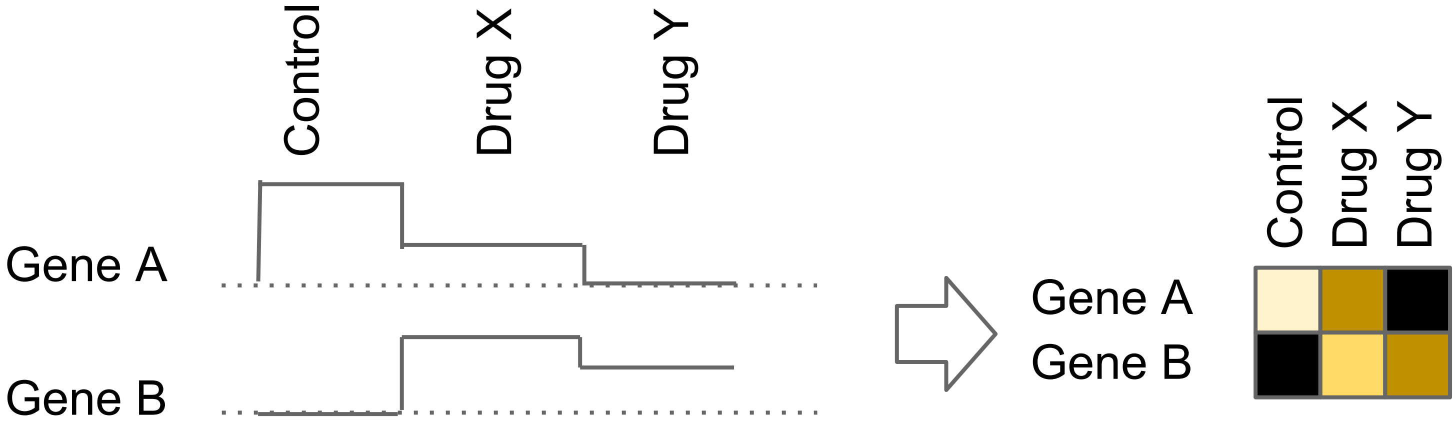
Heat²map: enriched differential gene expression heatmap

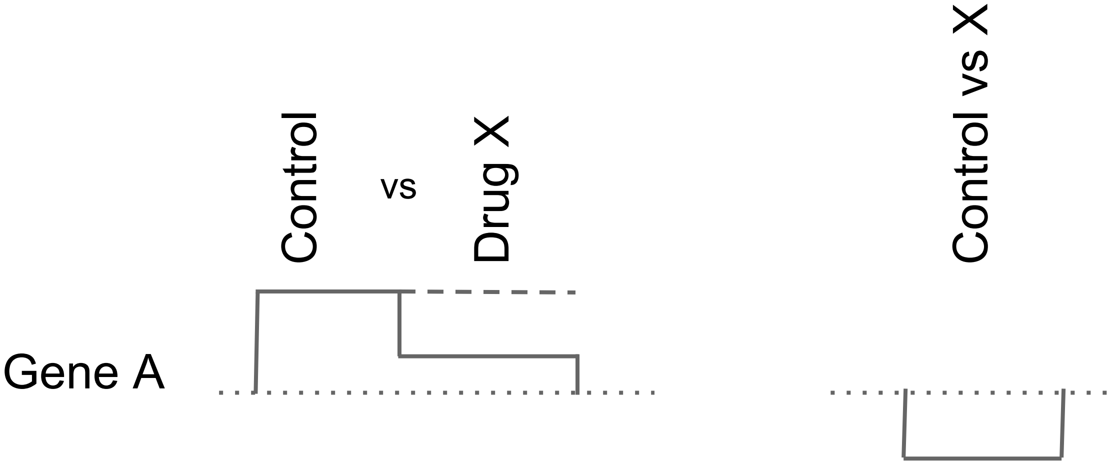
Project Update

November 14, 2011
Anton Zoubarev

Gene expression



Differential gene expression



Control

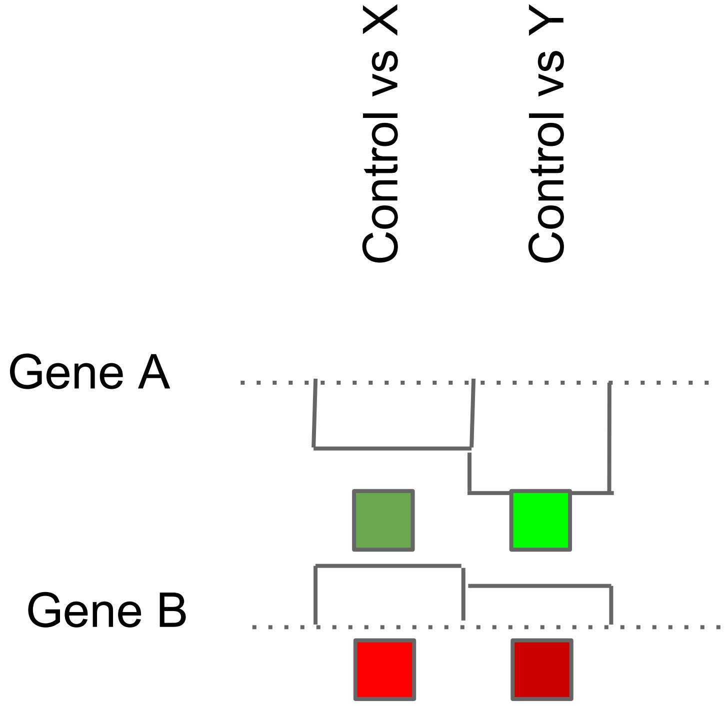
vs

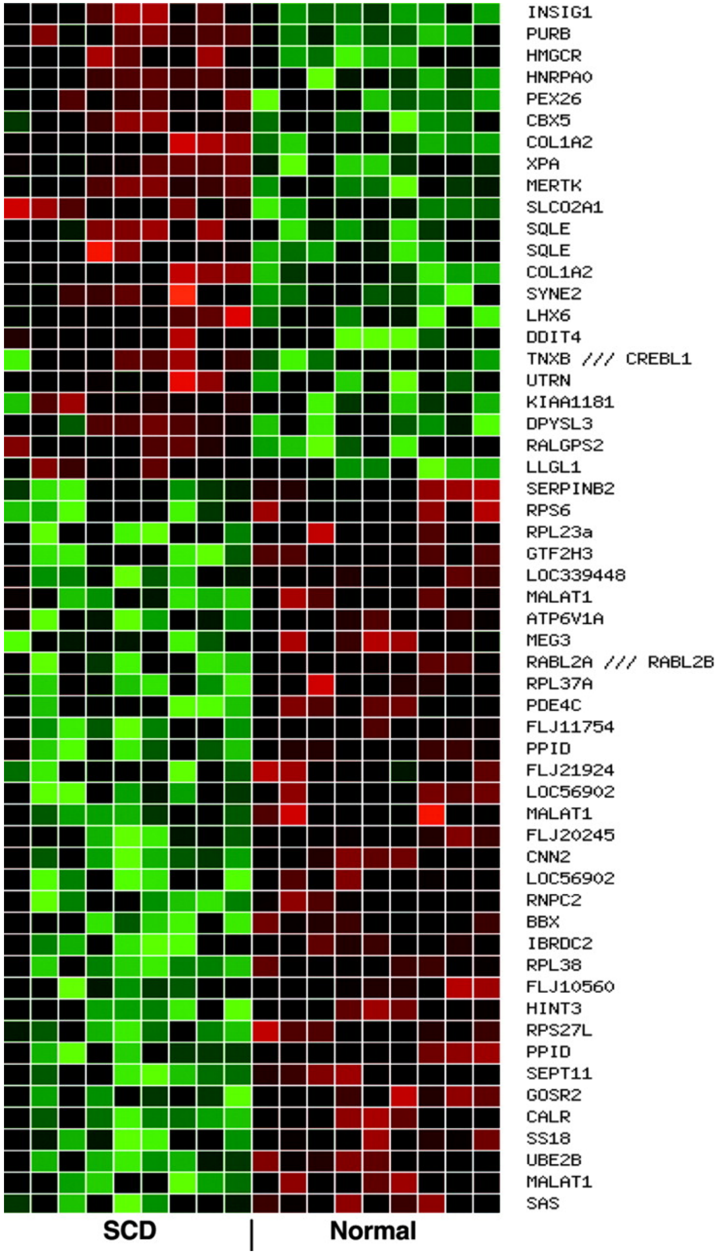
Drug X

Control vs X

Gene A

Differential gene expression heatmap

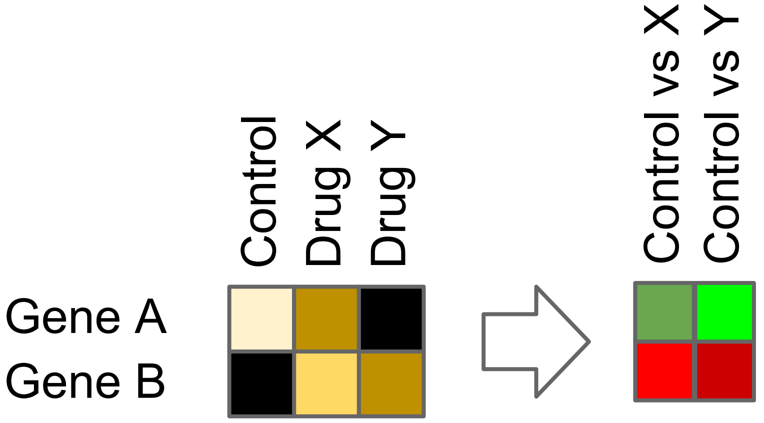




http://physiolgenomics.physiology.org/content/21/3/293/F1.expansion.html

What else do we need to encode?

- Uncertainty (statistical signifcance)
- We lose some information like expression levels. Can we bring them back?



Tasks

Find **differences, similarities, variability** between **groups of genes or conditions** in terms of direction of change in gene expression, its magnitude, its statistical significance, and baseline expression level

Data Set

Attribute	Type	Range
direction of change in gene expression	binary	up down
magnitude of change (baseline vs condition)	quantitative (log transformed)	0..6
p value (a measure of statistical significance)	quantitative	0..1
expression level at the baseline	quantitative	varies

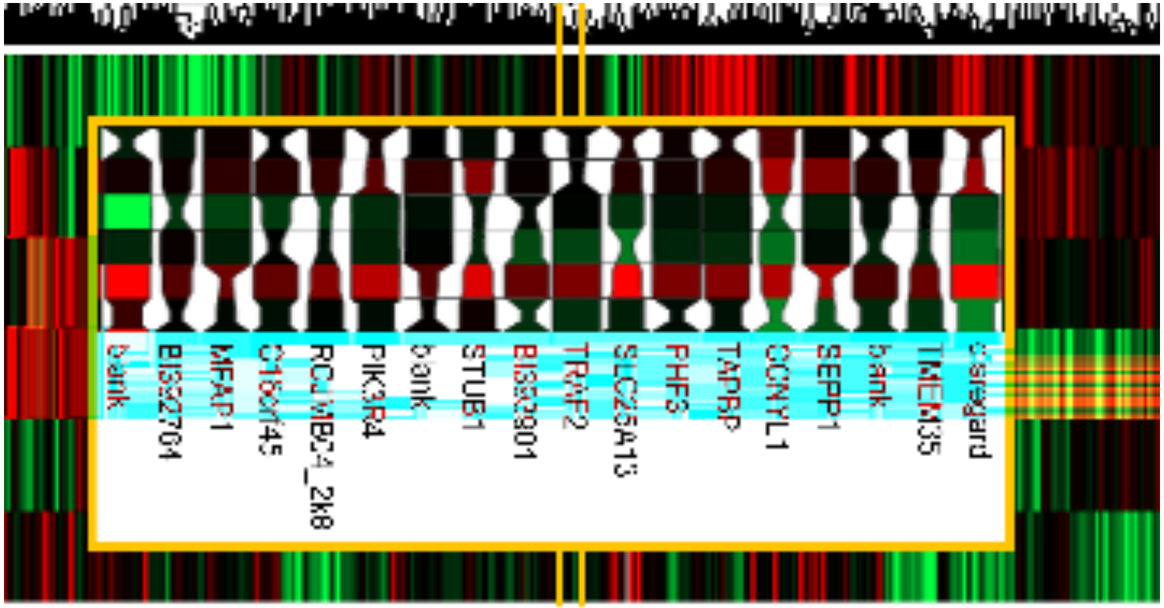
Data Set

Attribute	Type	Range
direction of change in gene expression	binary	up down
magnitude of change (baseline vs condition)	quantitative (log transformed)	~~0..6~~ 5 bins are ok
p value (a measure of statistical significance)	~~quantitative~~ ordinal	~~0..1~~ not signifcant low medium high
expression level at the baseline	~~quantitative~~ binary	~~varies~~ on/off modulation

are correlated

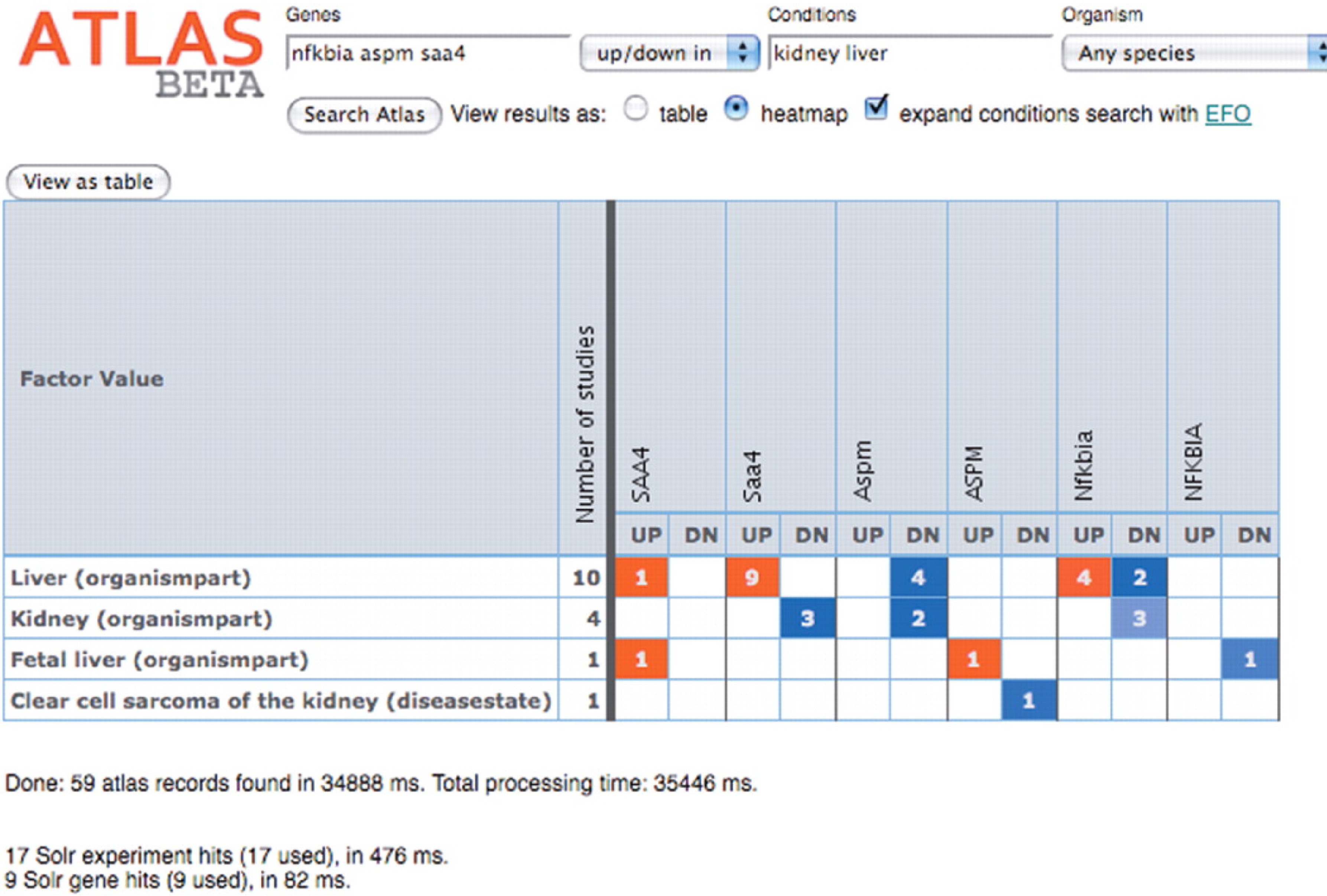
Dataset size : 200 x 200 cells = 8 x 5 pixels for each cell

Previous work



Enriched Heatmaps for Visualizing Uncertainty in Microarray Data
by Clemens Holzhüter, Hans-Jörg Schulz, and Heidrun Schumann - poster at Eurographics VCBM 2010

Previous work



Genes: nfkbia aspm saa4

up/down in

Conditions: kidney liver

Organism: Any species

View results as: ○ table ● heatmap ☑ expand conditions search with EFO

Factor Value	Number of studies	SAA4 UP	SAA4 DN	Saa4 UP	Saa4 DN	Aspm UP	Aspm DN	ASPM UP	ASPM DN	Nfkbia UP	Nfkbia DN	NFKBIA UP	NFKBIA DN
Liver (organismpart)	10	1		9			4			4	2		
Kidney (organismpart)	4				3		2				3		
Fetal liver (organismpart)	1	1						1					1
Clear cell sarcoma of the kidney (diseasestate)	1								1				

Done: 59 atlas records found in 34888 ms. Total processing time: 35446 ms.

17 Solr experiment hits (17 used), in 476 ms.
9 Solr gene hits (9 used), in 82 ms.

ArrayExpress : http://www.ebi.ac.uk/arrayexpress/

Visual Encoding

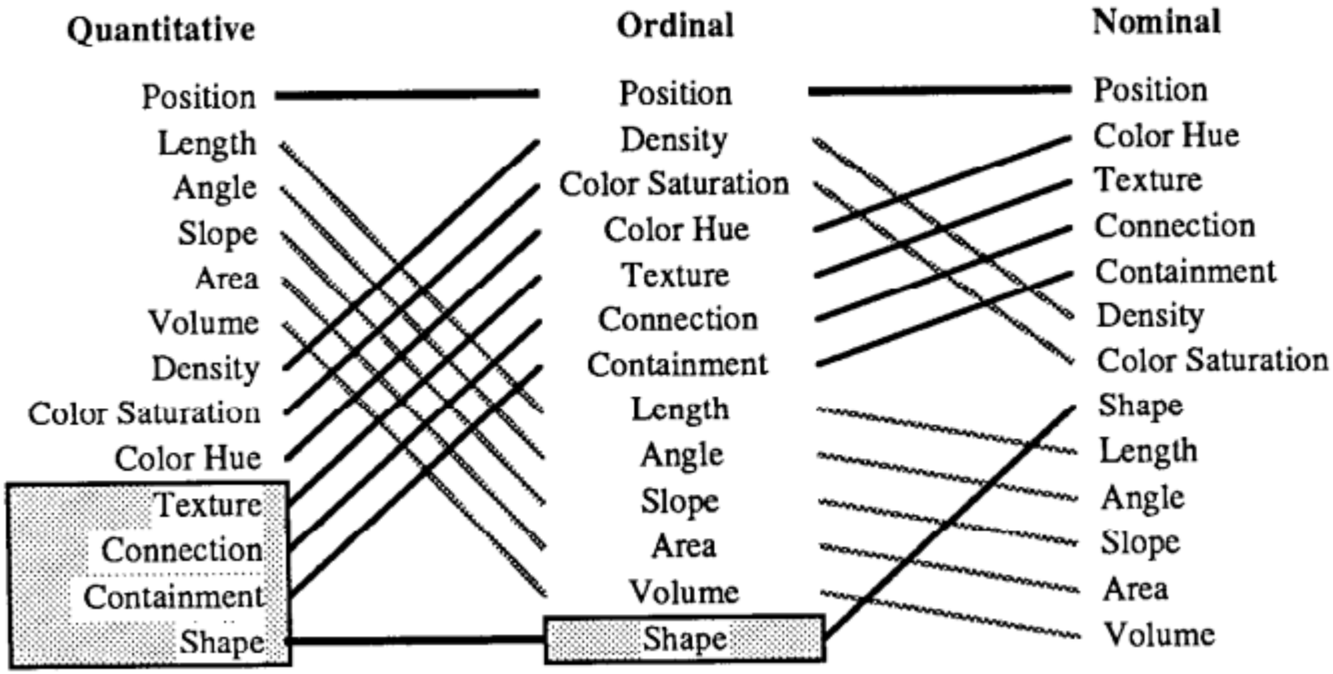


Fig. 15. Ranking of perceptual tasks. The tasks shown in the gray boxes are not relevant to these types of data.

Automating the Design of Graphical Presentations of Relational Information. Jock Mackinlay, ACM Transaction on Graphics, vol. 5, no. 2, April 1986, pp. 110-141.

Marks

Line (vertical or horizontal)
Rectangle

Visual Encoding

Attribute	Type	Range	Visual channel
direction of change in gene expression	binary	up down	position hue
magnitude of change	quantitative	5 bins	position length/area ligthness saturation
p value	ordinal	not signifcant low medium high	position length/area lightness saturation
expression level at the baseline	binary	on/off modulation	position hue

More constraints

An encoding may work for one cell but what about heatmap as whole?

Visual channel separability?

Use some of conventions established in the field (try to not surprise the user too much).

Progress

- Potential users interviewed
- Data type abstraction identified
- Simple heatmap drawing framework implemented
- Shrink the set of possible encodings even further
- Evaluate families of possible encodings
- Extract data
- Implement a couple of best solutions