

# Visualization of Gene Combinations

Christian Tominski & Heidrun Schumann  
University of Rostock, Germany

IV08, London, July 10, 2008

# Background

- Goal: Shed light on complex biol. systems
- Microarrays are fundamental technology
- Huge volumes of microarray data collected:  
Gene Expression Omnibus hosts ca. 250,000  
data sets  
([www.ncbi.nlm.nih.gov/geo/](http://www.ncbi.nlm.nih.gov/geo/), July 2008)
- Question: How to make sense of these data?

# Background

- Microarray data analysis involves many intricate steps

Computer Vision

Self-Organizing Maps

Normalization

Digitalization

Significance Analysis

Error correction

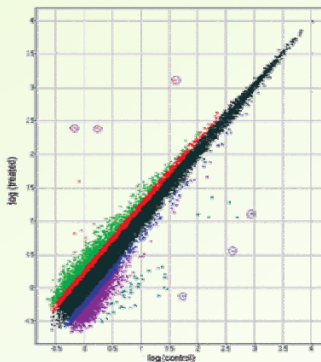
Clustering

Visualization

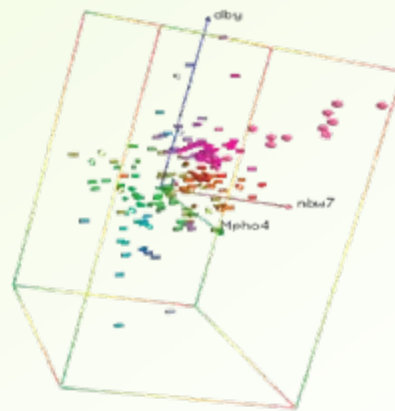
Principle Component Analysis

# Background

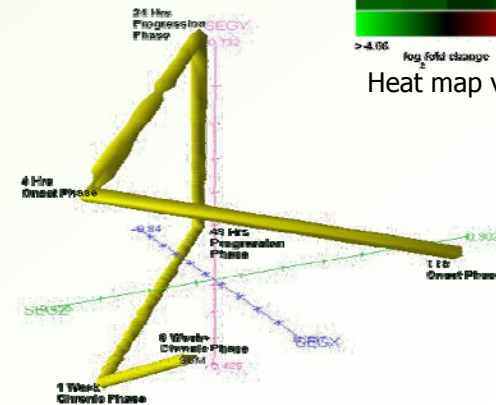
- Microarray data
  - Set of genes:  $G$
  - Set of samples:  $S$
  - Gene expression:  $exp_G : G \times S \rightarrow R$
- Many tools exist to visualize  $exp_G$



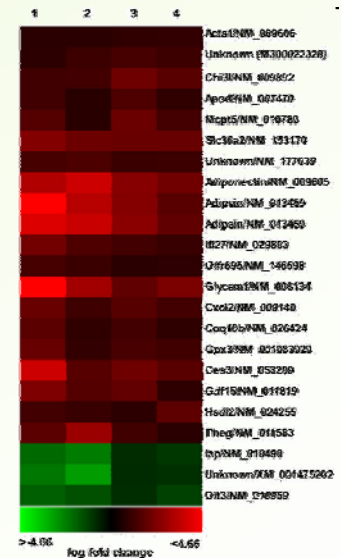
2D Scatter Plot



3D Scatter Plot



Time series visualization



Heat map visualization

# Background

- Visualization tools usually focus on expression of single genes
- To build groups or clusters, analytical methods are needed
- Difficulties:
  - Complex algorithms and costly computations
  - Difficult to understand and steer by user
- Our approach to support microarray exploration is:

***Gene Combinations + Visualization + Interaction***

# Gene Combinations

- Recall: set of genes  $G$ , set of Samples  $S$ , gene expression  $exp_G : G \times S \rightarrow R$
- **Gene combination:**  $GC \in P(G)$
- Expression of gene combinations:  
 $exp_{GC} : P(G) \times S \rightarrow R$
- Usually,  $exp_{GC}$  involves averaging  $exp_G$  of genes participating in a gene combination

# Visualization Challenges

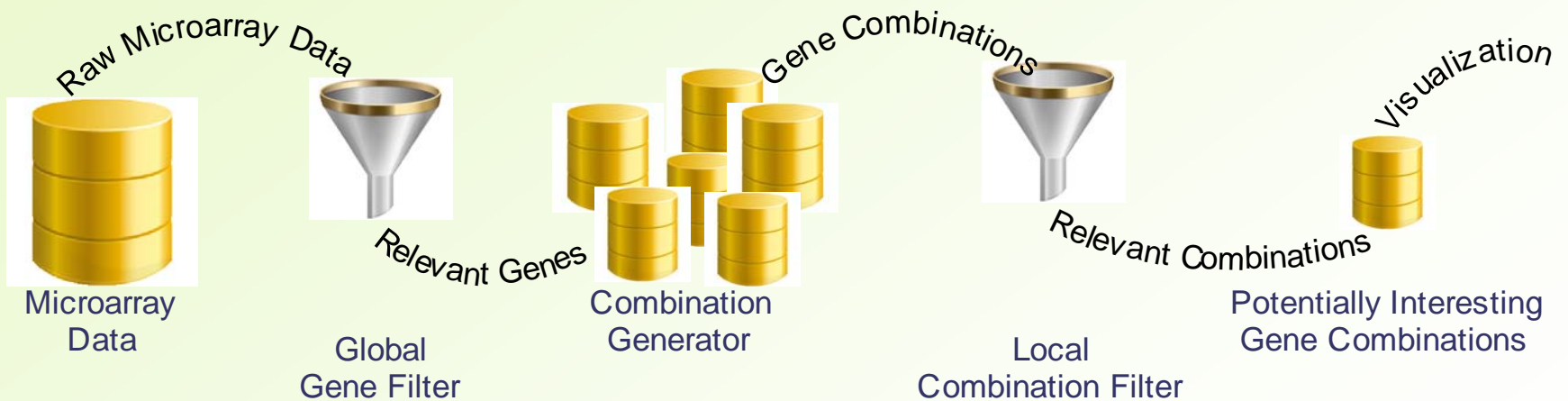
- Cannot visualize all gene combinations due to immense data volume
- Visual representation must reflect that gene combinations bear more information than single genes
- Users with biomedical background are accustomed to their tools and have certain expectations

# Proposed Solutions

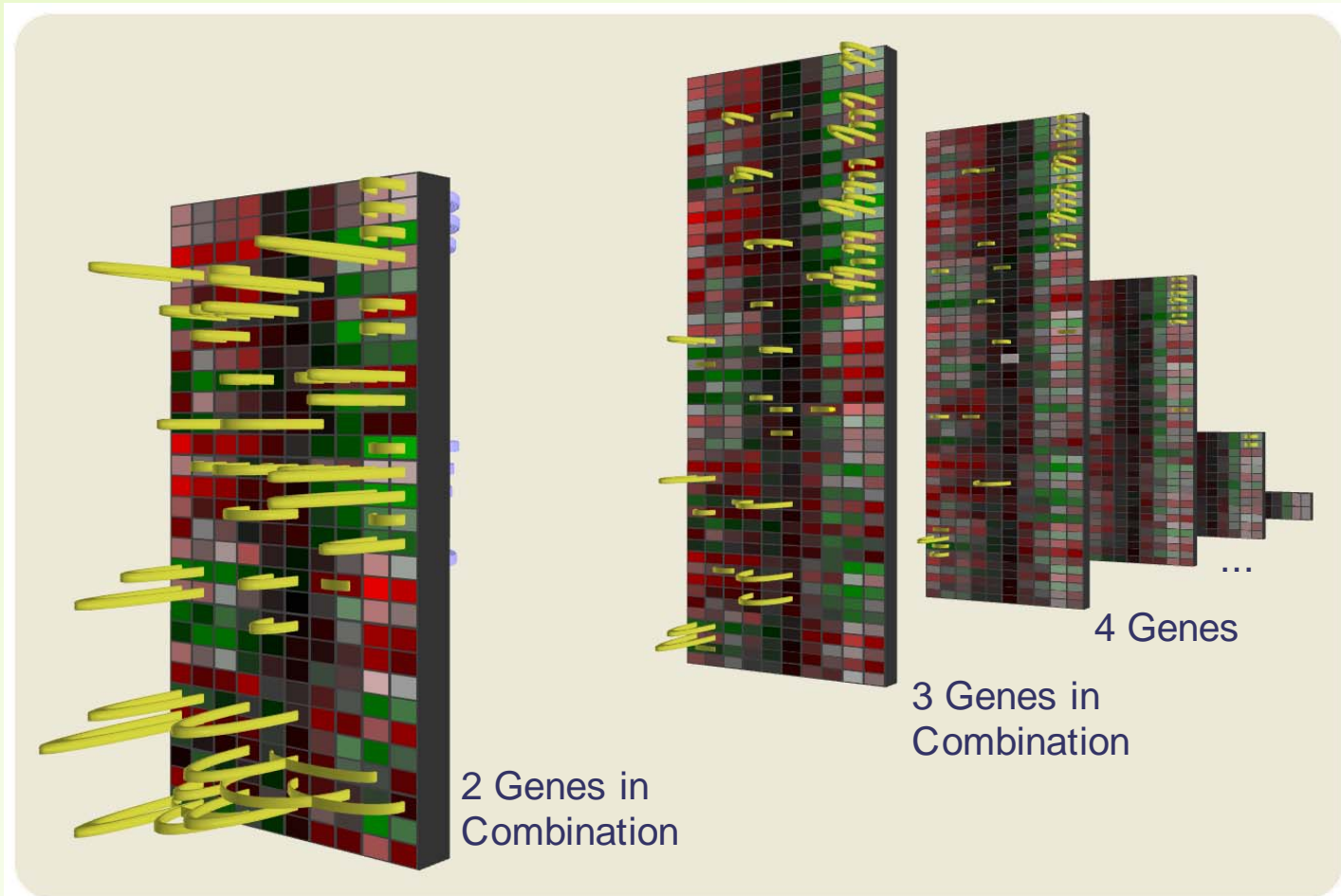
- Analysis pipeline
  - Pipeline of filters to reduce volume of data
  - Gene expression filter, gene combination filter, ...
- 3D Visualization
  - Based on heatmaps arranged in 3D
  - Novel color coding for aggregated expression values of gene combinations
  - Visual tools to support comparison
- Easy to use interaction techniques



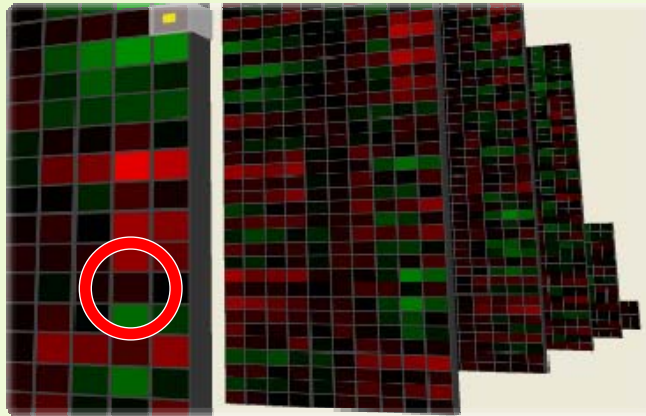
# Analysis Pipeline



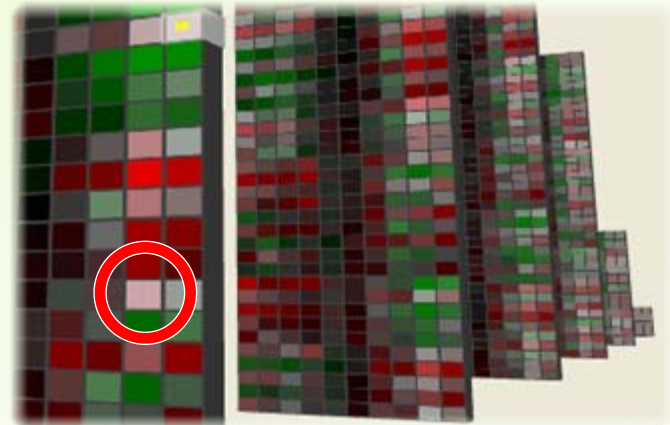
# Visualization



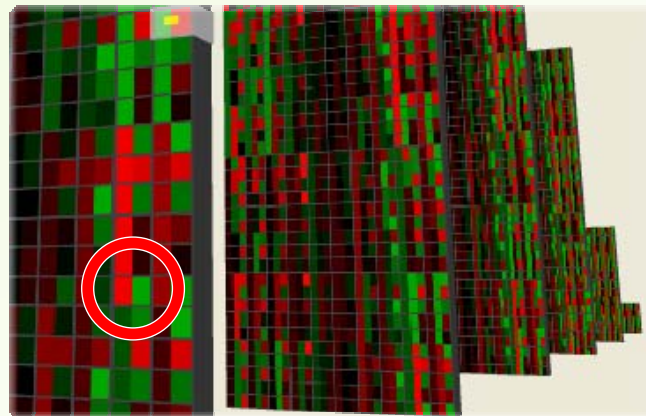
# Visualization



Classic Encoding

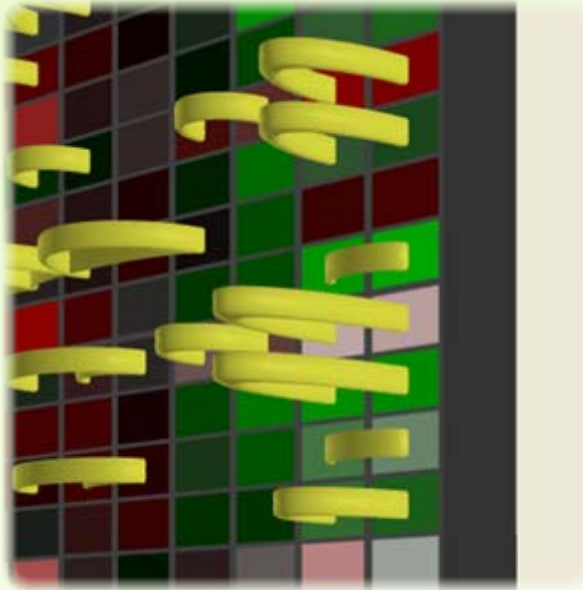


Variance Encoding

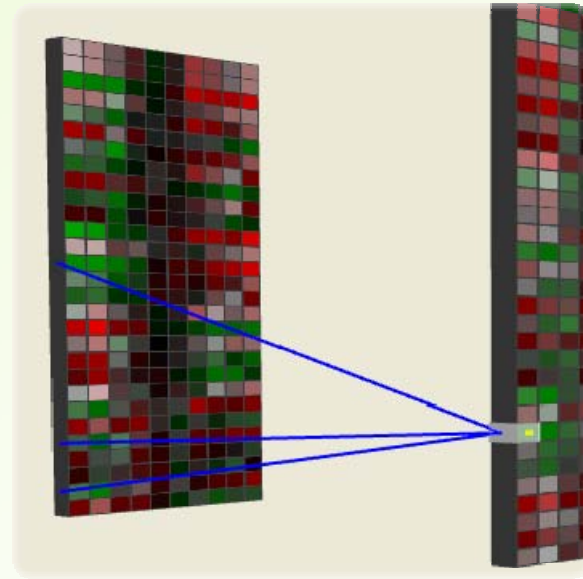


Small Multiples

# Visualization



Dis/Similarity Arcs



Gene Containment Links

# Interaction

**Let's switch to the demo!**

# Summary

- New concept for visually analyzing microarray data based on gene combinations
- Novel variance-aware color coding
- Interactive tool ViGeCo available as
  - Stand-alone application
  - Applet ([vcg.informatik.uni-rostock.de/~ct/ViGeCo/ViGeCo.html](http://vcg.informatik.uni-rostock.de/~ct/ViGeCo/ViGeCo.html))
  - Plugin for the Mayday Microarray Data Analysis Framework

# Future Work

- **Filtering**
  - Avoid situations where users stumble over too large data
  - Automatically find compromise between data size and interactivity
  - Integrate further filters from biomedical background
- **Scalability through integration with other tools**
  - Utilize facilities of Mayday framework
  - Connect to other data sources (e.g., Gene Expression Omnibus)
- **Improve usability**
  - Hyperlink from visualization to additional information sources