Visualization and Biology: Fertile Ground for Collaboration

Tamara Munzner

Department of Computer Science
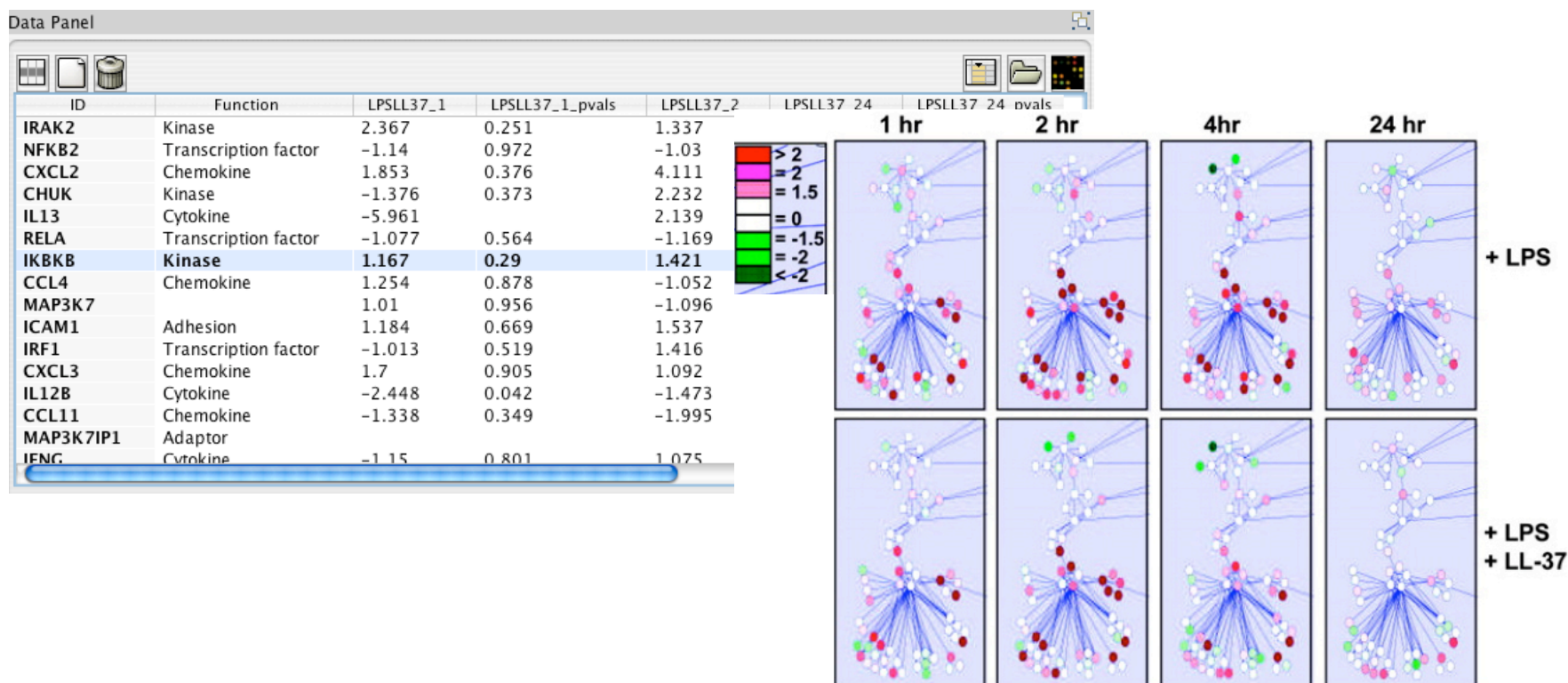
University of British Columbia

May 2010

http://www.cs.ubc.ca/~tmm/talks.html#bigdata10

Why do visualization?

- **pictures help us think**
  - substitute perception for cognition
 - external memory: free up limited cognitive/memory resources for higher-level problems

When should we bother doing vis?

- need a human in the loop
  - augment, not replace, human cognition
  - for problems that cannot be (completely) automated
- simple summary not adequate
  - statistics may not adequately characterize complexity of dataset distribution

Anscombe's quartet: same
- mean
- variance
- correlation coefficient
- linear regression line

http://upload.wikimedia.org/wikipedia/commons/b/b6/Anscombe.svg

3
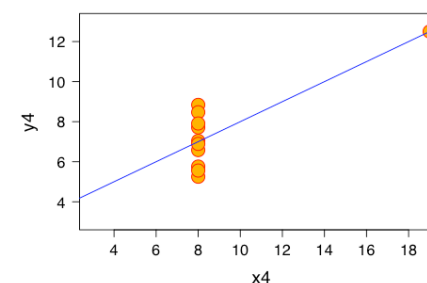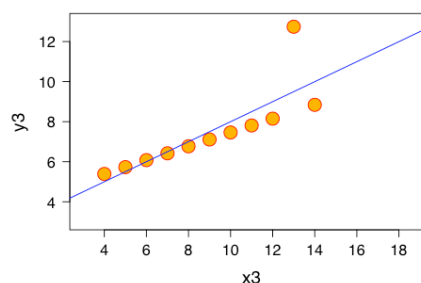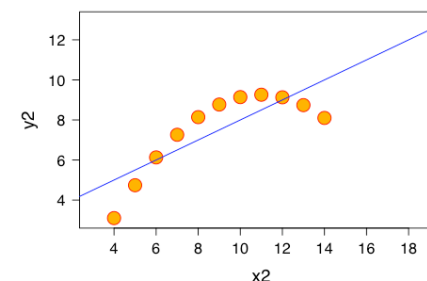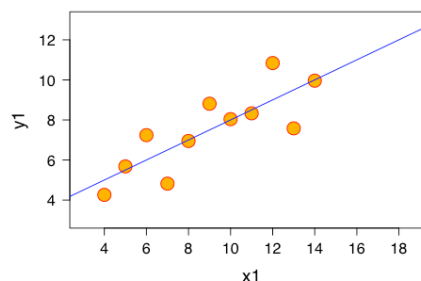
What does visualization allow?

- discovery vs. confirmation
  - discovering new things
    - hypothesis discovery, "eureka moment"
  - confirming conjectured things
    - hypothesis confirmation
  - contradicting conjectured things
    - especially (inevitably?) data cleansing

- discovery vs. speedup
  - novel capabilities
    - tool supports fundamentally new operations
  - **speedup**
    - tool accelerates workflow (most common!)

Good driving problems for vis research

- need for humans in the loop
- big data
- reasonably clear questions

- many areas of science are a great match
  - biology particularly appealing

Cerebral

collaboration with researchers at UBC Hancock Lab studying innate immunity

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

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

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

open-source software download (Cytoscape plugin)

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

deployed in InnateDB (mammalian innate immunity database)

http://www.innatedb.ca

Systems biology model

- graph G = {V, E}
  - V: proteins, genes, DNA, RNA, tRNA, etc.
 - E: interacting molecules
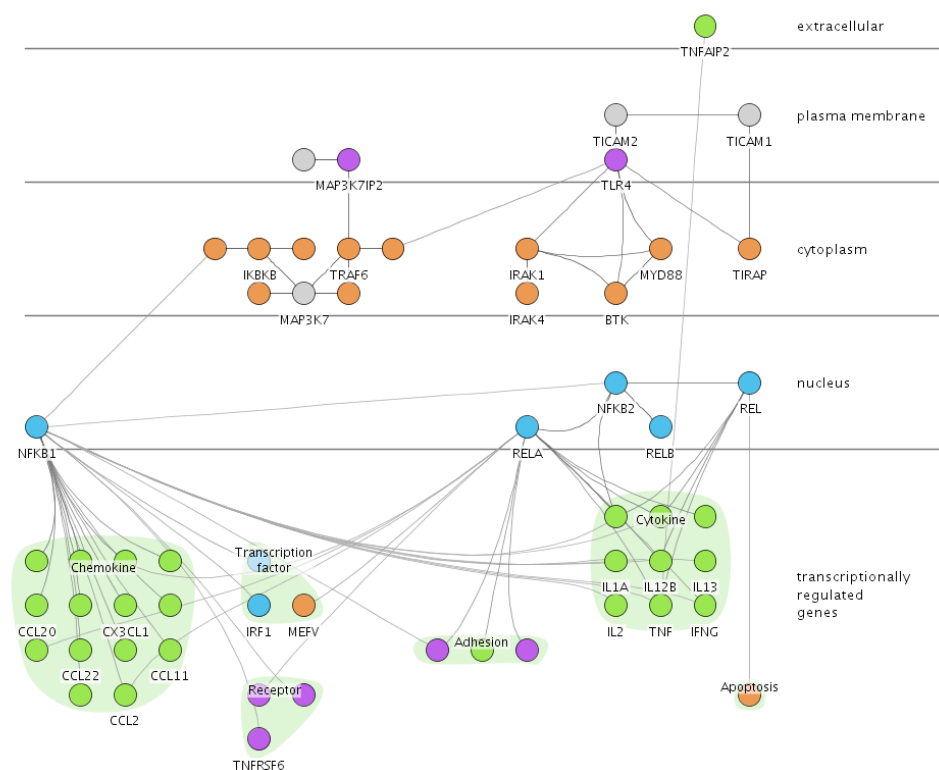
Model - Experiment cycle

- conduct experiments on cells
- interpret results in current graph model
- propose modifications to refine model

- vis tool to accelerate workflow?

Goal: Integrate model with measurements

- **system model**
  - interaction graph G = {V, E}
  - meta-data for each v in V
    - labels, biological attributes

- **experimental measurements**
 - multiple floats for each v in V
    - microarray data
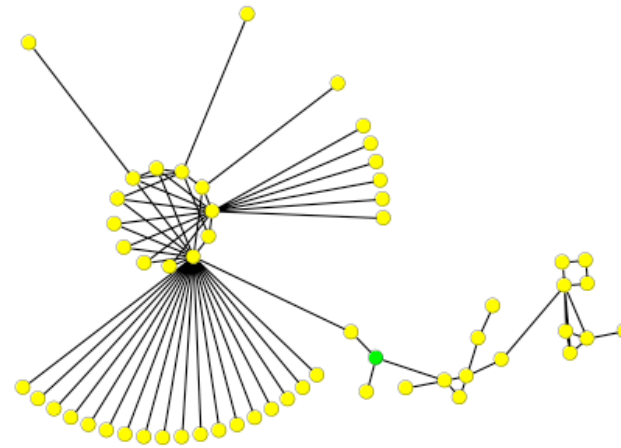
| ID	Function	LPSLL37_1	LPSLL37_1_pvals	LPSLL37_2	LPSLL37_24	LPSLL37_24_pvals
IRAK2	Kinase	2.367	0.251	1.337	−1.553	
NFKB2	Transcription factor	−1.14	0.972	−1.03	1.303	0.807
CXCL2	Chemokine	1.853	0.376	4.111	−1.019	0.745
CHUK	Kinase	−1.376	0.373	2.232	1.194	0.387
IL13	Cytokine	−5.961		2.139	−1.236	0.601
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					
IFNG	Cytokine	−1.15	0.801	1.075	1.053	0.521

9

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 for problem complexity
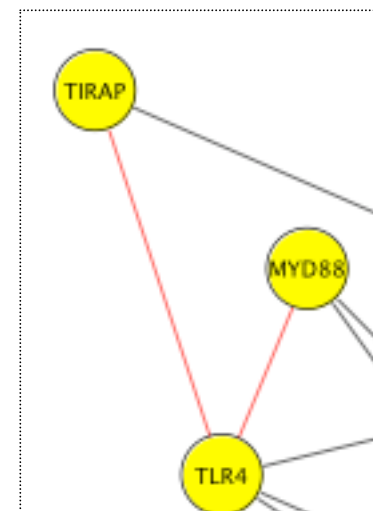
TLR4 biomolecule: E=74, V=54

- very local view

Immune system: E=1263, V=760

- bigger picture, target size for Cerebral
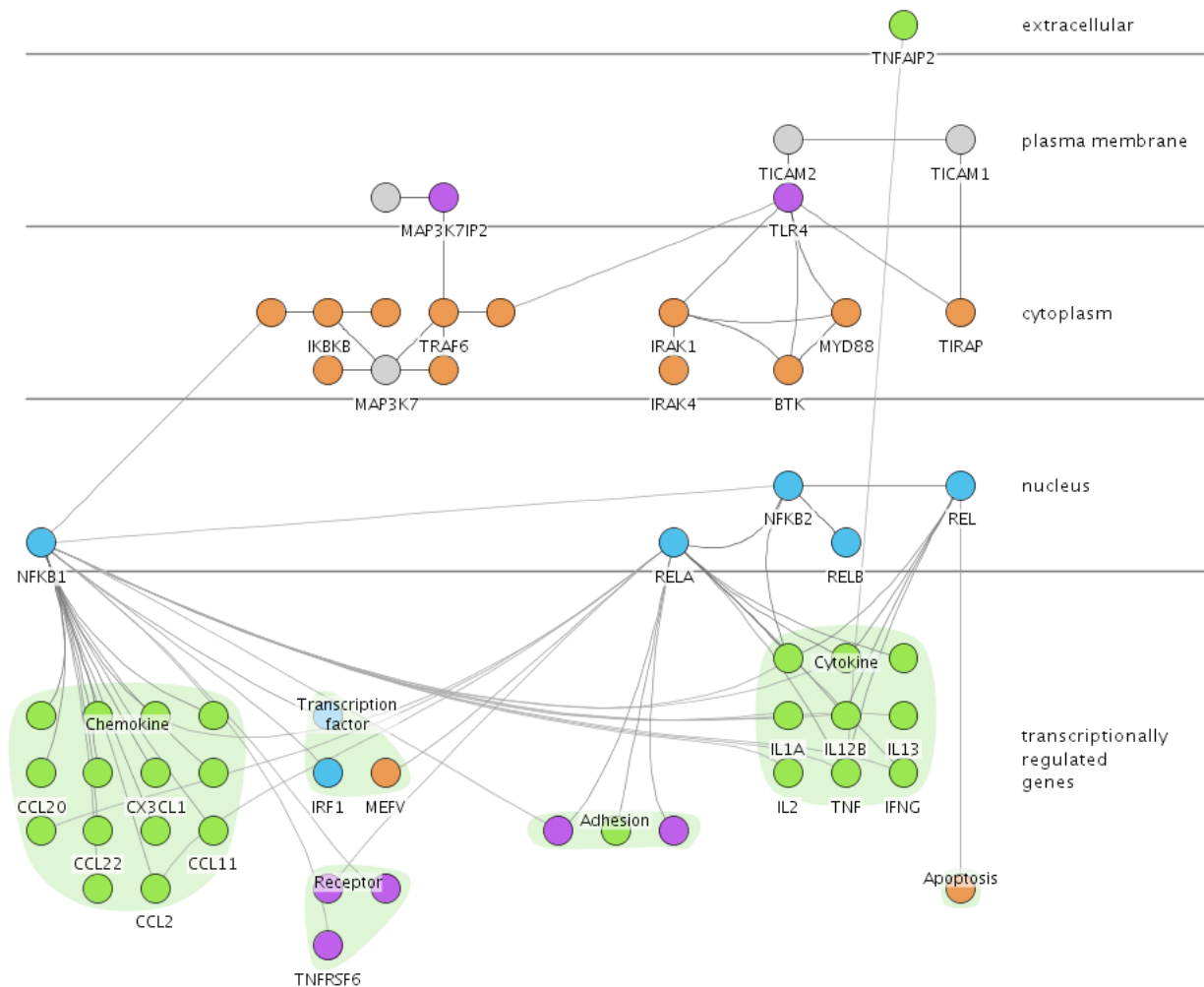
Human interactome: E~50,000, V~10,000

- too complex, beyond scope of tool
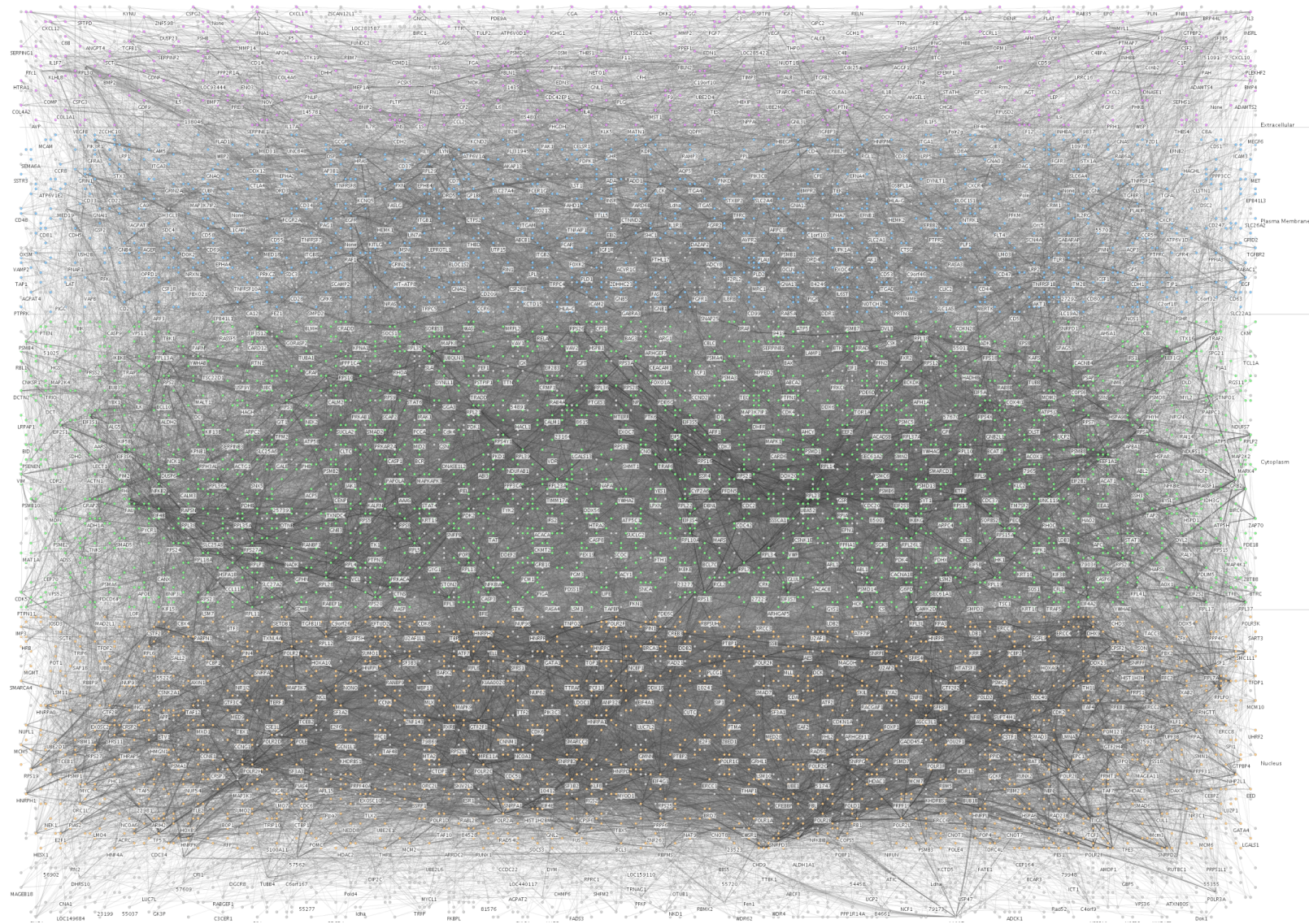
Cerebral video
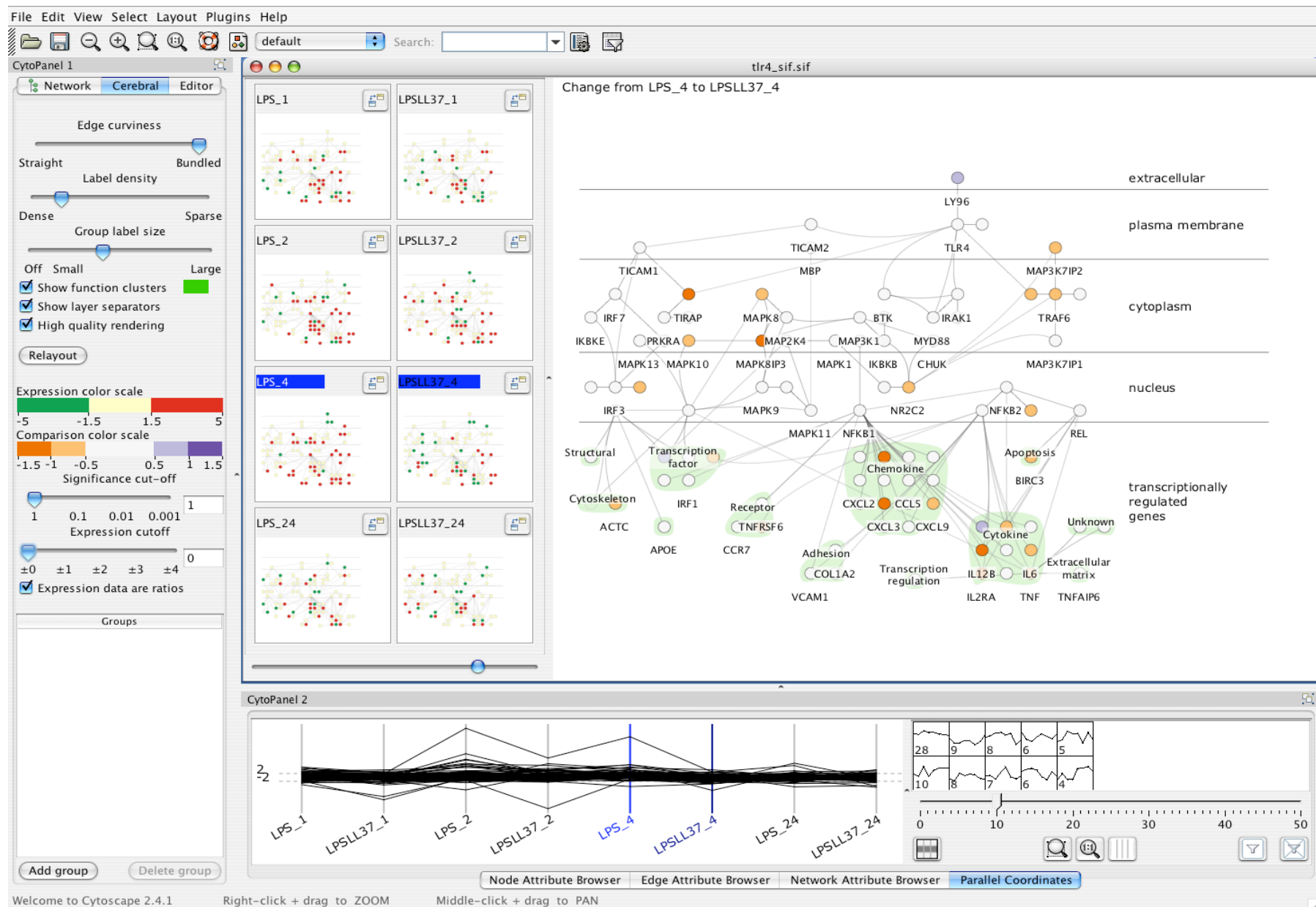
Encoding and interaction design decisions

- create custom graph layout
    - guided by biological metadata
- use small multiple views
    - one view per experimental condition
- show measured data in graph context
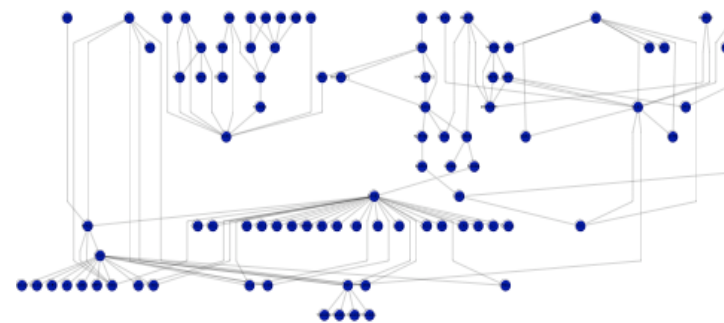   - not in isolation

Choice 1: Create custom graph layout

- graph layout heavily studied
  - given graph G={V,E}, create layout in 2D/3D plane
  - hundreds of papers
  - annual Graph Drawing conf.
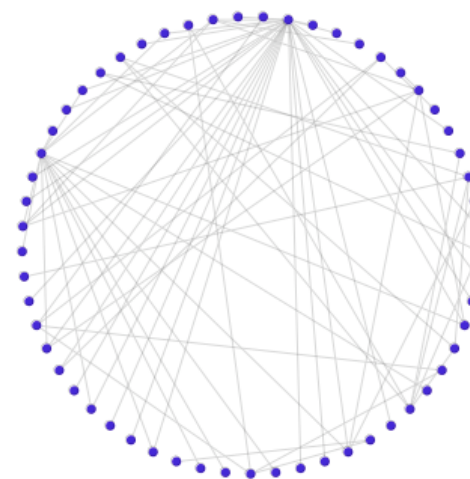
Circular (Six and Tollis, 1999)
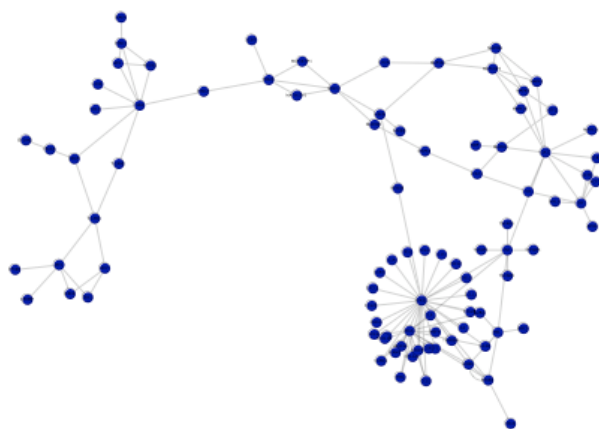
Force-directed
(Fruchterman and Reingold, 1991)

Hierarchical (Sugiyama 1989)

Existing layouts did not suit immunologists

- graph drawing goals
  - visualize graph structure
- biologist goals
  - visualize biological knowledge
  - some relationships happen to form a graph
  - cell location also relevant

Biological cells divided by membranes

- interactions generally occur within a compartment

- interaction location often known as part of model
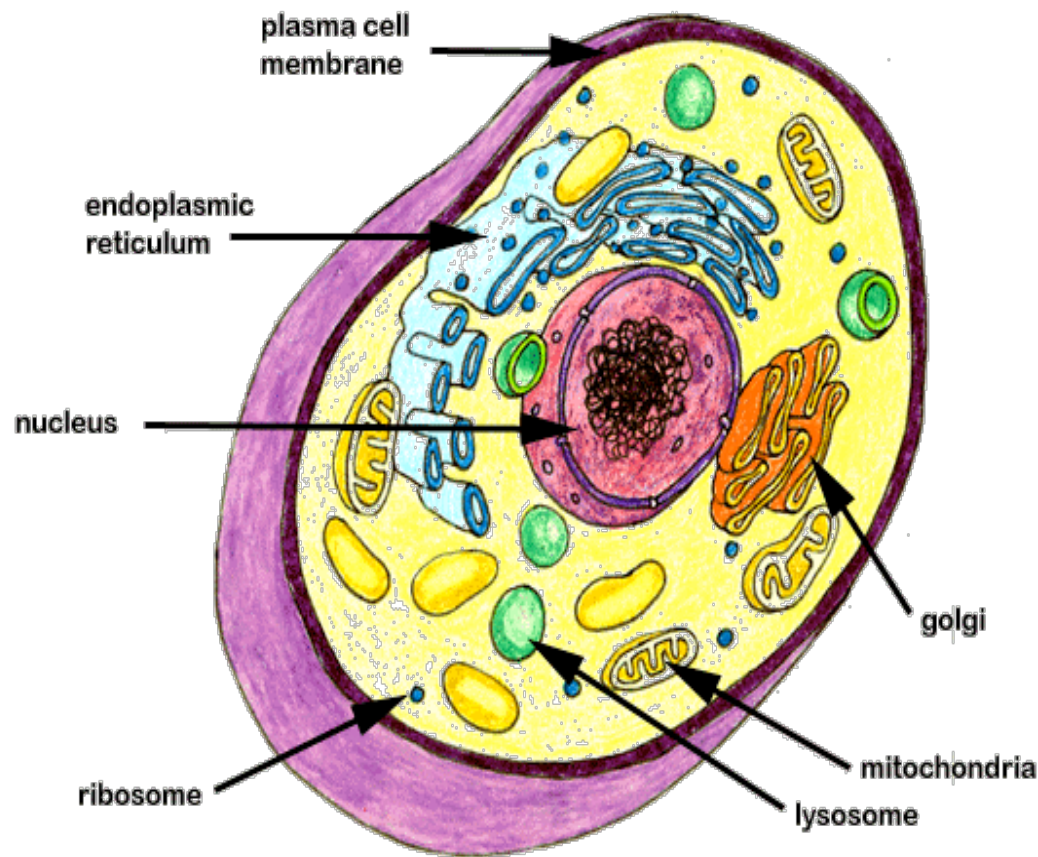
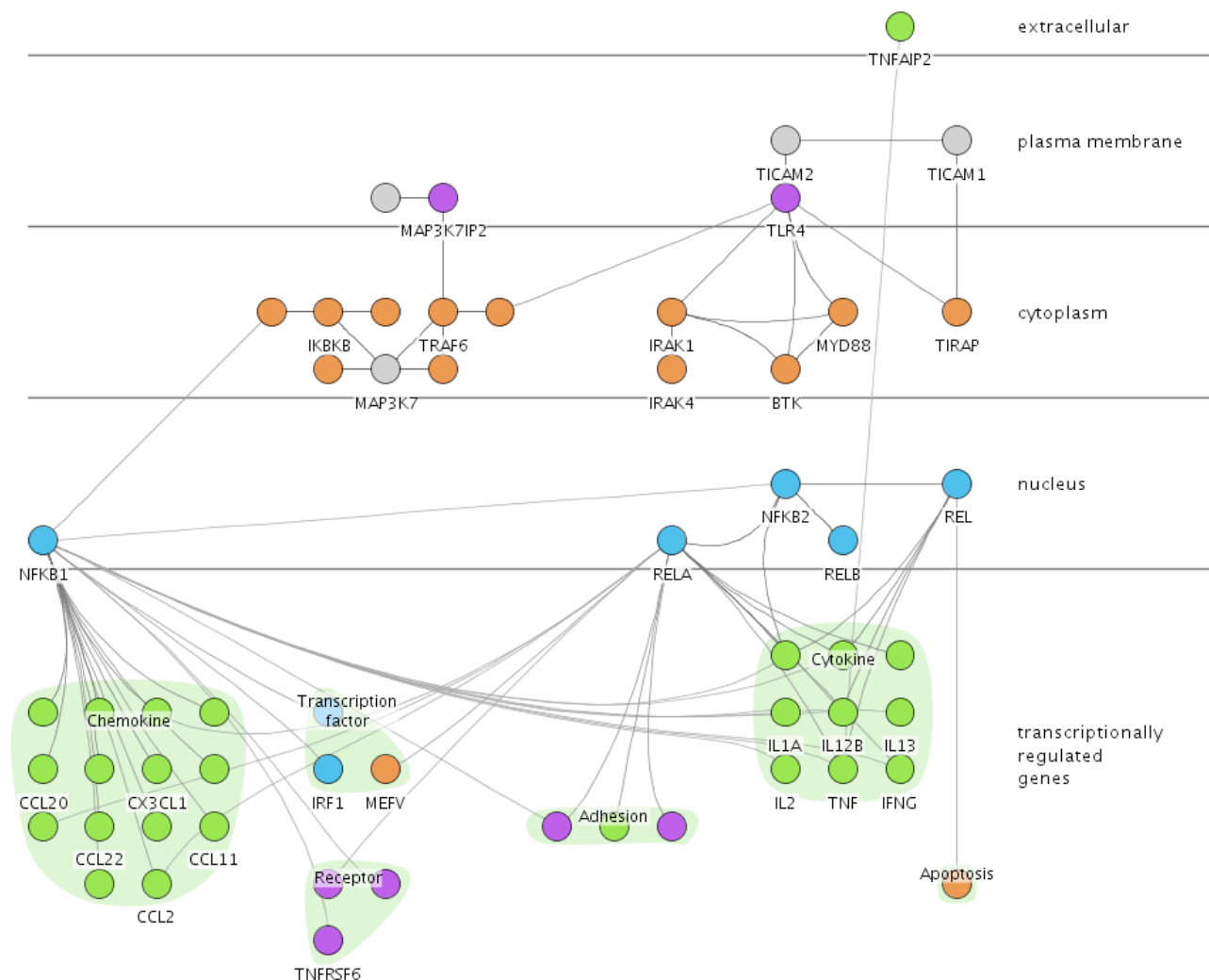Image credit: Dr.G Weaver, Colorado University at Denver

Hand-drawn diagrams

http://www.nature.com/nri/focus/tlr/nri1397.html
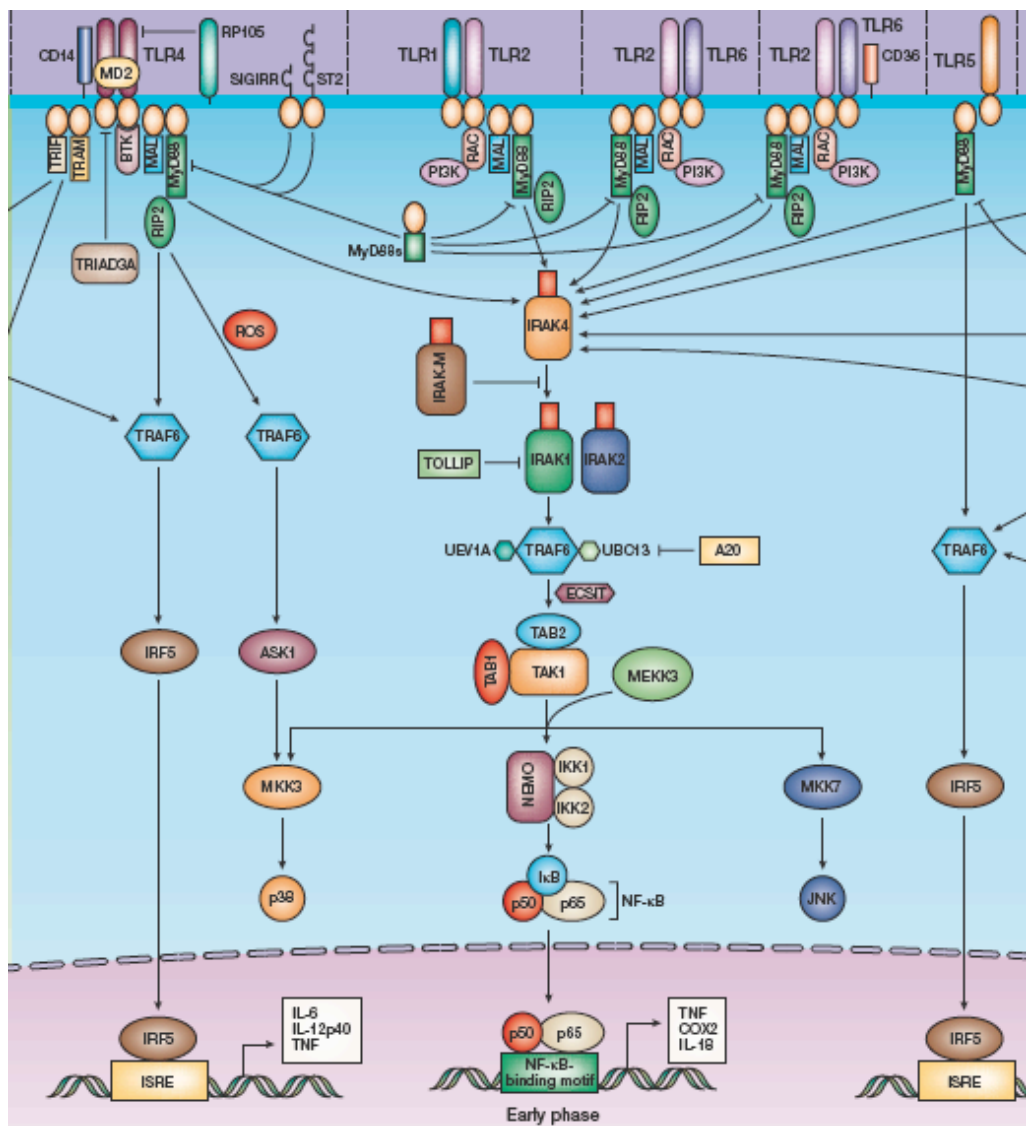
- cellular location spatially encoded vertically

- infeasible to create by hand in era of big data

19

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

20

Choice 2: Use small multiple views

- one graph instance per experimental condition
  - same spatial layout
  - color differently, by condition
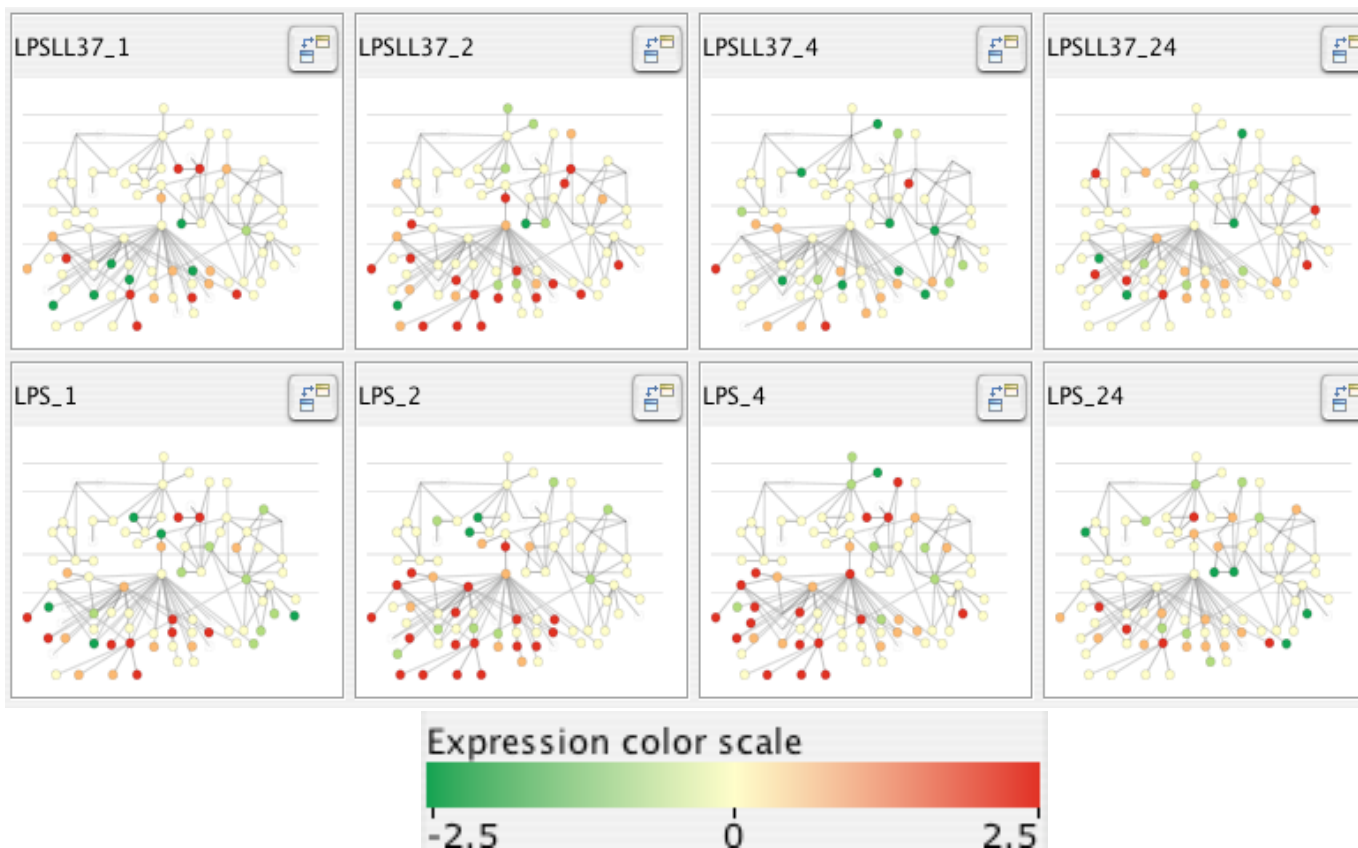
Why not animation?

- global comparison difficult
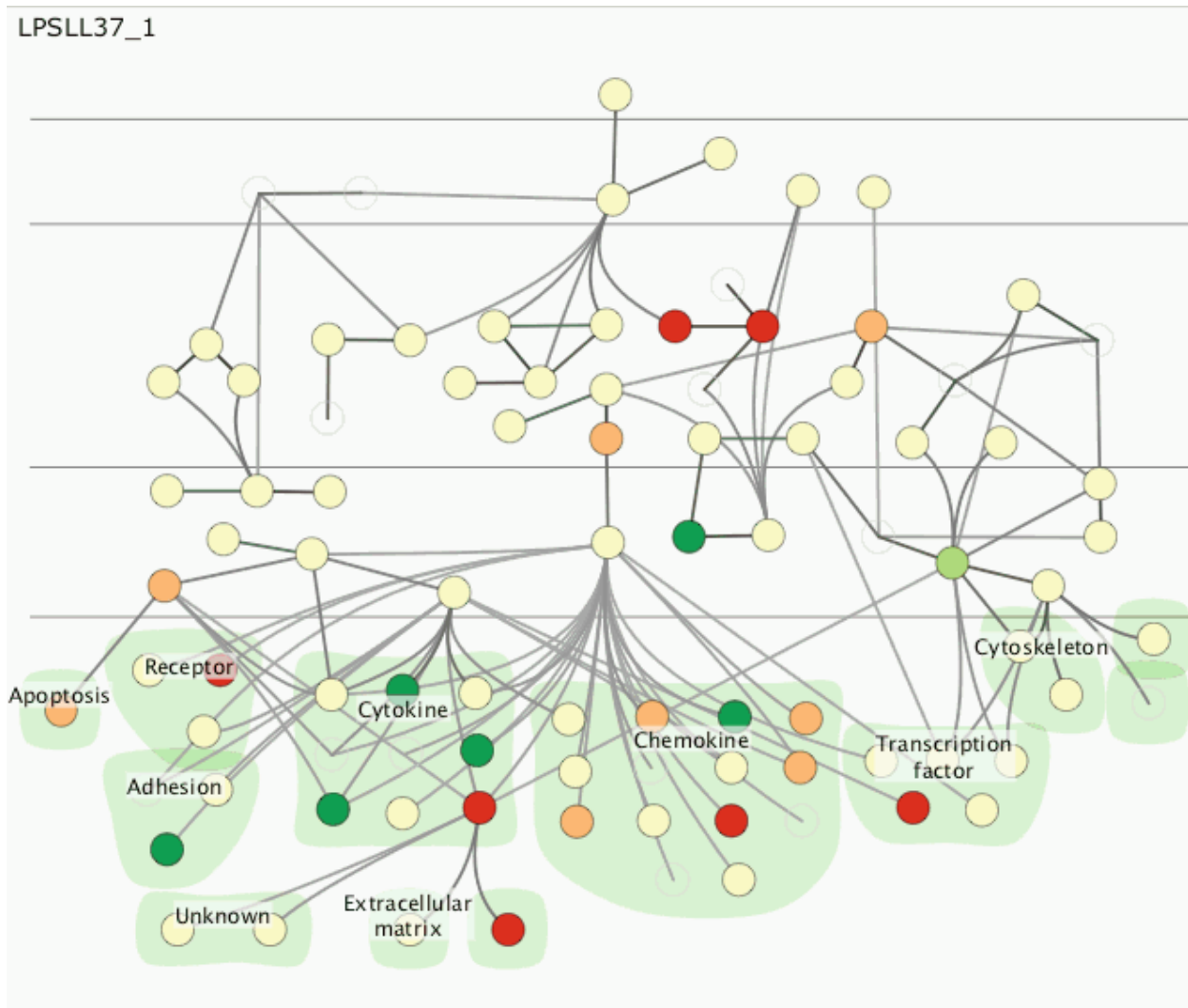
Why not animation?

- ## limits of human visual memory
  - – compared to side by side visual comparison

- Zooming versus multiple window interfaces: Cognitive costs of visual comparisons. Matthew Plumlee and Colin Ware. *ACM Trans. Computer-Human Interaction (ToCHI)*,13(2):179-209, 2006.

- Animation: can it facilitate? Barbara Tversky, Julie Bauer Morrison, and Mireille Betrancourt. *International Journal of Human-Computer Studies,* 57(4):247-262, 2002.

- Effectiveness of Animation in Trend Visualization. George Robertson, Roland Fernandez, Danyel Fisher, Bongshin Lee, John Stasko. IEEE Trans. Visualization and Computer Graphics 14(6):1325-1332 (Proc. InfoVis 08), 2008.

Why not glyphs?

- embed multiple conditions as a chart inside node
- clearly visible when zoomed in
- but cannot see from global view
  - only one value shown in overview
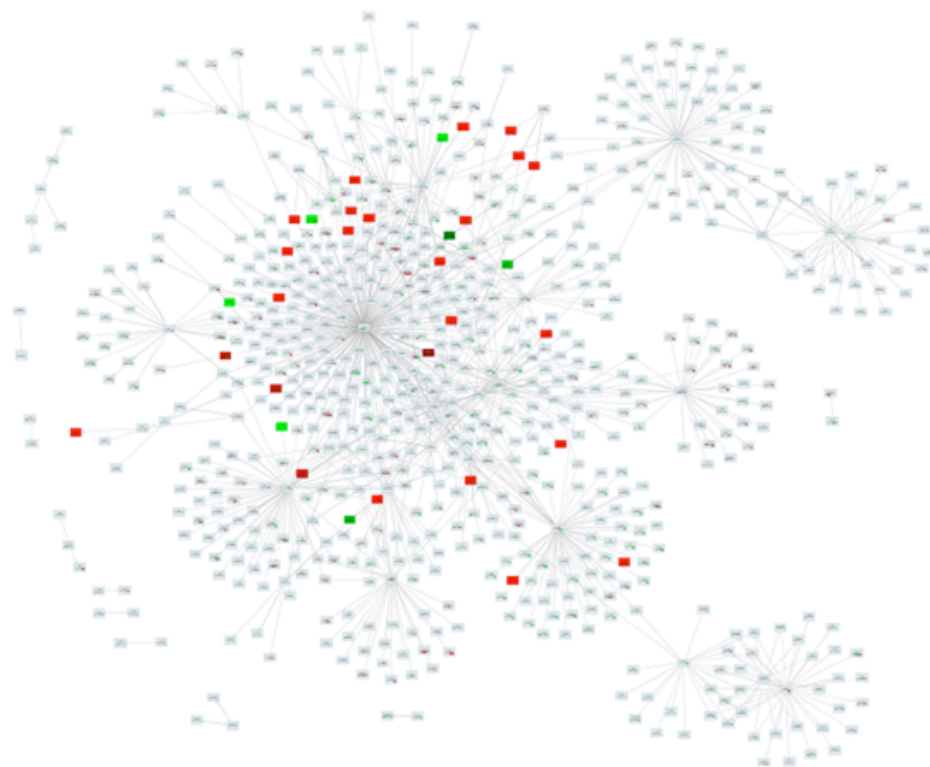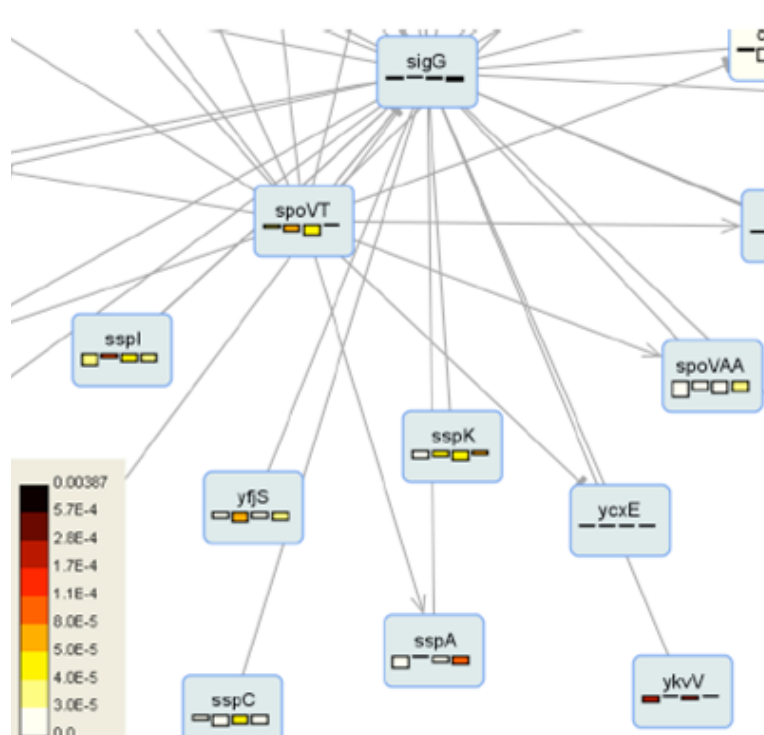
[M. A. Westenberg, S. A. F. T. van Hijum, O. P. Kuipers, J. B. T. M. Roerdink. Visualizing Genome Expression and Regulatory Network Dynamics in Genomic and Metabolic Context. Computer Graphics Forum, 27(3):887-894, 2008.]

Choice 3: Show measurements and graph

- why not measurements alone?
  - data driven hypothesis: gene expression clusters indicate similar function in cell?

- clusters are often untrustworthy artifacts!
  - noisy data: different clustering alg. → different results
  - measured data alone potentially misleading
  - **show in context of graph model**
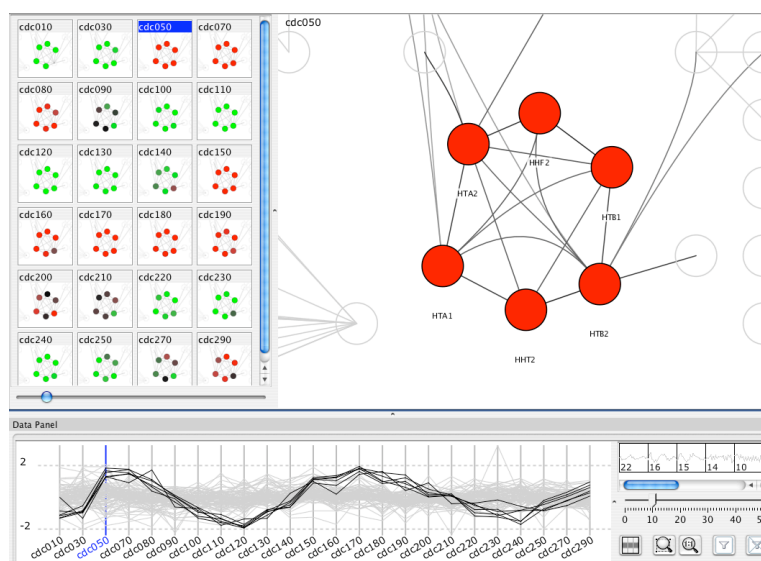
Adoption by biologists

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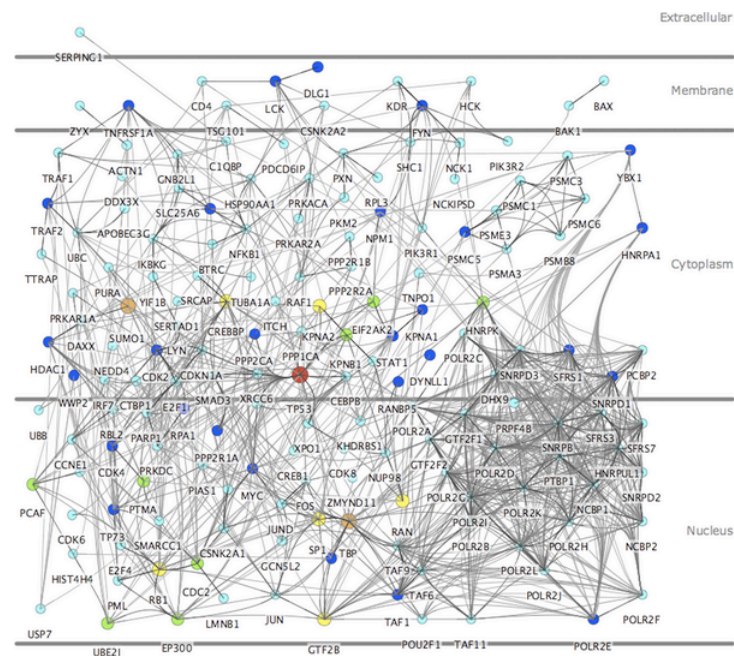
- Liqun He et al. The glomerular transcriptome and a predicted protein-protein interaction network. Journal of the American Society of Nephrology, 19(2):260-268, 2008.
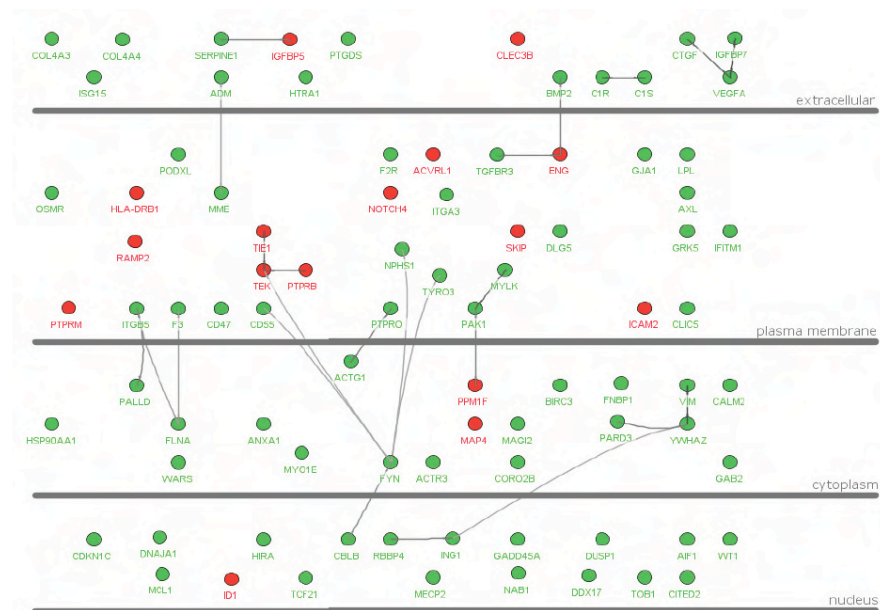
InnateDB links to Cerebral

- InnateDB: facilitating systems-level analyses of the mammalian innate immune response
  - David J Lynn, Geoffrey L Winsor, Calvin Chan, Nicolas Richard, Matthew R Laird, Aaron Barsky, Jennifer L Gardy, Fiona M Roche, Timothy H W Chan, Naisha Shah, Raymond Lo, Misbah Naseer, Jaimmie Que, Melissa Yau, Michael Acab, Dan Tulpan, Matthew D Whiteside, Avinash Chikatamarla, Bernadette Mah, Tamara Munzner, Karsten Hokamp, Robert E W Hancock, Fiona S L Brinkman. Molecular Systems Biology 2008; 4:218

  - http://innatedb.ca

Data cleansing example

- incorrect edge across many compartments
  - in well studied dataset
  - not obvious with other layouts
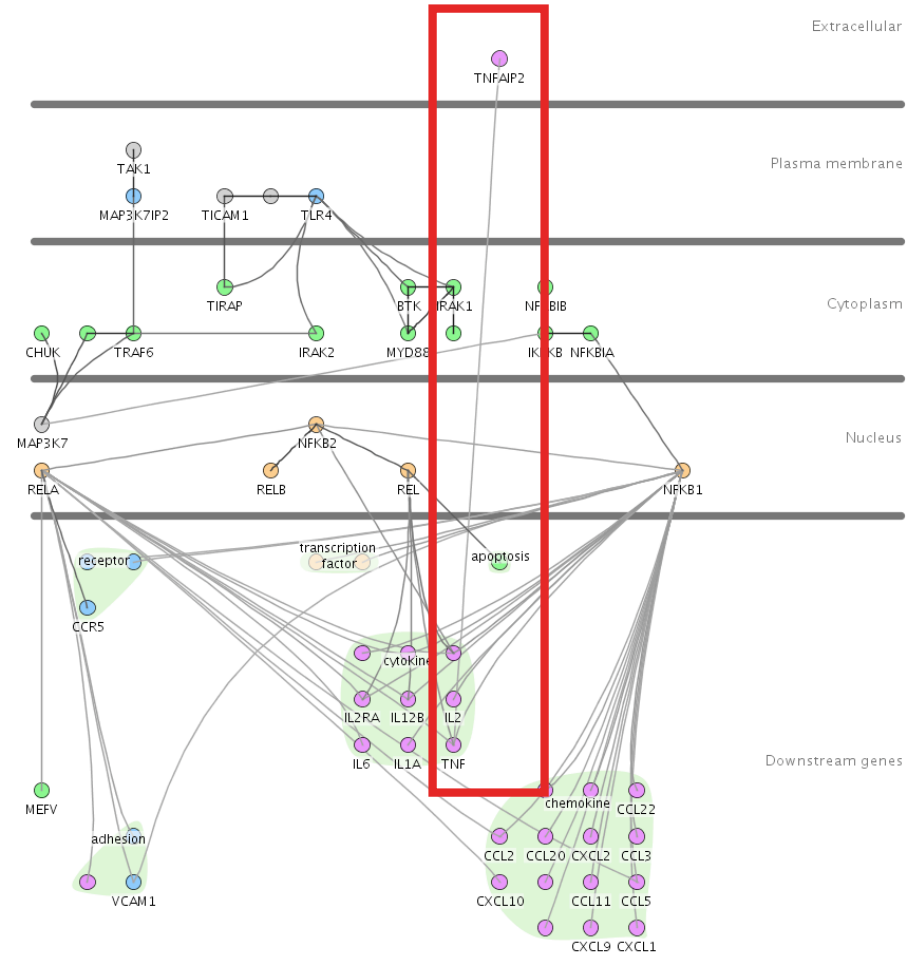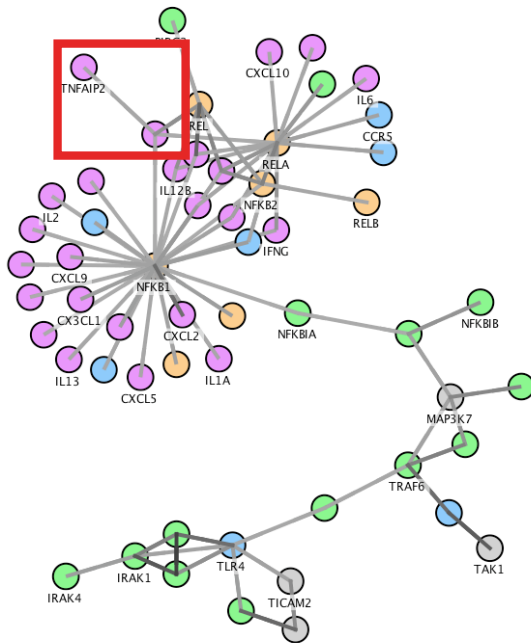
Cerebral summary

- supports interactive exploration of multiple experimental conditions in graph context

- provides familiar representation by using biological metadata to guide graph layout

TreeJuxtaposer: scalable tree comparison

collab with UT-Austin Hillis Lab: phylogenetics

TreeJuxtaposer: Scalable Tree Comparison using Focus+Context with
Guaranteed Visibility.
Tamara Munzner, François Guimbretière, Serdar Tasiran, Li Zhang, Yunhong Zhou.
ACM Trans. Graphics 22(3): 453-462, 2003 (Proc. SIGGRAPH 2003).

http://www.cs.ubc.ca/labs/imager/tr/2003/tj

open-source software, video

http://olduvai.sourceforge.net/tj
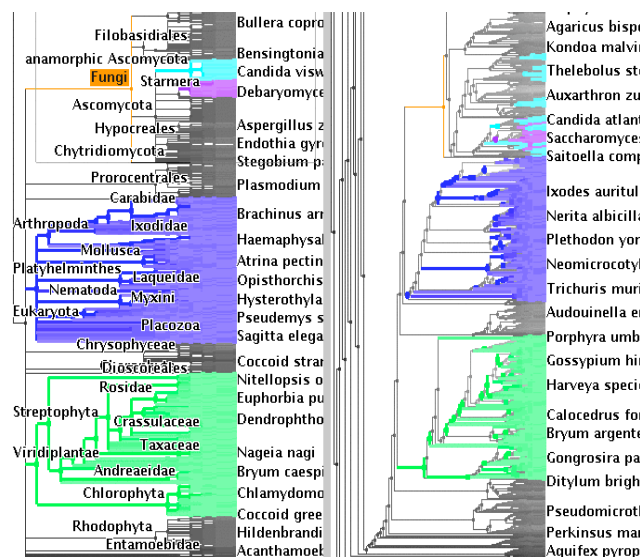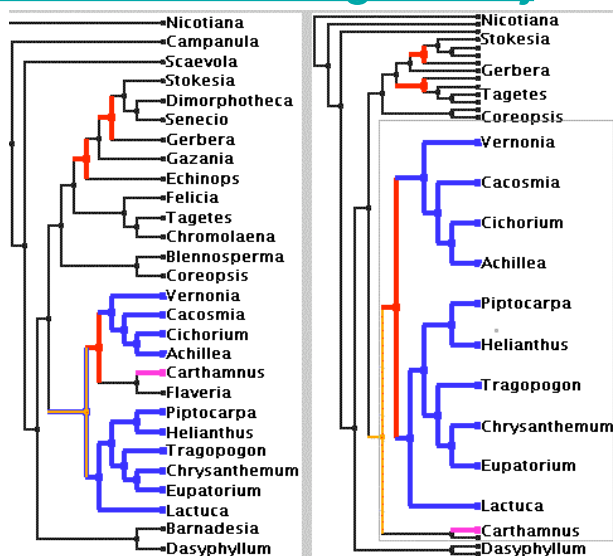
MizBee: multiscale synteny browser

collab with researchers at Broad Institute: synteny

MizBee: A Multiscale Synteny Browser
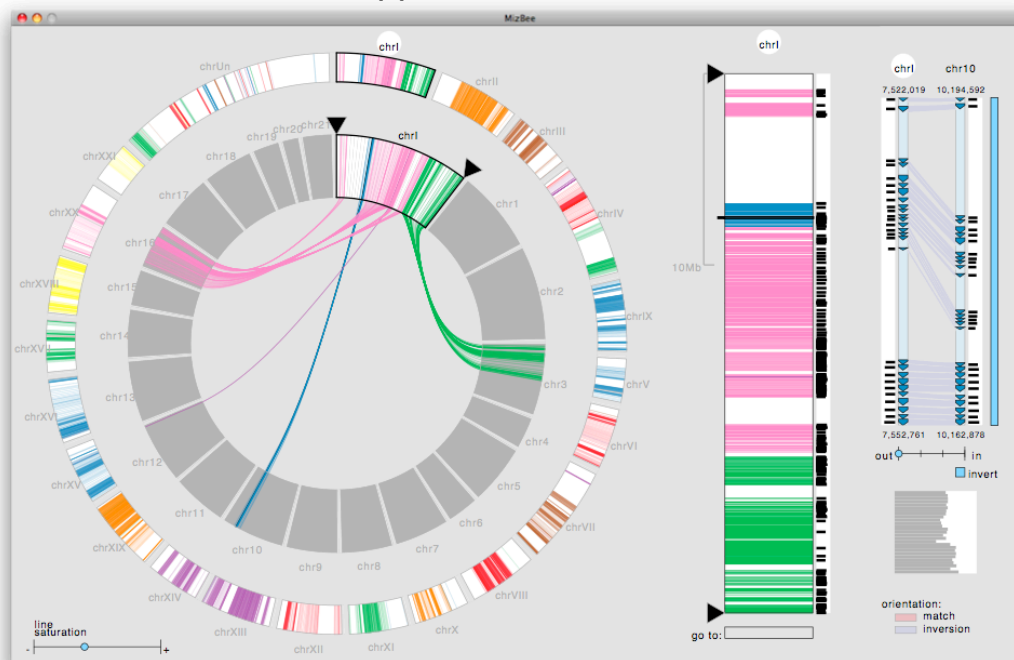
Miriah Meyer, Computer Science, Harvard

Tamara Munzner, Computer Science, UBC

Hanspeter Pfister, Computer Science, Harvard

IEEE Trans. Visualization and Computer Graphics (Proc. InfoVis 09), 15(6):897-904, 2009.

http://www.mizbee.org

paper, open-source software and executable applet, video

Pathline: comparative functional genomics

collab with Broad Regev lab: yeast regulatory networks

Pathline: A Tool for Comparative Functional Genomics.
Miriah Meyer, Computer Science, Harvard
Bang Wong, Broad Institute
Mark Styczynski, Chemical and Biomolecular Engineering, Georgia Tech
Tamara Munzner, Computer Science, UBC
Hanspeter Pfister, Computer Science, Harvard
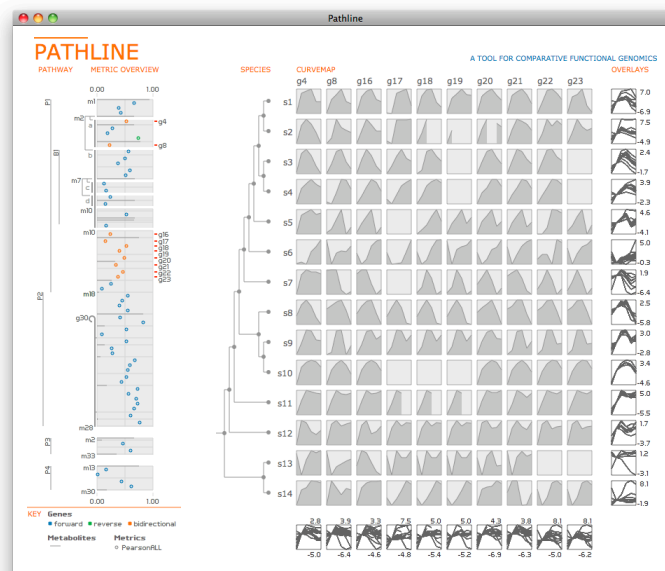Computer Graphics Forum (Proceedings of EuroVis 2010), to appear.

http://iic.harvard.edu/~miriah/publications/pathline.pdf
http://www.pathline.org
software to be released soon

multiple
  pathways
  genes/metabolites
  species

Vast opportunities

- young field, still much to be done

- think about your current workflow
  - what could you speed up by swapping in perception for cognition?
  - exploit the familiar, yet consider breadth of design alternatives

- finding some friendly vis collaborators

  - IEEE VisWeek 2010 (Vis, InfoVis, VAST)
    Oct 11-16, Salt Lake City
    http://vis.computer.org/VisWeek2010

  - EuroVis 2010: Jun 9-11, Bordeaux France
    http://eurovis2010.labri.fr/

More information

- this talk
  http://www.cs.ubc.ca/~tmm/talks.html#bigdata10

- papers, talks, videos...
  http://www.cs.ubc.ca/~tmm

- visualization intro book chapter
  http://www.cs.ubc.ca/~tmm/papers.html#akpchapter