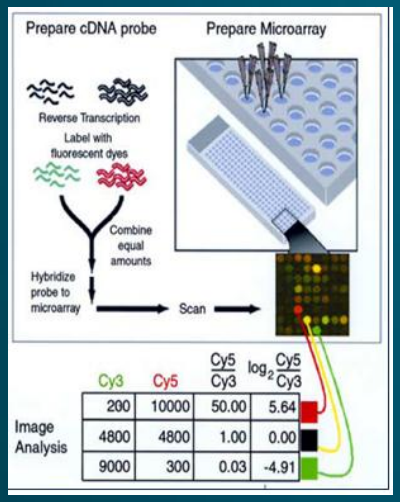
VISUALIZATION FOR HIGH THROUGHPUT BIOLOGICAL DATA

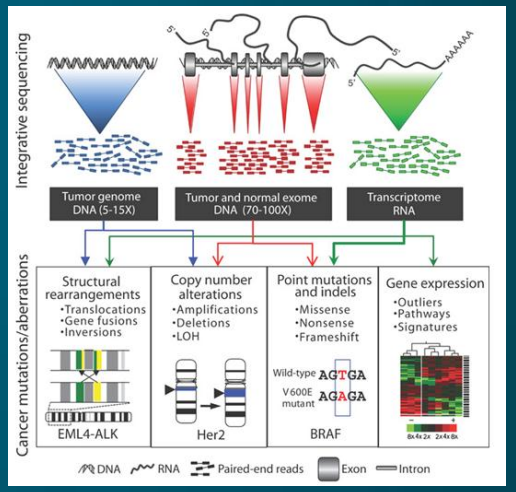
Leishi Zhang[1], Jasna Kuljis[2] and Xiaohui Liu[2]
[1] University of Konstanz, Germany
[2] Brunel University, UK

HIGH THROUGHPUT EXPERIMENTS

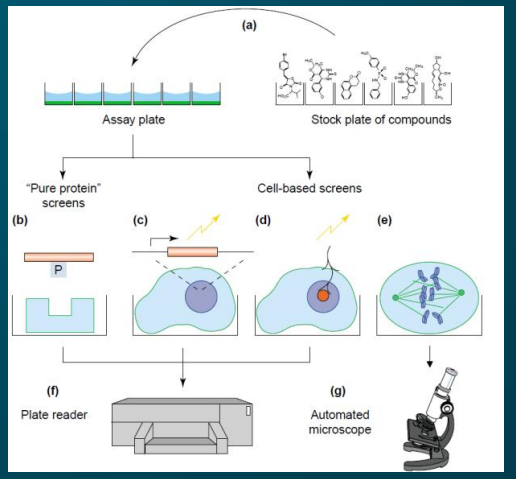
DNA microarrays



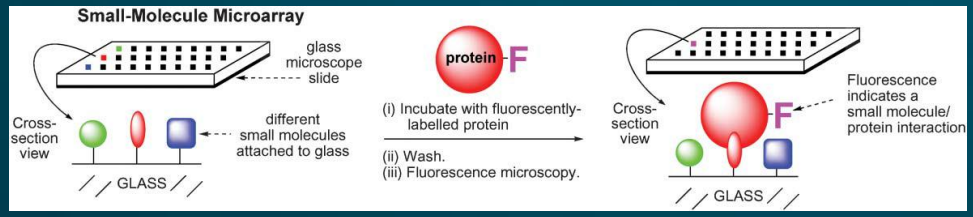
High throughput sequencing



High content screening



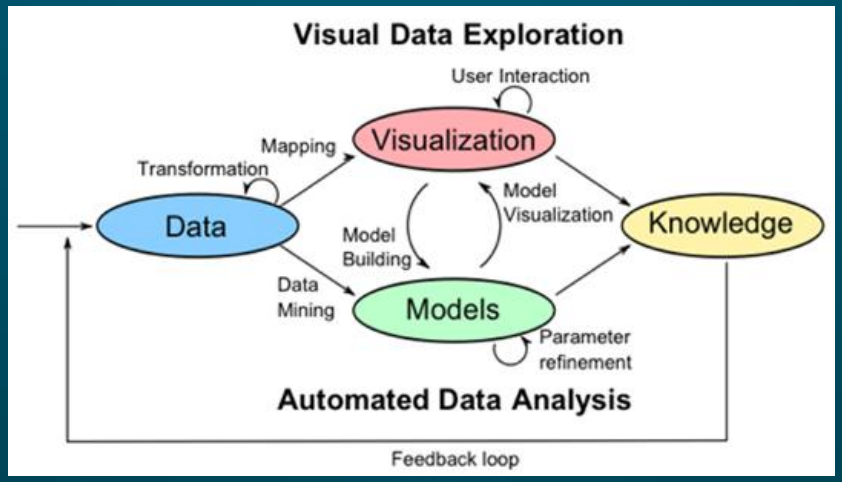
Small molecule microarrays



HT DATA

- *large*
- *high-dimensionality*
- *heterogeneous*

How to make sense out of the data? – visual analytics
automated data analysis + interactive visualizations



VISUAL ANALYTICS FOR HT DATA

Visual Analytics combines

- *automated data analysis (statistics and data mining methods)*
- *interactive visualization (visual parameters, graphical representations, and human computer interactions)*

In this talk, I will discuss…

- *existing visualization techniques*
- *open issues*

VISUALIZATION DESIGN

Information seeking paradigm

"Overview First, Zoom and Filter, Details on Demand"

- Ben Shneiderman, 1996

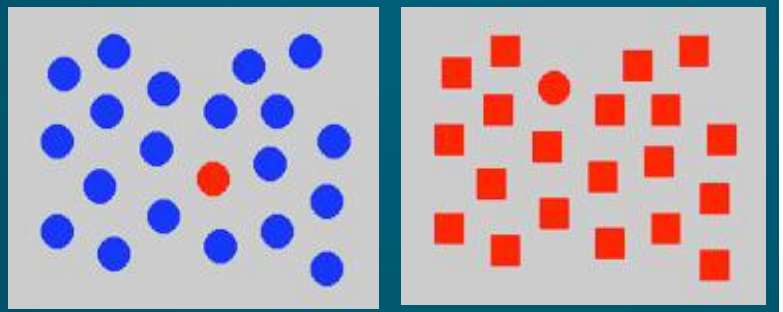
What is important?

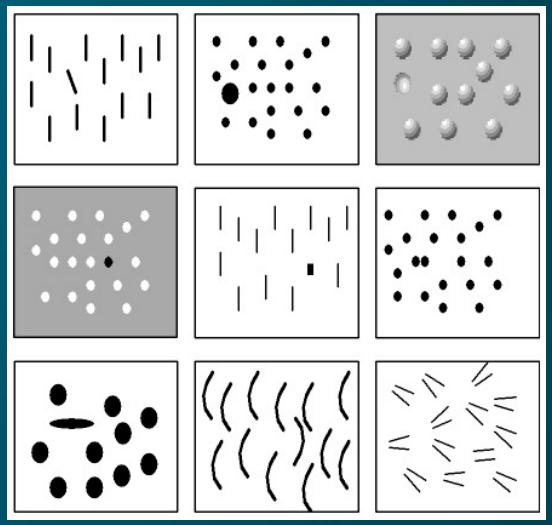
- *providing overview as well as details*
- *showing patterns and relations*
- *supporting dynamic queries*

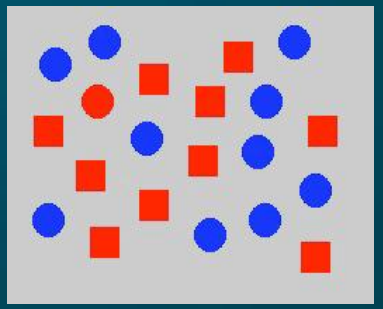
VISUAL PARAMETER DESIGN

Various visual channels

- *color*
- *Shape*
- *size*
- *position*
- *texture*
- *…*
- *Challenge: how to effectively use/combine different visual parameters to show interesting part of the data?*



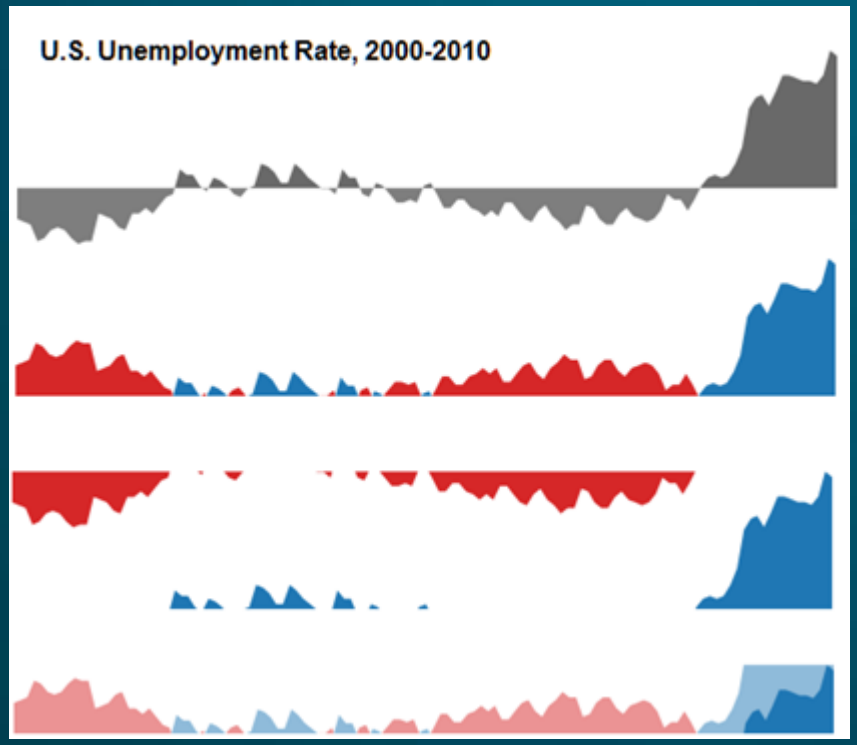




?…

VISUAL PARAMETER DESIGN

Play with the parameters



line?

line+ colors?

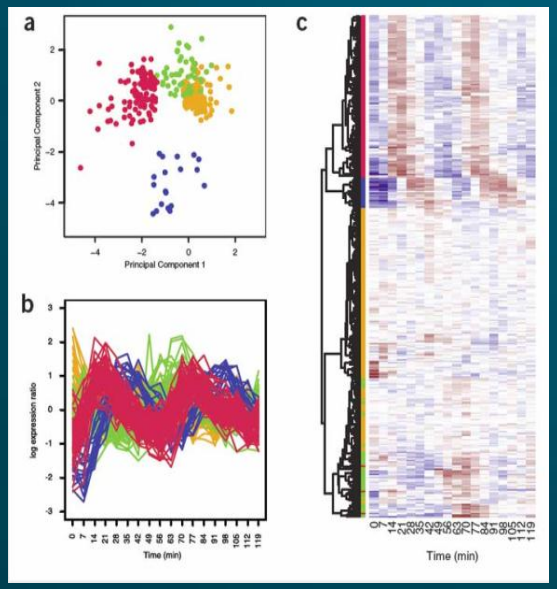
offset?

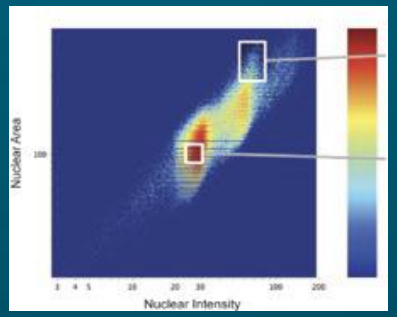
two tone colors?…

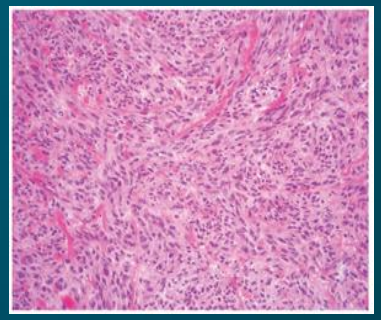
GRAPHICAL REPRESENTATION DESIGN

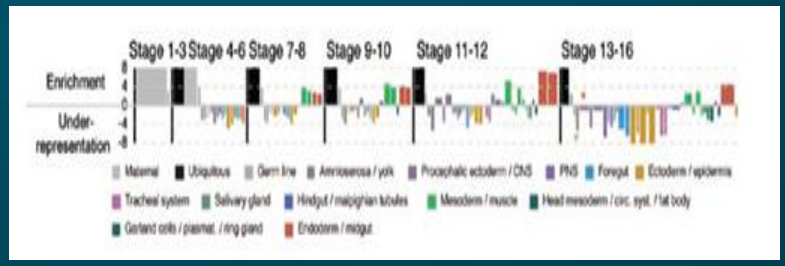
Challenge: given the large data, how to design graphical representations which:

- *highlight patterns and relations*
- *show both overview and details*



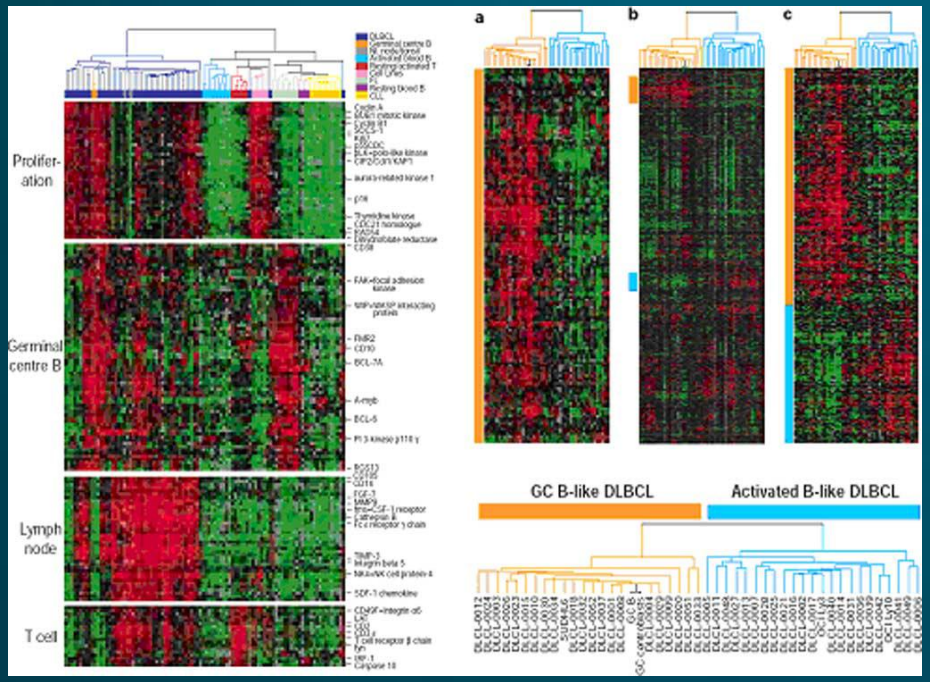




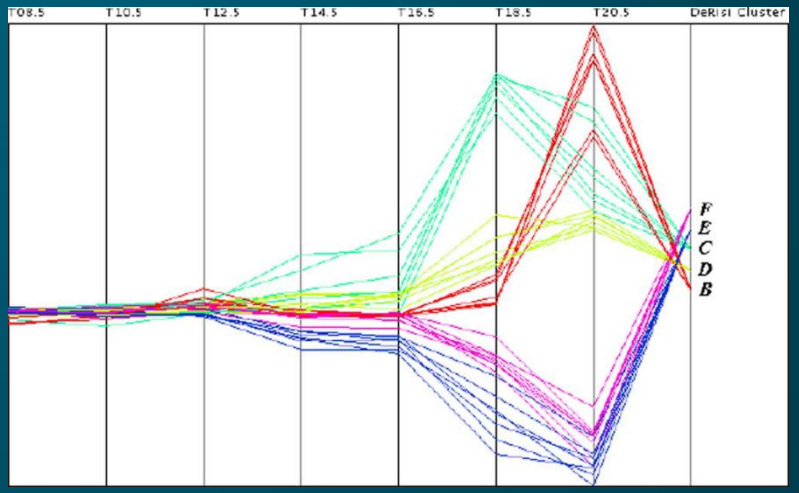


GR DESIGN - OVERVIEW (1)

Mapping data values – heatmap vs. parallel coordinates



no overlap

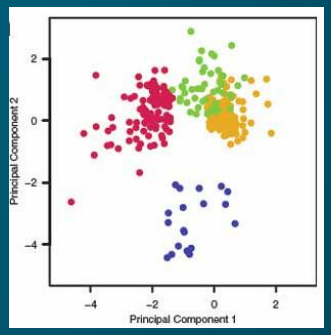


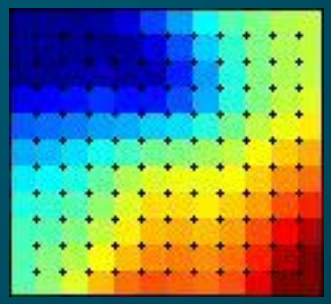
easy to see value differences

GR DESIGN - OVERVIEW (2)

Mapping distance/similarity between objects to a 2D/3D display as scatterplot or grids: dimension reduction

- *Projection Pursuit*
- *Principle Component Analysis*
- *Multi Dimensional Scaling*
- *Self Organising Map*
- *ISOMAP*
- *Locally Linear Embedding*
- *Stochastic Neighbourhood Embedding*
- *Generative Topographic Mapping*
- *…*

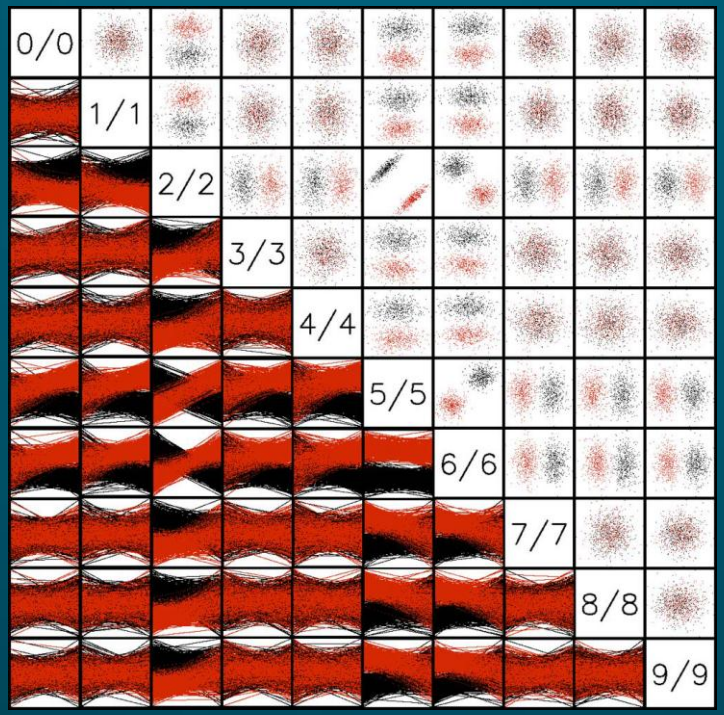




*Focus: best approximate the **structure** (pairwise distance, and/or neighborhood info.) of data in the low dimensional visual space*

GR DESIGN - OVERVIEW (3)

Divide & display: small multiples

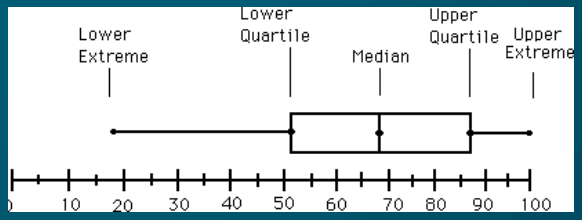


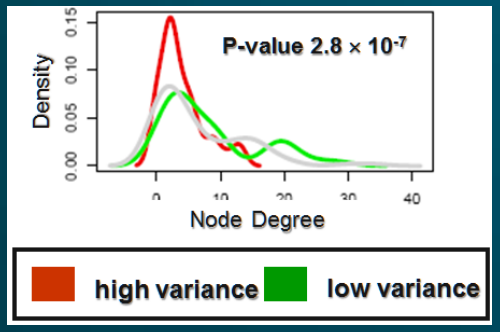
show details of data
dimensions ordering is crucial

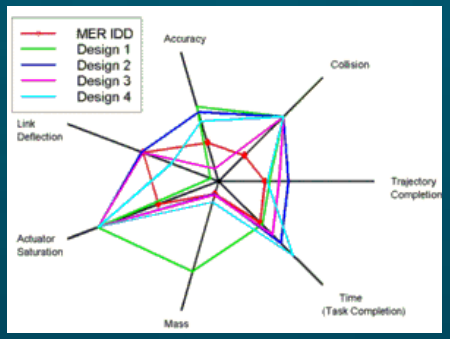
GR DESIGN – DETAILED INFO (1)

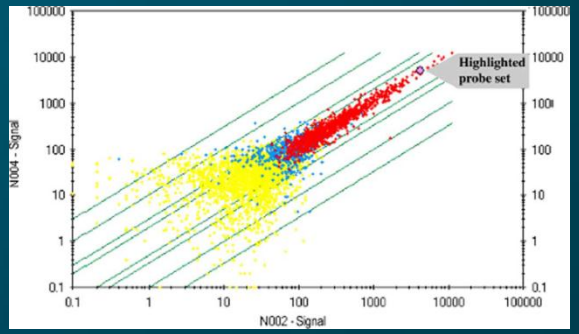
Data summarization, detailed comparison and correlation analysis

- *density plot*
- *box-and-whisker plot*
- *radar/spider plot*
- *correlation plot*
- *…*









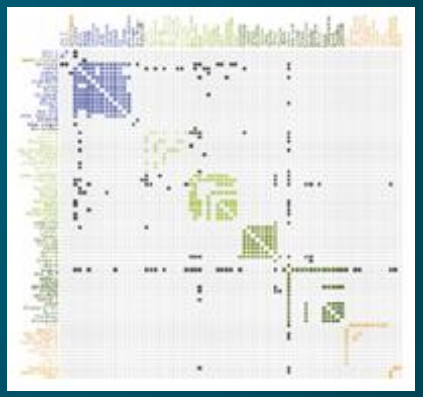
provide good support for statistical analysis, and comparison between subsets of data

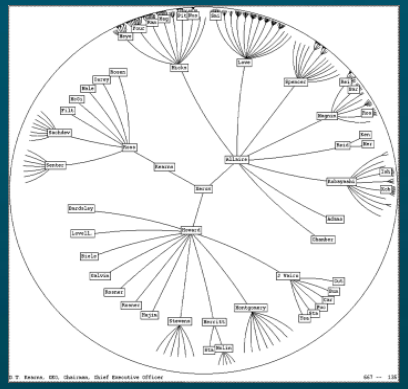
GR DESIGN – DETAILED INFO (2)

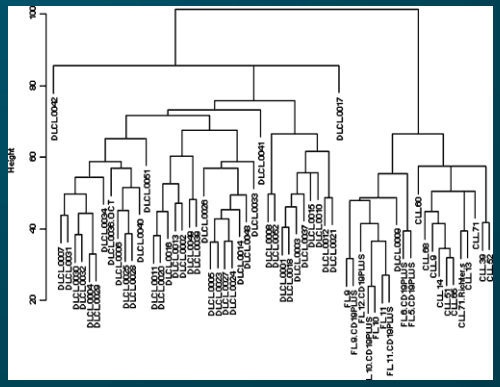
Links and relations

- *force-directed*
- *matrix view*
- *treemap*
- *hyperbolic view*
- *dendrogram*
- *…*

none –hierarchical relations







hierarchical relations

HUMAN COMPUTER INTERACTIONS

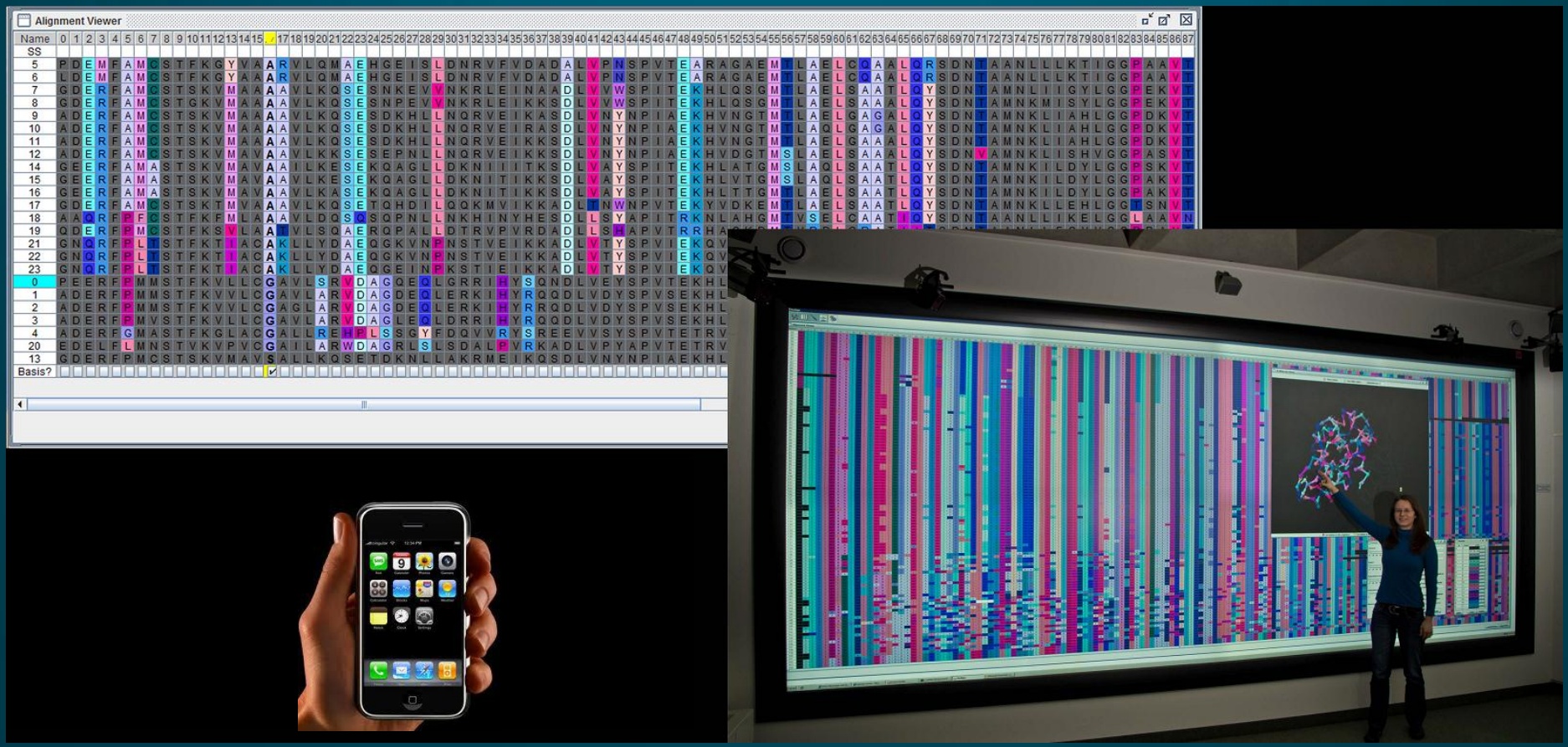
Design interactive user interface

- *zooming*
- *panning*
- *linking and brushing*
- *…*

Typically a HT Data Analysis tool integrates multiple visualization panels with linking and brushing and other mouse/keyboard functions to support dynamic query and detail-on-demand visual data analysis

OPEN ISSUES

- *Scalability*



OPEN ISSUES

- *Scalability (hardware, software)*
- *Visualizing uncertainties in data*
- *Visualizing evolving changes*
- *Evaluating quality of visual representations*

Thank you very much for your attention! ☺