Visualizing Biomolecular Data
with the Caleydo Framework

Alexander Lex

Graz University of Technology, Austria

CALEYDO

TU Graz

My Background: Austria

Our Institute

Institute for Computer Graphics and Vision
- Staff of about 80 people
- 4 Professors, 16 Post-Docs, 50 PhD students

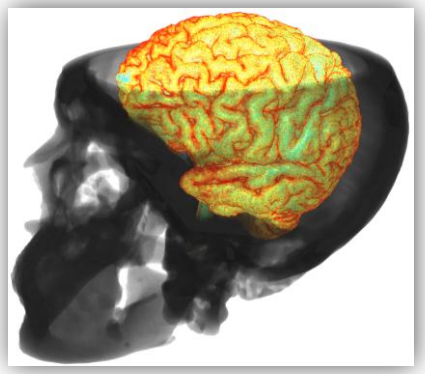
Computer Vision Group:
- Prof. Horst Bischof

Computer Graphics Group:
- Prof. Dieter Schmalstieg

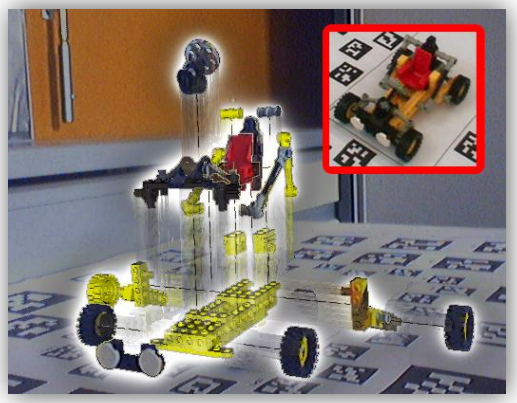
Research Areas

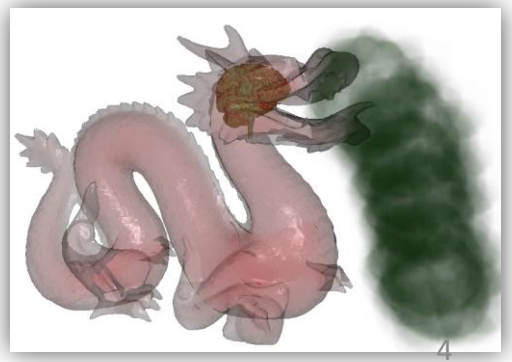
Augmented Reality

Multi-Display Environments



Volume Rendering / Medical Visualization





CALEYDO

Information Visualization and Visual Analytics

That’s what I’m talking about today!

InfoVis Research Topics

1. Visualization for biomolecular data
2. Connecting heterogeneous data sets

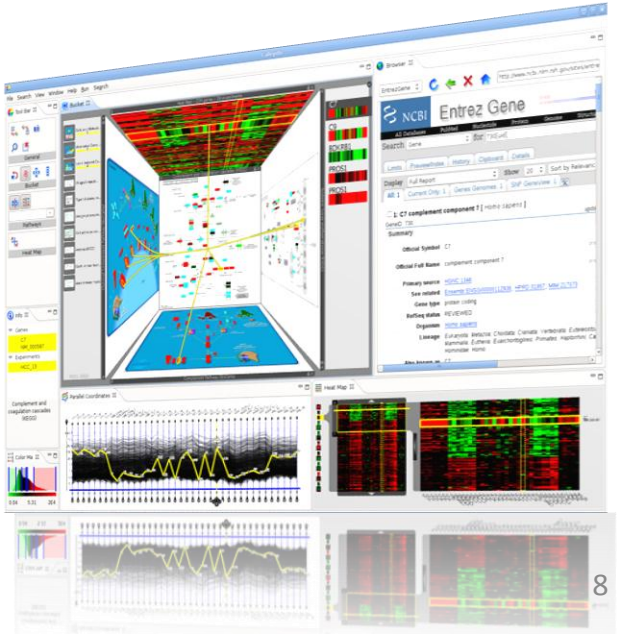
Find new visualization techniques and develop software that helps to understand the growing amounts of data in the life sciences

THE  PROJECT

What is Caleydo?

Visualization software developed with two goals:

1. Platform for creating novel visualization techniques
2. Tool that should actually help biologists



Features

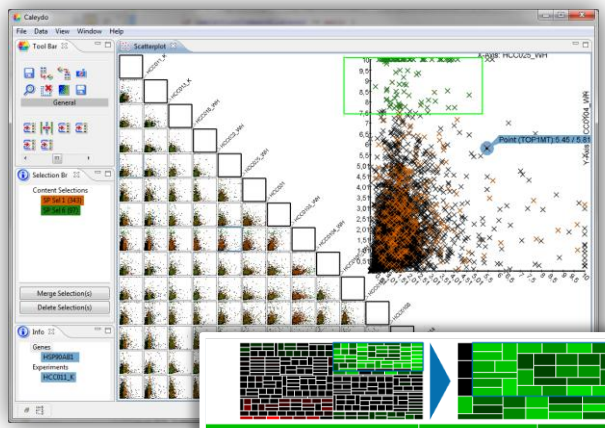
Various visualization techniques for tabular data and pathways

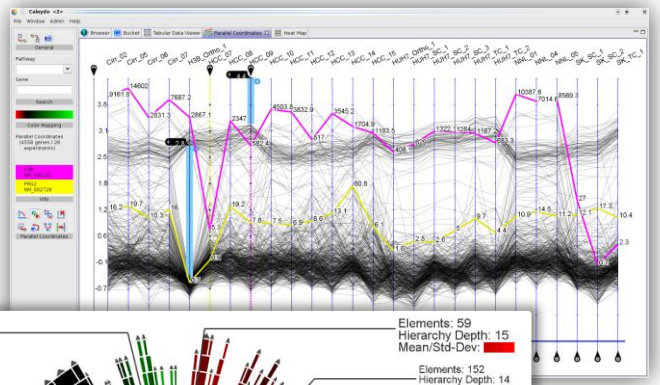
Runs on Windows and Linux (Mac yet to come)

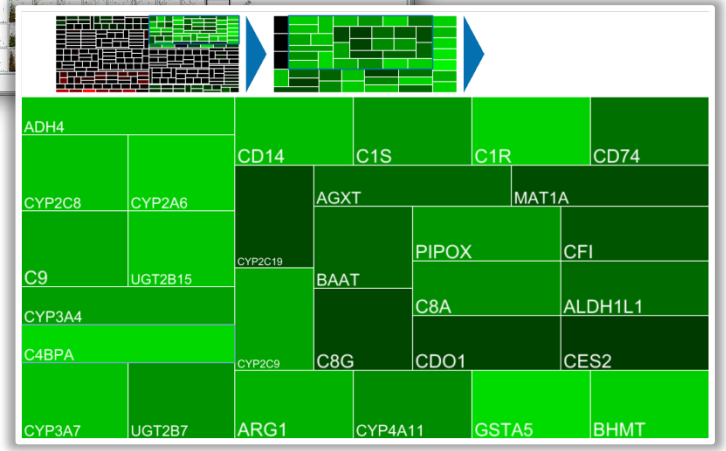
Analytical features:

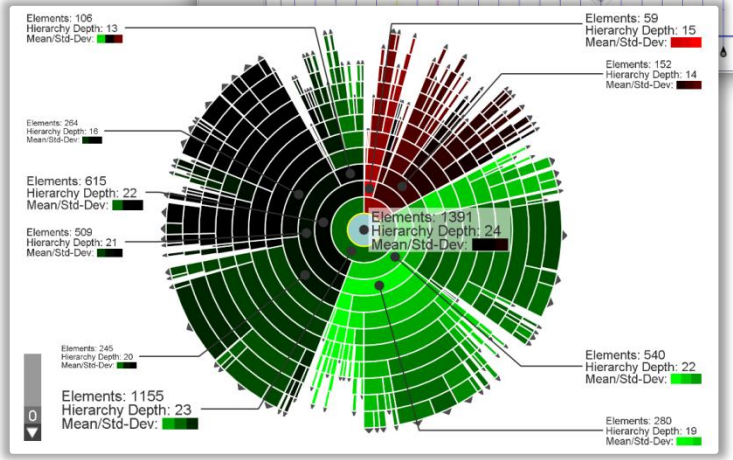
- R and WEKA interfaces
- Native clustering
- Fold Change, T-Tests, Variance Filters

Numerous Techniques









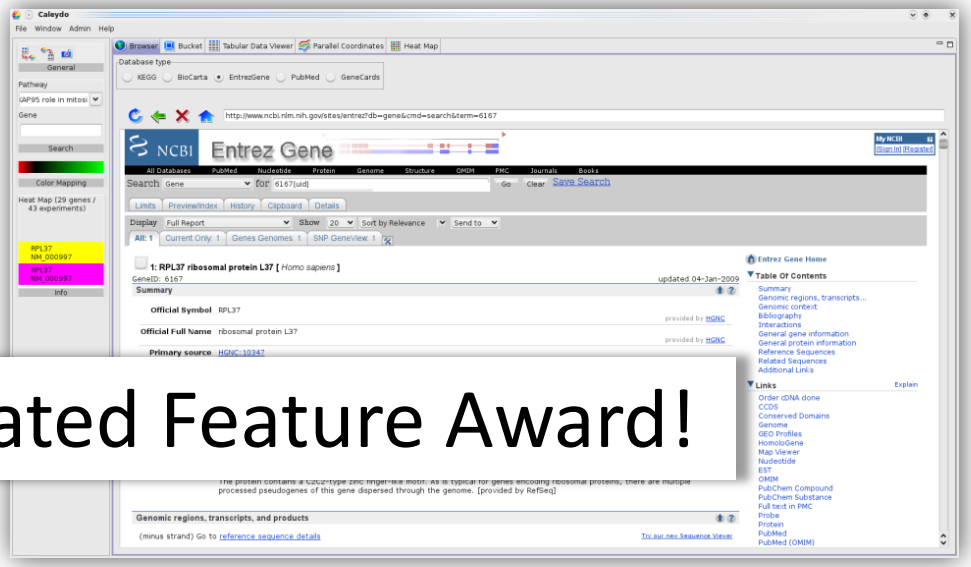
Linked Browser

Connected to major medical web databases

PubMed, Entrez,

KEGG, BioCarta,

GeneCards…



Most Appreciated Feature Award!

CALEYDO

VISUALIZATION FOR BIOMOLECULAR DATA

Research Topic 1

CALEYDO

Heat Maps

Common representation

Color coded magnitude of value

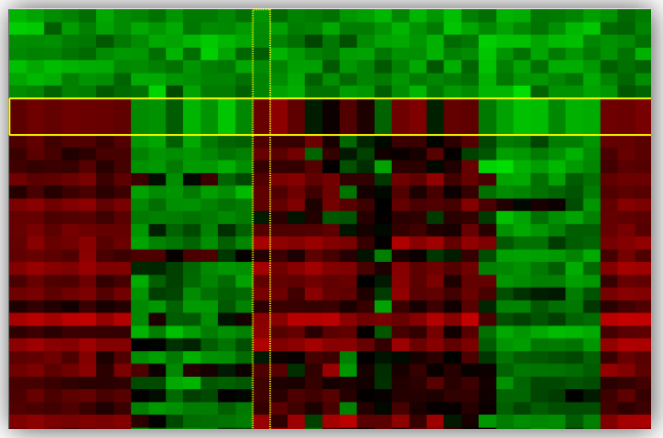
Filtering and clustering used

Problem:

do not scale well

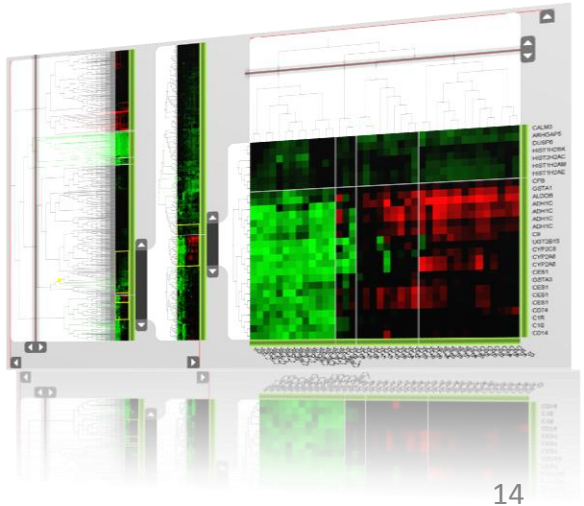
Solution:

hierarchical approach



[Lex, PacificVis 2010]

HIERARCHICAL HEAT MAP




Tool Bar
Bucket
Heat Map
Parallel Coordinates
Info
Genes
CYP2C9
NM_000771
Experiments
NNL_04
Browser
Entrez
NCBI
Entrez Gene
Search
Gene
for
Limits
Preview/Index
History
Clipboard
Details
Display
Full Report
Show
Sort by
1: ADH6 alcohol dehydrogenase 6 (class V) [Homo sapiens]
GeneID: 130
Summary
The heat map uses three levels of detail to visualize a focus and its context.

Works well, but…

Tables often have homogeneous subgroups of dimensions

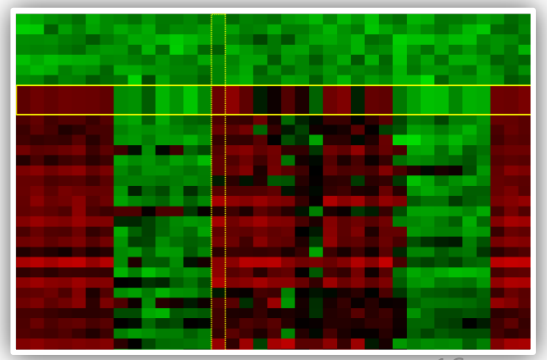
e.g., replicates, time-series, genotypes

Common task for multidimensional data:

Compare those groups

Traditional approach:

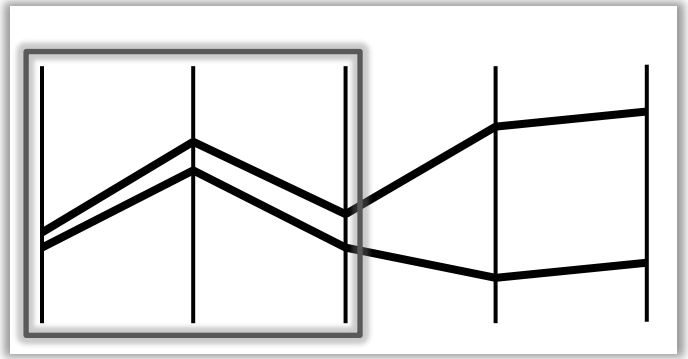
Filter, cluster all,
visualize with Heat Maps!



Clustering Inhomogeneous Data

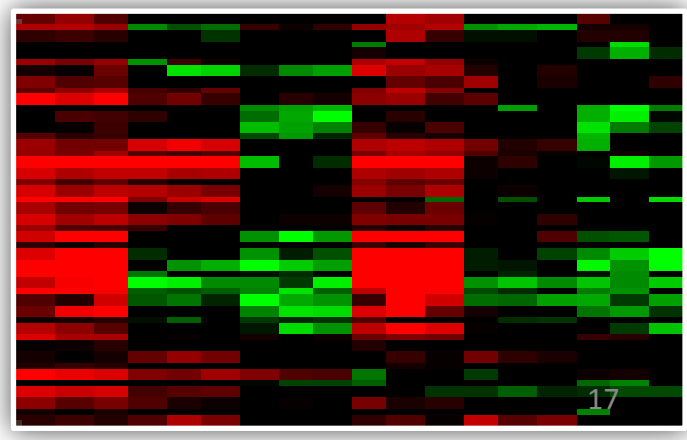
The Problem:

Clustering all dimensions at once obscures relations in homogeneous groups



The Solution:

Divide & Conquer!

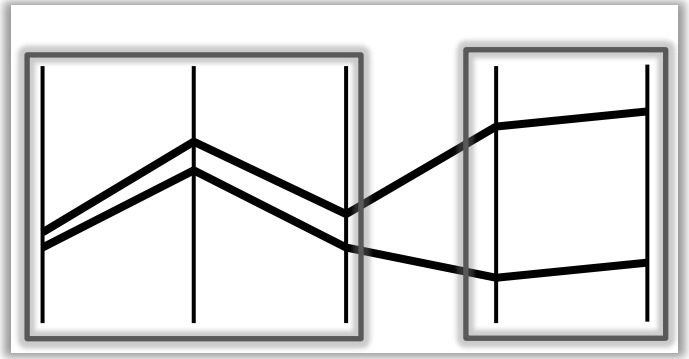


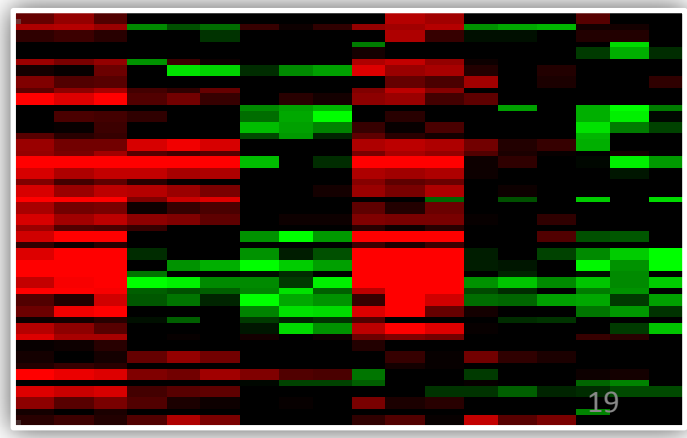
[Lex, InfoVis 2010]

MATCHMAKER

Clustering Inhomogeneous Data

Let's treat the groups separately!





Related Problem: Comparing Clustering Algorithms

Each run is one homogeneous group

Choice of algorithm, parameters and distance measures are important

No good quality metrics for clustering algorithms

- Visual assessment is best solution

The Process

1,1	1,2	1,3	1,4	1,5	1,6	1,7
2,1	2,2	2,3	2,4	2,5	2,6	2,7
3,1	3,2	3,3	3,4	3,5	3,6	3,7
4,1	4,2	4,3	4,4	4,5	4,6	4,7
5,1	5,2	5,3	5,4	5,5	5,6	5,7
6,1	6,2	6,3	6,4	6,5	6,6	6,7

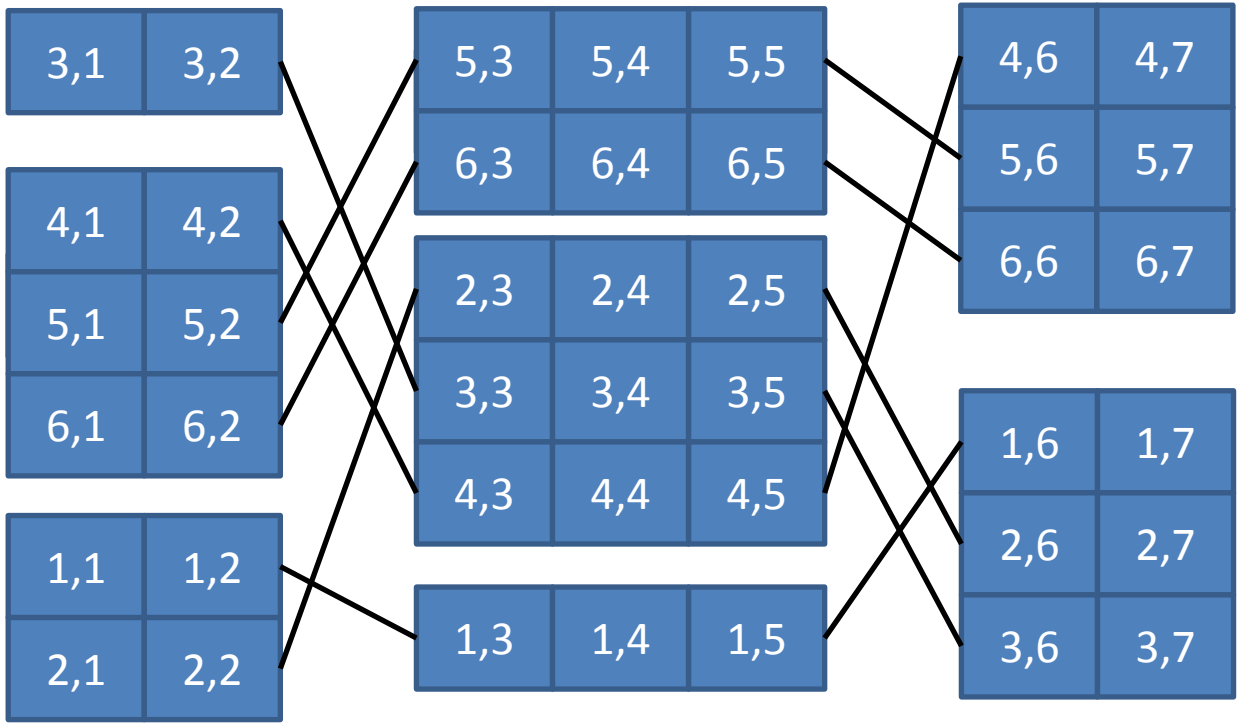
The Process

1,1	1,2
2,1	2,2
3,1	3,2
4,1	4,2
5,1	5,2
6,1	6,2

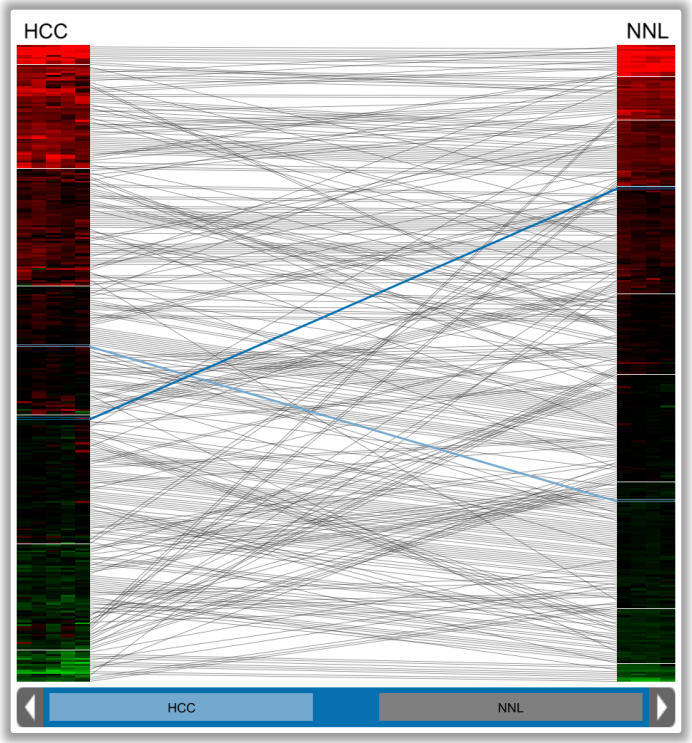
1,3	1,4	1,5
2,3	2,4	2,5
3,3	3,4	3,5
4,3	4,4	4,5
5,3	5,4	5,5
6,3	6,4	6,5

1,6	1,7
2,6	2,7
3,6	3,7
4,6	4,7
5,6	5,7
6,6	6,7

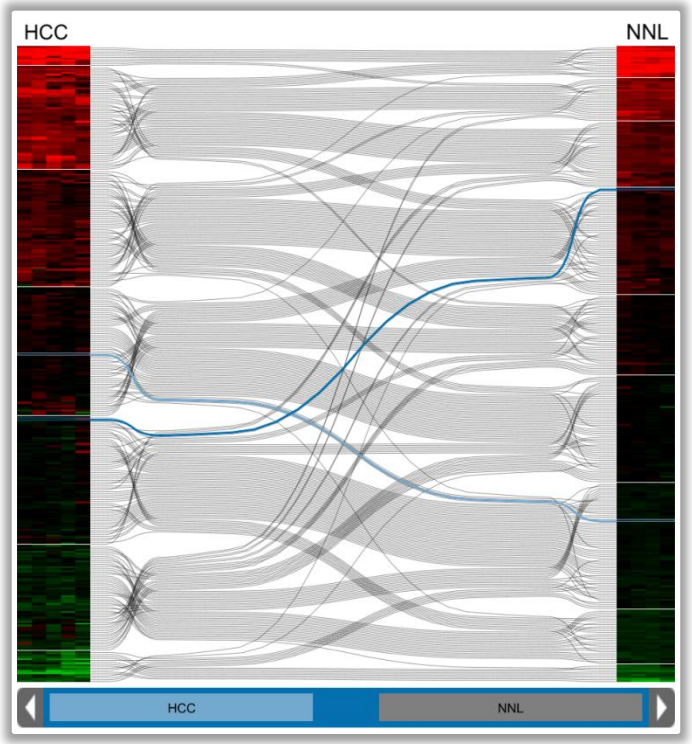
The Process


3,1
3,2
4,1
4,2
5,1
5,2
6,1
6,2
1,1
1,2
2,1
2,2
5,3
5,4
5,5
6,3
6,4
6,5
2,3
2,4
2,5
3,3
3,4
3,5
4,3
4,4
4,5
1,3
1,4
1,5
4,6
4,7
5,6
5,7
6,6
6,7
1,6
1,7
2,6
2,7
3,6
3,7

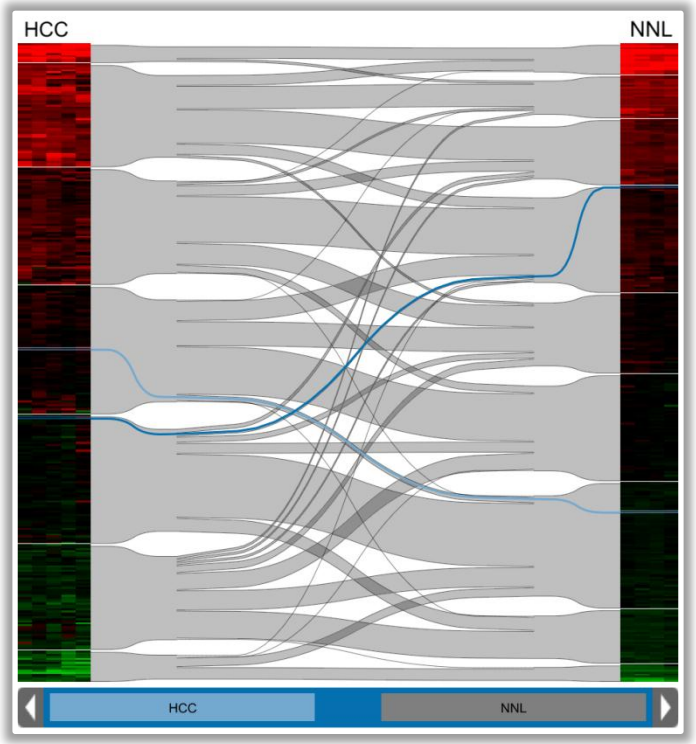
Immediate Result

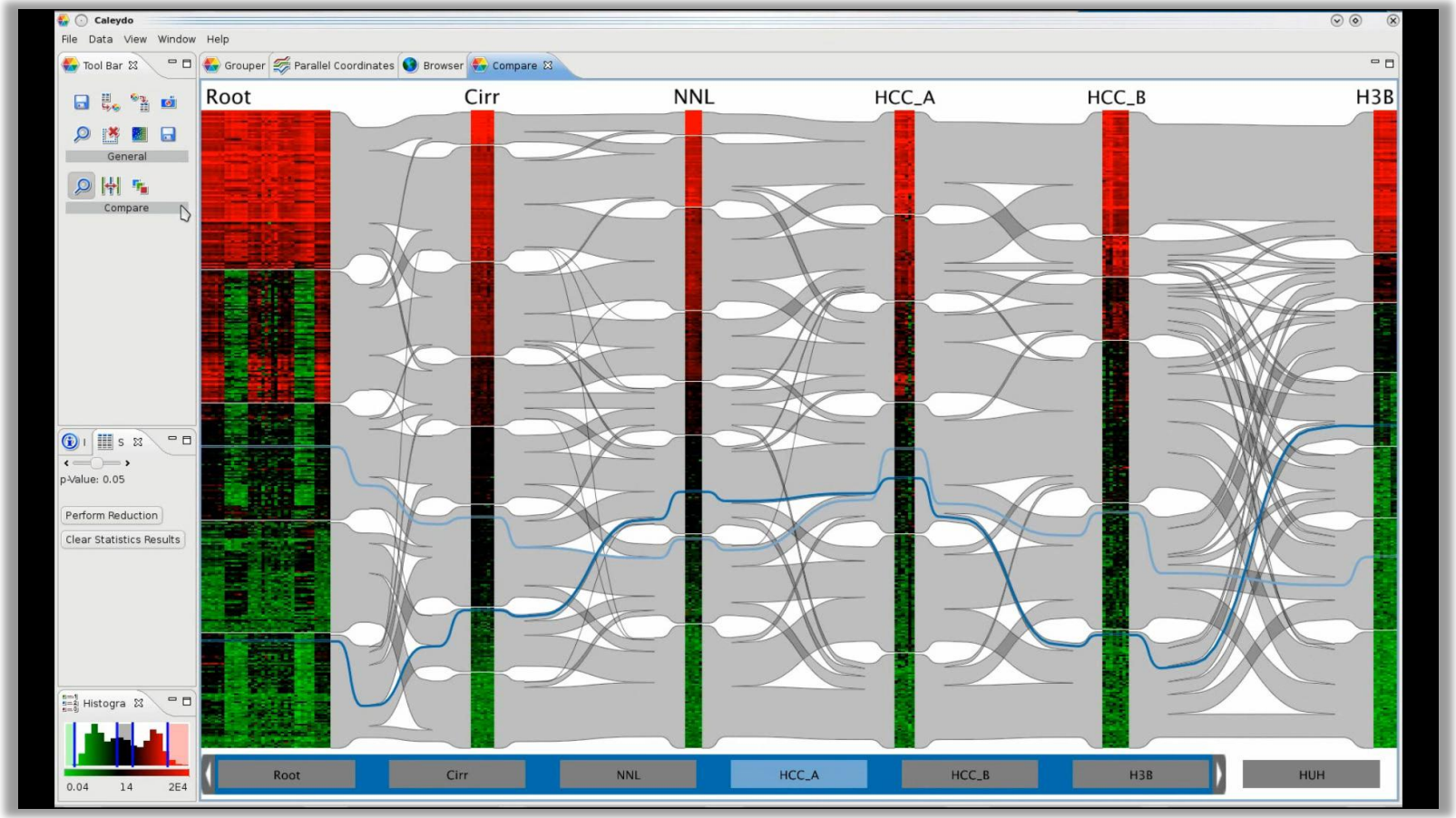


Using Splines

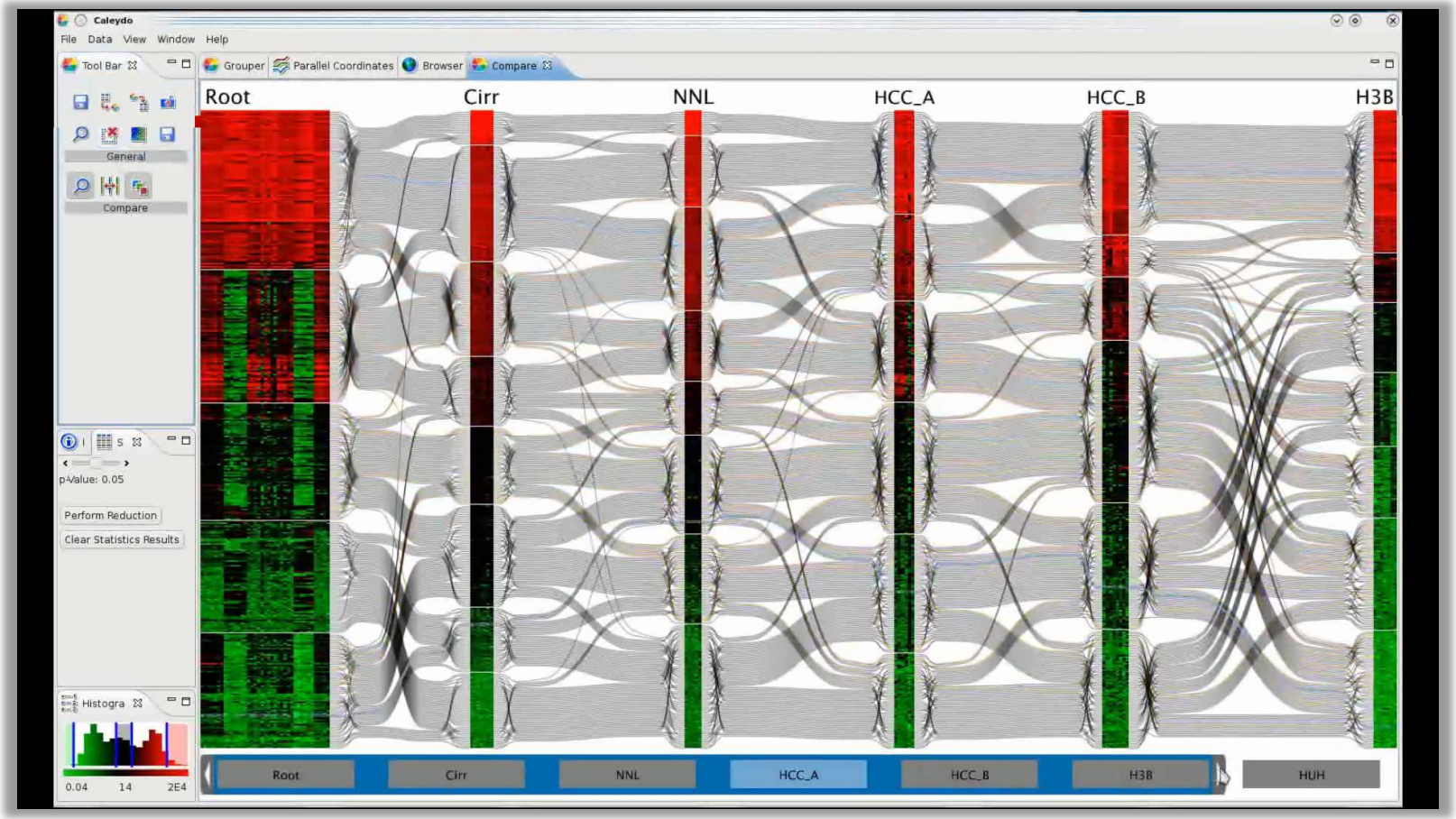


Using Pipes

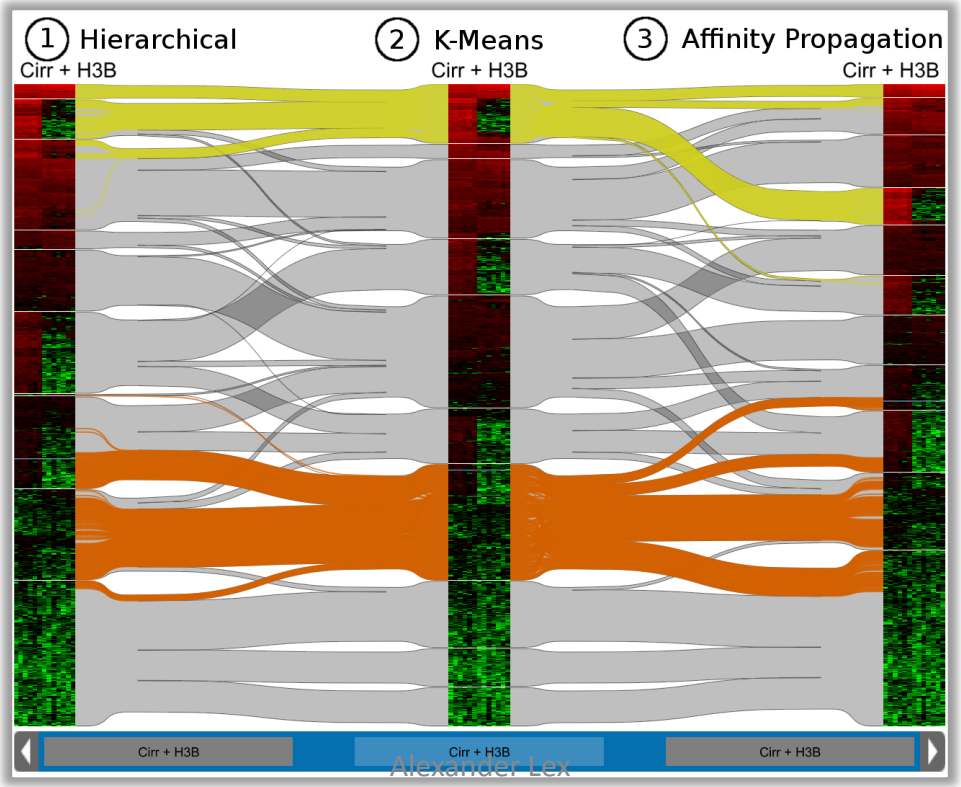



Caleydo
File Data View Window Help
Tool Bar
Grouper
Parallel Coordinates
Browser
Compare
General
Compare
Root
Cirr
NNL
HCC_A
HCC_B
H3B
p-Value: 0.05
Perform Reduction
Clear Statistics Results
Histogra
0.04
14
2E4
Root
Cirr
NNL
HCC_A
HCC_B
H3B
HUH

CALEYDO


Caleydo
File Data View Window Help
Tool Bar
General
Compare
Grouper
Parallel Coordinates
Browser
Compare
Root
Cirr
NNL
HCC_A
HCC_B
H3B
p-Value: 0.05
Perform Reduction
Clear Statistics Results
Histogra
0.04
14
2E4
Root
Cirr
NNL
HCC_A
HCC_B
H3B
HUH

Example: Cluster Algorithm Comparison



CALEYDO

The next step: Abstraction

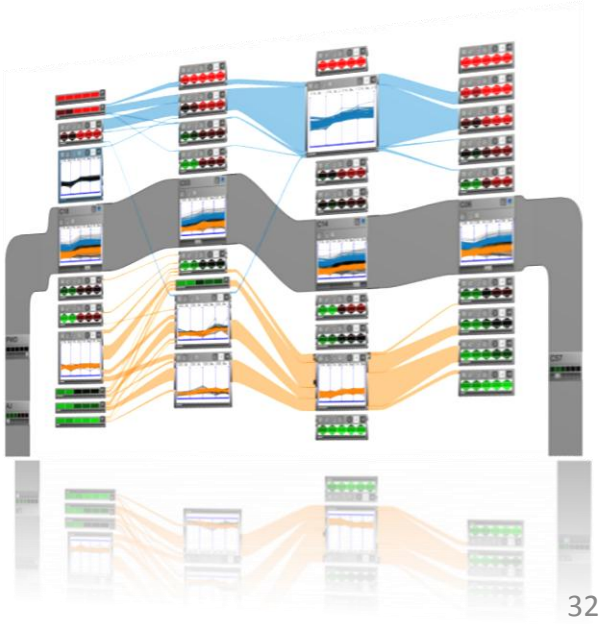
(Clustered) Datasets have homogeneous and inhomogeneous regions

No sense in wasting space for homogeneous parts

Use best visualization technique for each part and task!

[Lex, InfoVis 2011]

VISBRICKS



What is a Brick?

Shows part of data set

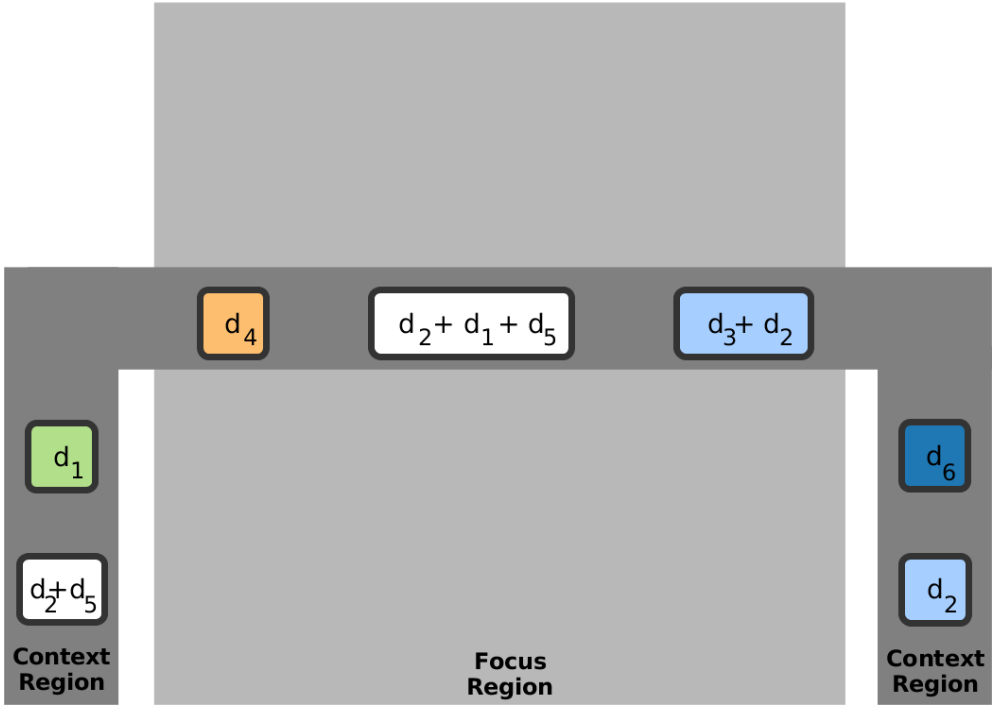
Multiform property:

- different levels of abstraction
- different visualization techniques

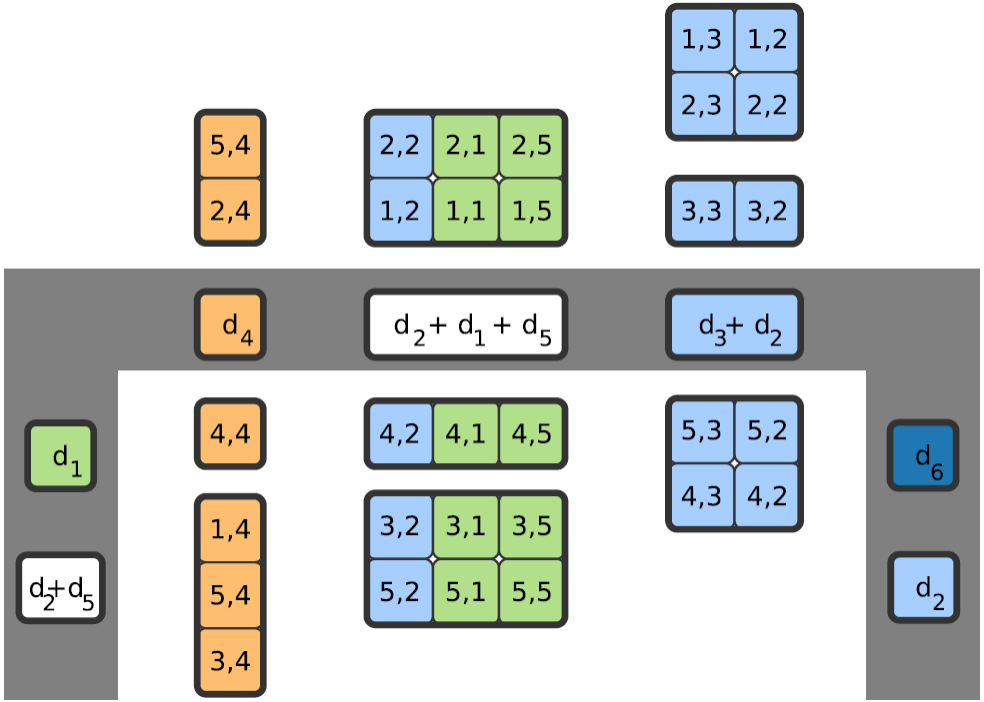
Two types of bricks:

- Dimension bricks
- Cluster bricks

Arrangement, Dimensions Bricks



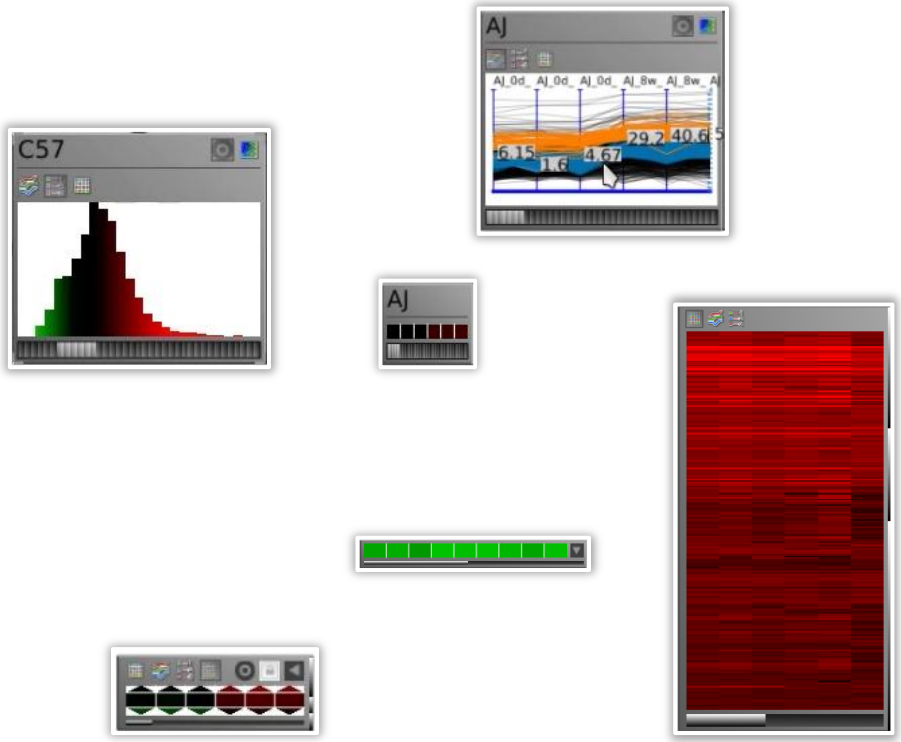
Added Cluster Bricks

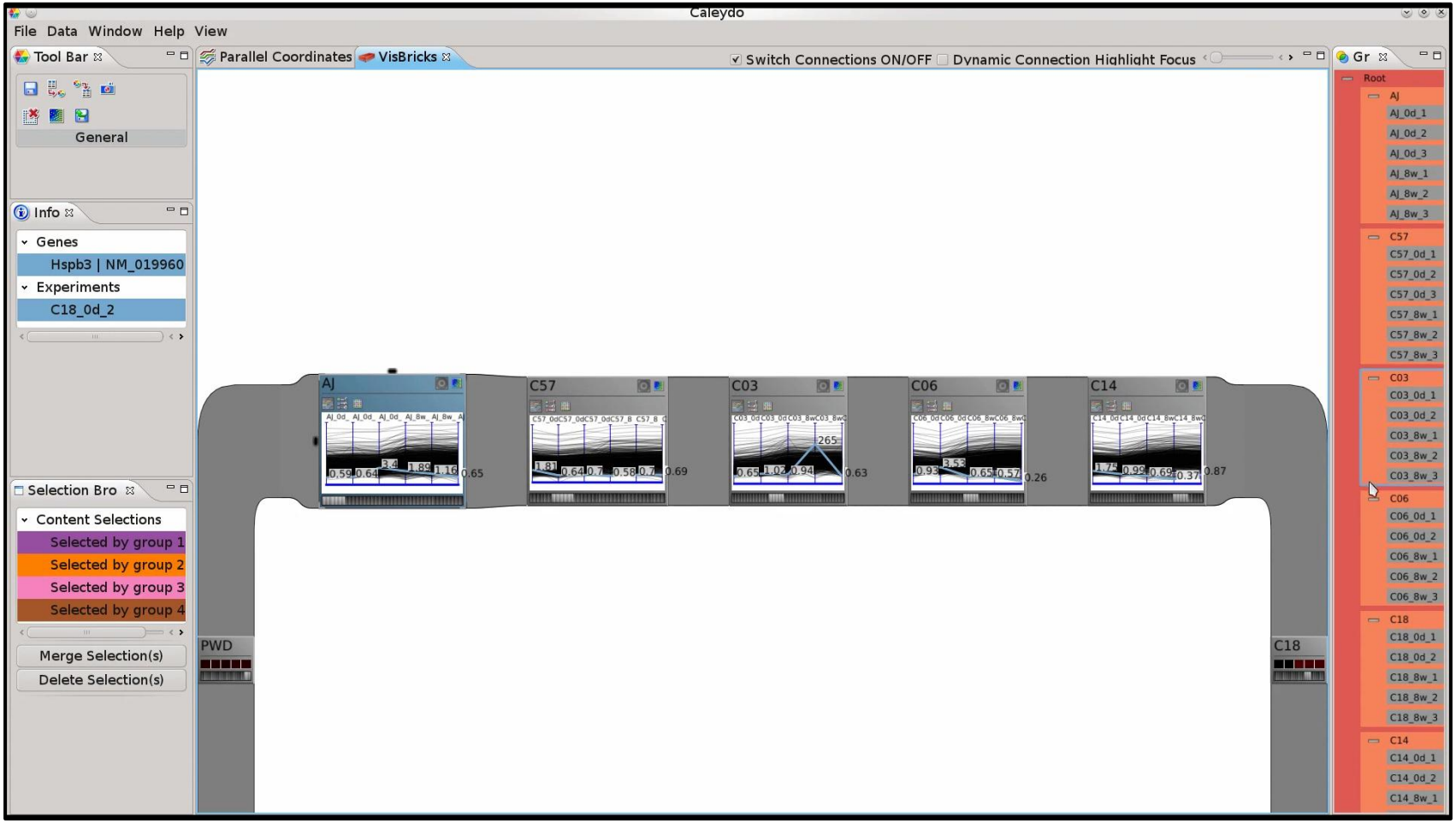


Examples for Bricks

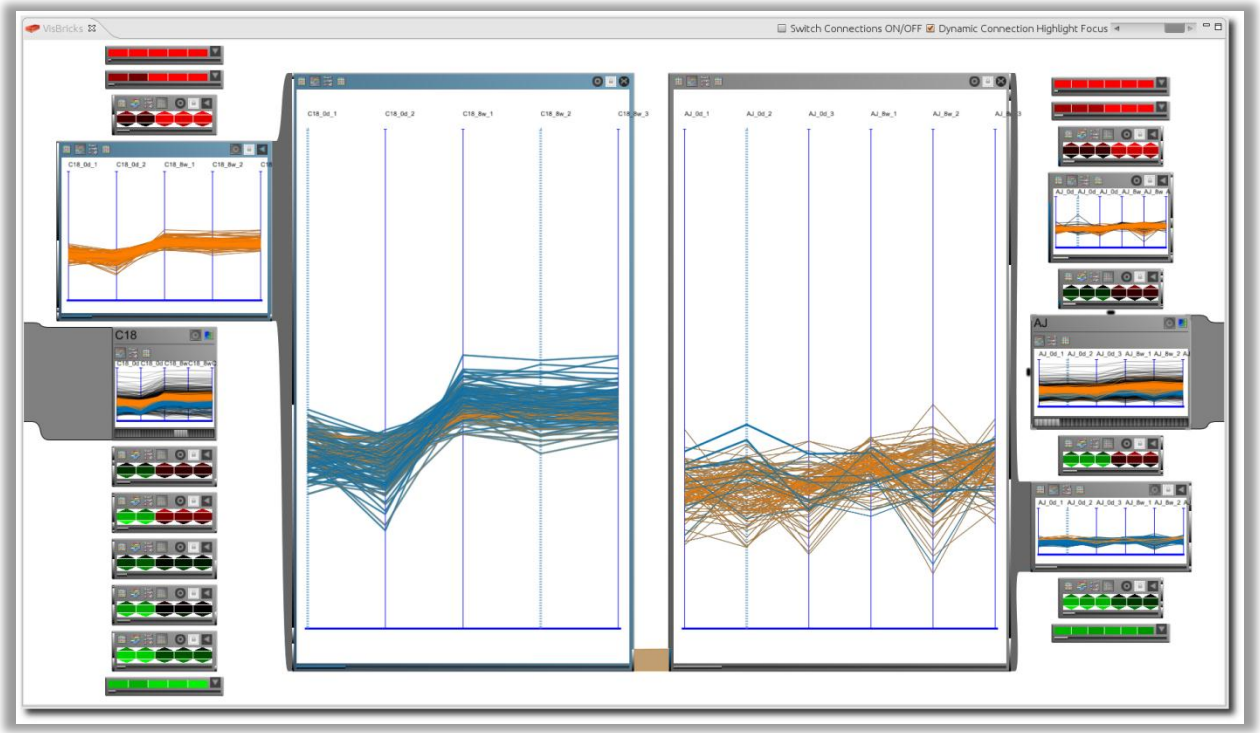
Dimension Bricks

Cluster Bricks



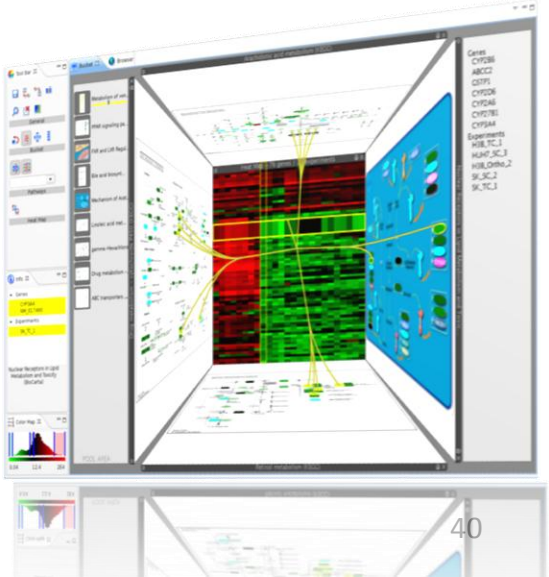

Caleydo
File Data Window Help View
Tool Bar
General
Info
Genes
Hspb3 | NM_019960
Experiments
C18_0d_2
Selection Bro
Content Selections
Selected by group 1
Selected by group 2
Selected by group 3
Selected by group 4
Merge Selection(s)
Delete Selection(s)
Parallel Coordinates
VisBricks
Switch Connections ON/OFF
Dynamic Connection Highlight Focus
Gr
Root
AJ
C57
C03
C06
C14
C18
PWD

Exploring Details



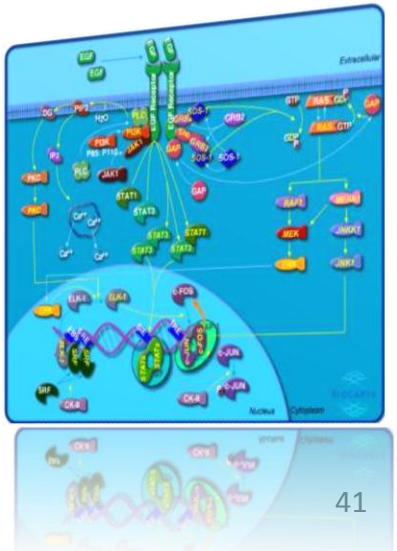
[Lex, PacificVis 2010]

PATHWAY EXPLORATION CALEYDO BUCKET

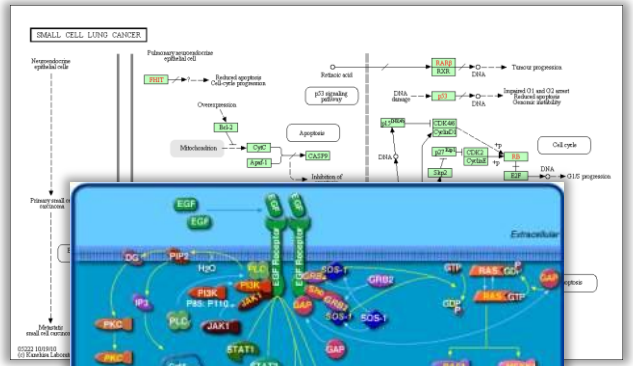


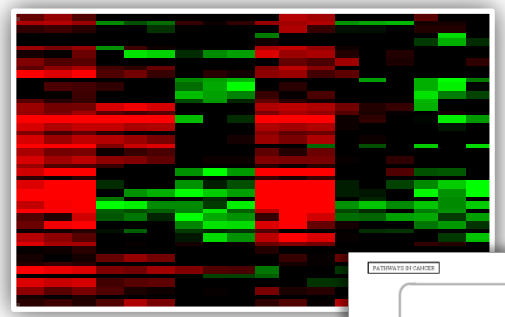
Pathway Visualization Goals

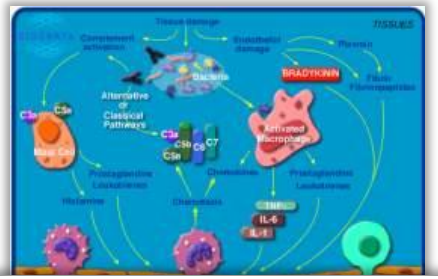
1. Map gene expression of multiple samples
2. Show the relations between pathways

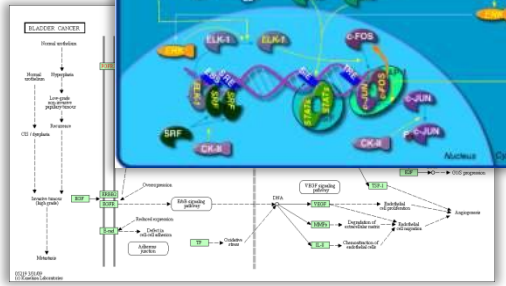


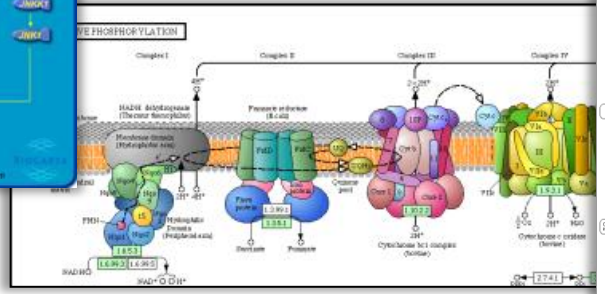
How to manage all these?

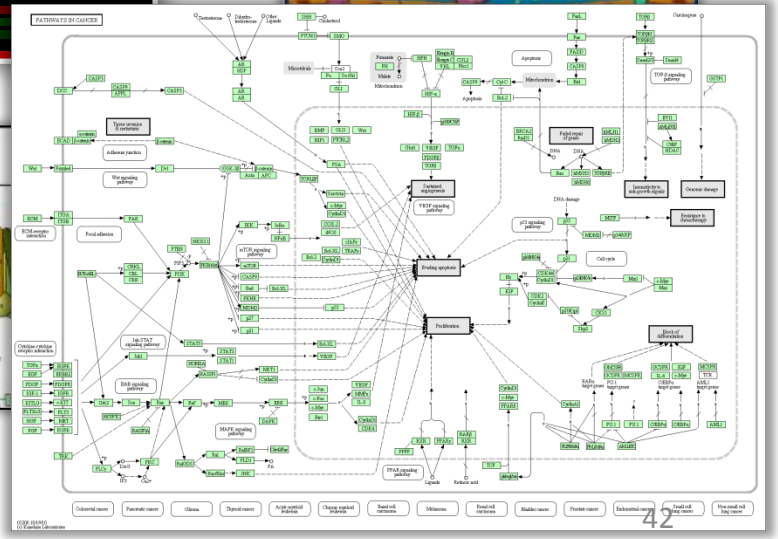




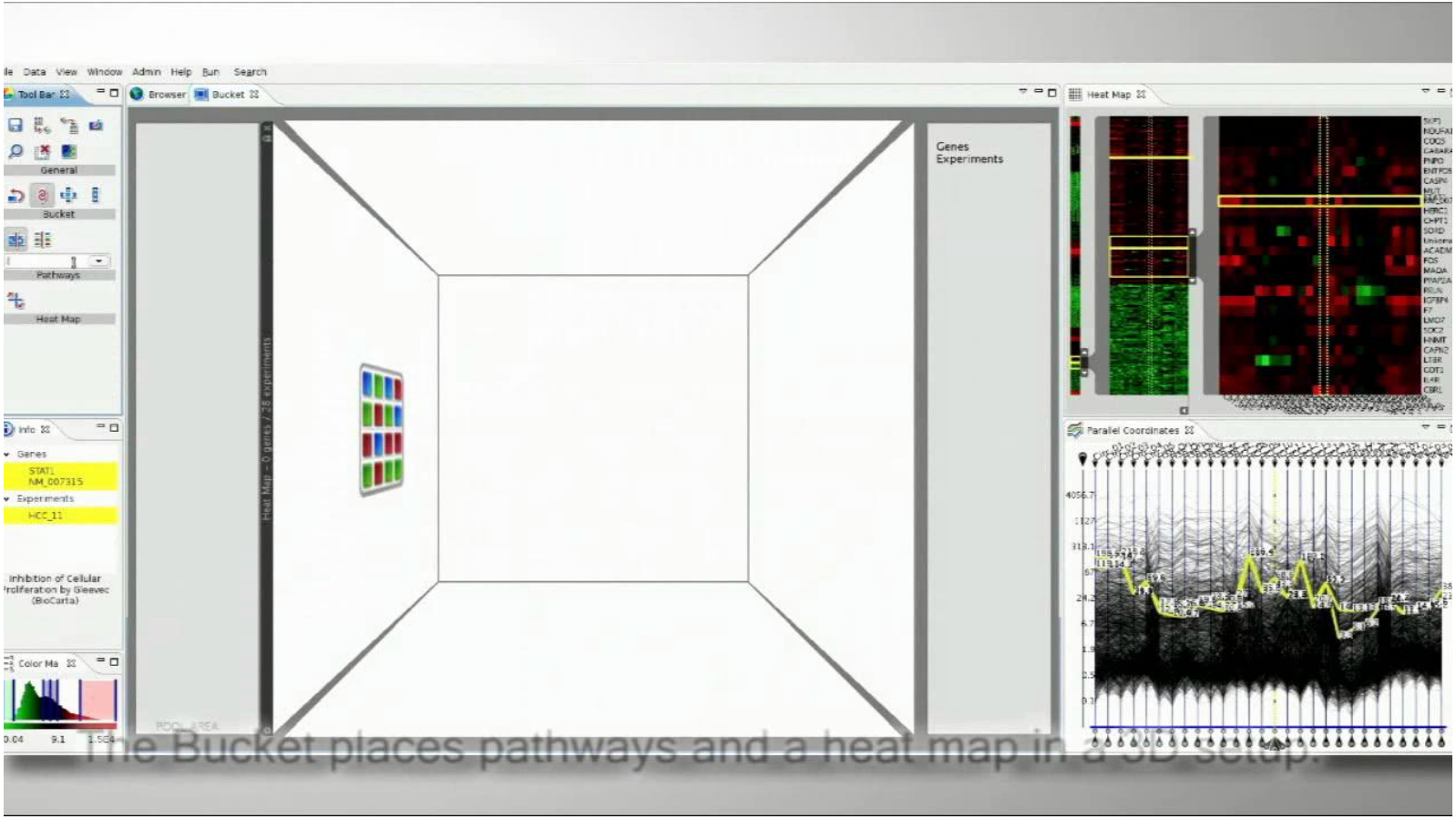








CALEYDO

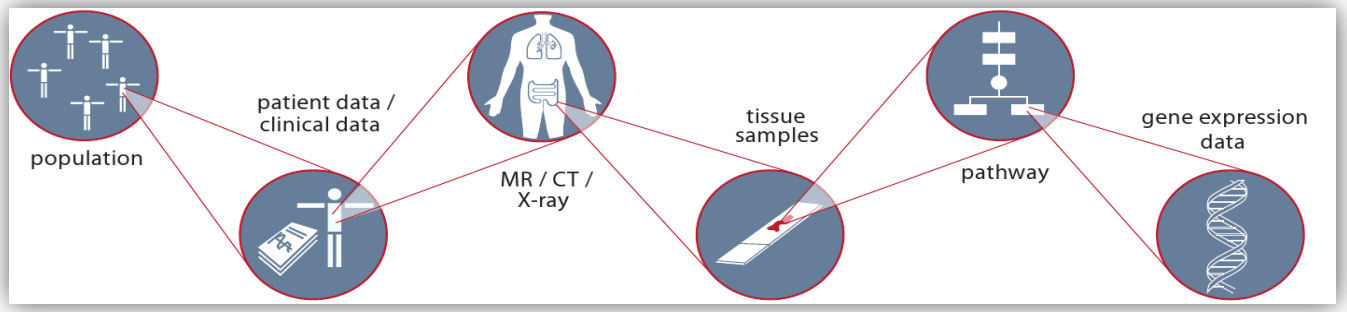

Tool Bar
General
Bucket
Pathways
Heat Map
Browser
Bucket
Genes
Experiments
Heat Map
Parallel Coordinates
Info
Genes
STAT1
NM_007315
Experiments
HCC_11
Inhibition of Cellular
Proliferation by Gleevec
(BioCarta)
Color Ma
The Bucket places pathways and a heat map in a 3D setup.

CONNECTING HETEROGENEOUS DATA SETS

Research Topic 2

Heterogeneous Data Analysis

Massive amounts of data
Different abstraction levels
Different sources
Different formats

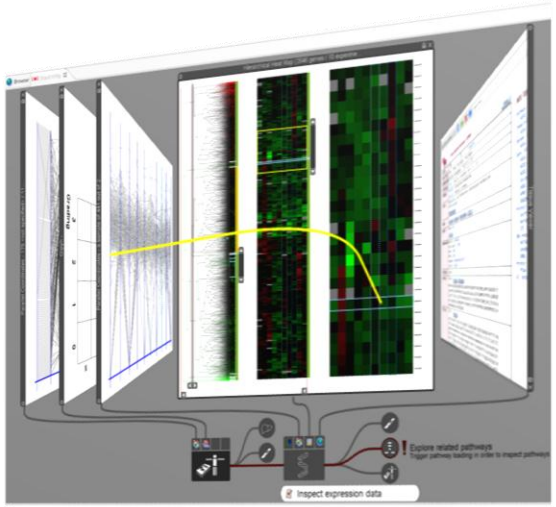


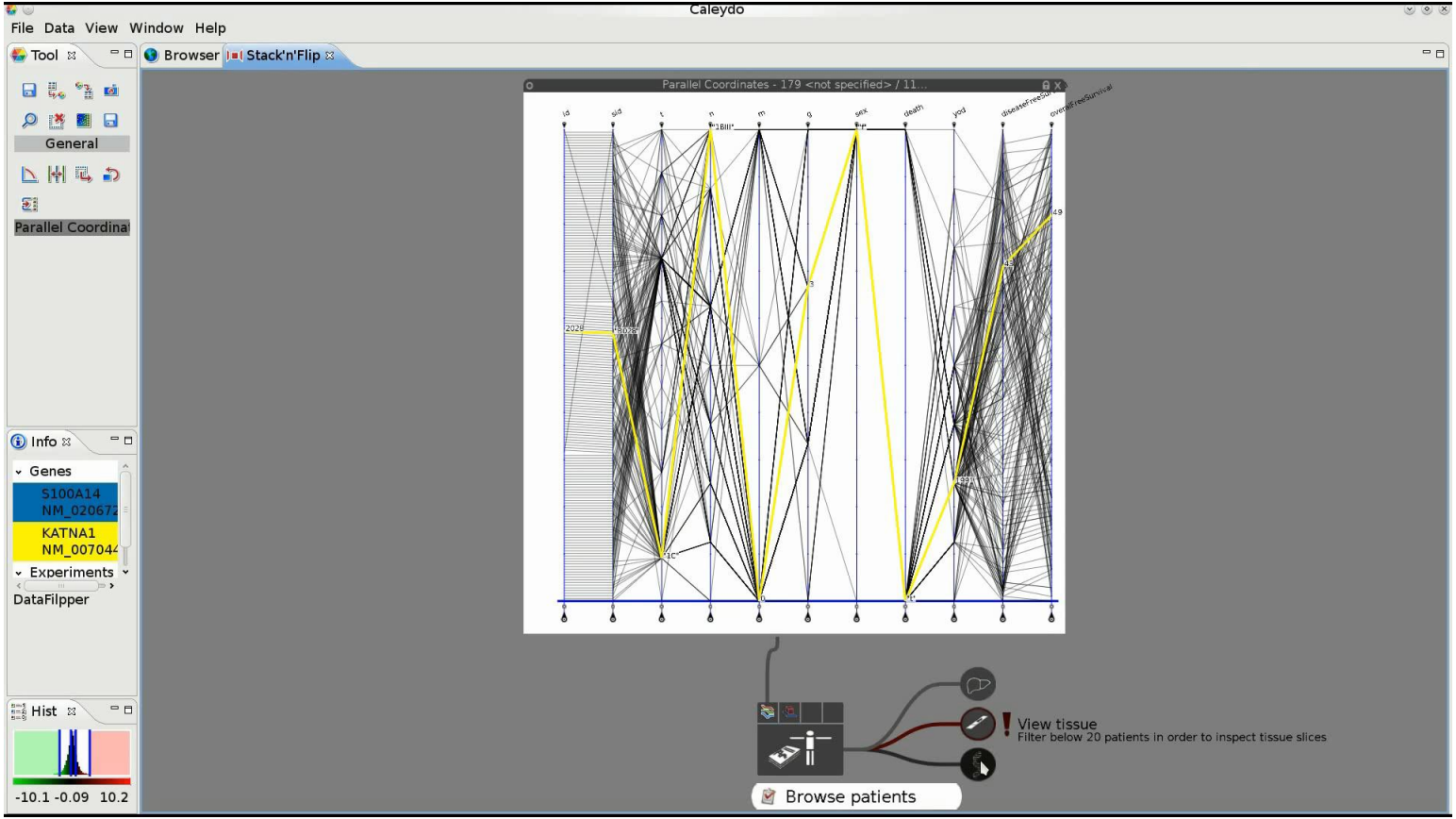
[Streit, CoVis 2009]

How to integrate data in a seamless analysis session?

[Streit, TVCG 2011]

GUIDANCE ACROSS MULTIPLE DATA SOURCES




Caleydo
File Data View Window Help
Tool
Browser
Stack'n'Flip
General
Parallel Coordina
Parallel Coordinates - 179 <not specified> / 11...
Info
Genes
S100A14
NM_020672
KATNA1
NM_007044
Experiments
DataFilpper
Hist
-10.1 -0.09 10.2
View tissue
Filter below 20 patients in order to inspect tissue slices
Browse patients

Super Application?

Expects a Super Application that can visualize everything

Not Feasible! Solution: use existing applications

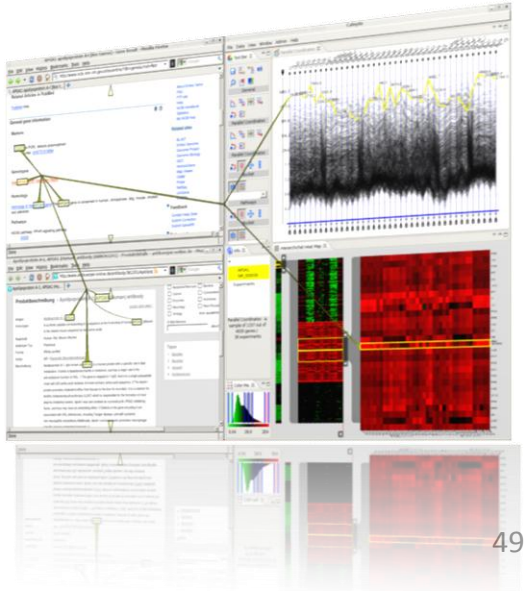
Downsides:

- not integrated
- no highlighting, linking, etc.

Can we solve this?

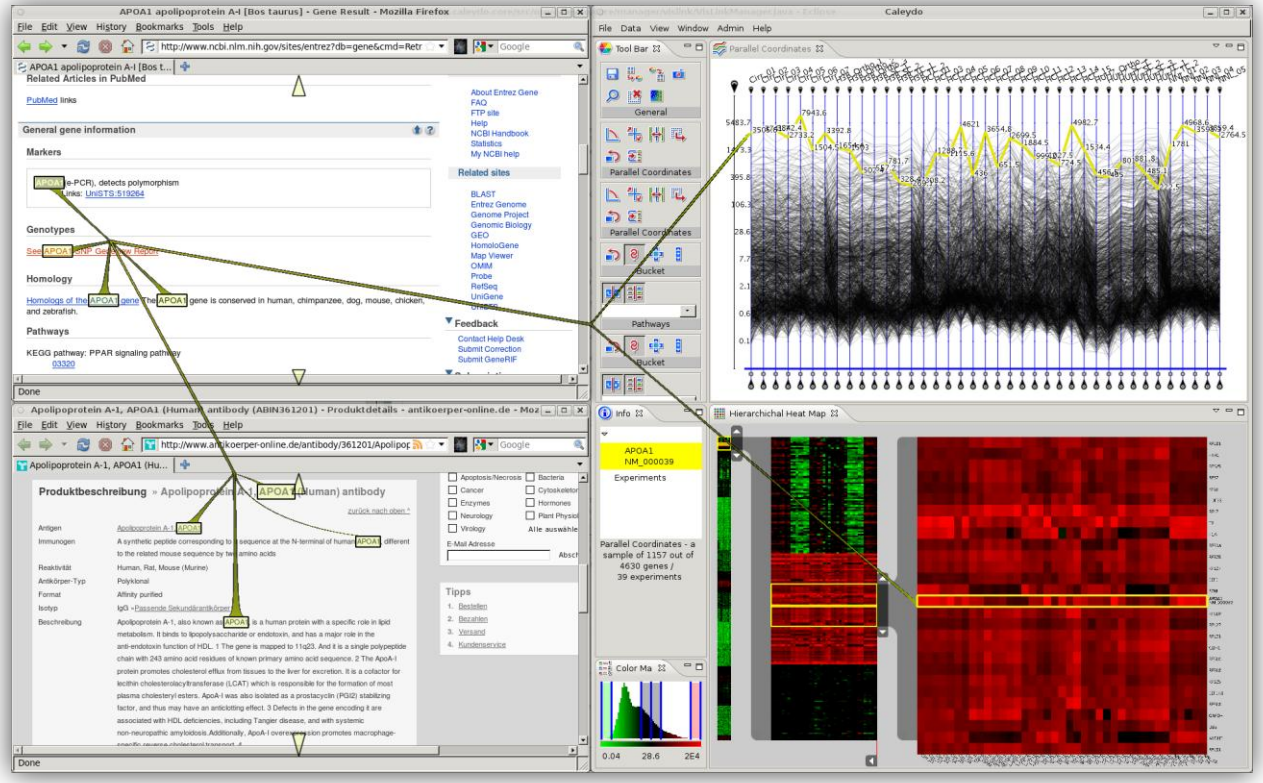
[Waldner, GI 2010] – best paper award

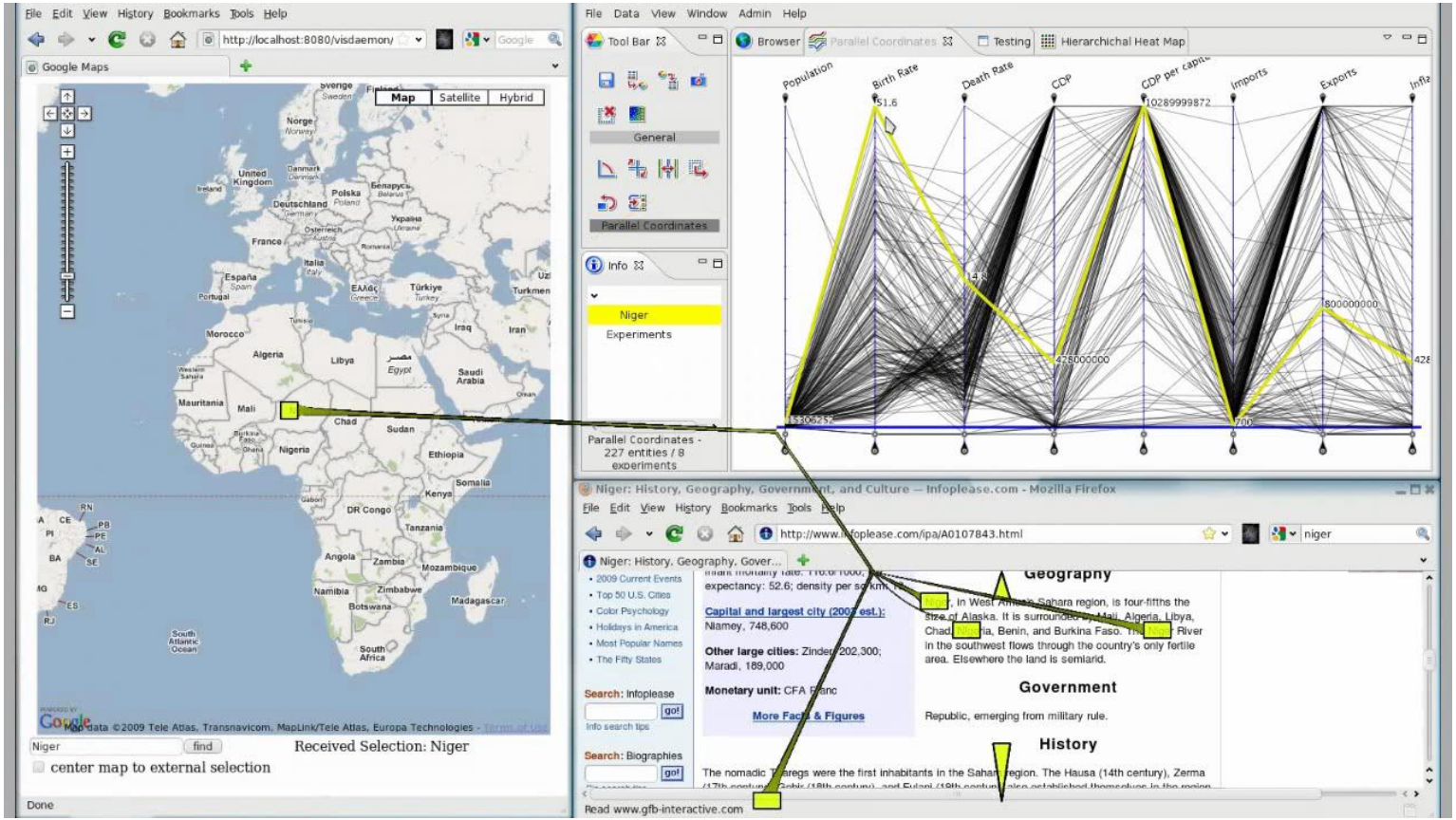
VISUAL LINKING ACROSS APPLICATIONS



CALEYDO

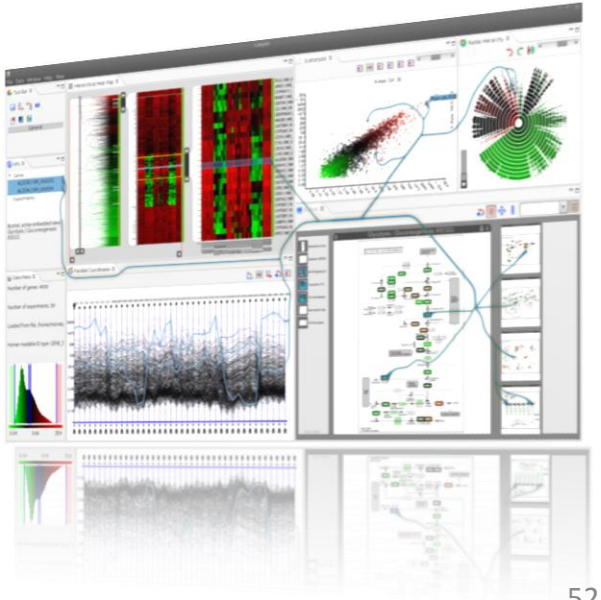
Visual Links Across Applications



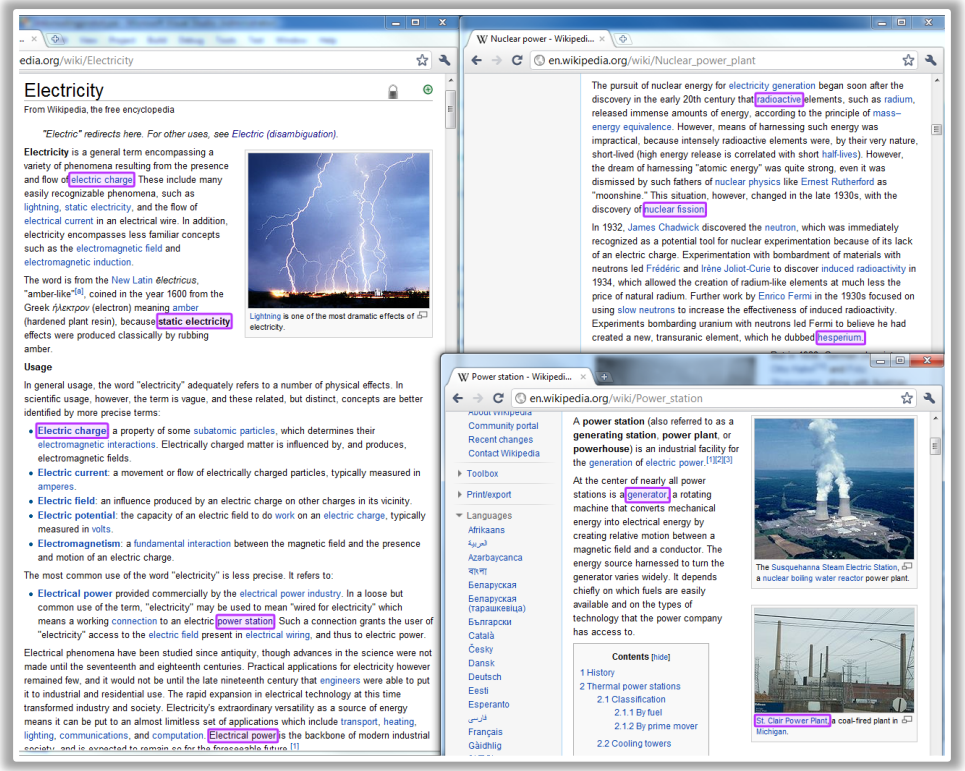

File Edit View History Bookmarks Tools Help
Google Maps
Map
Satellite
Hybrid
Received Selection: Niger
center map to external selection
Done
File Data View Window Admin Help
Tool Bar
General
Parallel Coordinates
Info
Niger
Experiments
Parallel Coordinates - 227 entities / 8 experiments
Browser
Parallel Coordinates
Testing
Hierarchichal Heat Map
Population
Birth Rate
Death Rate
GDP
Imports
Exports
51.6
10289999872
Niger: History, Geography, Government, and Culture — Infoplease.com - Mozilla Firefox
Capital and largest city (2007 est.):
Niamey, 748,600
Other large cities: Zinder 202,300; Maradi, 189,000
Monetary unit: CFA franc
More Facts & Figures
Geography
Government
Republic, emerging from military rule.
History
Read www.gfb-interactive.com

[Steinberger, InfoVis 2011]

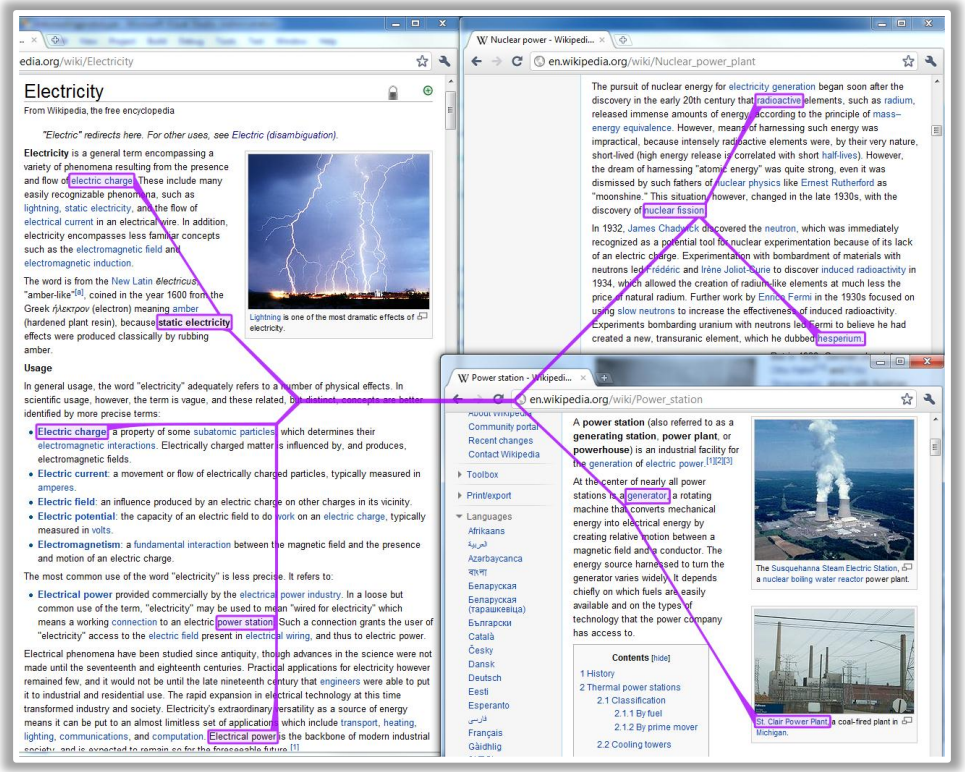
CONTEXT-AWARE VISUAL LINKS



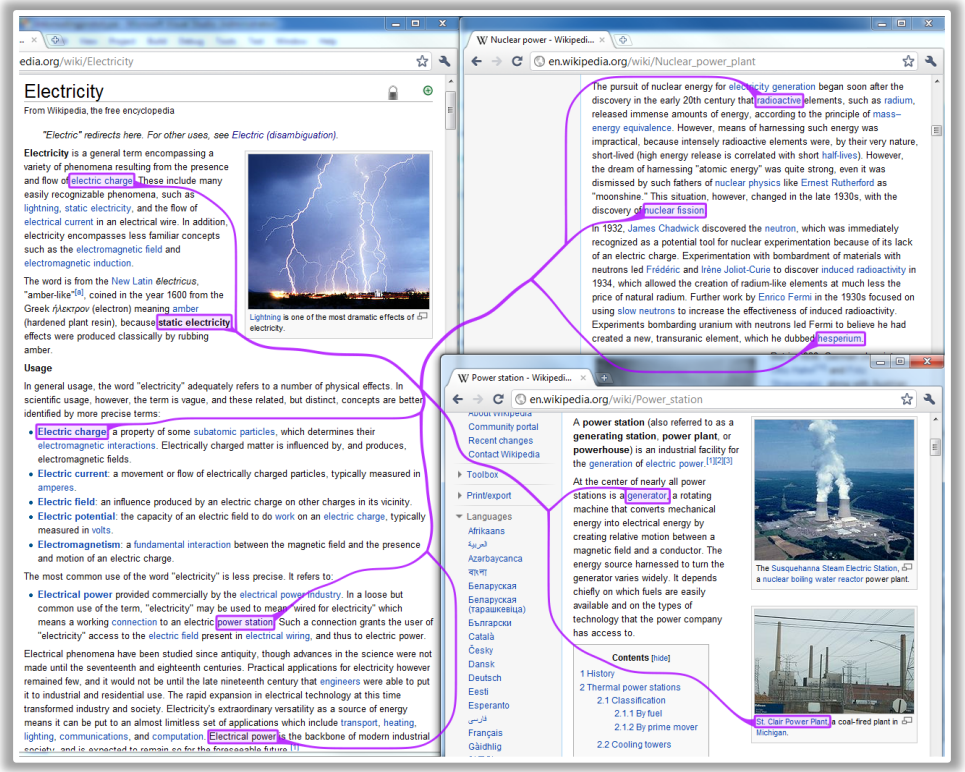
Simple Highlighting



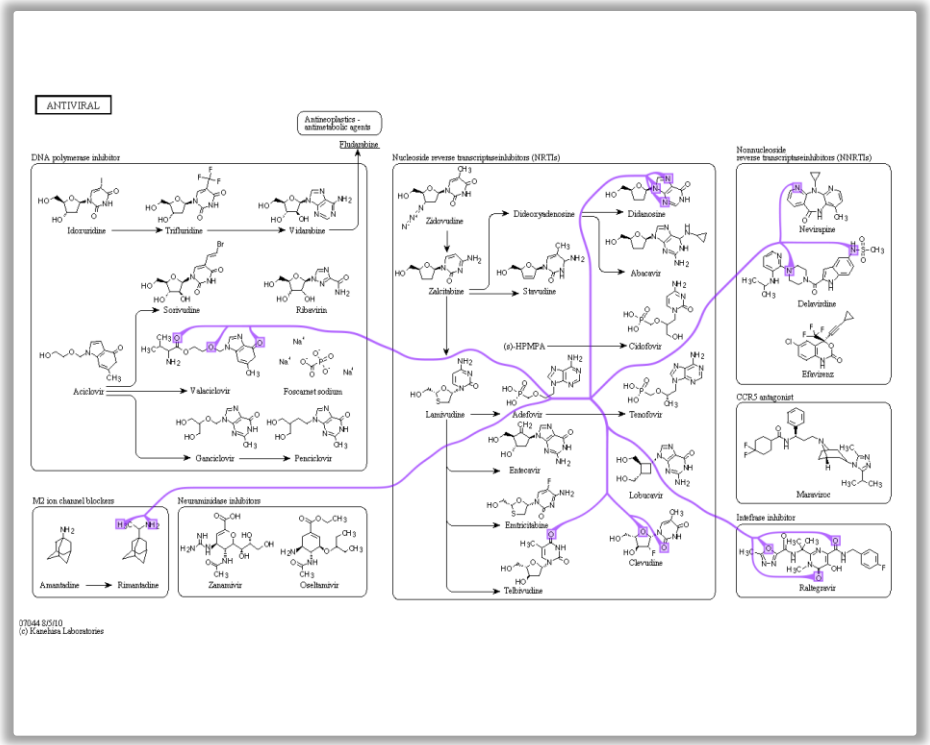
Traditional Visual Links



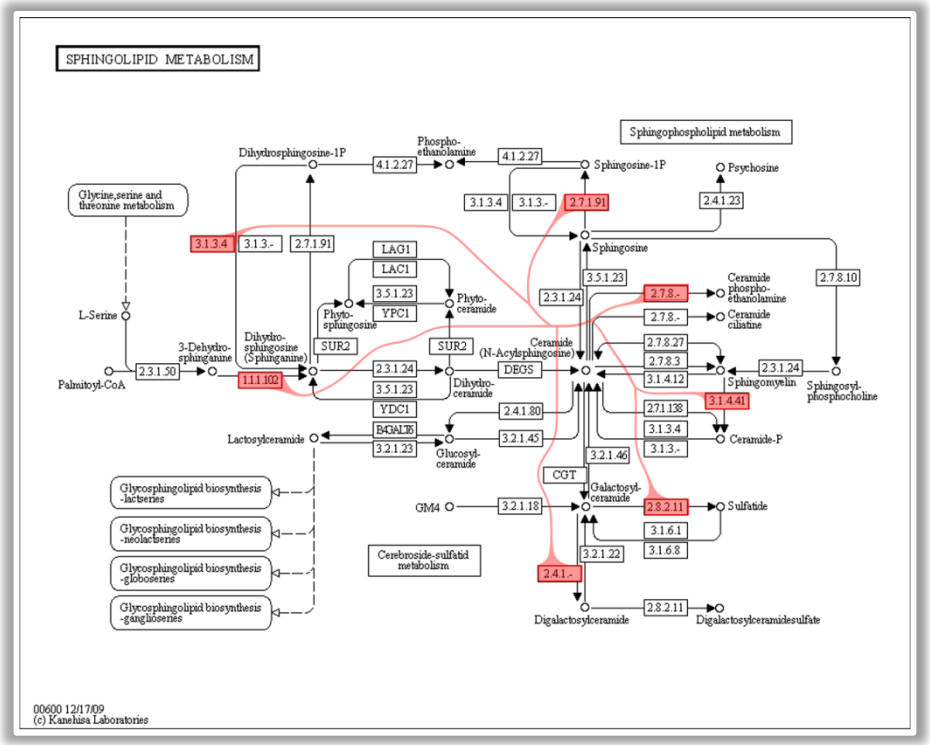
Context-Aware Visual Links



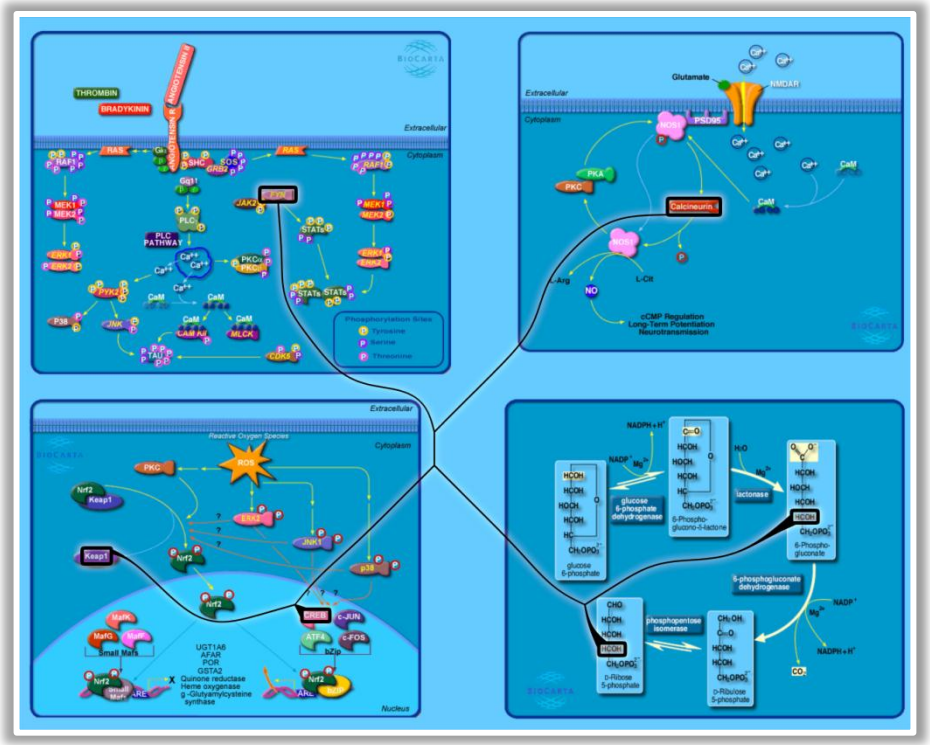
Kegg Pathway



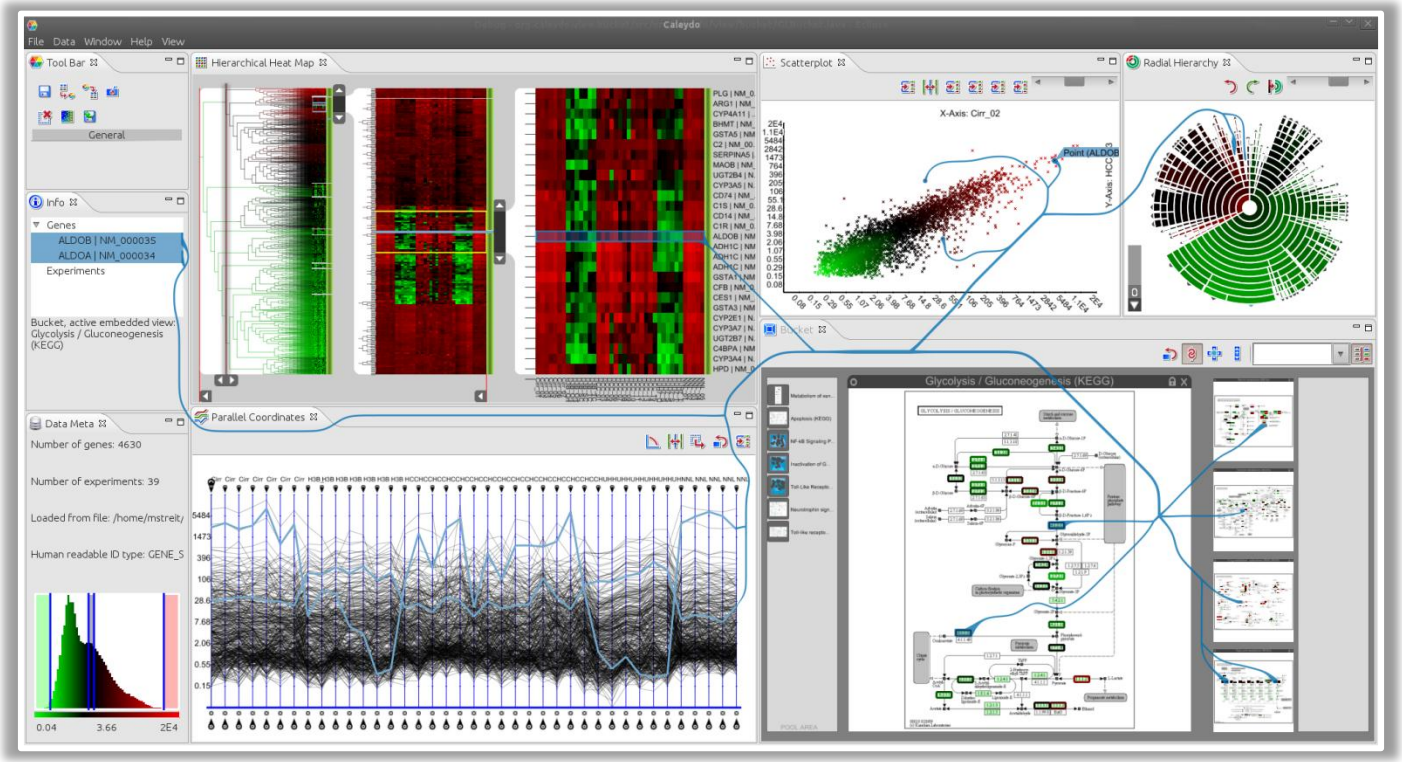
Kegg Pathway



Multiple Biocarta Pathways

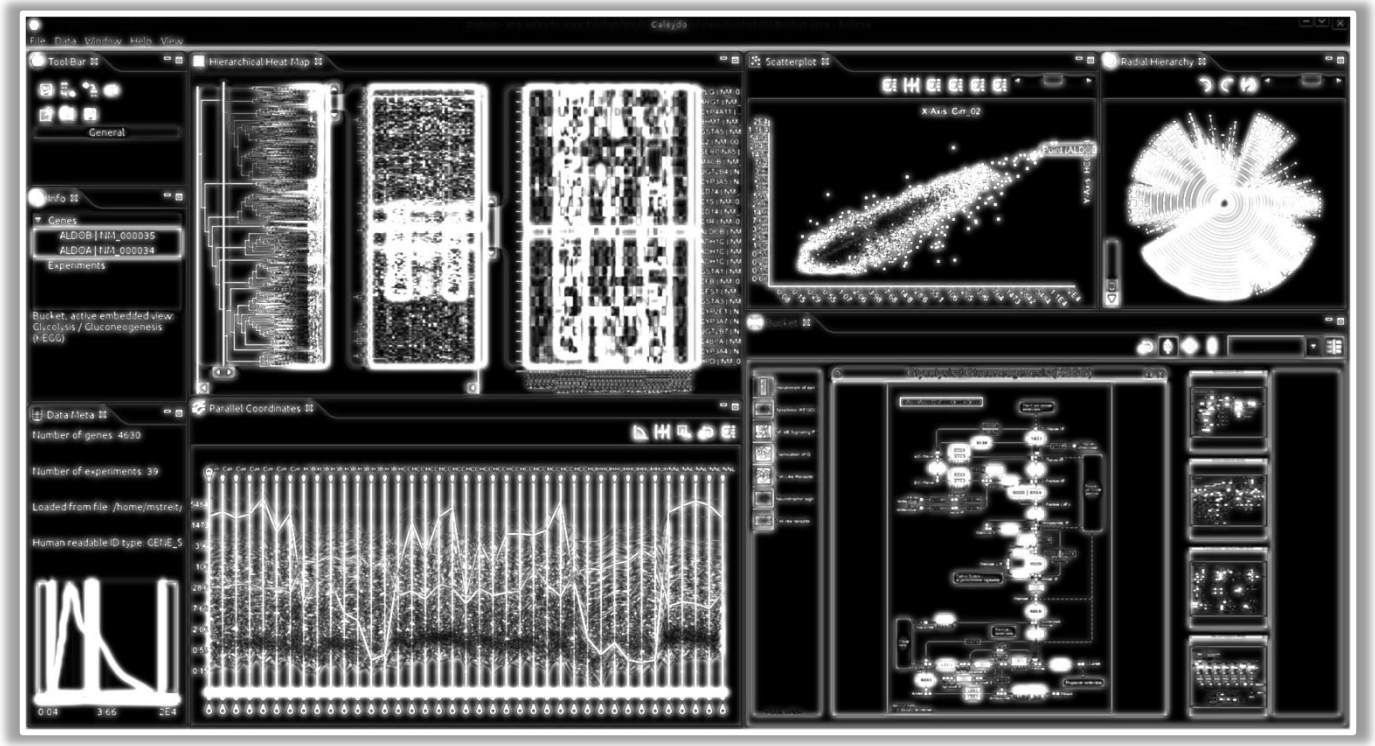


Caleydo

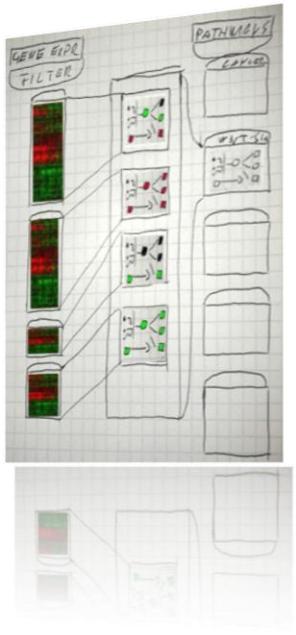


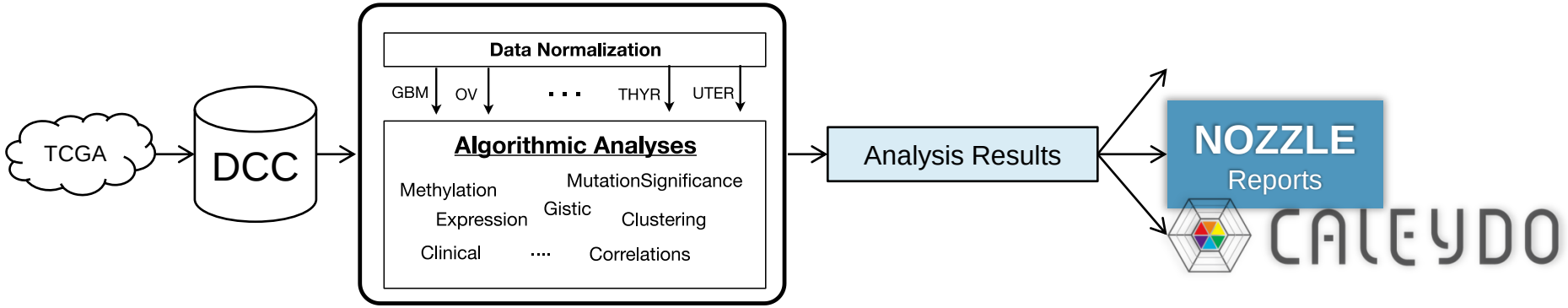
CALEYDO

How it works


Caleydo
File Data Window Help View
Tool Bar
General
Info
Genes
ALDOB | NM_000035
ALDOA | NM_000034
Experiments
Bucket, active embedded view:
Glycolysis / Gluconeogenesis
(4 EGO)
Data Meta
Number of genes: 4630
Number of experiments: 39
Loaded from file: /home/mstreit/
Human readable ID type: GENE_S
Hierarchical Heat Map
Parallel Coordinates
Scatterplot
X Axis: Cnt_02
Radial Hierarchy

PLANNED WORK HERE AT CBMI





Project Title Ideas:

OI – Omics Integrator

Frazzlomics

Wyee

Vision: Integrated Visualization of Multiple Datasets from Firehose

1. Analyze consistency of grouping

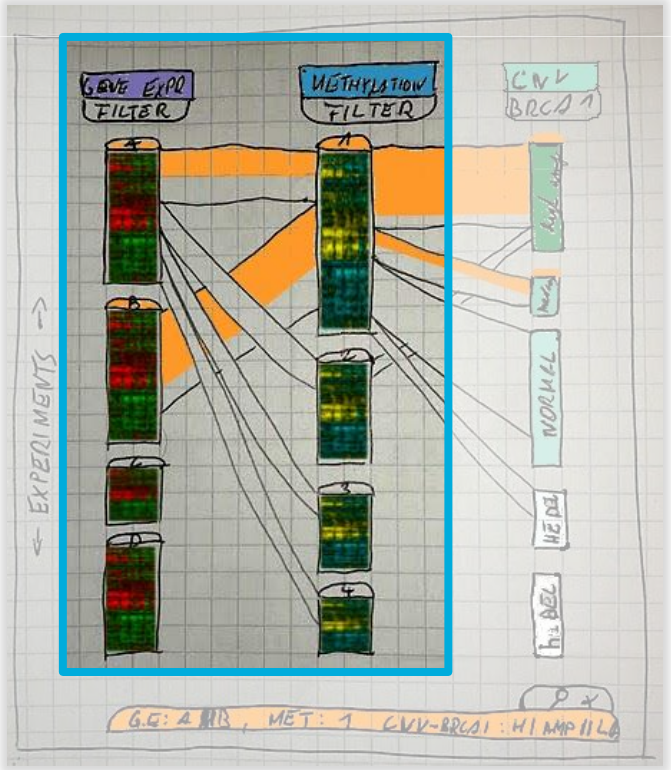
   across datasets

   between different results (e.g. 3 or 4 clusters?)

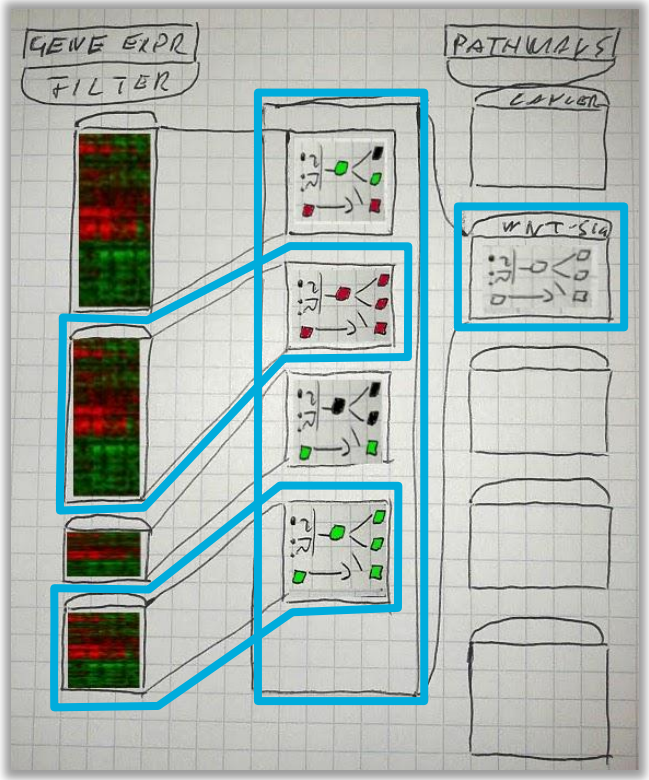
2. Analyze relations between grouping and structural variation data

3. Explore relations between pathways and groups

Showing Relationships



Integrating Pathways



Thank You!
Contributors:

Marc Streit
Prof. Dieter Schmalstieg
Hans-Jörg Schulz
Manuela Waldner
Markus Steinberger
Heidrun Schuhmann
Christian Partl
Thomas Geymayer
Bernhard Schlegl
Werner Puff

Michael Lafer
Jürgen Pillhofer
Prof. Kurt Zatloukal
Karl Kashofer
Helmut Doleisch
Heimo Müller
Stefan Sauer
Willhelm Steiner

CALEYDO