

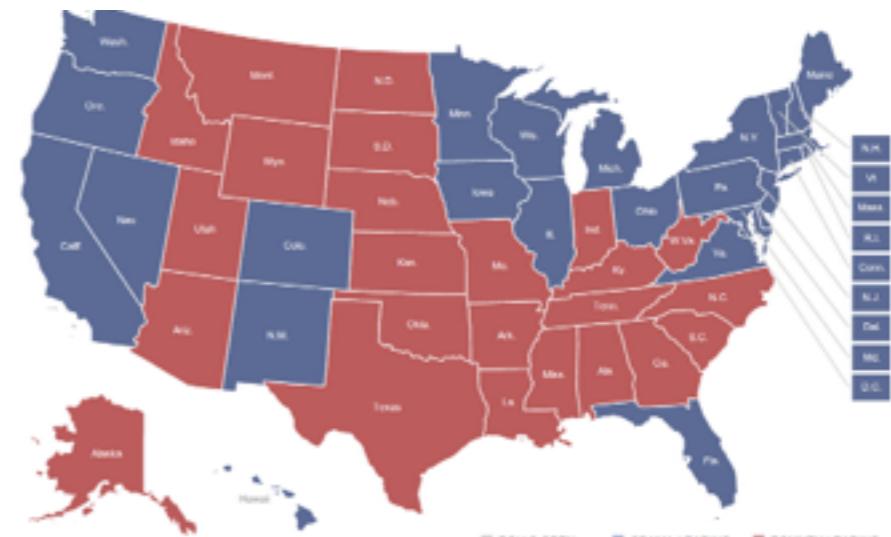
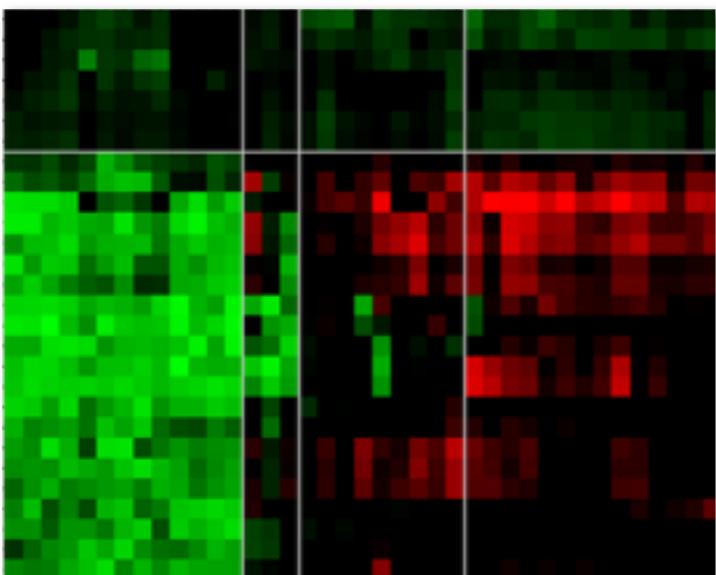
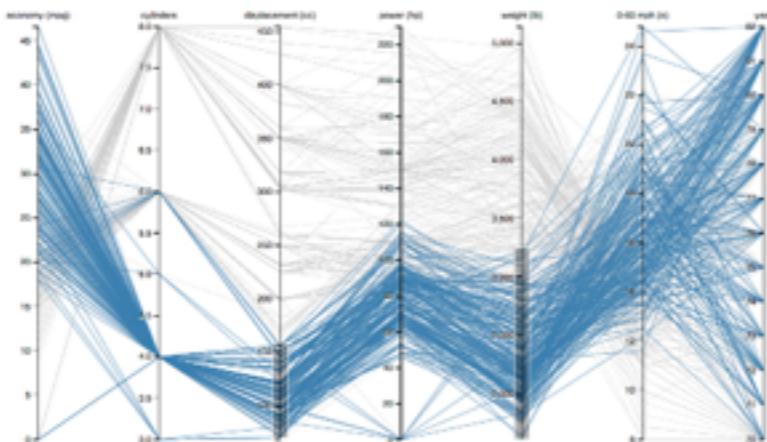
CS 109: Data Science

Big Data Visualization

Marc Streit

mstreit@seas.harvard.edu

On Tuesday



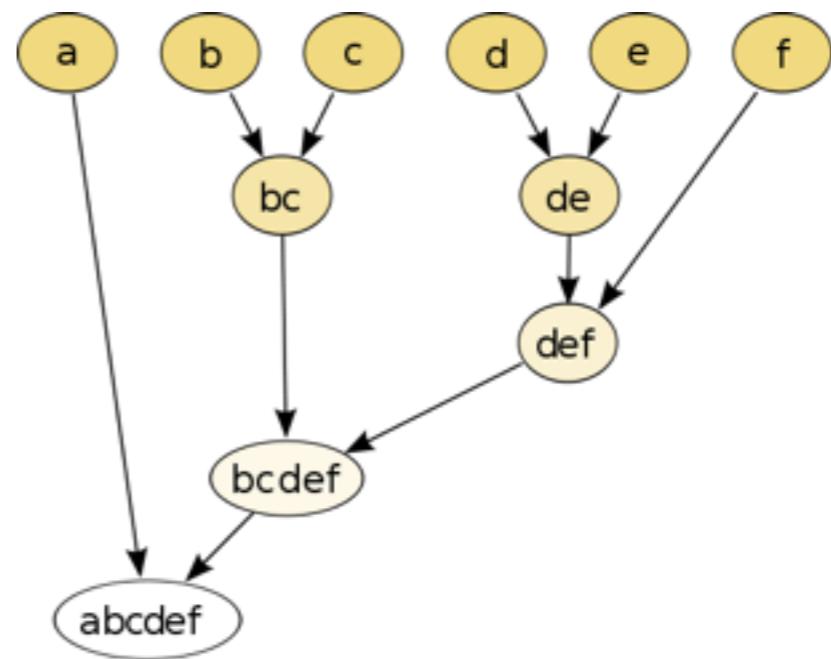
Multi-Dimensional
Data Visualization

Map & Text
Visualization

Today



Big Data
Visualization



Clustering,
Dimensionality
Reduction

Big Data Visualization



Big Data?

**“Big Data is like teenage sex:
everyone talks about it,
nobody really knows how to do it,
everyone thinks everyone else is doing it,
so everyone claims they are doing it.”**

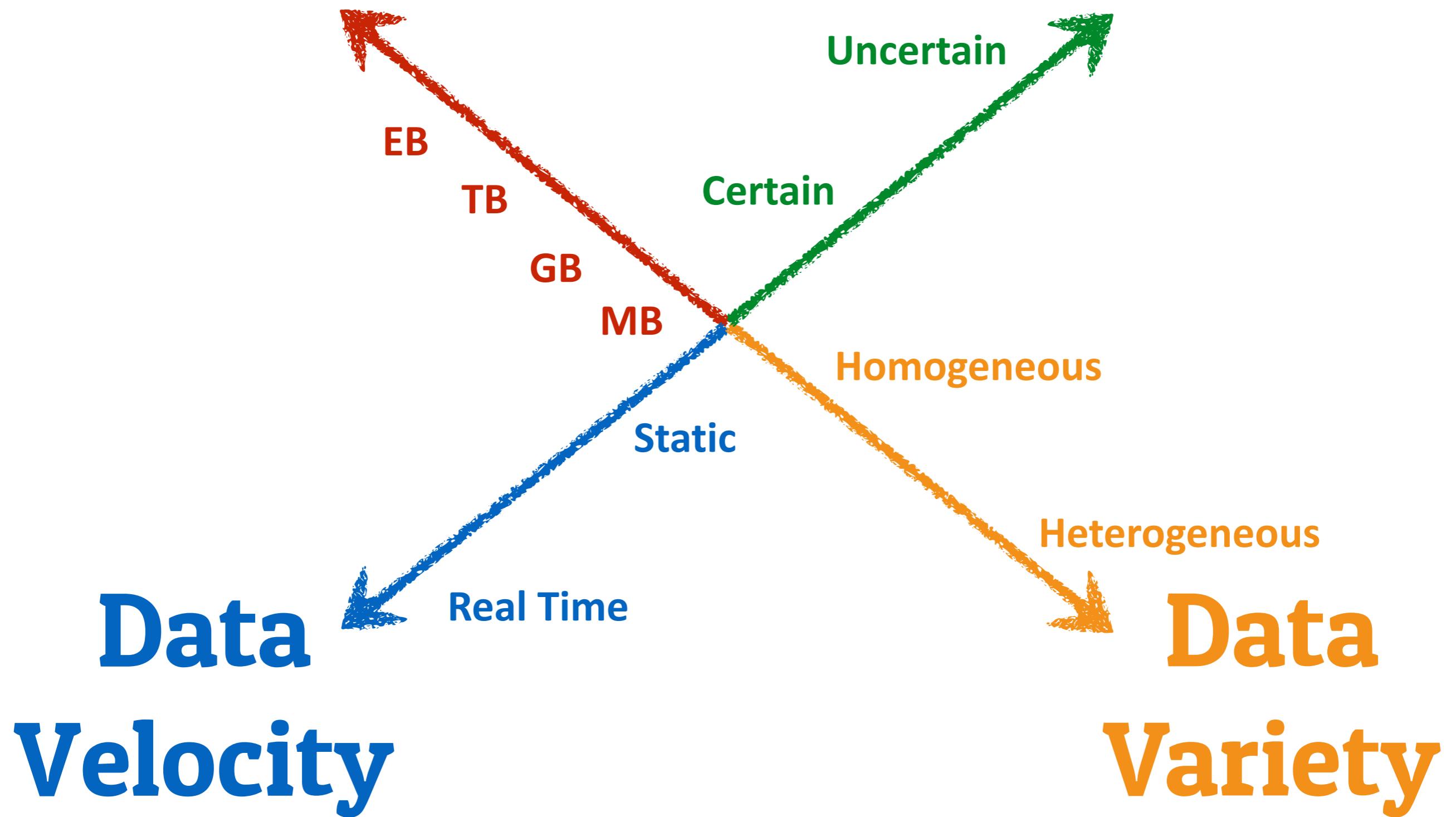
-Dan Ariely

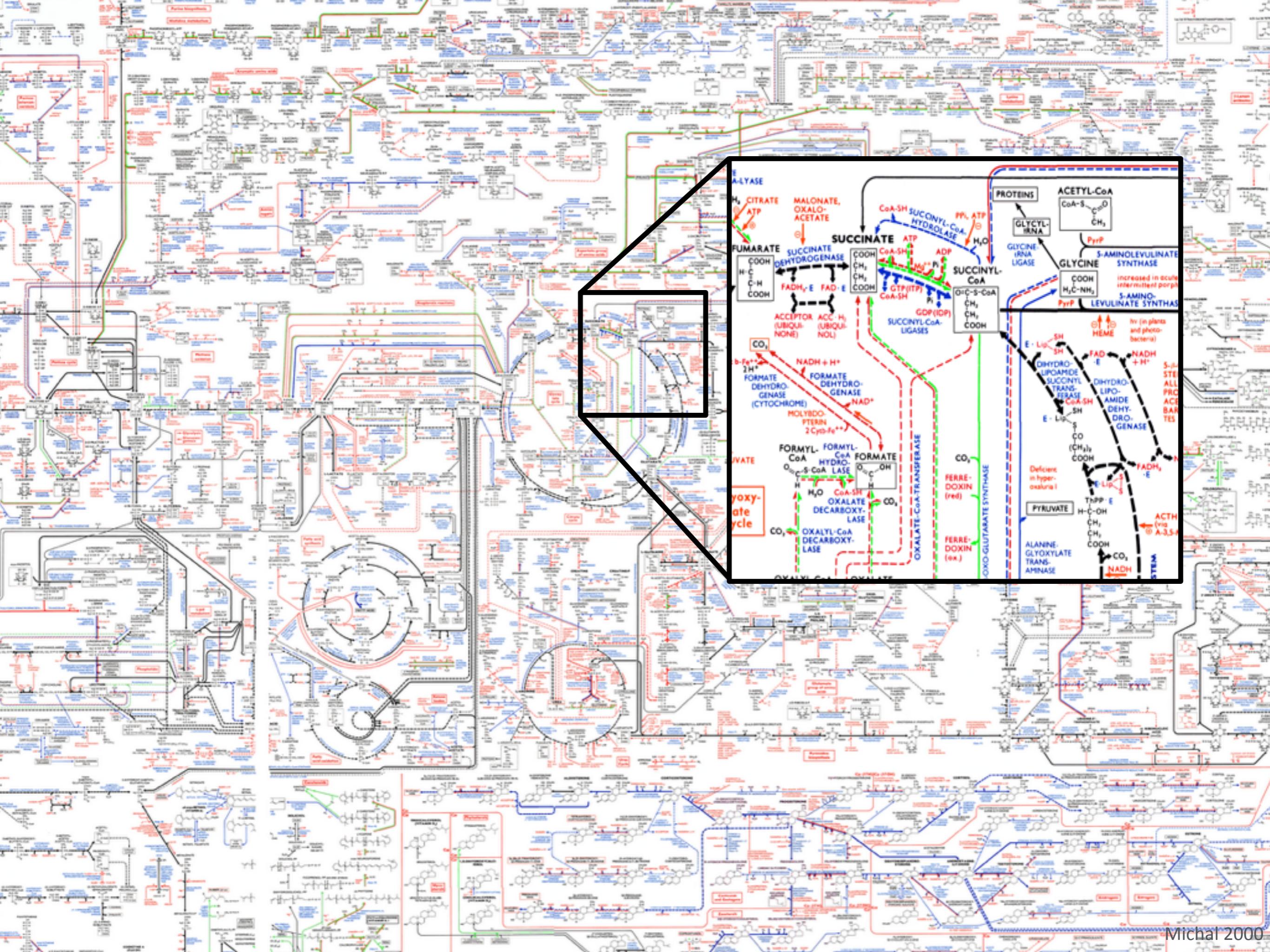
4 Vs
of
Big Data

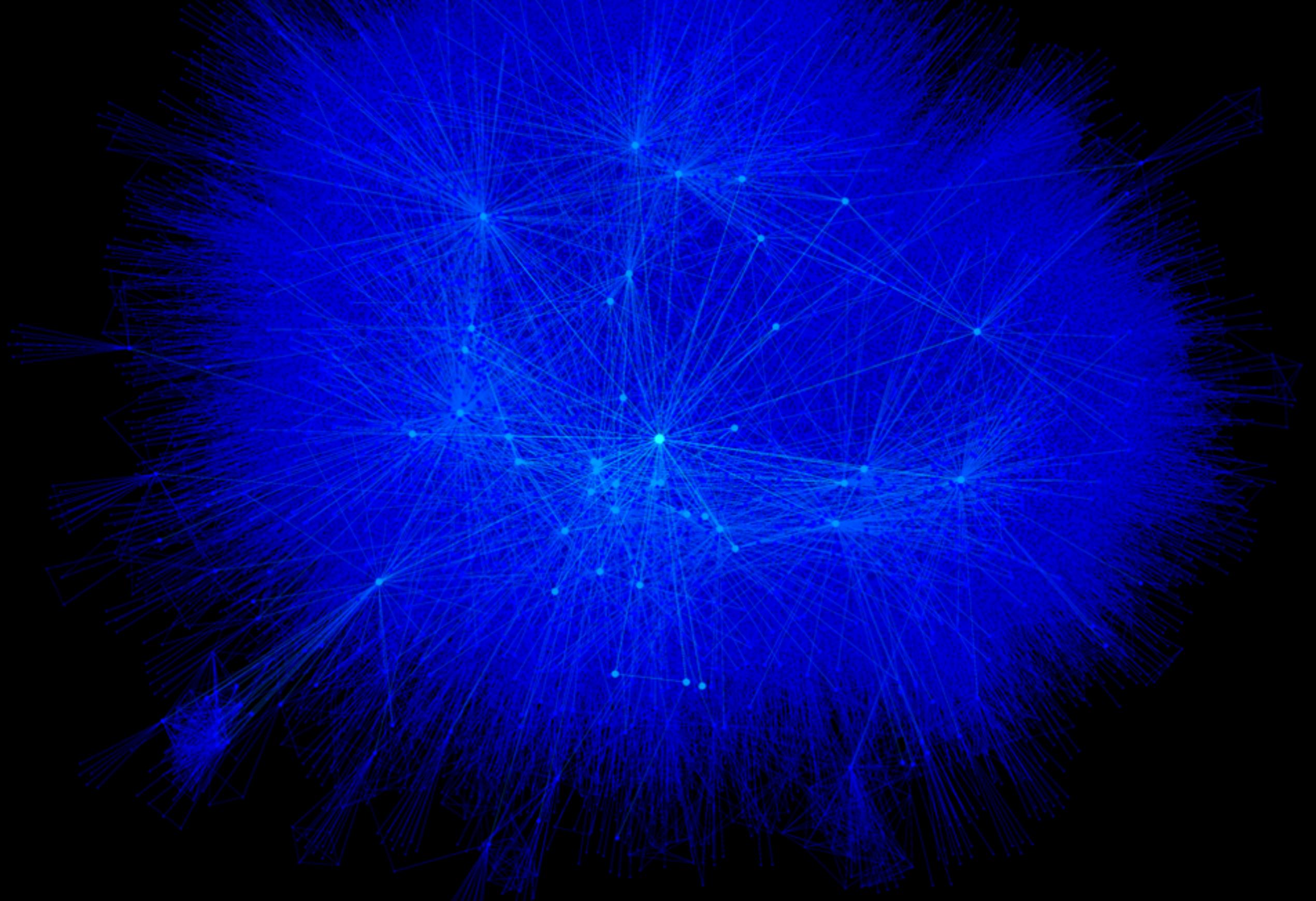
as defined by Gartner Group

Data Volume

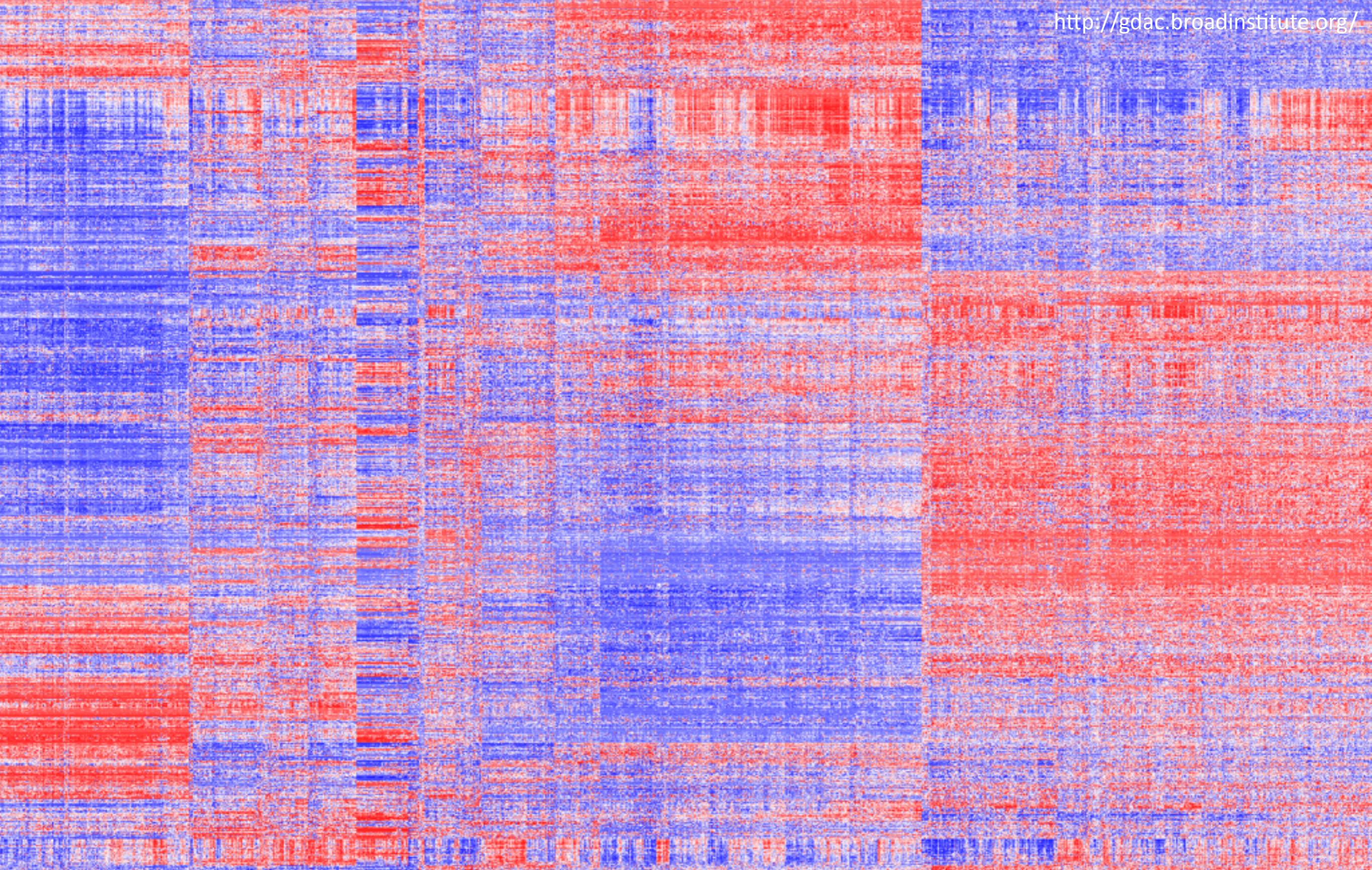
Data Veracity





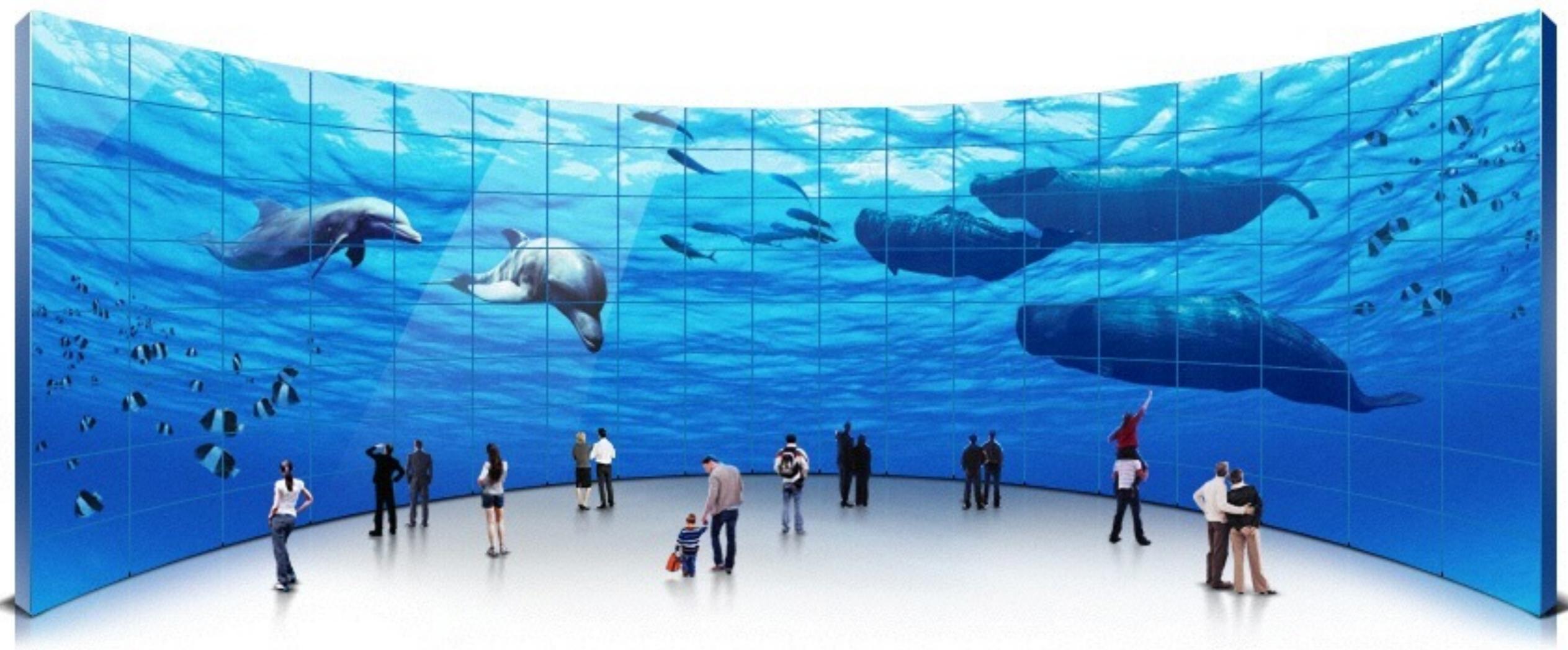


Giant Hairball



#data_points > #pixels

Get more pixels?!



Samsung - 250 large format displays

What else can we do?

Get even more pixels? ...well, not really.

Better: Don't show all information

Pragmatic

Natural display limitations: size, weight, cost

Technological

Resolution of eye better than of display

Human

Not able to perceive all information at once

Ways to deal with too much information

Temporal Partitioning

Navigation: Pan, Rotate

Geometric/Semantic Zooming

Abstraction

Spatial Partitioning

Multiple Coordinated Views

Overview + Detail

Temporal Partitioning: Panning

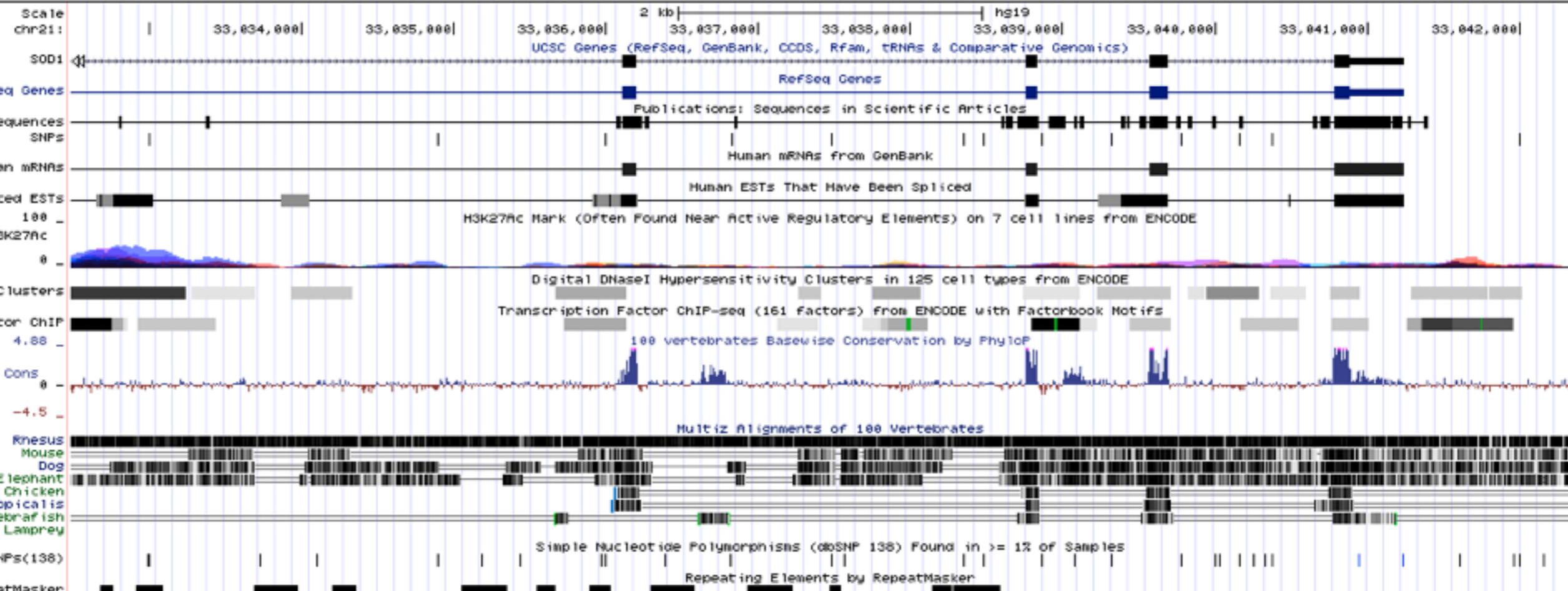
UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly

move <<< << < > >> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x 100x

chr21:33,032,483-33,042,456 9,974 bp. enter position, gene symbol or search terms

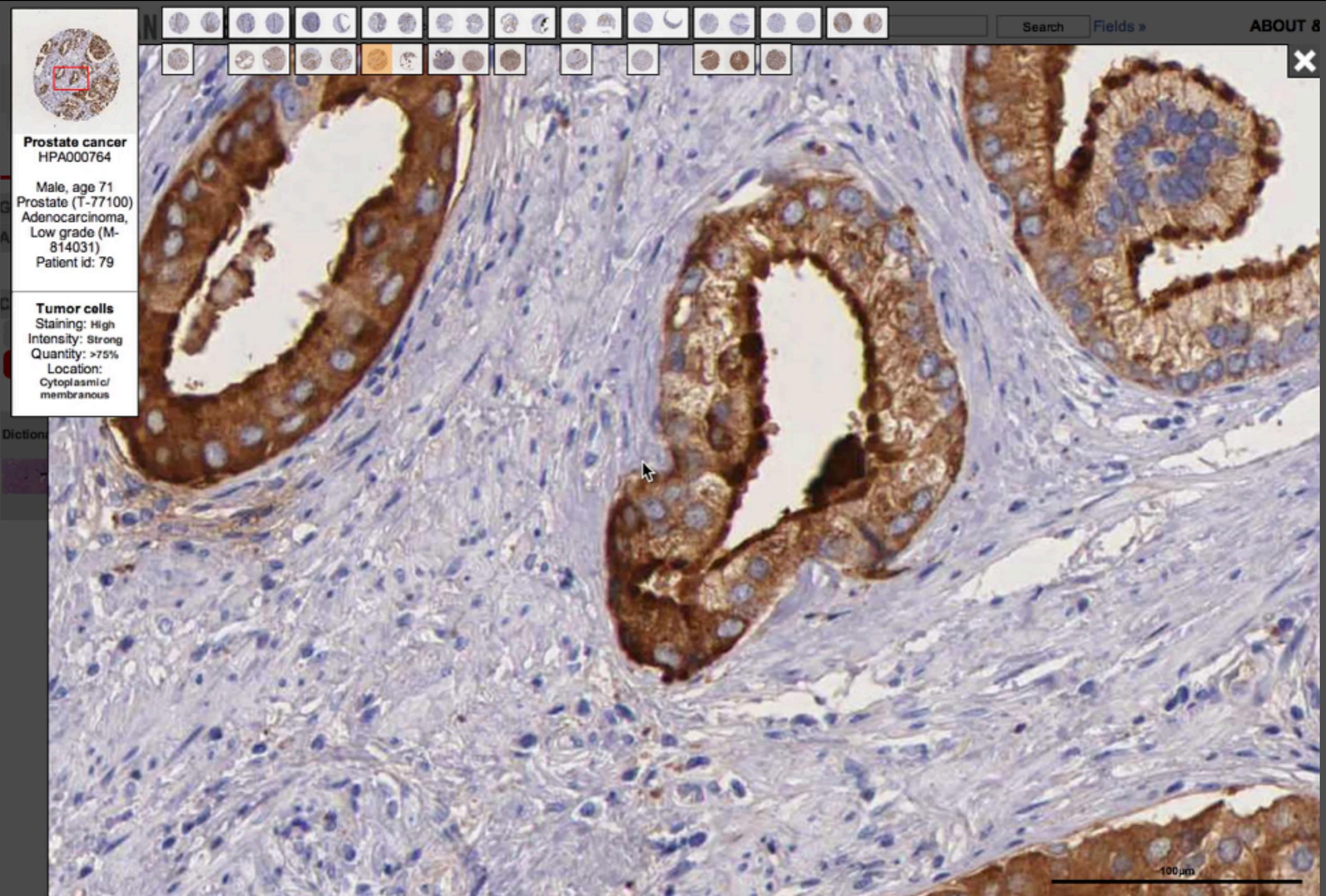
go

chr21 (q22.11) 21p13 21p12 21p11.2 q11.2 21q21.1 21q21.2 21q21.3 21q22.11 q22.12q22.13 21q22.2 21q22.3



Example: Human Protein Atlas Project

<http://www.proteinatlas.org/>

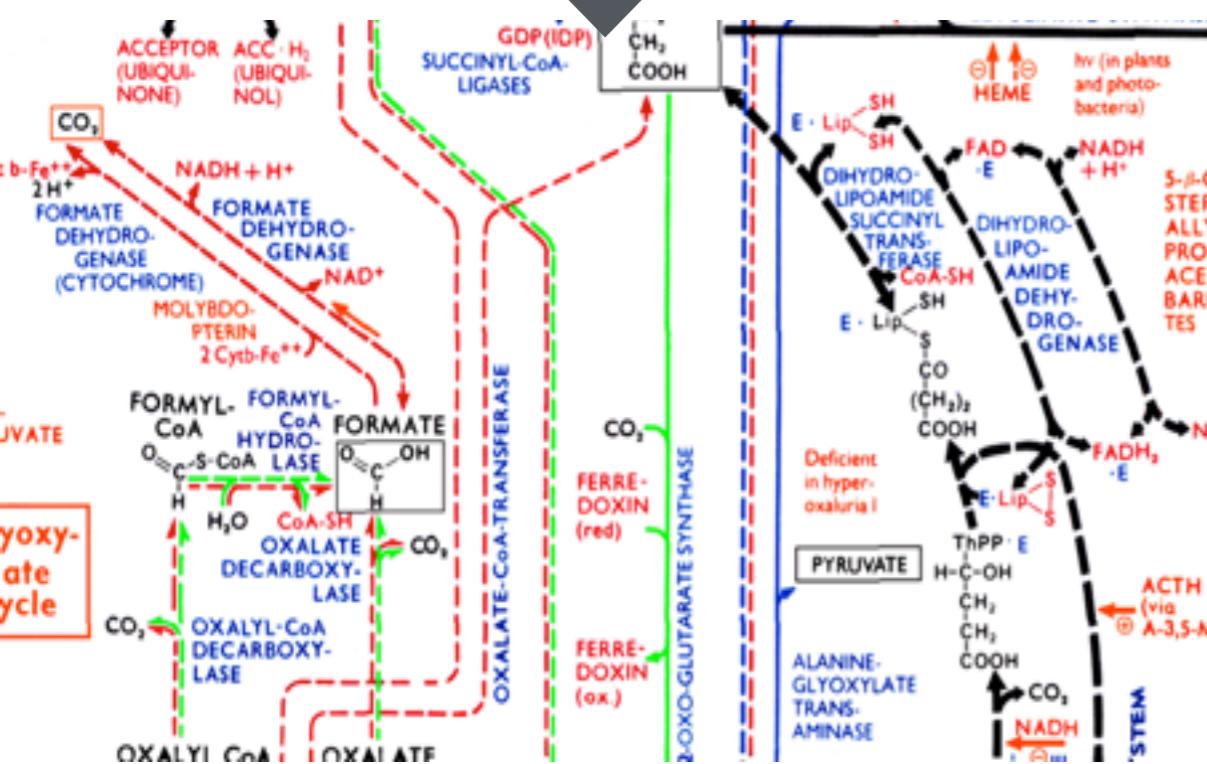
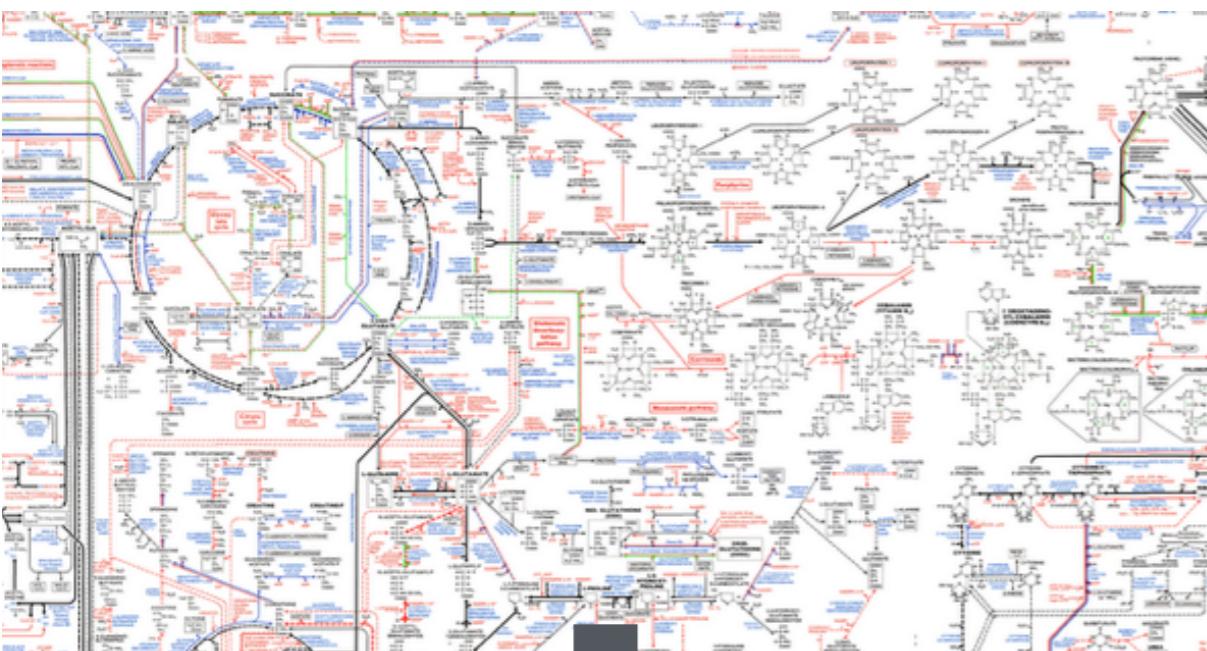


Temporal Partitioning: Rotation

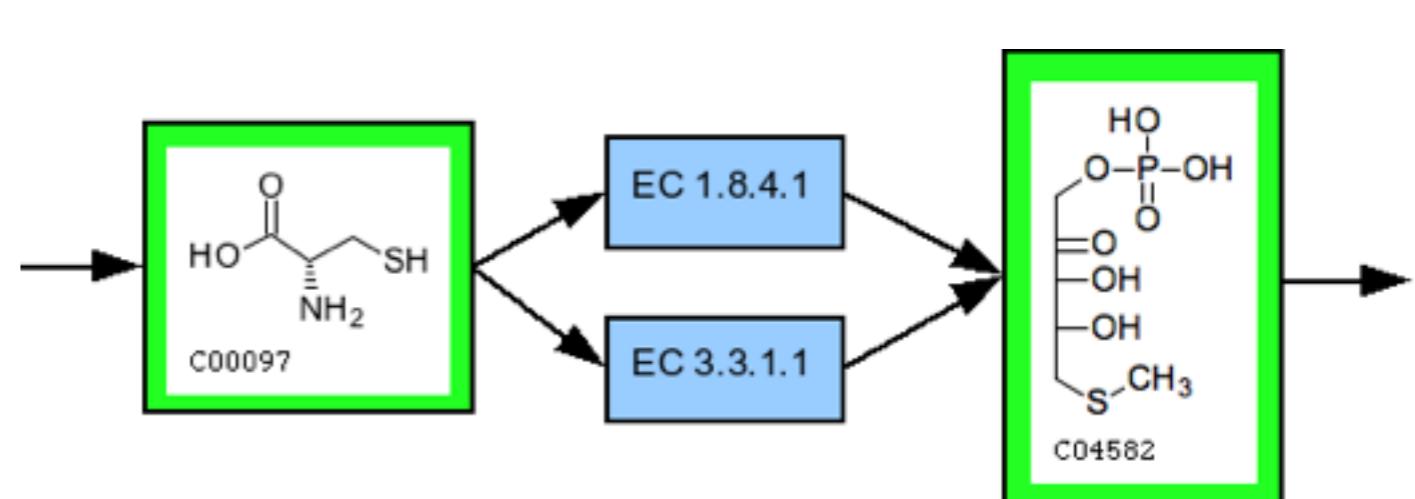


Temporal Partitioning: Zooming

Geometric vs. semantic zooming

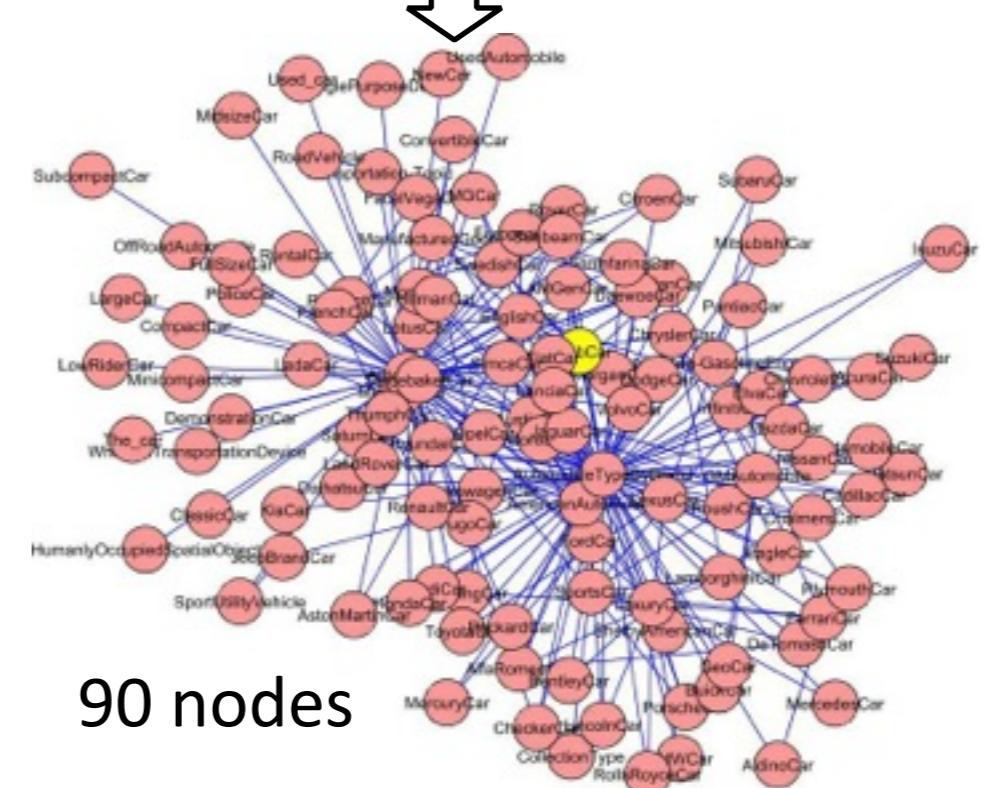
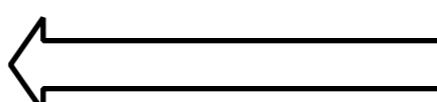
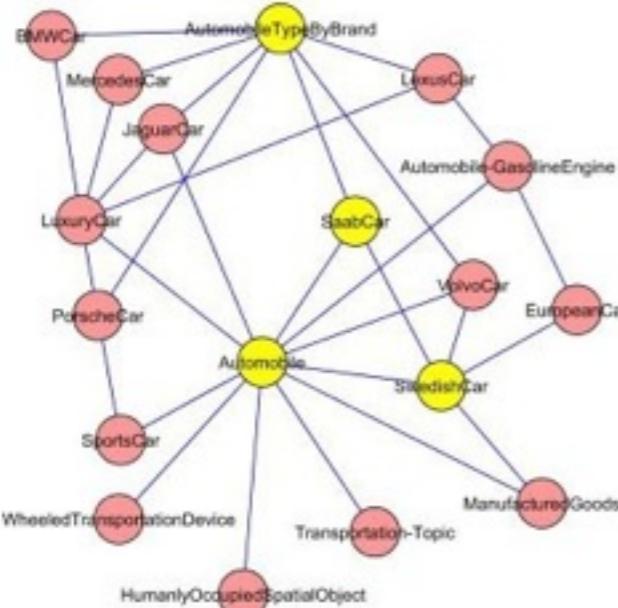
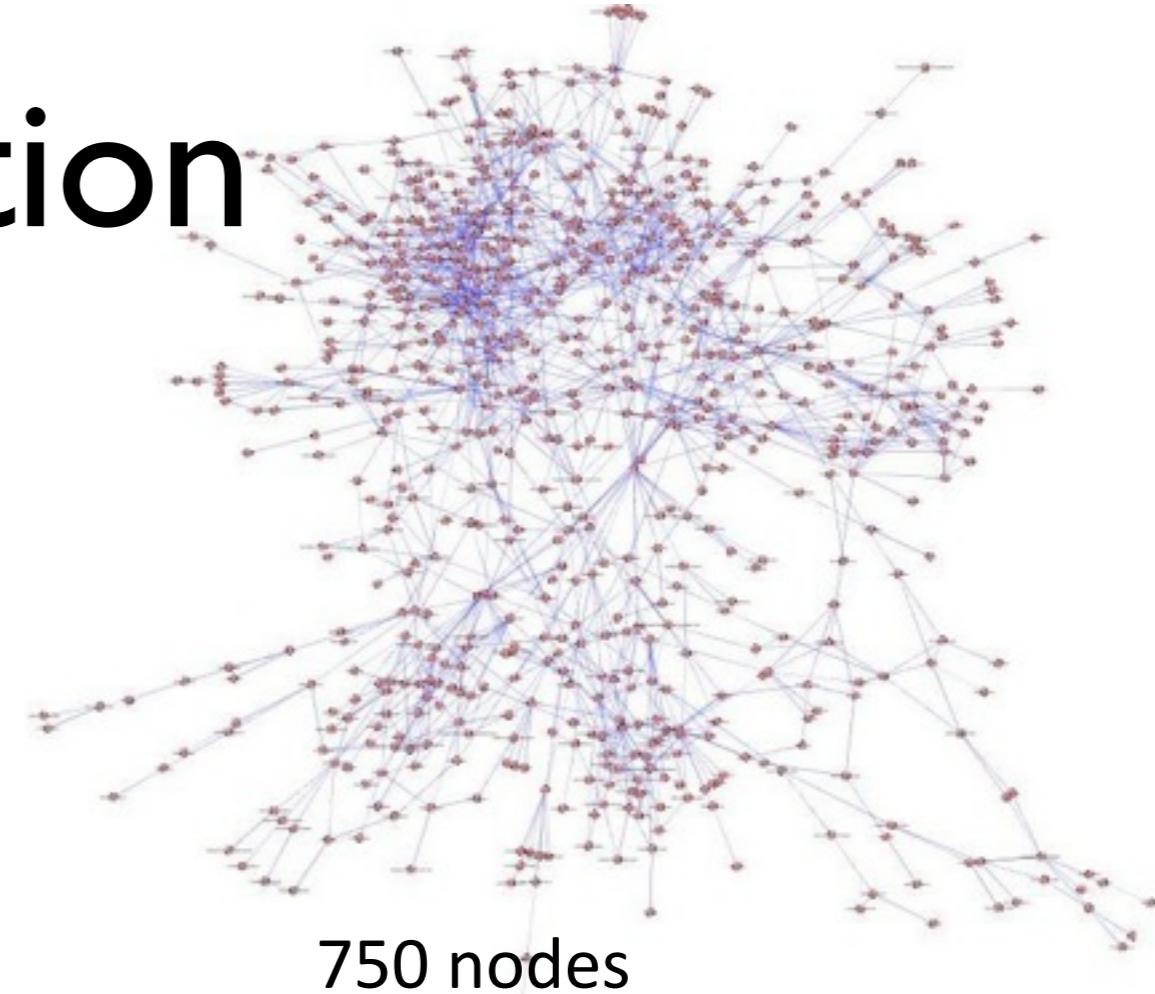
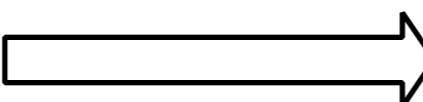
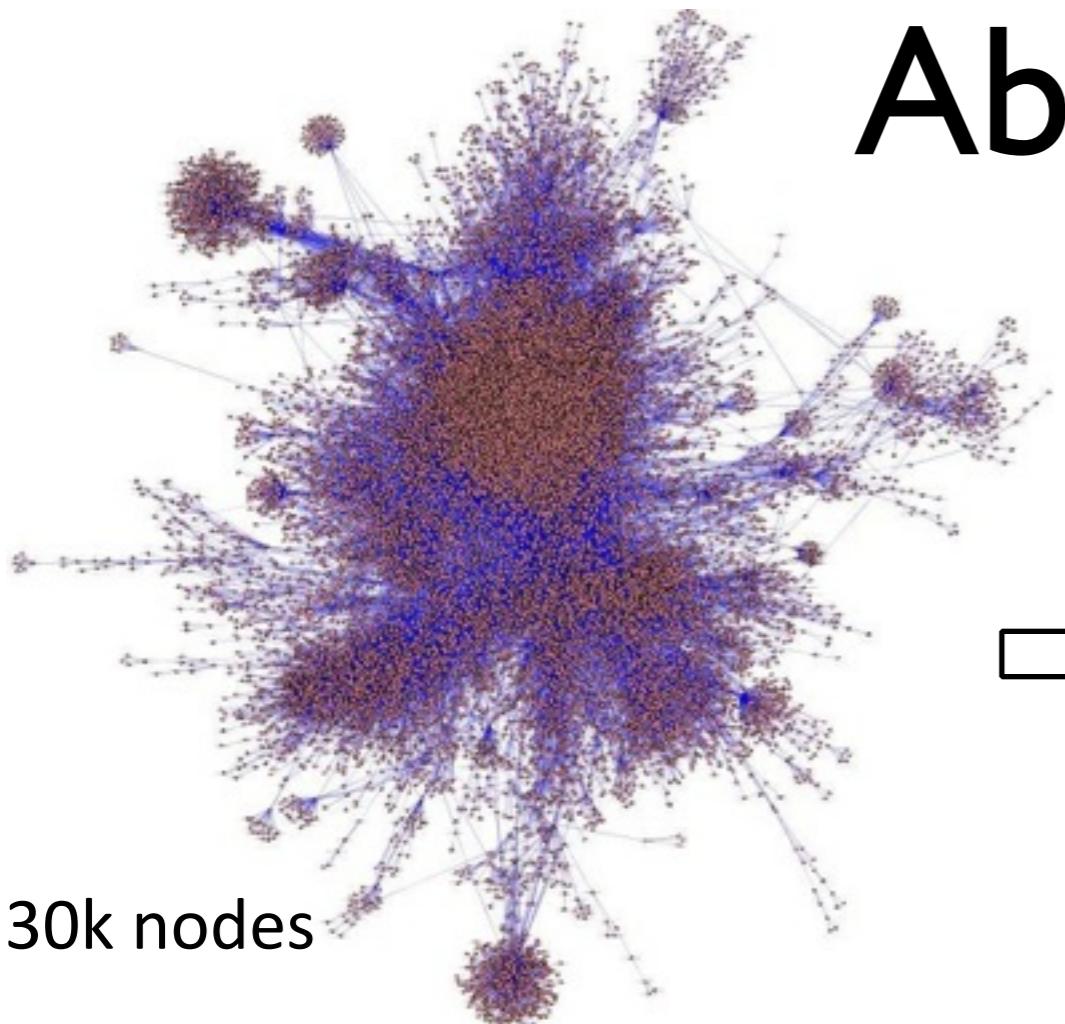


Geometric Zooming



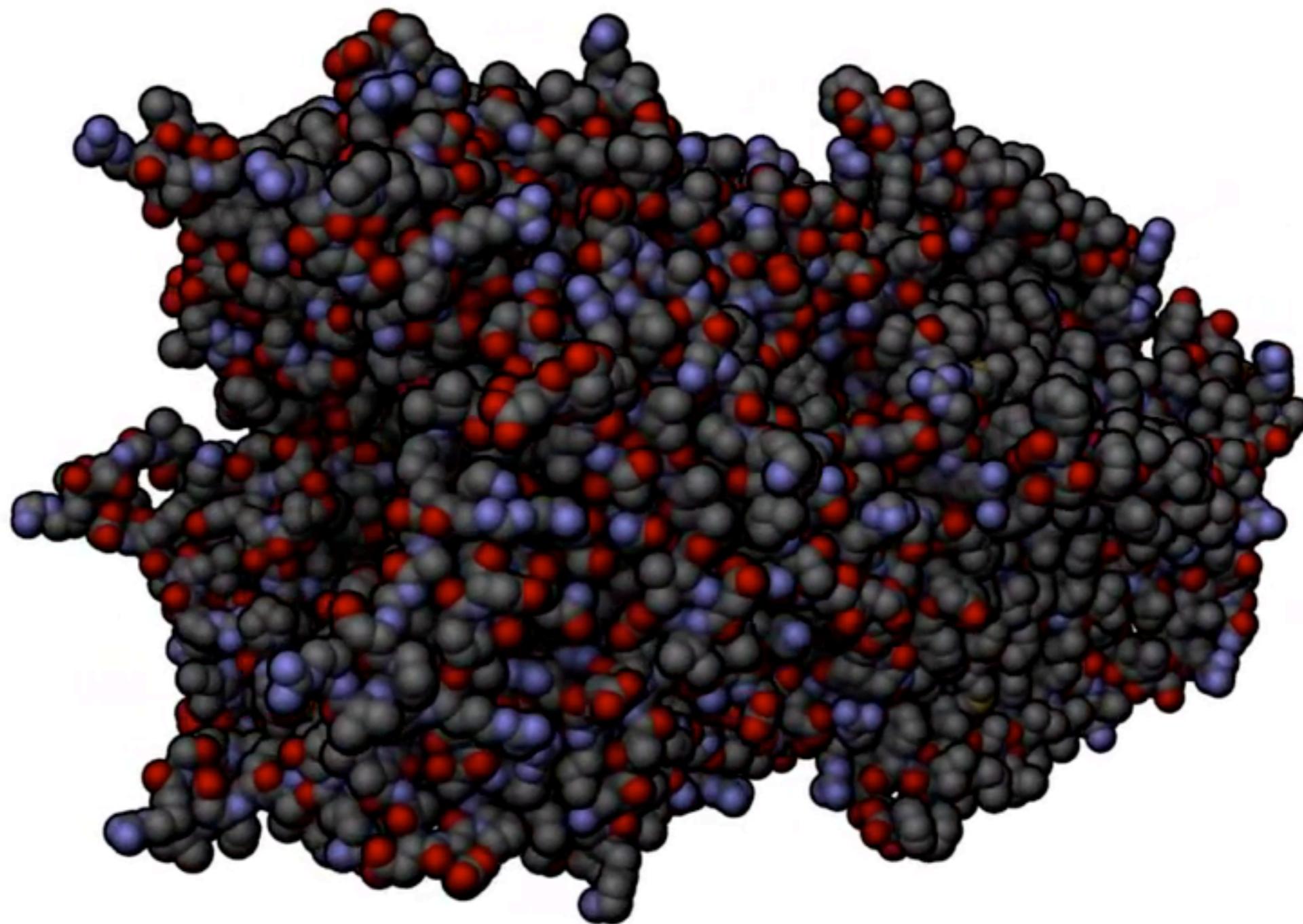
Semantic Zooming

Abstraction

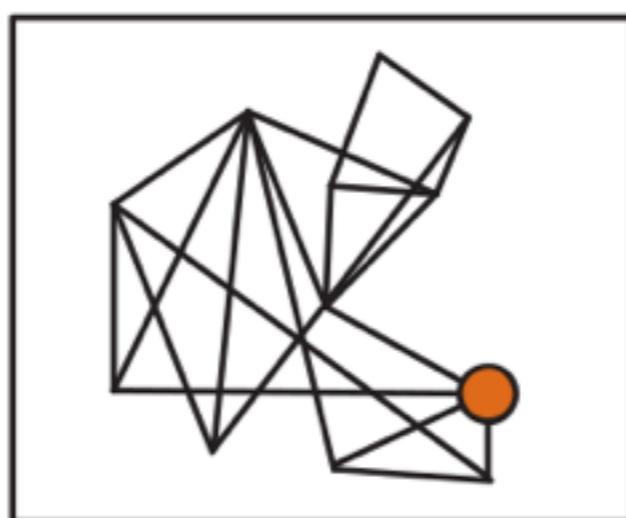
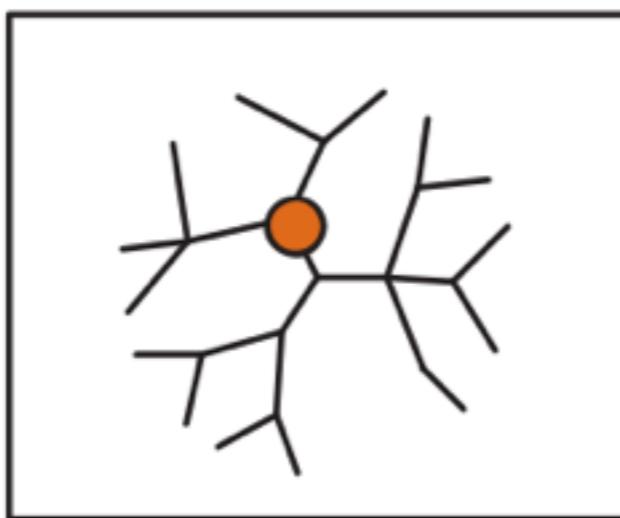


Example:

Continuous Abstraction



Spatial Partitioning: Multiple Coordinated Views (MCV)



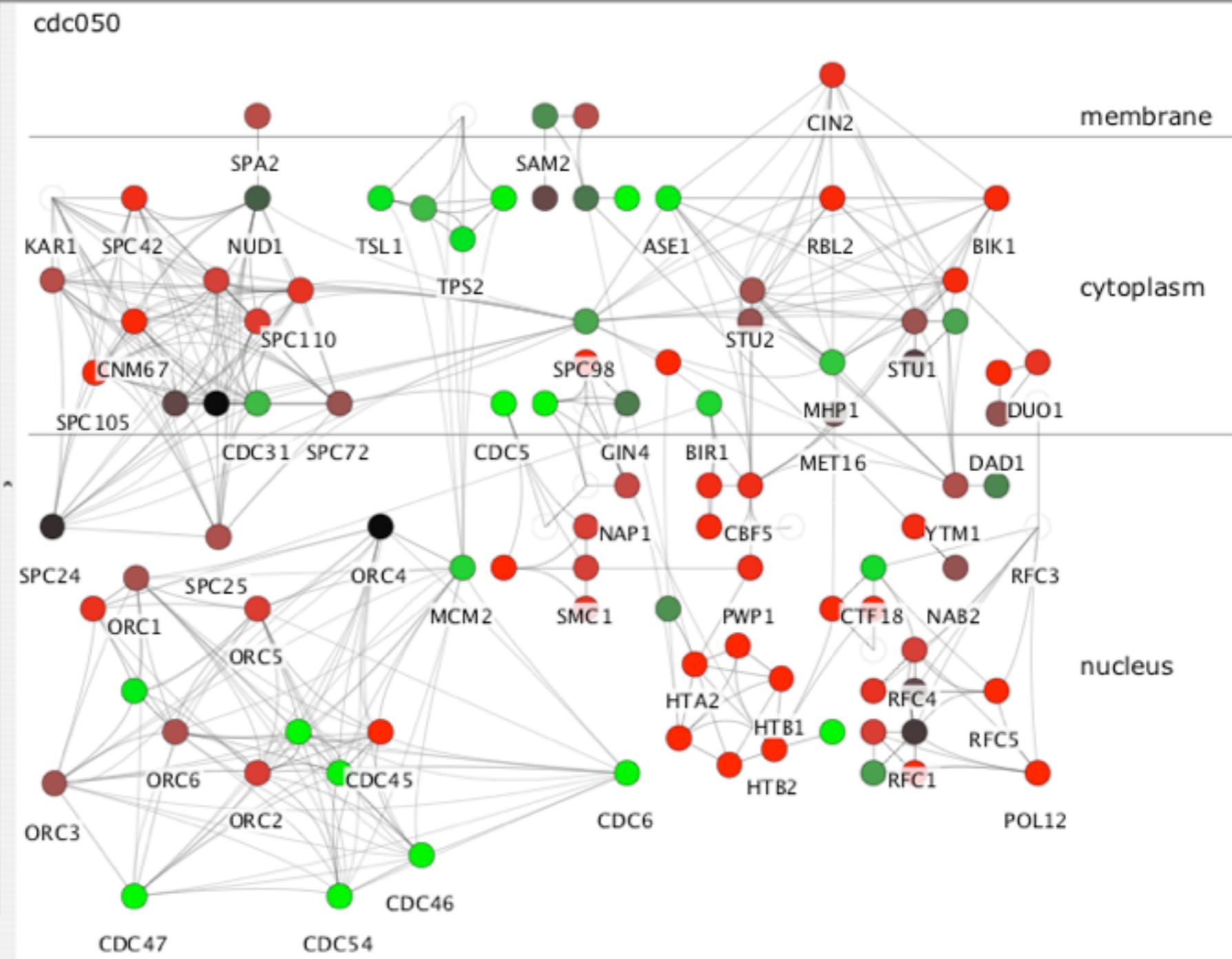
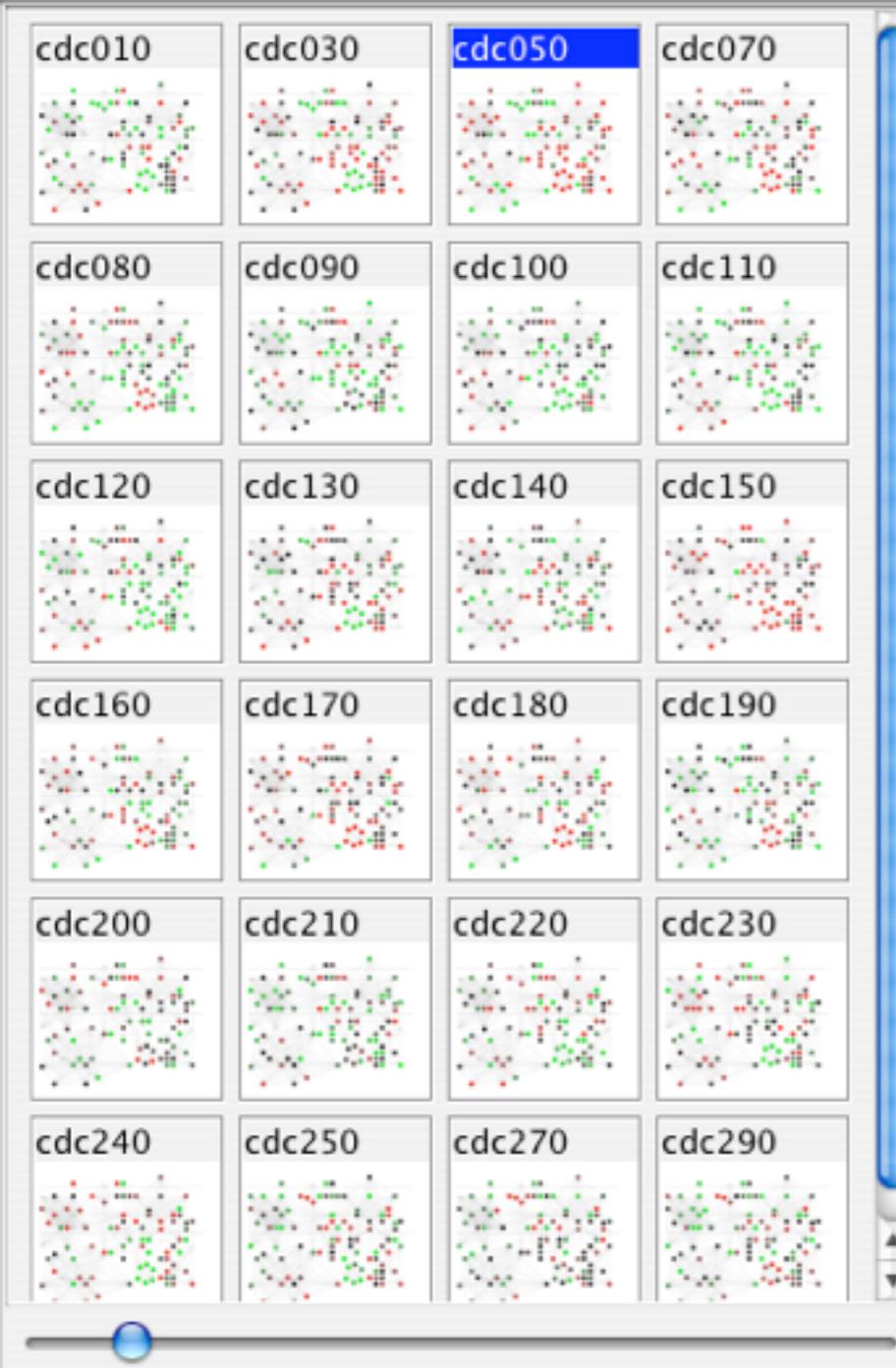
Colins and Carpendale 2007

MCV Type 1:

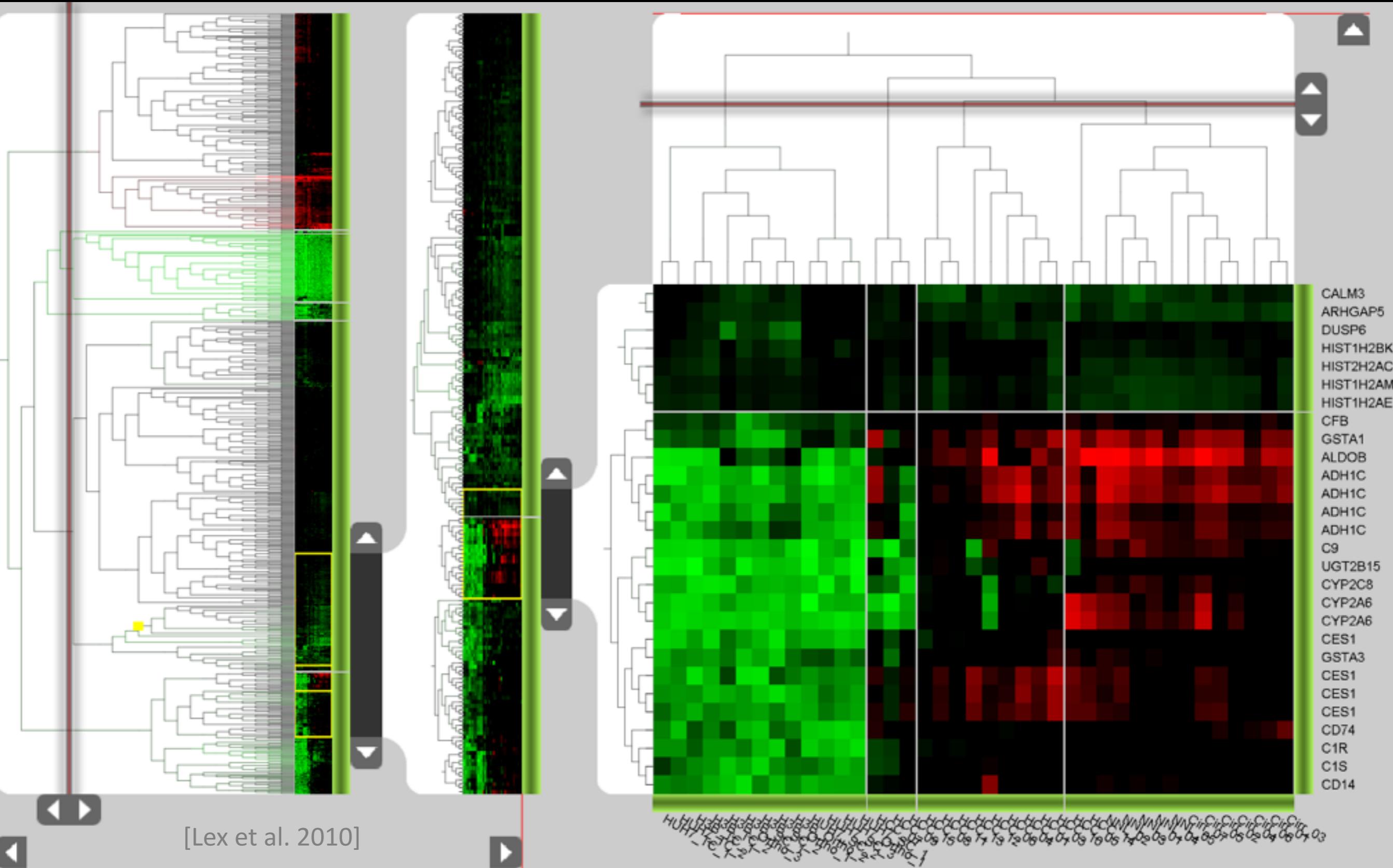
Different vis. techniques showing the same data



MCV Type 2: Same vis. technique showing different data



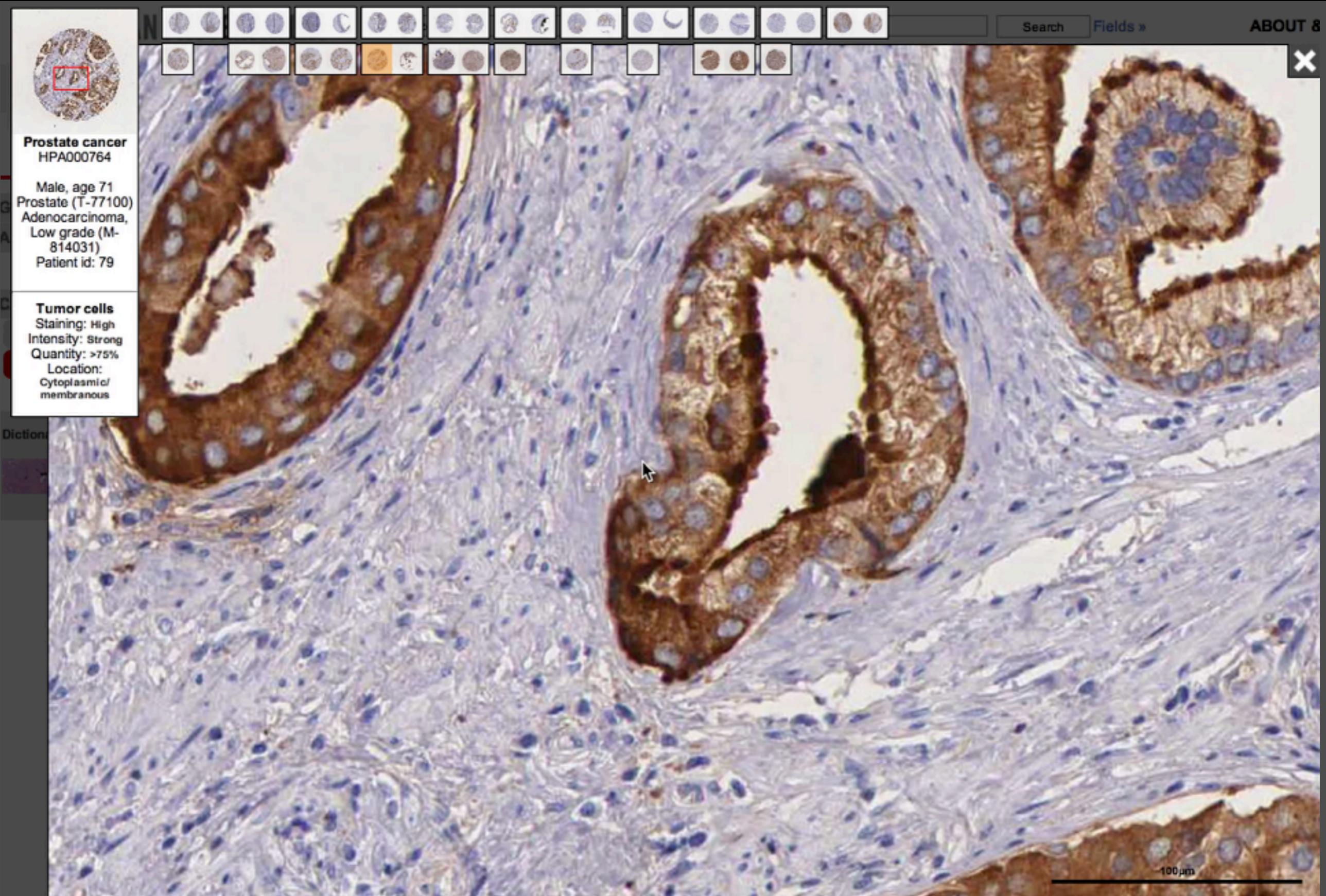
MCV Type 3: Overview + Detail



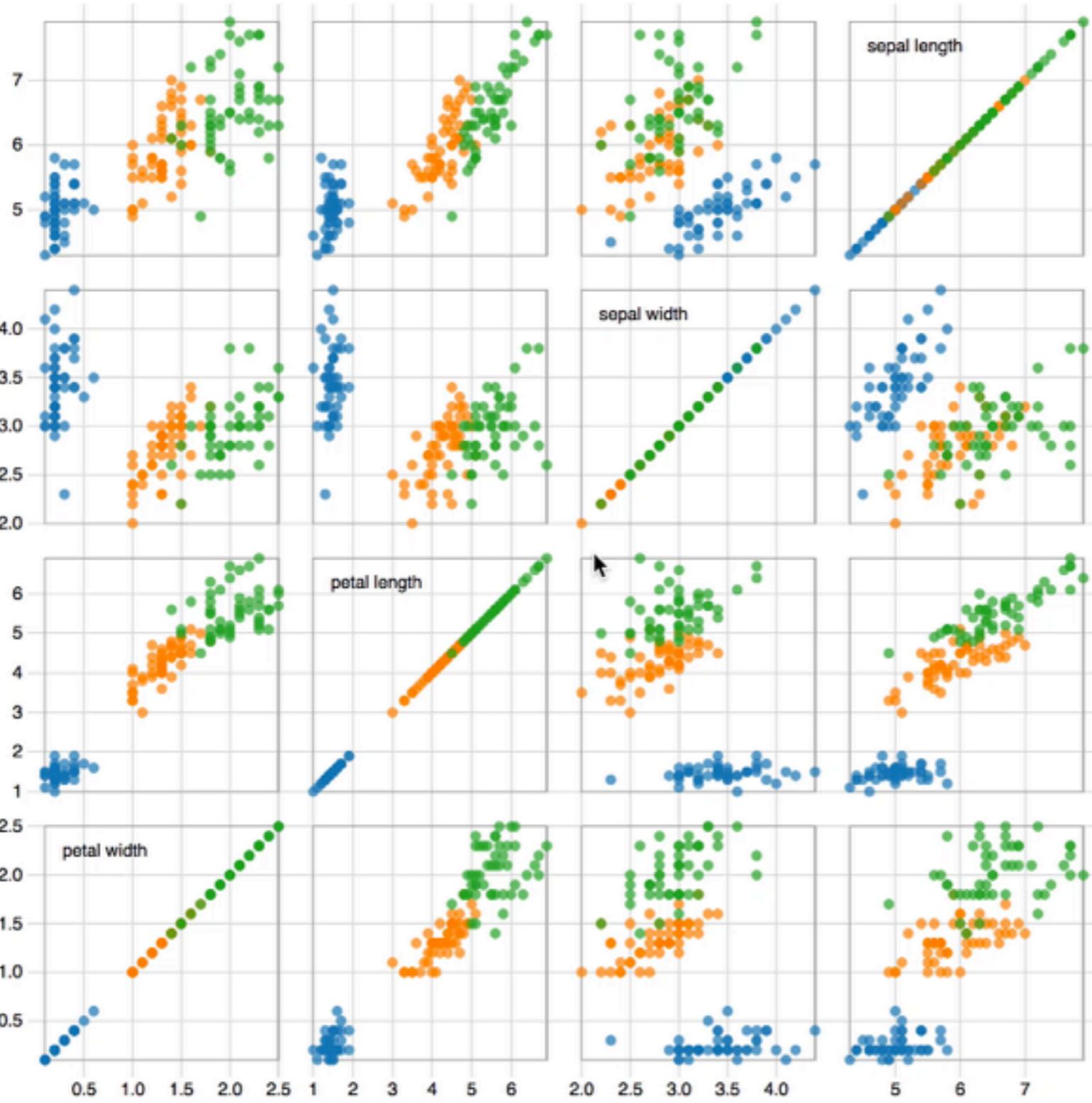
[Lex et al. 2010]

Example: Human Protein Atlas Project

<http://www.proteinatlas.org/>

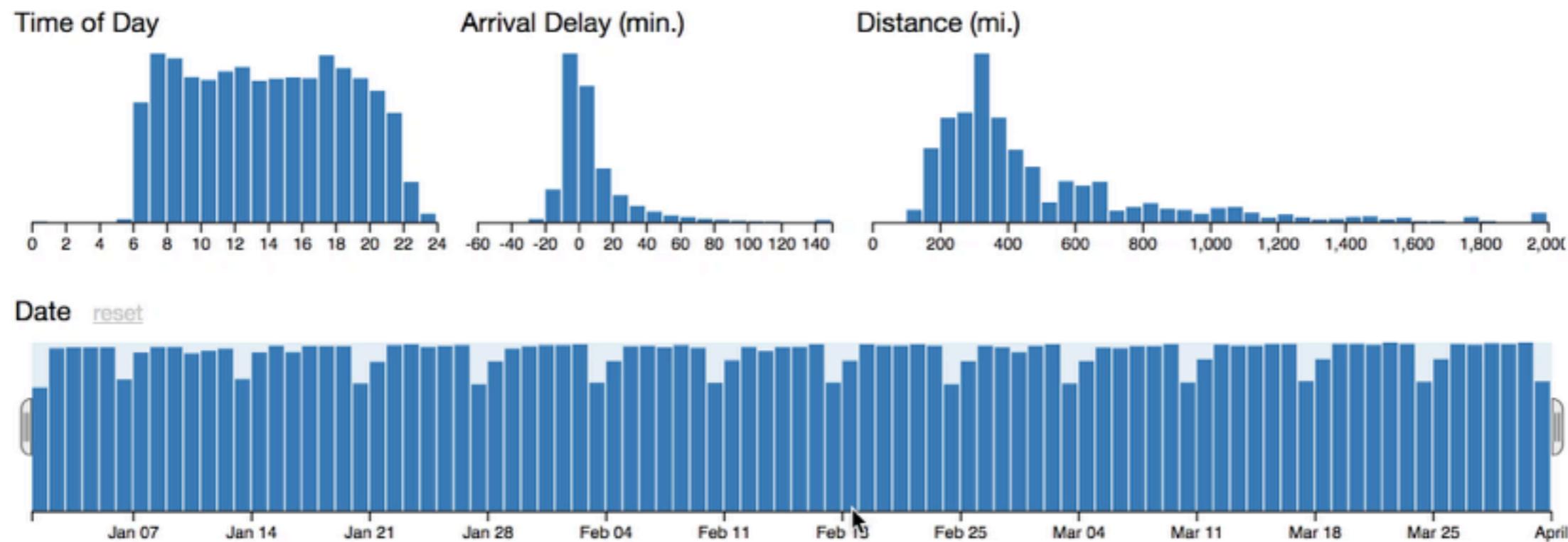


Selection / Filtering



Iris Dataset

Filtering Example



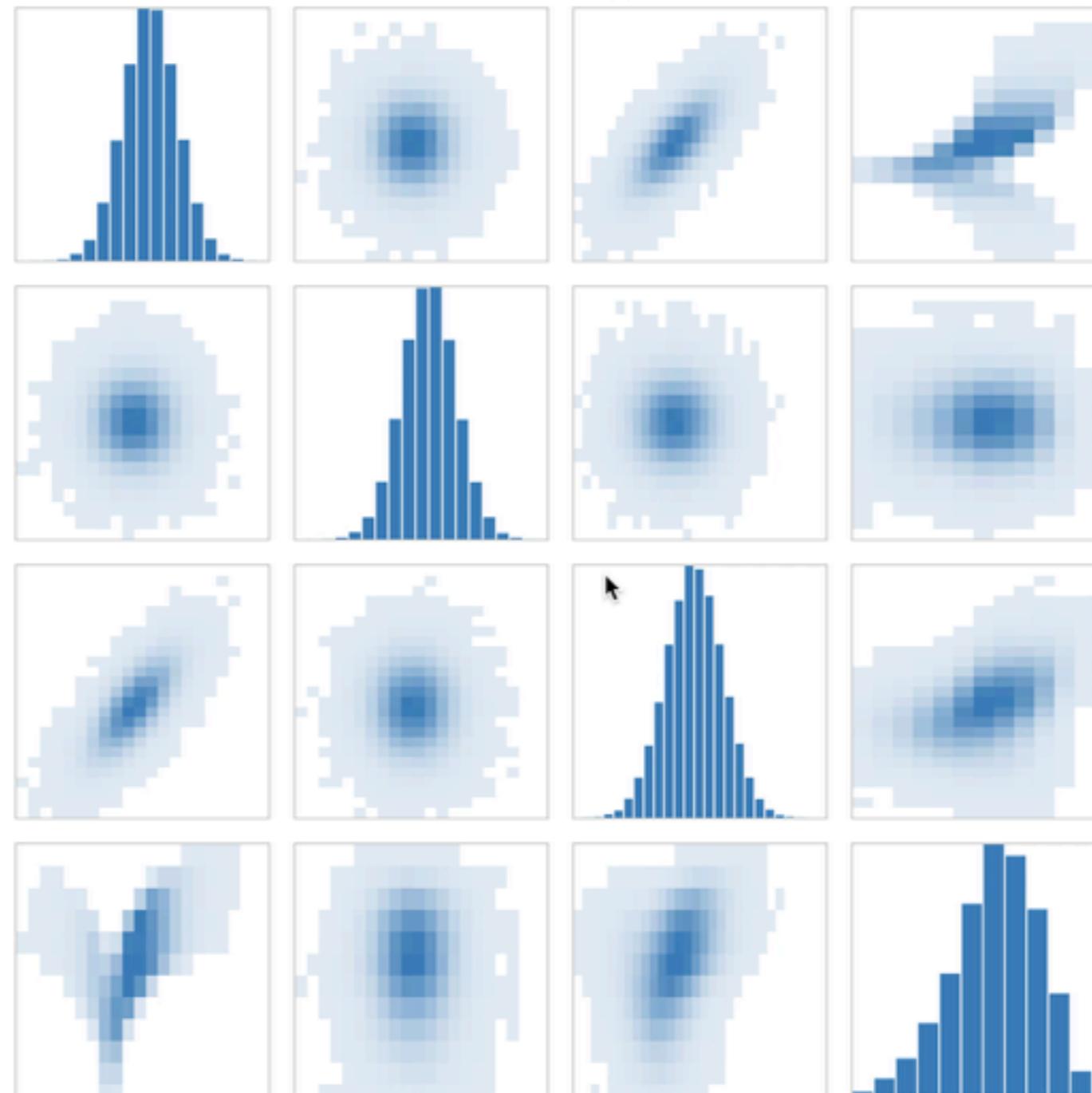
March 31, 2001

231,083 of 231,083 flights selected.

10:57 PM	MSY	HOU	303 mi.	+29 min.
10:48 PM	STL	MCO	880 mi.	+125 min.
10:37 PM	BNA	RDU	443 mi.	+106 min.
10:30 PM	HOU	MSY	303 mi.	-8 min.
10:10 PM	OMA	MDW	423 mi.	-7 min.
10:00 PM	HOU	MSY	303 mi.	-7 min.
09:55 PM	MSY	MCO	550 mi.	-16 min.
09:51 PM	MCO	MSY	550 mi.	+70 min.
09:45 PM	BNA	BWI	588 mi.	-5 min.

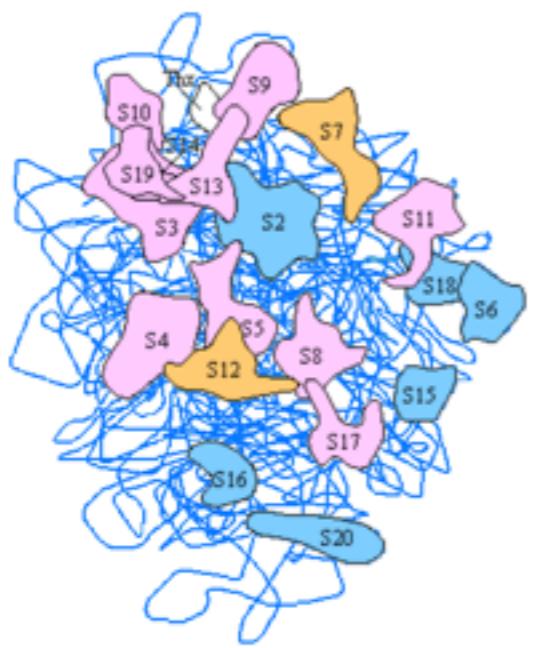
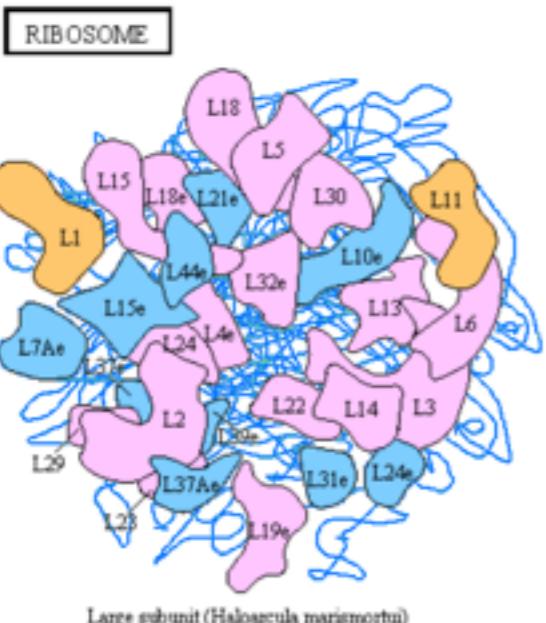
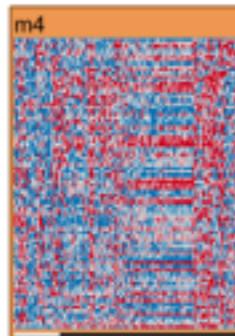
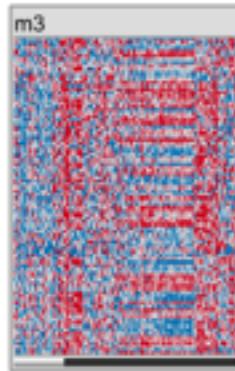
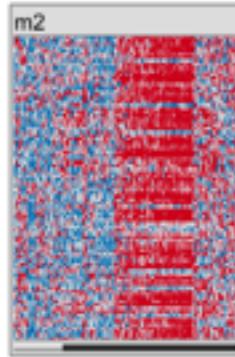
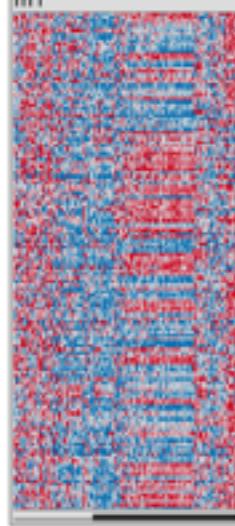
SPLOM Aggregation

Interactive Binned Scatterplot Matrix Dimensions: 4 Bins: 20 Data Points: 100k





Spot the difference!



03010 7/29/13
(c) Kanehisa Laboratories

Ribosomal RNAs

Bacteria / Archaea	23S	5S	16S
Eukaryotes	23S	5S	5.8S
	23S	5S	18S

Ribosomal proteins

EF-Tu	S10	L3	L4	L23	L2	S19	L22	S3	RP-L16	L29	
	S20e	L3e	L4e	L23Ae	L8e	S15e	L17e	S3e		L35e	
										L10e	L7/L12 stalk

S17	L14	L24		L5	S14	S8	L6		L18	S5	L30	L15	SecY
S11e	L23e	L26e	S4e	L11e	S29e	S15Ae	L9e	L32e	L19e	L5e	S2e	L7e	L27Ae

IF1	L36	S13	S11	S4	RpoA							
L34e	L14e	S18e	S14e	S9e	L18e	L17	L13	S9				

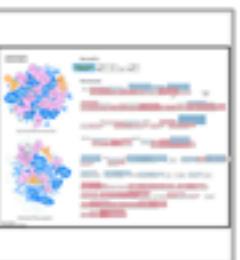
EF-TuG	S7	S12		L7A	RpoC,B							
	S5e	S23e	L30e	L7Ae	L7M12	L12	L10	L1	L11			
					LPI,LP2	LP0	L10Ae	L12e				

EF-Ts	S2	S4e	IF2	S15	IF3	L35	L20	L34	RF1	L31	L32	19	S18	S6
	S4e			S13e										

L28	L33		IF3	L35	L20	L34		RF1	L31	L32	19	S18	S6	
L28	L33			L21	L27	PtsY,Ffh	S16	L19	S1	S20	S21	L25		

L10e	L13e	L15e	L21e	L24e	L31e	L35Ae	L37e	L37Ae	L39e	L40e	L41e	L44e	
S3Ae	S6e	S8e	S17e	S19e	S24e	S25e	S26e	S27e	S27Ae	S28e	S30e		LX

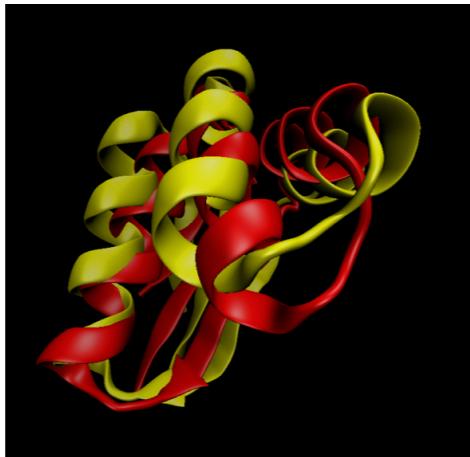
L6e	L18Ae	L22e	L27e	L28e	L29e	L36e	L38e					
S7e	S10e	S12e	S21e									



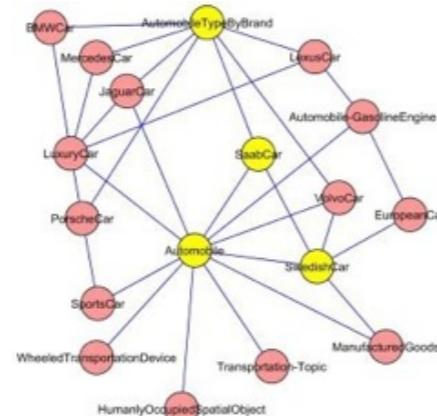
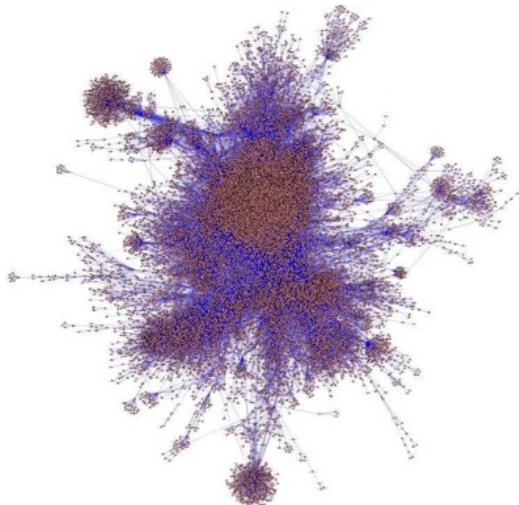
stratomex.caleydo.org

Spot the difference!

Summary: Key Concepts

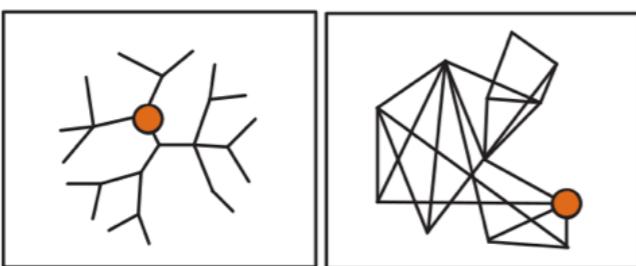


Navigation



Abstraction

Semantic/Geometric Zooming



Multiple Coordinated views

Questions

