INSTITUTE FOR Systems Biology

Modeling and Inference of Transcriptional Regulatory Networks

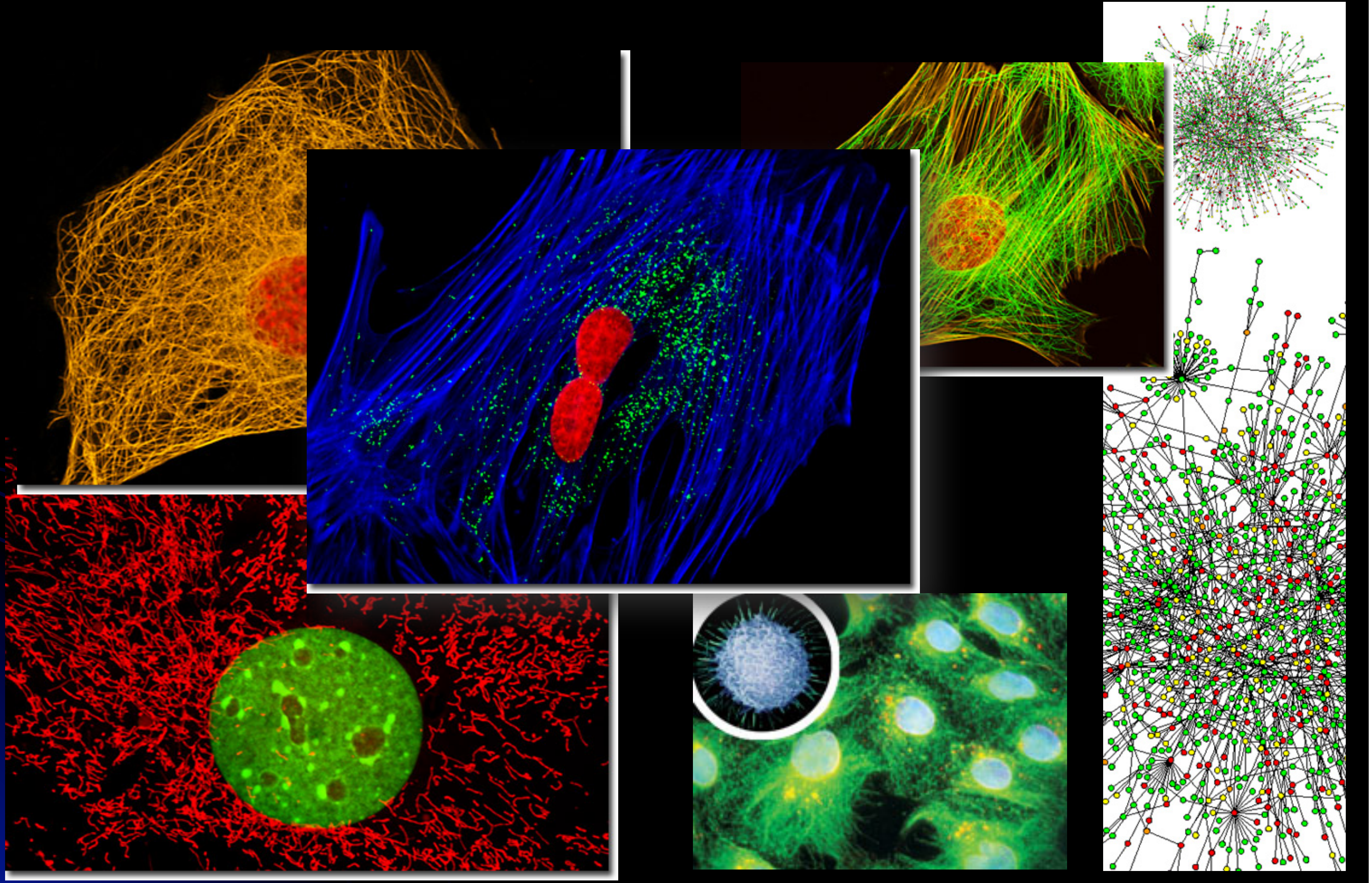
ilya shmulevich

INSTITUTE FOR Systems Biology

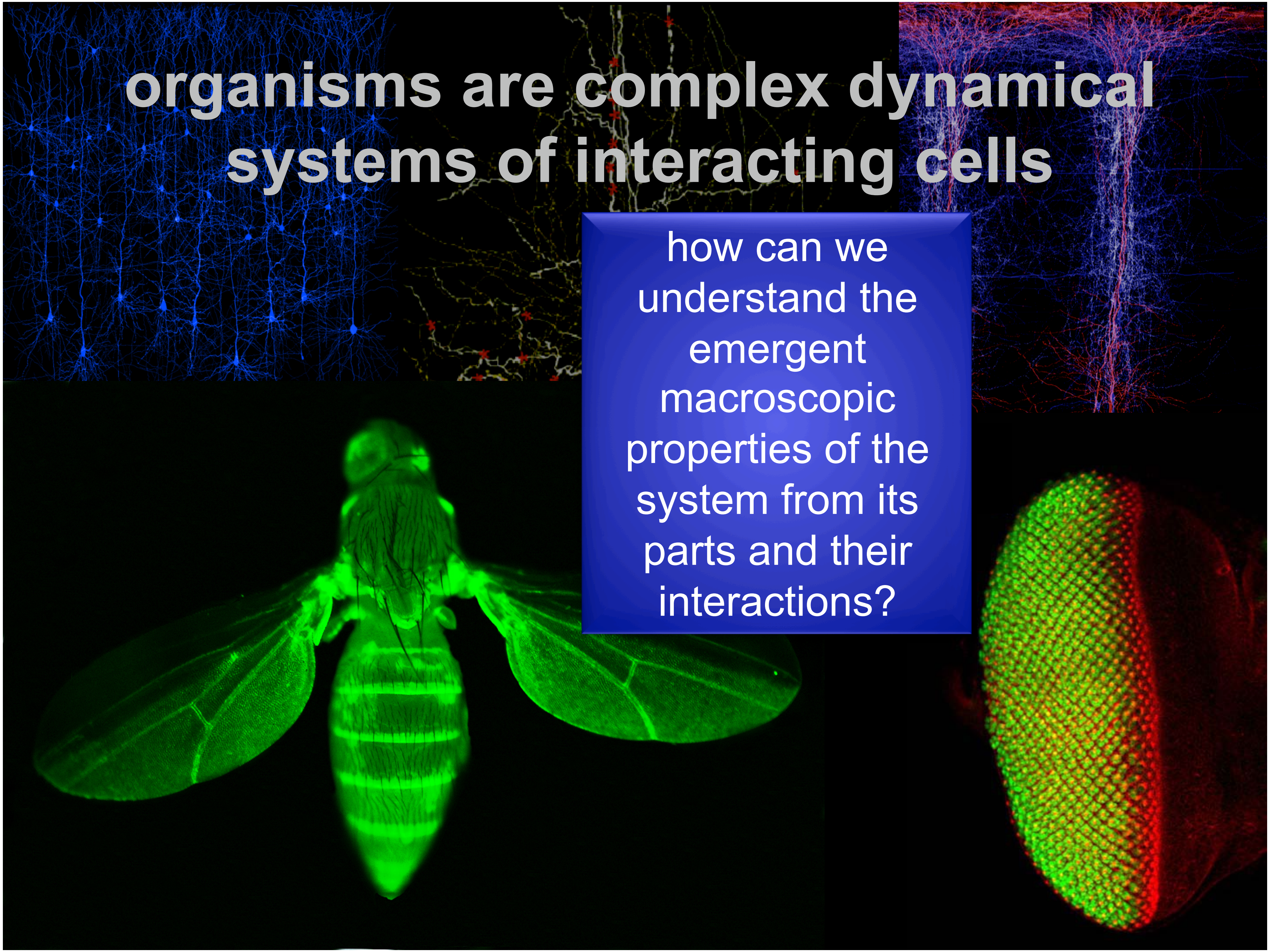
FRED HUTCHINSON CANCER RESEARCH CENTER

A LIFE OF SCIENCE

cells are complex dynamical systems of interacting molecules



organisms are complex dynamical systems of interacting cells

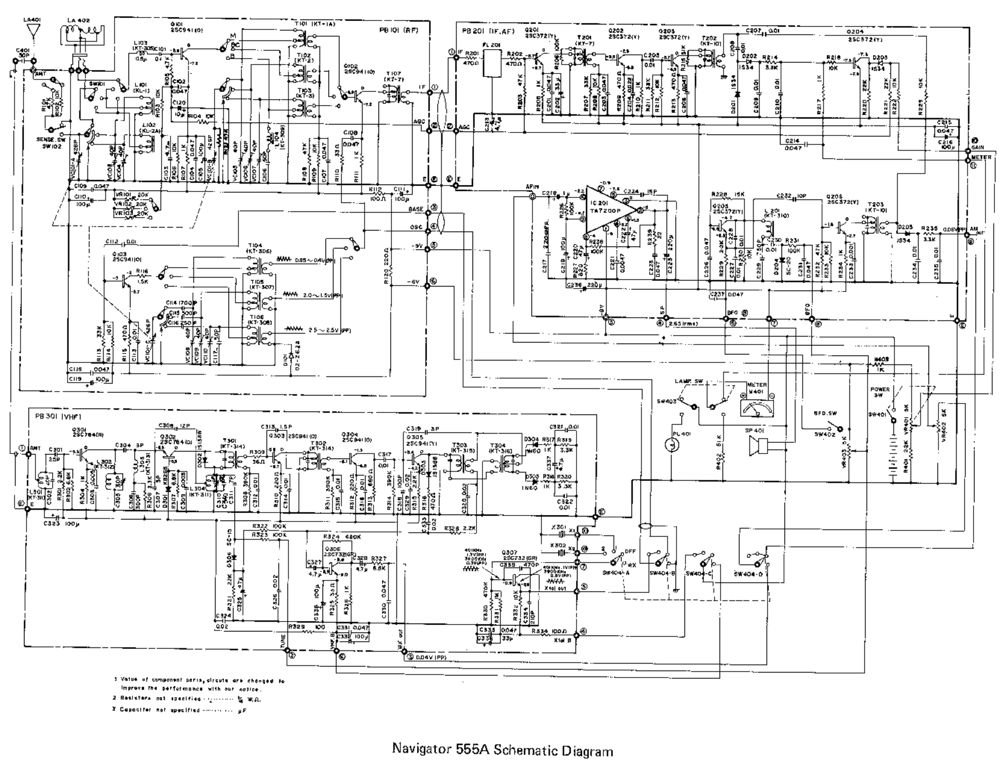


how can we understand the emergent macroscopic properties of the system from its parts and their interactions?

genetic networks

Complex dynamical regulatory networks among genes and their products control cell behaviors such as:

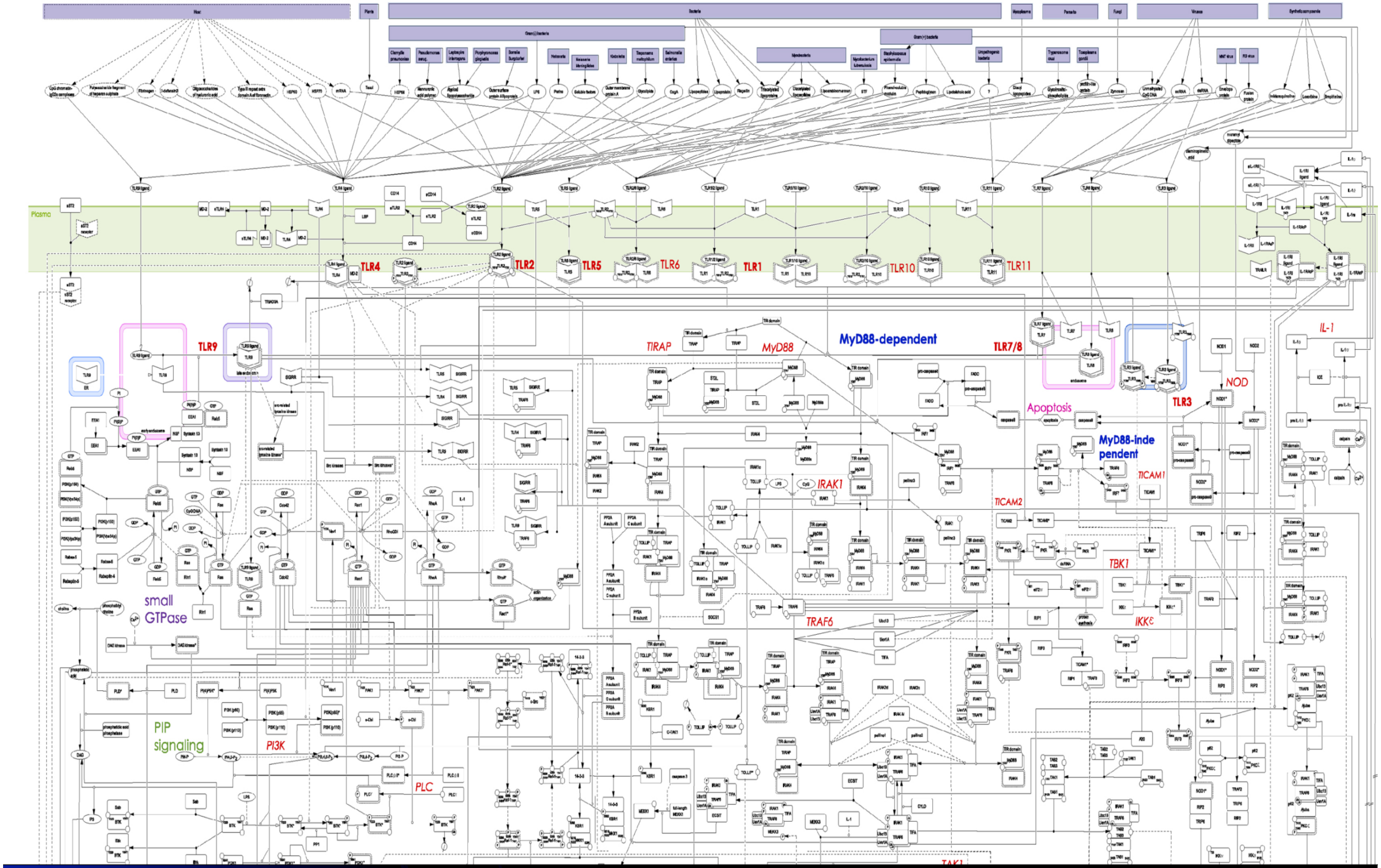
- cell cycle
- apoptosis
- cell differentiation
- communication between cells in tissues



Navigator 555A Schematic Diagram

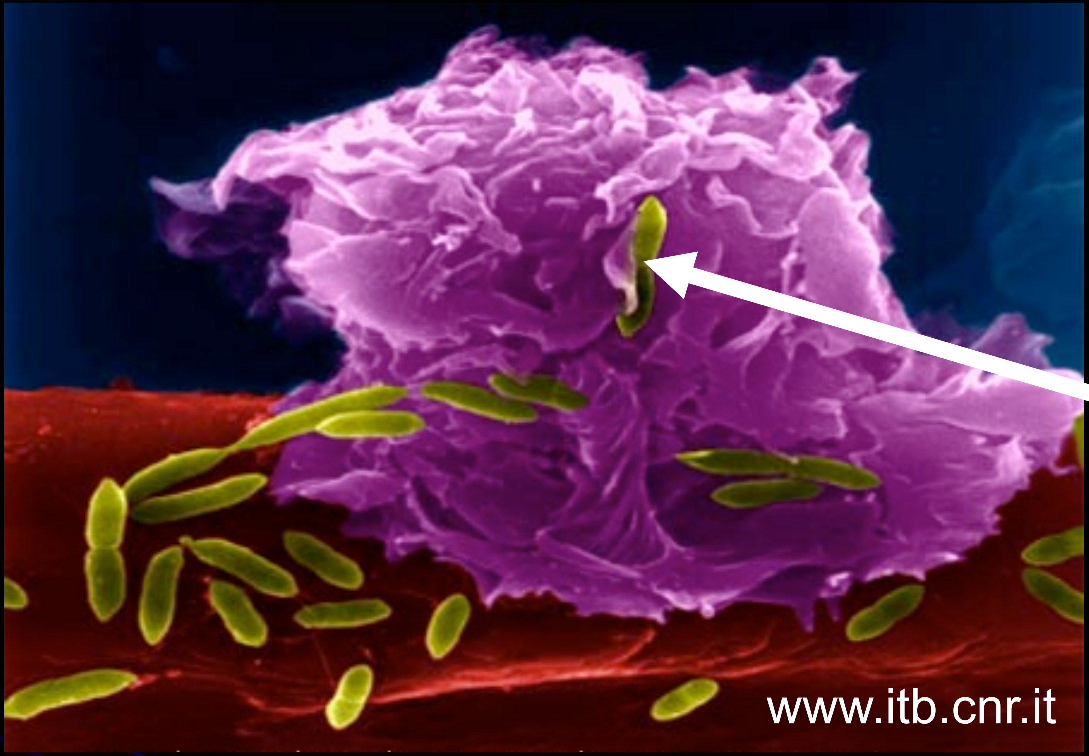
Analogy: circuit diagram

Map of the TLR signaling pathway in the macrophage



Mol Syst Biol. 2:2006.0015

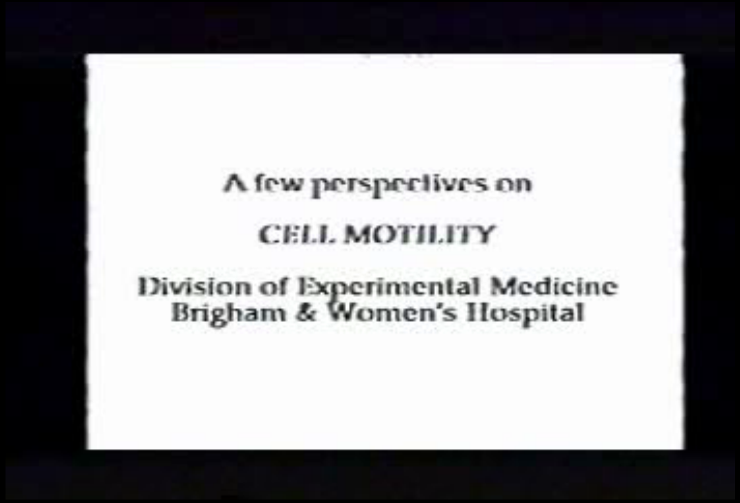
Macrophages are key immune cells



www.itb.cnr.it

bacterium
being engulfed

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



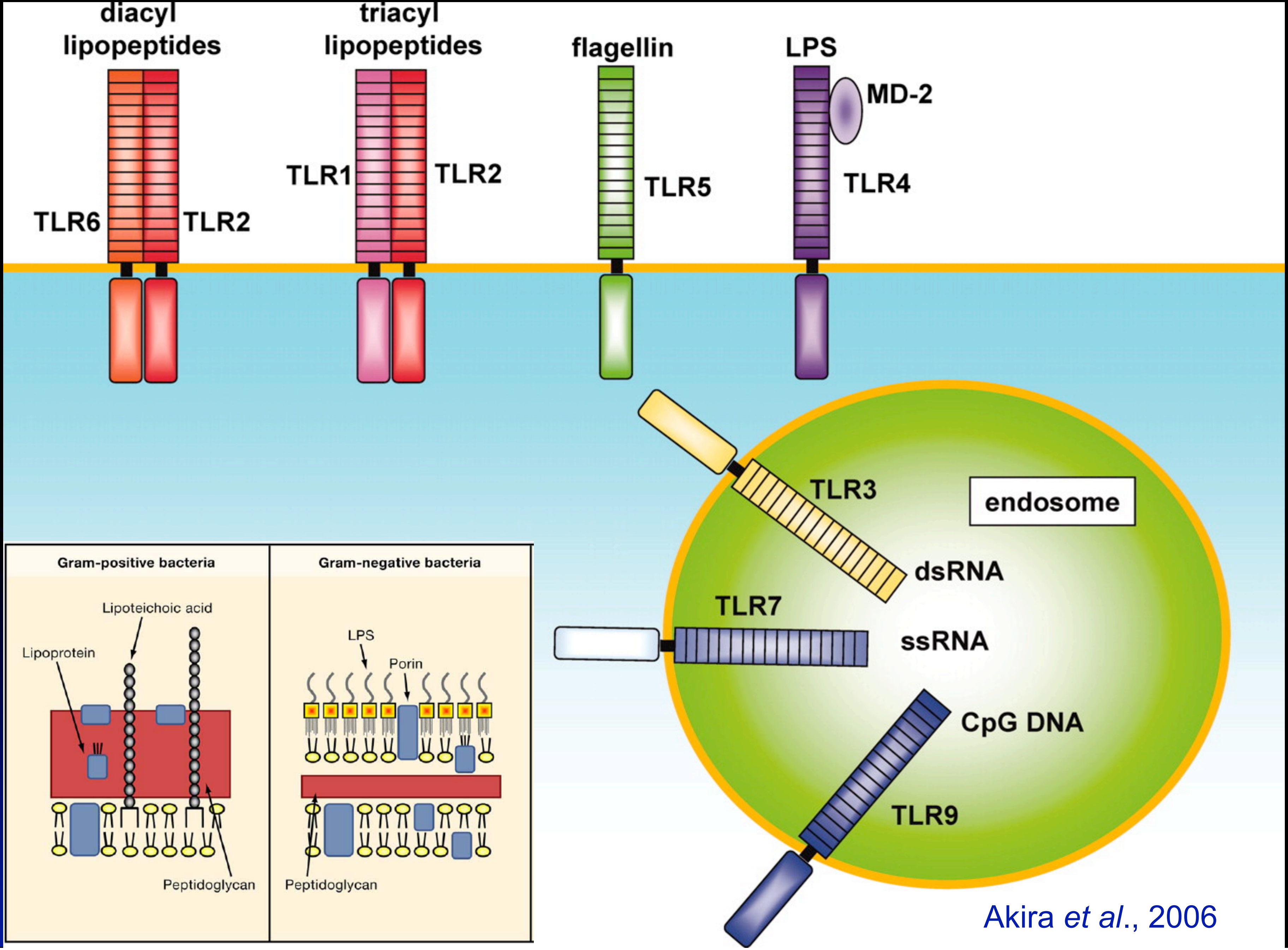
A few perspectives on

CELL MOTILITY

Division of Experimental Medicine
Brigham & Women's Hospital

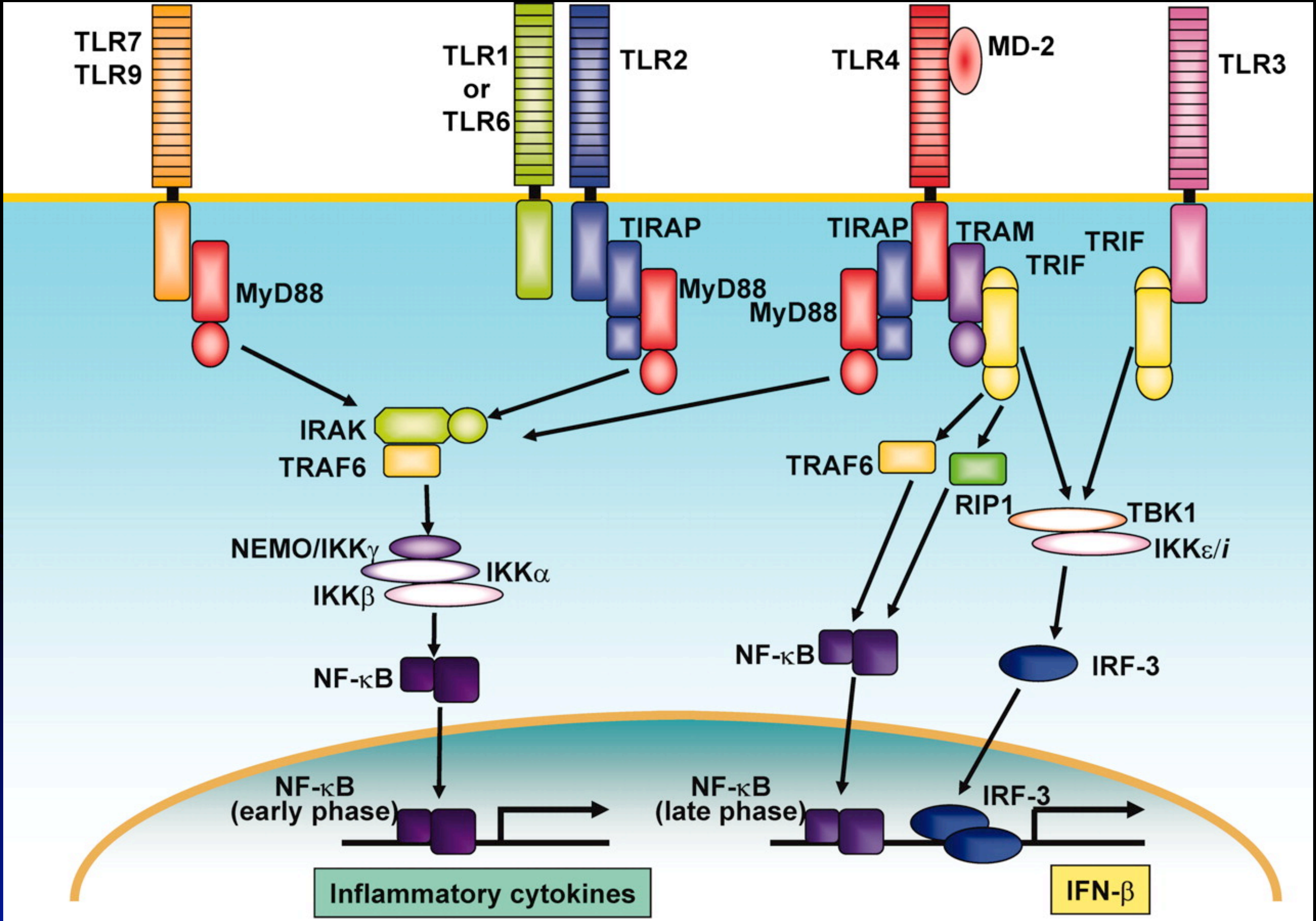
(thanks to Alan Aderem)

TLRs and their associated PAMPs



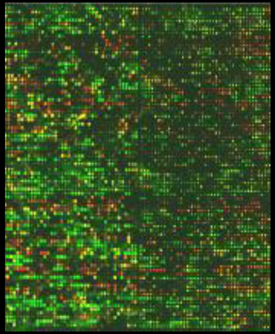
Akira *et al*., 2006

TLR signaling pathway

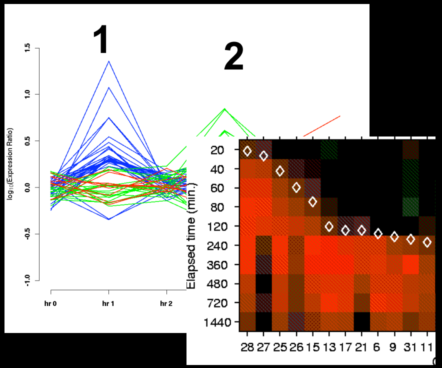


Akira *et al.*, 2006

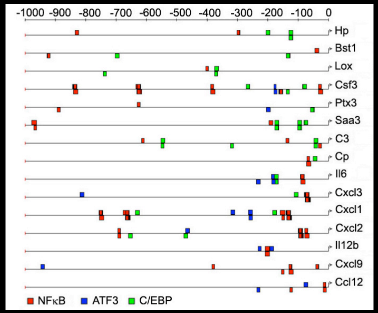
Transcriptional Networks: Strategy



Expression data

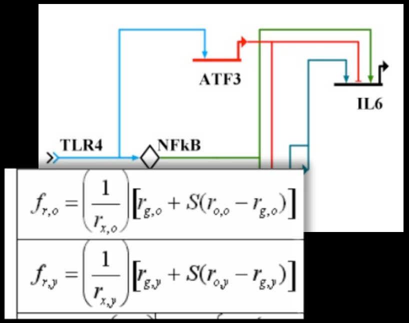


Temporal Analysis of Expression Data



Promoter analysis of binding sites

ChIP to validate targets

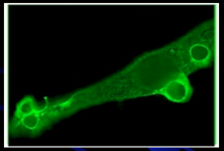


$$f_{r,o} = \left(\frac{1}{r_{x,o}}\right)\left[r_{g,o} + S(r_{o,o} - r_{g,o})\right]$$

$$f_{r,y} = \left(\frac{1}{r_{x,y}}\right)\left[r_{g,y} + S(r_{o,y} - r_{g,y})\right]$$

Modeling to predict function

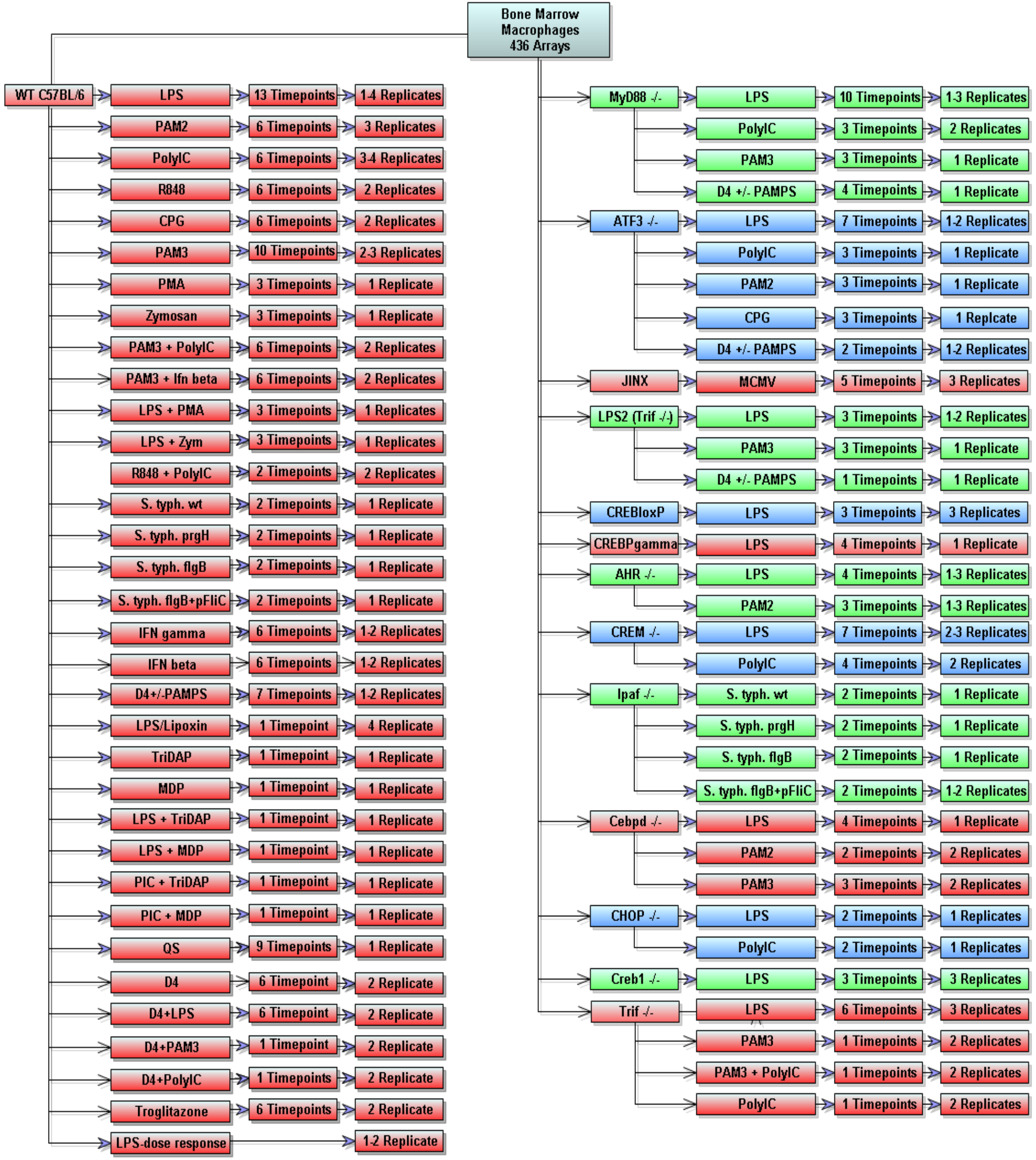
in vitro



Biological Validation

in vivo

The ingredients



Bone Marrow Macrophages 436 Arrays

WT C57BL/6	LPS	13 Timepoints	1-4 Replicates
	PAM2	6 Timepoints	3 Replicates
	PolyIC	6 Timepoints	3-4 Replicates
	R848	6 Timepoints	2 Replicates
	CPG	6 Timepoints	2 Replicates
	PAM3	10 Timepoints	2-3 Replicates
	PMA	3 Timepoints	1 Replicate
	Zymosan	3 Timepoints	1 Replicate
	PAM3 + PolyIC	6 Timepoints	2 Replicates
	PAM3 + Ifn beta	6 Timepoints	2 Replicates
	LPS + PMA	3 Timepoints	1 Replicates
	LPS + Zym	3 Timepoints	1 Replicates
	R848 + PolyIC	2 Timepoints	2 Replicates
	S. typh. wt	2 Timepoints	1 Replicate
	S. typh. prgH	2 Timepoints	1 Replicate
	S. typh. flgB	2 Timepoints	1 Replicate
	S. typh. flgB+pFliC	2 Timepoints	1 Replicate
	IFN gamma	6 Timepoints	1-2 Replicates
	IFN beta	6 Timepoints	1-2 Replicates
	D4+/-PAMPS	7 Timepoints	1-2 Replicates
	LPS/Lipoxin	1 Timepoint	4 Replicate
	TriDAP	1 Timepoint	1 Replicate
	MDP	1 Timepoint	1 Replicate
	LPS + TriDAP	1 Timepoint	1 Replicate
	LPS + MDP	1 Timepoint	1 Replicate
	PIC + TriDAP	1 Timepoint	1 Replicate
	PIC + MDP	1 Timepoint	1 Replicate
	QS	9 Timepoints	1 Replicate
	D4	6 Timepoint	2 Replicate
	D4+LPS	6 Timepoint	2 Replicate
	D4+PAM3	1 Timepoint	2 Replicate
	D4+PolyIC	1 Timepoints	2 Replicate
	Troglitazone	6 Timepoints	2 Replicate
	LPS-dose response		1-2 Replicate
MyD88 -/-	LPS	10 Timepoints	1-3 Replicates
	PolyIC	3 Timepoints	2 Replicates
	PAM3	3 Timepoints	1 Replicate
	D4 +/- PAMPS	4 Timepoints	1 Replicate
ATF3 -/-	LPS	7 Timepoints	1-2 Replicates
	PolyIC	3 Timepoints	1 Replicate
	PAM2	3 Timepoints	1 Replicate
	CPG	3 Timepoints	1 Replcate
	D4 +/- PAMPS	2 Timepoints	1-2 Replicates
JINX	MCMV	5 Timepoints	3 Replicates
LPS2 (Trif -/-)	LPS	3 Timepoints	1-2 Replicates
	PAM3	3 Timepoints	1 Replicate
	D4 +/- PAMPS	1 Timepoints	1 Replicate
CREBloxP	LPS	3 Timepoints	3 Replicates
CREBPgamma	LPS	4 Timepoints	1 Replicate
AHR -/-	LPS	4 Timepoints	1-3 Replicates
	PAM2	3 Timepoints	1-3 Replicates
CREM -/-	LPS	7 Timepoints	2-3 Replicates
	PolyIC	4 Timepoints	2 Replicates
Ipaf -/-	S. typh. wt	2 Timepoints	1 Replicate
	S. typh. prgH	2 Timepoints	1 Replicate
	S. typh. flgB	2 Timepoints	1 Replicate
	S. typh. flgB+pFliC	2 Timepoints	1-2 Replicates
Cebpd -/-	LPS	4 Timepoints	1 Replicate
	PAM2	2 Timepoints	2 Replicates
	PAM3	3 Timepoints	2 Replicates
CHOP -/-	LPS	2 Timepoints	1 Replicates
	PolyIC	2 Timepoints	1 Replicates
Creb1 -/-	LPS	3 Timepoints	3 Replicates
Trif -/-	LPS	6 Timepoints	3 Replicates
	PAM3	1 Timepoints	2 Replicates
	PAM3 + PolyIC	1 Timepoints	2 Replicates
	PolyIC	1 Timepoints	2 Replicates

130 microarray experiments (253 arrays)

Time courses out to 8 hours (24 for LPS)

Seven mouse strains:

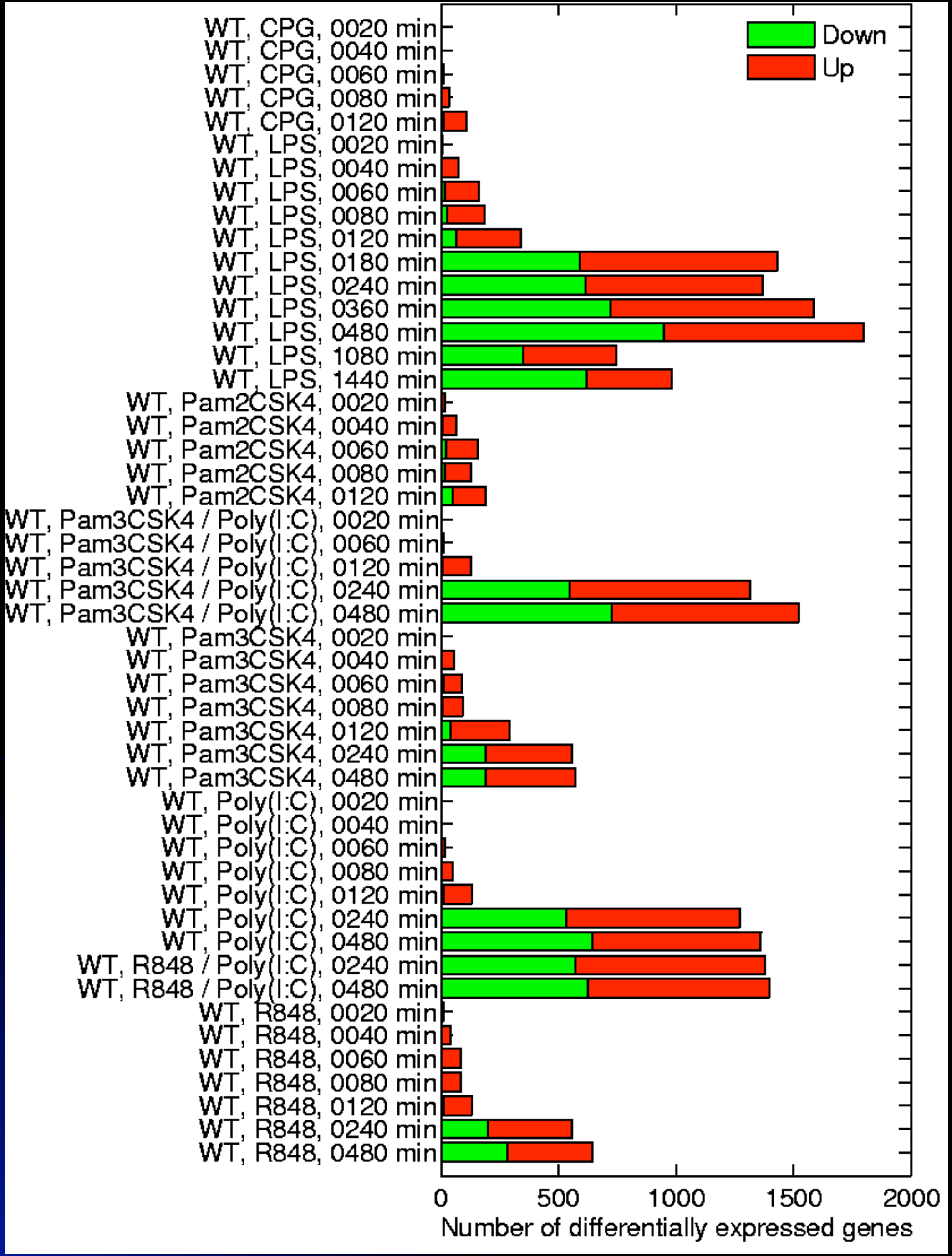
WT, Ahr$^{(-/-)}$, Atf3$^{(-/-),}$ Crem$^{(-/-)}$, Cebpd$^{(-/-)}$, Myd88$^{(-/-)}$, Ticam1$^{(Lps2/Lps2)}$

Combinations of six stimuli

LPS, Pam$_3$CSK$_4$, Pam$_2$CSK$_4$, poly I:C, CpG, R848, T091317

Mouse genome promoters (UCSC)

TRANSFAC Professional 10.3	Curated list of ~1800 human TFs

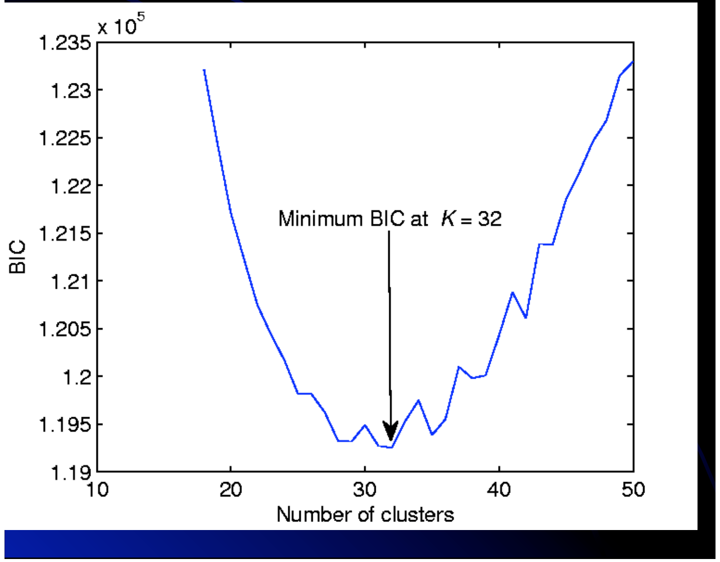


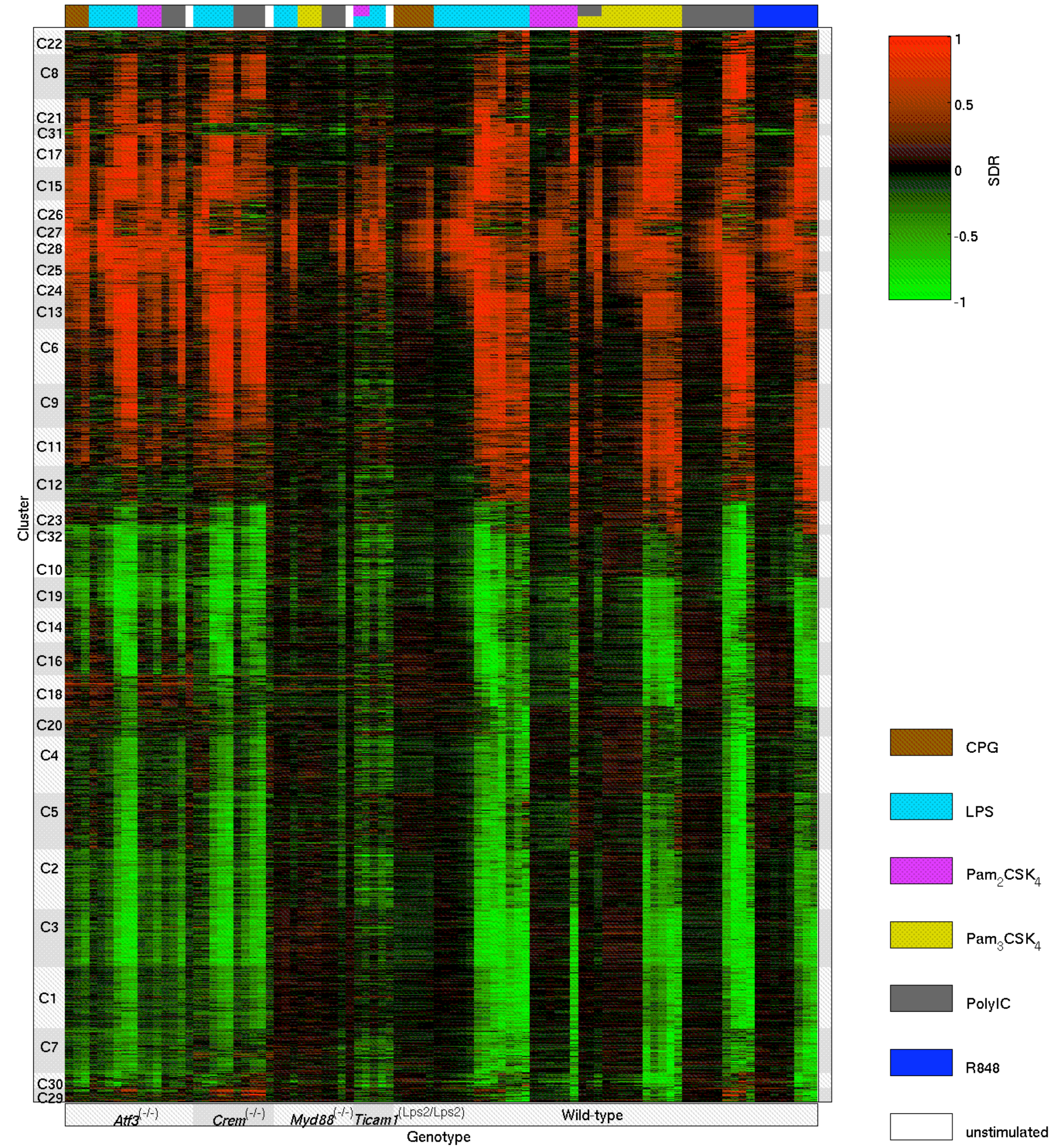
Number of differentially expressed genes vs. elapsed time (by stimulus)

Total number: 2,562

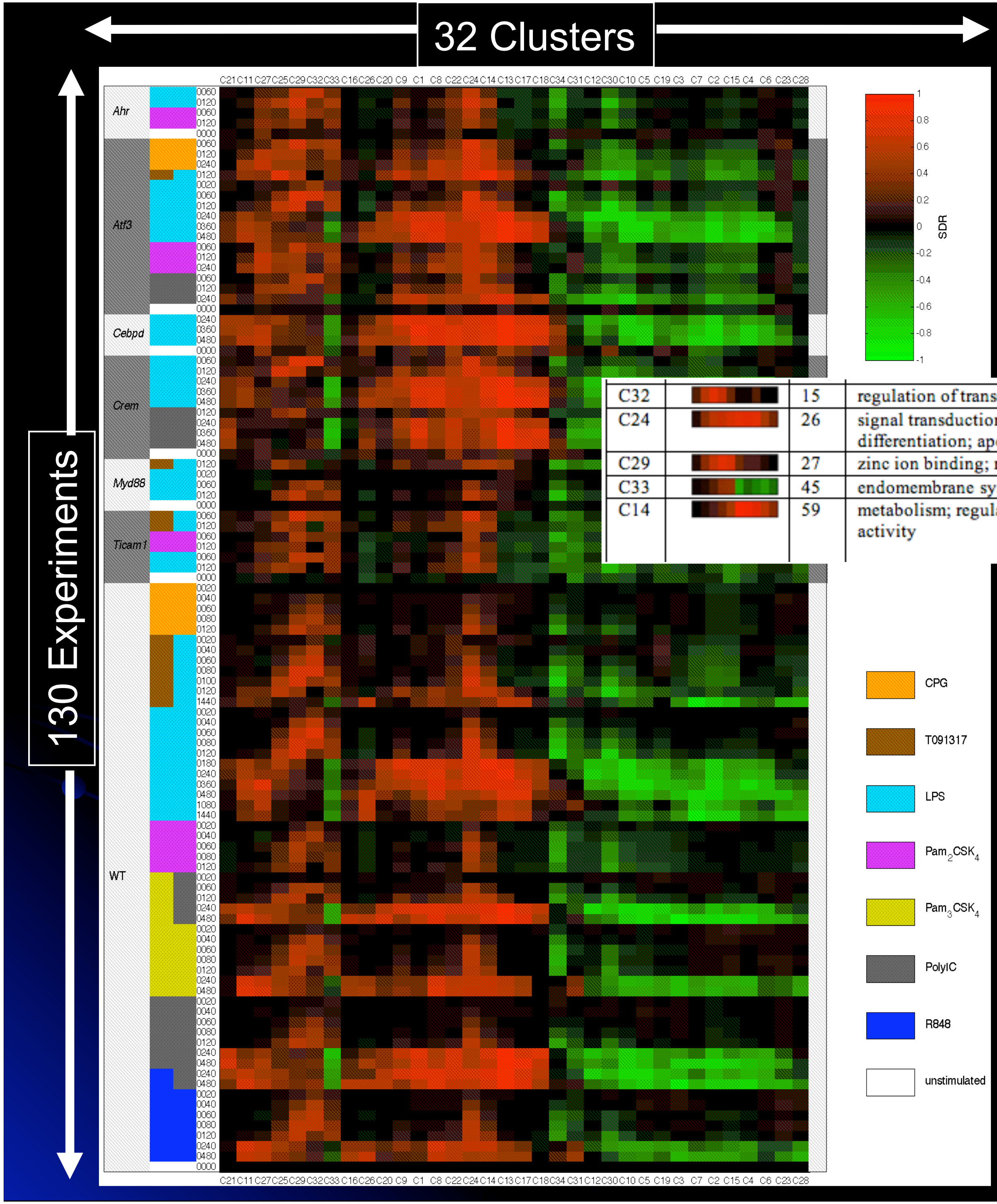
S. Ramsey, PLoS Comp Bio, 4,3, 2008.

clustering of gene expression profiles

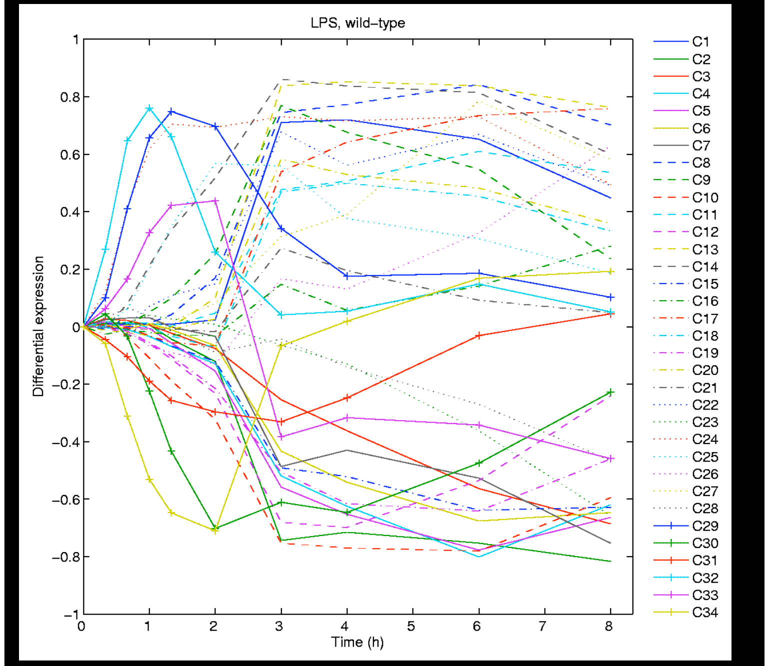




Cluster summarized expression profiles

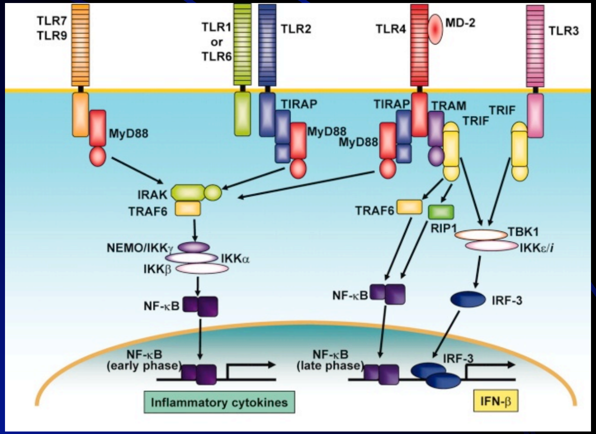


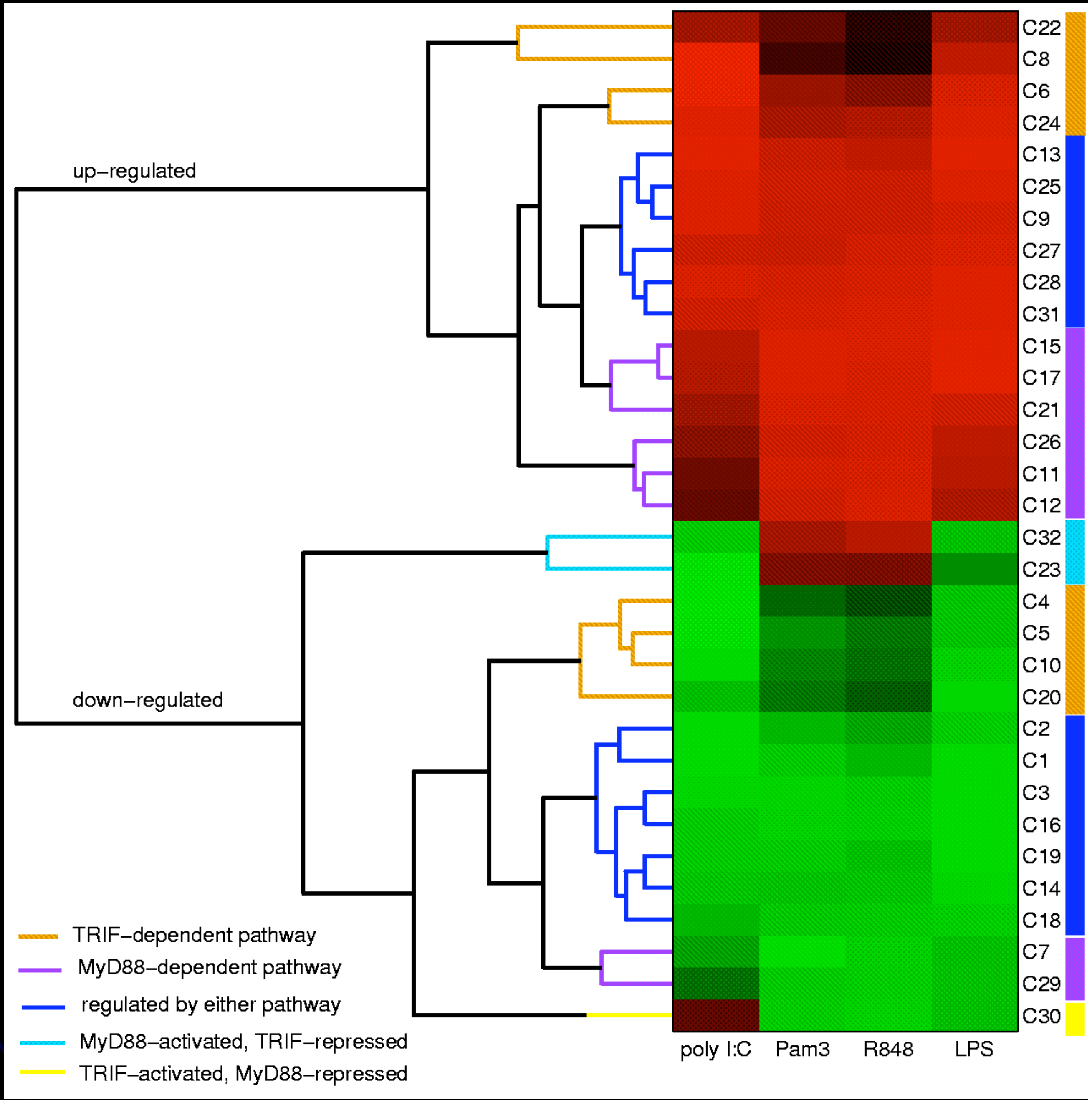
C32		15	regulation of transcription; response to stress	Btg2, Ewsr1, Fos, Hipk1, Tgif
C24		26	signal transduction; regulation of transcription; cell differentiation; apoptosis; cytokine activity	Atf3, Ccrn4l, Etv3, Irf1, Junb, Klf6, Nfkbiz, Prdm1
C29		27	zinc ion binding; regulation of transcription	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



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

Stimulus-specific cluster responses

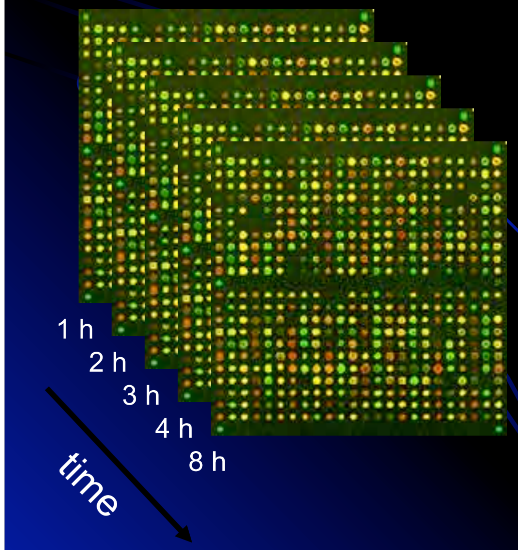




S. Ramsey, PLoS Comp Bio, 4,3, 2008.

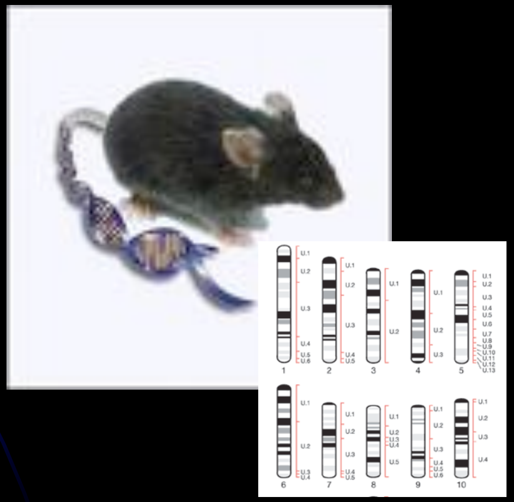
In conjunction with transcription factor binding site prediction, we can use the *timing* of expression to identify induced transcription factors that are associated with downstream groups of genes that they regulate.

Expression *dynamics*



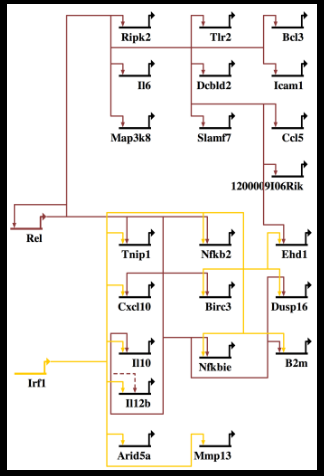
\+

Genomic sequence data



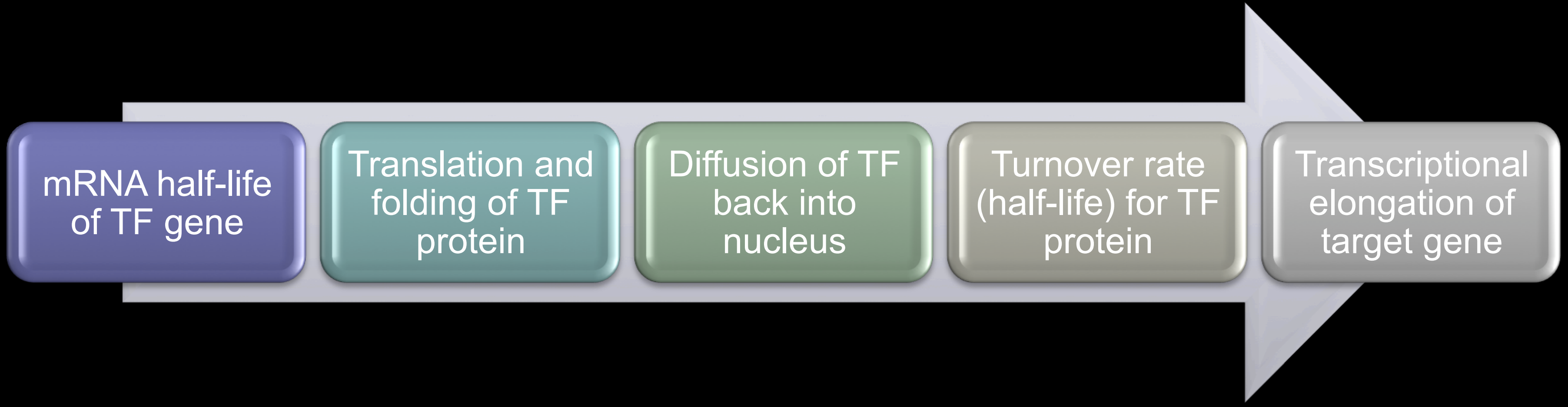
=

Transcriptional network



S. Ramsey

Components of the gene-gene transcriptional time delay



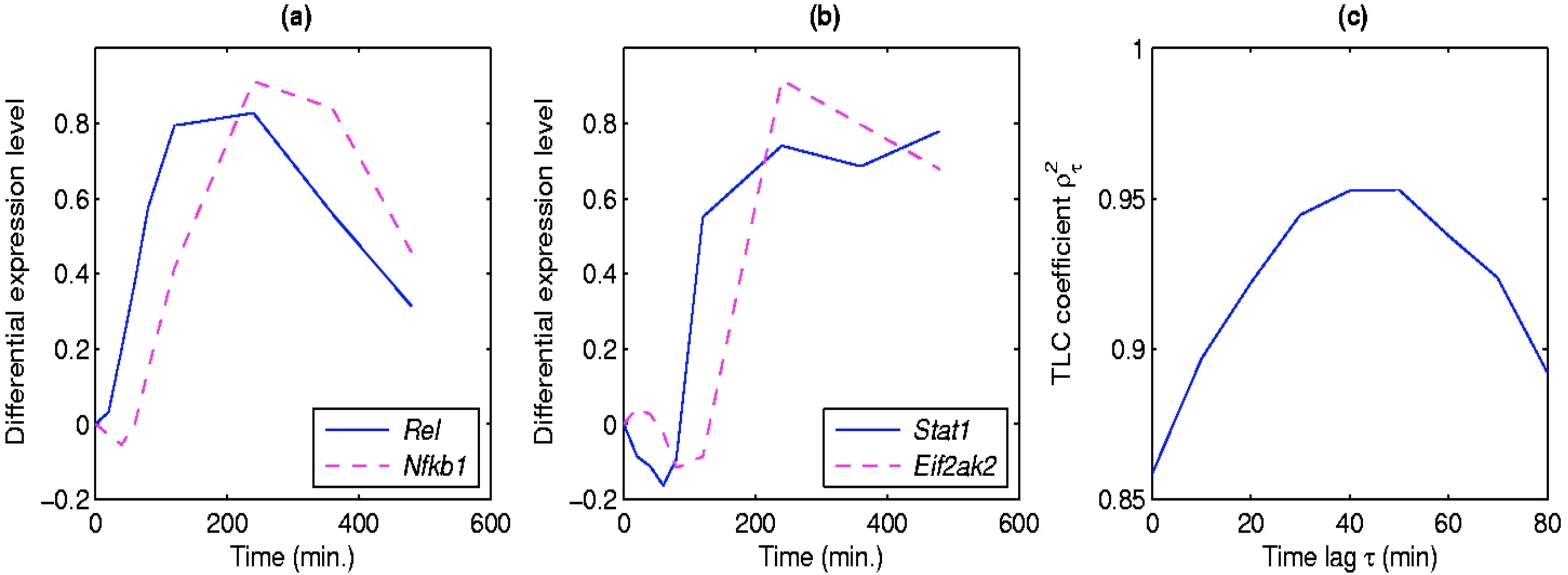
mRNA half-life of TF gene

Translation and folding of TF protein

Diffusion of TF back into nucleus

Turnover rate (half-life) for TF protein

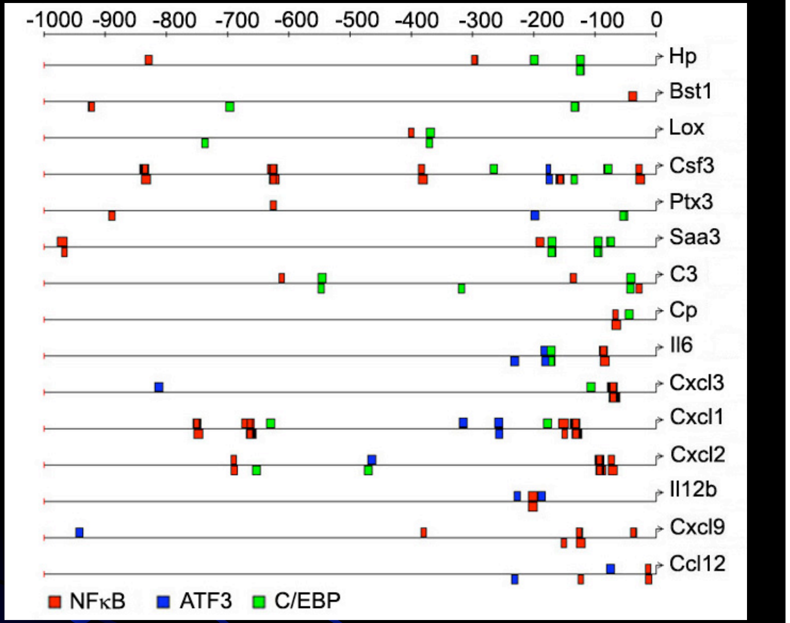
Transcriptional elongation of target gene



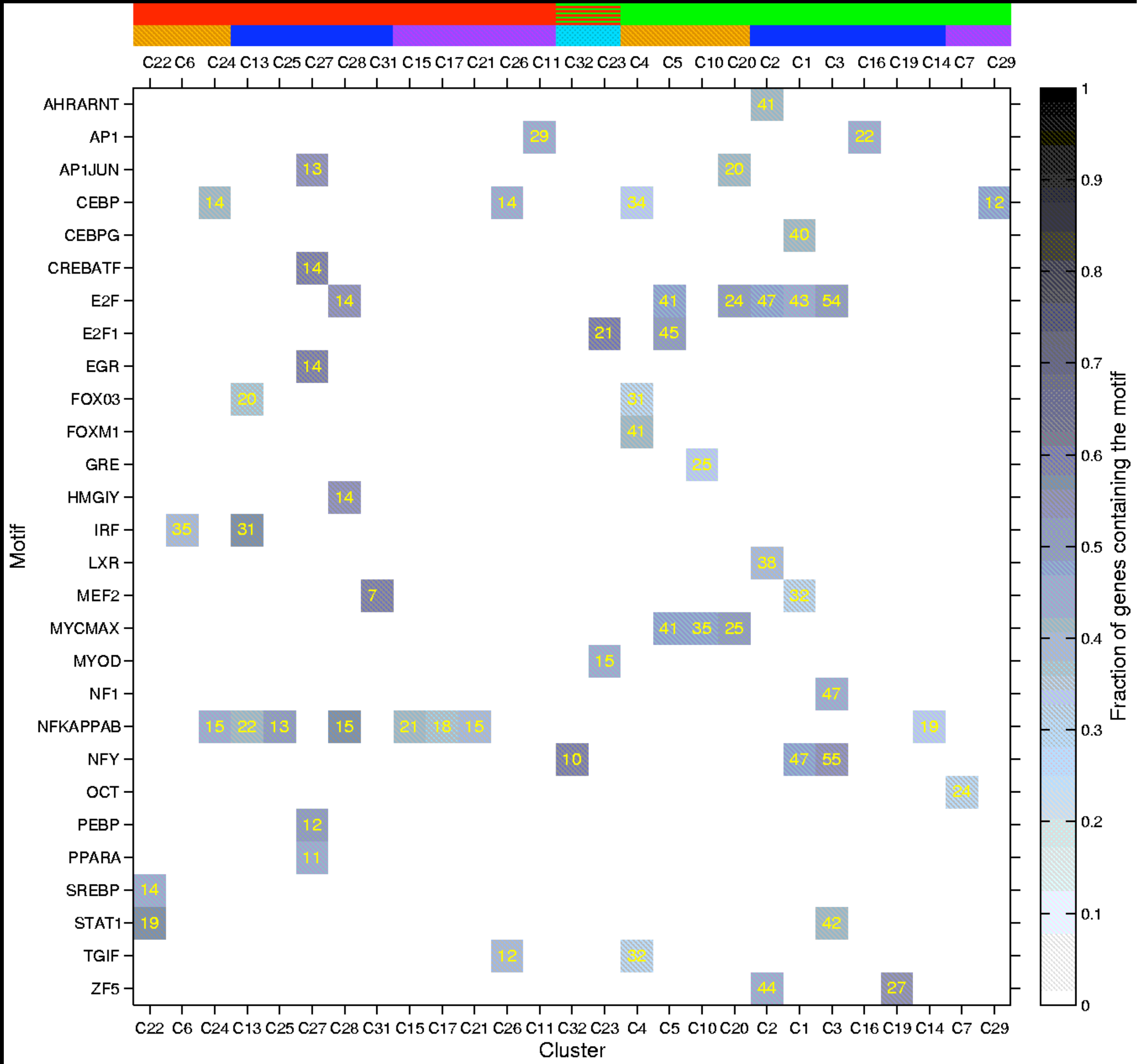
S. Ramsey, PLoS Comp Bio, 4,3, 2008.

Scanning for TF binding motif matches

Scanning:



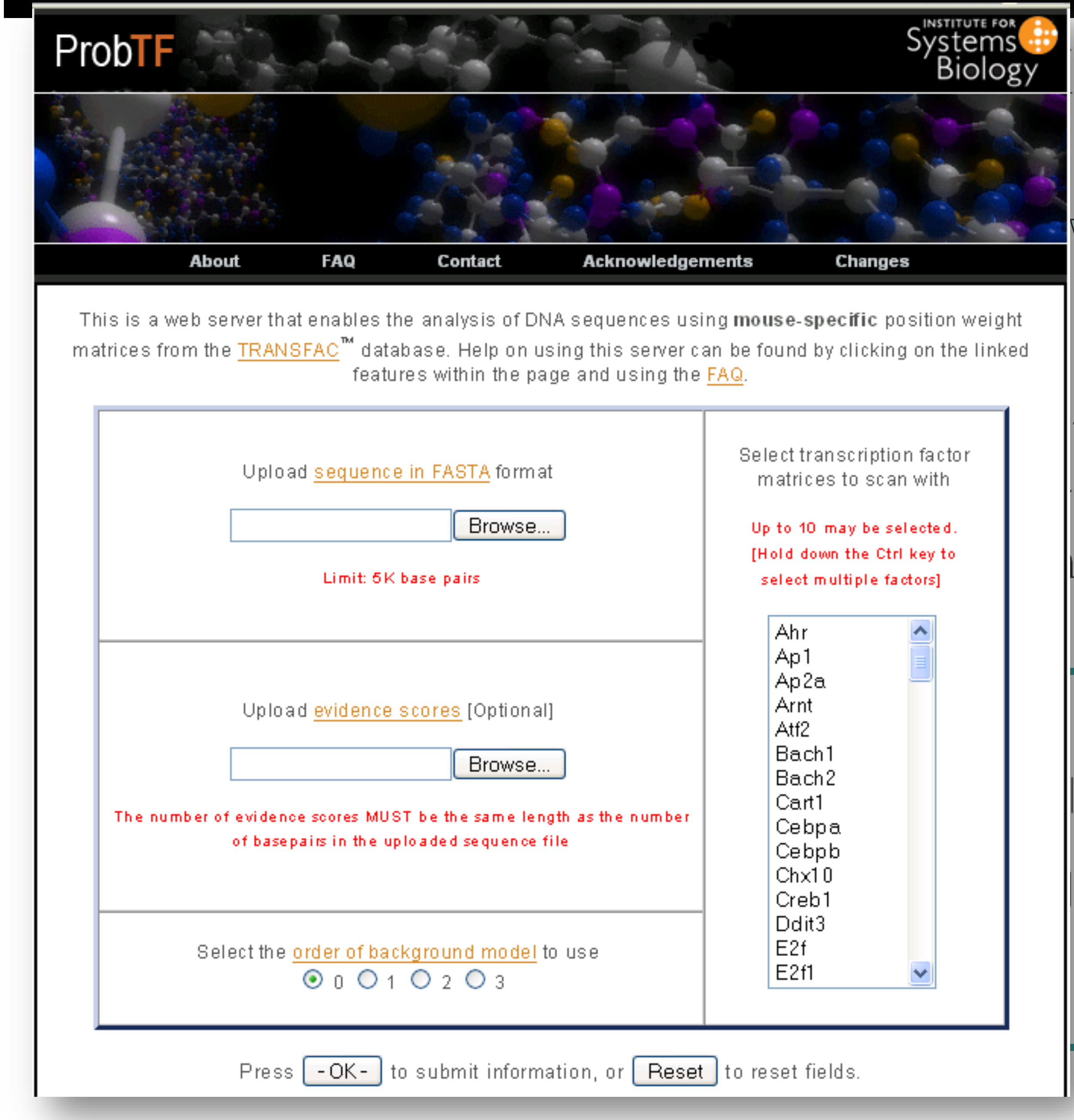
Motif enrichment matrix:



TF recognition site motif:

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

Probabilistic Framework for Transcription Factor Binding Site Prediction



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 - OK - to submit information, or Reset to reset fields.

www.probtf.org

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

round model

Data Fusion

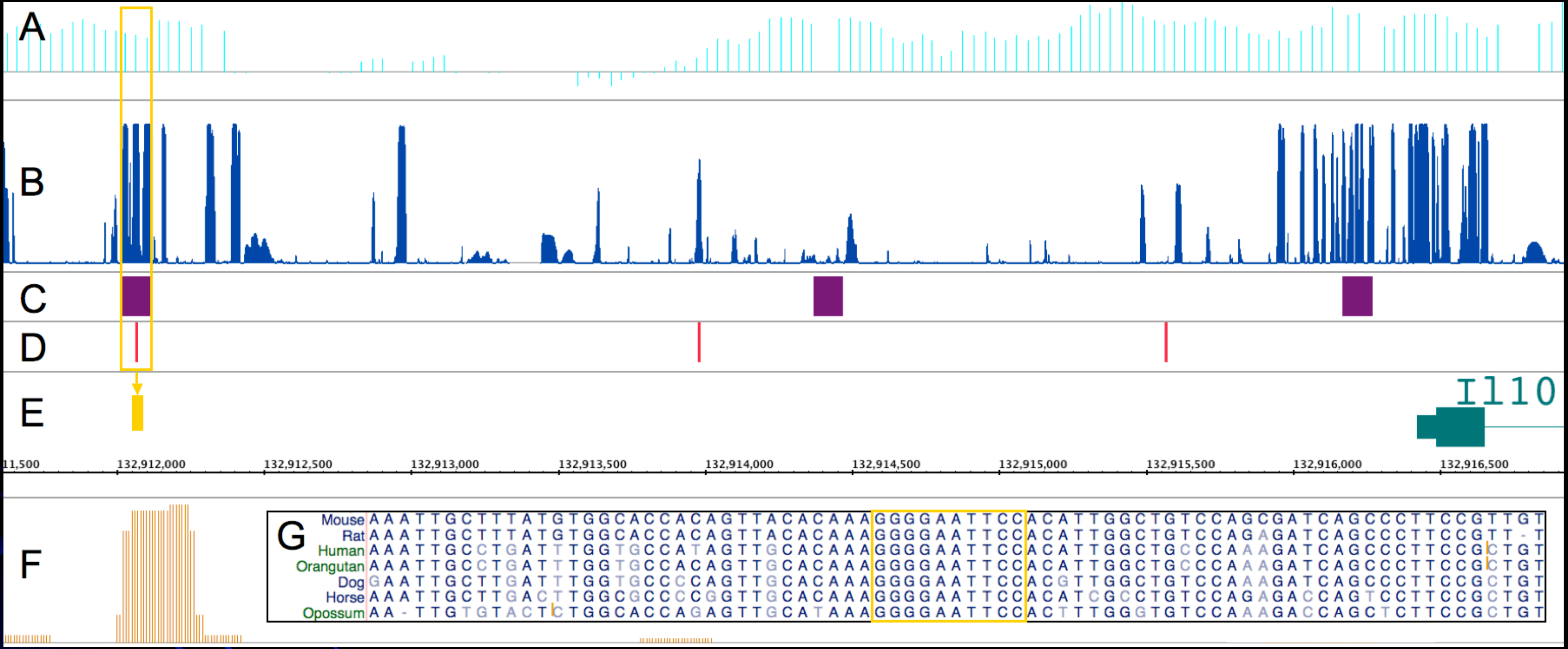
$$|A,\pi,\Theta,\phi) = P(S|A,\pi,\Theta,\phi)P(\mathcal{D}|A,\pi)$$

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

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

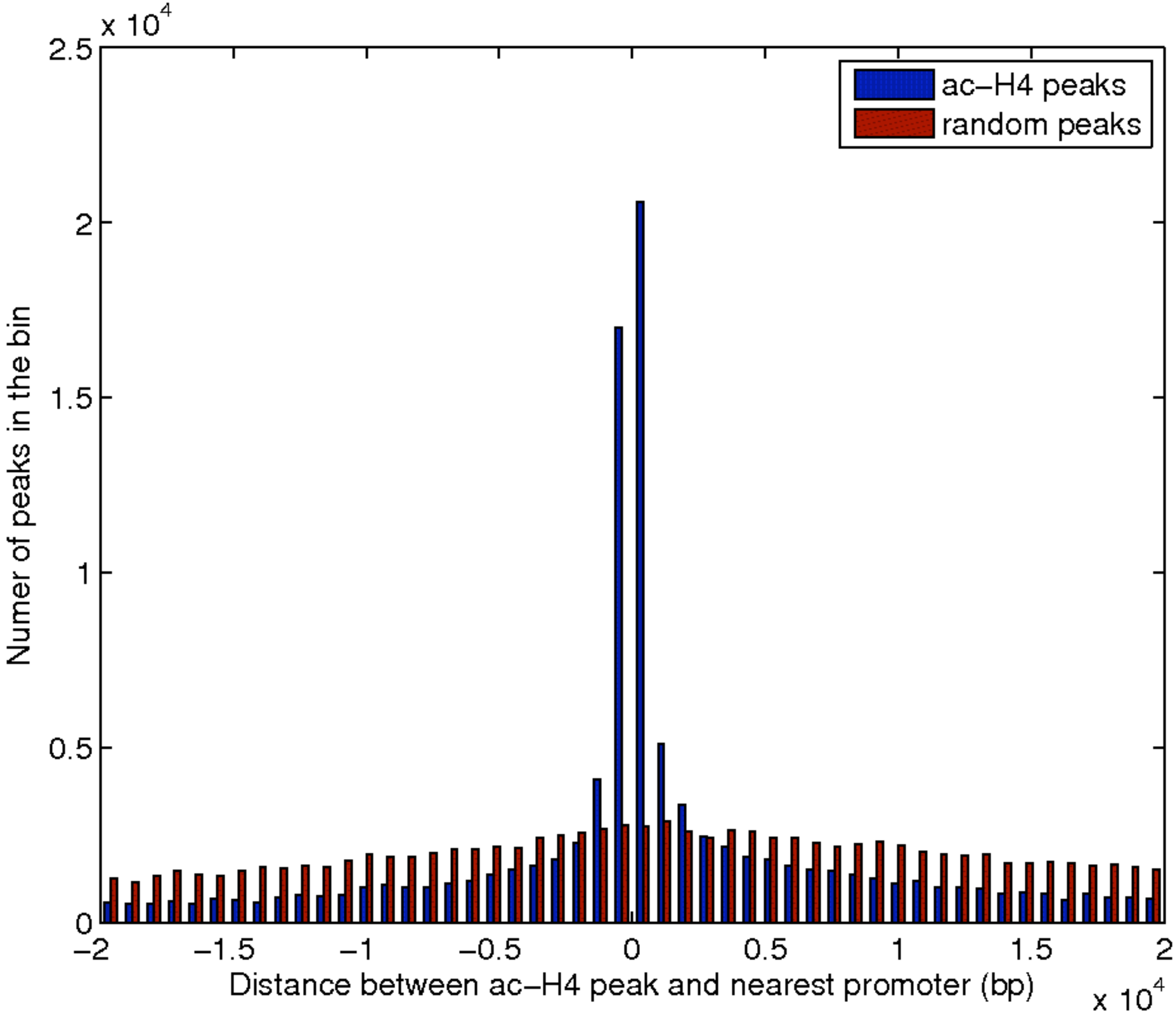
Lähdesmäki *et al* PLoS ONE, 3:3, e1820, 2008.

ChIP-seq binding event correlated with sequence-based and epigenetic evidence for cis-regulatory function

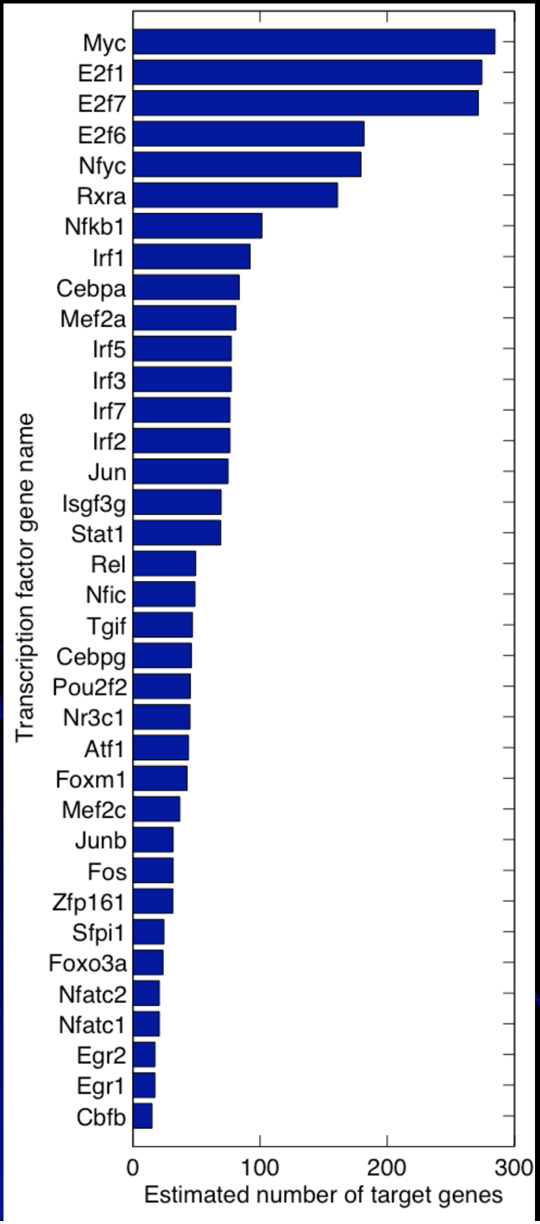


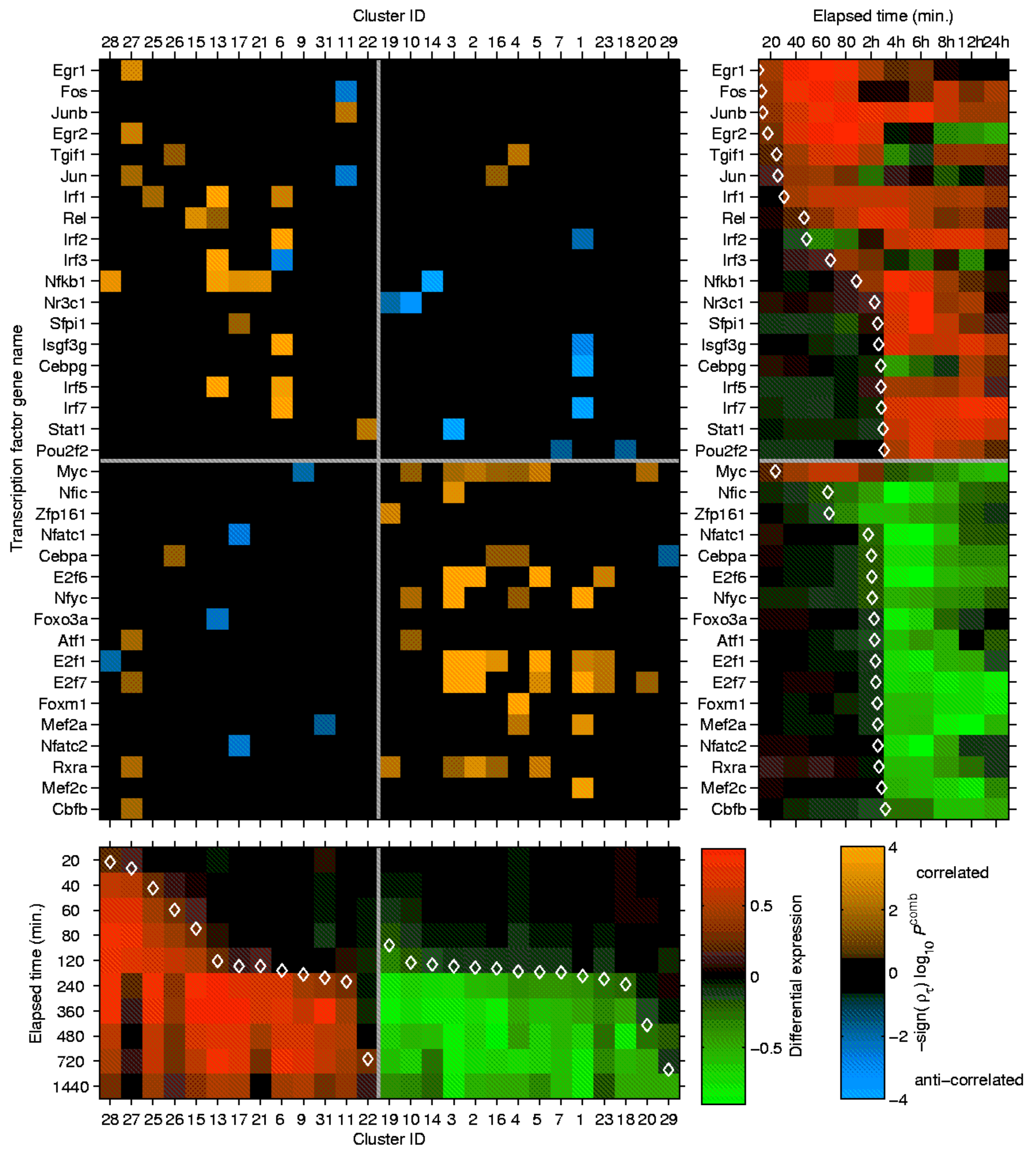
S. Ramsey

ac-H4 peaks are associated with transcription start sites



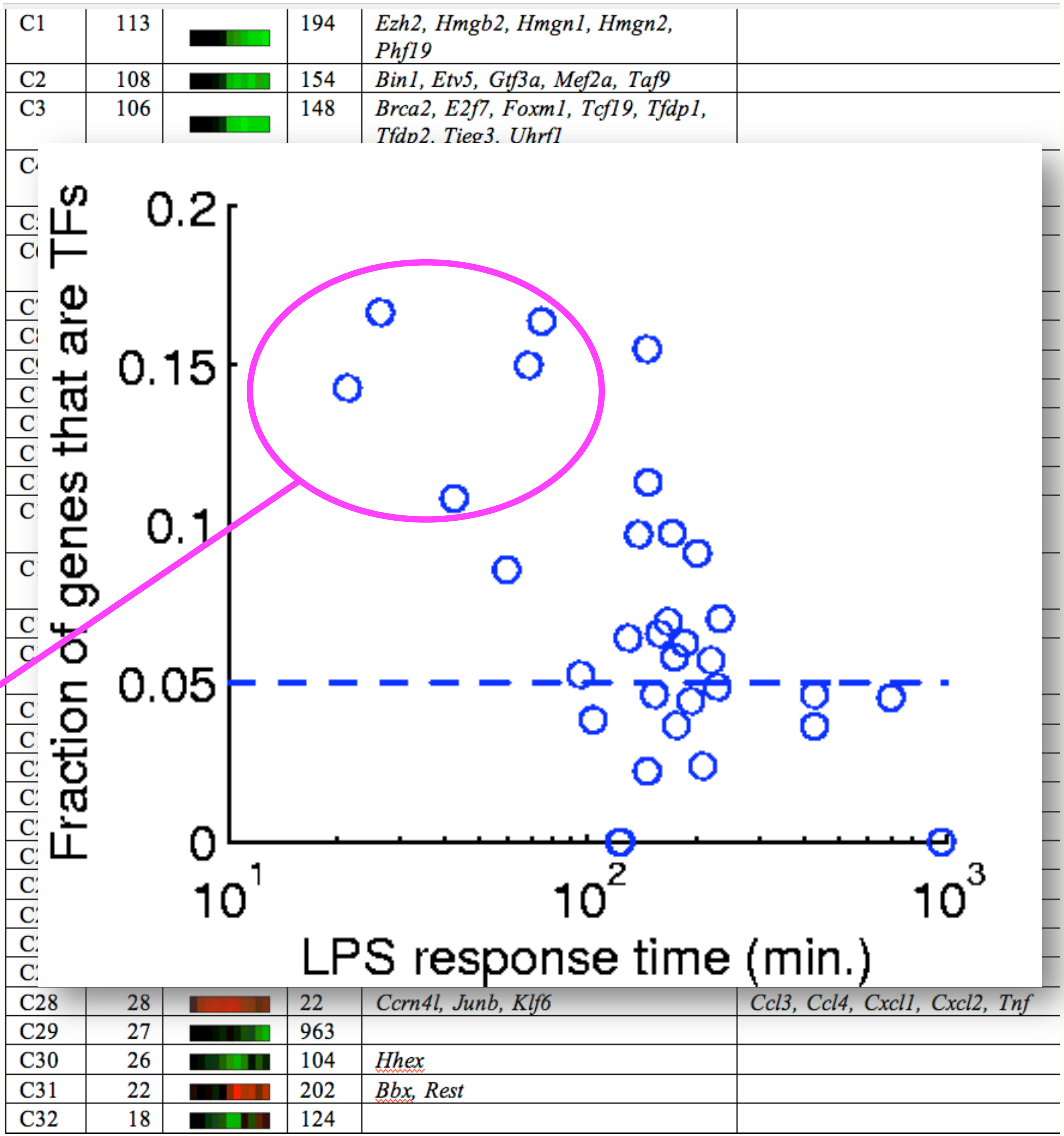
Map network





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

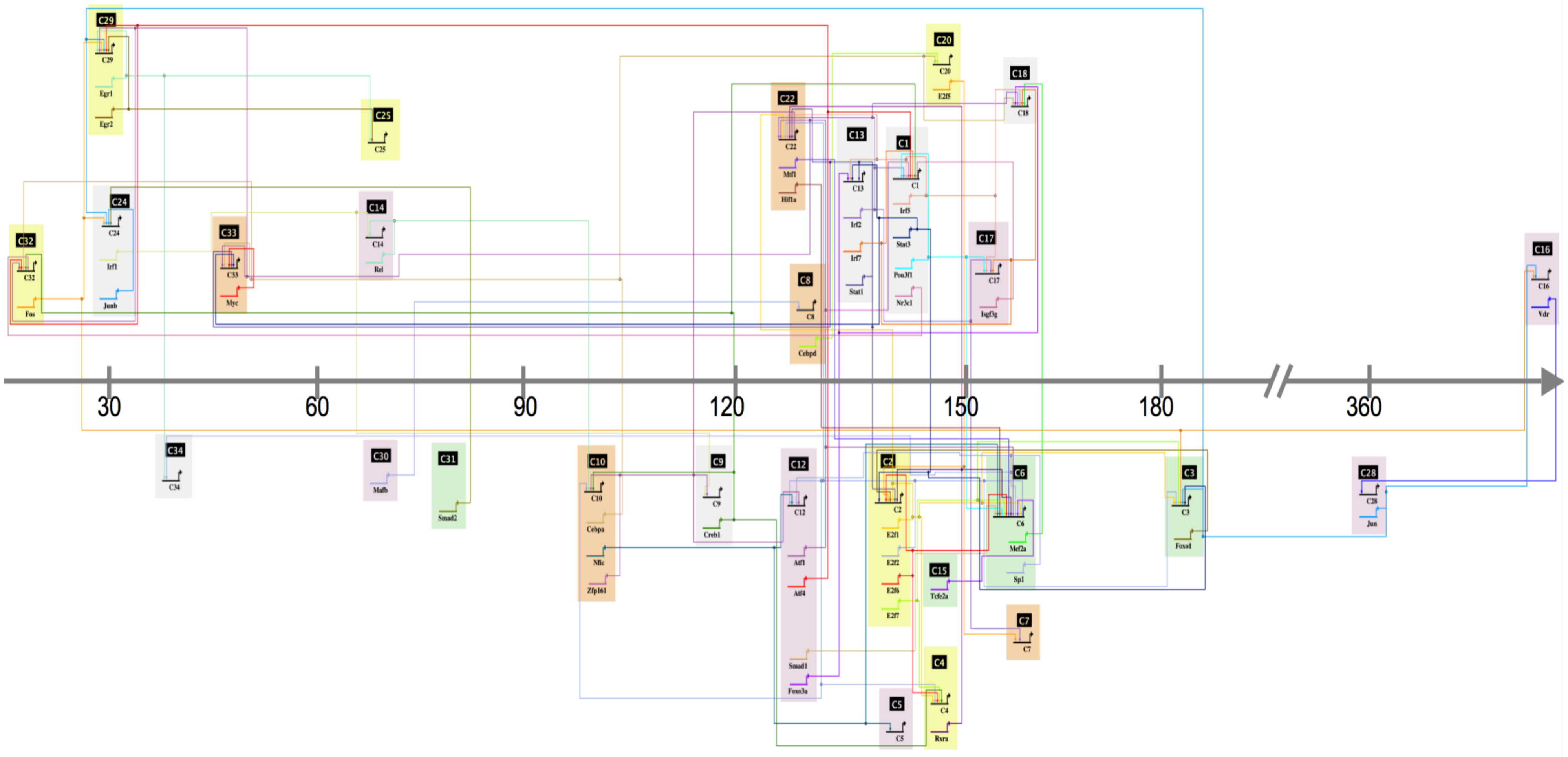
TF genes and cytokines within the clusters



C1	113	🟩	194	*Ezh2, Hmgb2, Hmgn1, Hmgn2, Phf19*	
C2	108	🟩	154	*Bin1, Etv5, Gtf3a, Mef2a, Taf9*	
C3	106	🟩	148	*Brca2, E2f7, Foxm1, Tcf19, Tfdp1, Tfdp2, Tieg3, Uhrf1*	
C28	28	🟥	22	*Ccrn4l, Junb, Klf6*	*Ccl3, Ccl4, Cxcl1, Cxcl2, Tnf*
C29	27	🟩	963		
C30	26	🟩	104	*Hhex*	
C31	22	🟥	202	*Bbx, Rest*	
C32	18	🟩	124		

Suggests early-induced TF genes may be regulating later genes

Temporal representation of the TF-cluster association network



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

ChIP-on-chip validation

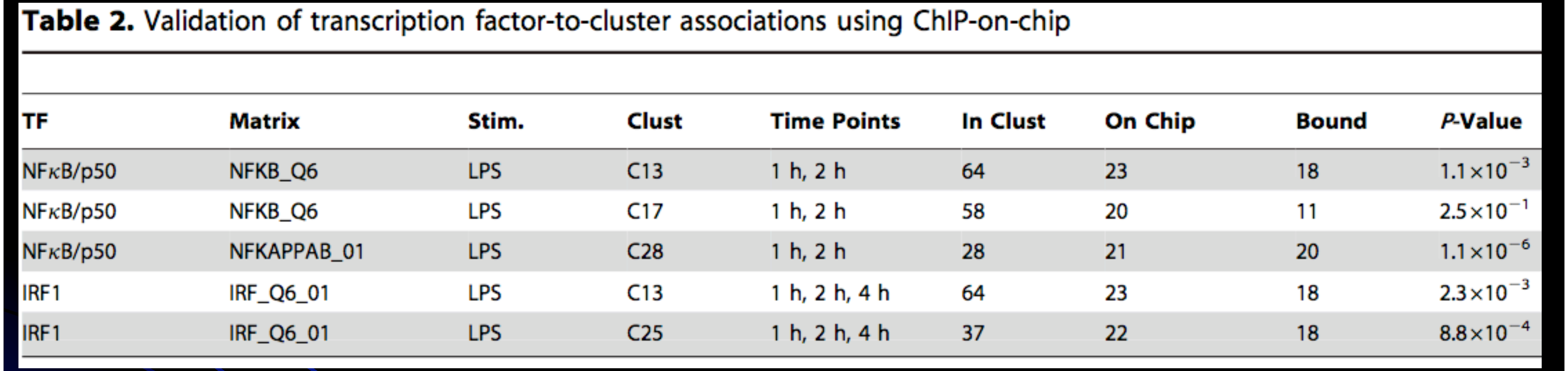
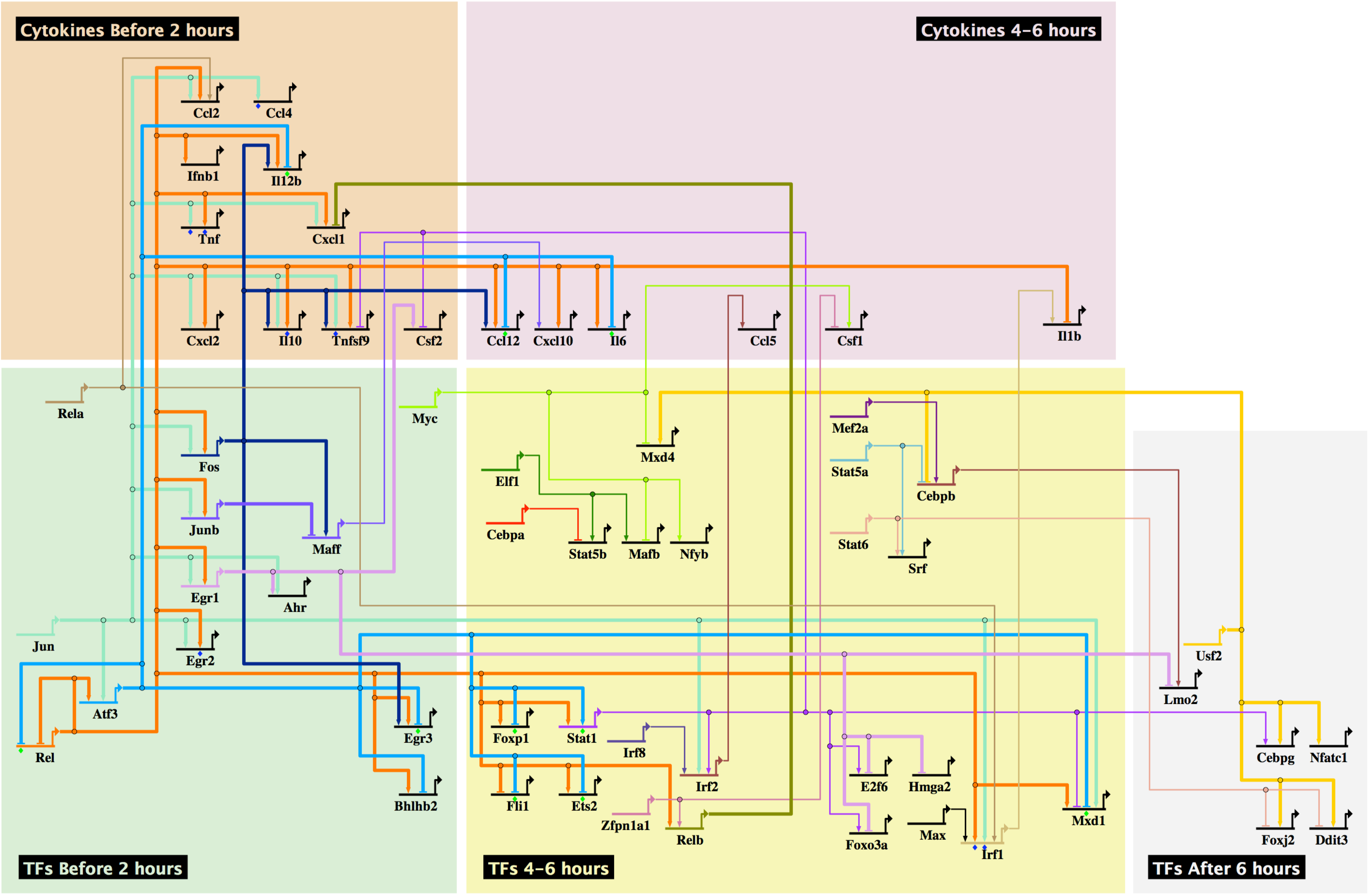


Table 2. Validation of transcription factor-to-cluster associations using ChIP-on-chip

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}

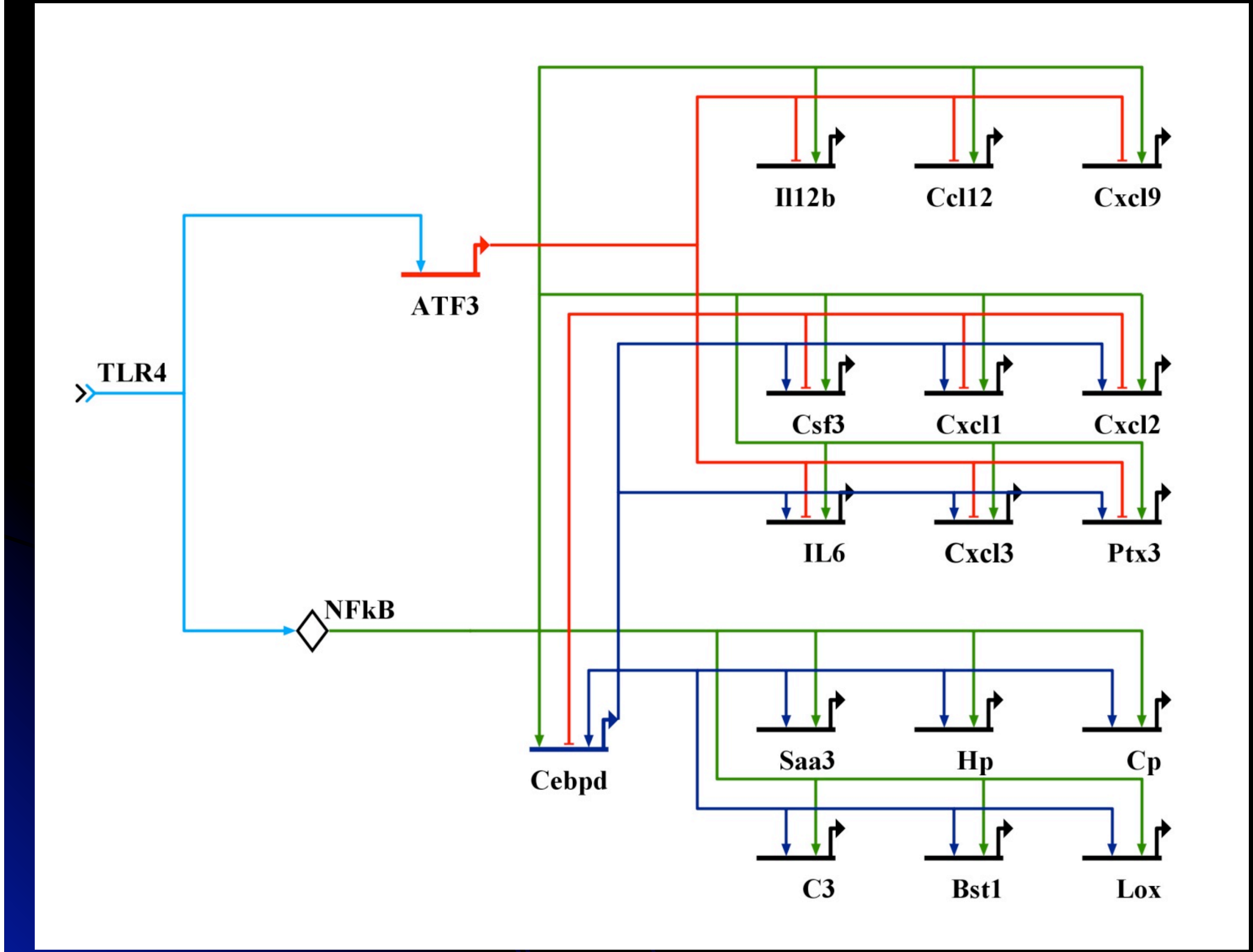
S. Ramsey, PLoS Comp Bio, 4,3, 2008.

Predicted Regulation of Cytokines and Transcription Factors



V. Thorsson

Characterization of an NFκB, CEBPδ, and ATF3 transcriptional sub-network



NFκB initiated

ATF3 attenuated

NFκB initiated

CEBPδ amplified

ATF3 attenuated

NFκB initiated

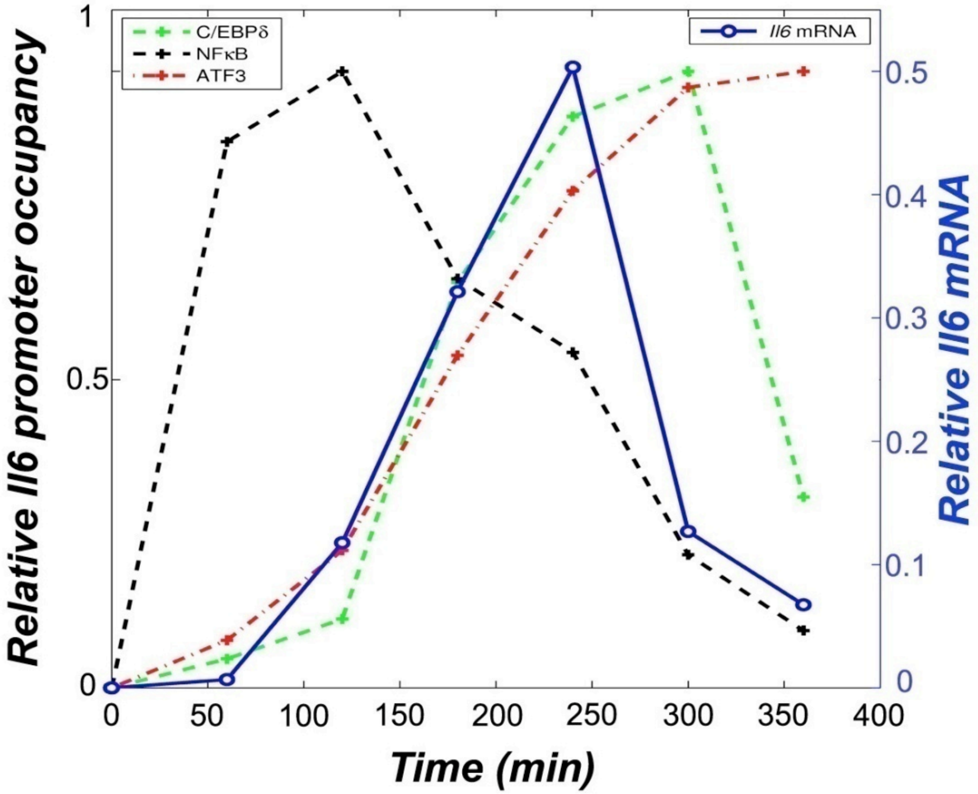
CEBPδ amplified

Nat Immun 10:437, 2009

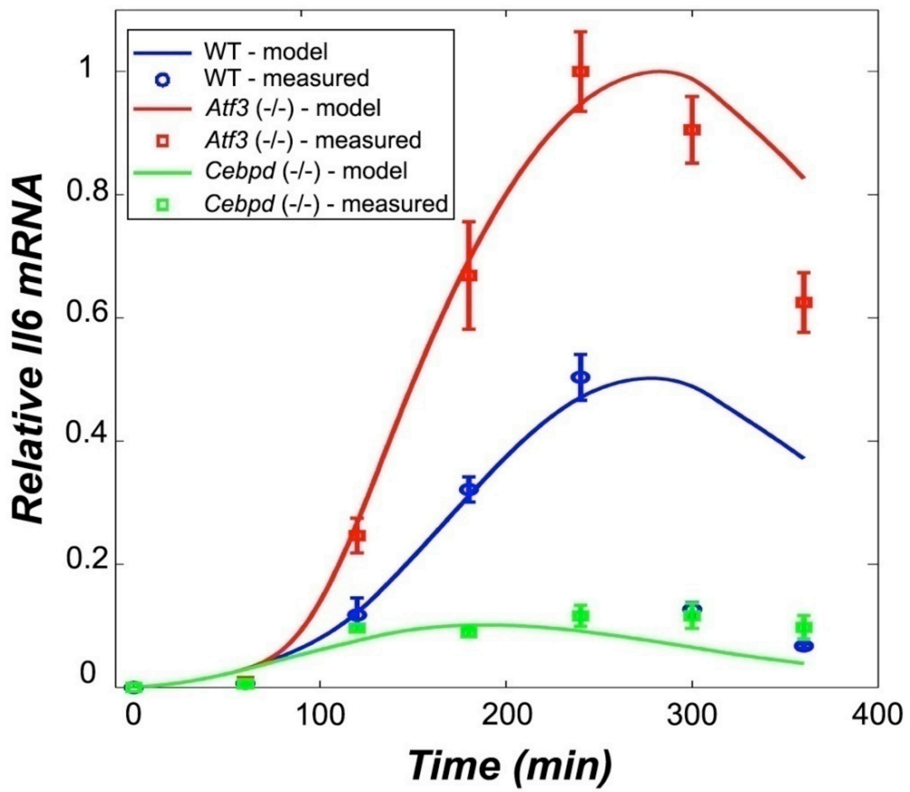
V. Litvak, S. Ramsey, M. Nykter

A model of the regulation of IL6 transcription by NFkB, ATF3 and CEBPδ

c



d

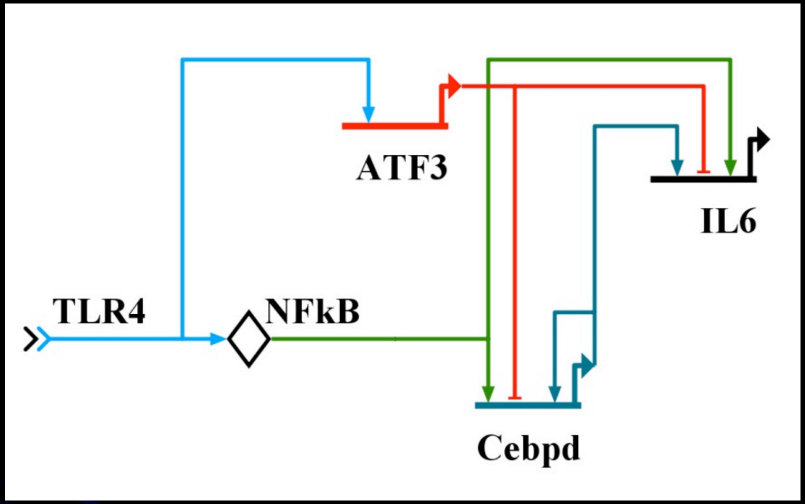


$$\tau \frac{d[Il6]}{dt} = -[Il6] + \frac{k_1\left[\mathrm{NFkB}\right]^{n_1} + k_2\left[\mathrm{NFkB}\right]\left[\mathrm{CEBP}\right]}{1 + k_1\left[\mathrm{NFkB}\right]^{n_1} + k_2\left[\mathrm{NFkB}\right]\left[\mathrm{CEBP}\right] + k_3\left[\mathrm{ATF3}\right]}$$

Nat Immun 10:437, 2009

V. Litvak, S. Ramsey, M. Nykter

NFκB and CEBPδ represent a Feed Forward Loop for the regulation of IL6 transcription

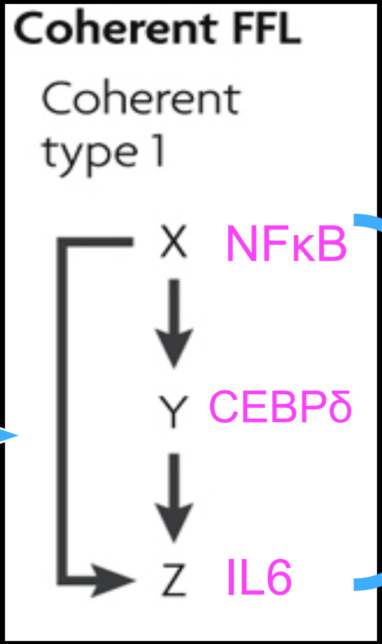


It is possible that CEBPδ acts as a check point to ensure that the system responds only to persistent inflammatory signals

A coherent FFL filters fluctuations in the environment; only persistent signals are able to activate the pathway

NFκB only

Low level production



NFκB and CEBPδ

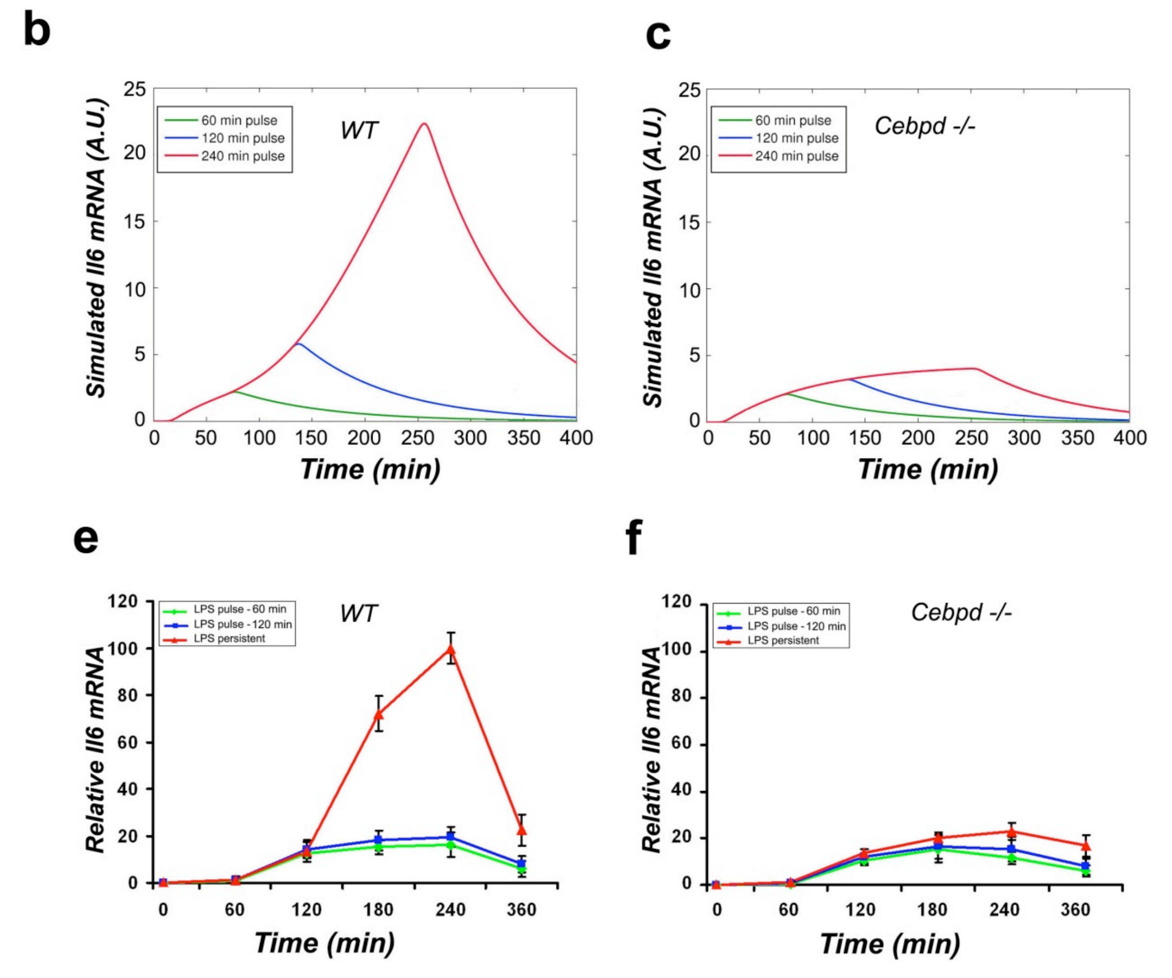
Amplified production

Nat Immun 10:437, 2009

V. Litvak, S. Ramsey, M. Nykter

Computational simulation of NFκB pulses

$$\tau \frac{d[Il6]}{dt} = -[Il6] + \frac{k_1[\text{NF}\kappa\text{B}]^{n_1} + k_2[\text{NF}\kappa\text{B}][\text{CEBP}\delta]}{1 + k_1[\text{NF}\kappa\text{B}]^{n_1} + k_2[\text{NF}\kappa\text{B}][\text{CEBP}\delta] + k_3[\text{ATF3}]}$$



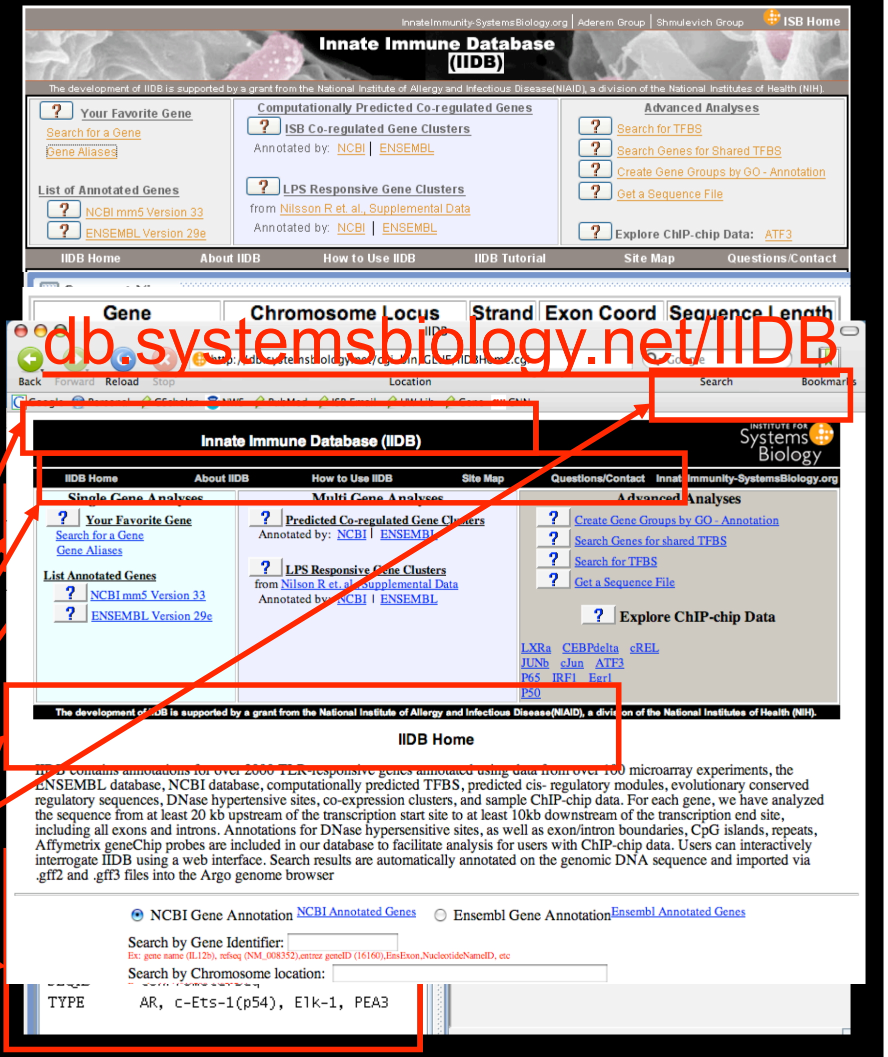
Nat Immun 10:437, 2009

V. Litvak, S. Ramsey, M. Nykter

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

[illegible]



db.systemsbiology.net/IIDB

The development of IIDB is supported by a grant from the National Institute of Allergy and Infectious Disease(NIAID), a division of the National Institutes of Health (NIH).

IIDB Home

IIDB contains annotations for over 2000 TLR-responsive genes annotated using data from over 100 microarray experiments, the ENSEMBL database, NCBI database, computationally predicted TFBS, predicted cis- regulatory modules, evolutionary conserved regulatory sequences, DNase hypertensive sites, co-expression clusters, and sample ChIP-chip data. For each gene, we have analyzed the sequence from at least 20 kb upstream of the transcription start site to at least 10kb downstream of the transcription end site, including all exons and introns. Annotations for DNase hypersensitive sites, as well as exon/intron boundaries, CpG islands, repeats, Affymetrix geneChip probes are included in our database to facilitate analysis for users with ChIP-chip data. Users can interactively interrogate IIDB using a web interface. Search results are automatically annotated on the genomic DNA sequence and imported via .gff2 and .gff3 files into the Argo genome browser

● NCBI Gene Annotation NCBI Annotated Genes ○ Ensembl Gene Annotation Ensembl Annotated Genes

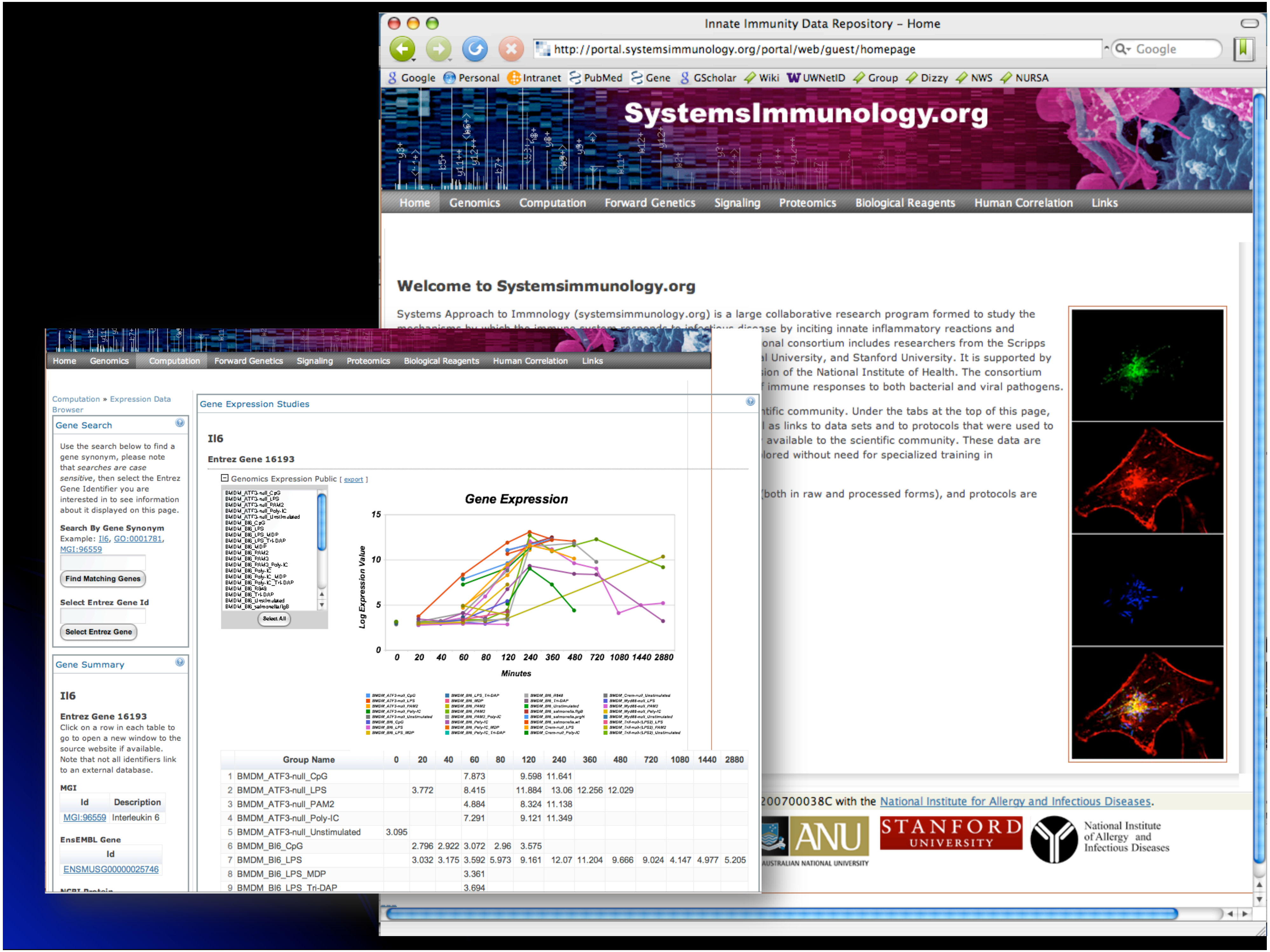
Search by Gene Identifier:

Ex: gene name (IL12b), refseq (NM_008352),entrez geneID (16160),EnsExon,NucleotideNameID, etc

Search by Chromosome location:

```
TYPE      AR, c-Ets-1(p54), Elk-1, PEA3
```

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



Welcome to Systemsimmunology.org

Systems Approach to Immnology (systemsimmunology.org) is a large collaborative research program formed to study the
ase by inciting innate inflammatory reactions and
onal consortium includes researchers from the Scripps
l University, and Stanford University. It is supported by
ion of the National Institute of Health. The consortium
f immune responses to both bacterial and viral pathogens.

ntific community. Under the tabs at the top of this page,
l as links to data sets and to protocols that were used to
available to the scientific community. These data are
lored without need for specialized training in

(both in raw and processed forms), and protocols are

200700038C with the National Institute for Allergy and Infectious Diseases.

Computation » Expression Data Browser

Gene Search

Use the search below to find a gene synonym, please note that *searches are case sensitive*, then select the Entrez Gene Identifier you are interested in to see information about it displayed on this page.

Search By Gene Synonym
Example: Il6, GO:0001781, MGI:96559

Select Entrez Gene Id

Gene Summary

Il6

Entrez Gene 16193
Click on a row in each table to go to open a new window to the source website if available. Note that not all identifiers link to an external database.

MGI

Id	Description
MGI:96559	Interleukin 6

EnsEMBL Gene

Id
ENSMUSG00000025746

NCBI Protein

Gene Expression Studies

Il6

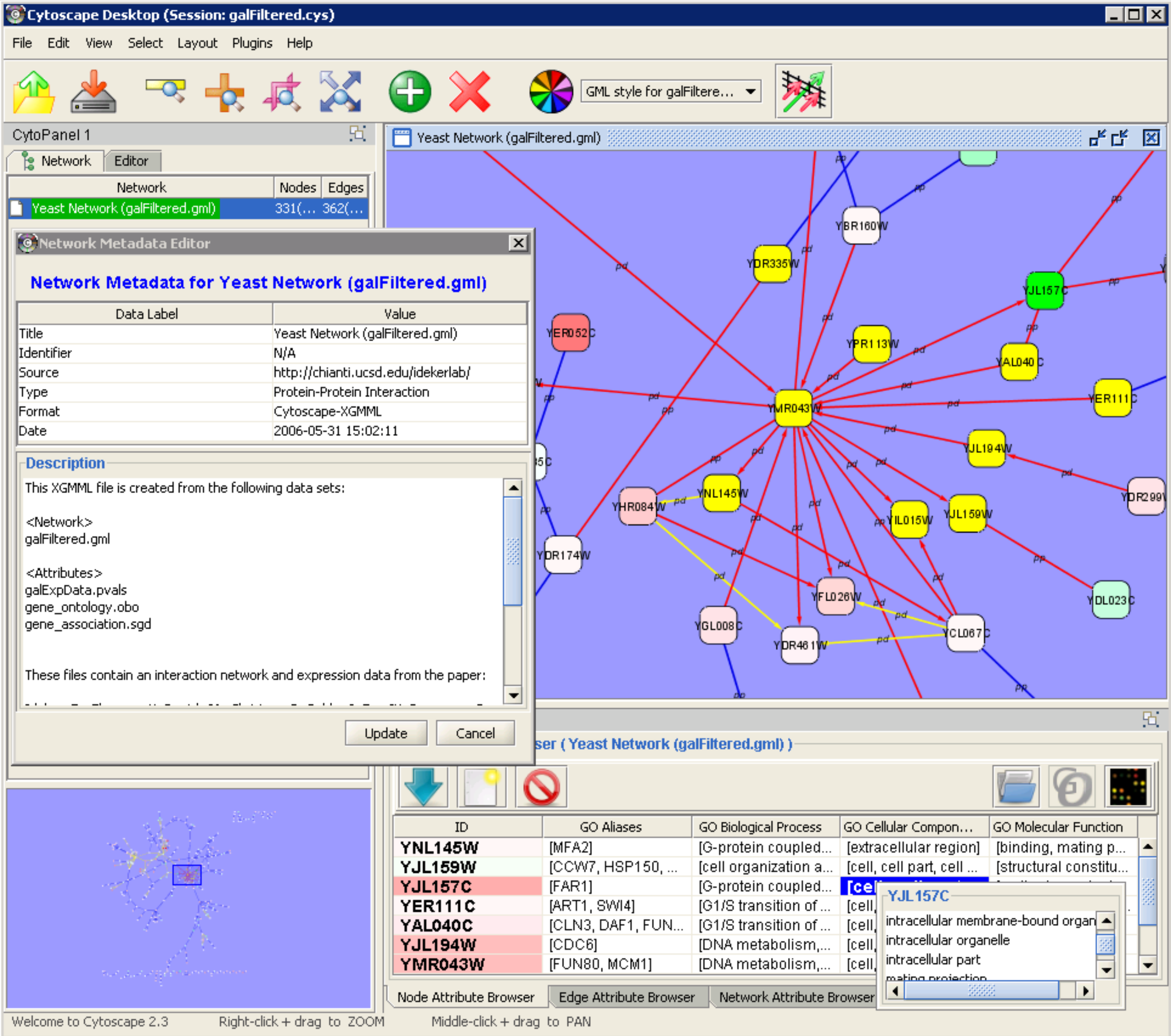
Entrez Gene 16193

☐ Genomics Expression Public [export]

BMDM_ATF3-null_CpG
BMDM_ATF3-null_LPS
BMDM_ATF3-null_PAM2
BMDM_ATF3-null_Poly-IC
BMDM_ATF3-null_Unstimulated
BMDM_Bl6_CpG
BMDM_Bl6_LPS
BMDM_Bl6_LPS_MDP
BMDM_Bl6_LPS_Tri-DAP
BMDM_Bl6_MDP
BMDM_Bl6_PAM2
BMDM_Bl6_PAM3
BMDM_Bl6_PAM3_Poly-IC
BMDM_Bl6_Poly-IC
BMDM_Bl6_Poly-IC_MDP
BMDM_Bl6_Poly-IC_Tri-DAP
BMDM_Bl6_R848
BMDM_Bl6_Tri-DAP
BMDM_Bl6_Unstimulated
BMDM_Bl6_salmonella.flgB

	Group Name	0	20	40	60	80	120	240	360	480	720	1080	1440	2880
1	BMDM_ATF3-null_CpG				7.873		9.598	11.641						
2	BMDM_ATF3-null_LPS		3.772		8.415		11.884	13.06	12.256	12.029				
3	BMDM_ATF3-null_PAM2				4.884		8.324	11.138						
4	BMDM_ATF3-null_Poly-IC				7.291		9.121	11.349						
5	BMDM_ATF3-null_Unstimulated	3.095												
6	BMDM_Bl6_CpG		2.796	2.922	3.072	2.96	3.575							
7	BMDM_Bl6_LPS		3.032	3.175	3.592	5.973	9.161	12.07	11.204	9.666	9.024	4.147	4.977	5.205
8	BMDM_Bl6_LPS_MDP				3.361									
9	BMDM_Bl6_LPS_Tri-DAP				3.694									

Cytoscape



Network	Nodes	Edges
Yeast Network (galFiltered.gml)	331(...	362(...

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

Description

This XGMML file is created from the following data sets:

<Network>
galFiltered.gml

<Attributes>
galExpData.pvals
gene_ontology.obo
gene_association.sgd

These files contain an interaction network and expression data from the paper:

ser (Yeast Network (galFiltered.gml))

ID	GO Aliases	GO Biological Process	GO Cellular Compon...	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...	[ce	
YER111C	[ART1, SWI4]	[G1/S transition of ...	[cell,	
YAL040C	[CLN3, DAF1, FUN...	[G1/S transition of ...	[cell,	
YJL194W	[CDC6]	[DNA metabolism,...	[cell,	
YMR043W	[FUN80, MCM1]	[DNA metabolism,...	[cell,	

YJL157C
intracellular membrane-bound organ
intracellular organelle
intracellular part

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

Analysis plug-in modules

Implemented in Java

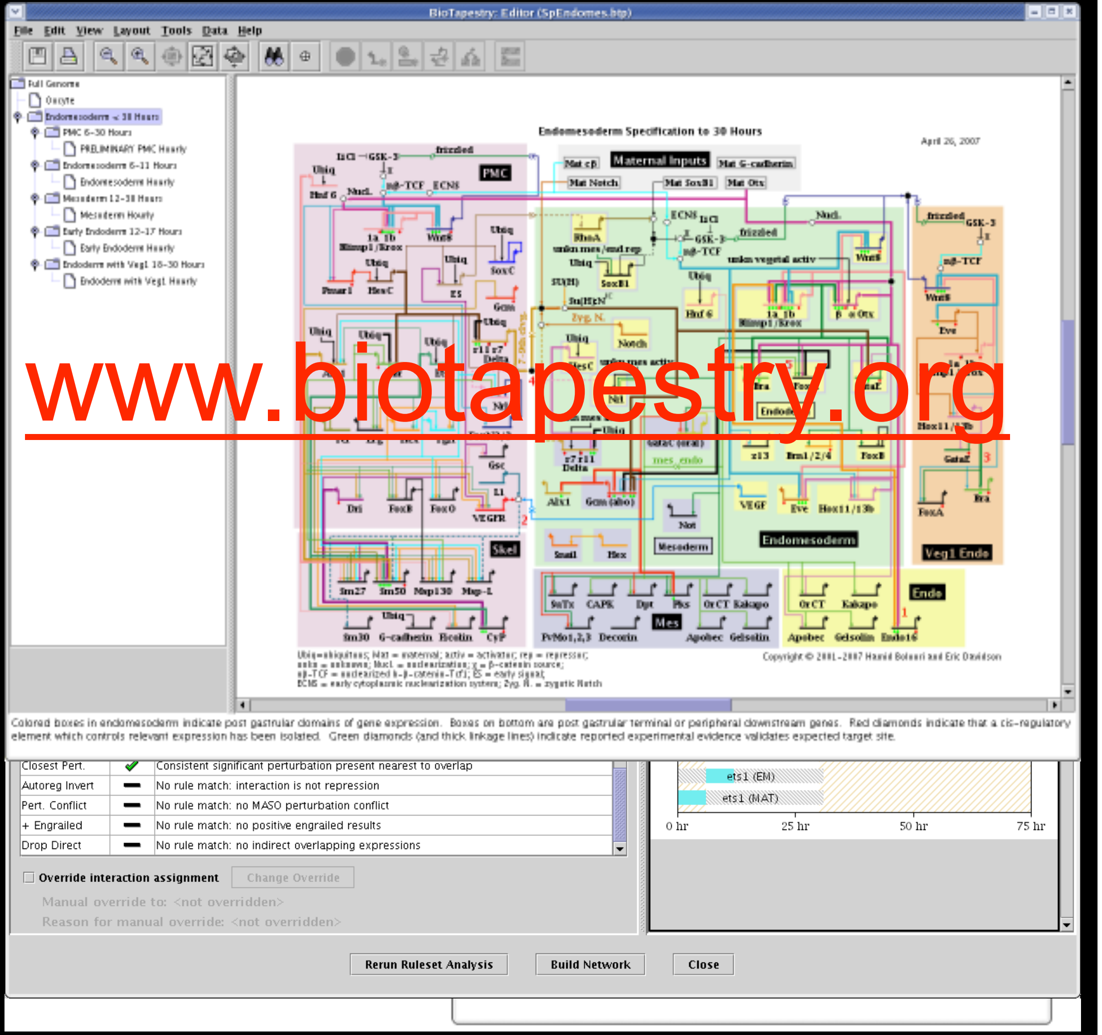
(networks, attributes, network metadata, etc.)

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)



Colored boxes in endomesoderm indicate post gastrular domains of gene expression. Boxes on bottom are post gastrular terminal or peripheral downstream genes. Red diamonds indicate that a cis-regulatory element which controls relevant expression has been isolated. Green diamonds (and thick linkage lines) indicate reported experimental evidence validates expected target site.

Closest Pert.	✓	Consistent significant perturbation present nearest to overlap
Autoreg Invert	—	No rule match: interaction is not repression
Pert. Conflict	—	No rule match: no MASO perturbation conflict
+ Engrailed	—	No rule match: no positive engrailed results
Drop Direct	—	No rule match: no indirect overlapping expressions

☐ Override interaction assignment [Change Override]

Manual override to: <not overridden>

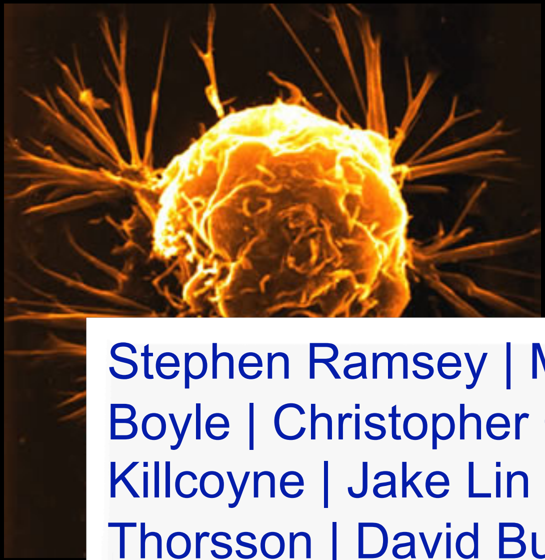
Reason for manual override: <not overridden>

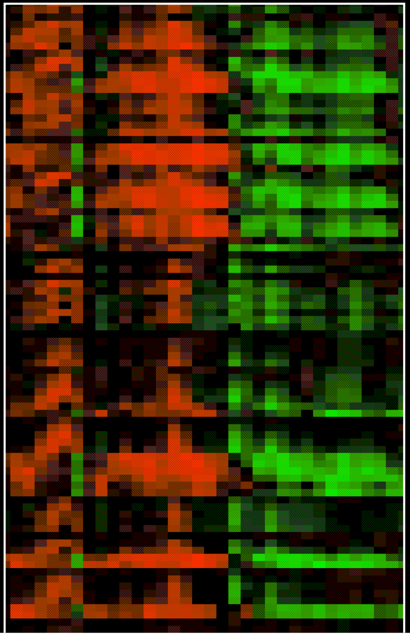
[Rerun Ruleset Analysis] [Build Network] [Close]

www.biotapestry.org

W. Longabaugh

thank you





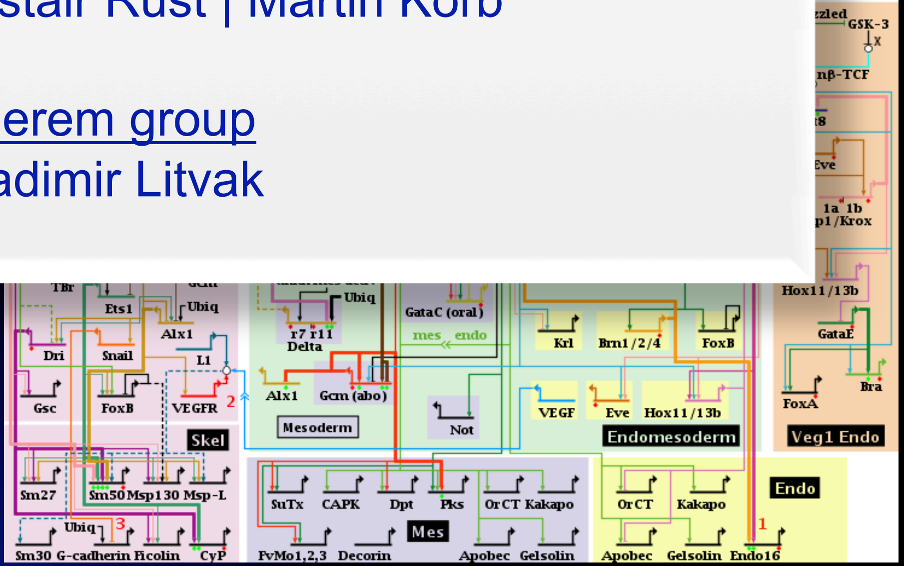
Stephen Ramsey | Matti Nykter | John Boyle | Christopher Cavnor | Sarah Killcoyne | Jake Lin | Vesteinn Thorsson | David Burdick | Hector Rovira | Madhumati Gundapuneni | Roger Kramer | Bill (William) Longabaugh | Harri Lähdesmäki | Alistair Rust | Martin Korb

<u>Aderem group</u>
Vladimir Litvak

NIH/NIAID U54 AI54253 &
HHSN272200700038C

NIH/NIGMS P50 MO-76547

NIH/NIGMS GM072855



$$P(S|A,\theta,\phi) = \underbrace{\prod_{i_1=1}^{a_1-1}\phi(s_{i_1})}_{\text{background } 1}\ \underbrace{\prod_{j_1=a_1}^{a_1+\ell-1}\theta(s_{j_1}, j_1-a_1+1)}_{\text{motif instance } 1}$$

$$\times\cdots\times\ \underbrace{\prod_{j_c=a_c}^{a_c+\ell-1}\theta(s_{j_c}, j_c-a_c+1)}_{\text{motif instance } c}\ \underbrace{\prod_{i_{c+1}=a_c+\ell}^{L}\phi(s_{i_{c+1}})}_{\text{background } c+1}.$$