

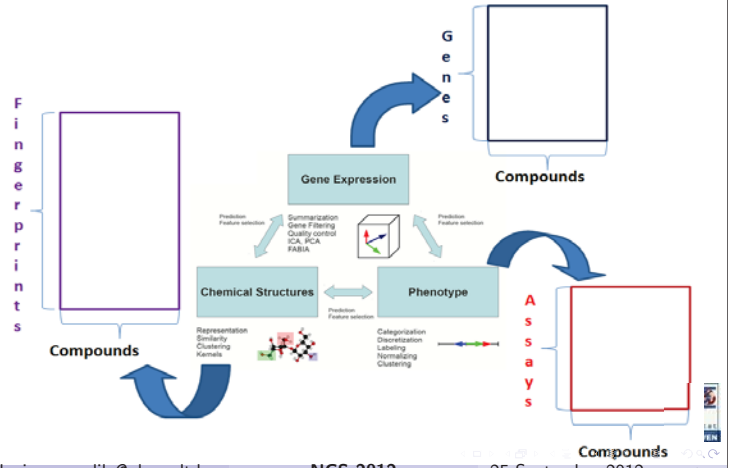
Integrated analysis of microarray data, biological data and chemical data in drug development using Multiple Factor Analysis

Nolen Joy Perualila
I-Biostat
Hasselt University, Belgium

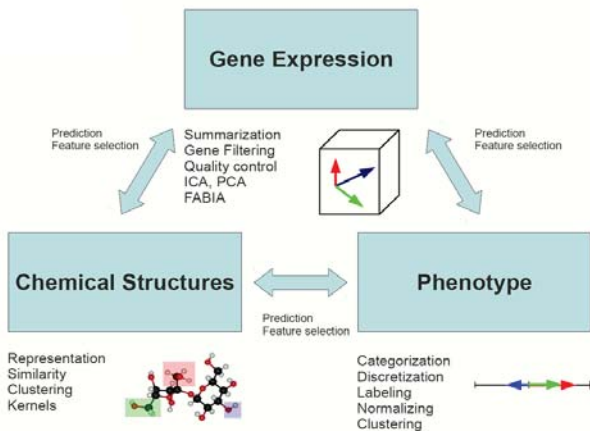
Non-Clinical Statistics Conference
Potsdam, Germany

25 September 2012

The Setting with Data Structure

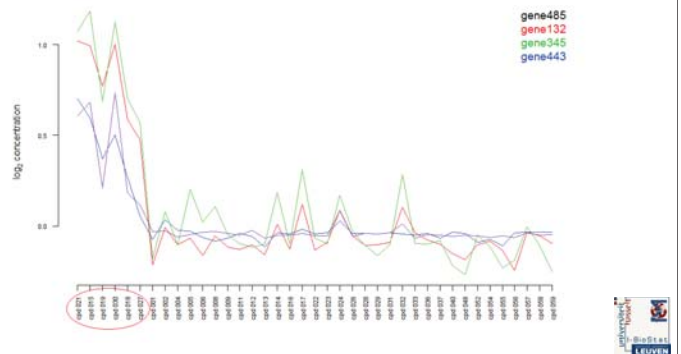


The Setting



Current Knowledge: Lead Genes and Compounds

- A cluster of compounds based on a set of genes.



Objectives

- identify subset of genes and bioassays that are pointing out to the same process or involved in a given pathway.
- find fingerprint features inducing higher/lower gene expression level and bioassay readouts in each pathway.

Method: Multiple Factor Analysis¹

- jointly analyze (quantitative and/or qualitative) data sets
- combine active variables defined on the same samples and bring out main factors(dimensions) of data variability.
- add supplementary variables

¹Escofier and Pagés (2002)

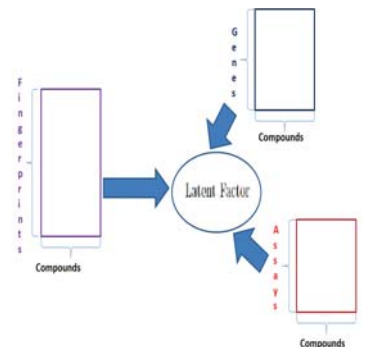
DataSet: common dimension

	Gene Expression	BioAssay	Fingerprint Features
	G1 G2 G3 ... GJ	B1 B2 ... BK	N= <chem>c1ccccc1</chem> ... N
Cpd1			1 0 1 ... 1
Cpd2			0 1 0 ... 1
Cpd3			
...			
CpdI			

- I = only 39 compounds(cpd)
- J = 566 Informative Genes (gene)
- K = 8 Bioassays (bio)
- L = 135 unique Fingerprint Features (fpf)

Method: Multiple Factor Analysis¹

- jointly analyze (quantitative and/or qualitative) data sets
- combine active variables defined on the same samples and bring out main factors(dimensions) of data variability. e.g Gene Expression and BioActivity.
- add supplementary variables e.g fingerprint features

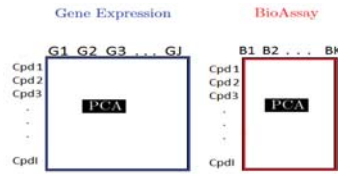


¹Escofier and Pagés (2002)

Method: MFA

Steps Involved:

- 1 Perform Principal Component Analysis (PCA) separately



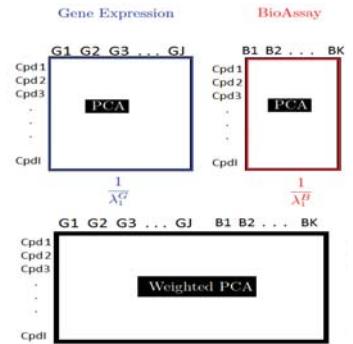
Method: MFA

Steps Involved:

- 1 Perform Principal Component Analysis (PCA) separately

-Weigh each variable.

- 2 Perform PCA on merged data

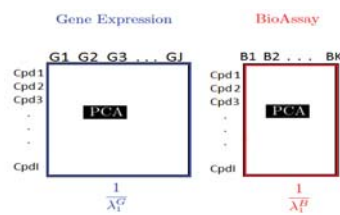


Method: MFA

Steps Involved:

- 1 Perform Principal Component Analysis (PCA) separately

-Weigh each variable.



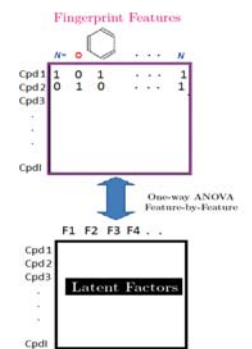
Method: MFA

Steps Involved:

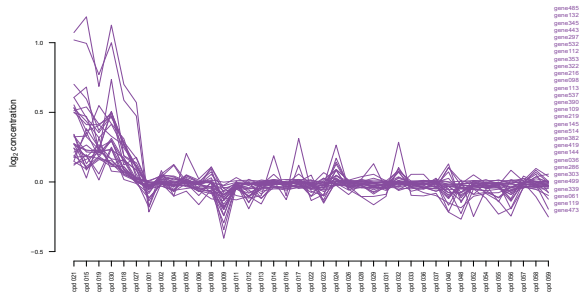
- 1 Perform Principal Component Analysis (PCA) separately

-Weigh each variable.

- 2 Perform PCA on merged data
- 3 do ANOVA for each derived factors controlling for the P/A of each fingerprint feature.

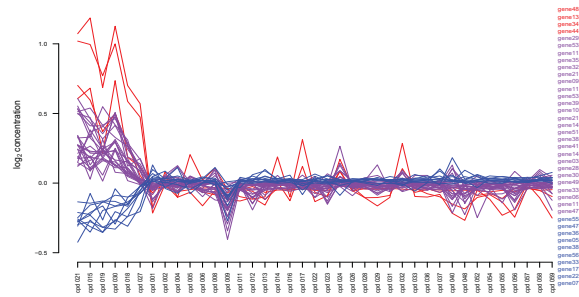


Upregulated Genes for Factor 2



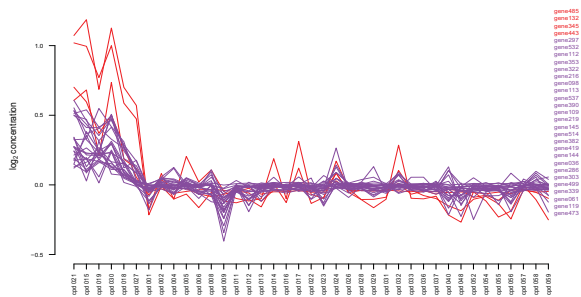
- Lead compounds give elevated expression levels for genes that are highly correlated to Factor 2.
- Other compound activities can also be noted for this set of genes

Genes for Factor 2



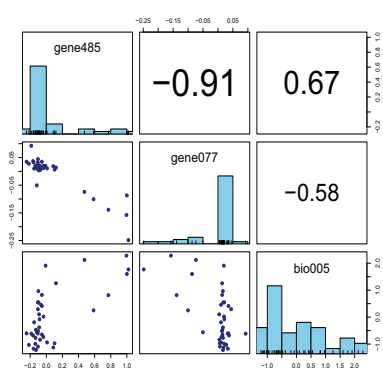
- Lead compounds have lower expression level than other compounds for some genes.

Upregulated Genes for Factor 2

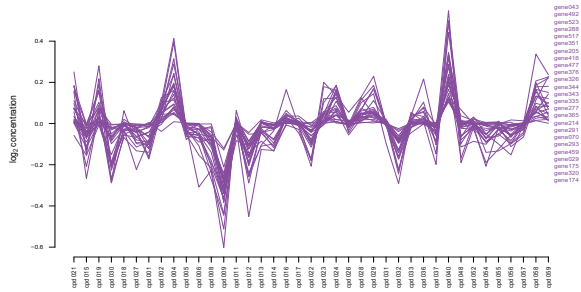


- The lead genes are part of the 2nd factor.
- Aside from the lead genes, there are other genes that have the same profiles across the compounds.

Genes and Bioassays for Factor 2

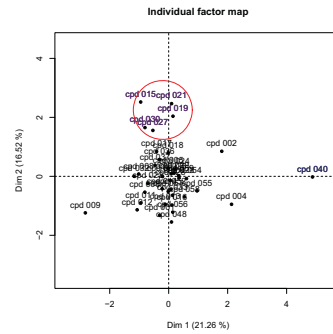


Upregulated Genes for Factor 1

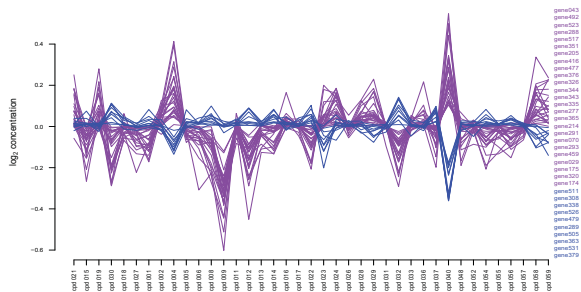


- Another set of genes that give the same profiles across compounds

MFA Results

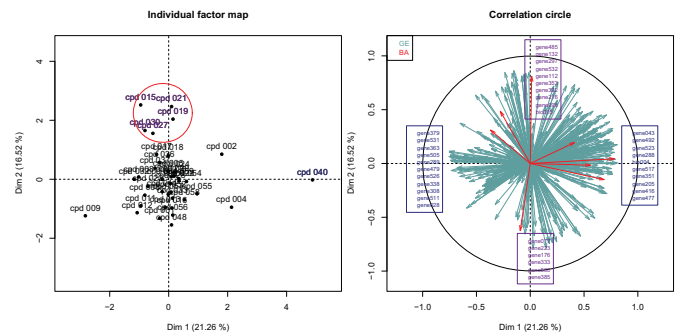


Genes for Factor 1



- One compound largely contributes to the structure of this group.

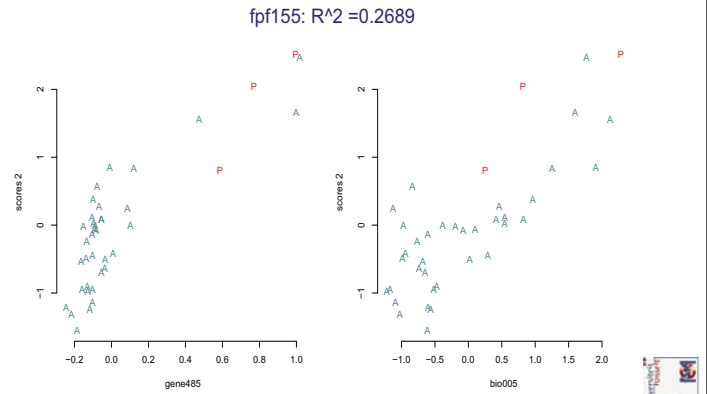
MFA Results



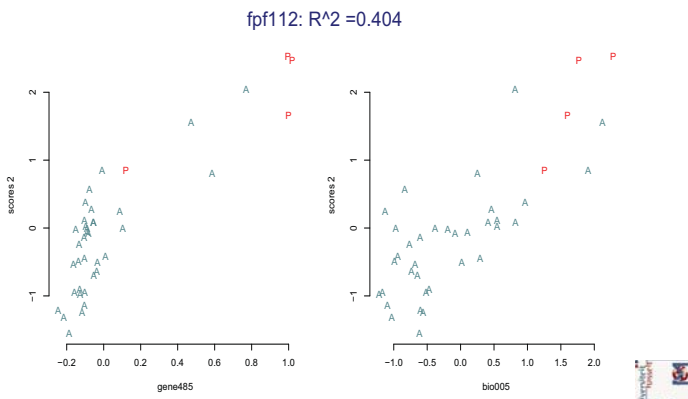
Top FPFs for Factor 2

CH	R2	p.value
fpf112	0.4040	0.0000
fpf058	0.2840	0.0005
fpf155	0.2689	0.0007
fpf020	0.2379	0.0016
fpf006	0.2379	0.0016
fpf144	0.2269	0.0022
fpf205	0.2202	0.0026
fpf199	0.2153	0.0029
fpf241	0.2090	0.0034
fpf177	0.1883	0.0058
fpf151	0.1703	0.0090

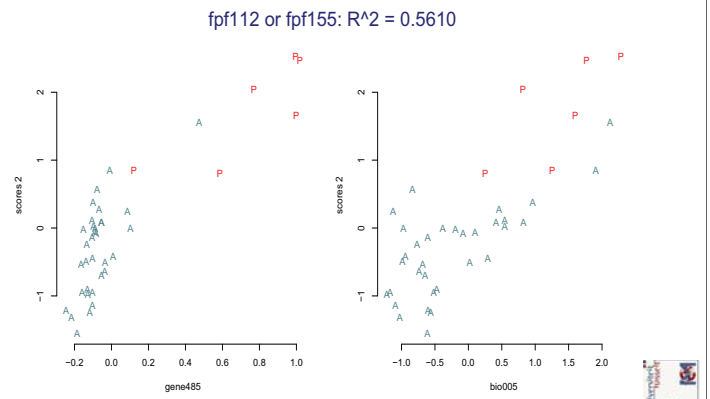
plot of gene/bioassay vs factor scores



plot of gene/bioassay vs factor scores



combination of 2 FPFs



Discussion

- starting with a known structure of interest, we are able to find
 - 1 more relevant genes
 - 2 pathways: Gene-Expression- BioAssay
 - 3 identify most discriminating fingerprint features
- in the future, FPF \rightarrow Target Prediction
- integration of Gene Expression, BioAssay, and Target Prediction Data



Research Team

- Interuniversity Institute for Biostatistics and statistical Bioinformatics (I-BioStat), Hasselt University
 - 1 Martin Otava
 - 2 Nolen Joy Perualila
 - 3 Pushpike Thilakarathne
 - 4 Tatsiana Khamiakova
 - 5 Ziv Shkedy
- Wolfson Institute for Research and Wellbeing, Durham University
 - 1 Adetayo Kasim

