

Big Data, Visualization, and Systems Biology

Tamara Munzner

University of British Columbia

Department of Computer Science

Big data and models

- Does big data imply all-data-no-model future for science?
 - No!!
 - Typical Wired hype...
- Big data leads to better models
- Example from system biology and visualization
 - Conduct experiments on cells
 - Interpret results in current model
 - Propose modifications to the model

Biomolecular interactions are selective

- Cell densely packed with biomolecules
- Interactions rare
- Model interactions as a graph

E. Coli

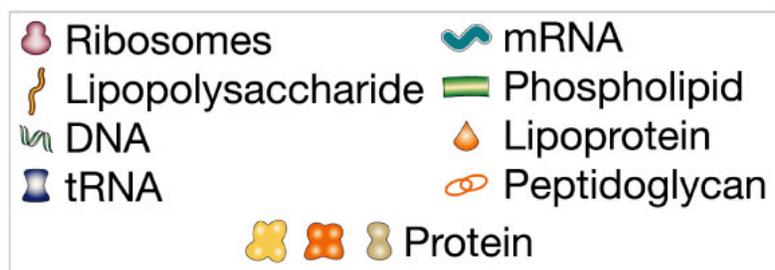
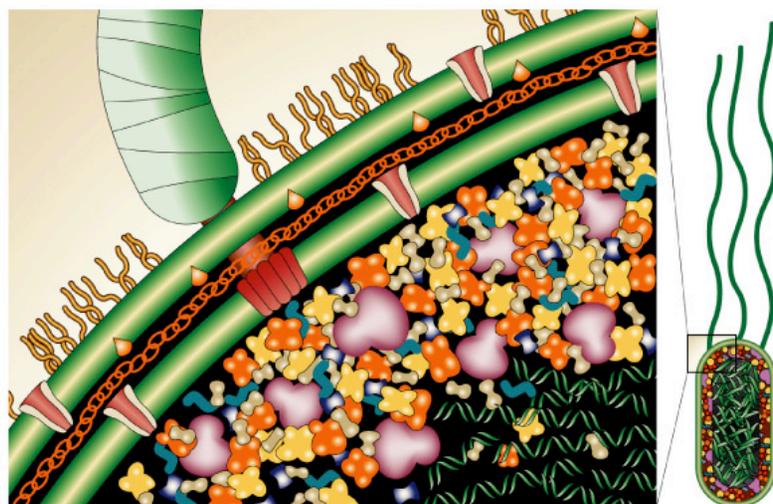
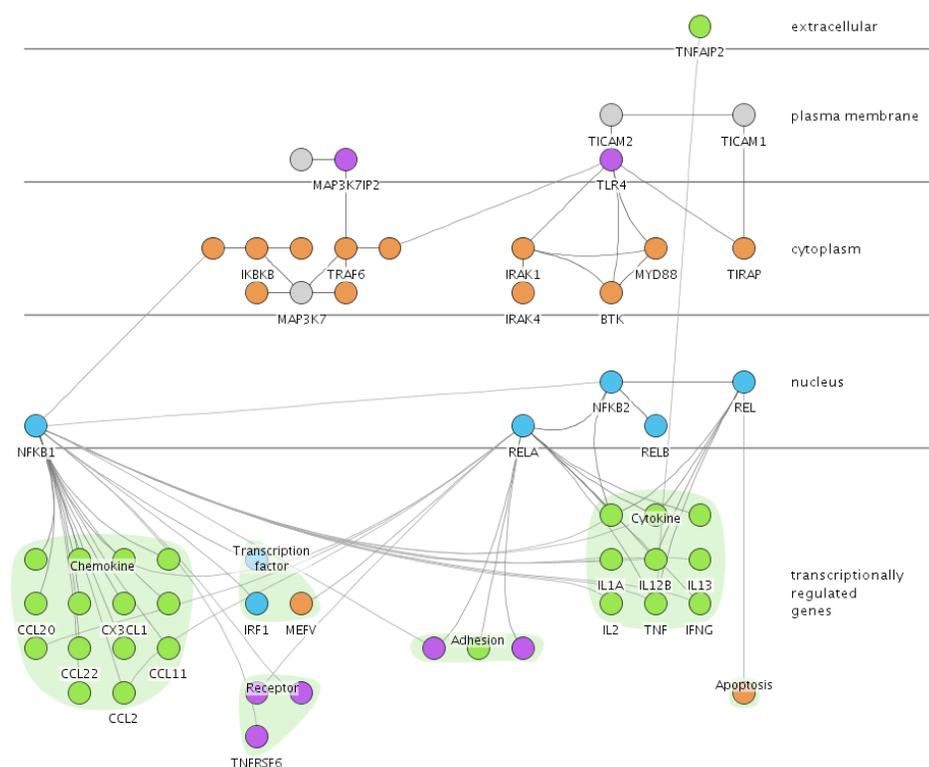


Image from Nature Publishing group

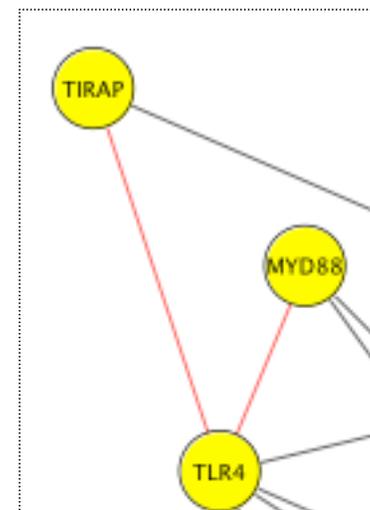
Systems biology model

- Graph $G = \{V, E\}$
 - V : proteins, genes, DNA, RNA, tRNA, etc.
 - E : interacting molecules

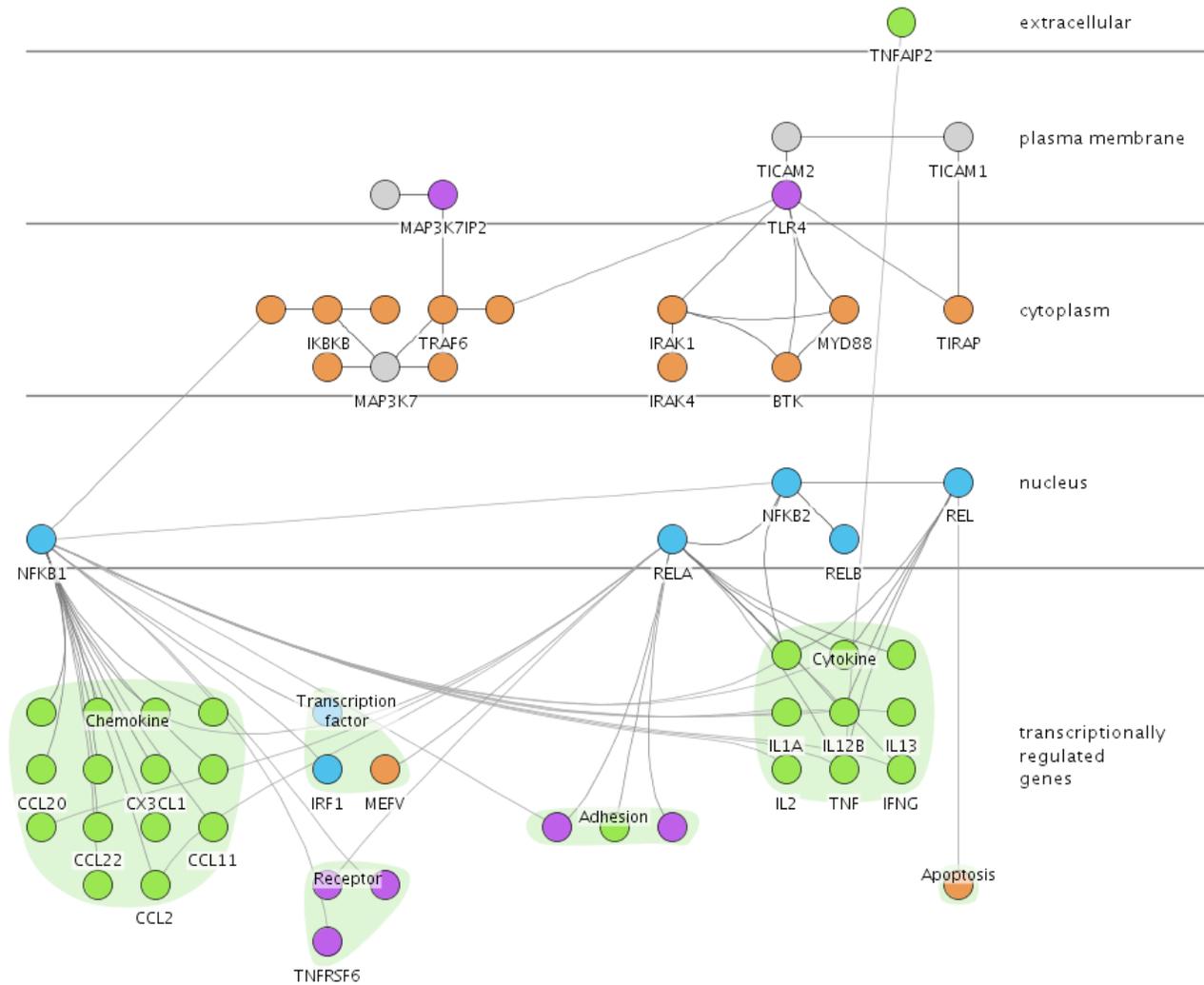


Model summarizes extensive lab work

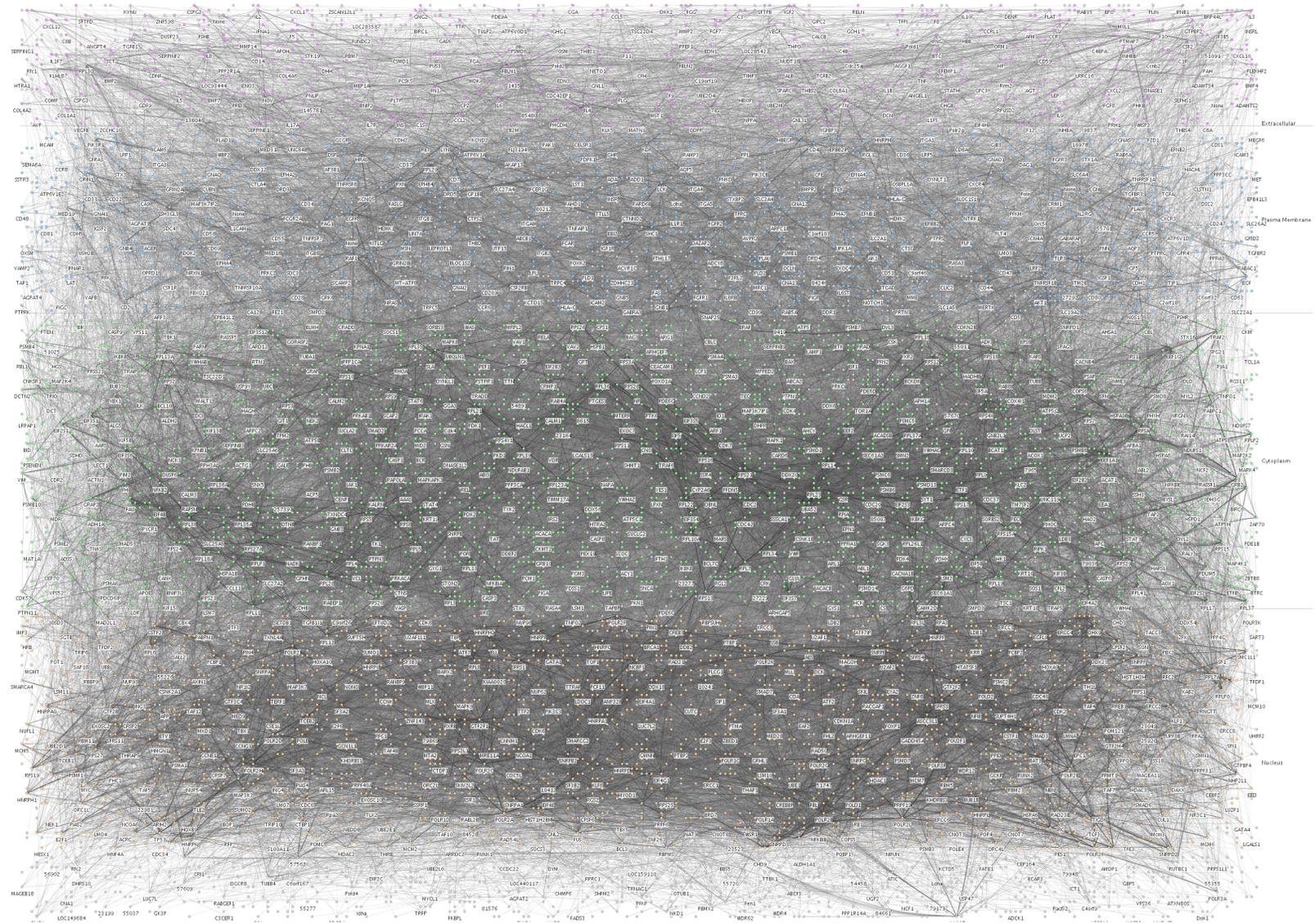
- Graphs come from hand-curated databases
 - Dynamic, change with each new publication
- Each edge has provenance from experimental evidence
 - TIRAP: an adapter molecule in the Toll signaling pathway. *Horng T, Barton GM, Medzhitov R.*
 - Mal (MyD88-adapter-like) is required for Toll-like receptor-4 signal transduction. *Fitzgerald KA, Palsson-McDermott EM, Bowie AG, Jefferies CA, Mansell AS, Brady G, Brint E, Dunne A, Gray P, Harte MT, McMurray D, Smith DE, Sims JE, Bird TA, O'Neill LA.*
- Choose scope to manage complexity



TLR4 biomolecule: E=74, V=54

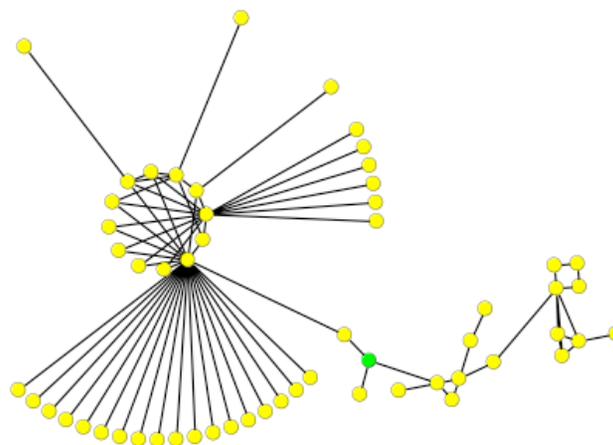


Human interactome: $E \sim 50,000$, $V \sim 10,000$



Goal: Overlay measurements on model

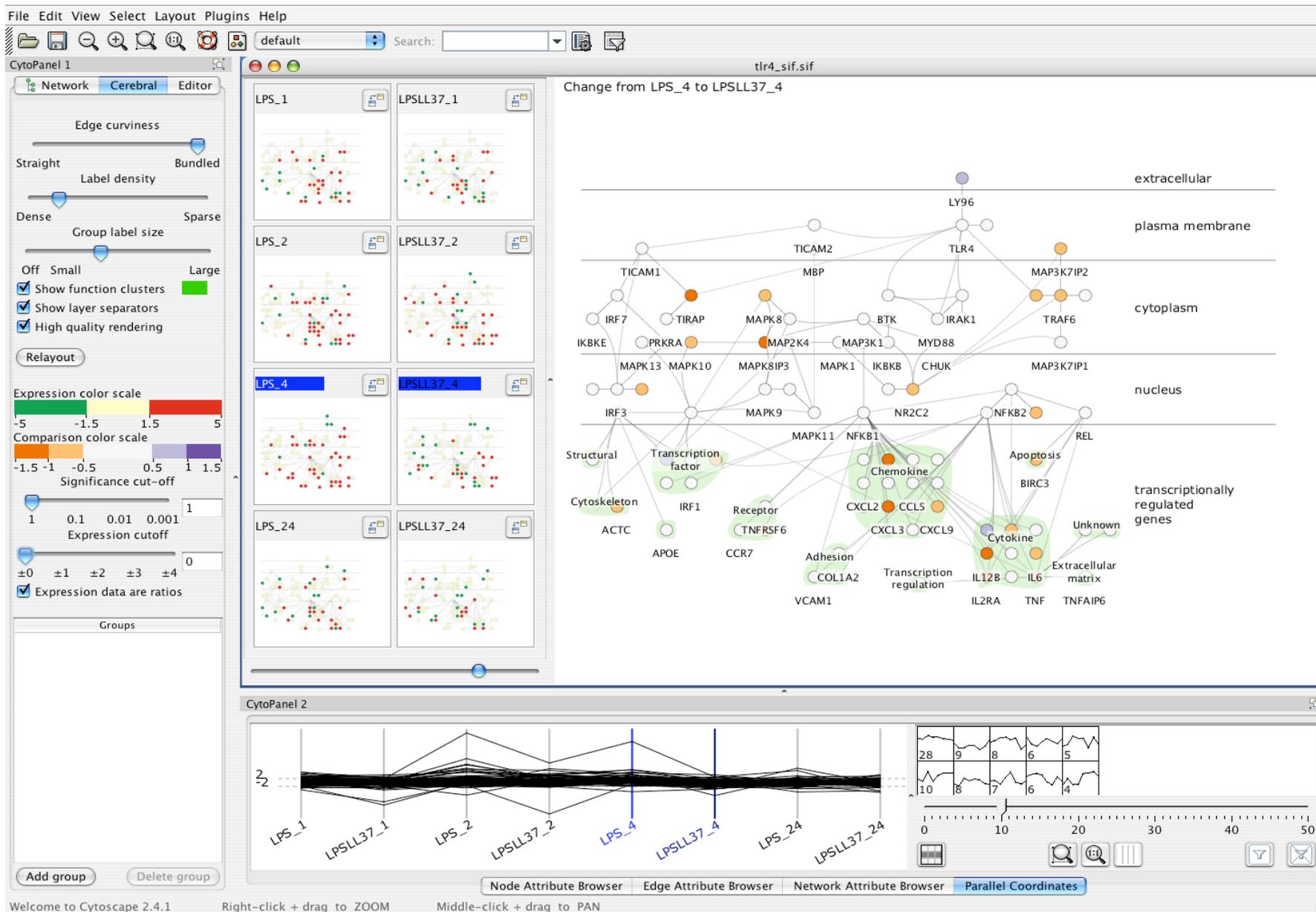
- Integrate
 - System model (graph)
 - Experimental measurements



Data Panel

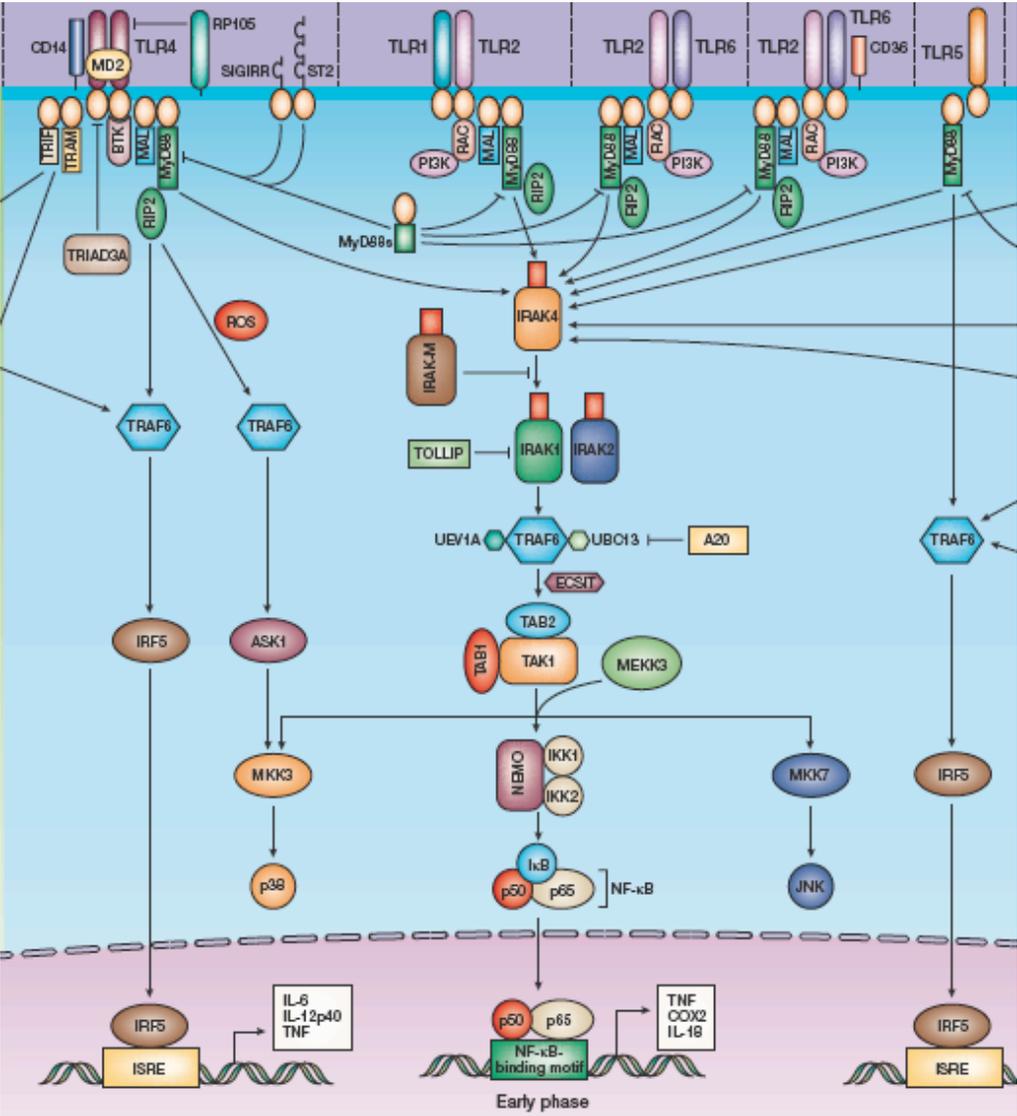
ID	Function	LPSLL37_1	LPSLL37_1_pvals	LPSLL37_2	LPSLL37_24	LPSLL37_24_pvals
IRAK2	Kinase	2.367	0.251	1.337	-1.553	
NFKB2	Transcription factor	-1.14	0.972	-1.03	1.303	0.807
CXCL2	Chemokine	1.853	0.376	4.111	-1.019	0.745
CHUK	Kinase	-1.376	0.373	2.232	1.194	0.387
IL13	Cytokine	-5.961		2.139	-1.236	0.601
RELA	Transcription factor	-1.077	0.564	-1.169	1.943	0.594
IKBKB	Kinase	1.167	0.29	1.421	-1.907	0.286
CCL4	Chemokine	1.254	0.878	-1.052	1.499	0.761
MAP3K7		1.01	0.956	-1.096	1.222	0.8
ICAM1	Adhesion	1.184	0.669	1.537	1.392	0.671
IRF1	Transcription factor	-1.013	0.519	1.416	1.081	0.995
CXCL3	Chemokine	1.7	0.905	1.092	-1.598	0.521
IL12B	Cytokine	-2.448	0.042	-1.473	-2.109	0.08
CCL11	Chemokine	-1.338	0.349	-1.995	-1.785	0.129
MAP3K7IP1	Adaptor					
JFNG	Cytokine	-1.15	0.801	1.075	1.053	0.521

Cerebral



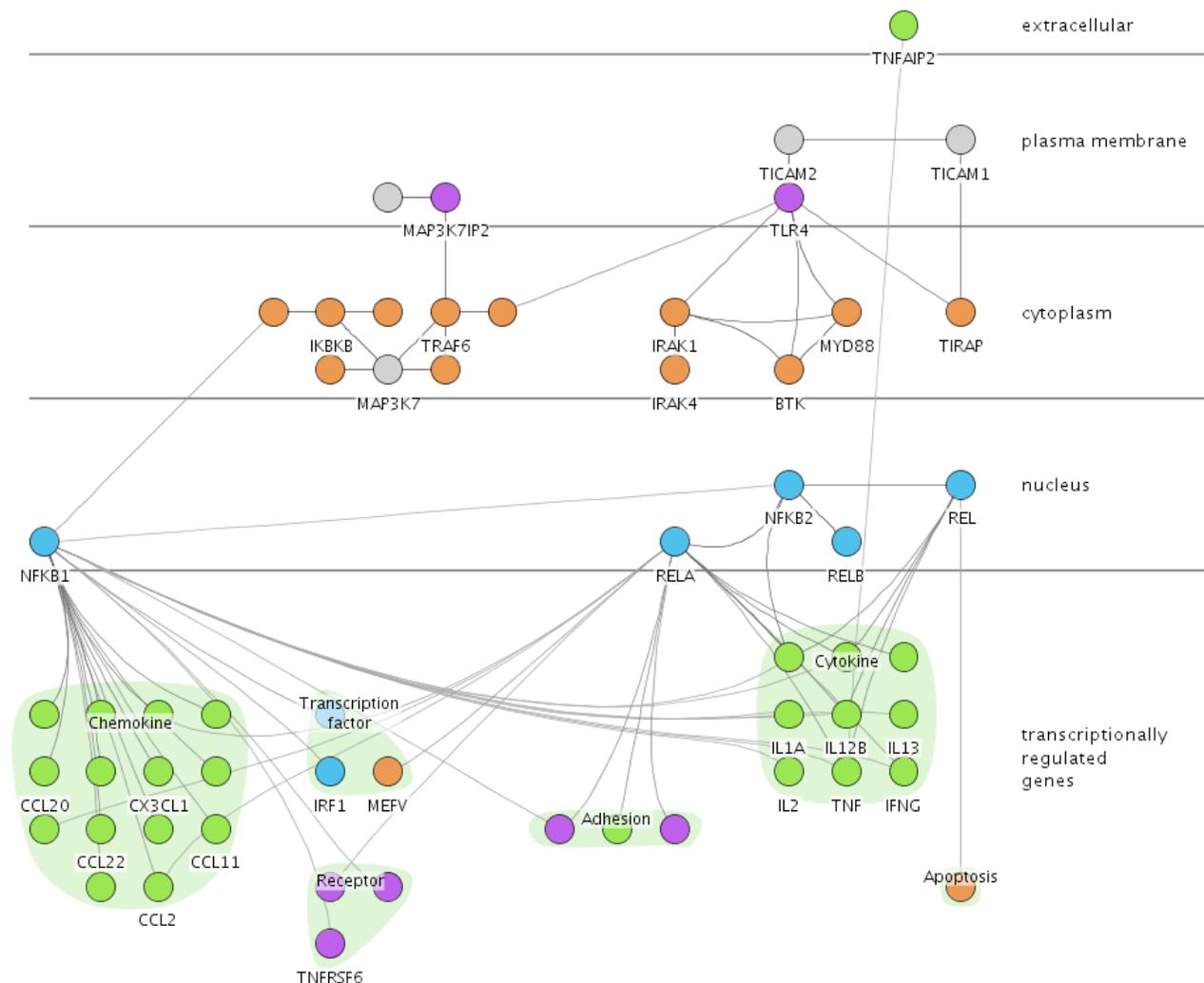
Video

Hand-drawn diagrams



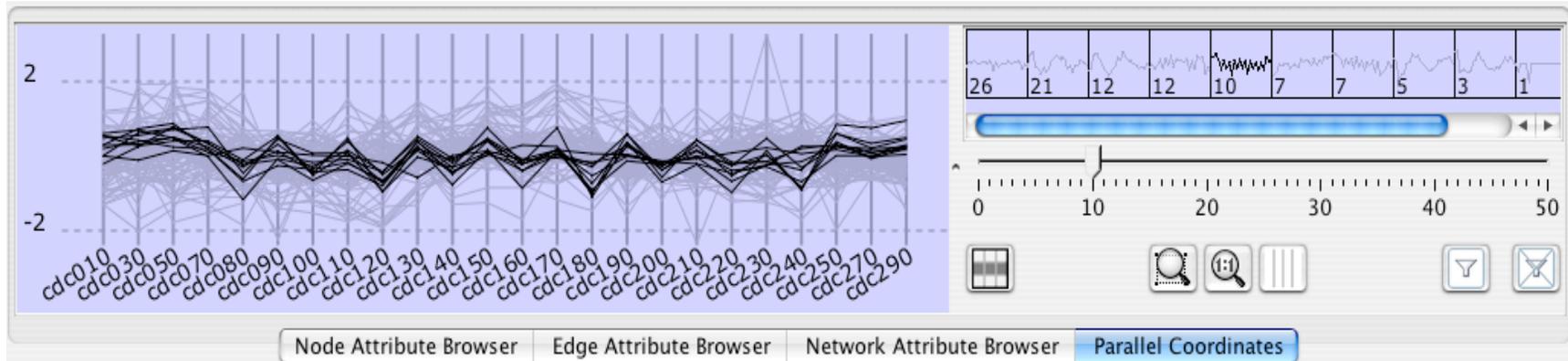
- Cellular location encoded spatially
- Infeasible to create by hand in era of big data

Cerebral layout using biological metadata



- Similar to hand-drawn
- Spatial position reveals location in cell
- Simulated annealing in $O(E\sqrt{V})$ vs. $O(V^3)$ time

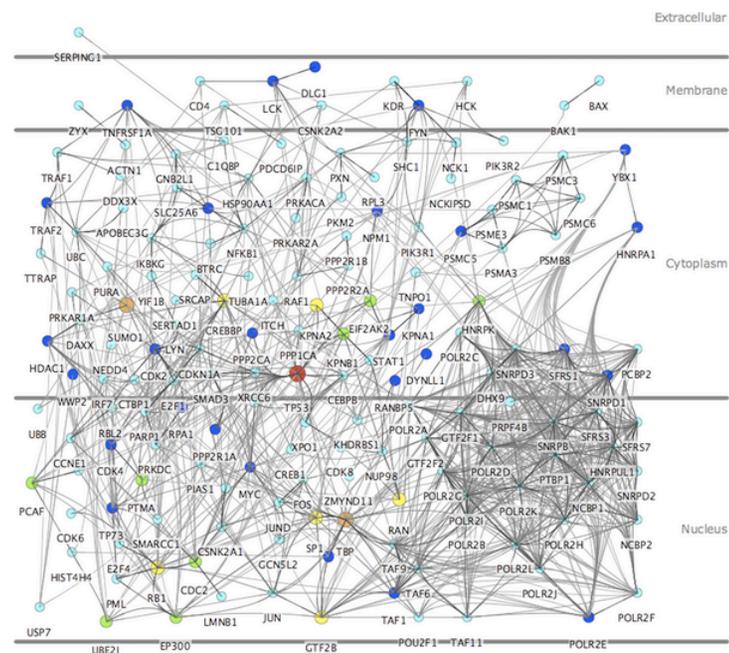
Measurement data alone insufficient



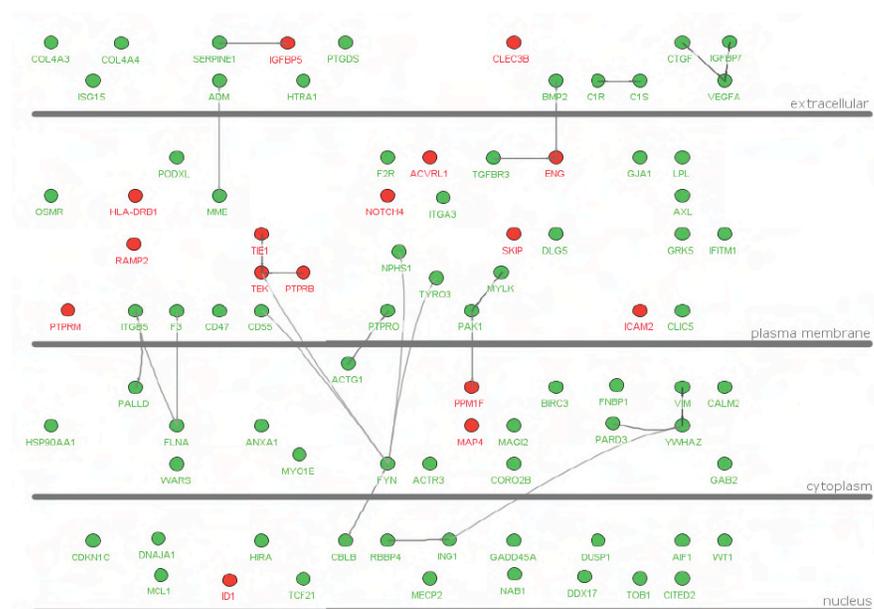
- Data driven hypothesis
 - Clusters indicate similar function?
 - Same pattern of gene expression → same role in cell?
- Clusters are often untrustworthy artifacts!
 - Data noisy
 - Different clustering algorithm → different results
 - **Show in context of graph model**

Adoption by biologists

- Matthew D Dyer, T. M Murali, and Bruno W Sobral. The landscape of human proteins interacting with viruses and other pathogens. PLoS Pathogens, 4(2):e32, 2008.



- Liqun He et al. The glomerular transcriptome and a predicted protein-protein interaction network. Journal of the American Society of Nephrology, 19(2):260-268, 2008.



More information

Cerebral: Visualizing Multiple Experimental Conditions on a Graph with Biological Context

Aaron Barsky, Computer Science, UBC

Tamara Munzner, Computer Science, UBC

Jennifer Gardy, Microbiology and Immunology, UBC

Robert Kincaid, Agilent Technologies

IEEE Transactions on Visualization and Computer Graphics (Proc. InfoVis 2008) 14(6) (Nov-Dec) 2008, p 1253-1260.

<http://www.cs.ubc.ca/labs/imager/tr/2008/cerebral/>

<http://www.cs.ubc.ca/labs/imager/th/2008/BarskyMscThesis/>

open-source software download

<http://www.pathogenomics.ca/cerebral/>

deployed in InnateDB (mammalian innate immunity database)

<http://www.innatedb.ca>