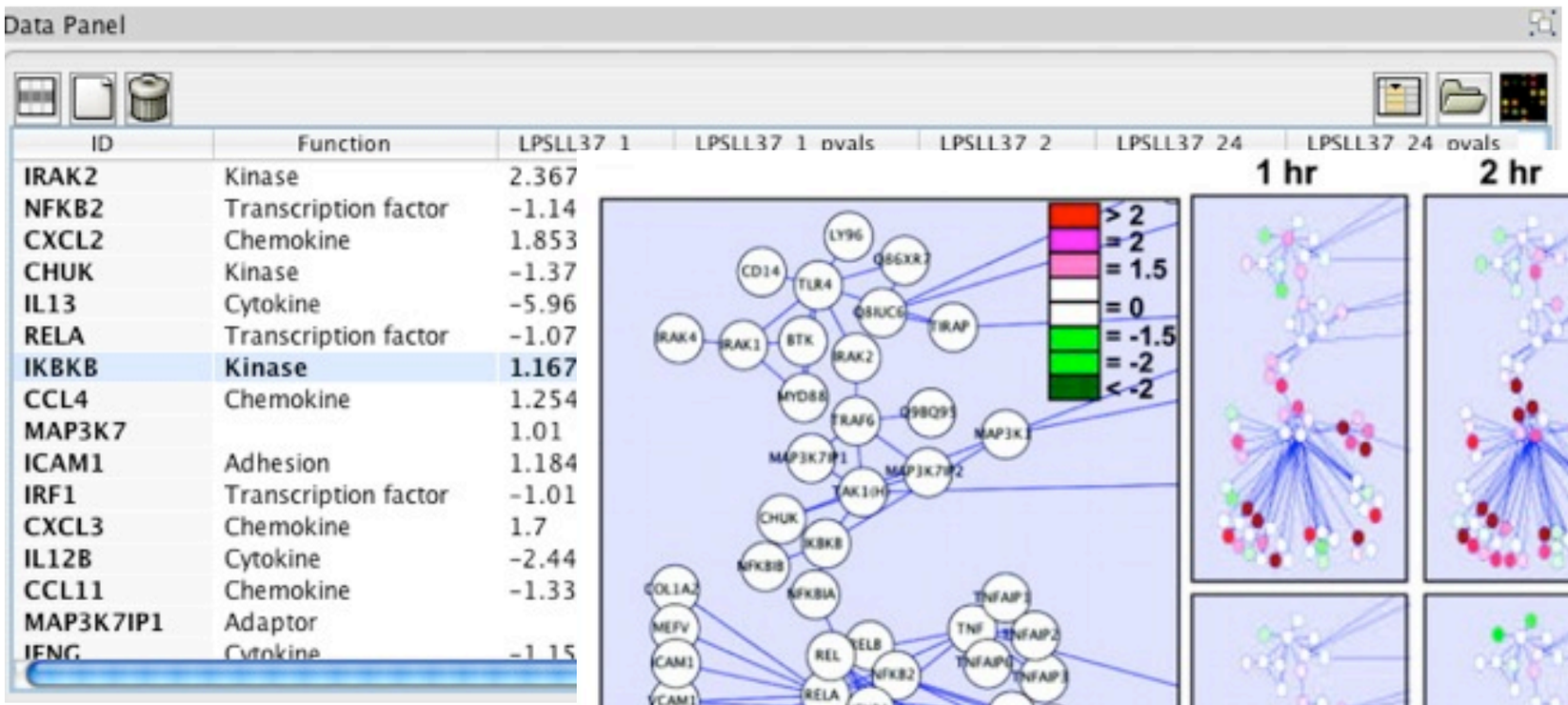
Biology is Destiny: Of Graphs and Genes

Tamara Munzner
University of British Columbia

UBC CS TechTrek 2012
21 Jan 2012

Why do visualization?

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



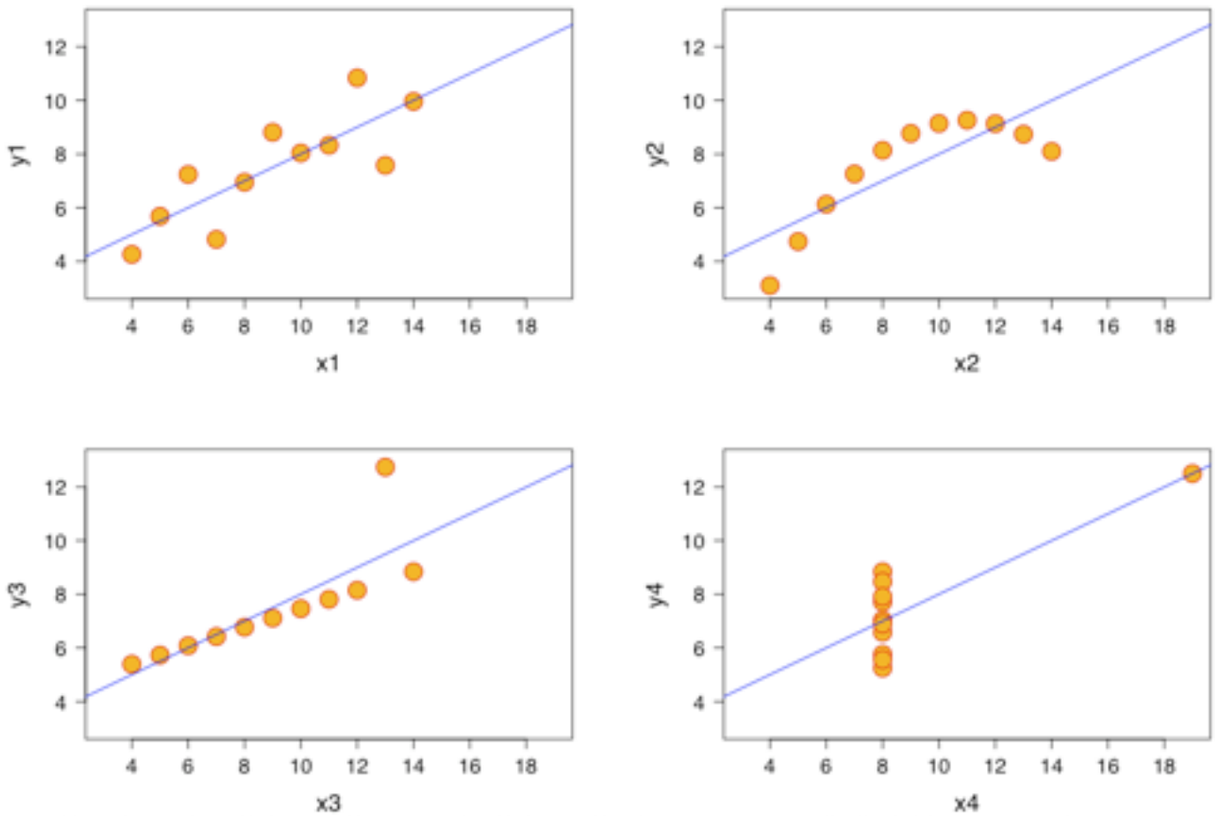
2

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

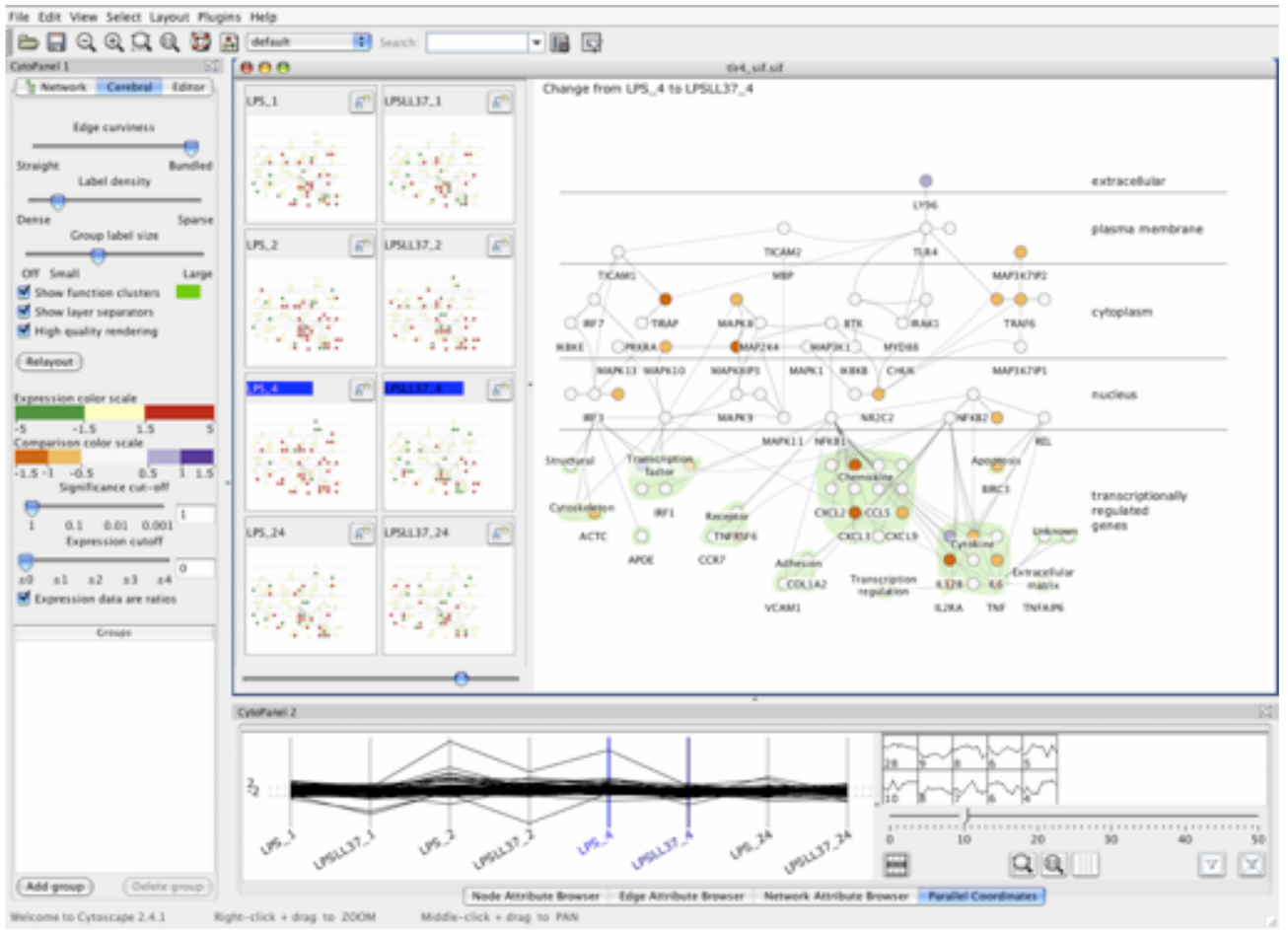
- mean
- variance
- correlation coefficient
- linear regression line



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

3

Cerebral



Comparing Multiple Experimental Conditions Within Biologically Meaningful Network Context

joint work with:

Aaron Barsky, Jennifer Gardy, Robert Kincaid

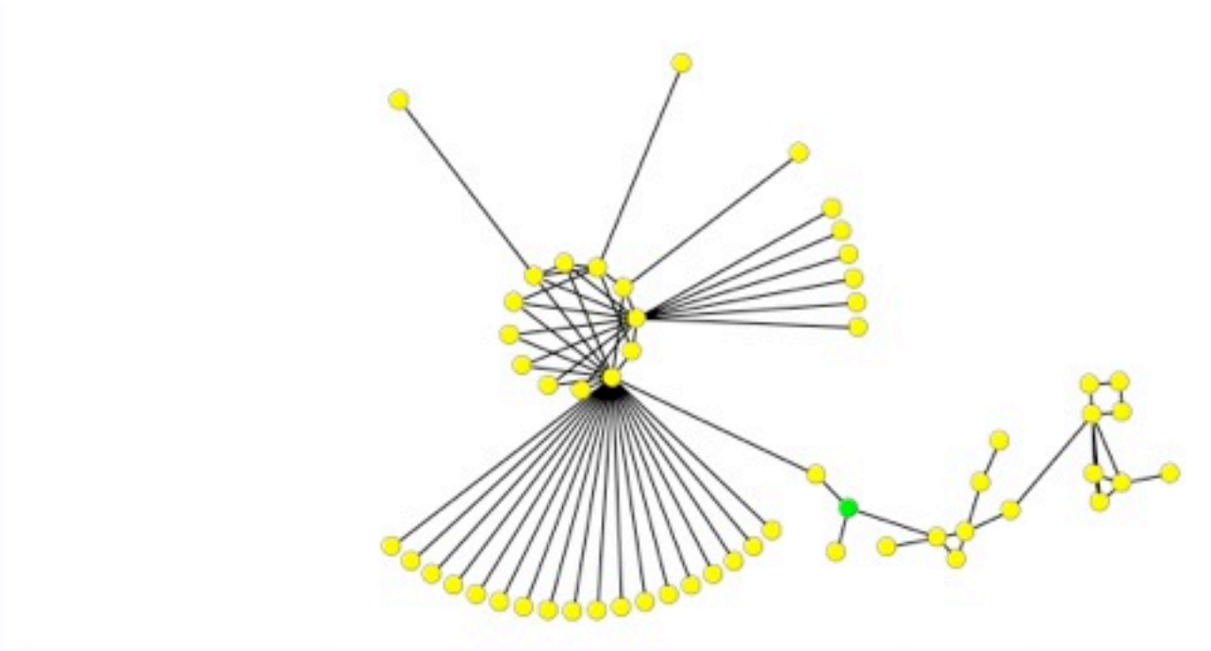
http://www.pathogenomics.ca/cerebral/

Cerebral: Visualizing Multiple Experimental Conditions on a Graph with Biological Context.
Barsky, Munzner, Gardy, Kincaid. IEEE InfoVis 2008.

4

Systems biology model

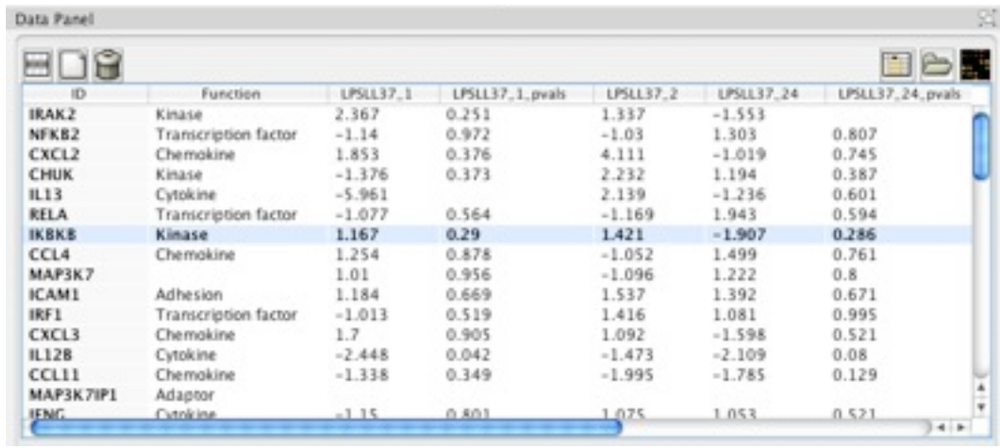
- graph G = {V, E}
  - V: proteins, genes, DNA, RNA, tRNA, etc.
    - metadata: labels, biological attributes
  - E: interacting molecules
    - known from previous research



5

Cycle: model - experiment

- conduct experiments on cells
  - –microarrays
  - –measurements for each vertex in graph



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

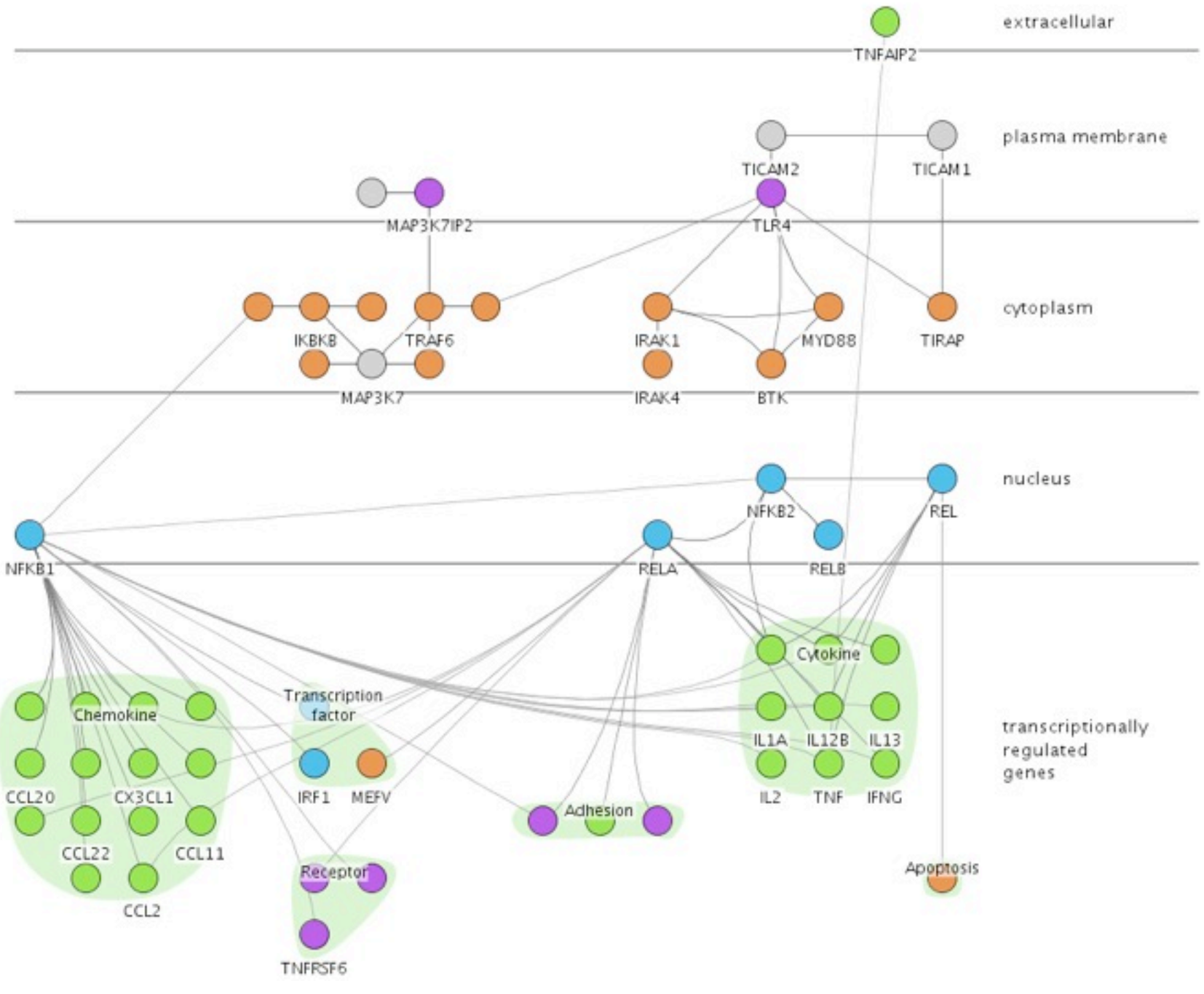
- interpret results in current graph model
- propose modifications to refine model

- vis tool to accelerate workflow
  - –integrated tool to see graph and measurements together
  - –choose scope for problem complexity

6

TLR4 biomolecule: E=74,V=54

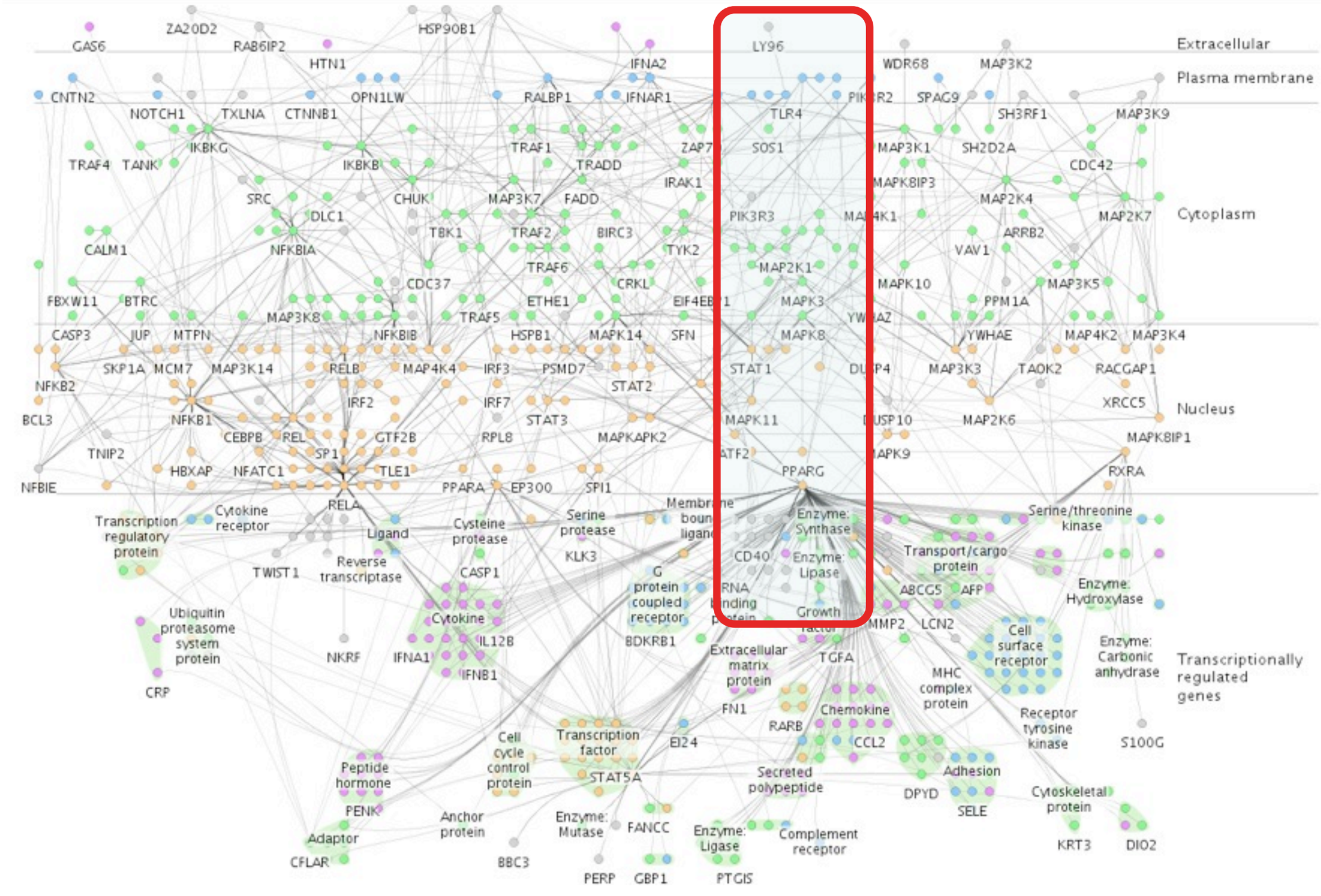
- very local view



7

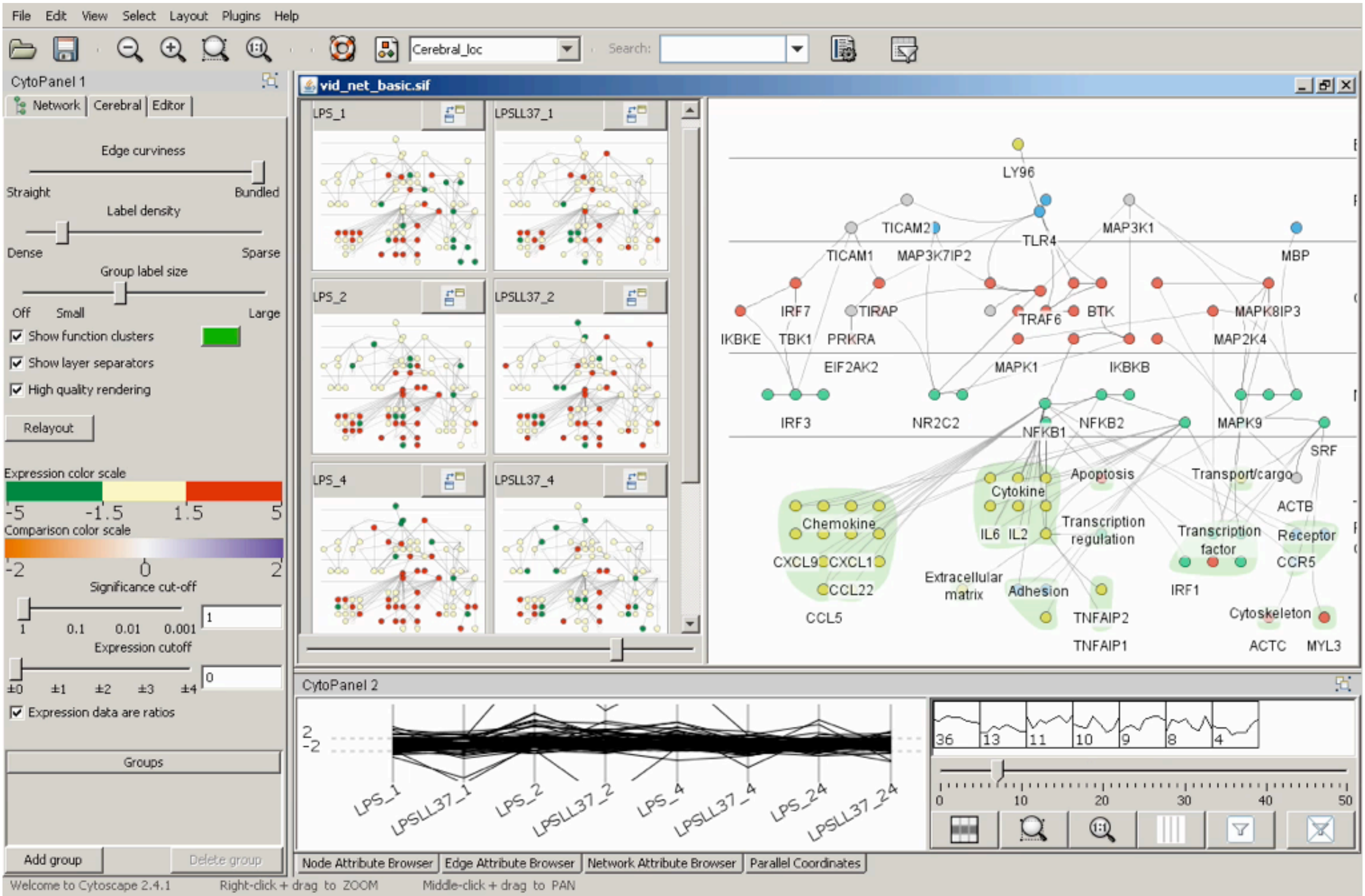
Immune system: E=1263,V=760

- bigger picture, target size for Cerebral



8

Cerebral video



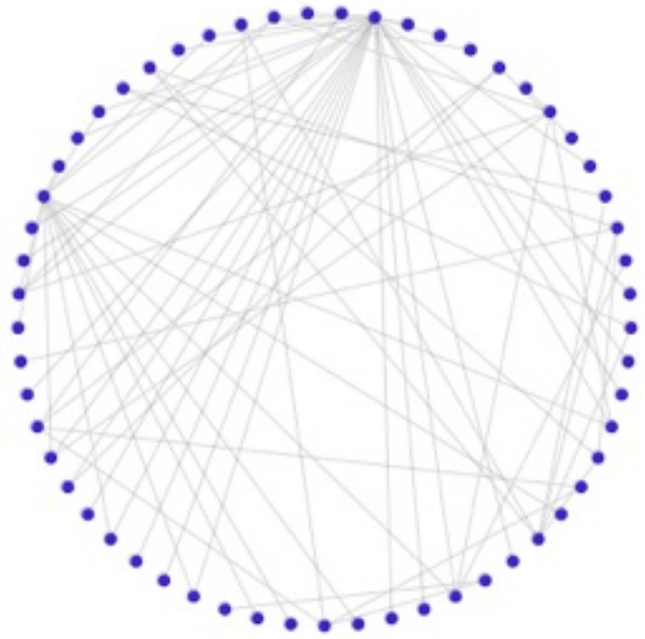
Encoding and interaction design decisions

- 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

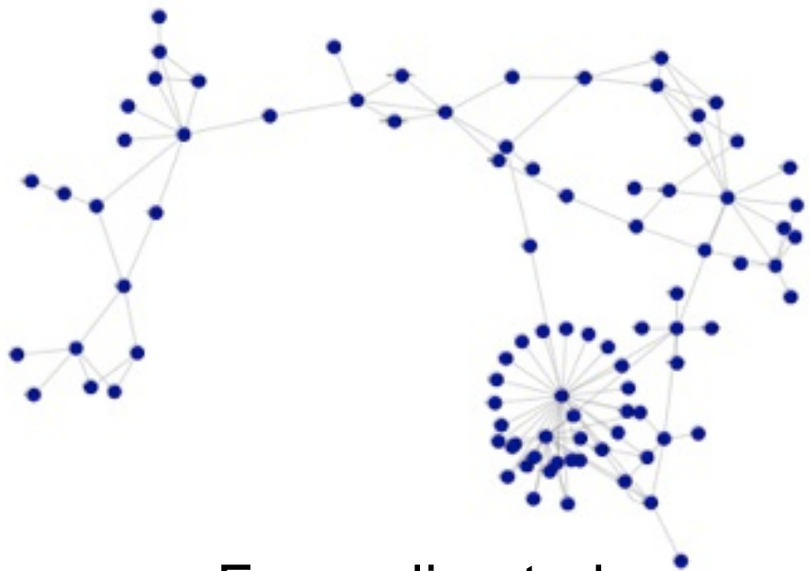
10

Choice: Create custom graph layout

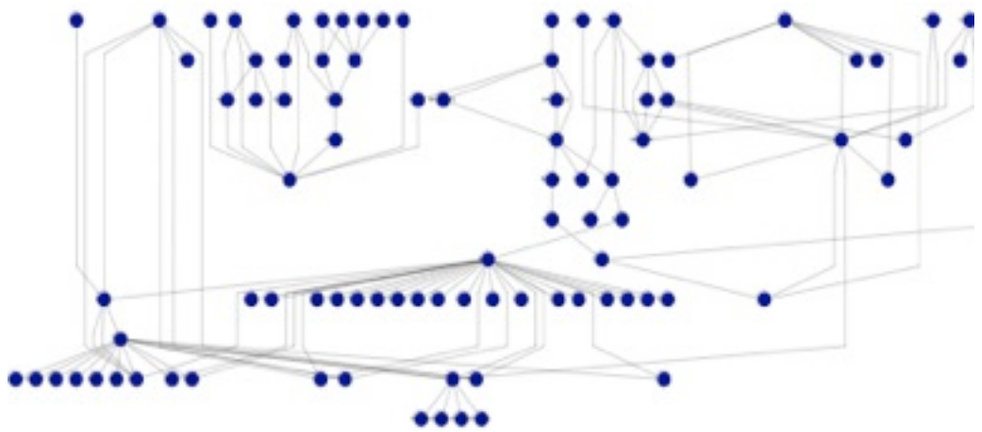
- graph layout heavily studied
  - given graph G={V,E},
    create layout in 2D/3D plane
  - hundreds of papers
  - annual Graph Drawing conf.



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

12

Biological cells divided by membranes

- interactions generally occur within a compartment
- interaction location often known as part of model

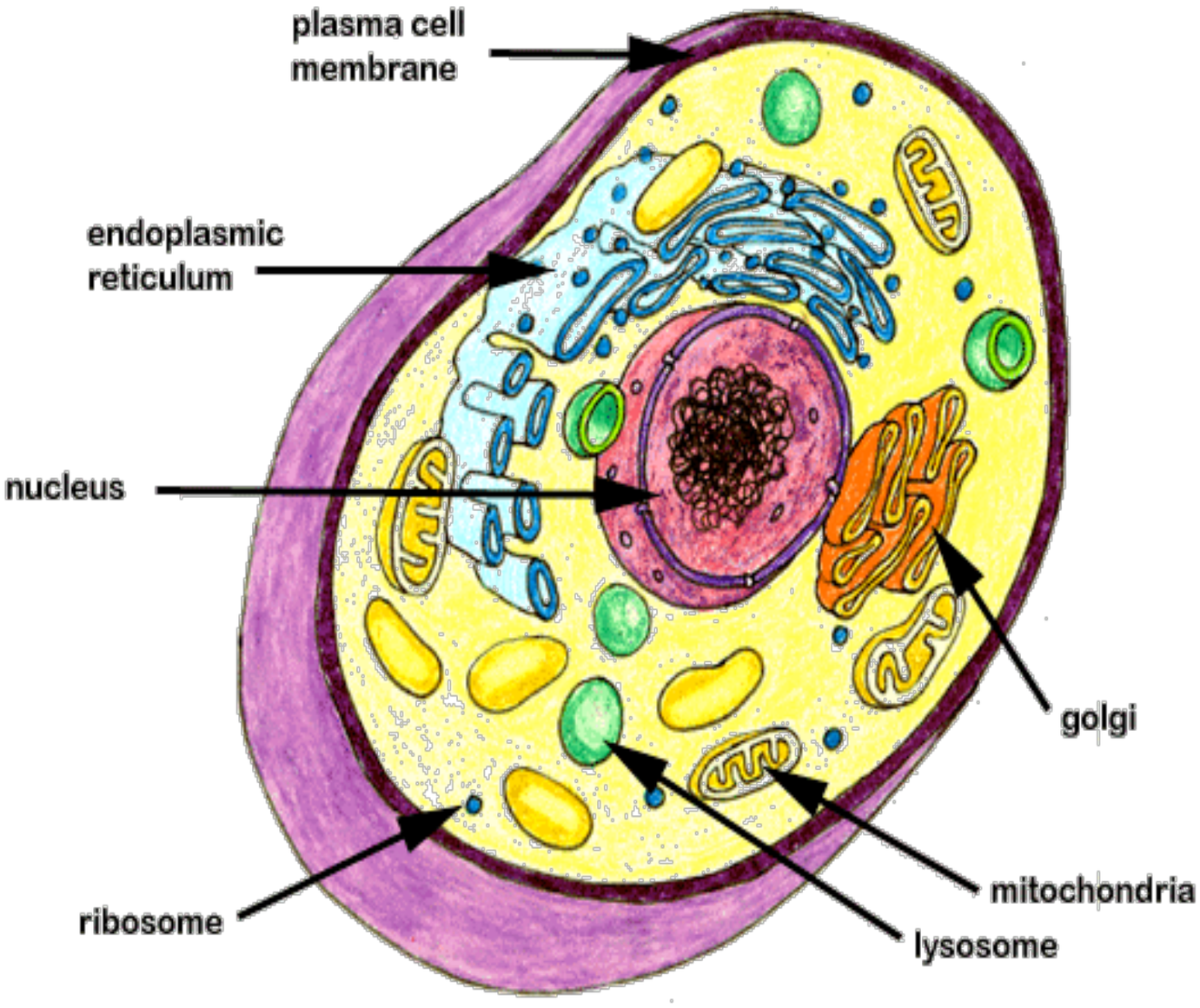
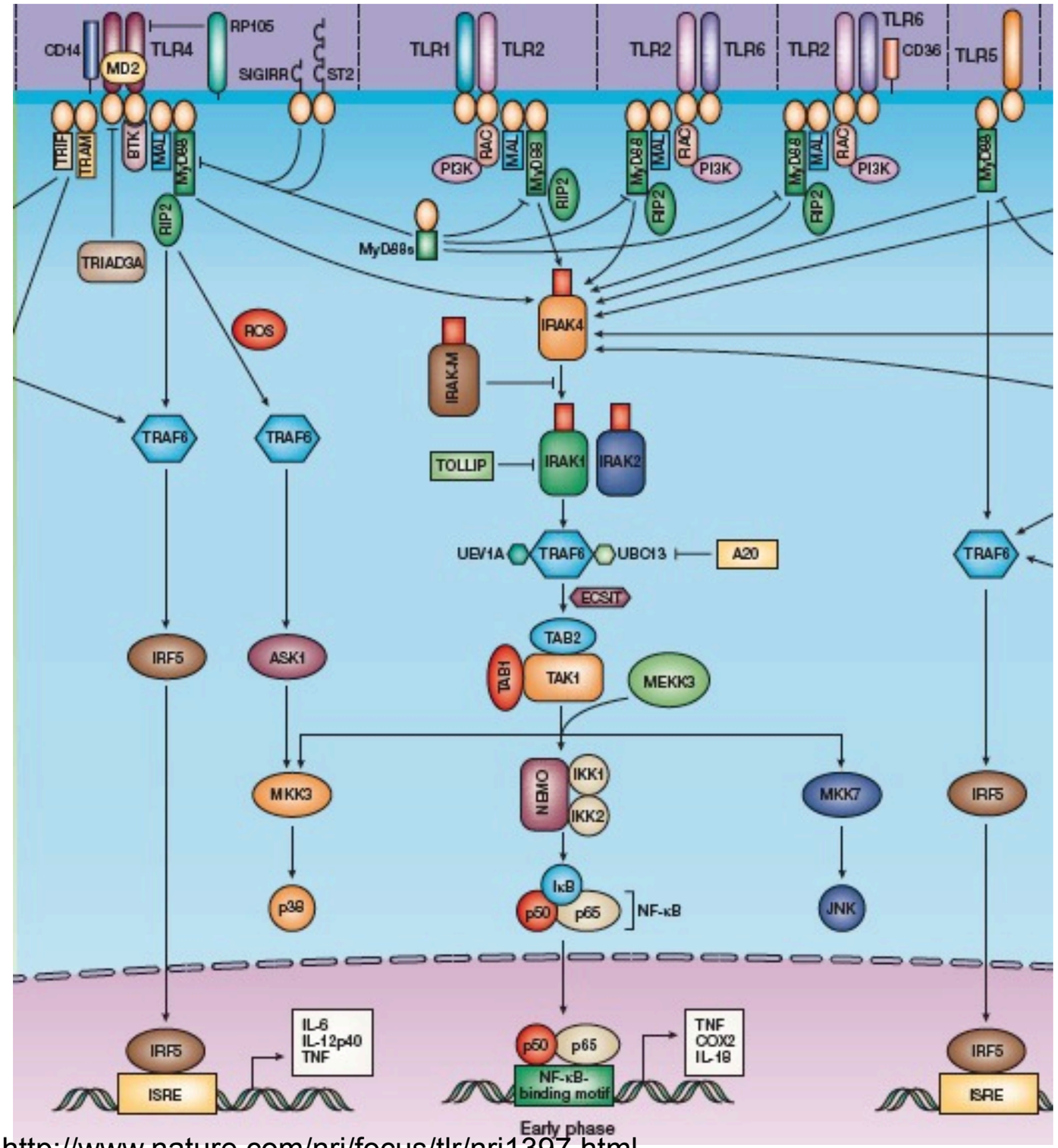


Image credit: Dr.G Weaver, Colorado University at Denver

13

Hand-drawn diagrams

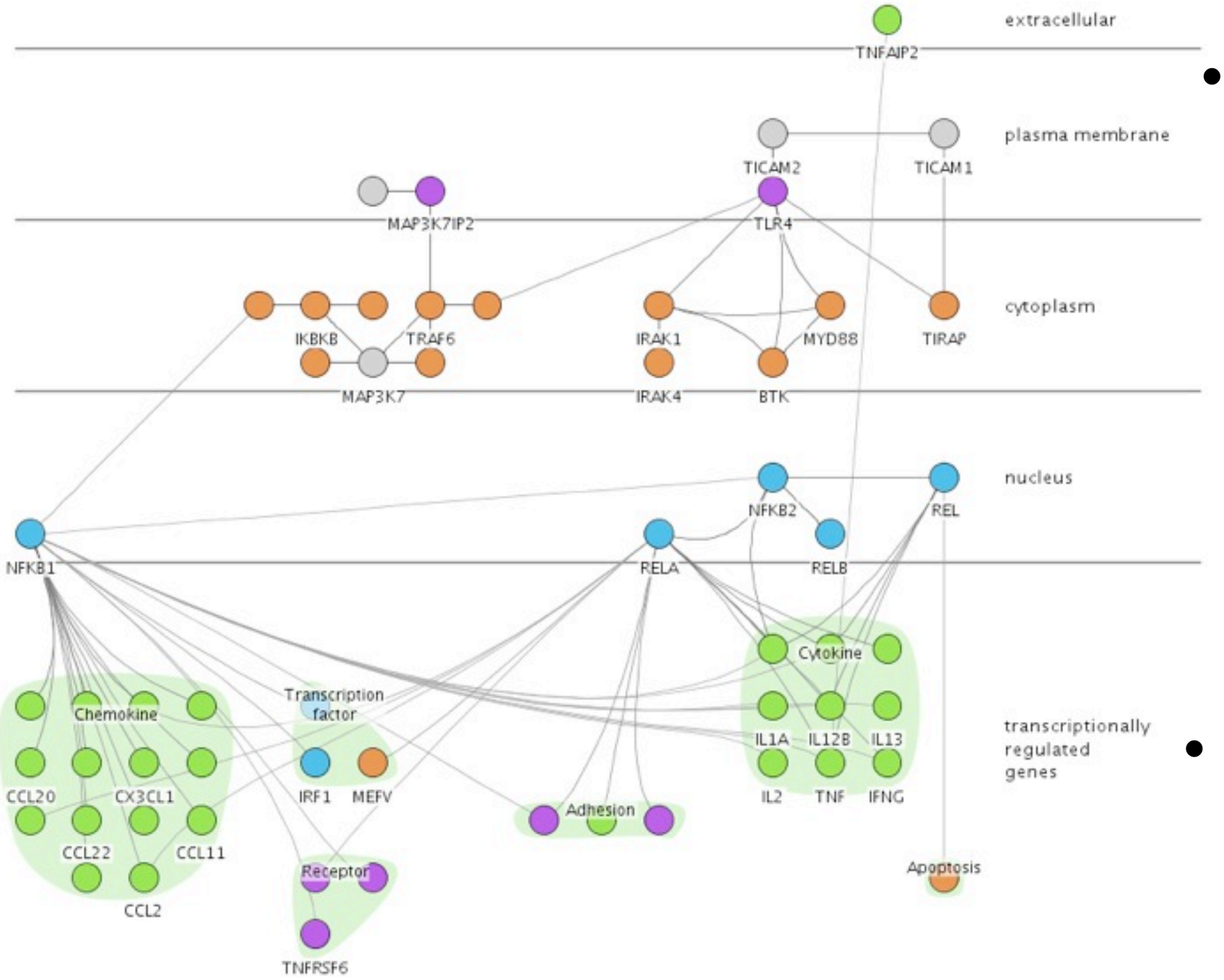


http://www.nature.com/nri/focus/tlr/nri1397.html

- cellular location spatially encoded vertically
- infeasible to create by hand in era of big data

14

Lay out using biological metadata



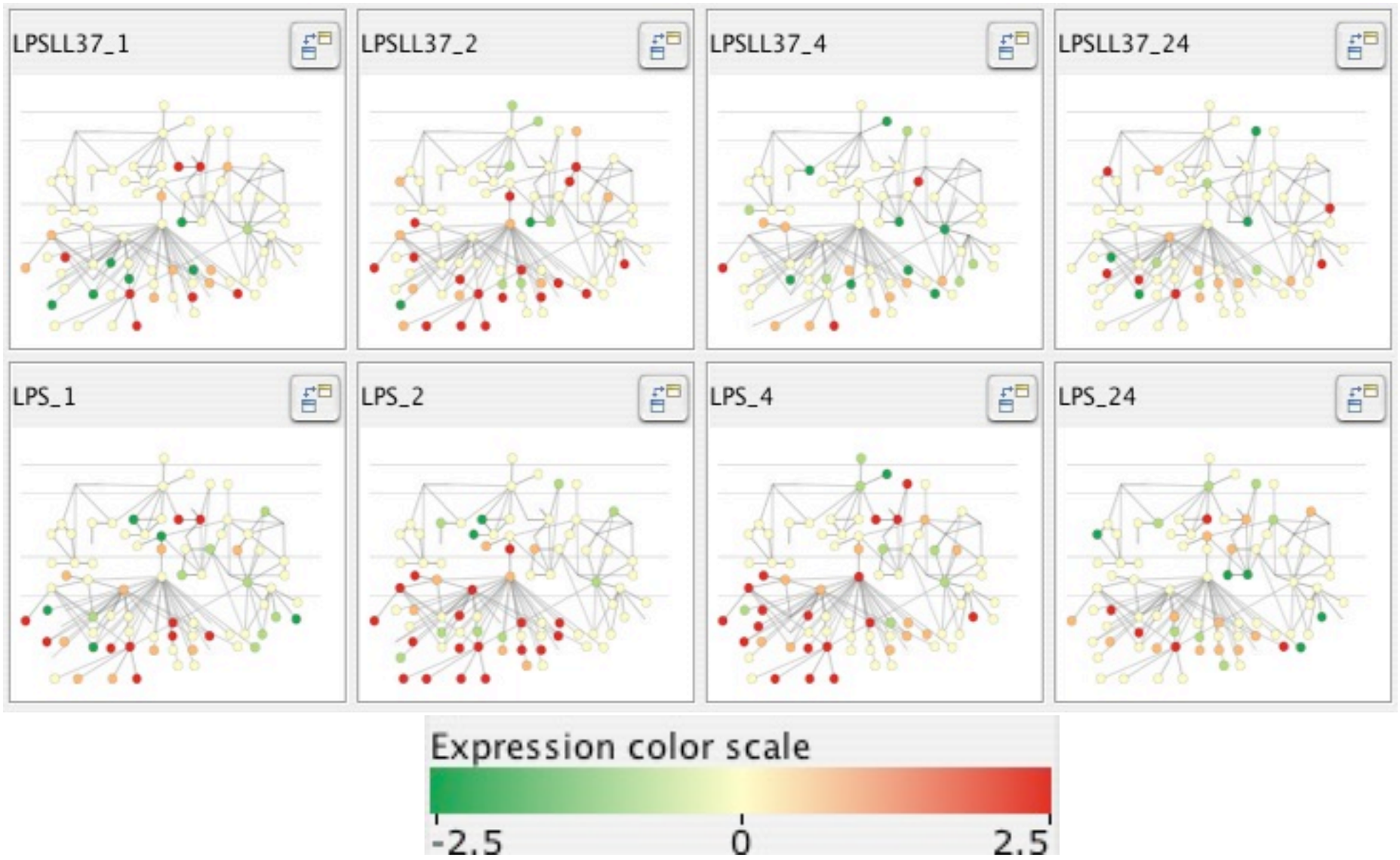
- similar to hand-drawn: spatial position reveals location in cell

- simulated annealing in $O(E\sqrt{V})$ vs. $O(V^3)$ time

15

Choice 2: Use small multiple views

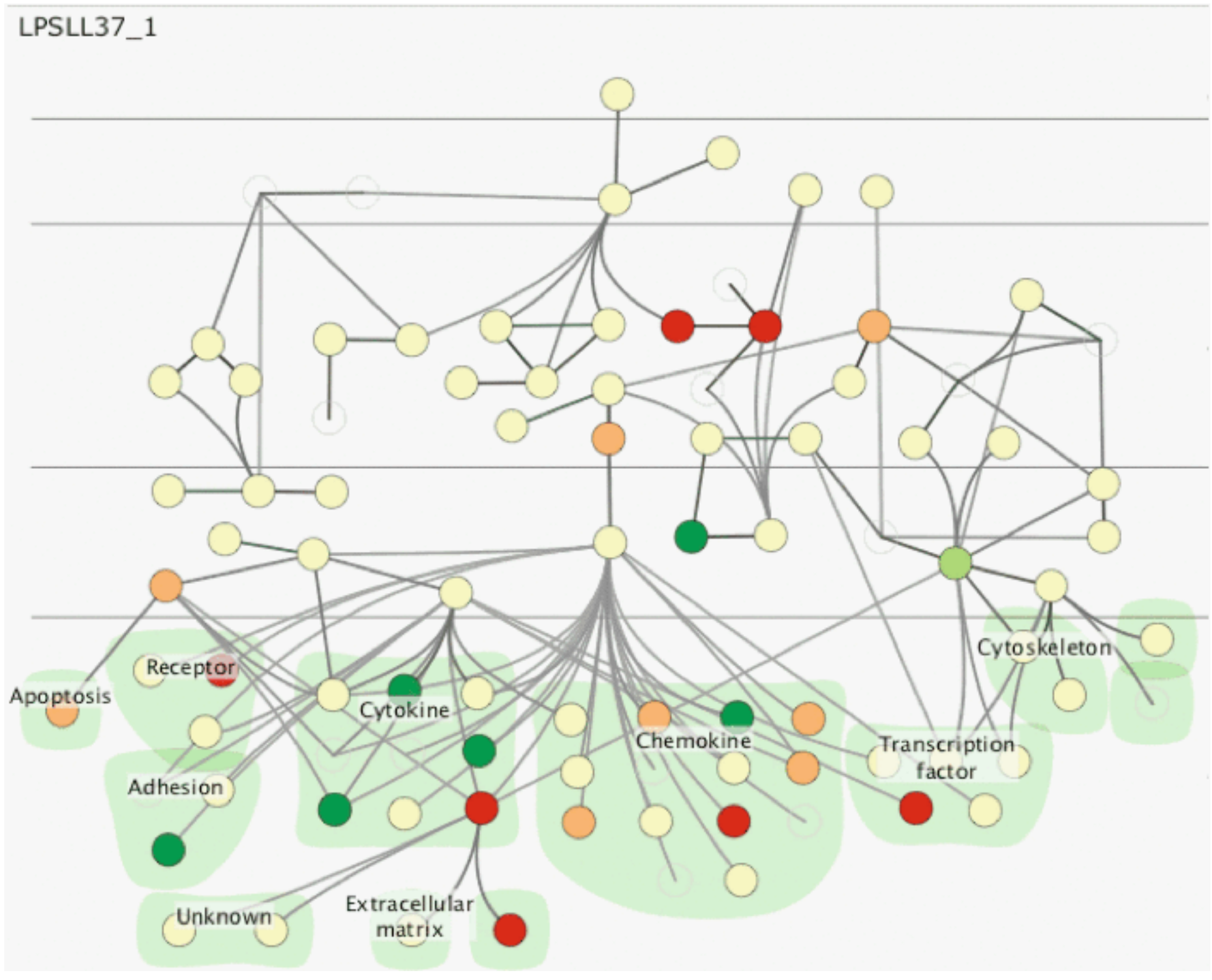
- one graph instance per experimental condition
  - same spatial layout
  - color differently, by condition



16

Why not animation?

- global comparison difficult



17

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.

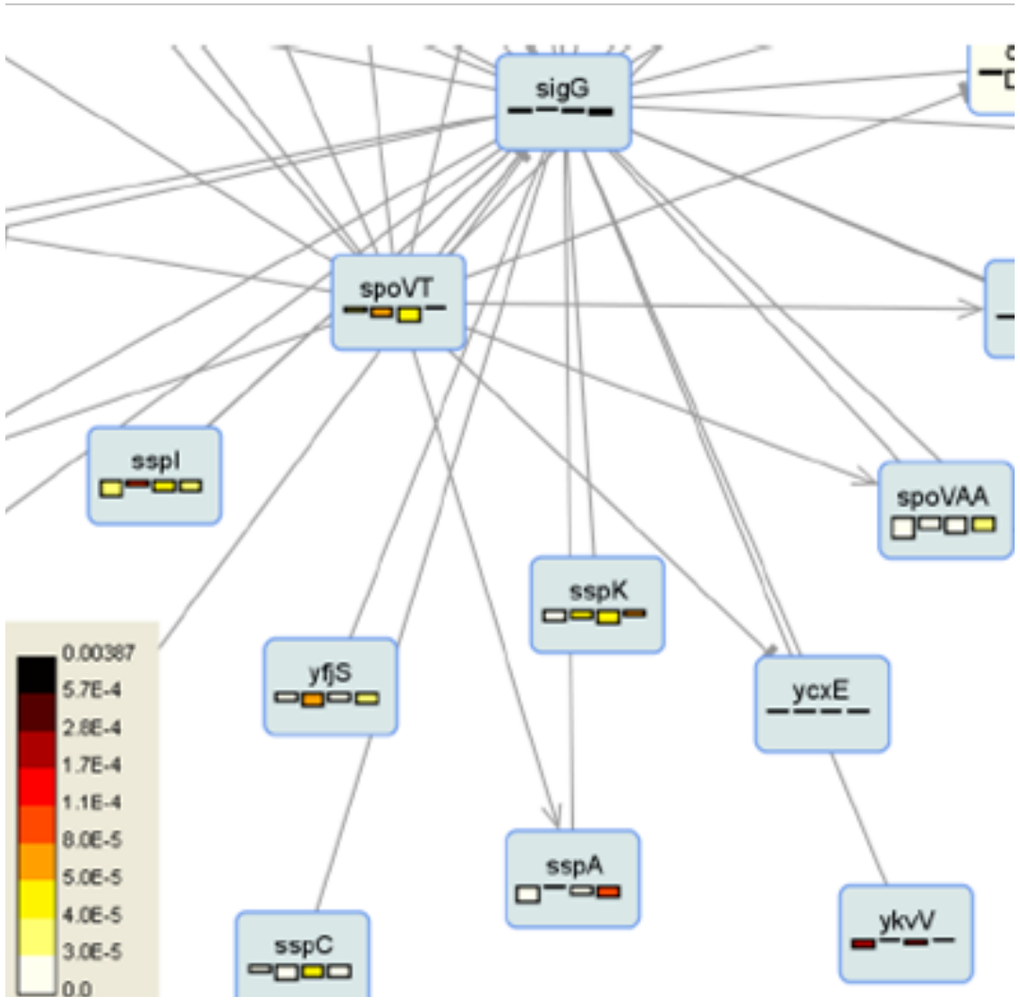
- Animation: can it facilitate? Barbara Tversky, Julie Bauer Morrison, and Mireille Betrancourt. *International Journal of Human-Computer Studies,* 57(4):247-262, 2002.

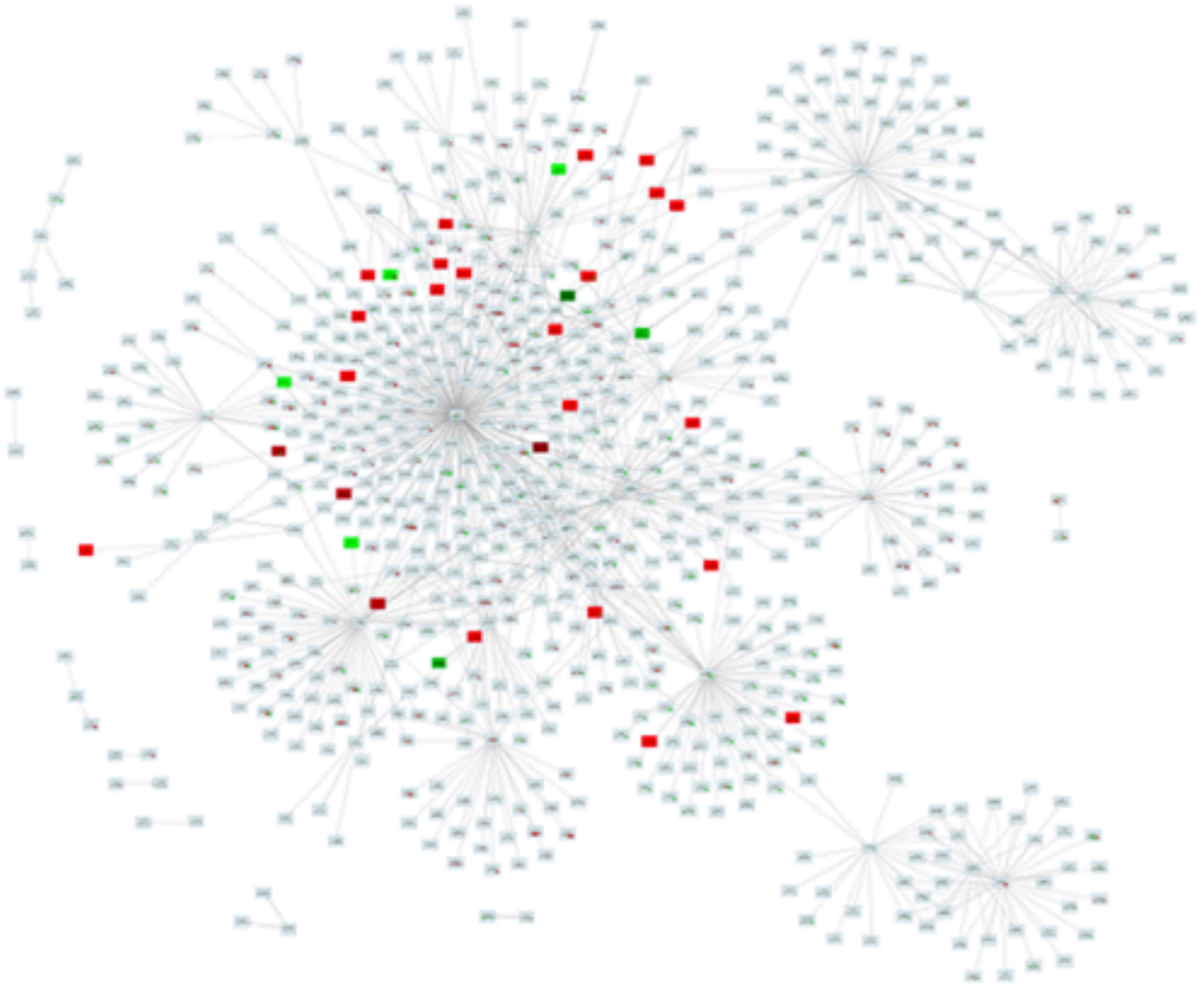
- Effectiveness of Animation in Trend Visualization. George Robertson, Roland Fernandez, Danyel Fisher, Bongshin Lee, John Stasko. IEEE Trans. Visualization and Computer Graphics 14(6):1325-1332 (Proc. InfoVis 08), 2008.

18

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



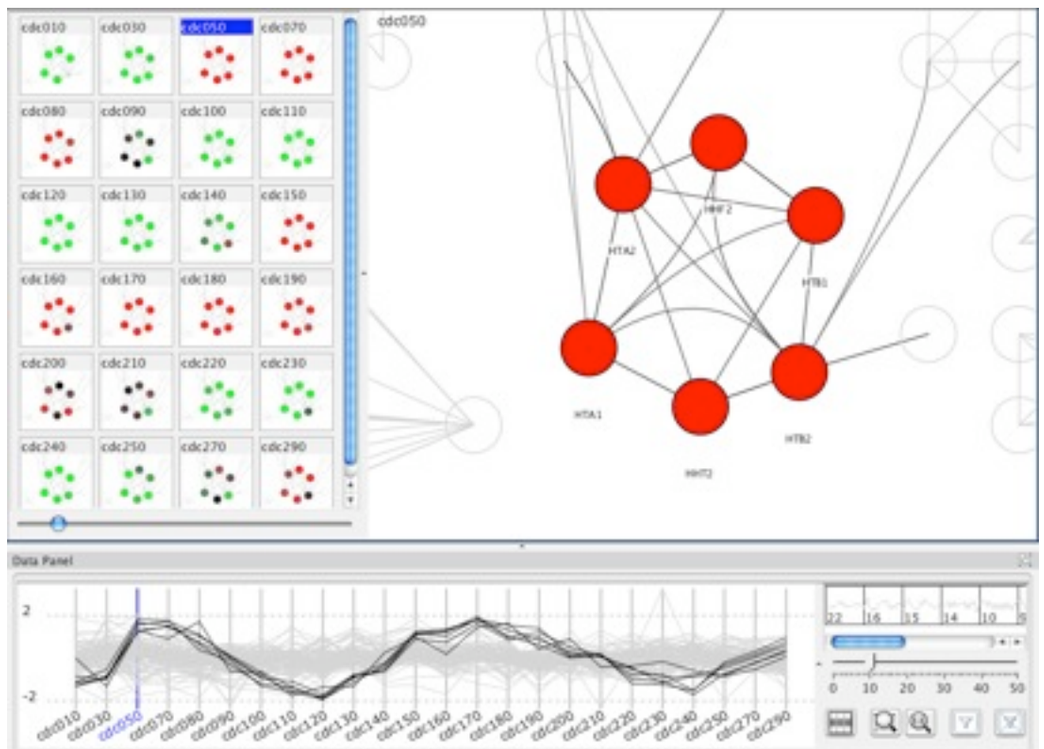


[M. A. Westenberg, S. A. F. T. van Hijum, O. P. Kuipers, J. B. T. M. Roerdink. Visualizing Genome Expression and Regulatory Network Dynamics in Genomic and Metabolic Context. Computer Graphics Forum, 27(3):887-894, 2008.]

19

Choice: Show measures 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**

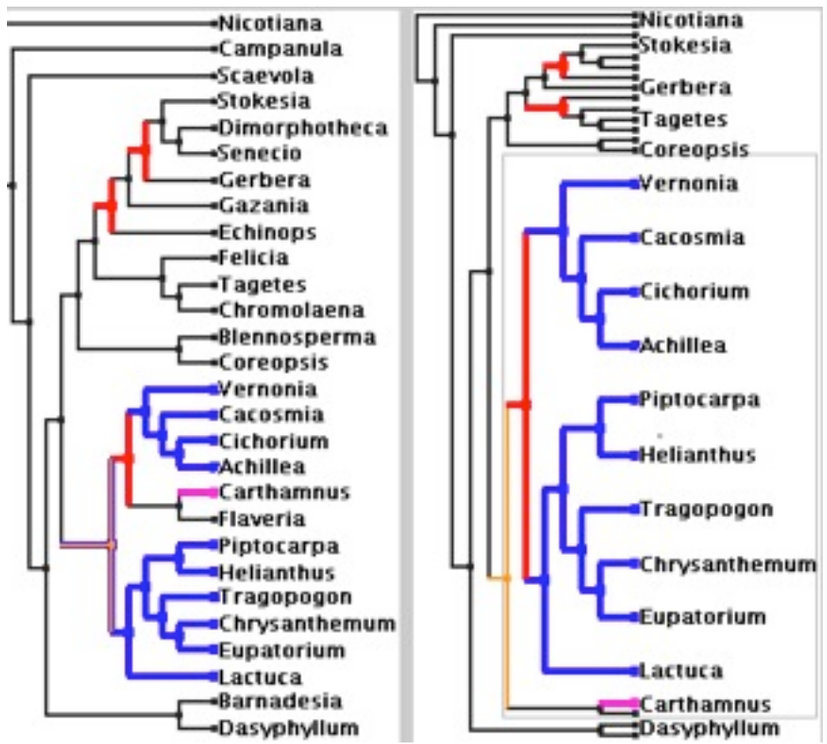


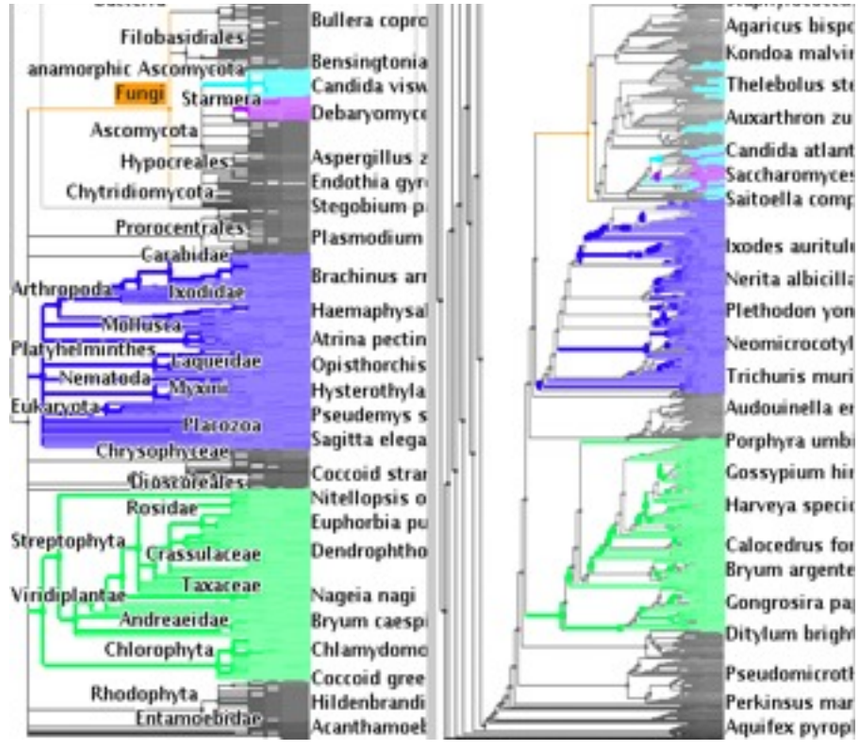
20

Contributions

- Cerebral
  - supports interactive exploration of multiple experimental conditions in graph context
  - provides familiar representation by using biological metadata to guide graph layout
- tool deployment
  - open source, Cytoscape plugin
  - used by target group of collaborators
    - showcased in http://innatedb.ca
  - many more independent adopters
    - 12+ bio lit citations with Cerebral diagrams so far

21





TreeJuxtaposer

Scalable Phylogenetic Tree Comparison

joint work with:

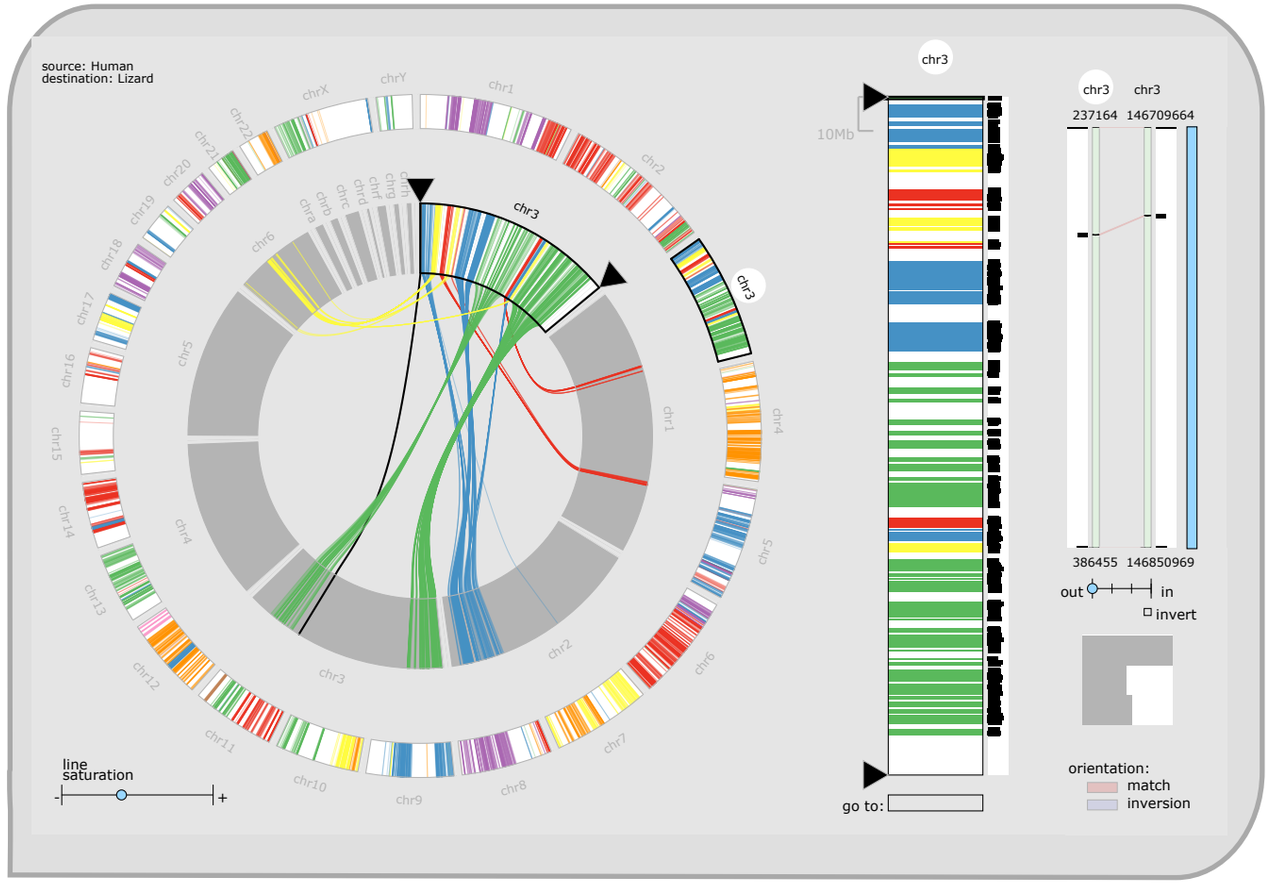
François Guimbretière, Serdar Tasiran, Li Zhang, Yunhong Zhou

http://olduvai.sf.net/tj

TreeJuxtaposer: Scalable Tree Comparison using Focus+Context with Guaranteed Visibility.
Munzner,, Guimbretière, Tasiran, Zhang, Zhou. ACM SIGGRAPH 2003.

22

MizBee



A Browser for Comparative Genomics Data

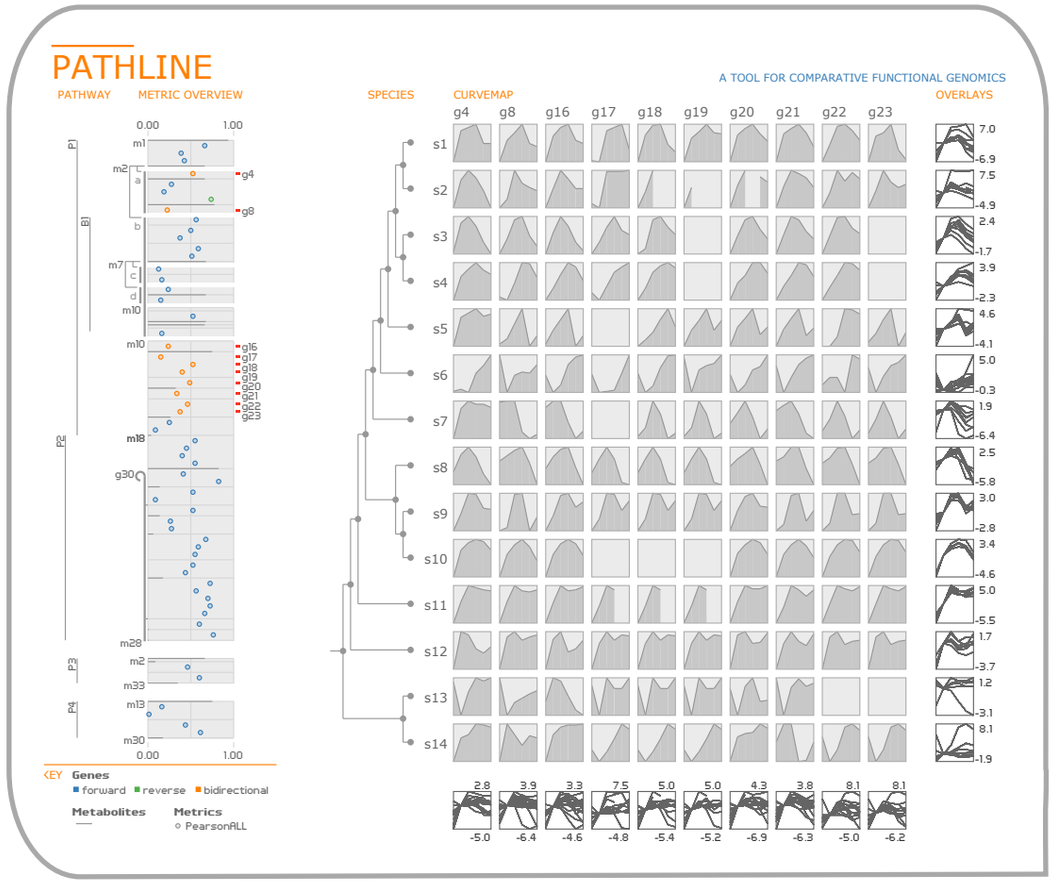
joint work with:
Miriah Meyer, Hanspeter Pfister

http://www.mizbee.org

MizBee: A Multiscale Synteny Browser.
Meyer, Munzner, Pfister, IEEE InfoVis 2009.

23

Pathline



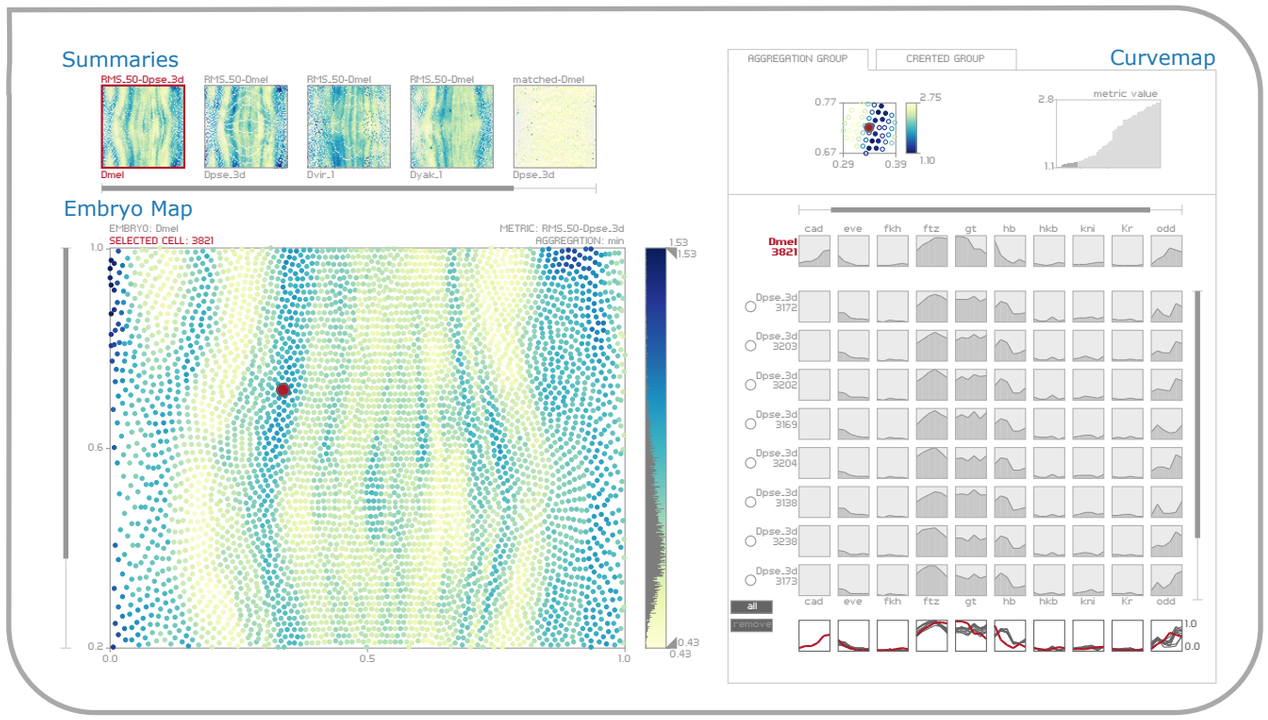
A Tool for Comparative Functional Genomics Data

joint work with:
Miriah Meyer, Bang Wong, Mark Styczynski, Hanspeter Pfister

http://www.pathline.org

Pathline: A Tool for Comparative Functional Genomics
Meyer, Wong, Styczynski, Munzner, Pfister, IEEE/Eurographics EuroVis 2010.

24



MulteeSum

A Tool for Exploring Space-Time Expression Data

joint work with:
Miriah Meyer, Angela DePace, Hanspeter Pfister

http://www.multeesum.org

MulteeSum: A Tool for Comparative Spatial and Temporal Gene Expression Data.
Meyer, Munzner, DePace, Pfister. IEEE InfoVis 2010.

25

More information

- principles in more depth: vis intro book chapter
  http://www.cs.ubc.ca/~tmm/papers.html#akpchapter

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

- this talk
  http://www.cs.ubc.ca/~tmm/talks.html#techtrek12

26