Order Restricted Clustering for Dose-Response Microarray Data

Adetayo Kasim

Interuniversity Institute for Biostatistics and Statistical Bioinformatics

Universiteit Hasselt & Katholieke Universiteit Leuven

Joint work with:

Dan Lin , Ziv Shkedy , An De Bondt, Willem Talloen, Hinrinch Gohlman and Luc Bijnens

Outline of Presentation

- Introduction
- δ - biclustering
- Clustering of Dose-response data
- Application to Data
- Conclusion

Introduction

Dose-response microarray experiments

- Monitoring of gene expression with respect to increasing dose of compound
- To establish a dose-response relationship.
- To determine the shape of the relationship
- To identify the minimum effective dose.

Introduction

Dose-Response Data

- A cell line was treated with 3 compounds
- 4 doses per compound
- 3 rats per dose
- 16,998 genes

Structure

$$\begin{pmatrix} y_{1,1} & y_{1,2} & \cdots & y_{1,p} \\ y_{2,1} & y_{2,2} & \cdots & y_{2,p} \\ \vdots & \vdots & \ddots & \vdots \\ y_{n,1} & y_{n,2} & \cdots & y_{n,p} \end{pmatrix}$$

Introduction

Related Works

- Testing for trend in dose-response microarray experiments – Lin *et al.,* (2007a)

- Classification of trends in dose-response microarray experiments using information theory selection methods. - Lin *et al.,* (2007b)

Introduction

Clustering

- Hierarchical clustering
- K-means
- Self organizing maps (SOM)

Introduction

Biclustering

- clustering of genes under subset of conditions
- Madeira and Oliveira, 2004 reviewed biclustering methods
- δ – biclustering (Cheng and Church, 2000)

δ - Biclustering

Model

$$x_{ij} = \mu + \alpha_i + \beta_j + r_{ij}$$

Similarity Score

$$H_{IJ} = \sum_{i,j} \frac{r_{ij}^2}{|\mathrm{I}||\mathrm{J}|}$$

δ - Criterion

$$H_{IJ} <= \delta$$

$$\delta >= 0$$

Clustering of Dose-response Data

Clustering of Dose-Response Data

Model

- For Each gene;

$$y_{jk} = \mu(d_j) + \varepsilon_{jk}$$

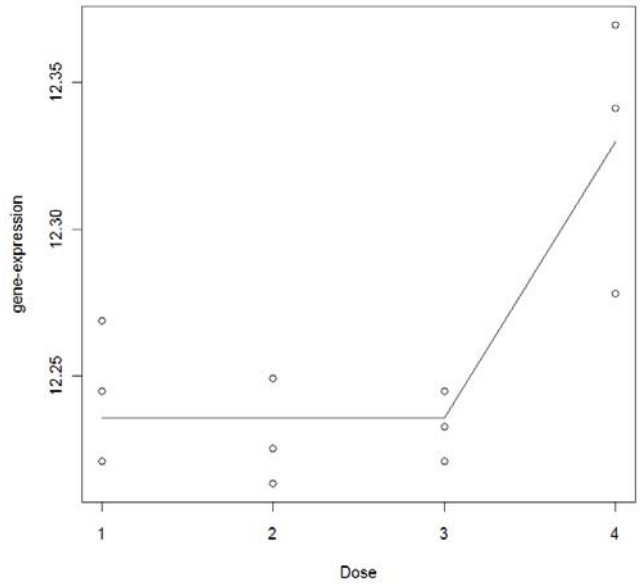
- Ordered Constraints

$$\mu_1 \leq \mu_2 \leq \cdots \mu_p$$

or

$$\mu_1 \geq \mu_2 \geq \cdots \mu_p$$

Example



Clustering of Dose-Response Data

- Clustering using observed data and isotonic means

$$y_{ijk} = \mu + \alpha_i + \beta_j^* + r_{ijk}$$

- Clustering using only isotonic Means

$$\mu_i(d_j) = \mu + \alpha_i + \beta_j^* + r_{ij}$$

Clustering of Dose-Response Data

δ- clustering

- clustering of genes under all conditions
- relative choice for delta
- specification of minimum members of a cluster (phi)

Choice of δ

$$\delta = \lambda * H$$

$$0 \leq \lambda \leq 1$$

Clustering of Dose-Response Data

Algorithm 1: δ - clustering

Input: Y, a matrix of real number; ϕ, minimum number of genes in a cluster; and λ: $0 \leq \lambda \leq 1$

Output: Y_{IJ}, a subset of Y with rows set I and Column set J with score not larger than δ or $I \leq \phi$

Initialization: $\delta = \lambda * H$, where H is the mean squared residue score of the observed data.

Clustering of Dose-Response Data

Algorithm 1: δ - clustering

Iteration :

1. Apply node deletion algorithm of Cheng and Church (2000) only in gene direction with fixed conditons/dose levels.
2. if mean squared residue score of the reduced matrix satisfies δ criterion or number of genes in the reduced matrix is at most ϕ, then output the reduced matrix as a cluster.
3. Delete members of cluster found in step 2.
4. Repeat Steps 1 to 3 on the non-clustered gene until every gene belongs to a cluster.

Clustering of Dose-Response Data

Algorithm 2: Order restricted clustering based on observed data and isotonic means

Input: Y, a matrix of real number; Y^* a matrix of isotonic means, ϕ minimum number of genes in a cluster; and λ:

$$0 \leq \lambda \leq 1$$

Output: Y^*_{IJ}, a subset of Y^* with rows set I and Column set J ; with score no larger than δ or $I \leq \phi$

Clustering of Dose-Response Data

Algorithm 2: Order restricted clustering based on observed data and isotonic means

Initialization: $\delta = \lambda * H$, where H is the mean squared residue score of the observed data.

Iteration :

1. Using global likelihood ratio statistics, assign each gene to a direction
2. Apply Algorithm 1 using model; $y_{ijk} = \mu + \alpha_i + \beta_j^* + r_{ijk}$

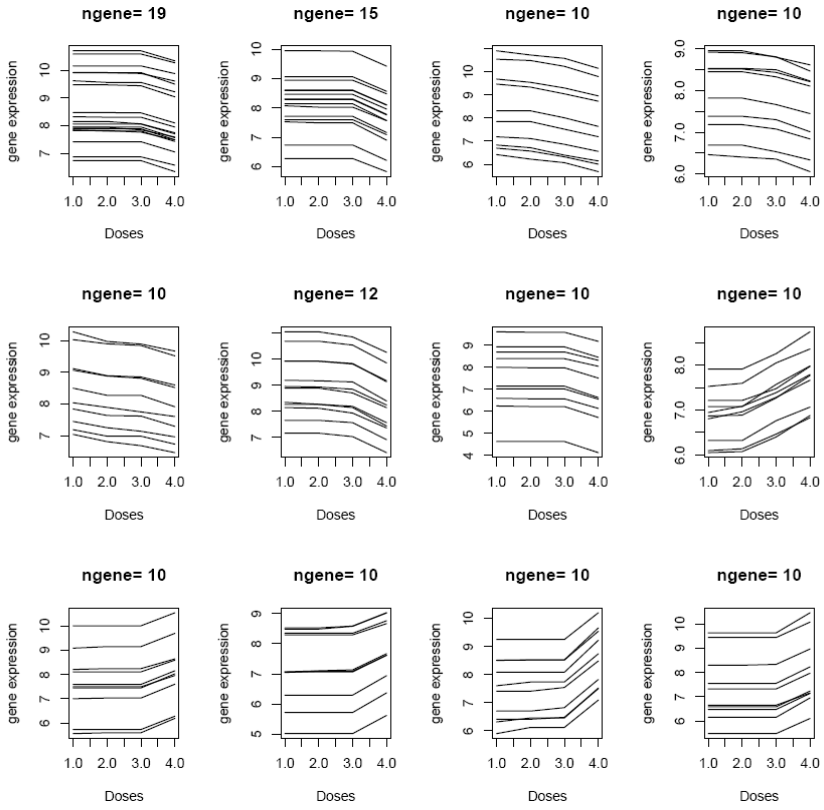
Application to Data

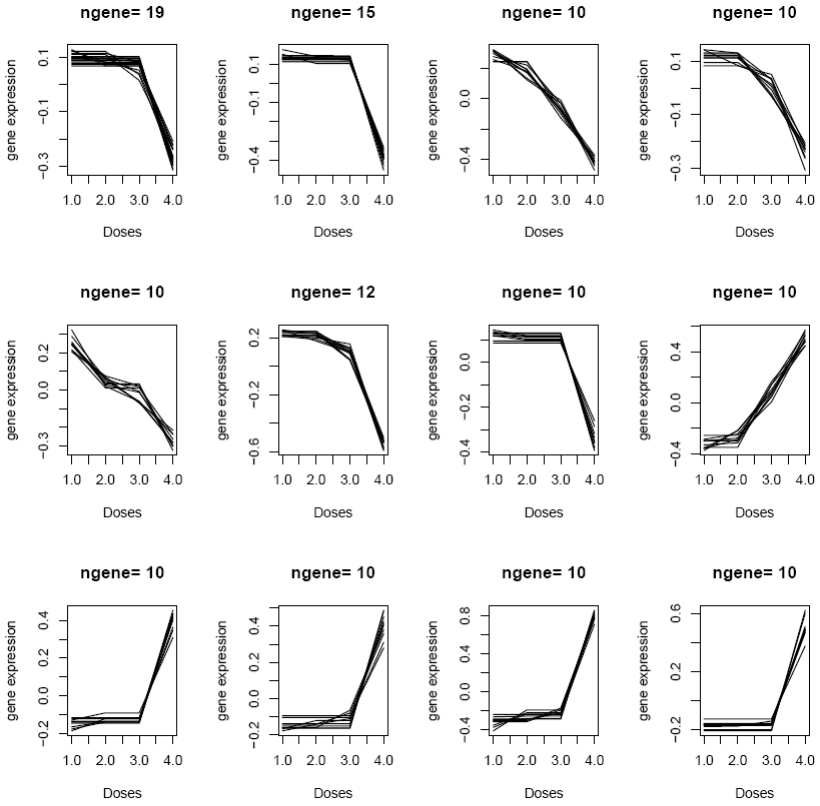
Initial Filtering

- Global likelihood ratio test – Lin *et al,* (2007)
- Clustering only significant genes

Application to Data

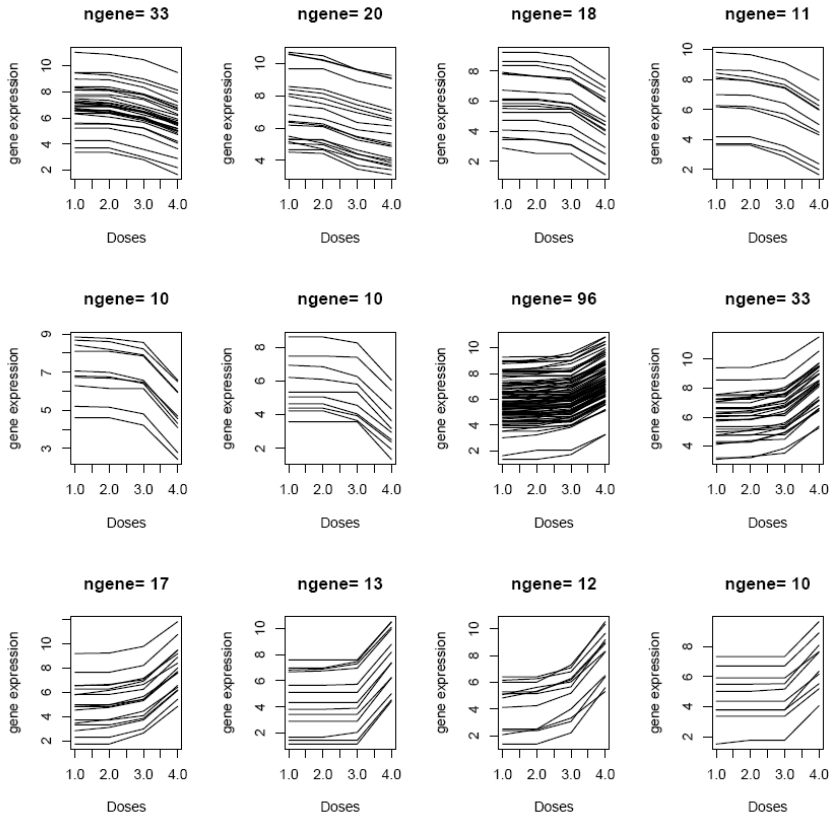
Observed Data and Isotonic Means

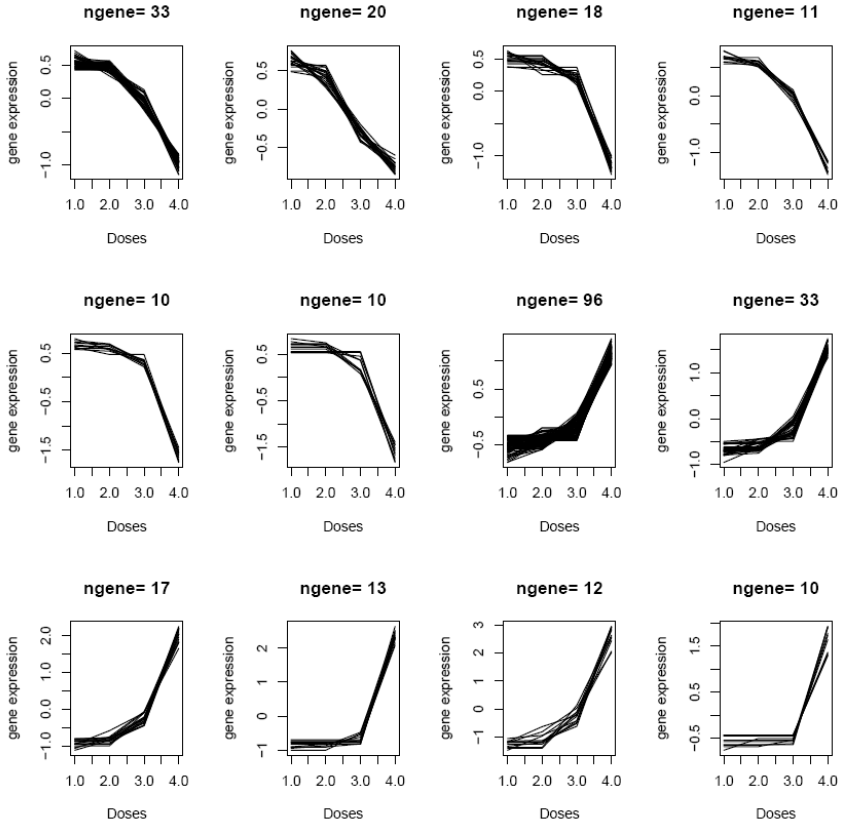




Application to Data

Isotonic Means

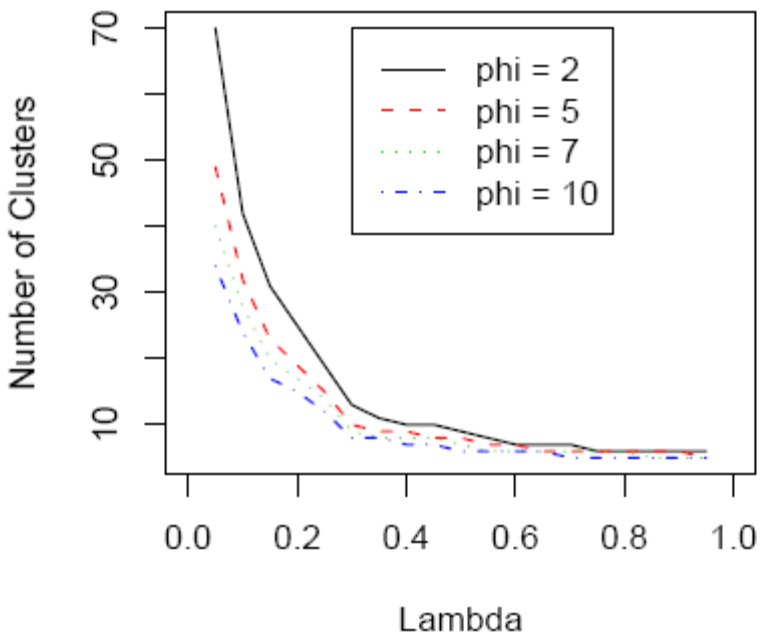




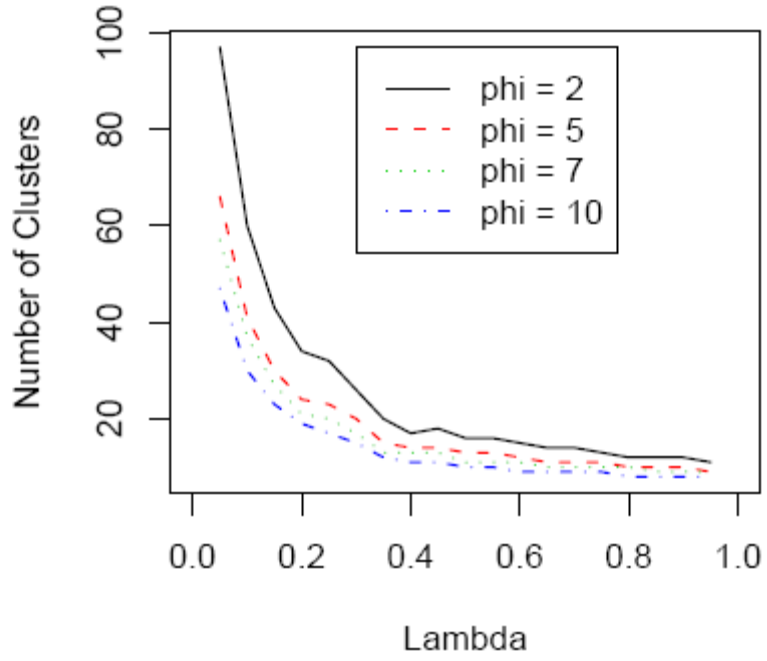
Application to Data

Choice of Lambda and phi

up



Down



Conclusion

- fast exploratory tool for dose-response microarray data
- resulting clusters have intrinsic ordering
- quality and number of clusters depends on choice of lambda and phi
- the method can be used with or without initial filtering

THANK YOU