

# Visualizing Biomolecular Data with the Caleydo Framework

**Alexander Lex**

Graz University of Technology, Austria

# 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



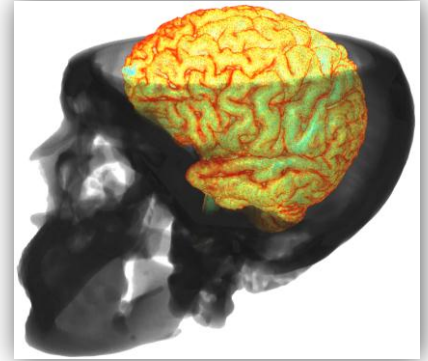
# Research Areas



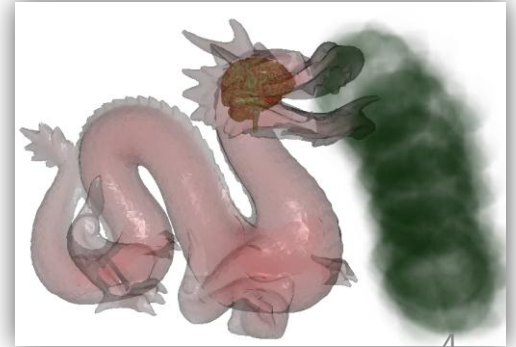
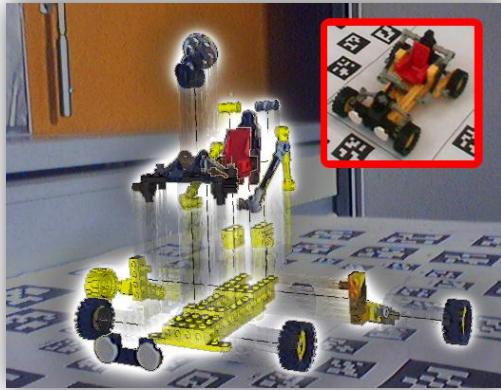
Augmented Reality



Multi-Display Environments



Volume Rendering /  
Medical Visualization



# 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

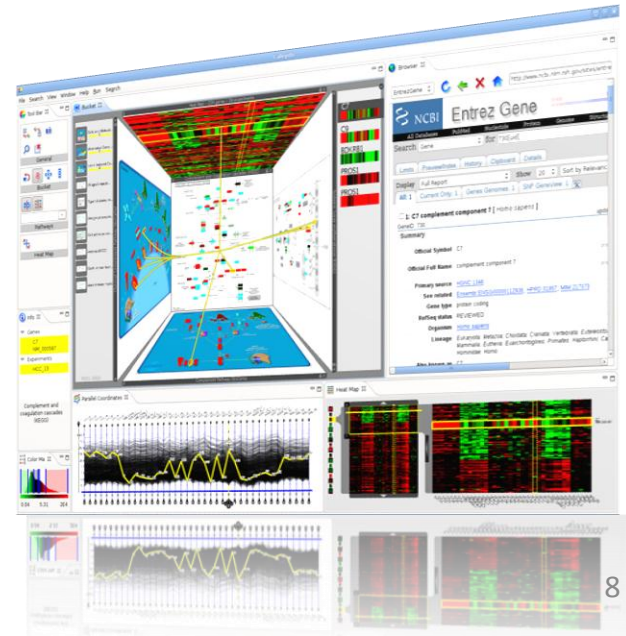
Find **new visualization techniques** and develop software that helps to understand the growing amounts of data in the life sciences

# THE CALEYDO 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)

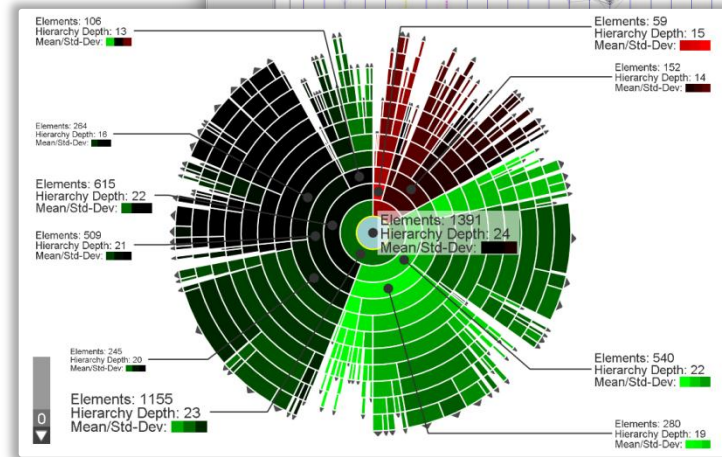
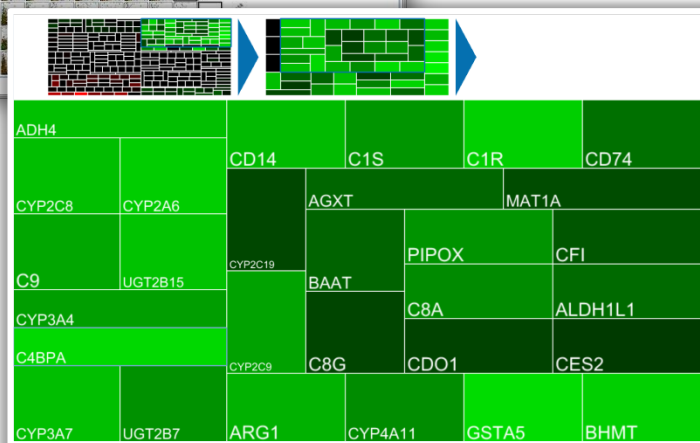
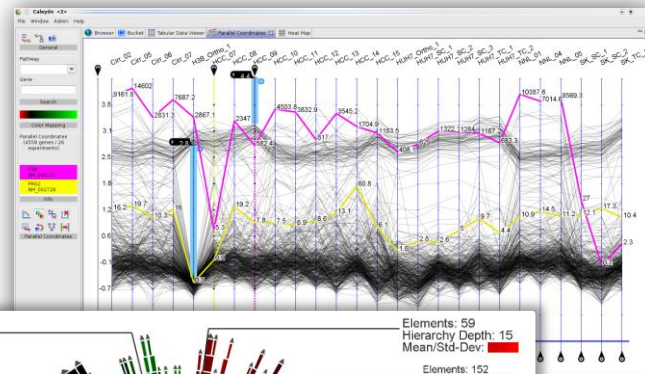
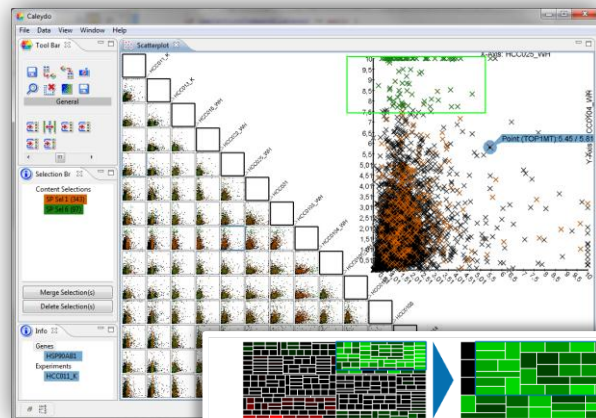
Analytical features:

- R and WEKA interfaces

- Native clustering

- Fold Change, T-Tests, Variance Filters

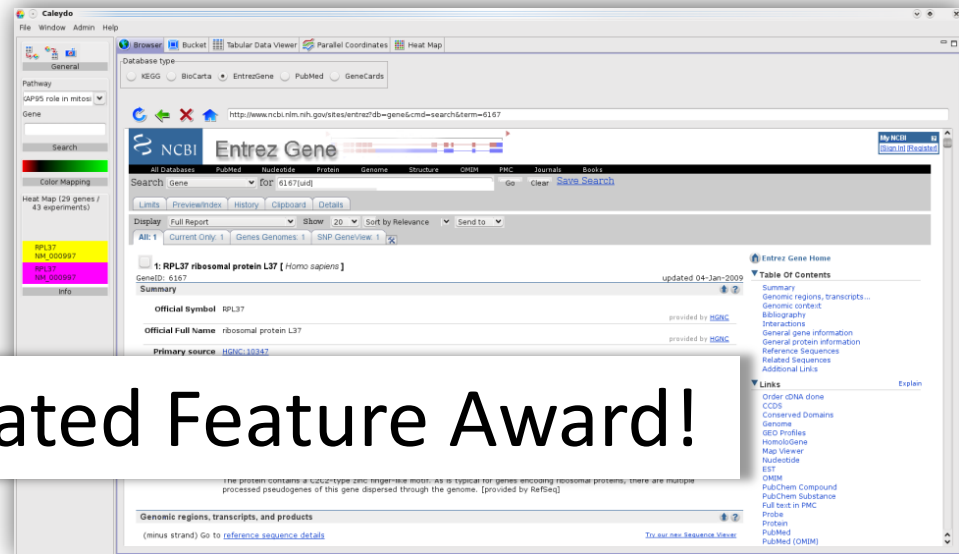
# Numerous Techniques



# Linked Browser

Connected to major medical web databases

PubMed, Entrez,  
KEGG, BioCarta,  
GeneCards...



Most Appreciated Feature Award!



# VISUALIZATION FOR BIOMOLECULAR DATA

Research Topic 1



# Heat Maps

Common representation

Color coded magnitude of  
value

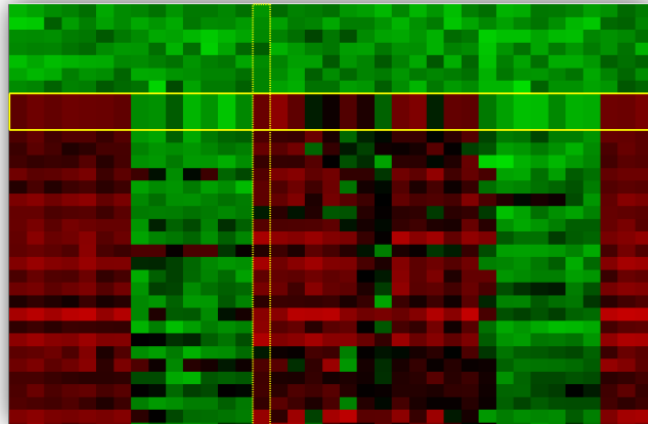
Filtering and **clustering** used

Problem:

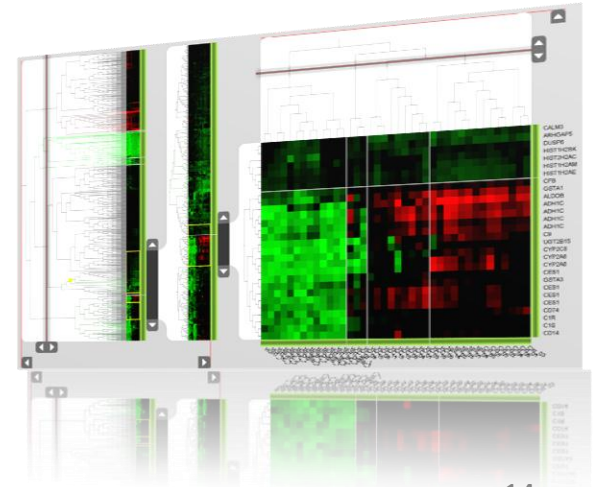
do not scale well

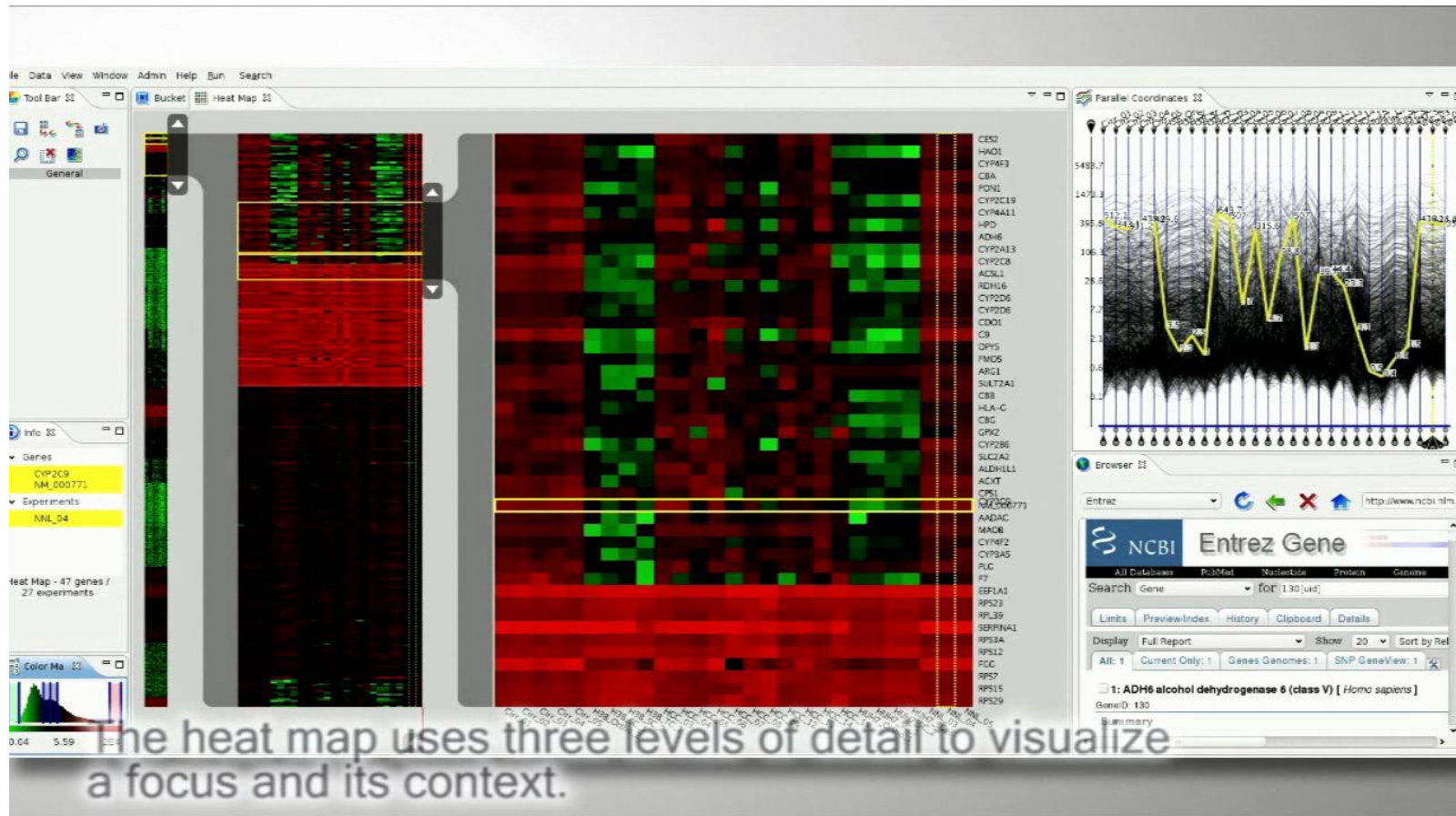
Solution:

**hierarchical approach**



# HIERARCHICAL HEAT MAP





# Works well, but...

Tables often have homogeneous subgroups of dimensions

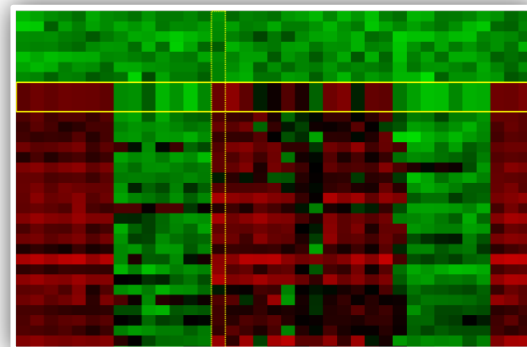
e.g., replicates, time-series, genotypes

Common task for multidimensional data:

Compare those groups

Traditional approach:

Filter, cluster all,  
visualize with Heat Maps!





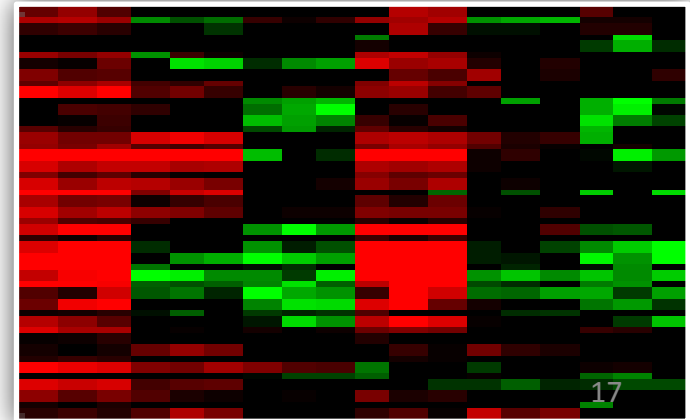
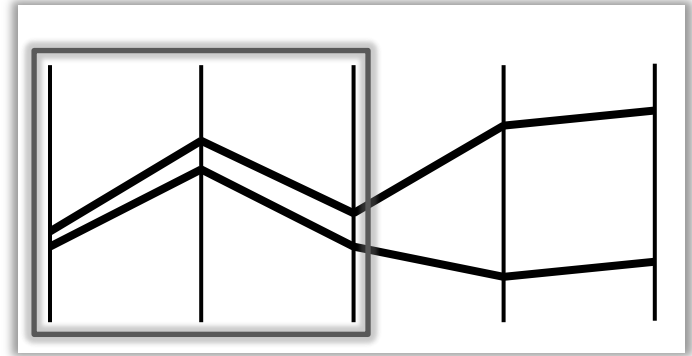
# Clustering Inhomogeneous Data

## The Problem:

Clustering all dimensions at once **obscures relations in homogeneous groups**

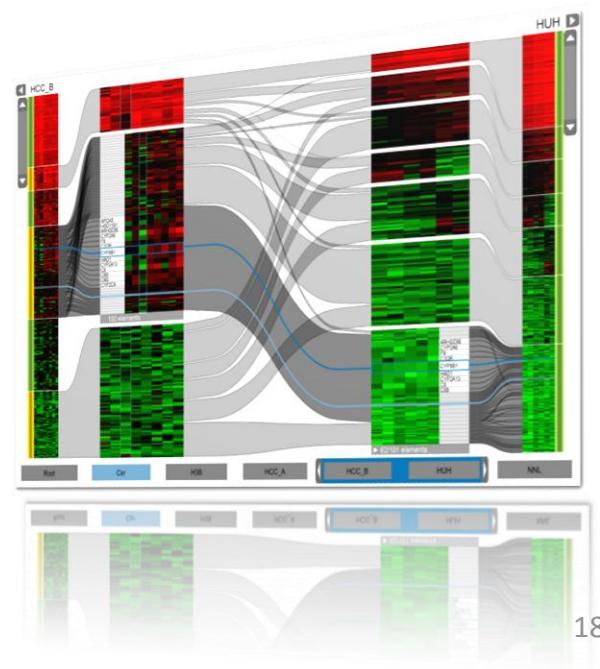
## The Solution:

Divide & Conquer!



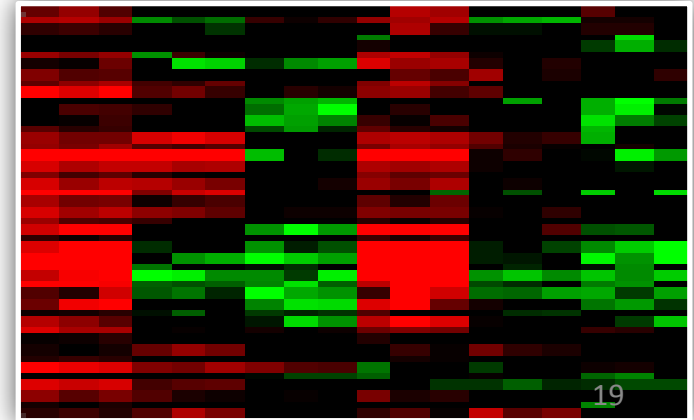
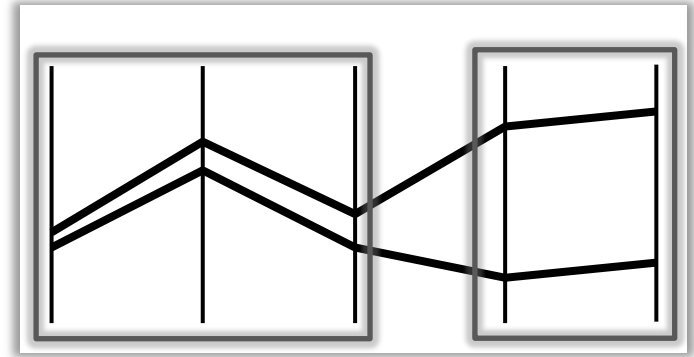
[Lex, InfoVis 2010]

# 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

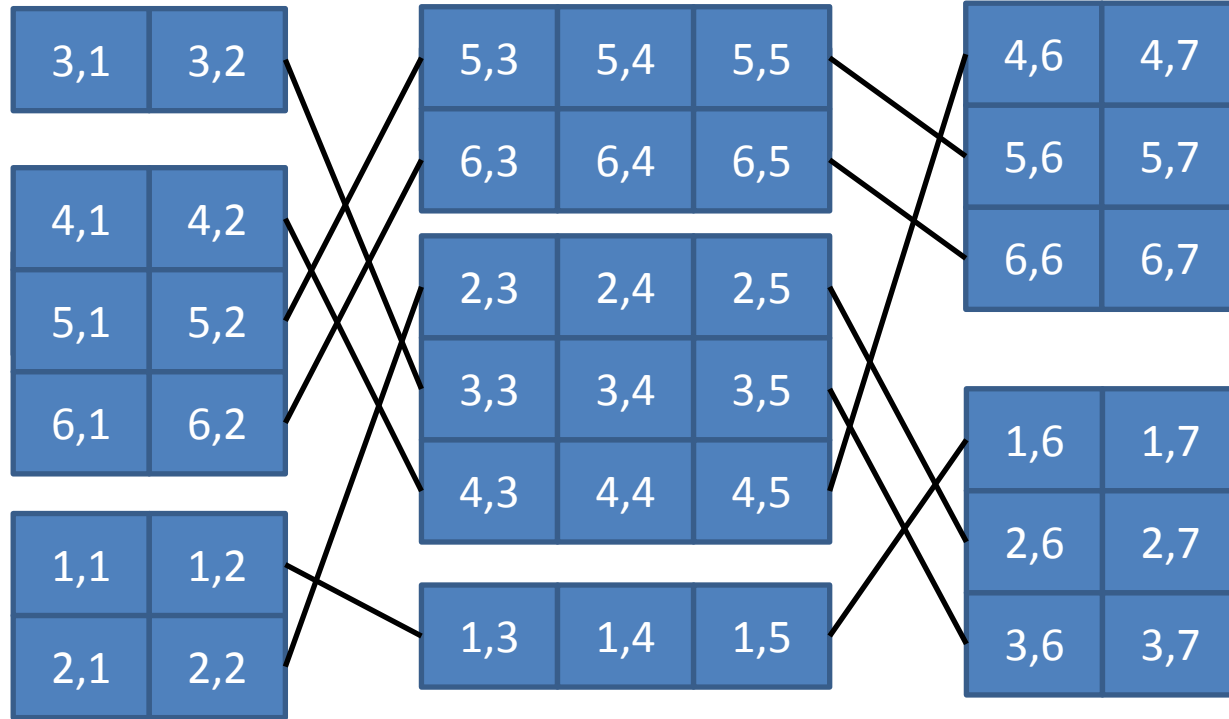
# The Process

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

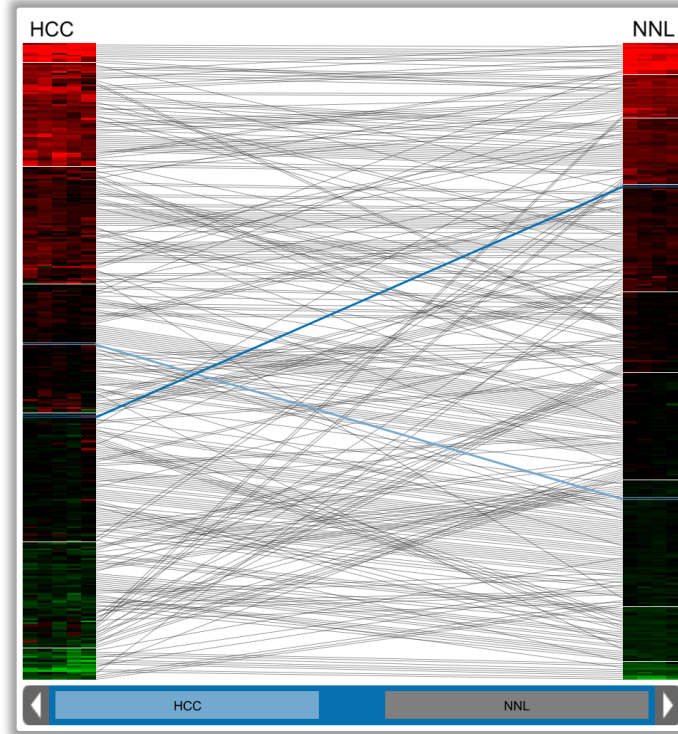
1,3	1,4	1,5
2,3	2,4	2,5
3,3	3,4	3,5
4,3	4,4	4,5
5,3	5,4	5,5
6,3	6,4	6,5

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

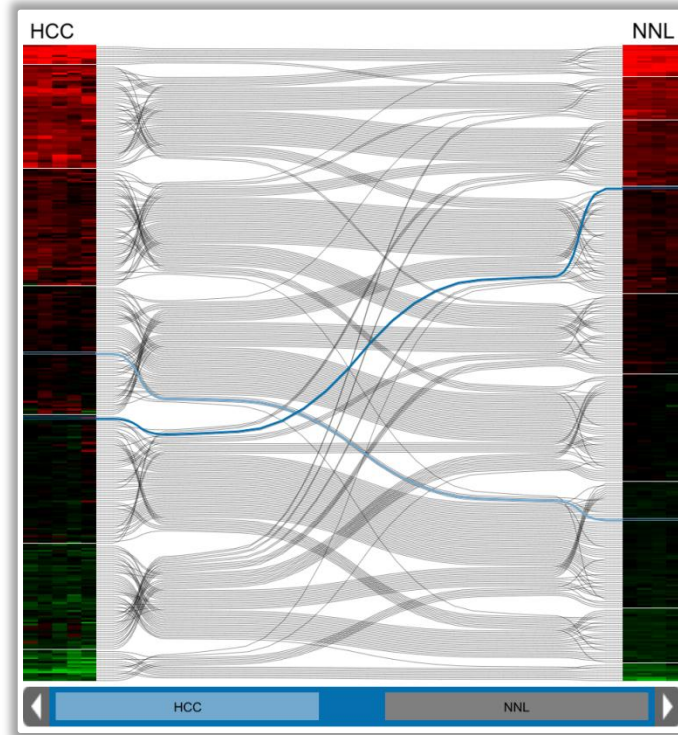


# Immediate Result

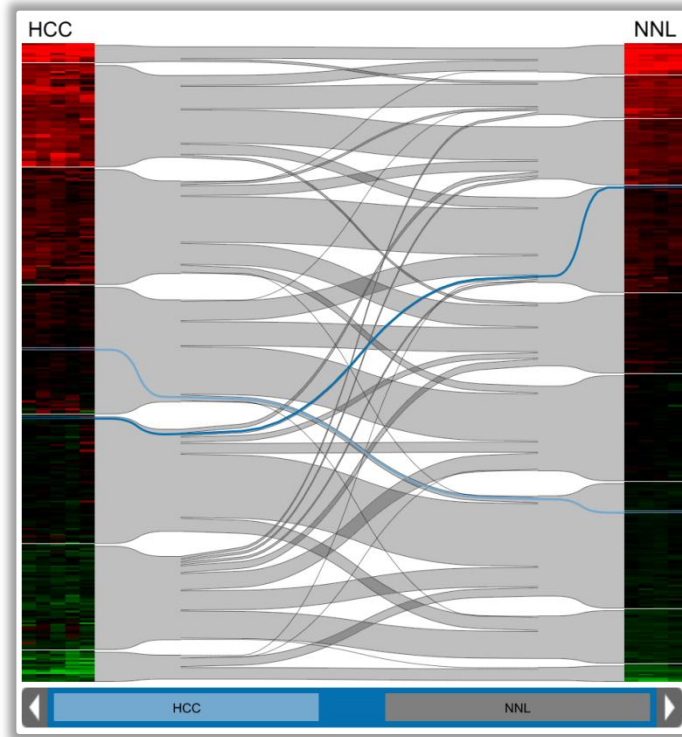


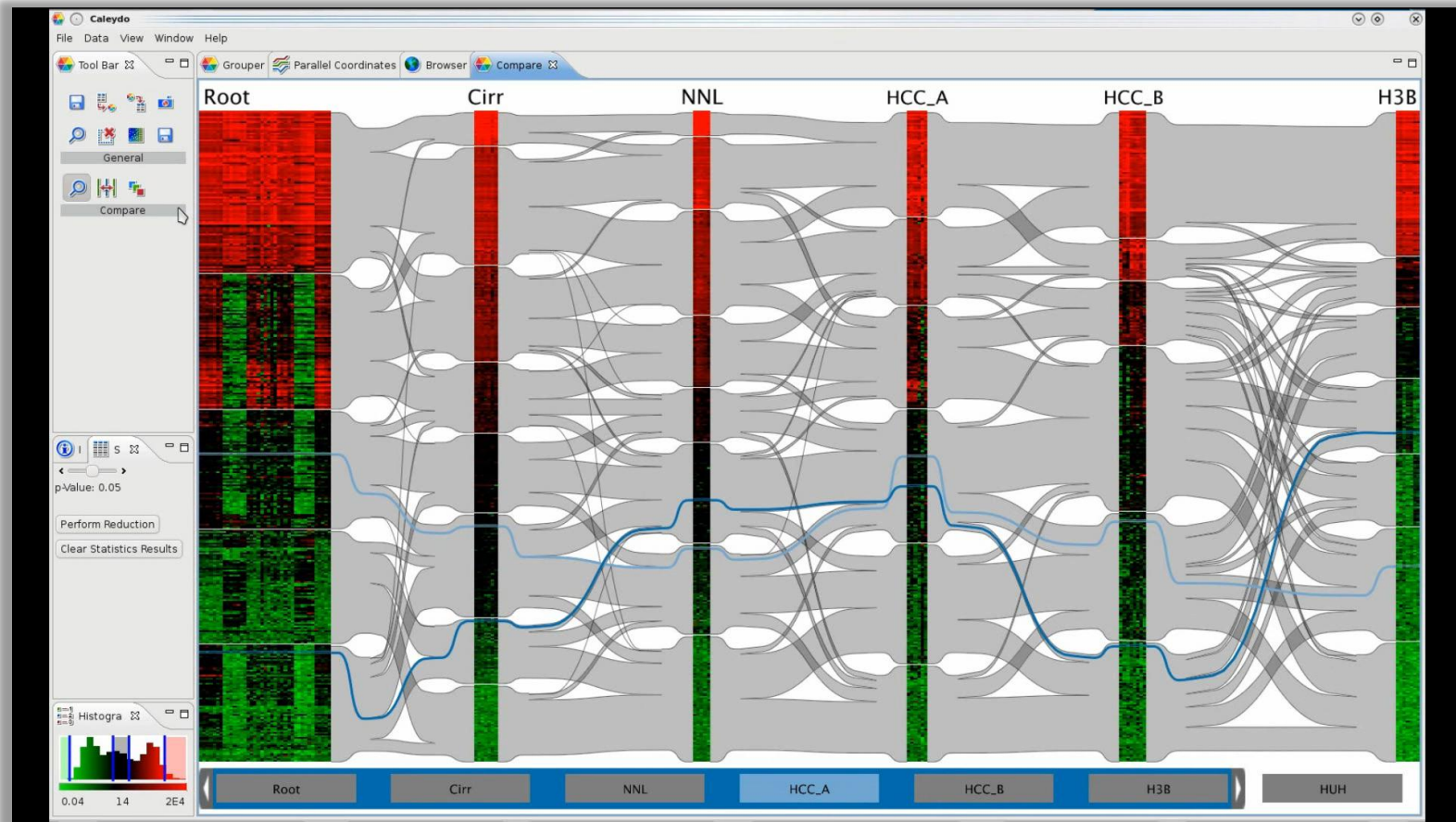


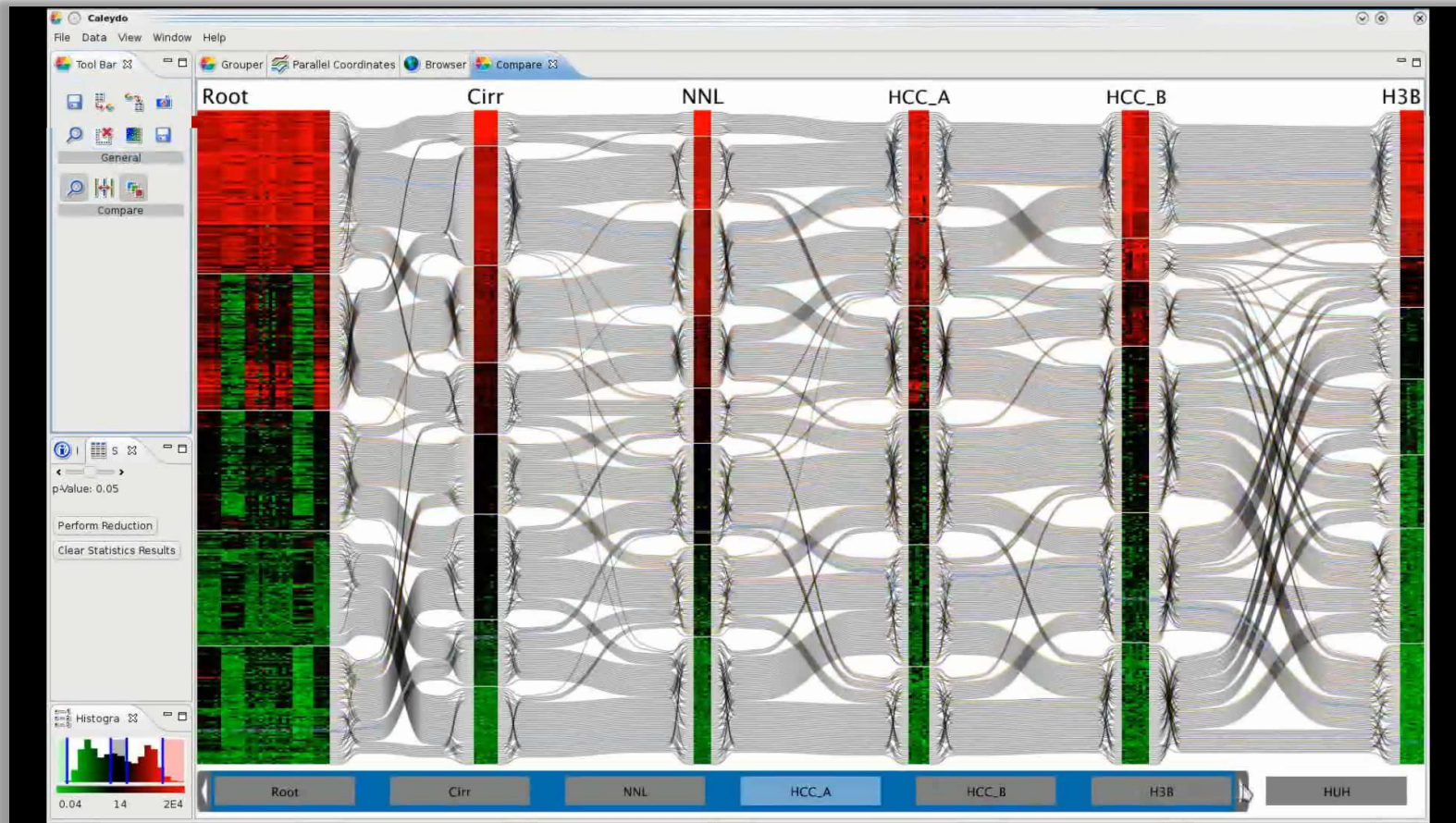
# Using Splines



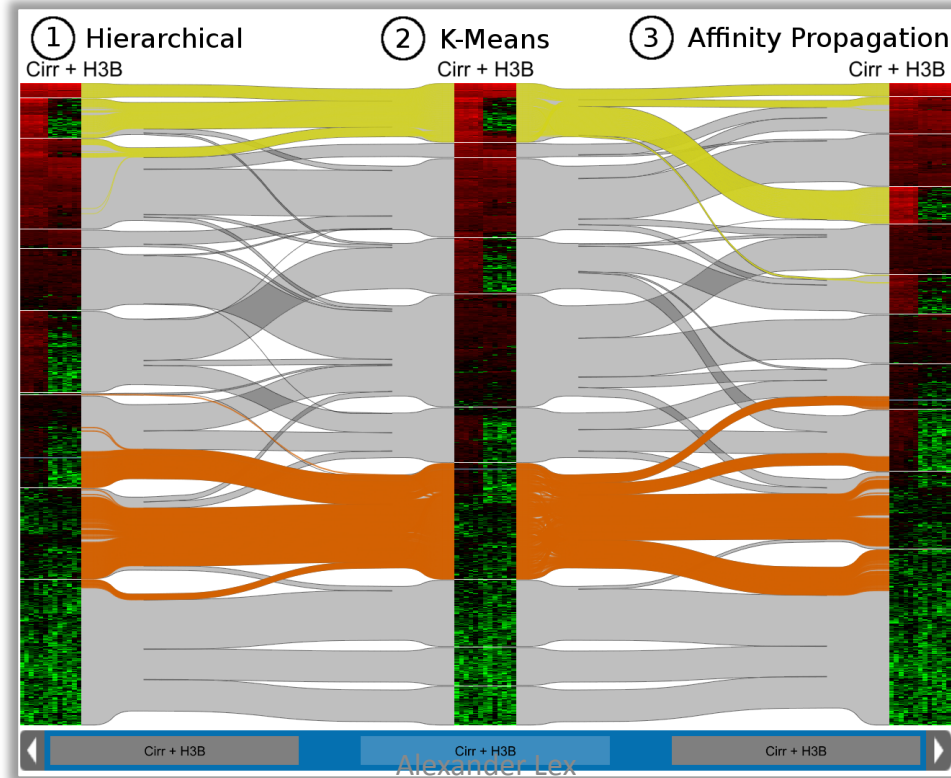
# Using Pipes







# Example: Cluster Algorithm Comparison



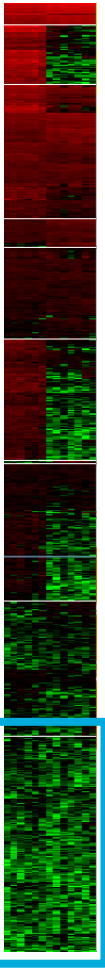


# The next step: Abstraction

(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**

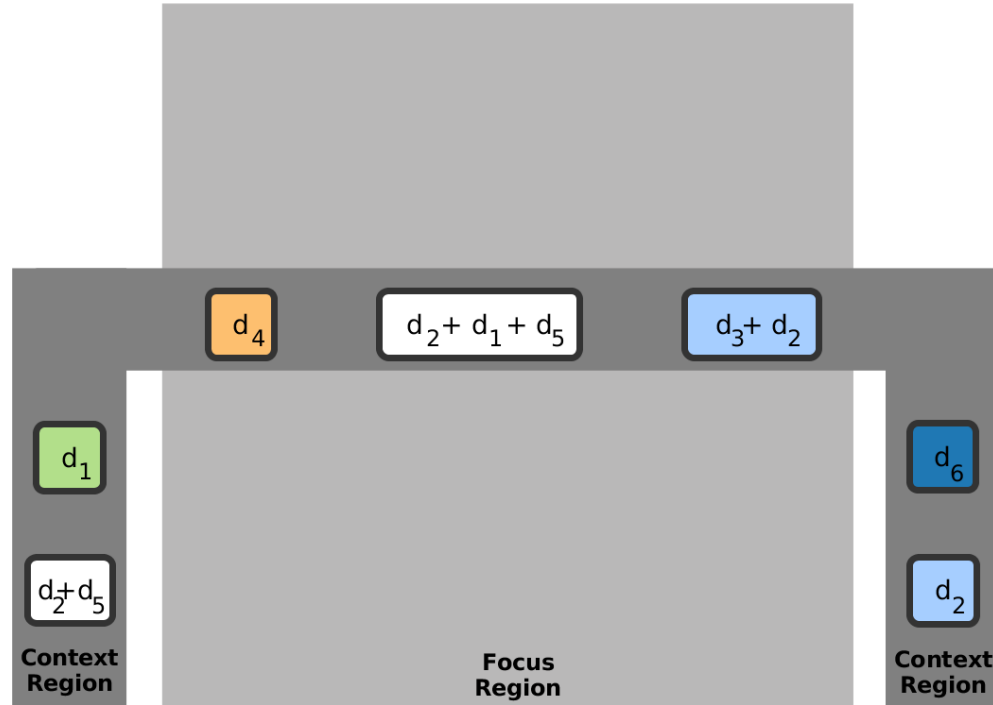
Two types of bricks:

- Dimension** bricks

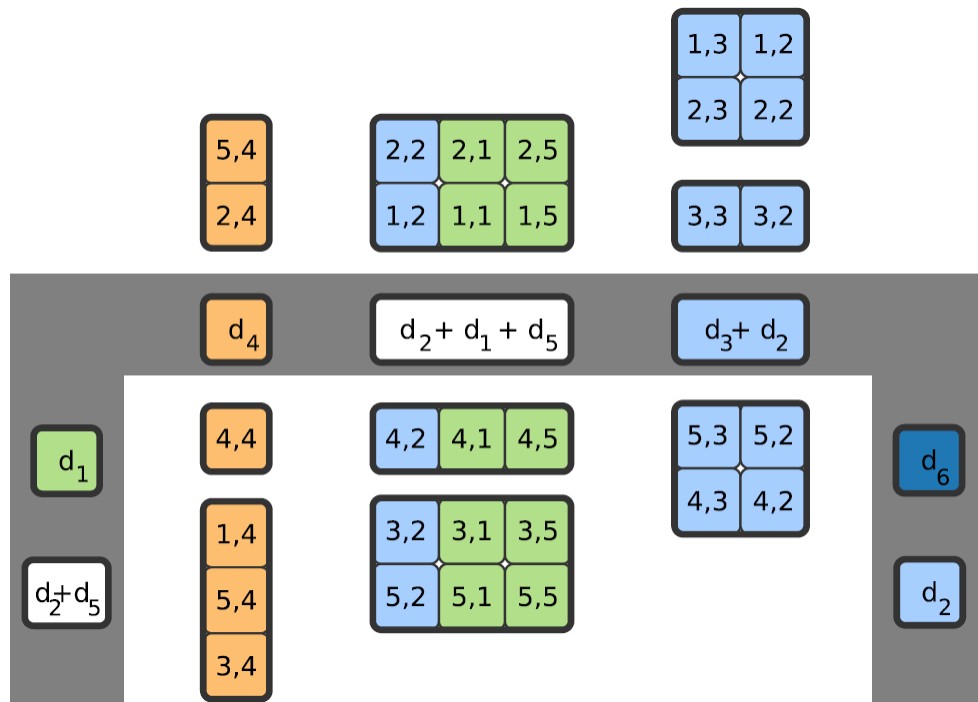
- Cluster** bricks



# Arrangement, Dimensions Bricks



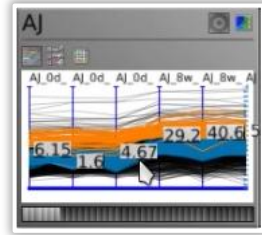
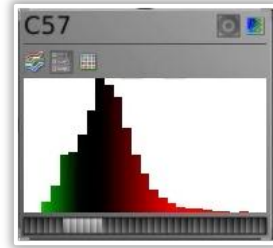
# Added Cluster Bricks

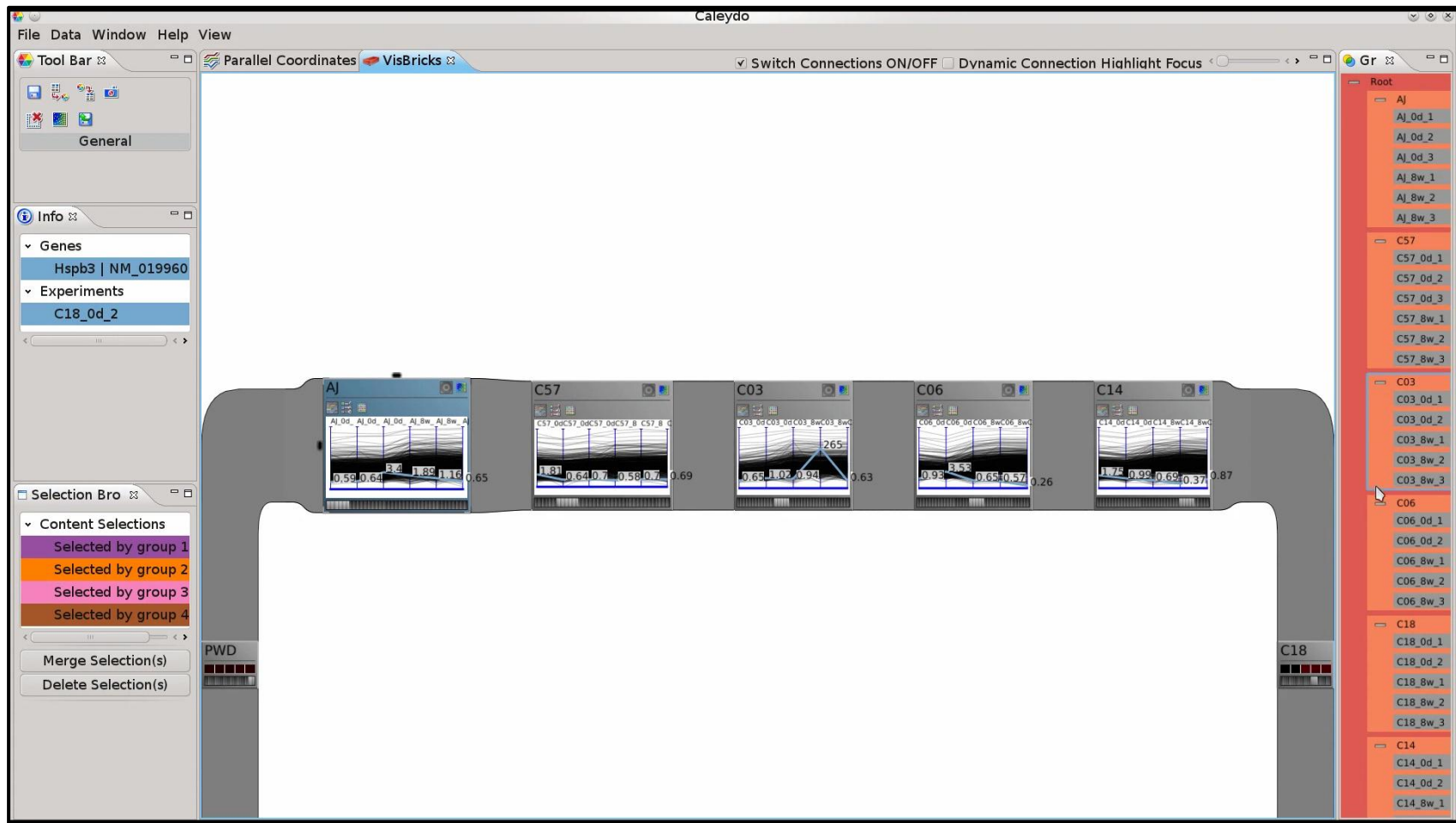


# Examples for Bricks

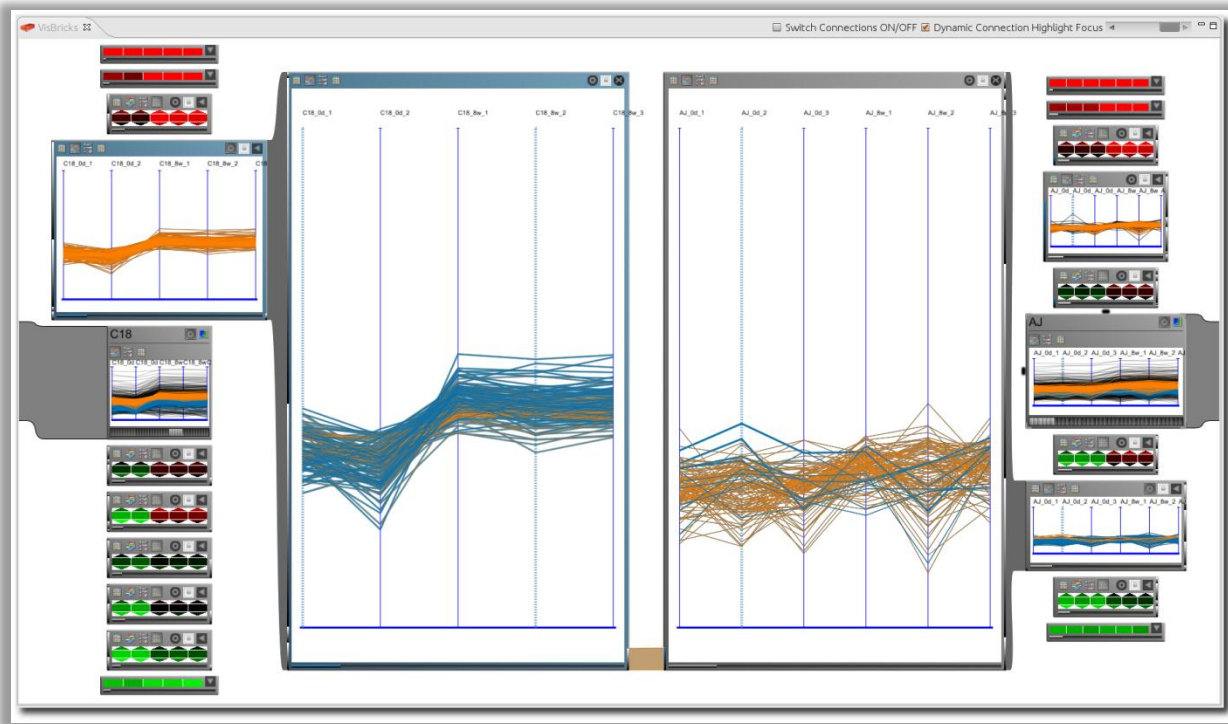
Dimension Bricks

Cluster Bricks





# Exploring Details

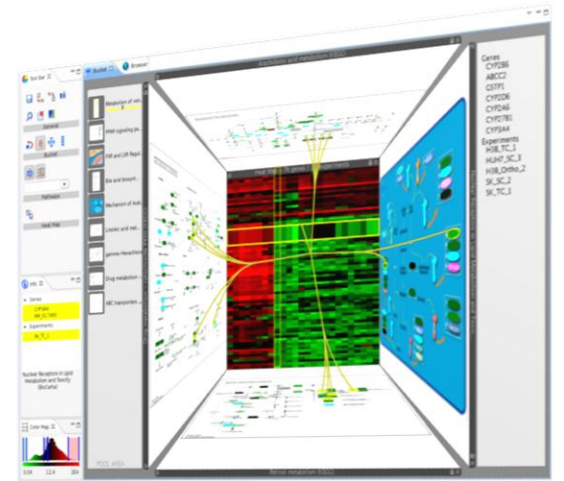


[Lex, PacificVis 2010]

# PATHWAY EXPLORATION CALEYDO BUCKET

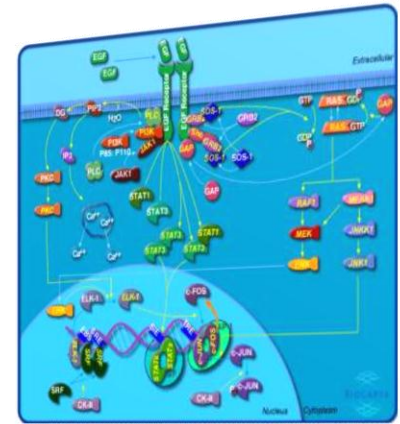


Alexander Lex

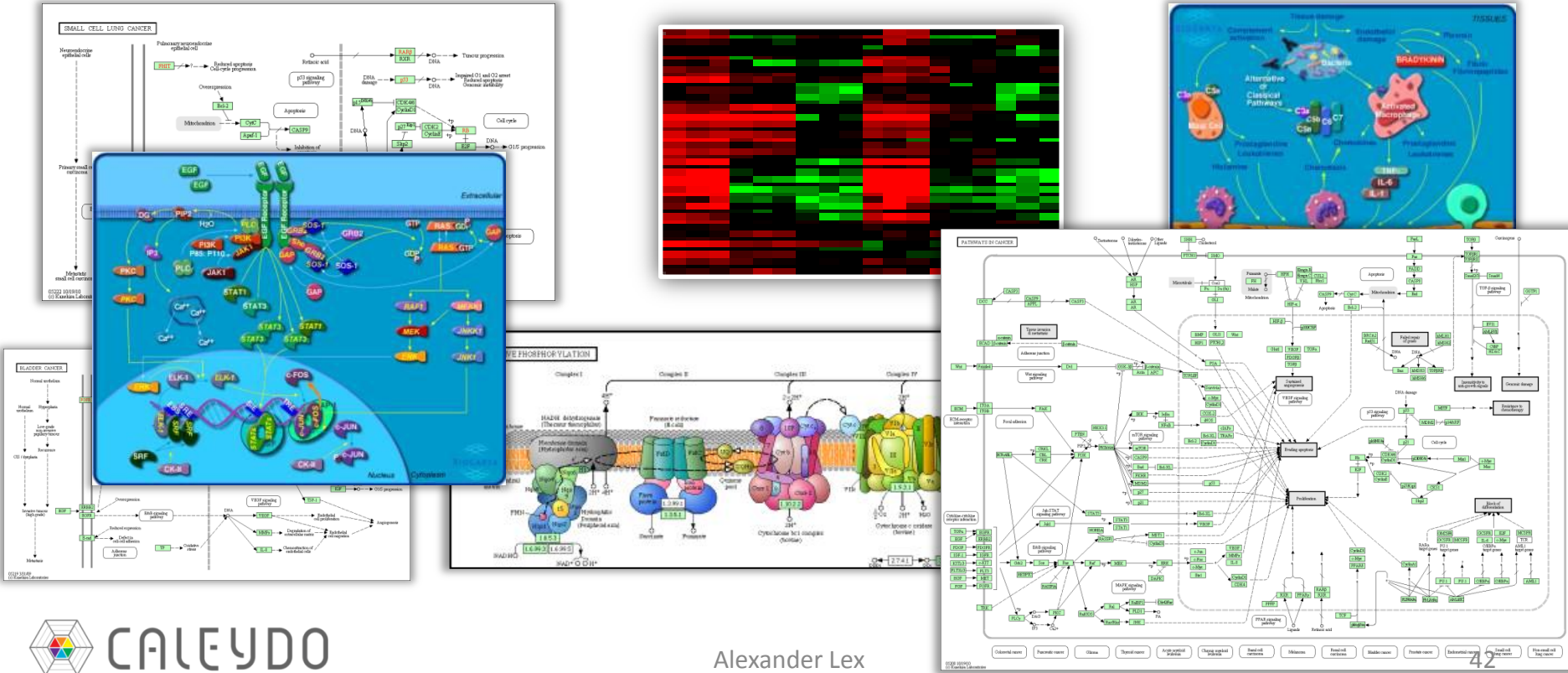


# Pathway Visualization Goals

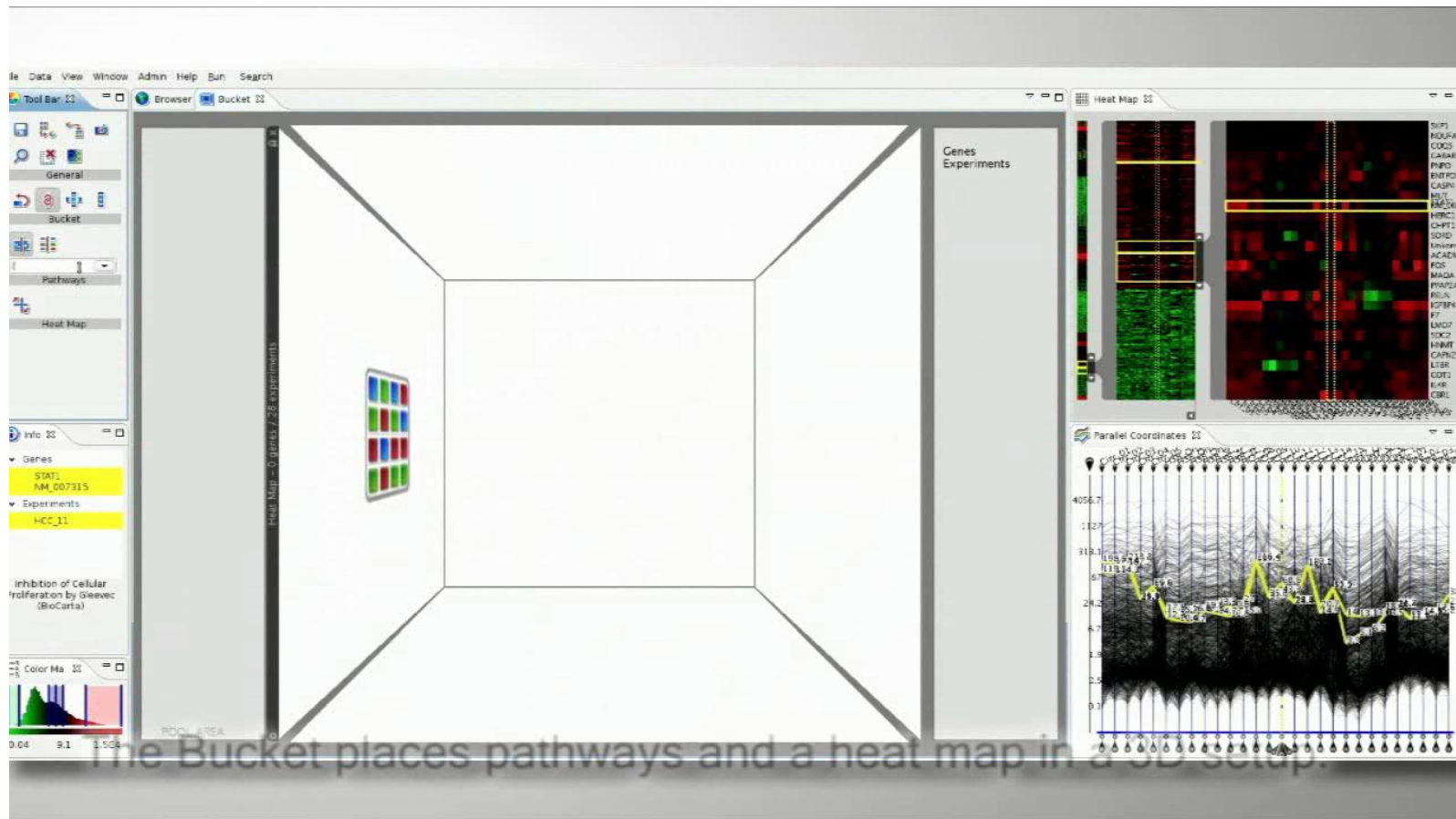
1. Map gene expression of multiple samples
2. Show the relations between pathways



# How to manage all these?





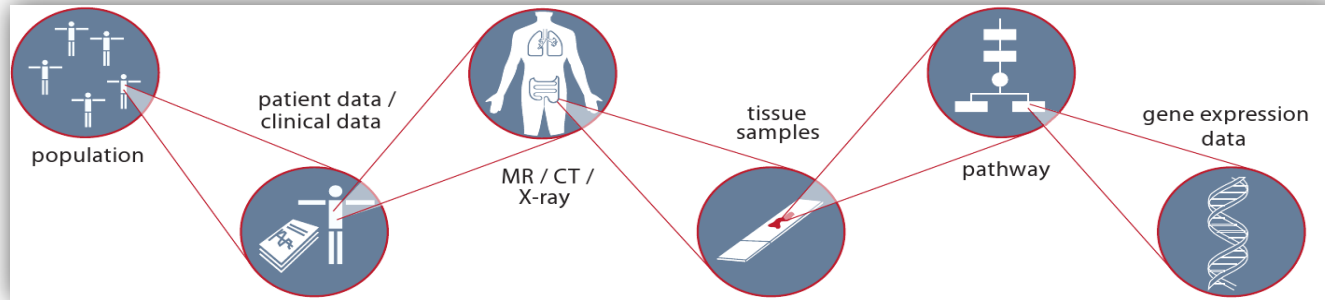


# CONNECTING HETEROGENEOUS DATA SETS

Research Topic 2

# Heterogeneous Data Analysis

Massive amounts of data  
Different abstraction levels  
Different sources  
Different formats

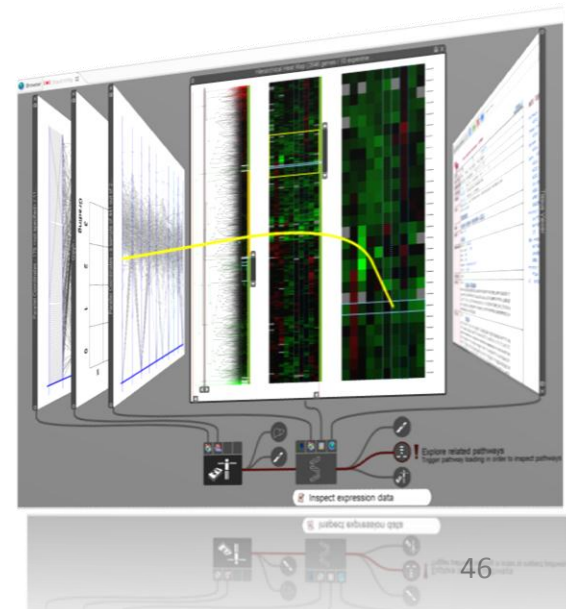


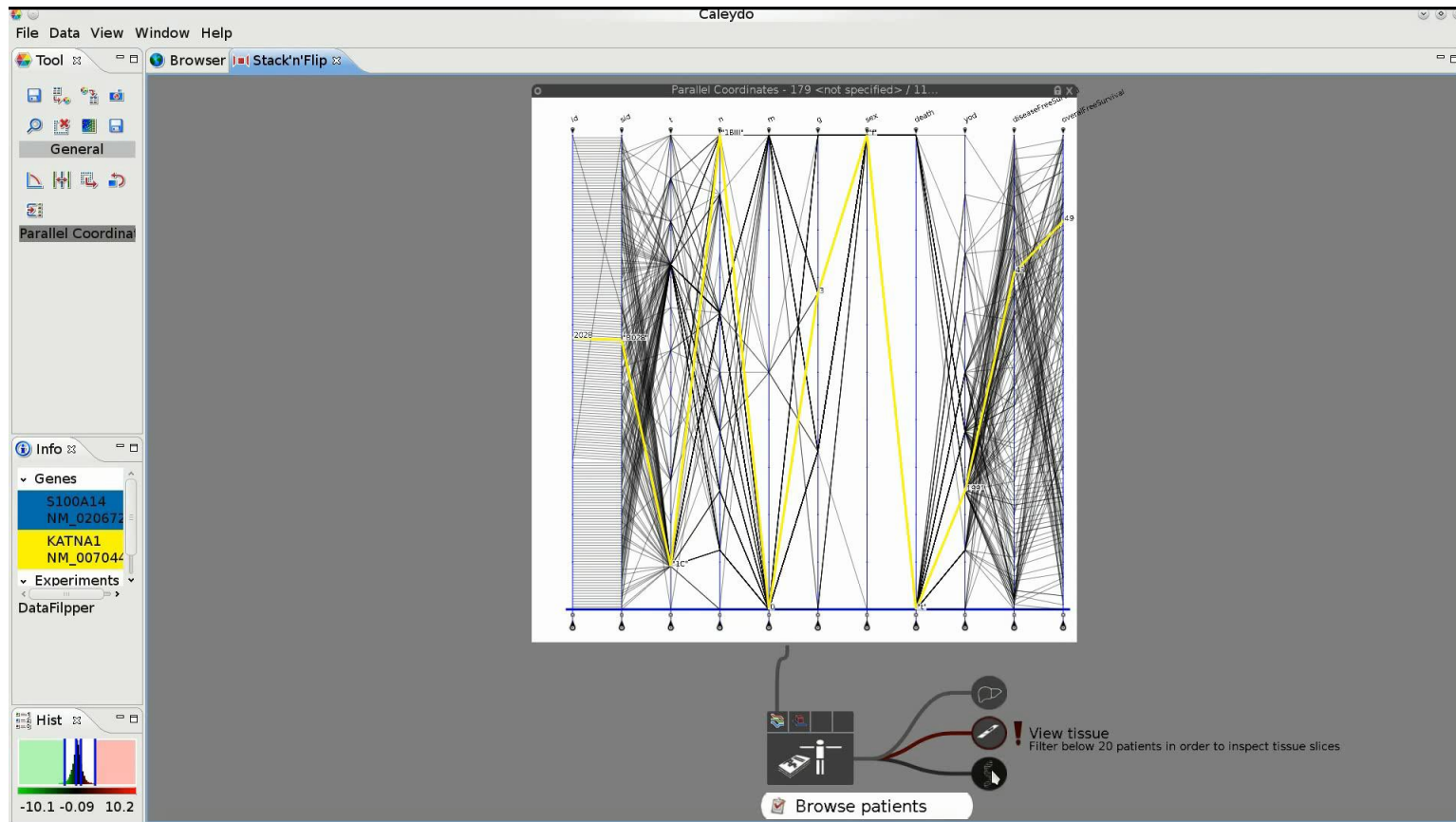
[Streit, CoVis 2009]

**How to integrate data in a seamless analysis session?**

[Streit, TVCG 2011]

# GUIDANCE ACROSS MULTIPLE DATA SOURCES





# Super Application?

Expects a **Super Application** that can visualize everything

Not Feasible! Solution: use existing applications

Downsides:

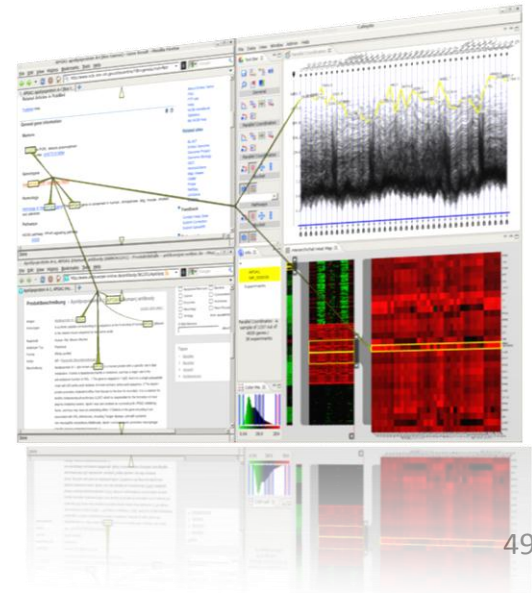
- not integrated

- no highlighting, linking, etc.

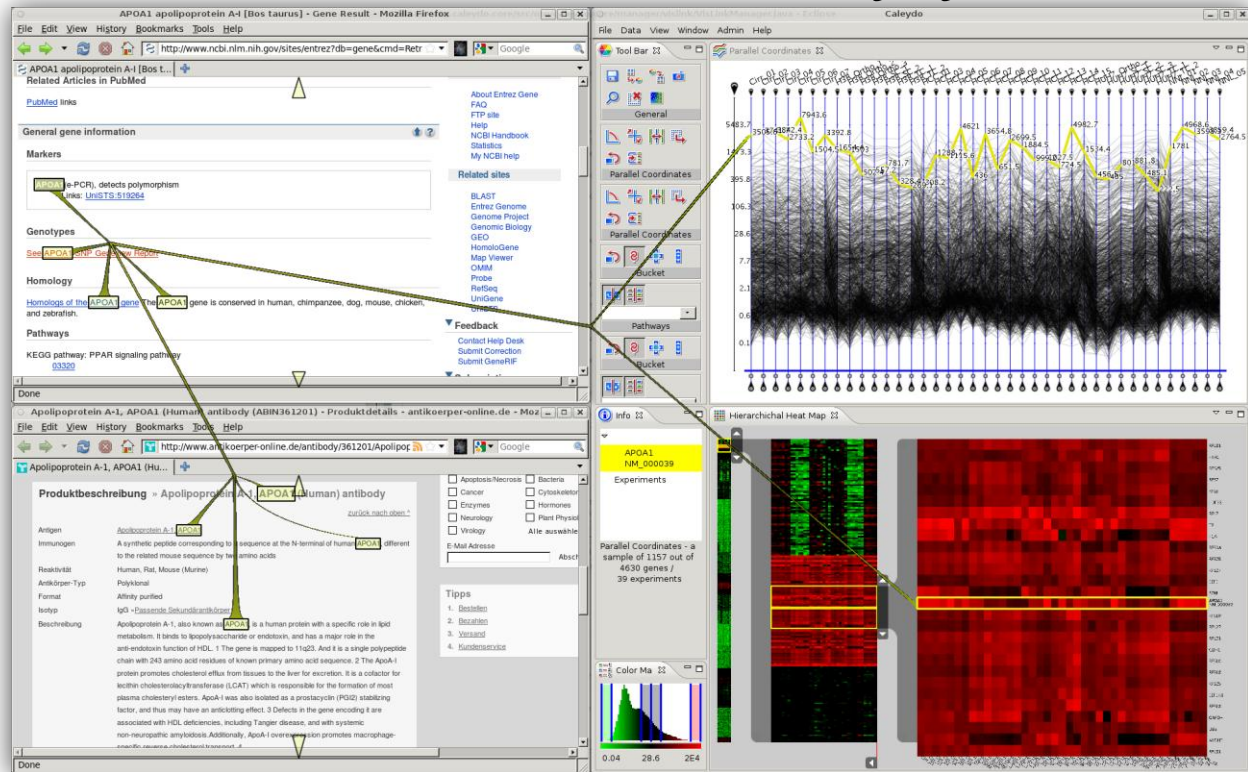
Can we solve this?

[Waldner, GI 2010] – best paper award

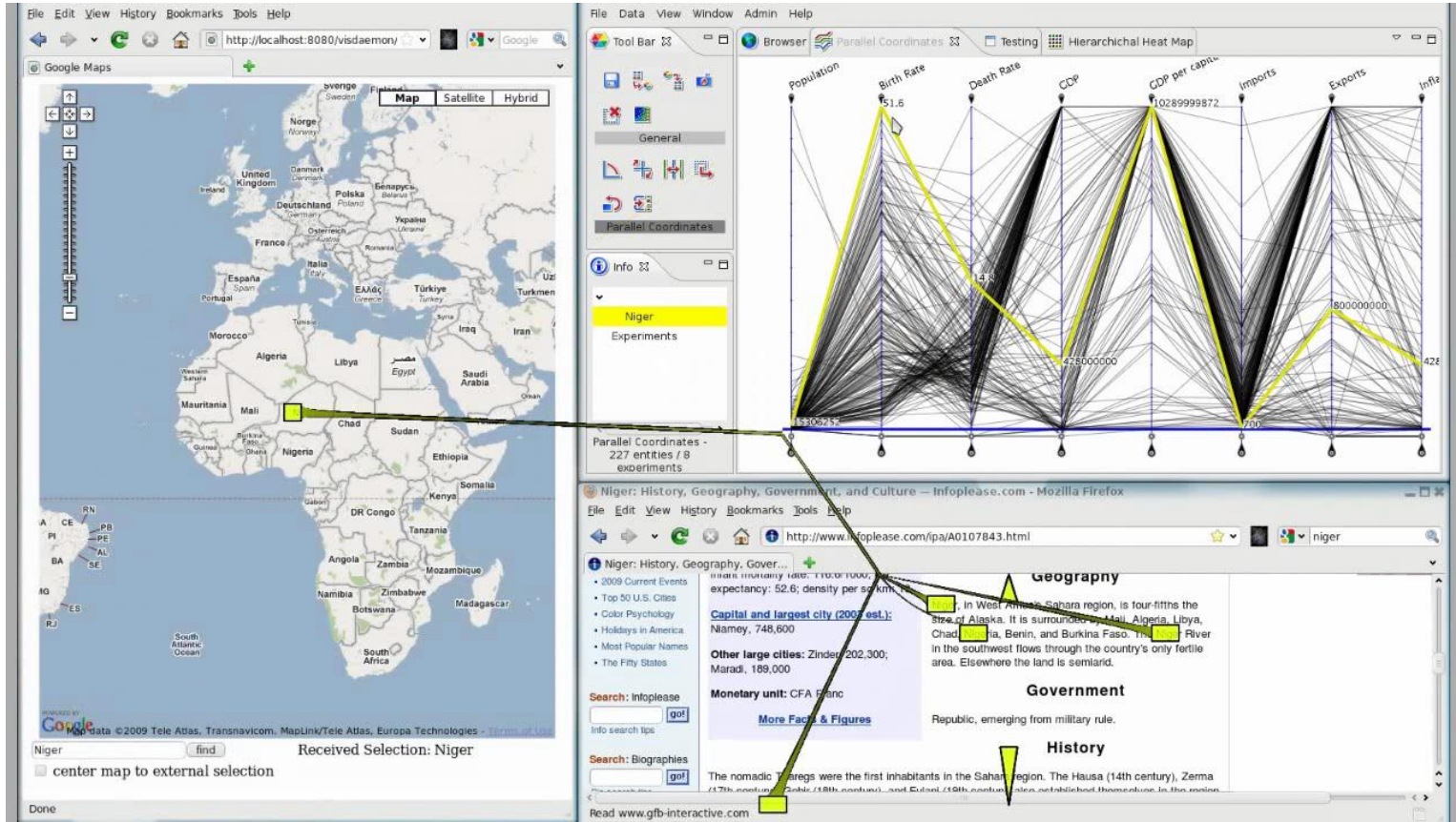
# VISUAL LINKING ACROSS APPLICATIONS



# Visual Links Across Applications

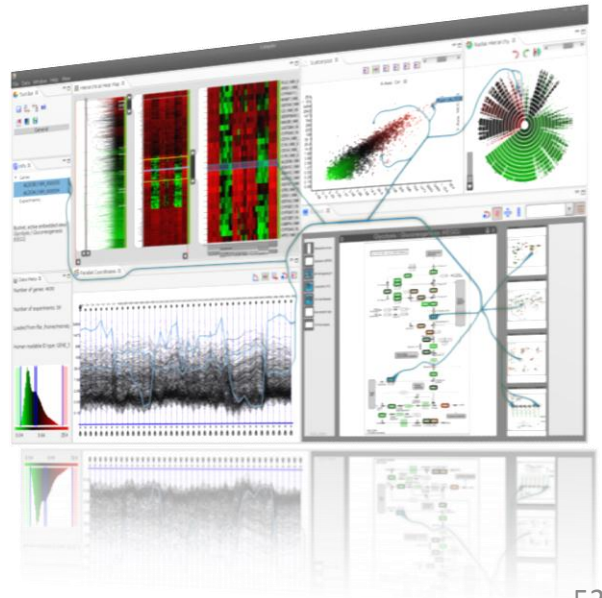




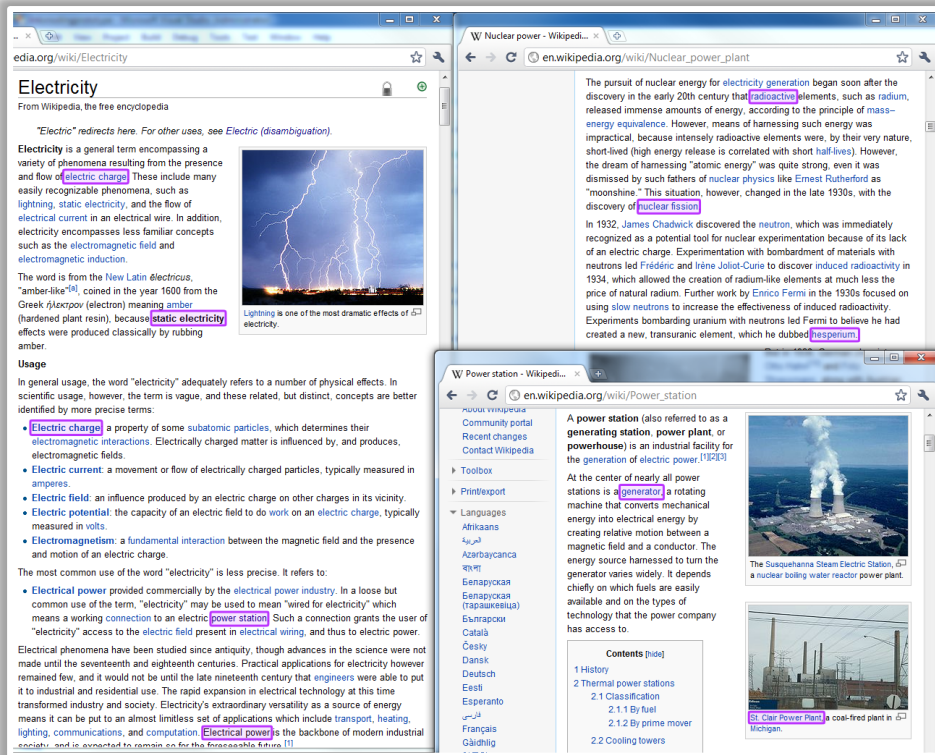


[Steinberger, InfoVis 2011]

# CONTEXT-AWARE VISUAL LINKS



# Simple Highlighting



# Traditional Visual Links

**Electricity**  
From Wikipedia, the free encyclopedia

*"Electric" redirects here. For other uses, see Electric (disambiguation).*

**Electricity** is a general term encompassing a variety of phenomena resulting from the presence and flow of **electric charge**. These include many easily recognizable phenomena, such as lightning, static electricity, and the flow of electrical current in an electrical wire. In addition, electricity encompasses less familiar concepts such as the electromagnetic field and electromagnetic induction.

The word is from the New Latin *electricus*, "amber-like"<sup>[4]</sup>, coined in the year 1600 from the Greek *ήλεκτρον* (electron) meaning *amber* (hardened plant resin), because **static electricity** effects were produced classically by rubbing amber.

**Usage**

In general usage, the word "electricity" adequately refers to a number of physical effects. In scientific usage, however, the term is vague, and these related, but distinct concepts are better identified by more precise terms:

- Electric charge** is a property of some subatomic particles which determines their electromagnetic interactions. Electrically charged matter is influenced by, and produces, electromagnetic fields.
- Electric current**: a movement or flow of electrically charged particles, typically measured in amperes.
- Electric field**: an influence produced by an electric charge on other charges in its vicinity.
- Electric potential**: the capacity of an electric field to do work on an electric charge, typically measured in volts.
- Electromagnetism**: a fundamental interaction between the magnetic field and the presence and motion of an electric charge.

The most common use of the word "electricity" is less precise. It refers to:

- Electrical power** provided commercially by the **electrical power industry**. In a loose but common use of the term, "electricity" may be used to mean "wired for electricity" which means a working connection to an electric **power station**. Such a connection grants the user of "electricity" access to the electric field present in **electrical wiring**, and thus to electric power.

Electric phenomena have been studied since antiquity, though advances in the science were not made until the seventeenth and eighteenth centuries. Practical applications for electricity however remained few, and it would not be until the late nineteenth century that engineers were able to put it to industrial and residential use. The rapid expansion in electrical technology at this time transformed industry and society. Electricity's extraordinary versatility as a source of energy means it can be put to an almost limitless set of applications which include transport, heating, lighting, communications, and computation. **Electrical power** is the backbone of modern industrial society, and is expected to remain so for the foreseeable future.<sup>[1]</sup>

**Nuclear power - Wikipedia**  
en.wikipedia.org/wiki/Nuclear\_power\_plant

The pursuit of nuclear energy for **electricity** generation began soon after the discovery in the early 20th century that **radioactive** elements, such as **radium**, released immense amounts of energy. According to the principle of **mass–energy equivalence**. However, means of harnessing such energy was impractical, because intensely radioactive elements were, by their very nature, short-lived (high energy release is correlated with short half-lives). However, the dream of harnessing "atomic energy" was quite strong, even it was dismissed by such fathers of nuclear physics like Ernest Rutherford as "moonshine." This situation however, changed in the late 1930s, with the discovery of **nuclear fission**.

In 1932, James Chadwick discovered the neutron, which was immediately recognized as a potential tool for nuclear experimentation because of its lack of an electric charge. Experimentation with bombardment of materials with neutrons led Frédéric and Irène Joliot-Curie to discover induced radioactivity in 1934, which allowed the creation of radium-like elements at much less the price of natural radium. Further work by Enrico Fermi in the 1930s focused on using slow neutrons to increase the effectiveness of induced radioactivity. Experiments bombarding uranium with neutrons led Fermi to believe he had created a new, transuranic element, which he dubbed **hesperium**.

**Power station - Wikipedia**  
en.wikipedia.org/wiki/Power\_station

A **power station** (also referred to as a **generating station**, **power plant**, or **powerhouse**) is an industrial facility for the generation of electric power.<sup>[1][2][3]</sup>

At the center of nearly all power stations is a **generator**, a rotating machine that converts mechanical energy into electrical energy by creating relative motion between a magnetic field and a conductor. The energy source harnessed to turn the generator varies widely. It depends chiefly on which fuels are easily available and on the types of technology that the power company has access to.

**Contents** [hide]

- History
- Thermal power stations
  - Classification
    - 1.1 By fuel
    - 1.2 By prime mover
  - 2.2 Cooling towers

The Susquehanna Steam Electric Station, a nuclear boiling water reactor power plant.

**St. Clair Power Plant** is a coal-fired plant in Michigan.



# Context-Aware Visual Links

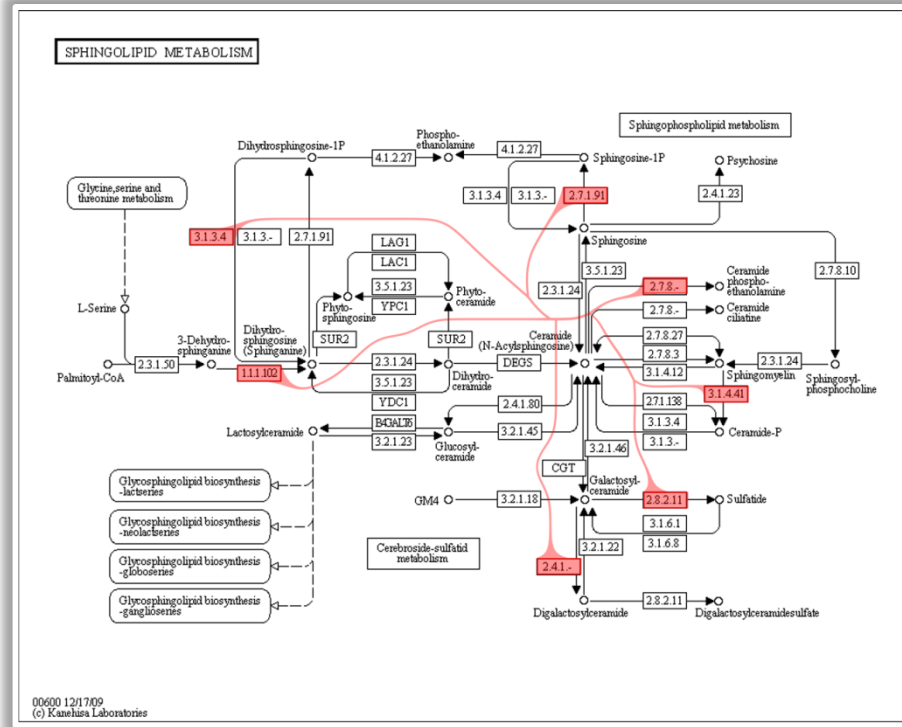
The image displays three browser windows illustrating context-aware visual links. The windows are:

- Electricity** (Wikipedia):
  - Text: "Electricity is a general term encompassing a variety of phenomena resulting from the presence and flow of **electric charge**. These include many easily recognizable phenomena, such as lightning, static electricity, and the flow of electrical current in an electrical wire. In addition, electricity encompasses less familiar concepts such as the electromagnetic field and electromagnetic induction."
  - Text: "The word is from the New Latin *electricus*, 'amber-like', coined in the year 1600 from the Greek *ήλεκτρον* (electron) meaning **amber** (hardened plant resin), because **static electricity** effects were produced classically by rubbing amber."
  - Text: "In general usage, the word 'electricity' adequately refers to a number of physical effects. In scientific usage, however, the term is vague, and these related, but distinct, concepts are better identified by more precise terms:"
  - List of terms:
    - Electric charge**: a property of some subatomic particles, which determines their electromagnetic interactions. Electrically charged matter is influenced by, and produces, electromagnetic fields.
    - Electric current**: a movement or flow of electrically charged particles, typically measured in amperes.
    - Electric field**: an influence produced by an electric charge on other charges in its vicinity.
    - Electric potential**: the capacity of an electric field to do work on an electric charge, typically measured in volts.
    - Electromagnetism**: a fundamental interaction between the magnetic field and the presence and motion of an electric charge.
  - Text: "The most common use of the word 'electricity' is less precise. It refers to:"
  - List of terms:
    - Electrical power** provided commercially by the **electrical power industry**. In a loose but common use of the term, 'electricity' may be used to mean 'wired for electricity' which means a working connection to an electric **power station**. Such a connection grants the user of 'electricity' access to the **electric field** present in **electrical wiring**, and thus to electric power.
  - Text: "Electrical phenomena have been studied since antiquity, though advances in the science were not made until the seventeenth and eighteenth centuries. Practical applications for electricity however remained few, and it would not be until the late nineteenth century that engineers were able to put it to industrial and residential use. The rapid expansion in electrical technology at this time transformed industry and society. Electricity's extraordinary versatility as a source of energy means it can be put to an almost limitless set of applications which include transport, heating, lighting, communications, and computation. **Electrical power** is the backbone of modern industrial society, and is expected to remain so for the foreseeable future. [1]"
  - Image: A lightning bolt striking a city at night.
- Nuclear power** (Wikipedia):
  - Text: "The pursuit of nuclear energy for **electricity** generation began soon after the discovery in the early 20th century that **radioactive** elements, such as **radium**, released immense amounts of energy, according to the principle of **mass-energy equivalence**. However, means of harnessing such energy was impractical, because intensely radioactive elements were, by their very nature, short-lived (high energy release is correlated with short **half-lives**). However, the dream of harnessing 'atomic energy' was quite strong, even it was dismissed by such fathers of **nuclear physics** like **Ernest Rutherford** as 'moonshine.' This situation, however, changed in the late 1930s, with the discovery of **nuclear fission**."
  - Text: "In 1932, James Chadwick discovered the neutron, which was immediately recognized as a potential tool for nuclear experimentation because of its lack of an electric charge. Experimentation with bombardment of materials with neutrons led Frédéric and Irène Joliot-Curie to discover induced radioactivity in 1934, which allowed the creation of **radium-like** elements at much less the price of natural radium. Further work by **Enrico Fermi** in the 1930s focused on using **slow neutrons** to increase the effectiveness of induced radioactivity. Experiments bombarding uranium with neutrons led Fermi to believe he had created a new, transuranic element, which he dubbed **hesperium**."
- Power station** (Wikipedia):
  - Text: "A **power station** (also referred to as a **generating station**, **power plant**, or **powerhouse**) is an industrial facility for the generation of **electric power**. [1][2][3]"
  - Text: "At the center of nearly all power stations is a **generator**, a rotating machine that converts mechanical energy into electrical energy by creating relative motion between a magnetic field and a conductor. The energy source harnessed to turn the generator varies widely. It depends chiefly on which fuels are easily available and on the types of technology that the power company has access to."
  - Text: "The Susquehanna Steam Electric Station, a nuclear boiling water reactor power plant."
  - Text: "St. Clair Power Plant, a coal-fired plant in Michigan."
  - Table of Contents:
    - 1 History
    - 2 Thermal power stations
    - 2.1 Classification
    - 2.1.1 By fuel
    - 2.1.2 By prime mover
    - 2.2 Cooling towers

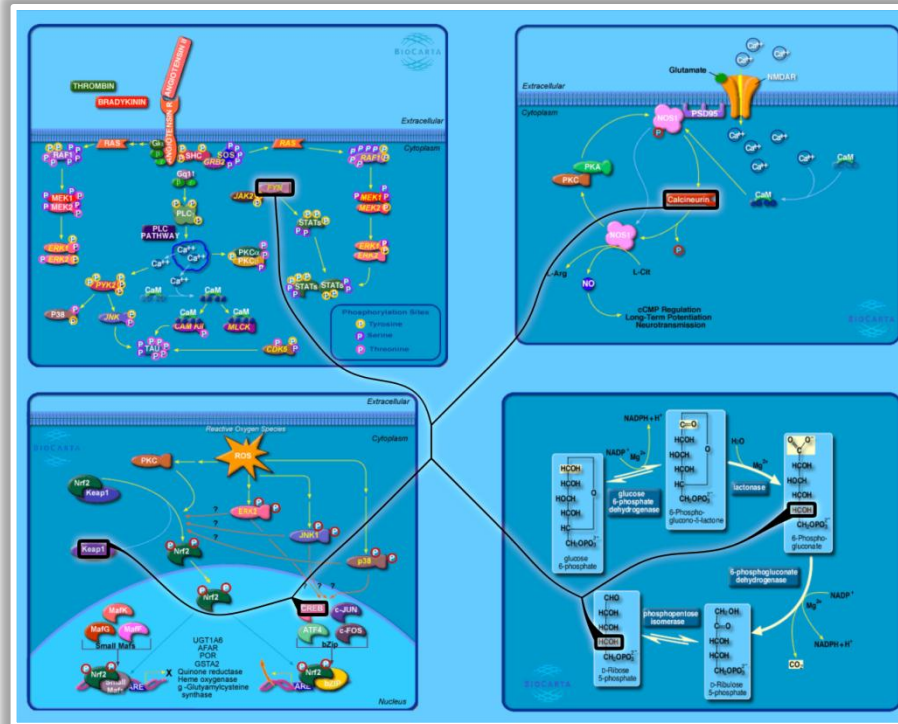
Purple lines connect related terms across the windows, such as "Electricity" to "Electricity", "Electric charge", "Electric field", "Electric potential", "Electromagnetism", "Electrical power", "power station", "generator", "radioactive", "nuclear fission", "slow neutrons", "hesperium", "radioactive elements", "mass-energy equivalence", "half-lives", "nuclear physics", "moonshine", "nuclear fission", "neutron", "Ernest Rutherford", "Joliot-Curie", "radioactivity", "radium-like", "Enrico Fermi", "slow neutrons", "radioactivity", "uranium", "transuranic element", "hesperium", "generator", "mechanical energy", "electrical energy", "magnetic field", "conductor", "energy source", "fuels", "technology", "Susquehanna Steam Electric Station", "St. Clair Power Plant", "coal-fired plant", "Michigan".

[illegible]

# Kegg Pathway

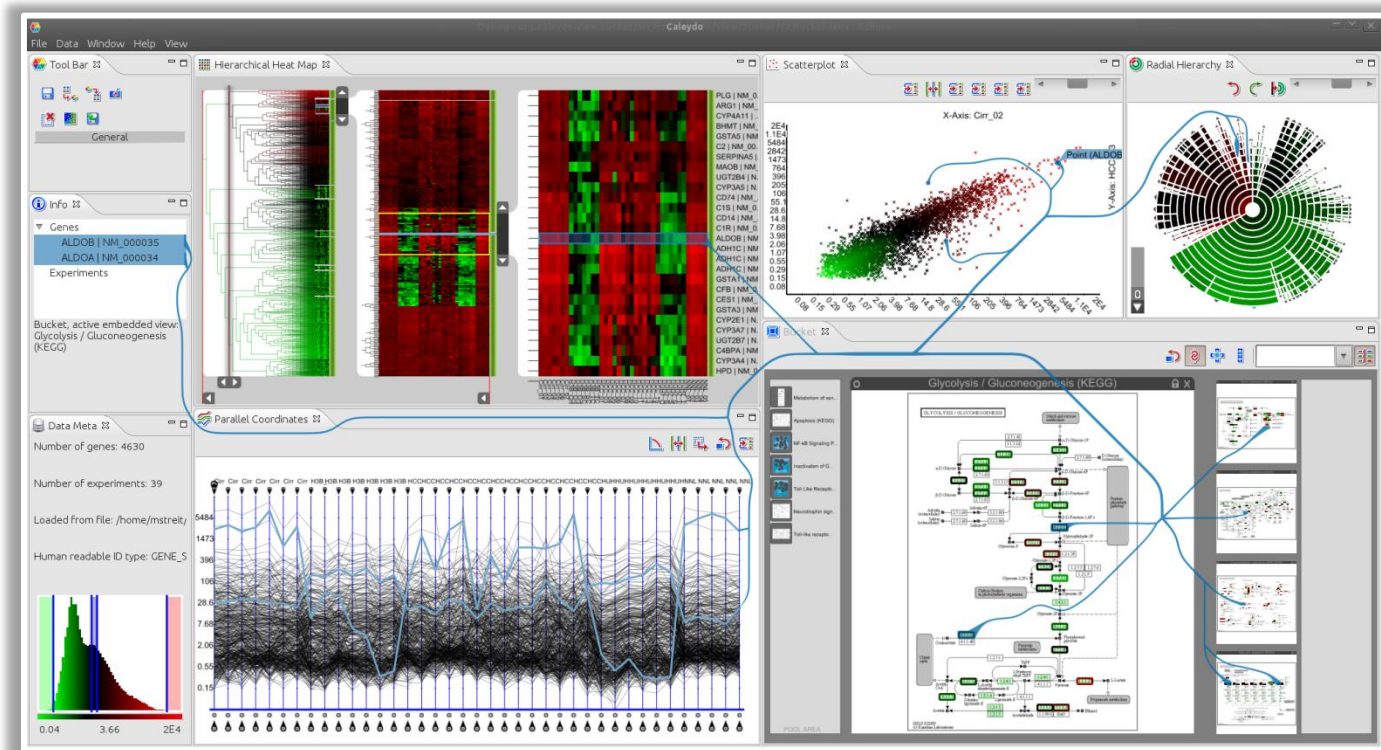


# Multiple Biocarta Pathways

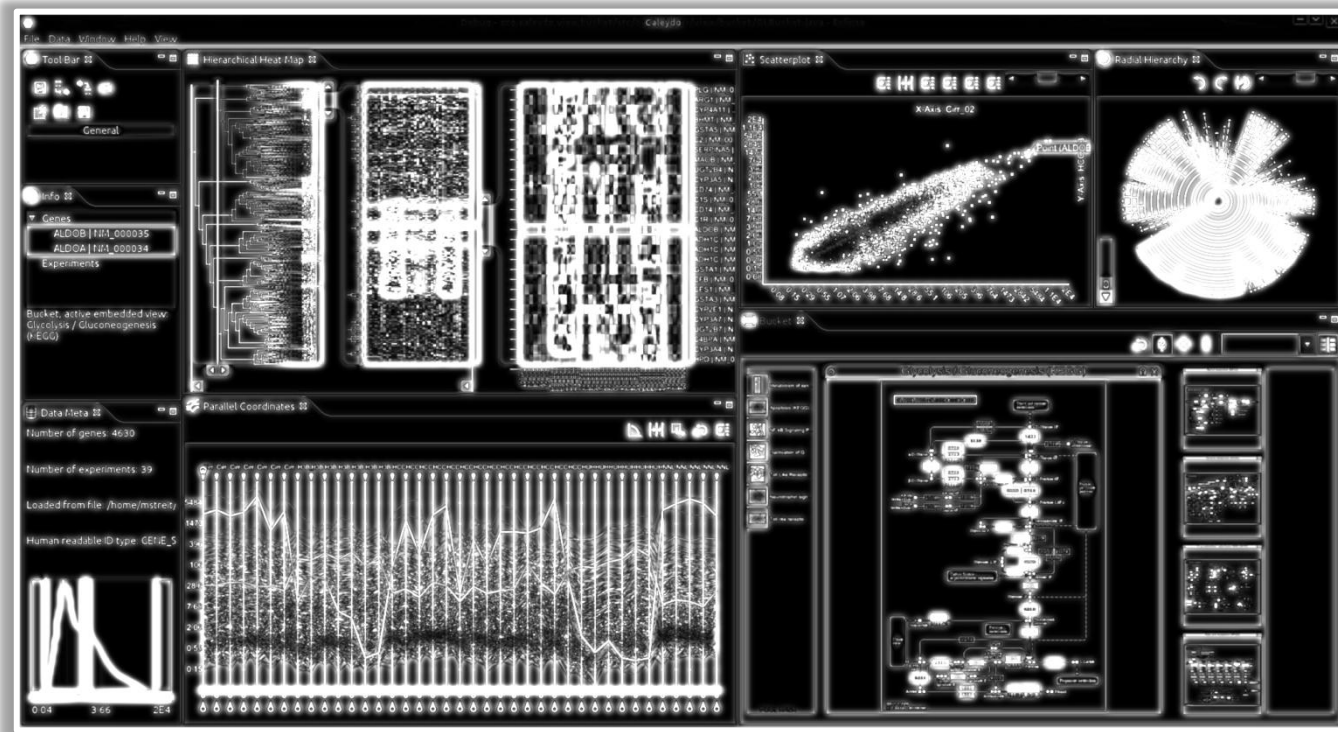




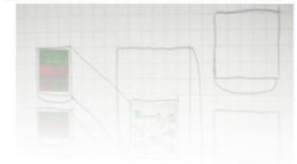
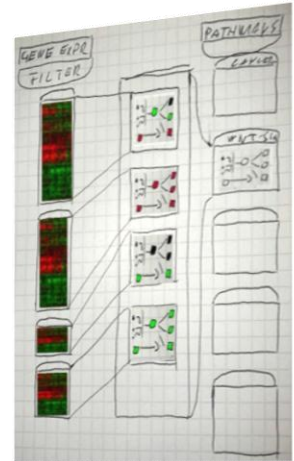
# Caleydo

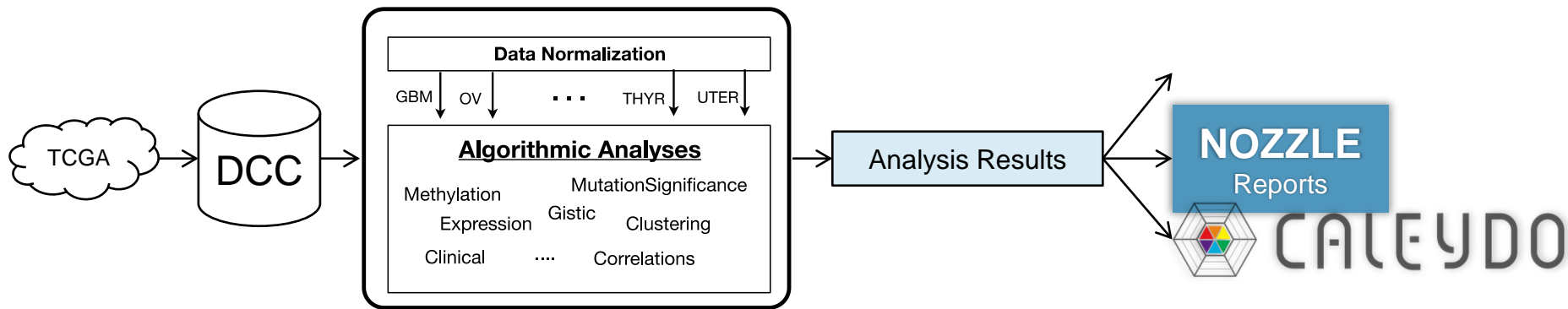


# How it works



# PLANNED WORK HERE AT CBMI





Project Title Ideas:

OI – Omics Integrator

Frazzlomics

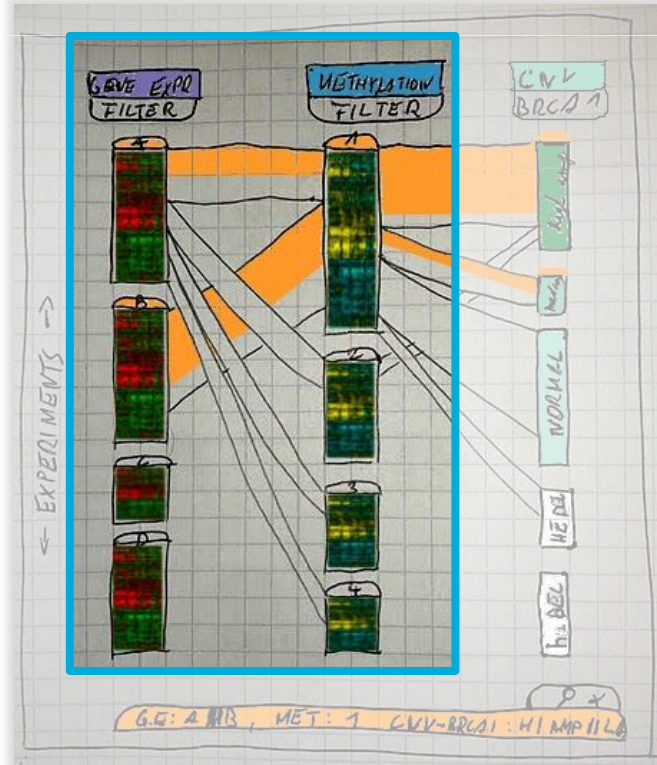
Wyee



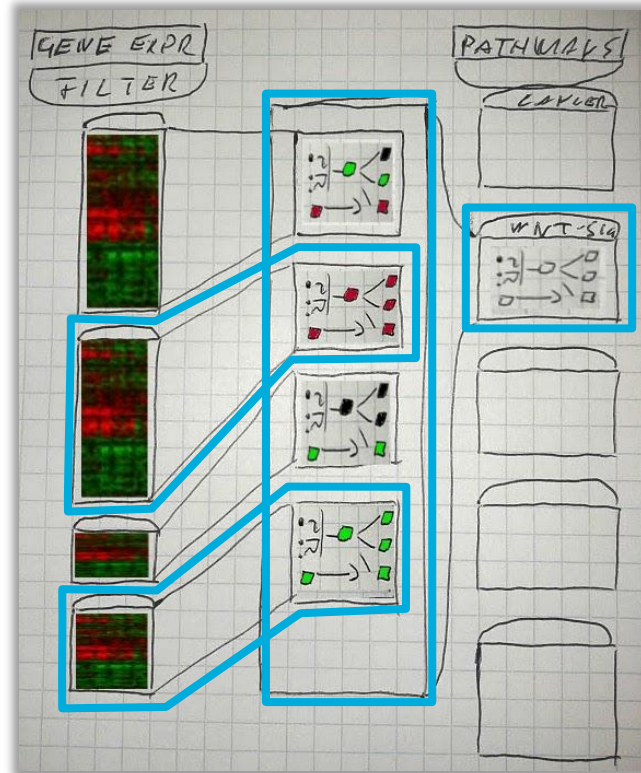
# Vision: Integrated Visualization of Multiple Datasets from Firehose

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

