# Order Restricted Clustering for Dose-Response Microarray Data

**Adetayo Kasim**

Durham University

# Outline of Presentation

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

Durham University

# Motivating Example

**Dose-Response Data - **Lin, *et al.*, ( 2007)

- Human epidermal squamous carcinomal cell line (A431)
- Treated with 3 compounds
- Per compound, 4 doses were given .
- Each treatment was given to 3 independent rats.
- 16,998 genes

Durham
University

# Motivating Example

## 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)
  - Depends on number of doses

Durham University

# Motivating Example

Unsupervised pattern discovery

- In high-dimensional data, one needs to specify a priori what one wants to detect.
- We expect dose-dependent changes to be of biologically interest.

Durham University

# Part I: Biclustering

# Introduction

## Biclustering

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

Durham University

# δ - Biclustering

**Model**

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

Durham University

# δ - Biclustering

**Mean Squared Residue Score**

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

Durham University

# δ - Biclustering

$$A = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 3 & 4 & 5 & 6 \\ 30 & 31 & 32 & 33 & 34 \\ 32 & 33 & 34 & 35 & 36 \\ 81 & 82 & 83 & 84 & 85 \\ 91 & 92 & 93 & 94 & 95 \end{pmatrix}$$

$$B = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 3 & 4 & 5 & 6 \\ 30 & 31 & 32 & 33 & 34 \\ 32 & 33 & 34 & 35 & 36 \\ 42 & 43 & 30 & 30 & 31 \\ 37 & 30 & 36 & 35 & 34 \end{pmatrix}$$

$$H_A = 0$$

$$H_B \neq 0$$

Durham University

# δ - Biclustering

## δ - Criterion

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

$$\delta >= 0$$

Durham University

# Part II: Clustering Dose-Response Relationship

Durham University

## Clustering of Dose-Response Data

**With Order restriction**

- For Each gene in a data matrix Y;

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

$$j = 0, 1, \ldots, J$$

$$k = 1, 2, \ldots, n_j$$

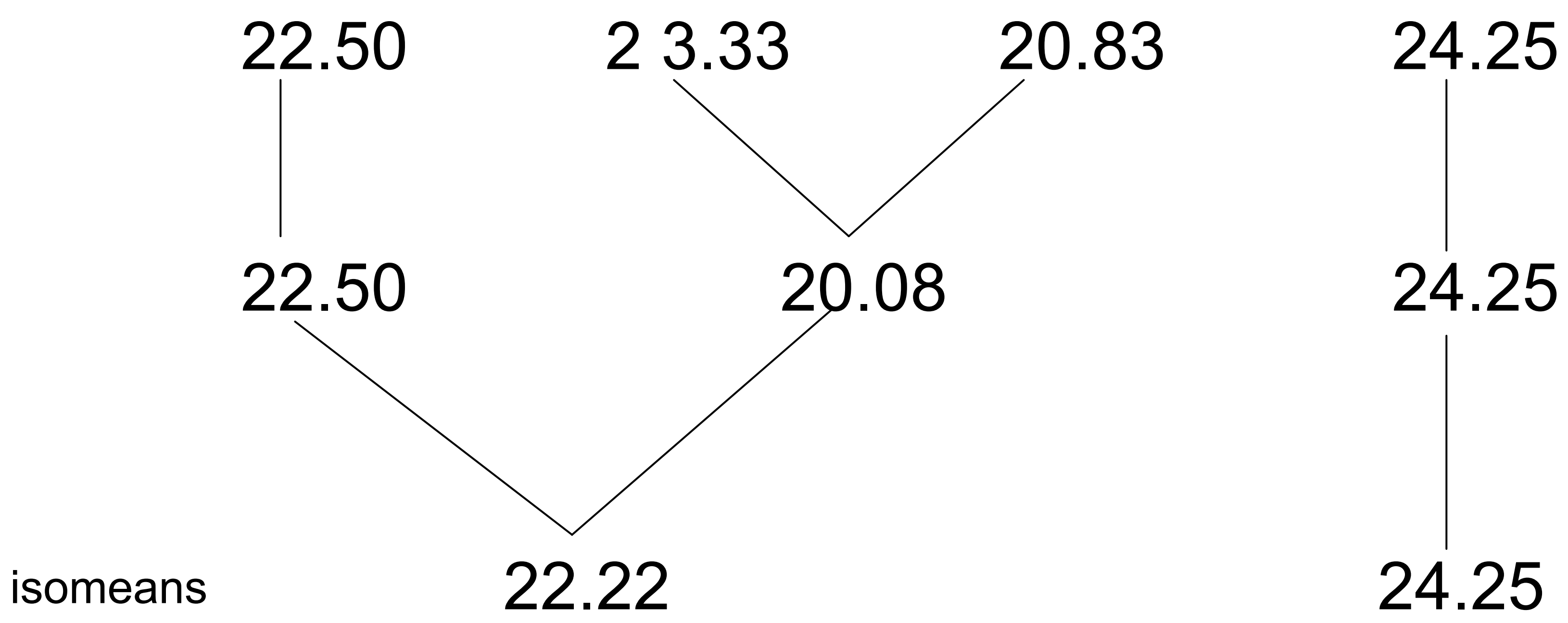
- Under monotone constraints

Durham University

# Clustering of Dose-Response Data

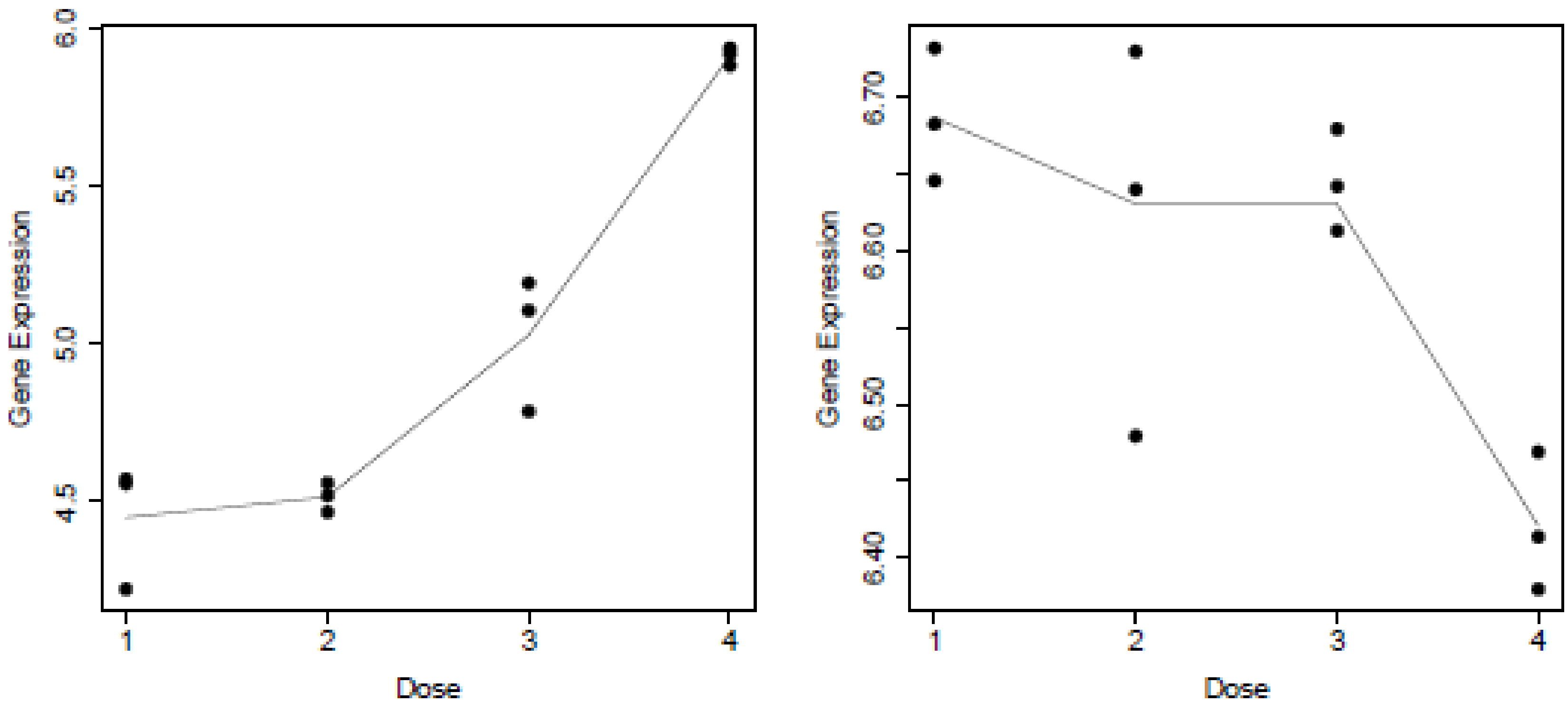
## Monotone increasing means

- Pool-adjacent-violators algorithm (PAVA)




Durham University

# Clustering of Dose-Response Data



# Clustering of Doses-Response Data

## δ - clustering

- Clustering of genes under all conditions
- Relative choice for lambda
- Specification of minimum number of genes in a cluster (φ)

Durham University

# Clustering of Dose-Response Data

**Model**

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

Durham University

## Clustering of Dose-Response Data

**Choice of $\delta$**

$$\delta = \lambda * H_P$$

$$0 <= \lambda <= 1$$

Durham University

# 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 – relative precision

**Output:** ***X*** - a subset of ***Y***

**Initialization:** $$\delta = \lambda * \mathrm{H}_\mathrm{p}$$

Durham University

# Clustering of Dose-Response Data

**Algorithm 1:** $\delta$ - 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 $\delta$ - criterion or number of genes in the reduced matrix is at most $\varphi$, 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.

Durham University

# Clustering of Dose-Response Data

**Algorithm 2**: Order restricted clustering

**Input:** ***Y**** **-** a matrix of isotonic means, φ - minimum number of genes in a cluster, and λ – relative precision .

**Output:** ***X**** - a subset of ***Y***

**Initialization:** $\delta = \lambda * \mathrm{H}_{\mathrm{p}}$

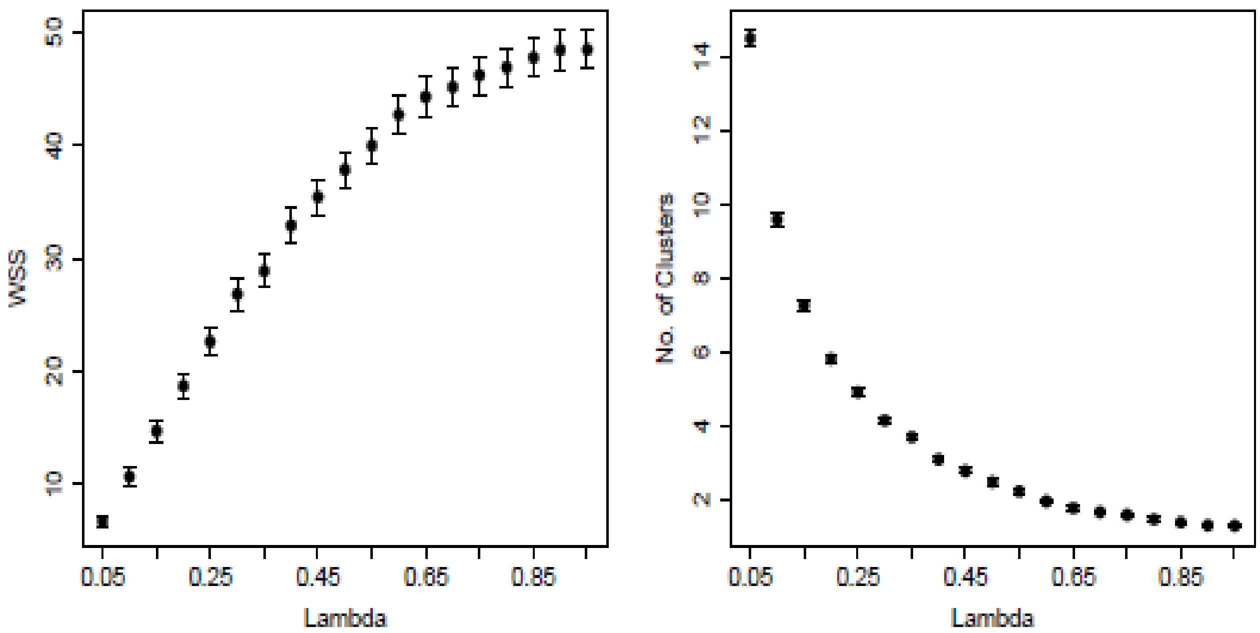
**Iteration :**

1. Determine monotone direction
2. Apply Algorithm 1

Durham University

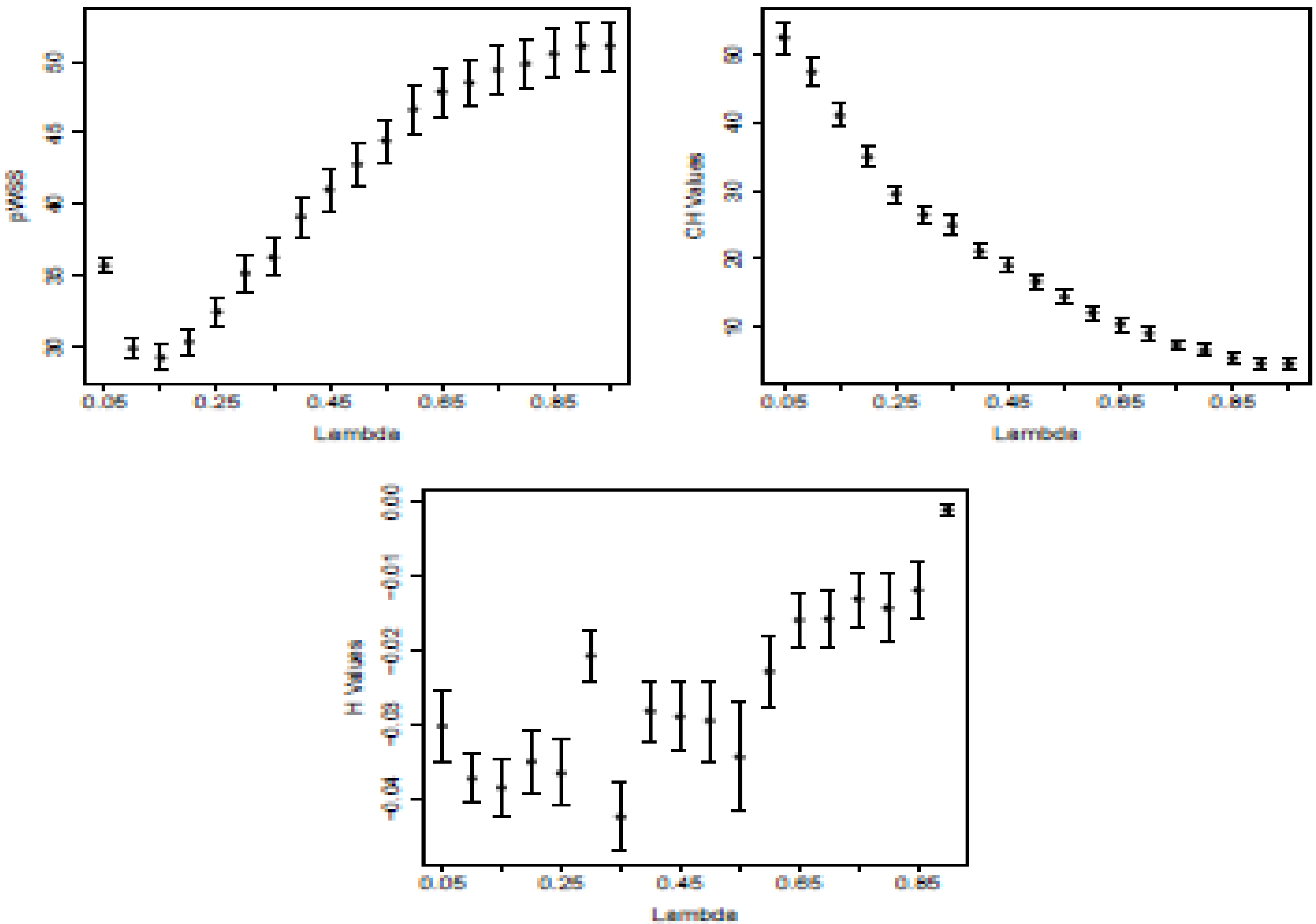
# Choice of  Clustering Parameters

- Within Cluster sums of squares (WSS)  and number of clusters




Durham University

# Choice of Clustering Parameters




Durham University

# Part III: Application to Data

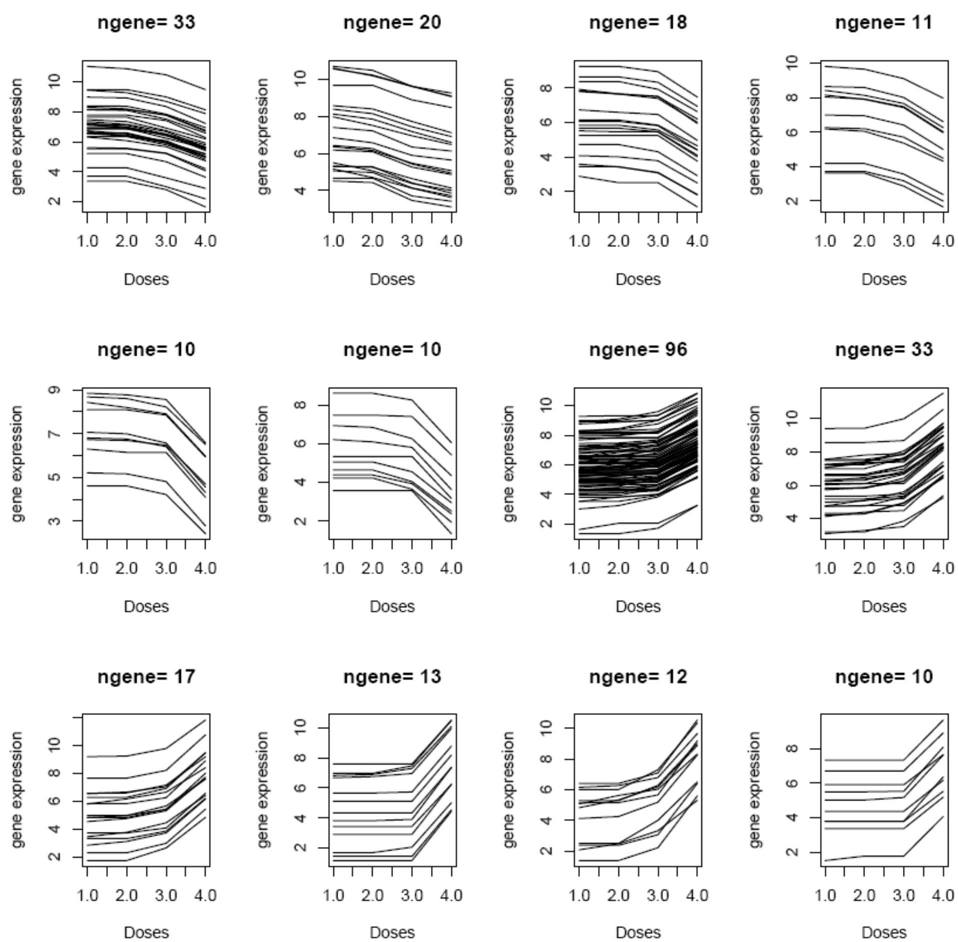
# Application to Data

## Initial Filtering

- $E^2$ is used to select genes with significant trends - Lin, *et al., (* 2007)
- Order restricted clustering is applied only on the significant genes

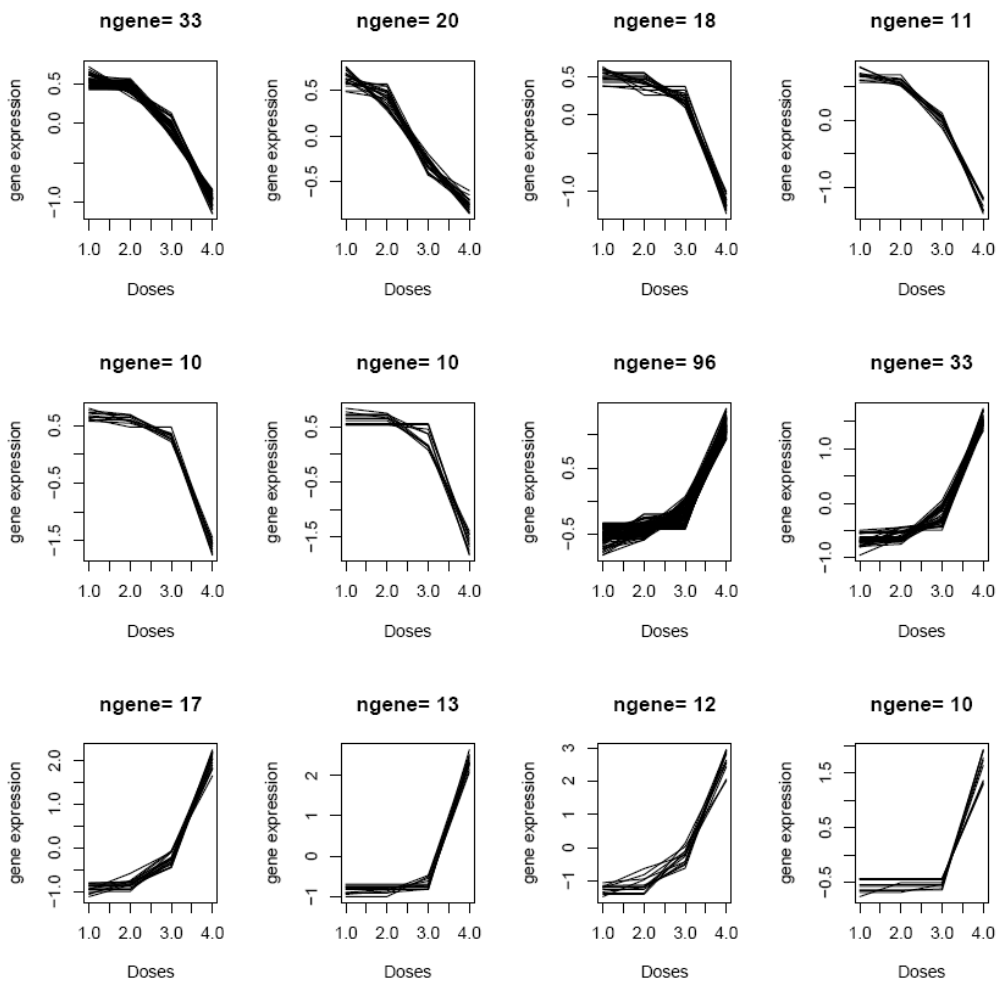
Durham University

# Application to Data




Durham University

# Application to Data




Durham University

# 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 $\varphi$.

Durham University

# Acknowledgement

**Hasselt University**

- Dan Lin
- Ziv Shkedy

**Open Analytics**

- Tobias Verbeke

**Janssen Pharmaceutica**

- Willem Talloen
- Hinrich Goelhmann
- An De Bondt
- Luc Bijnens


Durham University

# THANK YOU

Durham University