Modeling Dose-response Microarray Data in Early Drug Development Experiments

Dan Lin
I-Biostat, Hasselt University
09/29/2010

Overview

- Introduction to dose-response modeling in microarray experiments
- Focus: dose-response modeling+ applications
  - Test for trend
  - Classification of trends
- Concluding remarks and related research

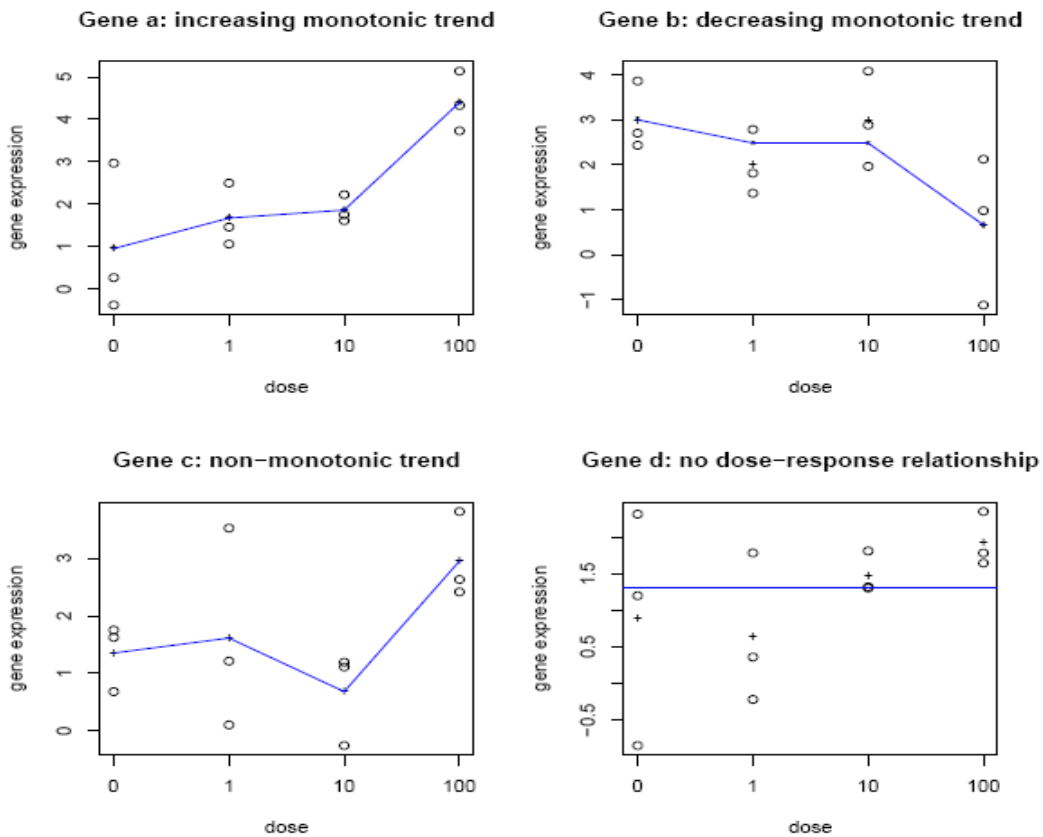
Introduction to Dose-response study

- Aim
  - To understand mechanism of action
  - To explore the desired properties

- Biological information from gene expression data create new opportunities for developing effective therapies
  - Identify drug target
  - Explore functions of genes/pathways in a dose-dependency manner

Application of Microarrays in Drug Discovery:

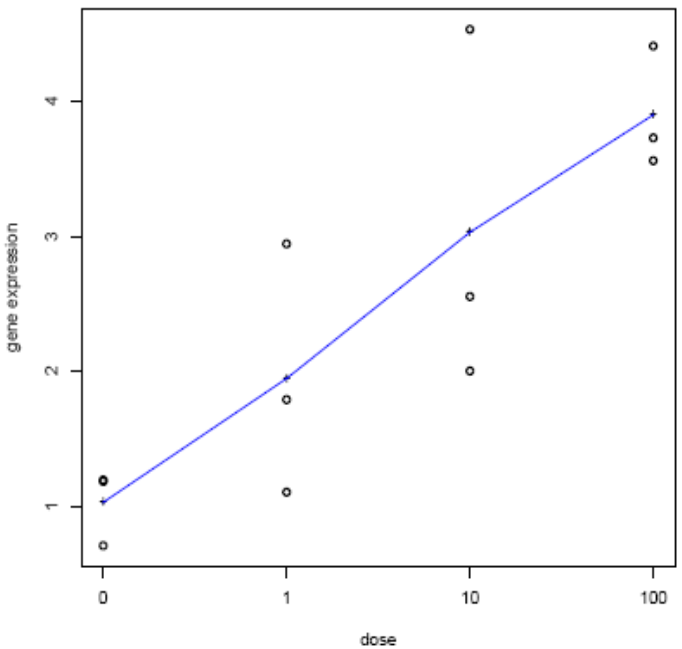
- Pharmacology study



Case Study for Dose-response Modeling

- Human epidermal squamous carcinoma cell-lines

	EGF (ng/ml)			
Dose	0	1	10	100
Control	3	3	3	3

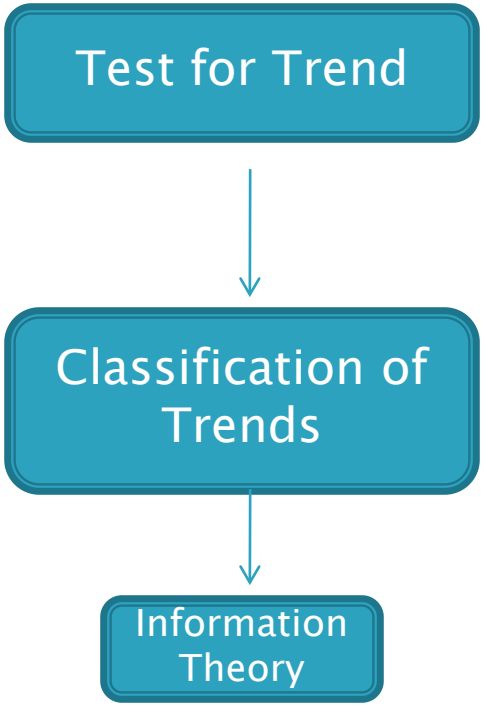


  - 16,998 genes measured on each array

Dose-response Modeling

- Two main research questions:
  - Is there a dose-response relationship?
    ⇨ Test for trend
  - what's the nature of dose-response relationship?
    ⇨ Classification of dose-response curve shapes

Dose-response Modeling



Test for Trend

- The setting:

$$y_{ij} = f(\theta, \mathbf{x_i}) + \epsilon_{\mathbf{ij}}$$

  - Non-parametric method: Isotonic Regression
    - Without the knowledge of dose-response shape
    - Mechanistic model

  - Parametric modeling: e.g. 4PL Regression
    - Prior knowledge of dose-response shape
    - Empirical model

Test for Trend

- For gene (i=1, …, m) with K dose (j=1,…,K)

  H_0: $\mu_1 = \mu_2 = \ldots = \mu_k$

  H_1^{Up}: $\mu_1 \leq \mu_2 \leq \ldots \leq \mu_k$

  or

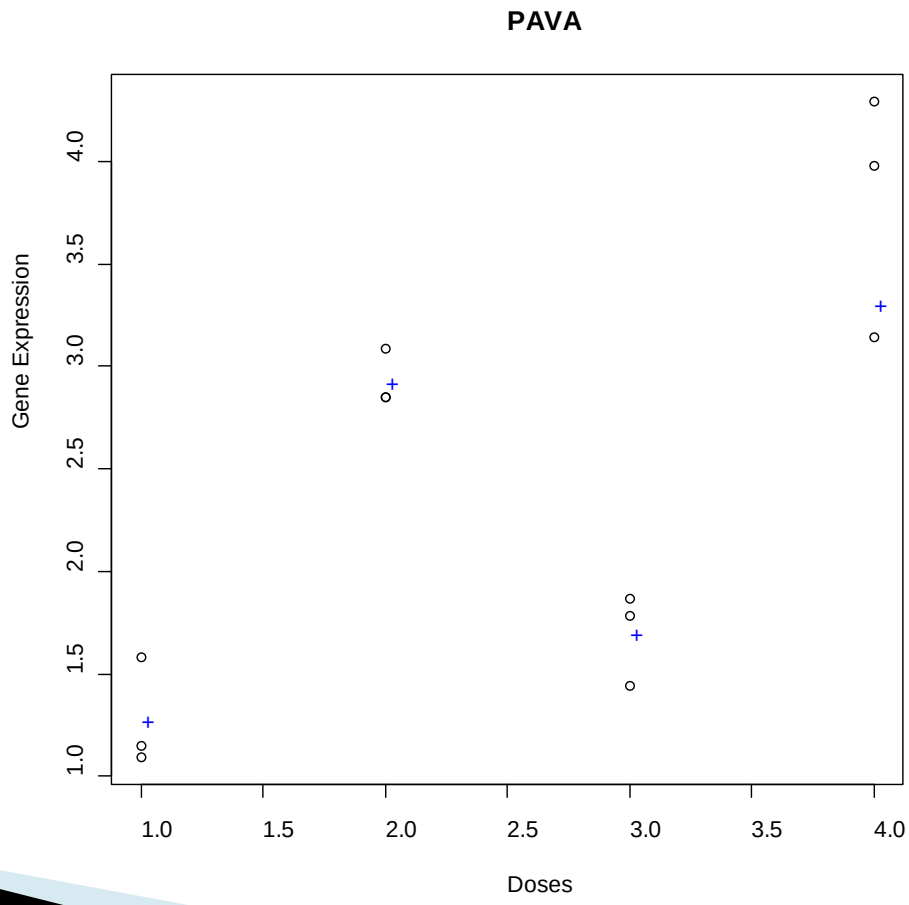
  H_1^{Down}: $\mu_1 \geq \mu_2 \geq \ldots \geq \mu_k$

  with at lease one inequality

- Pooled-adjacent-violator-algorithm (PAVA) to obtain estimates of the isotonic means $\hat{\mu}^*$

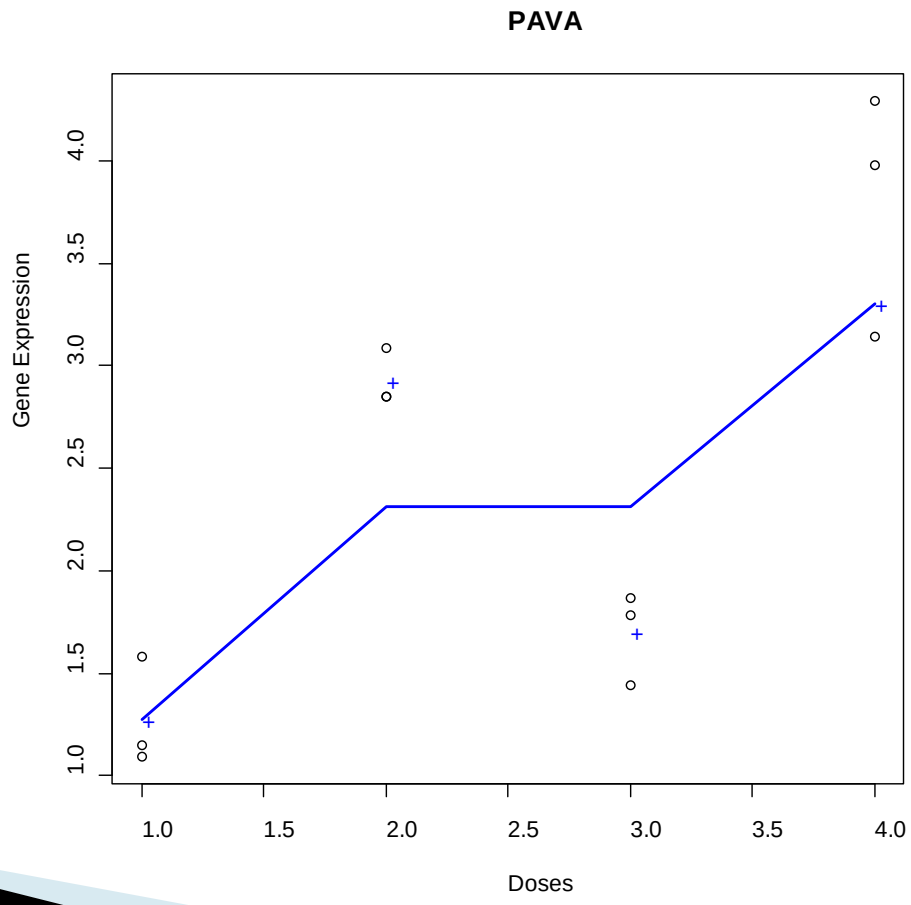
Estimation of Isotonic Means

- Data of one gene



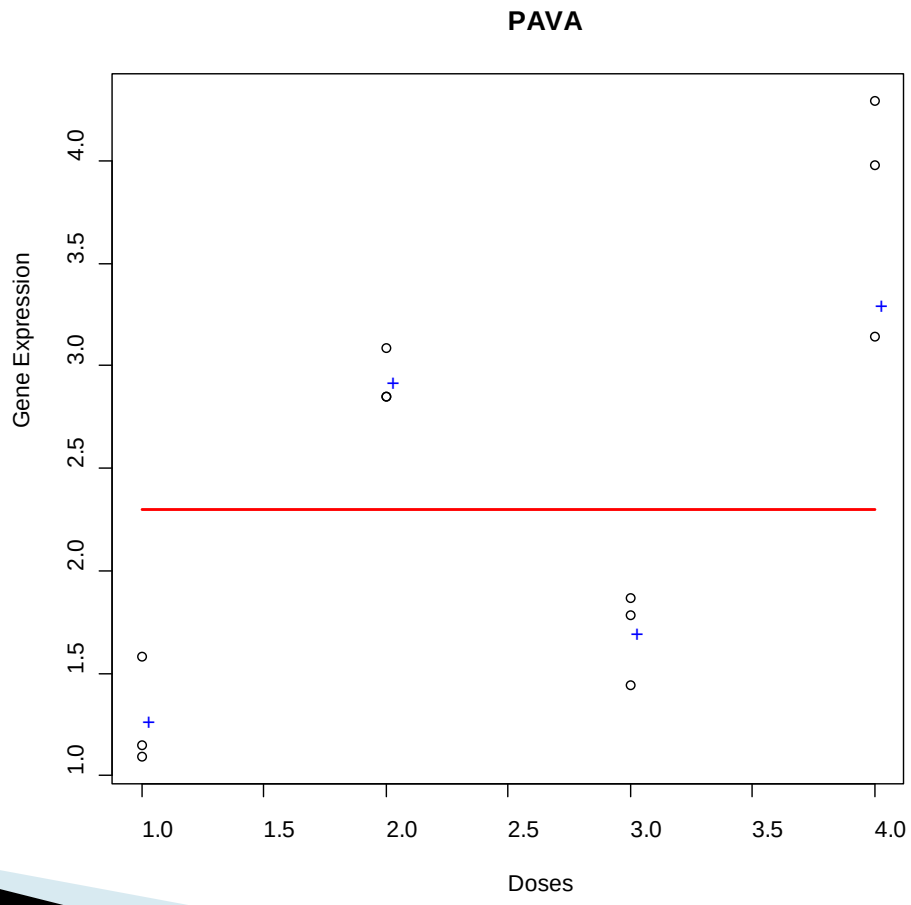
Estimation of Isotonic Means

- Under increasing constraints



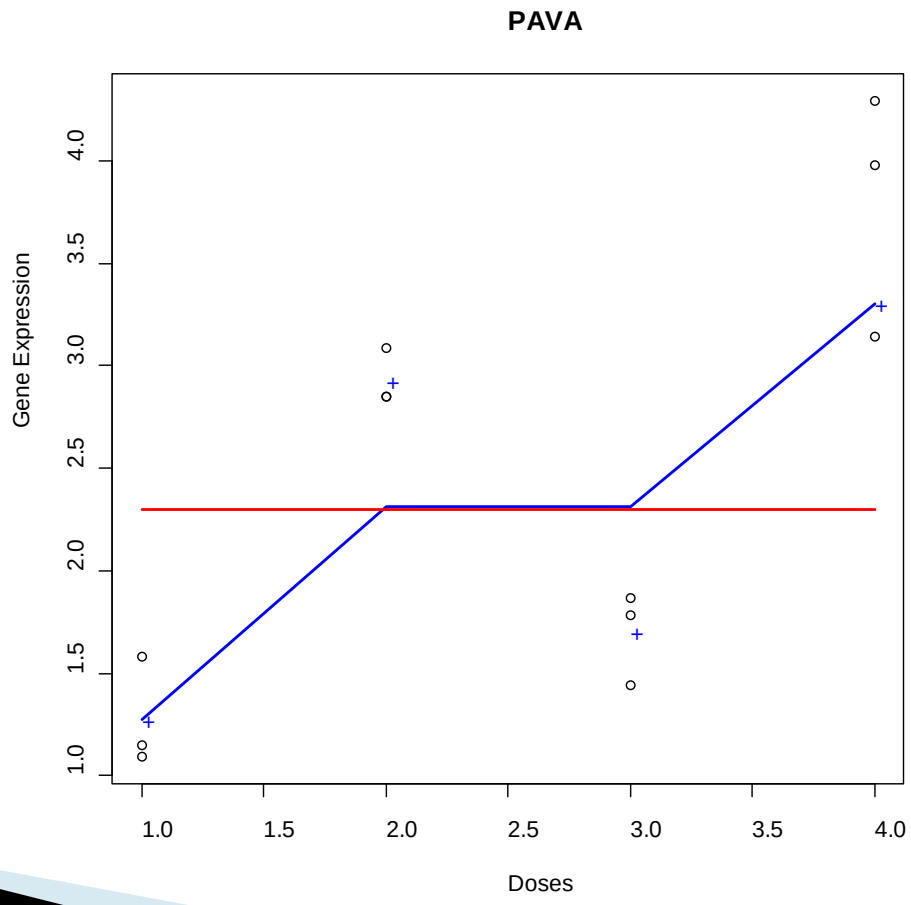
Estimation of Isotonic Means

- Under decreasing constraints

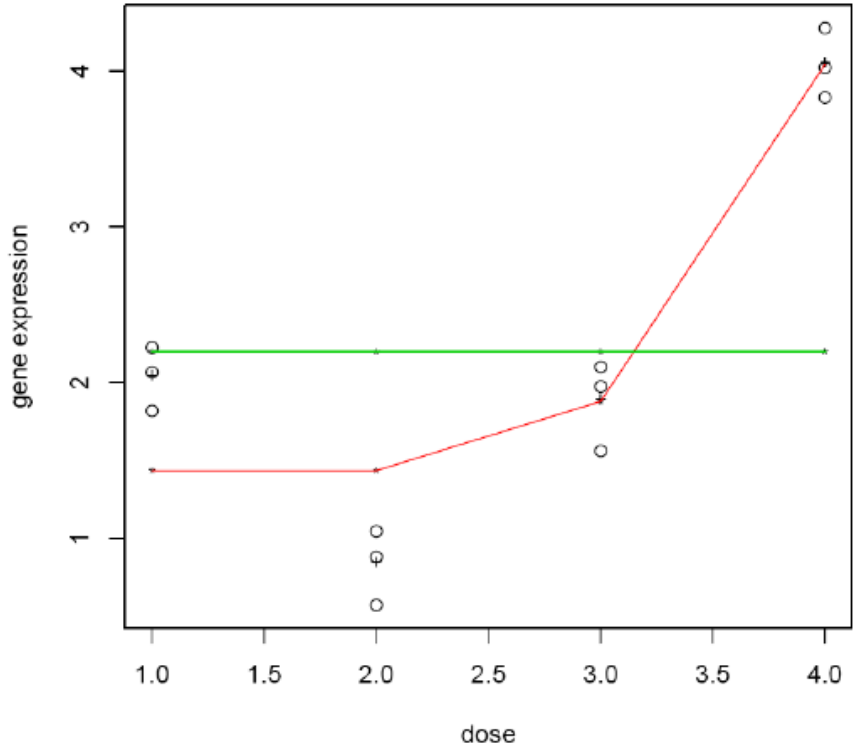


Estimation of Isotonic Means

- Under decreasing constraints



Test Statistics

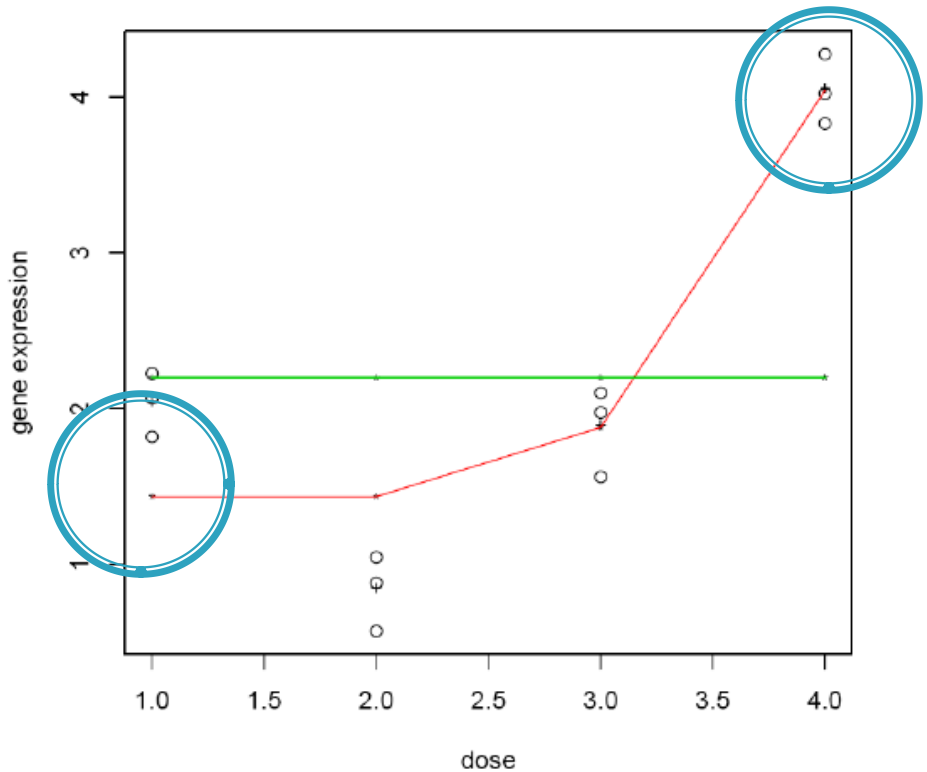


- LRT: $$\lambda^{2/N} = \frac{\hat{\sigma}^2_{H_1}}{\hat{\sigma}^2_{H_0}}$$

(Bartholomew 1959)

- Direction of trend is unknown in advance
- In practice, we calculate LRT statistics twice for each direction

Test Statistics



- LRT: $$\lambda^{2/N} = \frac{\hat{\sigma}^2_{H_1}}{\hat{\sigma}^2_{H_0}}$$

(Bartholomew 1959)

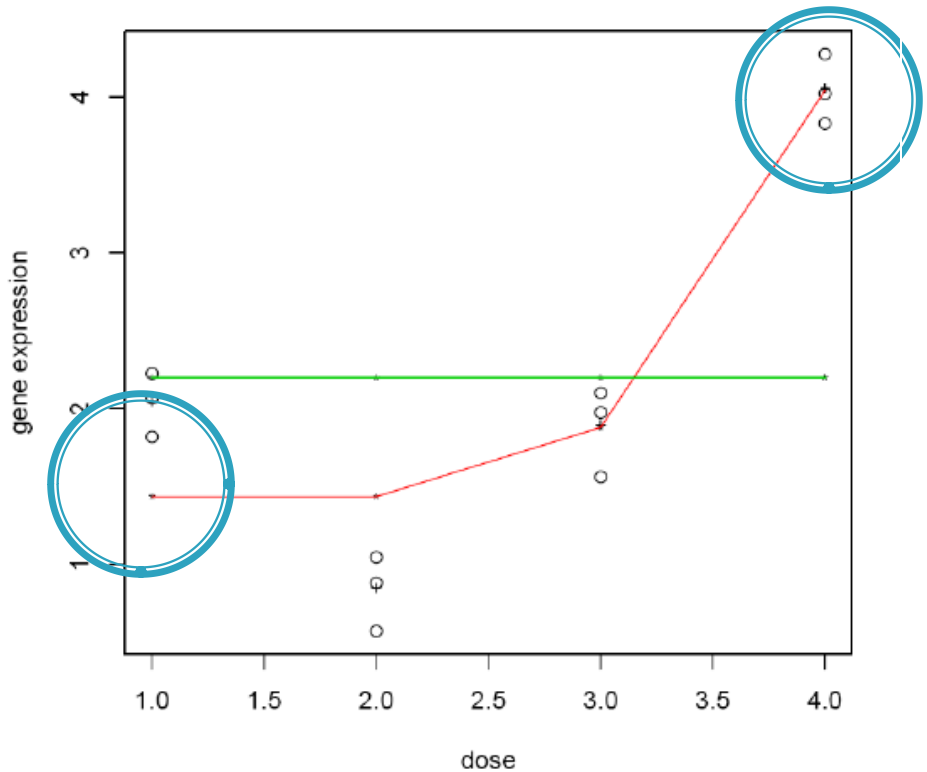
$$W = \frac{\hat{\mu}^*_K - \overline{X}_1}{s_t}$$

(Williams, 1971, 1972)

$$W' = \frac{\hat{\mu}^*_K - \hat{\mu}^*_1}{s_t}$$

(Marcus, 1976)

Test Statistics



$$M = \frac{\hat{\mu}_K^* - \hat{\mu}_1^*}{\sqrt{\sum_j (X_{jl} - \hat{\mu}_j^*)^2 / (N-K)}}$$

(Hu *et al.* 2005)

$$M = \frac{\hat{\mu}_K^* - \hat{\mu}_1^*}{\sqrt{\sum_{jl} (X_{jl} - \hat{\mu}_j^*)^2 / (N-J)}}$$

where $J = \text{unique}(\hat{\mu}^*)$

(*Lin et al.* 2007)

Directional Inference

- Two-sided p-values:

$$p = 2 \times min(p^{Up}, p^{Down})$$

- Determination of direction
  - If $p^{Up} \leq \alpha/2$, reject H_0 and declare H_1^{Up}
  - If $p^{Down} \leq \alpha/2$, reject H_0 and declare H_1^{Down}
  - If p^{Up} *and* $p^{Down} \leq \alpha/2$, reject H_0 and declare a non-monotonic trend

Multiple Testing Issue

- Testing of thousands of genes simultaneously increases the Type I error

	# not rejected	# rejected	Total
# true null	U	V	m_0
# false null	T	S	m_1
Total	W	R	m

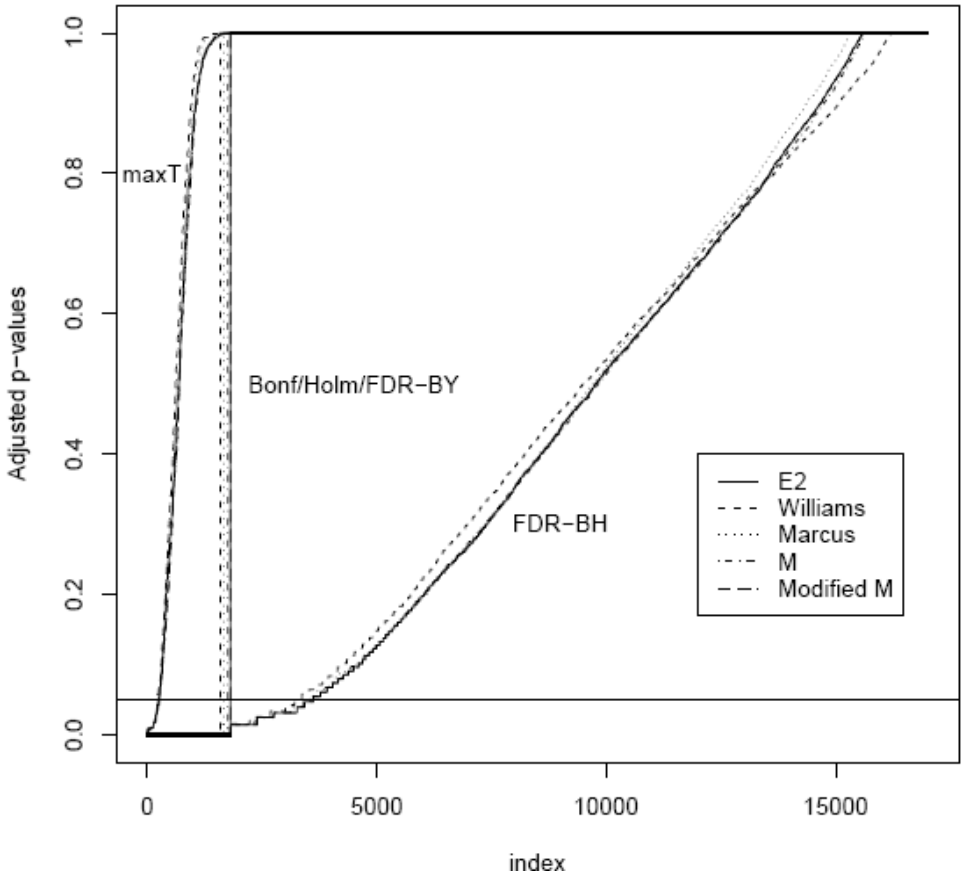
- Family-Wise Error Rate (FWER) = P(V>0)
- False Discovery Rate (FDR, Benjamini and Hochberg 1995)

$$Q = \begin{cases} V/R & R > 0 \\ 0 & R = 0 \end{cases}$$

Testing for Trend: Application

- Case study: data for EGF doses under control

Test	# sign
LRT	3499
W	3209
W’	3533
M	3562
M’	3567

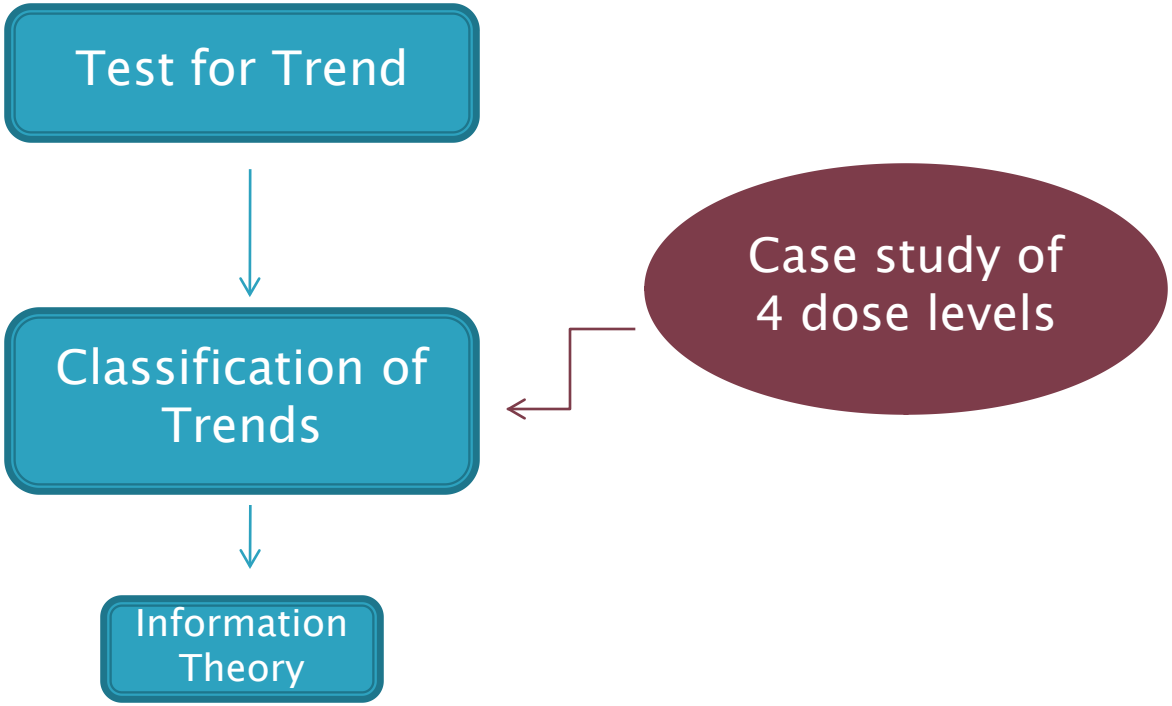


Testing for Trend: Conclusions

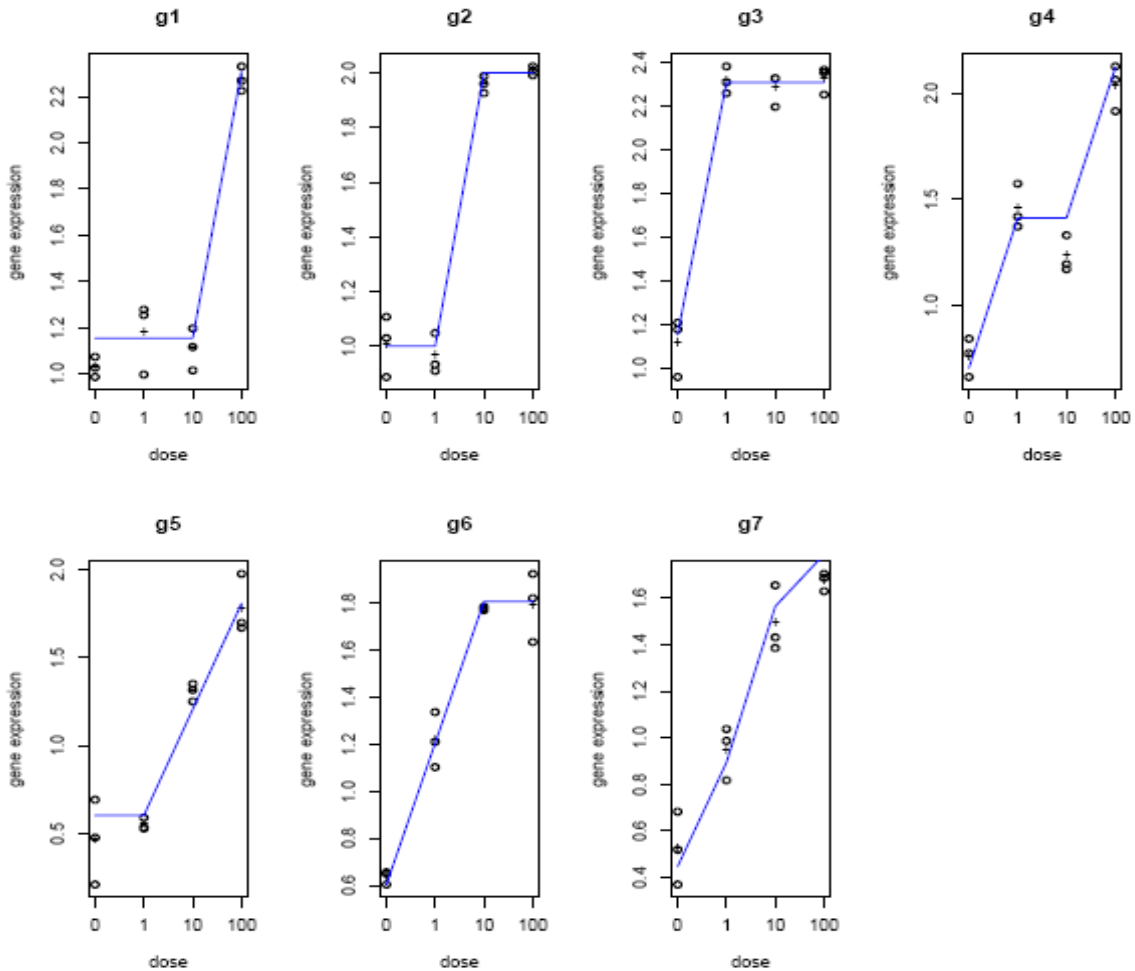
- Results:
  - Five test statistics show a similar number of significant findings
  - Due to the unknown distribution for the test statistics of the M and modified M tests, resampling-based procedures are employed
  - Simulation study has confirmed this finding, in which the LRT, M, and modified M yield slightly higher power
  -

Lin *et al.* (2007)

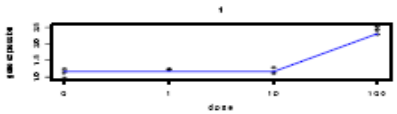
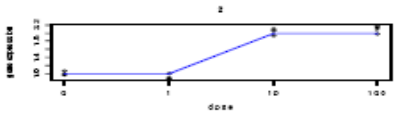
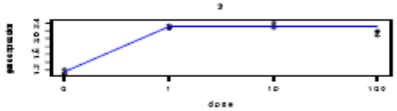
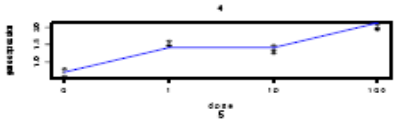
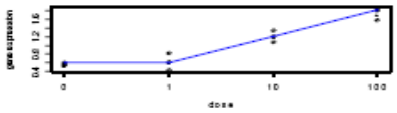
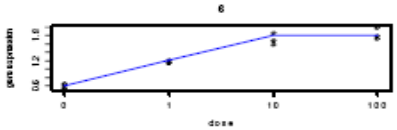
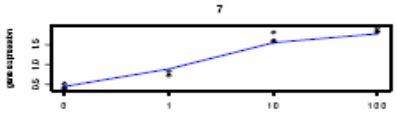
Dose-response Modeling



Classification Trends



Classification of Trends

Shapes	Alternatives	# parameters	MED
	$\mu_1 = \mu_2 = \mu_3 < \mu_4$	2	4
	$\mu_1 = \mu_2 < \mu_3 = \mu_4$	2	3
	$\mu_1 < \mu_2 = \mu_3 = \mu_4$	2	2
	$\mu_1 < \mu_2 = \mu_3 < \mu_4$	3	2
	$\mu_1 = \mu_2 < \mu_3 < \mu_4$	3	3
	$\mu_1 < \mu_2 < \mu_3 = \mu_4$	3	2
	$\mu_1 < \mu_2 < \mu_3 < \mu_4$	4	2

Classification of Trends Using Information Criteria

- Akaike information criterion (Akaike 1973, 1974)

$$AIC = -2\log l(\theta \mid D) + 2M$$

- Bayesian information criterion (Schwarz 1978)

$$BIC = -2\,\mathbf{log}\, l(\theta \mid D) + M\,\mathbf{log}(N)$$

- Order restricted information criterion (Anraku 1999)

$$ORIC = -2\,\mathbf{log}\, l(\theta \mid D) + \sum_{j=1}^{K} iP(j,k,w_j)$$

  - where $P(j, k, w_j)$ denotes the level probability that for given K doses under H_0 the isotonic regression will result in j unique isotonic means

Results of Information Criteria for Model Seletion

	Likelihood	AIC	BIC	ORIC
g_1	344	1528	1648	1348
g_2	25	307	369	221
g_3	14	106	126	86
g_4	343	370	337	253
g_5	885	823	715	655
g_6	178	170	149	120
g_7	1710	195	155	816

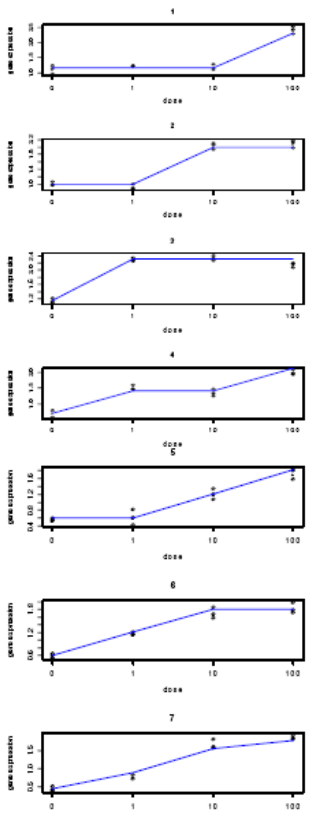
Lin *et al.* (2008)

Multiple Contrast Test

- Multiple contrast test (Mukerjee et al. 1986, 1987)
- Multiple contrast test can be used to test for trend, which shows similar results as the LRT
- Multiple contrasts are a nature link to select the best contrast for the dose-response curve
- Isotonic coefficients can be used to describe the dose-response relationship with corresponding shapes

Multiple Contrast Test

- Seven models:



$\mu1 = \mu2 = \mu3 < \mu4$	$c_1=(c_{11}, c_{21}, c_{31}, c_{41})$	$T_1^{sc} = \frac{\sum_{j=1}^{4} n_j c_{j1} \bar{X}_j}{s\sqrt{\sum_{j=1}^{4} n_j c_j}}$
$\mu1 = \mu2 < \mu3 = \mu4$	$c_2=(c_{12}, c_{22}, c_{32}, c_{42})$	T_2^{sc}
$\mu1 < \mu2 = \mu3 = \mu4$	$c_3=(c_{13}, c_{23}, c_{33}, c_{43})$	T_3^{sc}
$\mu1 < \mu2 = \mu3 < \mu4$	$c_4=(c_{14}, c_{24}, c_{34}, c_{44})$	T_4^{sc}
$\mu1 = \mu2 < \mu3 < \mu4$	$c_5=(c_{15}, c_{25}, c_{35}, c_{45})$	T_5^{sc}
$\mu1 < \mu2 < \mu3 = \mu4$	$c_6=(c_{16}, c_{26}, c_{36}, c_{46})$	T_6^{sc}
$\mu1 < \mu2 < \mu3 < \mu4$	$c_7=(c_{17}, c_{27}, c_{37}, c_{47})$	T_7^{sc}

Multiple Contrast Test

- The MCT statistic builds the maximum over seven of such single contrasts

$$T^{MC} = \max\{T_1^{sc}, T_2^{sc}, \ldots, T_7^{sc}\}$$

- Inference of MCT statistics can be made based on the *q*–multivariate T distribution
- Multiplicity adjustment: FWER for each gene

Multiple Contrast Test

- Take $\mu_1 < \mu_2 = \mu_3 < \mu_4$ for example (Abelson and Tukey 1963)

Inequality	Corner Pattern	Corner Vector	SSD
$\mu_1 < \mu_2$	$\mu_1 < \mu_2 = \mu_3 = \mu_4$	(1, 0, 0, 1)	3/4
$\mu_3 < \mu_4$	$\mu_1 = \mu_2 = \mu_3 < \mu_4$	(0, 0, 0, 1)	3/4

  - where $SSD = \sum_j (\mu_j - \bar{\mu})^2$

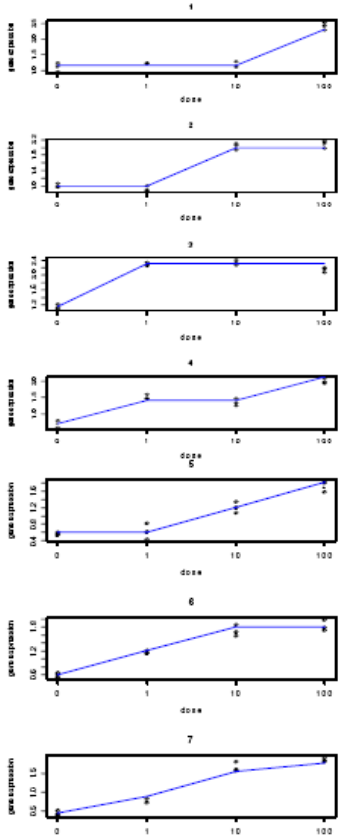
- To obtain the contrast coefficients by solving

$$(s)\ c_1 + c_2 + c_3 + c_4 = 1$$
$$(a) \quad c_2 + c_3 + c_4 = \sqrt{3/4}$$
$$(b) \quad c_4 = \sqrt{3/4}$$

$$\Rightarrow \quad c_1 = -0.866, \quad c_2 = c_3 = 0, \quad c_4 = 0.866$$

Multiple Contrast Test

- Seven models:



$\mu1 = \mu2 = \mu3 < \mu4$	c_1=(−0.2887, −0.2887, −0.2887, 0.866)
$\mu1 = \mu2 < \mu3 = \mu4$	c_2=(−0.5, −0.5, 0.5, 0.5)
$\mu1 < \mu2 = \mu3 = \mu4$	c_3=(−0.866, 0.2887, 0.2887, 0.2887)
$\mu1 < \mu2 = \mu3 < \mu4$	c_4=(−0.866, 0, 0, 0.866)
$\mu1 = \mu2 < \mu3 < \mu4$	c_5=(−0.5, −0.5, −0.134, 0.866)
$\mu1 < \mu2 < \mu3 = \mu4$	c_6=(−0.866, −0.134, 0.5, 0.5)
$\mu1 < \mu2 < \mu3 < \mu4$	c_7=(−0.886, −0.134, 0.134, 0.866)

Classification of Trends: Application

3499	AIC	BIC	ORIC
g_1	1528	1648	1348
g_2	307	369	221
g_3	106	126	86
g_4	370	337	253
g_5	823	715	655
g_6	170	149	120
g_7	195	155	816

3277	MCT
g_1	688
g_2	205
g_3	1515
g_4	60
g_5	463
g_6	93
g_7	253

Lin *et al.* (2010)

Classification of Trends: Conclusions

- The AIC and BIC tend to classify genes with simpler models
- The ORIC penalizes less on complex model (g_7)
- The MCT favors simpler models
- Simulation study is needed to compare the performance of these different approaches

Concluding Remarks

- Two stage analysis: to ensure the control of the FDR by the LRT in the first step and information criteria for model selection

- Unified analysis: MCT integrates two steps

References

Lin D., Shkedy Z., Yekutieli D., Dhammika A., Bijnens, L., Modeling Dose-response Microarray Data in Early Drug Development Experiments Using R, Springer (to appear in 2010).

- Parametric modeling
- Modeling averaging of parameters of interest from the models
- Bayesian approach for order restricted inference
- MCT ratio test
- FDR-adjusted CIs for ratio parameters
- Gene set analysis

Thank you!