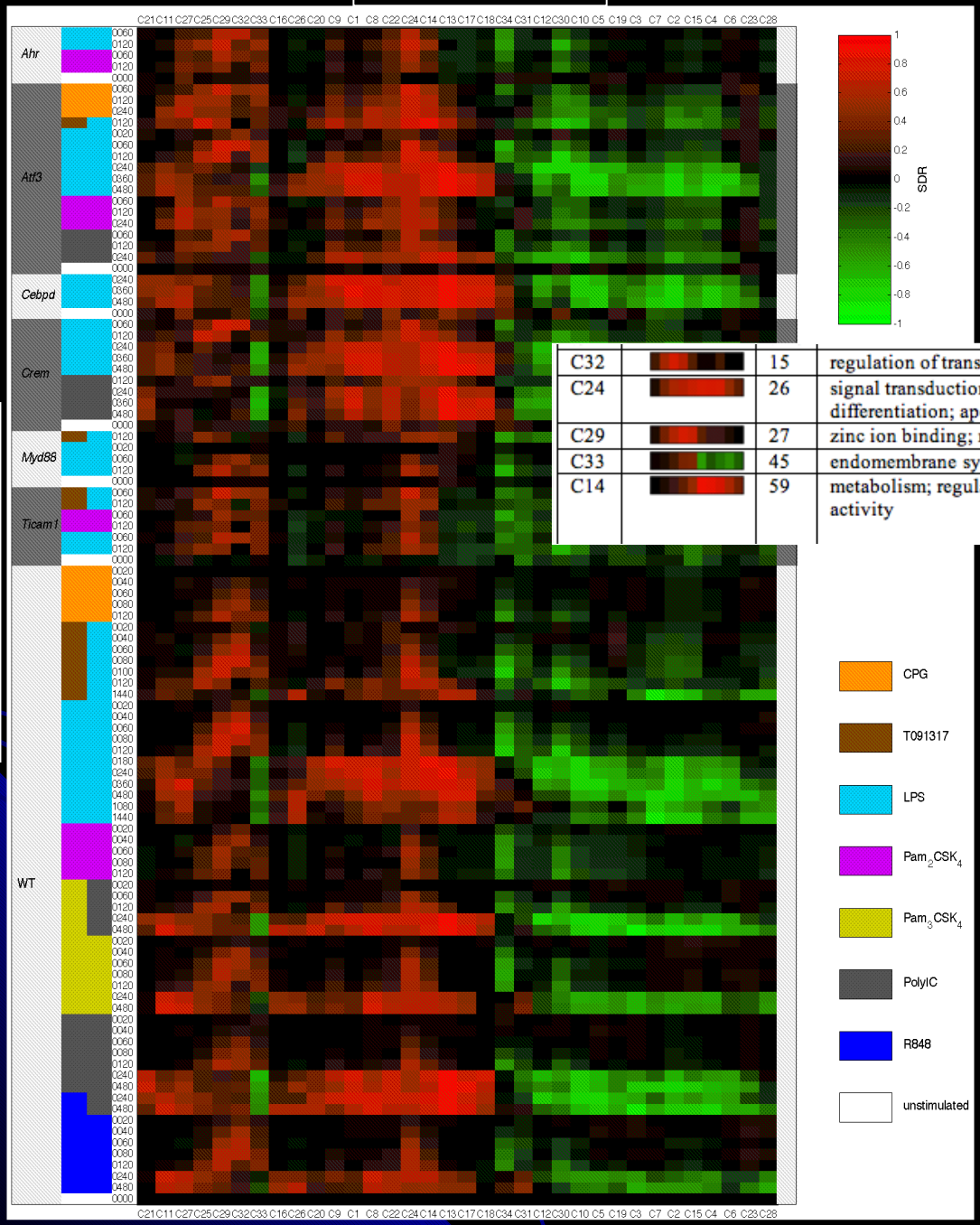
clustering of gene expression profiles





34 Clusters

130 Experiments



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, Ccrn4l, 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_2CSK_4

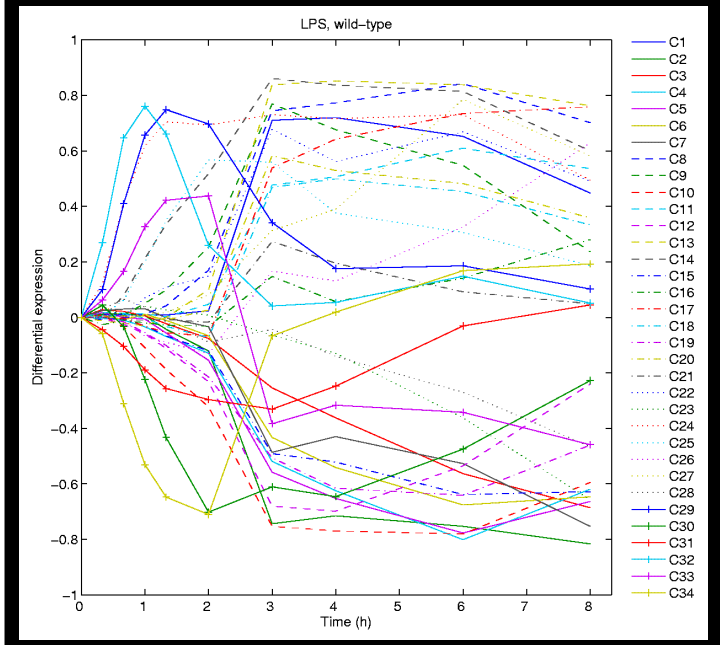
Pam_3CSK_4

PolyIC

R848

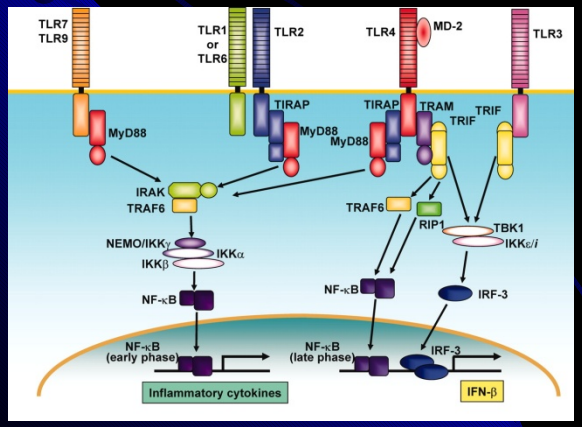
unstimulated

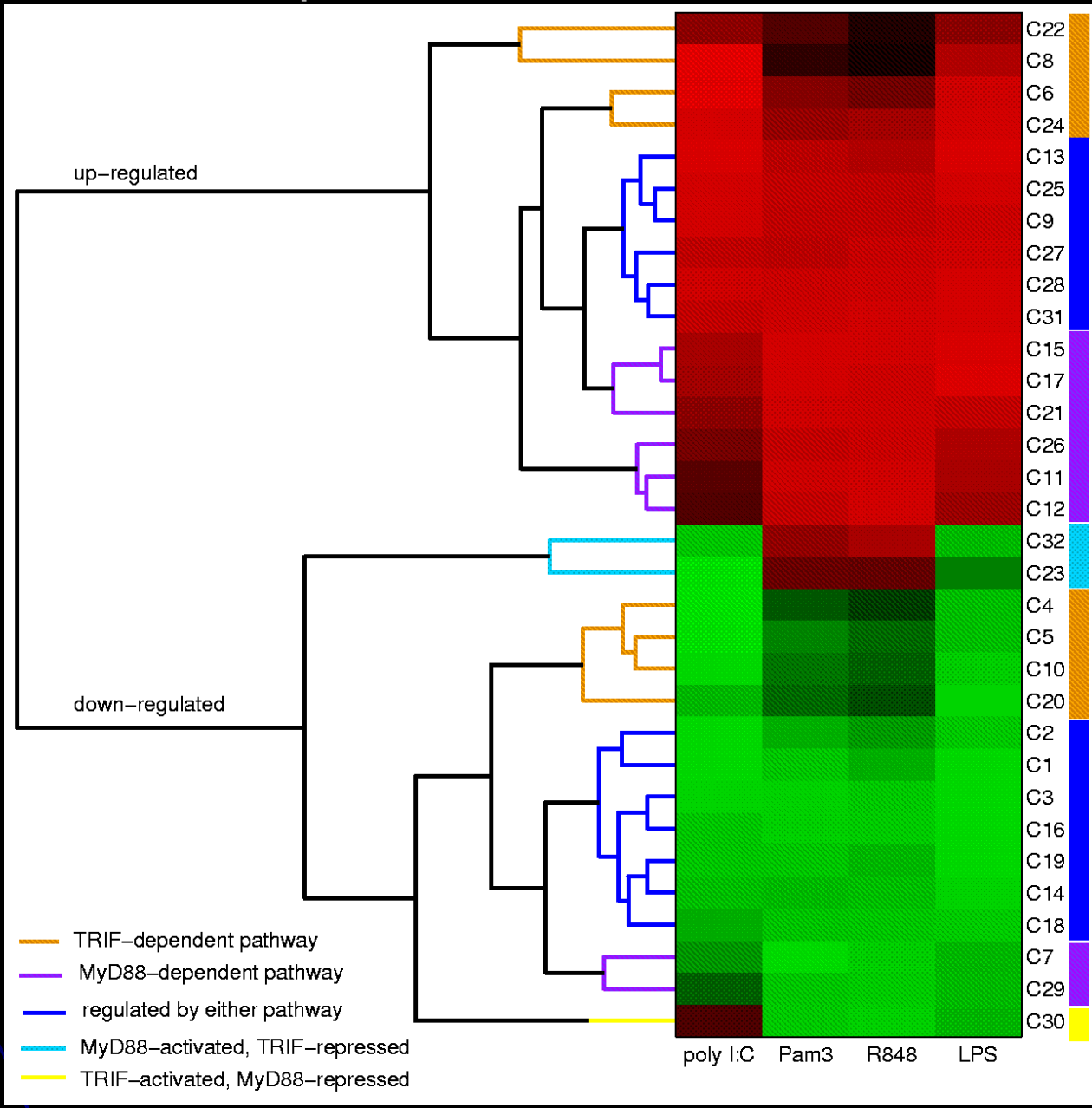
Cluster summarized expression profiles



S. Ramsey, PLoS Comp Bio, 4,3,, 2008.

Stimulus-specific cluster responses

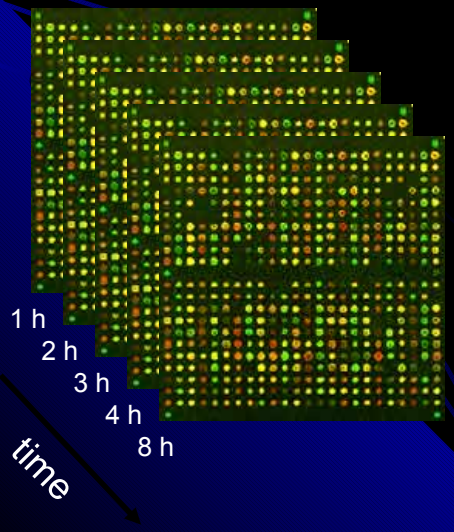




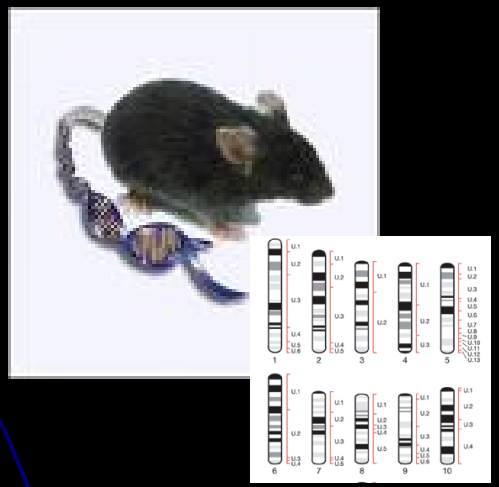
S. Ramsey, PLoS Comp Bio, 4,3, 2008.

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.

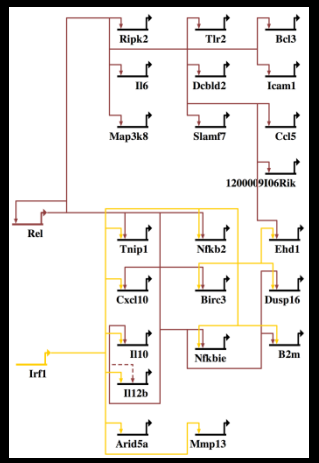
Expression *dynamics*



Genomic sequence data



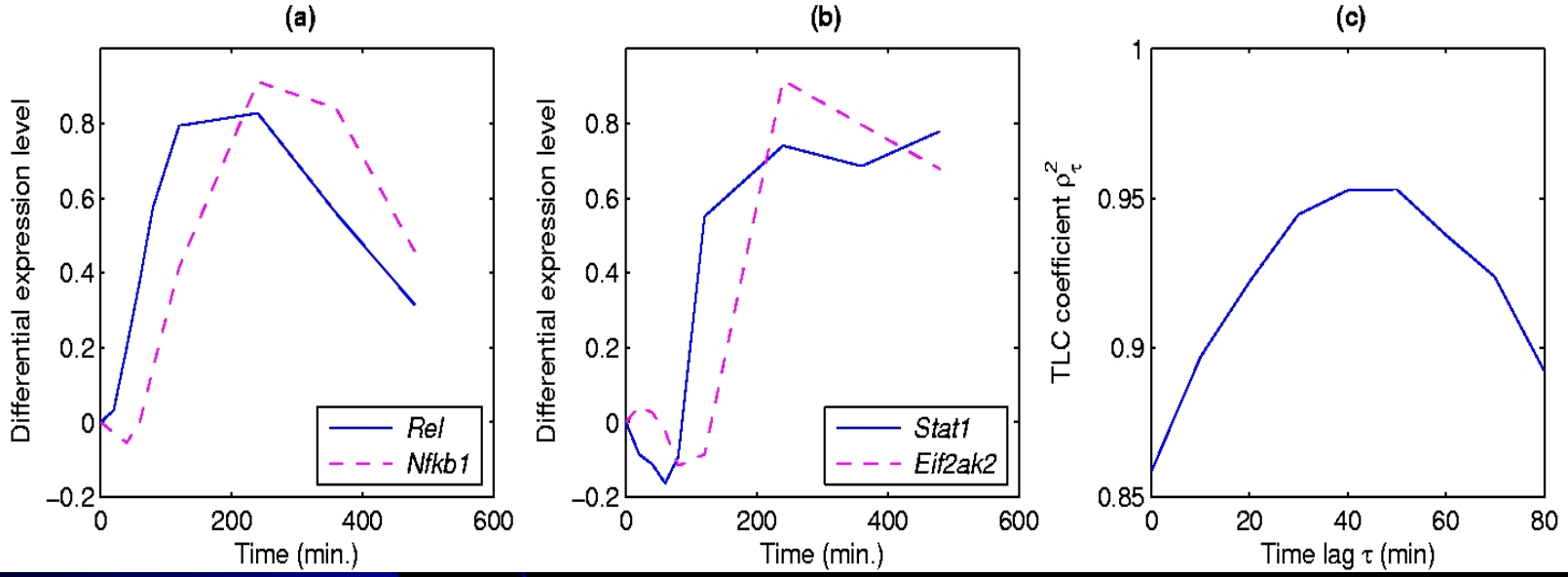
Transcriptional network



S. Ramsey

Components of the gene-gene transcriptional time delay

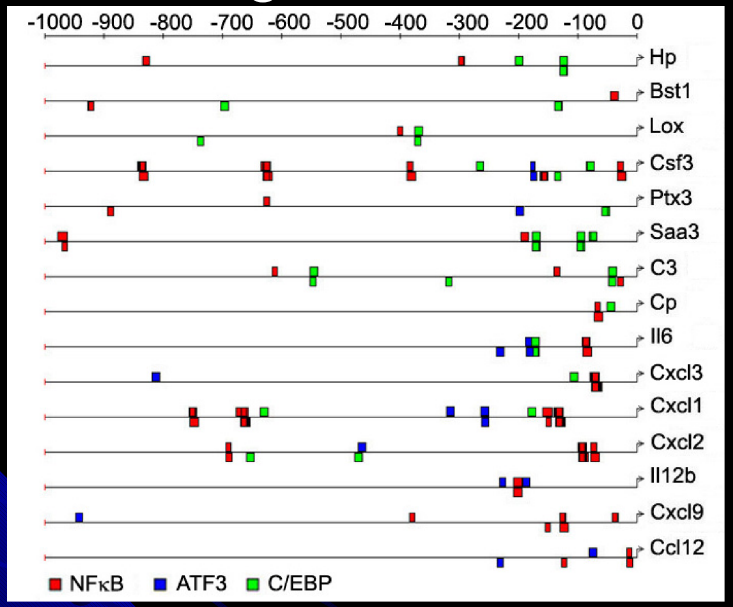
- mRNA half-life of TF gene
- Translation and folding of TF protein
- Diffusion of TF back into nucleus
- Turnover rate (half-life) for TF protein
- Transcriptional elongation of target gene



S. Ramsey, PLoS Comp Bio, 4,3, 2008.

Scan for TF binding motif matches

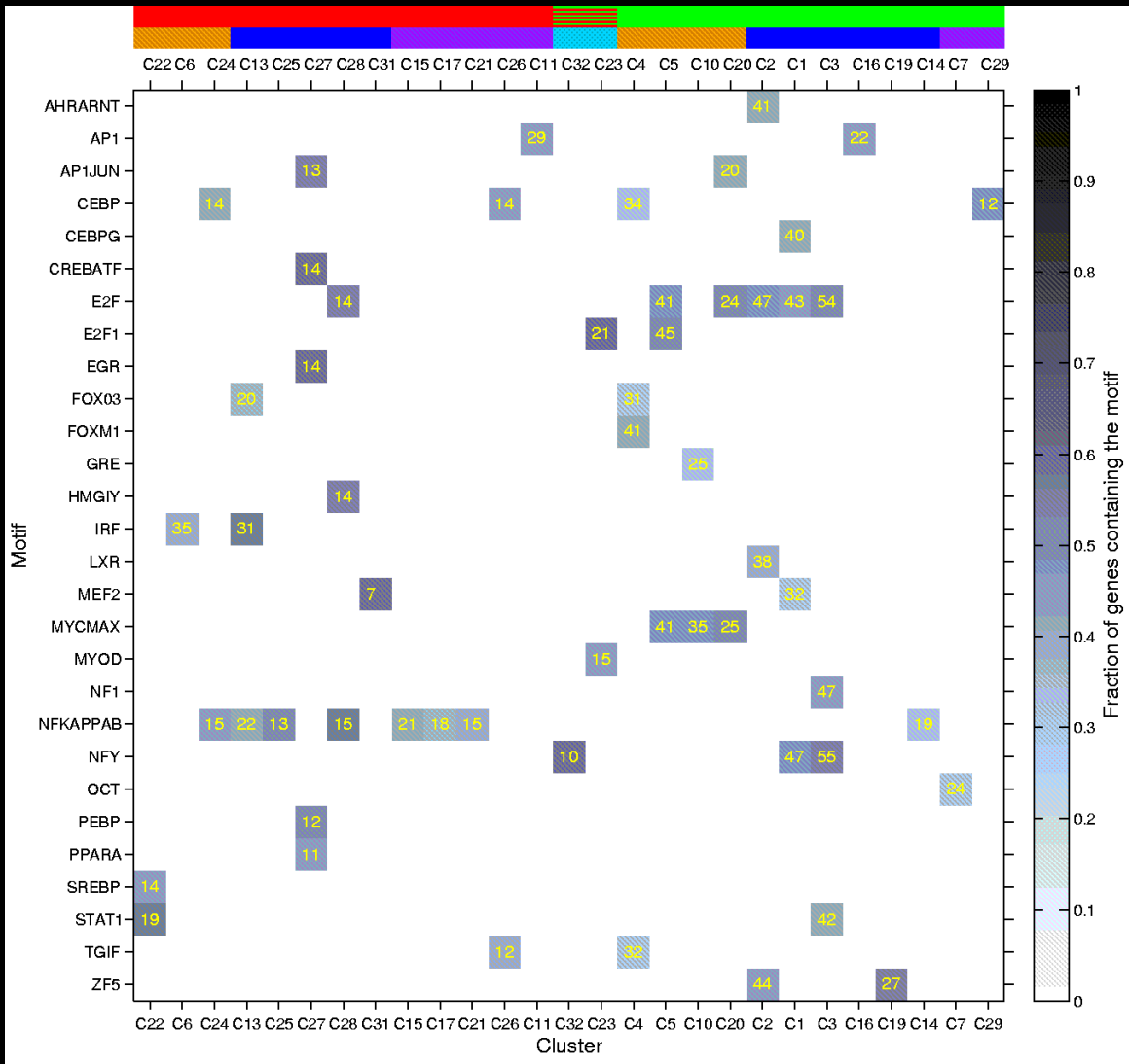
Scanning:



TF recognition site motif:

S. Ramsey, PLoS Comp Bio, 4,3, 2008.

Motif enrichment matrix:



Probabilistic Framework for Transcription Factor Binding Site Prediction

www.probtf.org

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

round model

Data Fusion

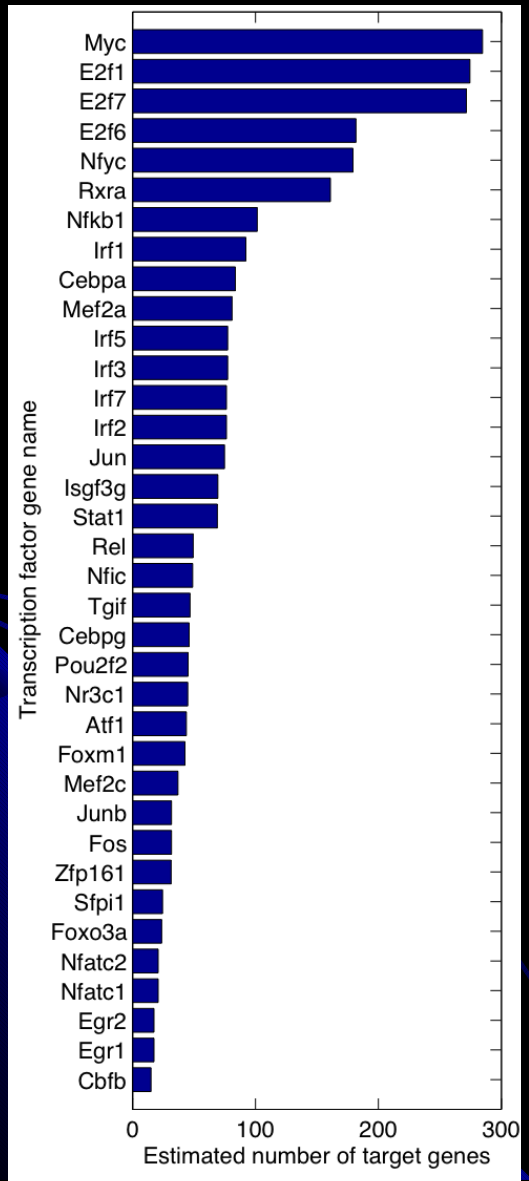
$$|A, \pi, \Theta, \phi) = P(S|A, \pi, \Theta, \phi)P(\mathcal{D}|A, \pi)$$

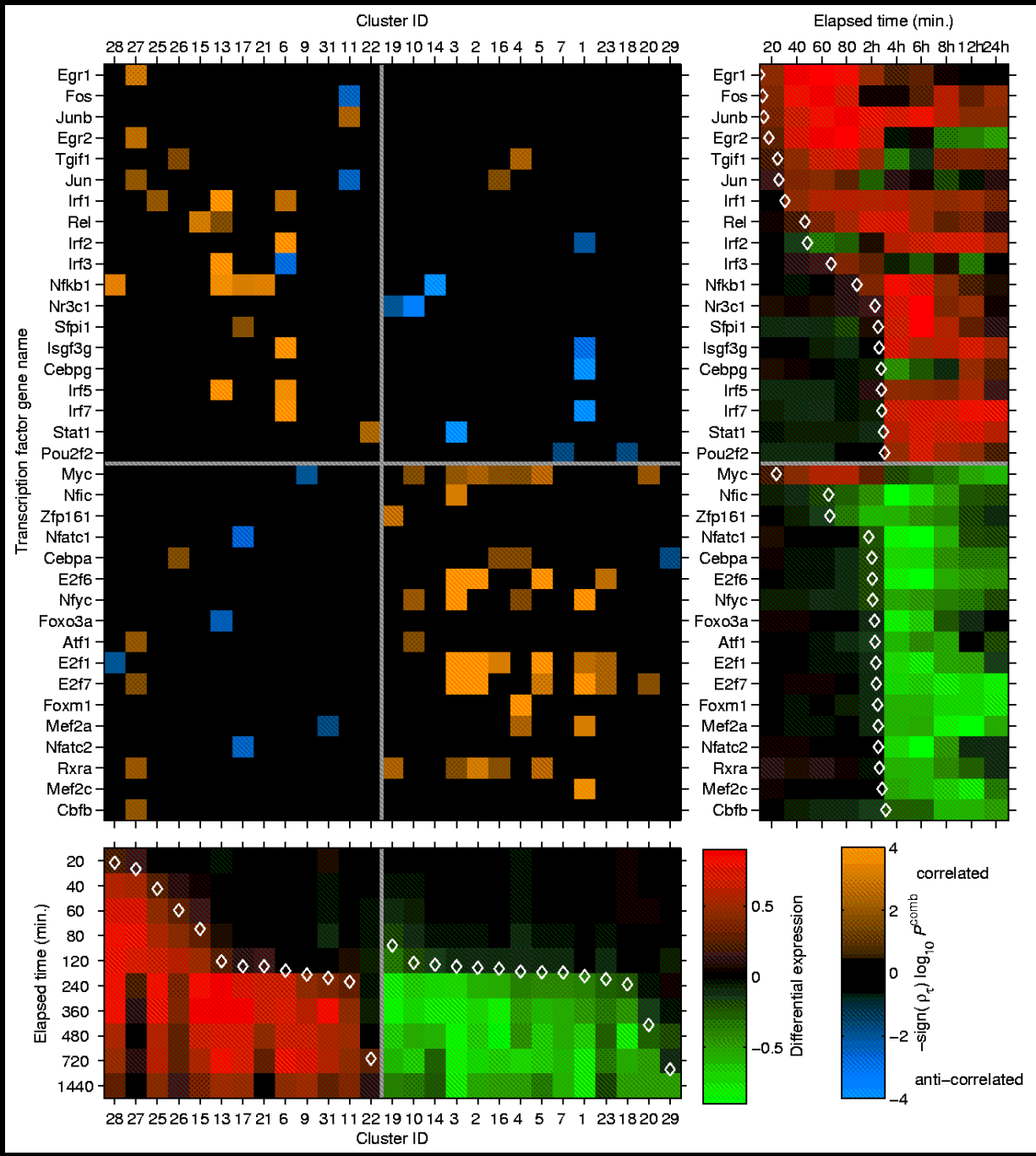
$$|S, \mathcal{D}) = \frac{P(S, \mathcal{D}|A, \pi)P(A, \pi)}{P(S, \mathcal{D})}$$

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

Lähdesmäki *et al* PLoS ONE, 3:3, e1820, 2008.

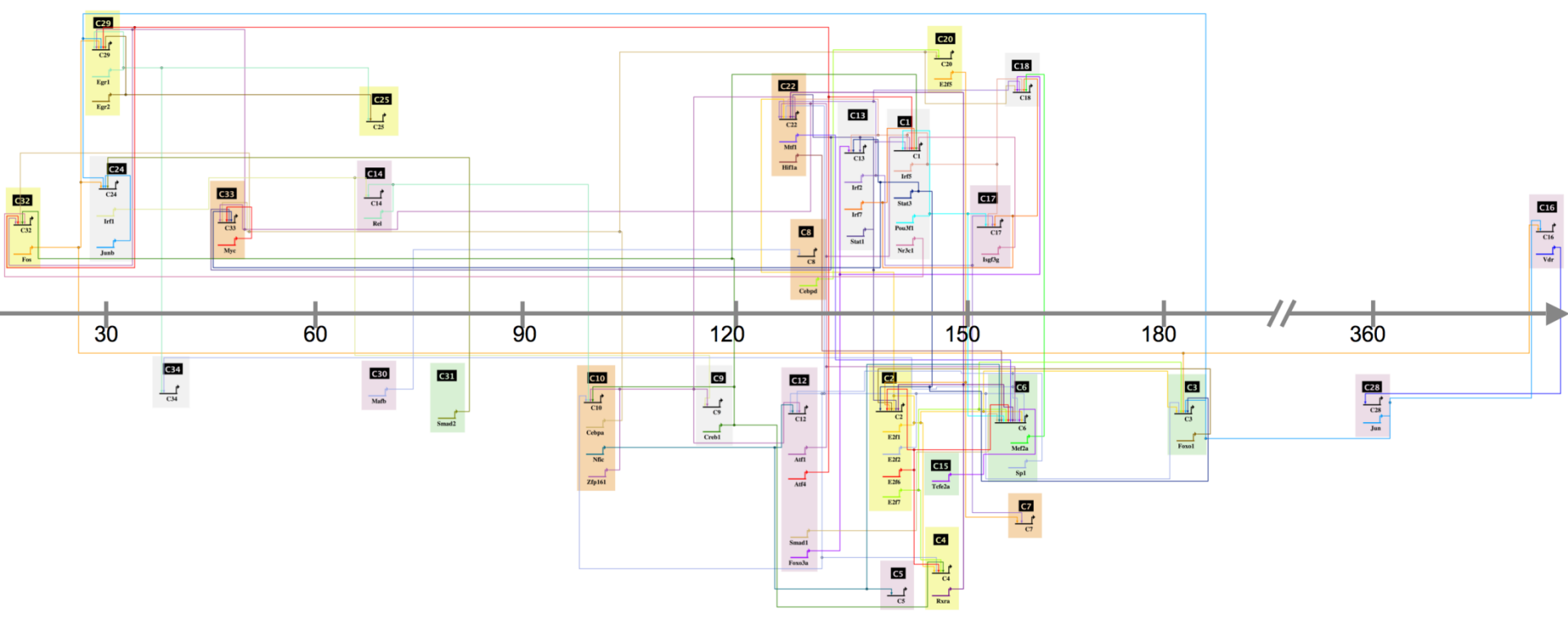
Map network





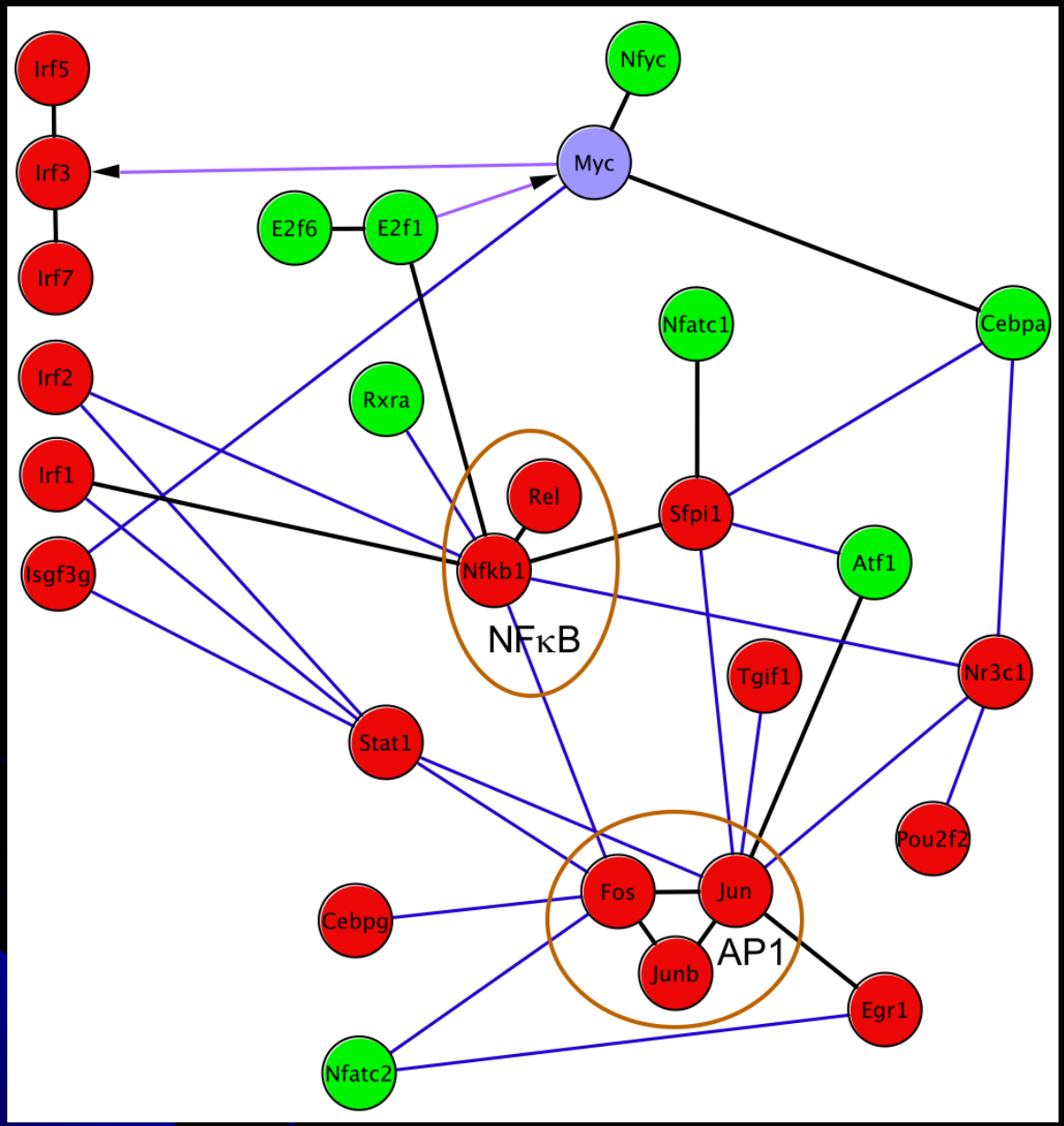
S. Ramsey, PLoS Comp Bio, 4,3, 2008.

TF-to-cluster association network



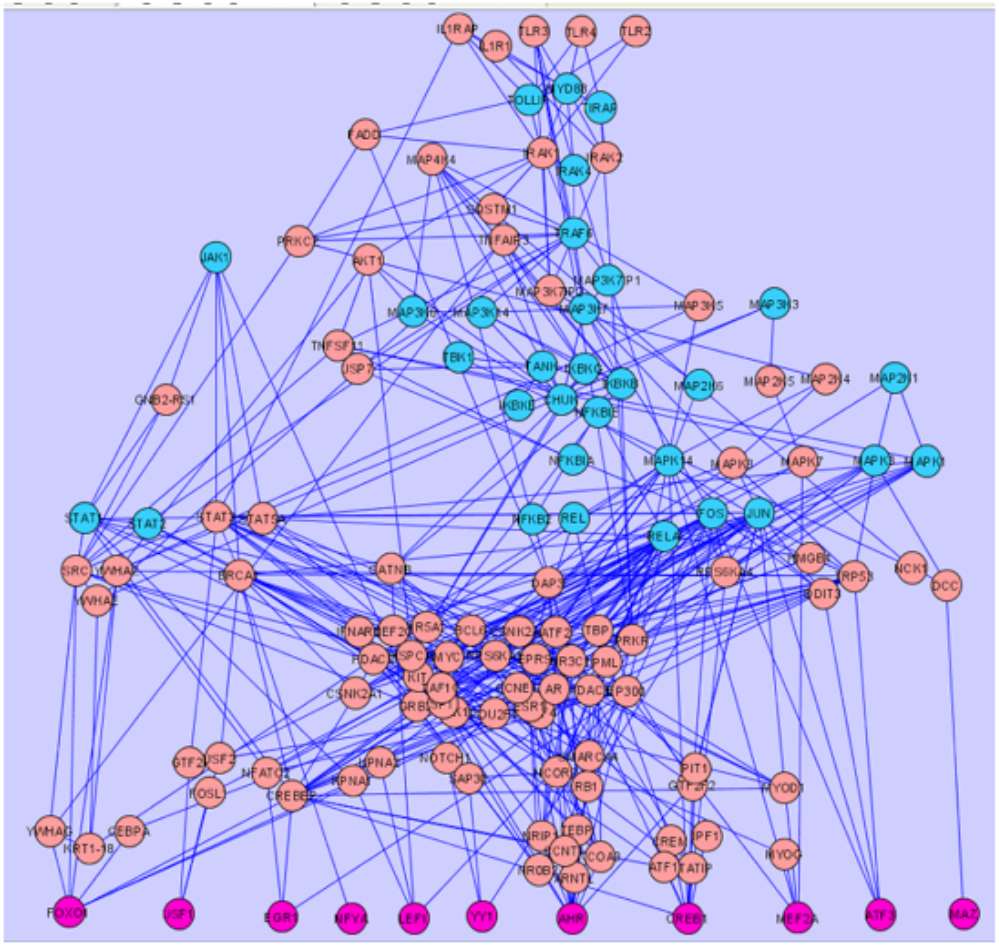
S. Ramsey, PLoS Comp Bio, 4,3, 2008.

TF interactions



S. Ramsey, PLoS Comp Bio, 4,3, 2008.

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



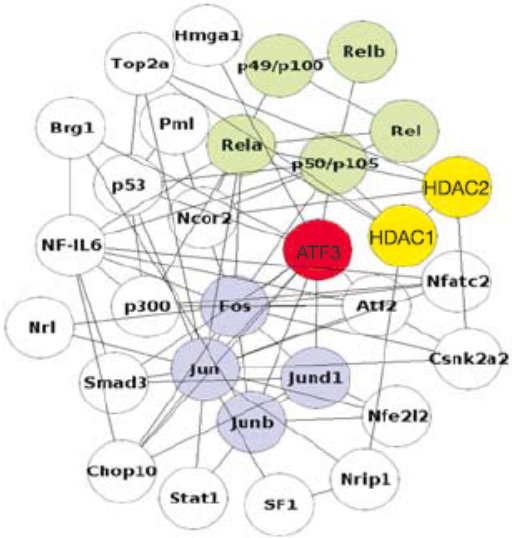
Known TLR Signaling pathway

(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-kB TF complexes.

B. Li, V. Thorsson

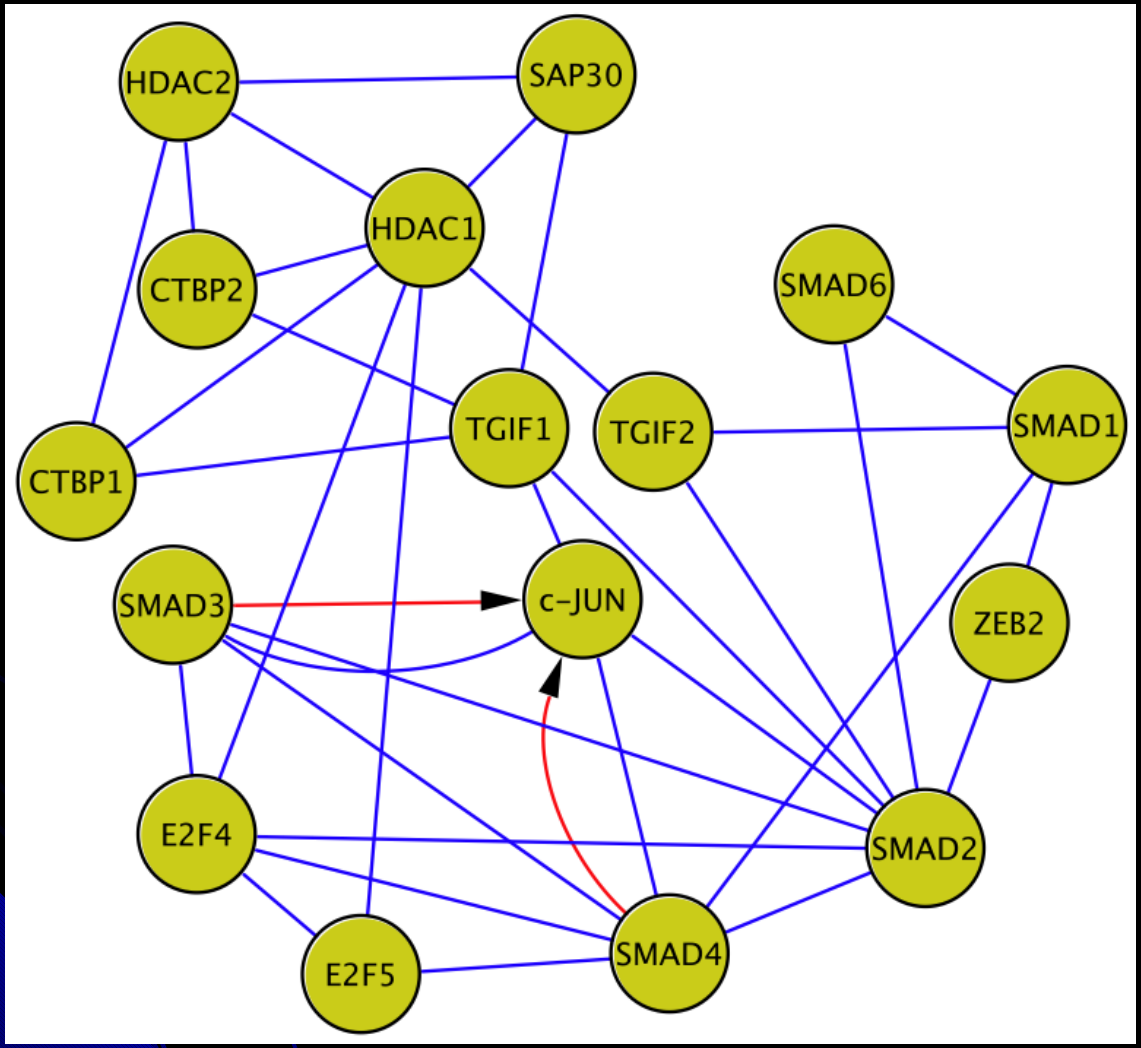
ChIP-on-chip validation

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	1.1×10^{-3}
NFκB/p50	NFKB_Q6	LPS	C17	1 h, 2 h	58	20	11	2.5×10^{-1}
NFκB/p50	NFKAPPAB_01	LPS	C28	1 h, 2 h	28	21	20	1.1×10^{-6}
IRF1	IRF_Q6_01	LPS	C13	1 h, 2 h, 4 h	64	23	18	2.3×10^{-3}
IRF1	IRF_Q6_01	LPS	C25	1 h, 2 h, 4 h	37	22	18	8.8×10^{-4}

S. Ramsey, PLoS Comp Bio, 4,3, 2008.

TGIF enhanceosome

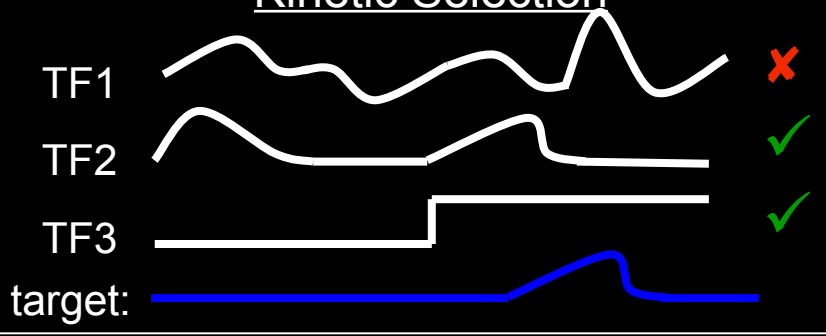


S. Ramsey, PLoS Comp Bio, 4,3, 2008.

Constrain Possible Regulators

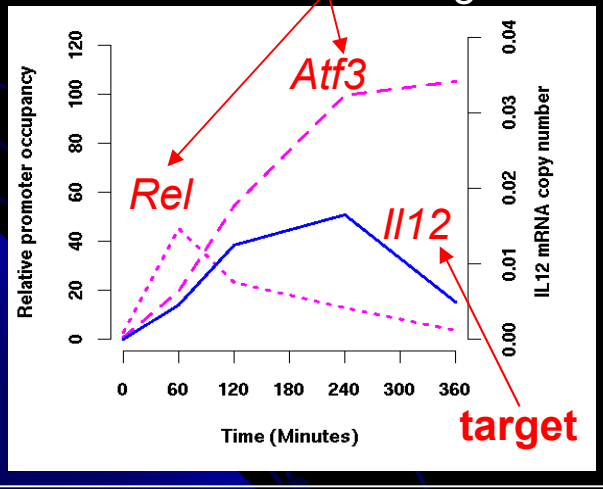
- Gene expression (timing, etc.)
- Promoter sequence analysis
- Immunoprecipitation data

Kinetic Selection

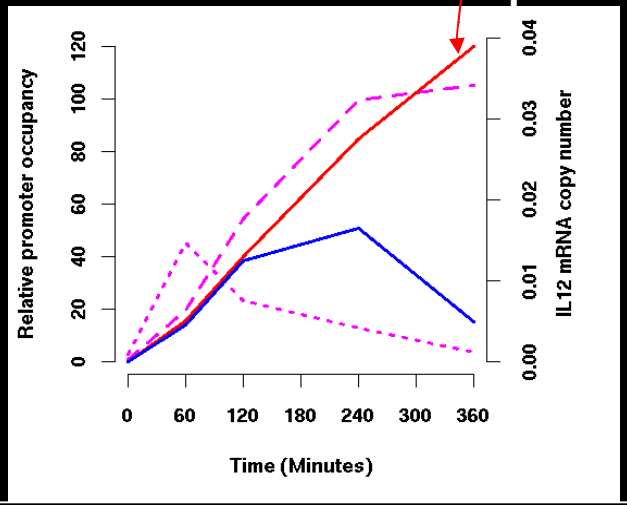


Parameter Fitting

one coefficient for each regulator:



Prediction



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

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

V. Thorsson

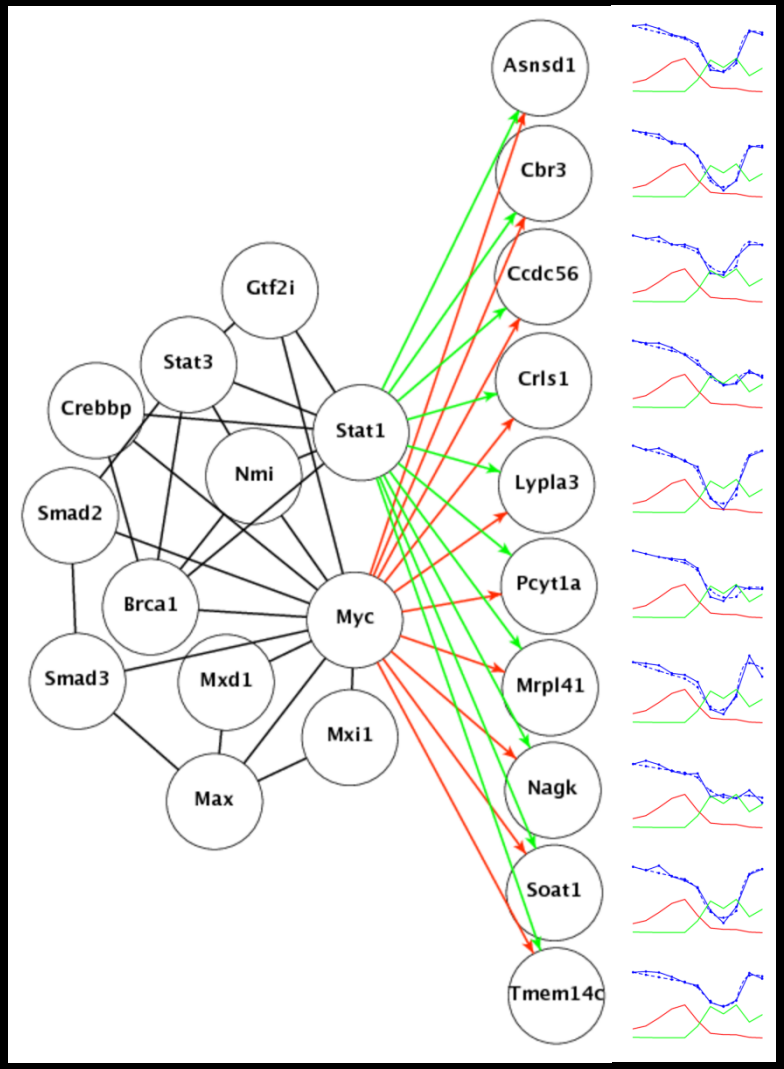
Network Predictions

Macrophage

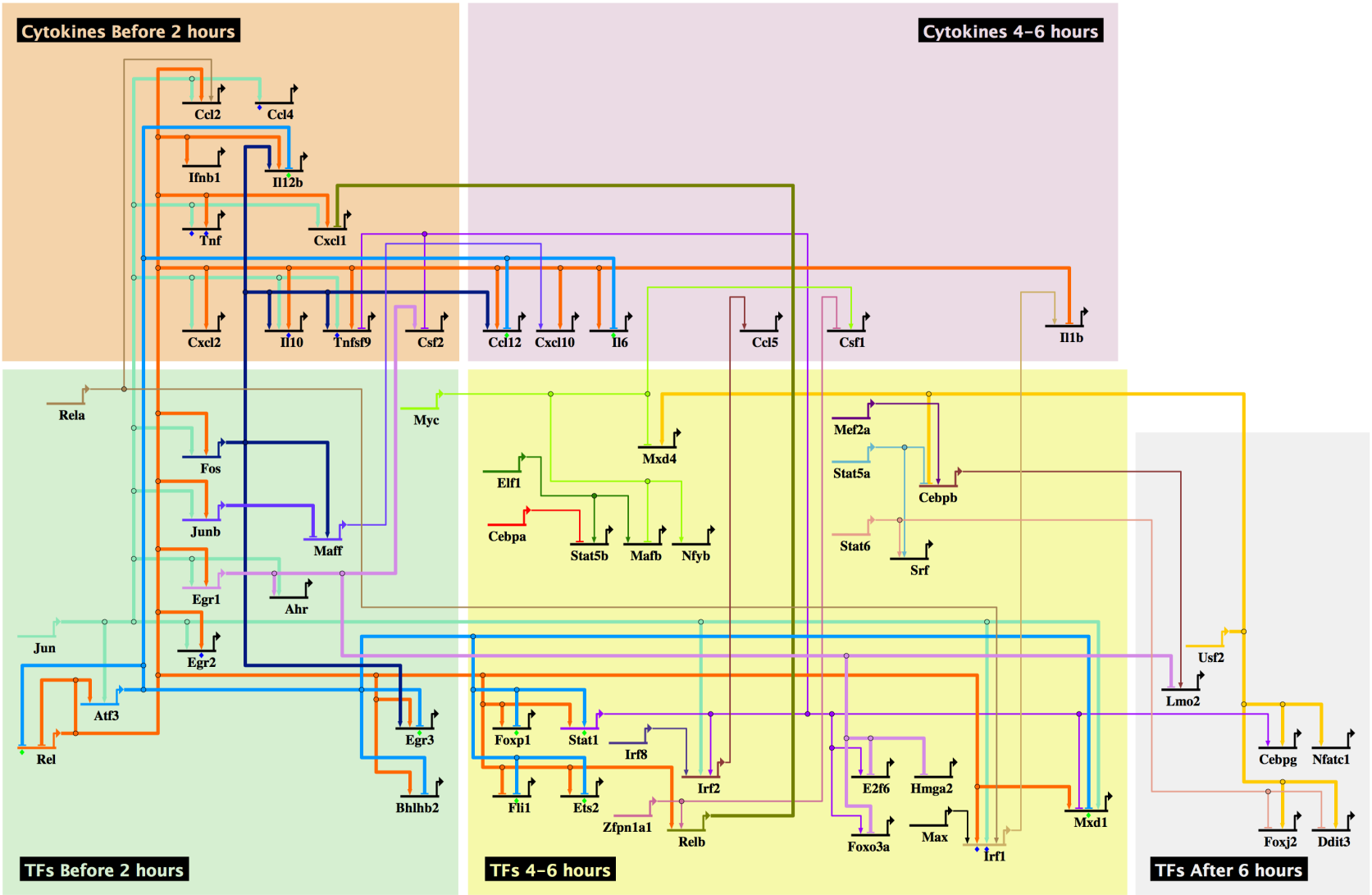
Predicted network contains co-regulated gene groups with shared:

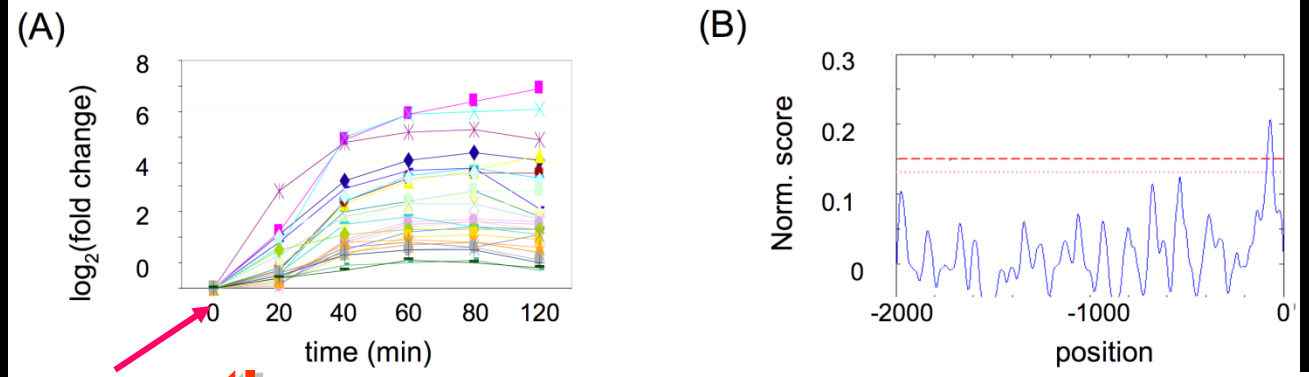
- Binding sites predictions, for single motifs or pairs of motifs
- Kinetics
- Predicted activators and repressors
- Predicted co-factors
- Function (GO categories)

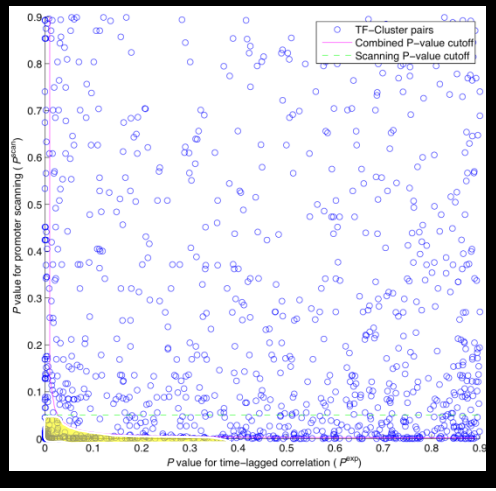
Predictions can be tested with directed binding and/or functional assays



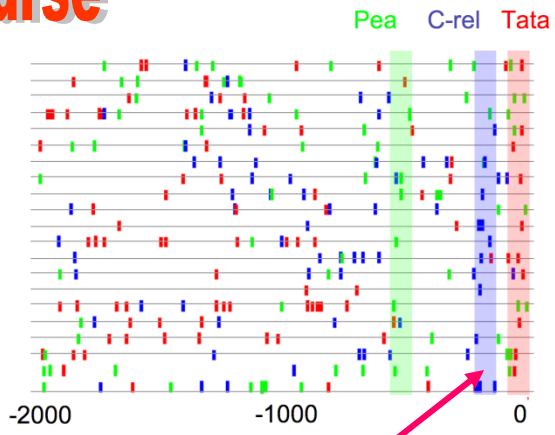
Predicted Regulation of Cytokines and Transcription Factors



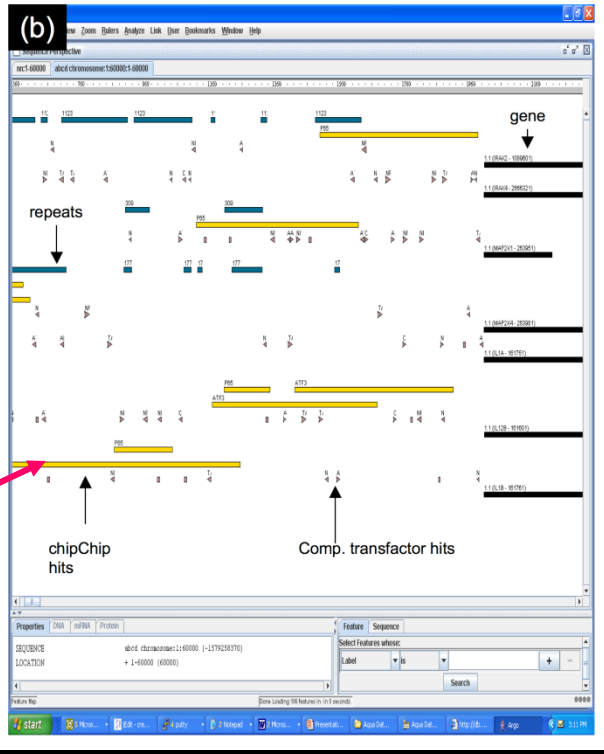




Microarray time-course data



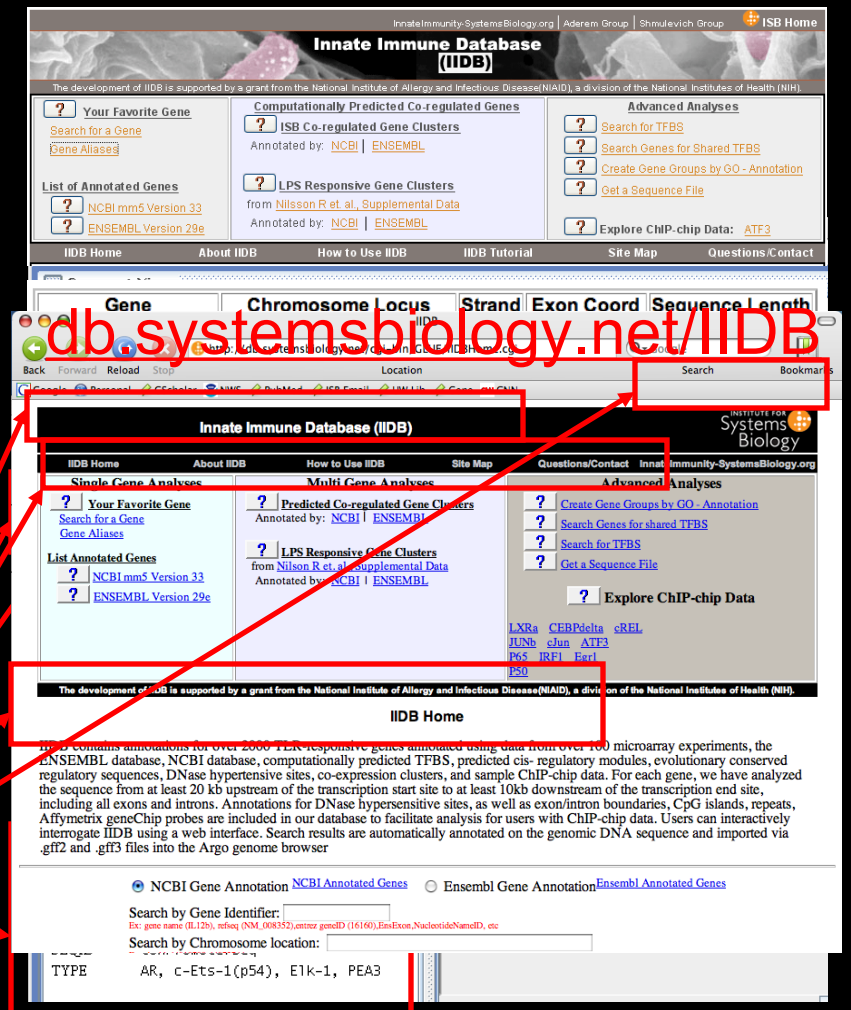
Combined density of predicted TF binding sites



ChIP-chip data

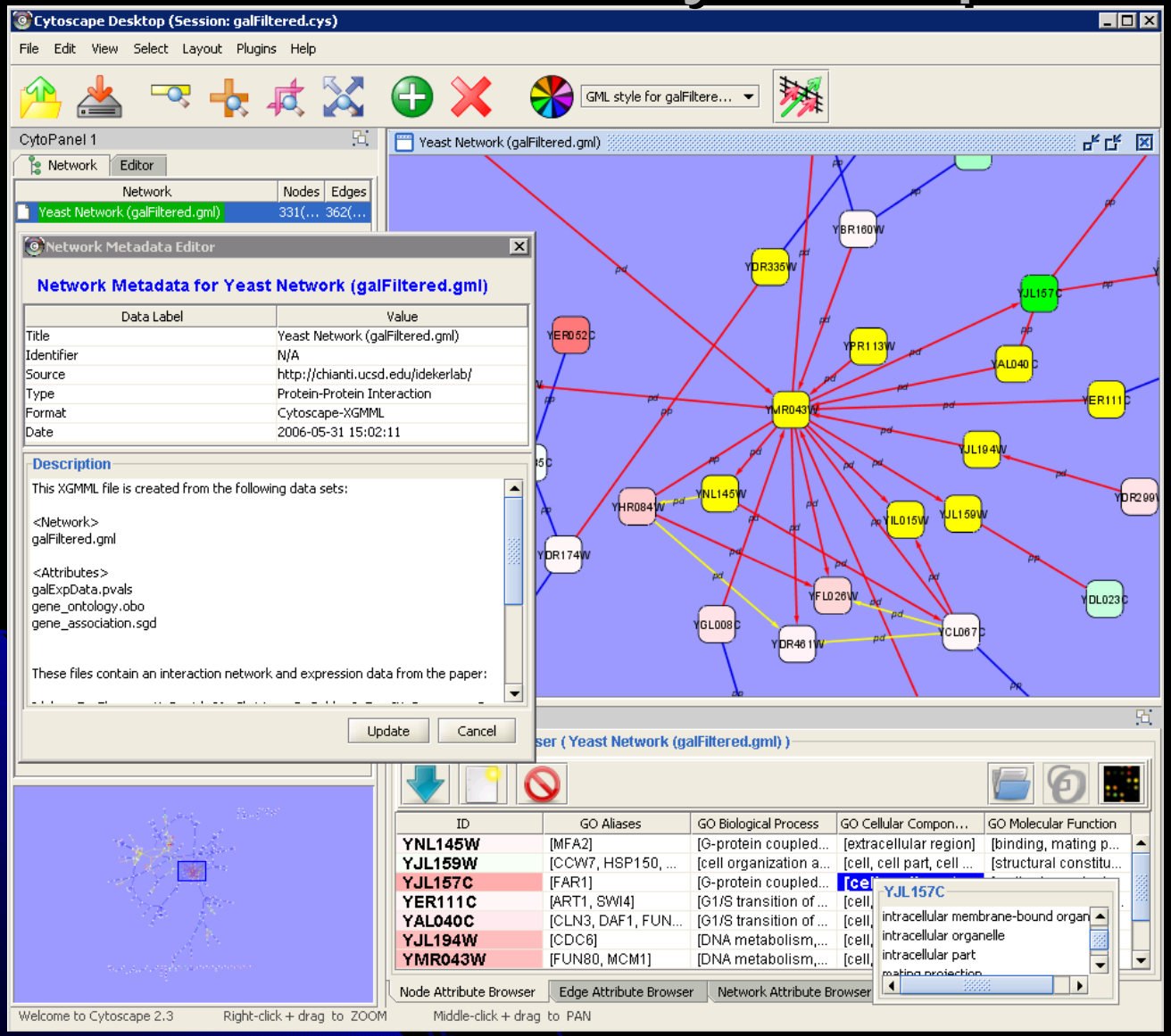
Innate Immune Database (IIDB)

- 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



Korb *et al*. *BMC Immunology*, 9:7, 2008.

Cytoscape



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/

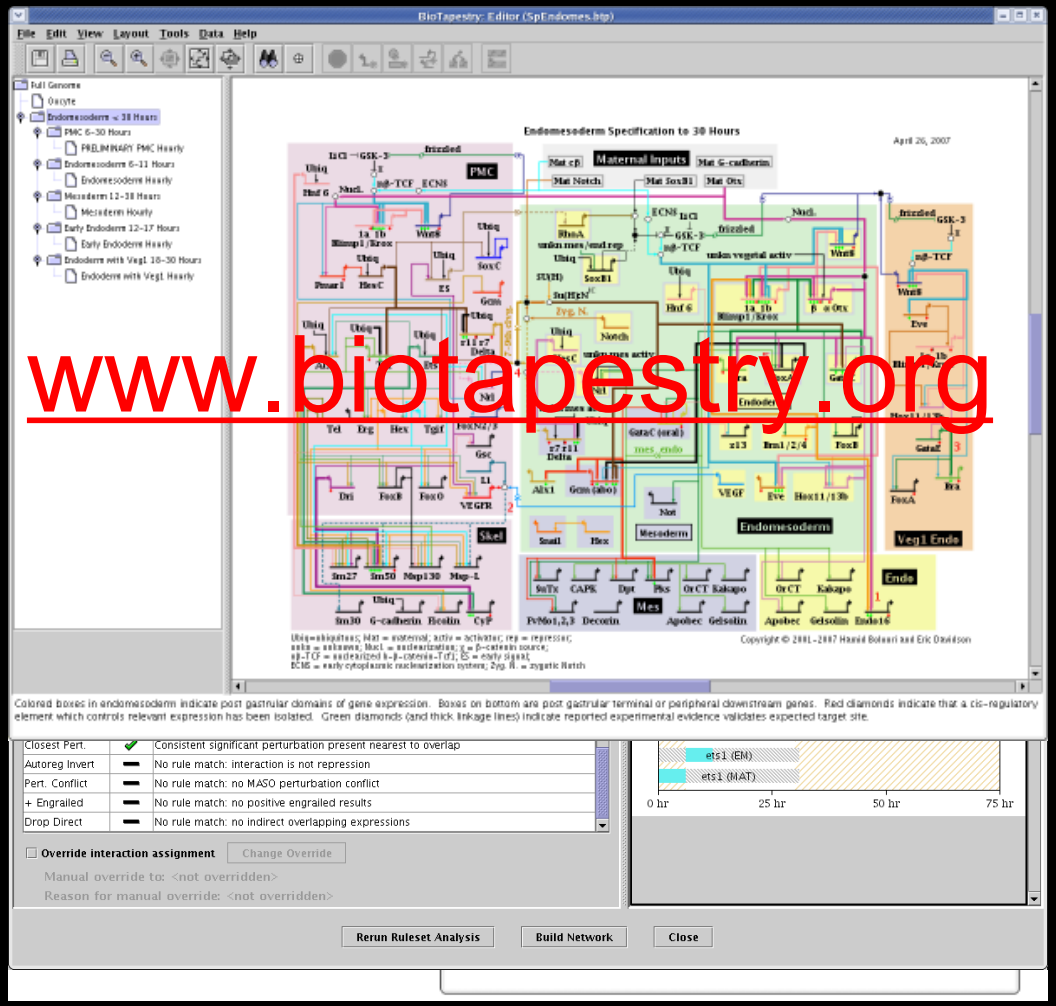
BioTapestry

Graphical application for building & visualising gene regulatory networks

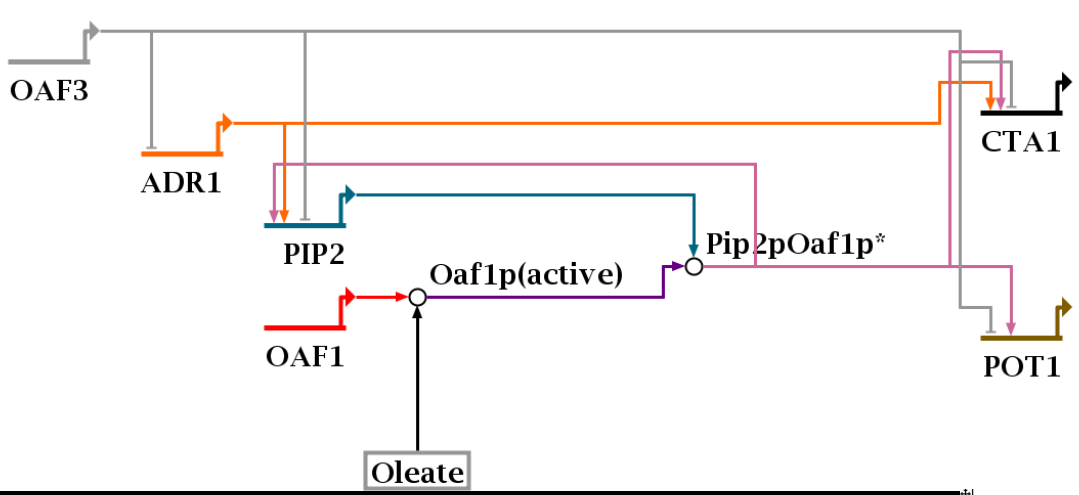
Hierarchical network model for spatially and temporally complex network activation programs

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

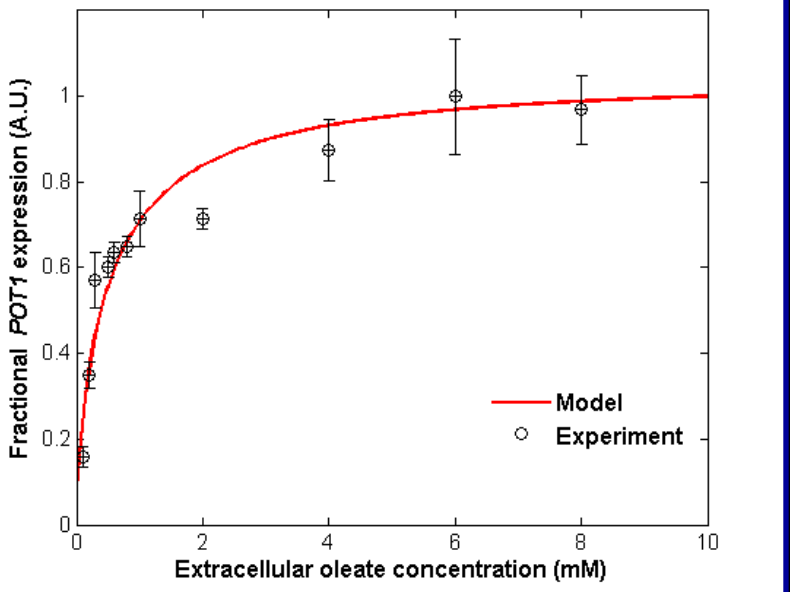
- Build networks from high-throughput data using worksheet feature (under development)



W. Longabaugh



modeling the oleate-response transcriptional network



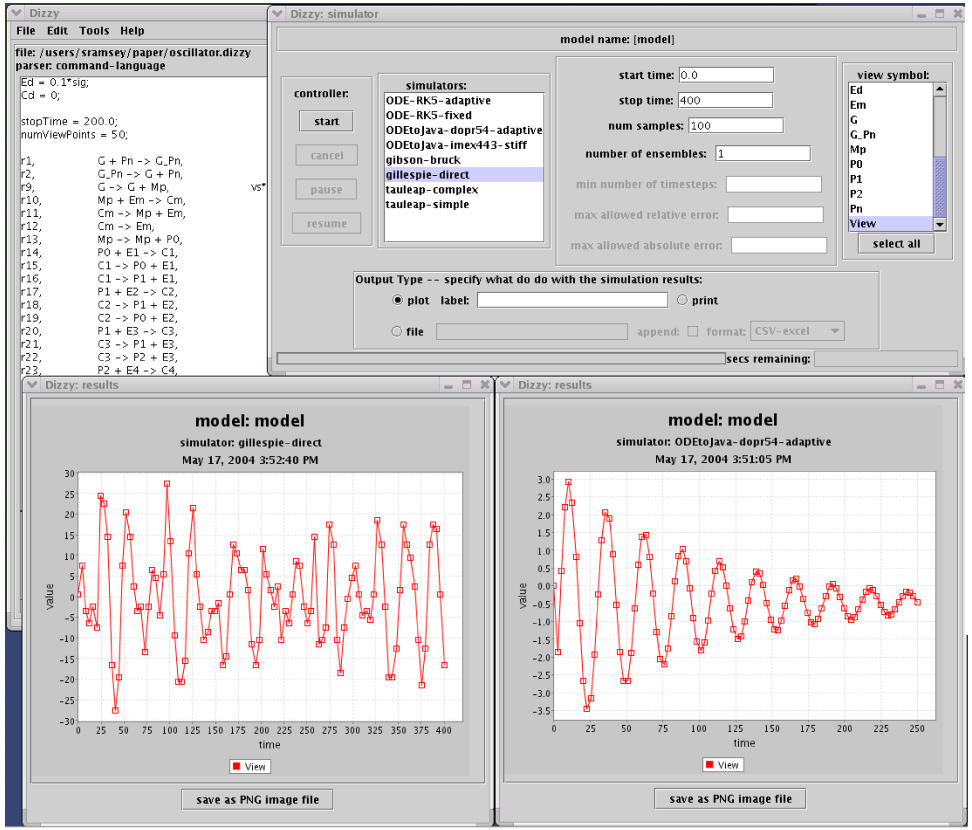
Steady-state dose-response

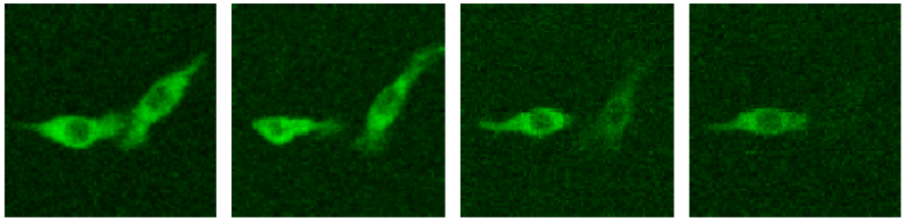
$\pi_o(t) = \frac{Fp_o(t)O_{ic}}{K_{M,o} + O_{ic}}$	Activation of Oaf1p by intracellular oleate (π_o denotes activated Oaf1p)
$h(t) = \left(\frac{Q}{2K_{D,h}}\right)\left(1 + \frac{Q}{K_{D,h}}(\pi_o + p_p) - \sqrt{\left(1 + \frac{Q}{K_{D,h}}(\pi_o + p_p)\right)^2 - 4\left(\frac{Q}{K_{D,h}}\right)\pi_o p_p}\right)$	Heterdimerization 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.
$\frac{dr_\gamma(t)}{dt} = k_{i,r,\gamma} f_{r,\gamma}(t - \tau_{lr}) - k_{d,r,\gamma} r_\gamma(t), \quad \forall \gamma \in \{a,c,o,y,p,f\}$	Gene expression
$\frac{dp_\gamma(t)}{dt} = k_{i,p,\gamma} r_\gamma(t - \tau_{lp}) - k_{d,p,\gamma} p_\gamma(t), \quad \forall \gamma \in \{a,c,o,y,p,f\}$	Protein expression
$f_{r,a} = \left(\frac{1}{r_{x,a}}\right)\left[r_{g,a} + S(r_{o,a} - r_{g,a})\right]$	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
$f_{r,o} = \left(\frac{1}{r_{x,o}}\right)\left[r_{g,o} + S(r_{o,o} - r_{g,o})\right]$	Fractional activity of *OAF1* gene
$f_{r,y} = \left(\frac{1}{r_{x,y}}\right)\left[r_{g,y} + S(r_{o,y} - r_{g,y})\right]$	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

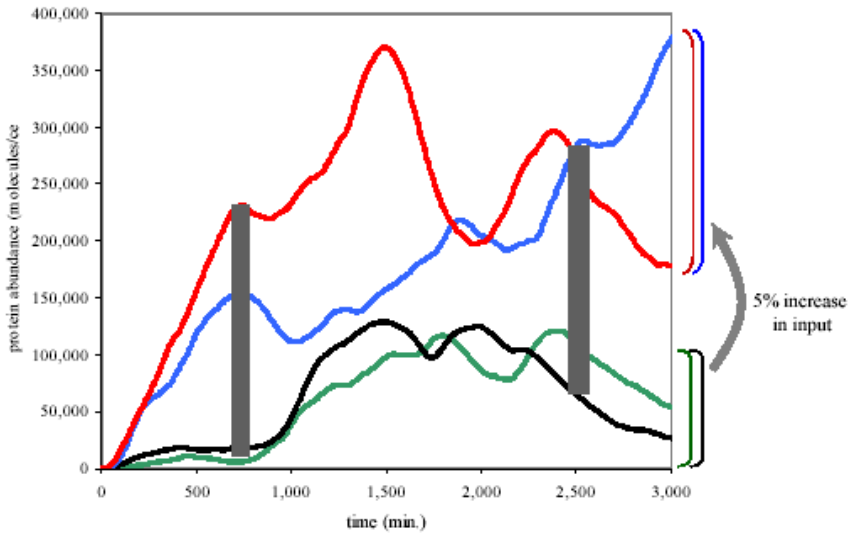
S. Ramsey, A. Ratushny, J. Smith, J. Aitchison

Network Simulation

Stochastic Simulation (Dizzy)



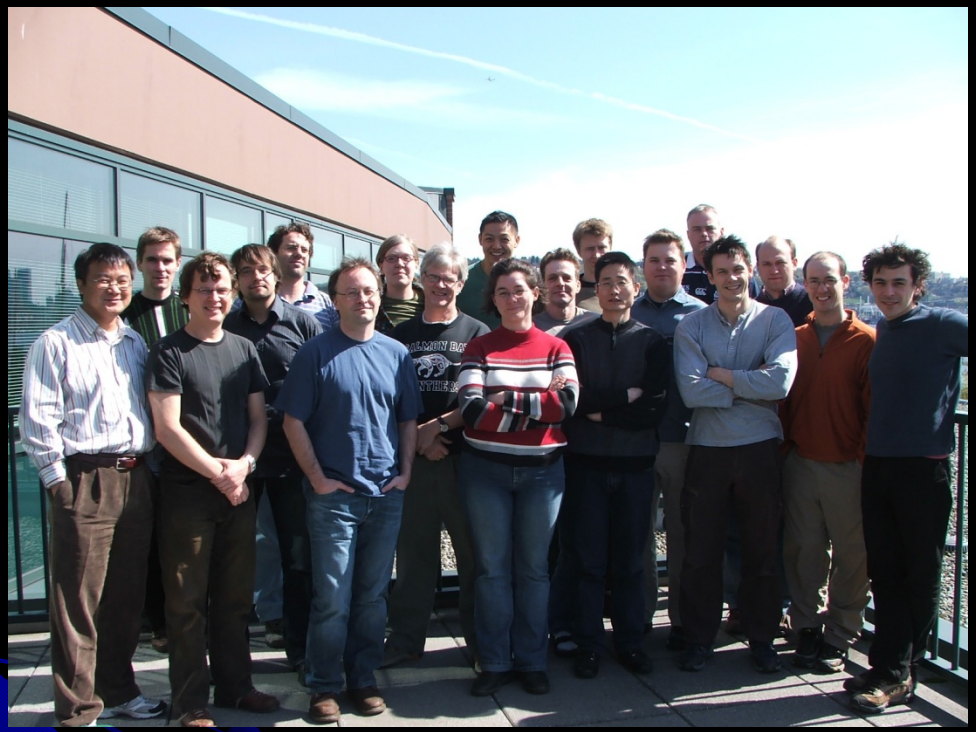




Individual macrophage cells show different levels of IkBa:GFP after stimulation by dsRNA (TLR3 ligand)

S. Ramsey

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