Dose-Response Modeling of Gene Expression Data in Microarray Experiments

Setia Pramana

Interuniversity Institute for Biostatistics and Statistical Bioinformatics,
Universiteit Hasselt, Diepenbeek, Belgium

Research Team

- I-Biostat
  - Setia Pramana
  - Dan Lin
  - Ziv Shkedy
  - Philippe Haldermans
- J&J Pharmaceutical Research and Development
  - An De Bondt
  - Hinrich Göhlmann
  - Willem Talloen
  - Luc Bijnens
  - Jose Pinheiro
  - Tobias Verbeke

Outline

- Introduction to Dose-response Studies
- Testing for Monotonic Trend
- Model Based
- Model Averaging
- Application
- Concluding Remarks

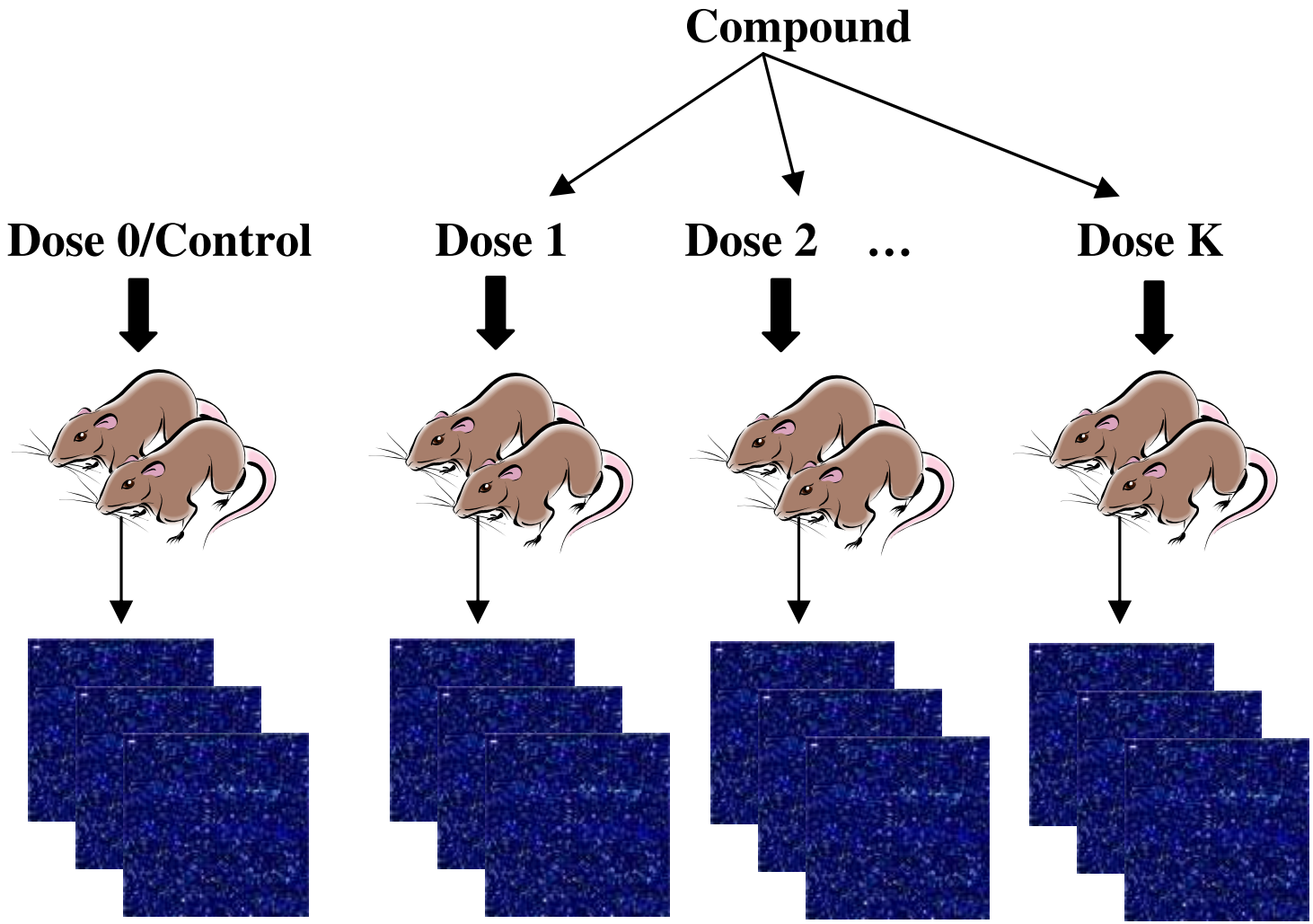
Dose-response (DR) studies: The fundamental study in drug discovery

- Good drugs: Strong effects on a specific biological pathways, minimal effects on all other pathways.
- Too high dose can be dangerous, too low dose decreases the chance of it showing effectiveness.
- DR studies:
  - Investigate the dependence of the response on doses: how the drug works? Has it the desired properties?
  - What is the shape of the relationship?
  - Discover a dose or a range of dose that are both efficacious and safe. Target dose: minimum effective dose (MED), maximally tolerated dose (MTD) or half maximal effective dose (ED_{50}).

Dose-response in Microarray Experiments

- Monitoring of gene expression with respect to increasing dose of a compound.
- To identify a subset of genes with overall dose related trend.
- To investigate the mechanism of action of potential drug in the entire genome.
- To compare between compounds using the gene expression information.

Dose-response in Microarray: The study

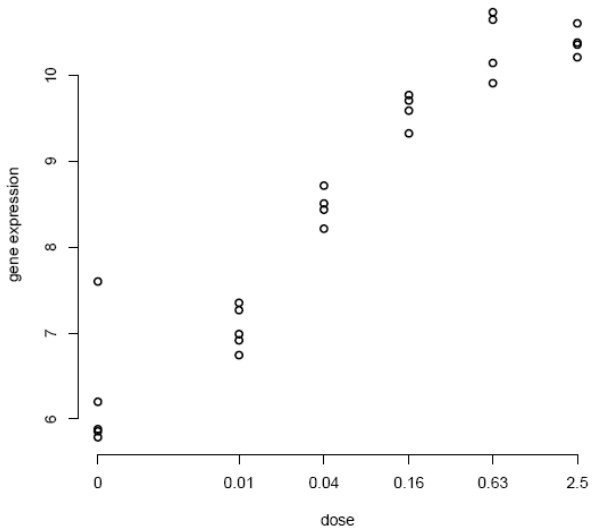


Dose-response in Microarray: Data Structure

$$X = \begin{array}{l} \textit{Gene 1} \\ \textit{Gene 2} \\ \cdot \\ \cdot \\ \cdot \\ \textit{Gene m} \end{array} \begin{bmatrix} x_{011}..x_{01n_0} & x_{111}..x_{11n_1} & \cdots & x_{k11}..x_{k1n_k} \\ x_{011}..x_{02n_0} & x_{121}..x_{12n_1} & \cdots & x_{k21}..x_{k2n_k} \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ x_{0m1}..x_{0mn_0} & x_{1m1}..x_{1mn_1} & \cdots & x_{km1}..x_{kmn_k} \end{bmatrix}$$

d_0 d_1 ….. d_k

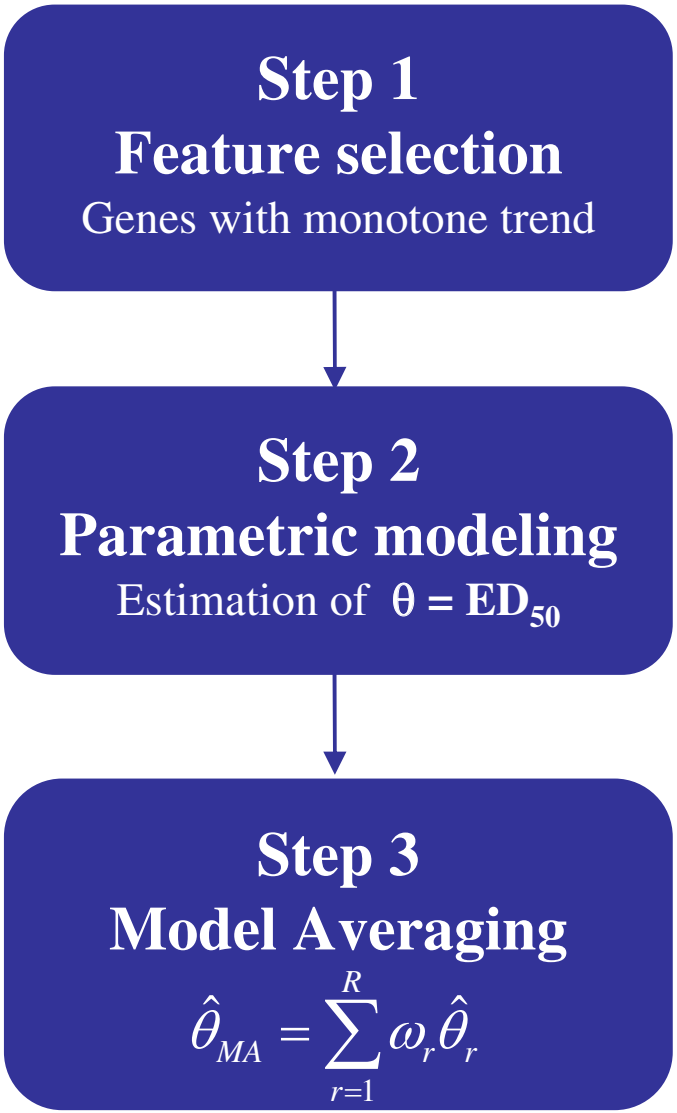
Dose levels



Dose-response in Microarray: Modeling

- No prior info about the dose-response shape, but it's assumed to be monotone.
- Monotone assumption is based on in general, increasing the dose of a harmful agent results a proportional increase in the incidence of an adverse effect and the severity of the effect.
- Genes have different shapes.

Modeling Framework



Step 1: Feature Selection: Testing for Monotonic Trend

- Gene specific test:

$$H_0: \quad \mu(d_0) = \mu(d_1) = \cdots = \mu(d_K)$$

$$H_1^{Up}: \quad \mu(d_0) \leq \mu(d_1) \leq \cdots \leq \mu(d_K)$$

or

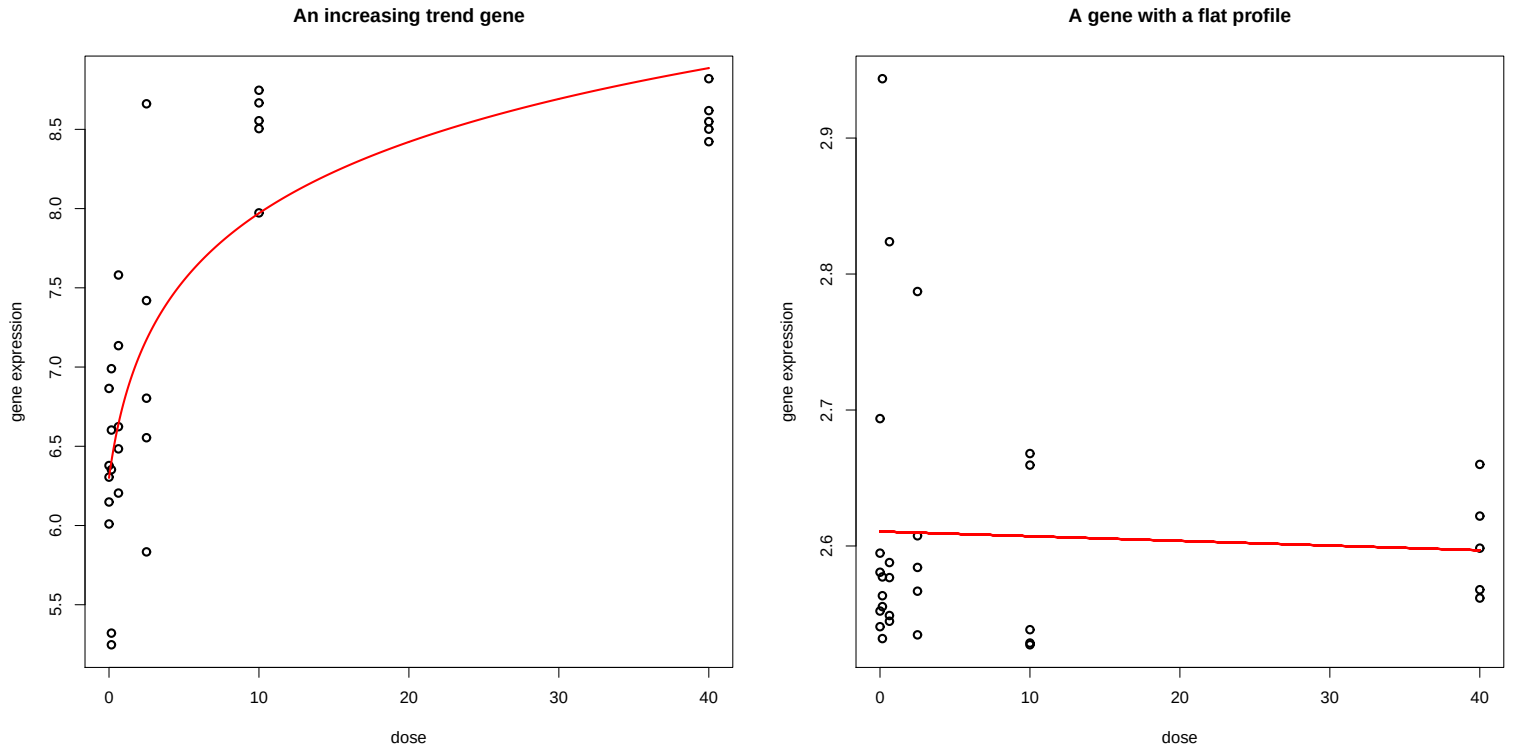
$$H_1^{Down}: \quad \mu(d_0) \geq \mu(d_1) \geq \cdots \geq \mu(d_K)$$

with at least one inequality.

- Test statistics: Likelihood Ratio Test ($\bar{E}_{01}^2$).

(Lin et al., 2007)

Step 2: Dose-response Modeling



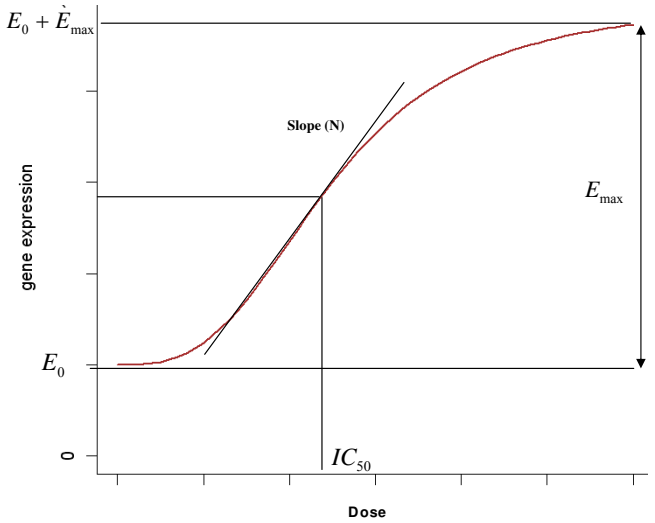
- For each differentially expressed gene:

$$Y_{ij} = f(d_i, \theta) + \varepsilon_{ij},\ i = 1, 2, \ldots, K,\ j = 1, 2, \ldots, n_i,$$

where $f(d_i, \theta)$: the dose-response model (e.g., E_{max}, Logistic).

Dose-response Modeling: Target Dose (ED_{50})

- From the DR model the ED_{50} is estimated.
- The ED_{50}: dose which induces a response halfway between the baseline and maximum.



- ED_{50} reflects the potency of the tested drug or compound.
- The ED_{50} is restricted to lie within the interval (d_1, d_k] to avoid problems arising from extrapolating beyond the dose range under investigation.

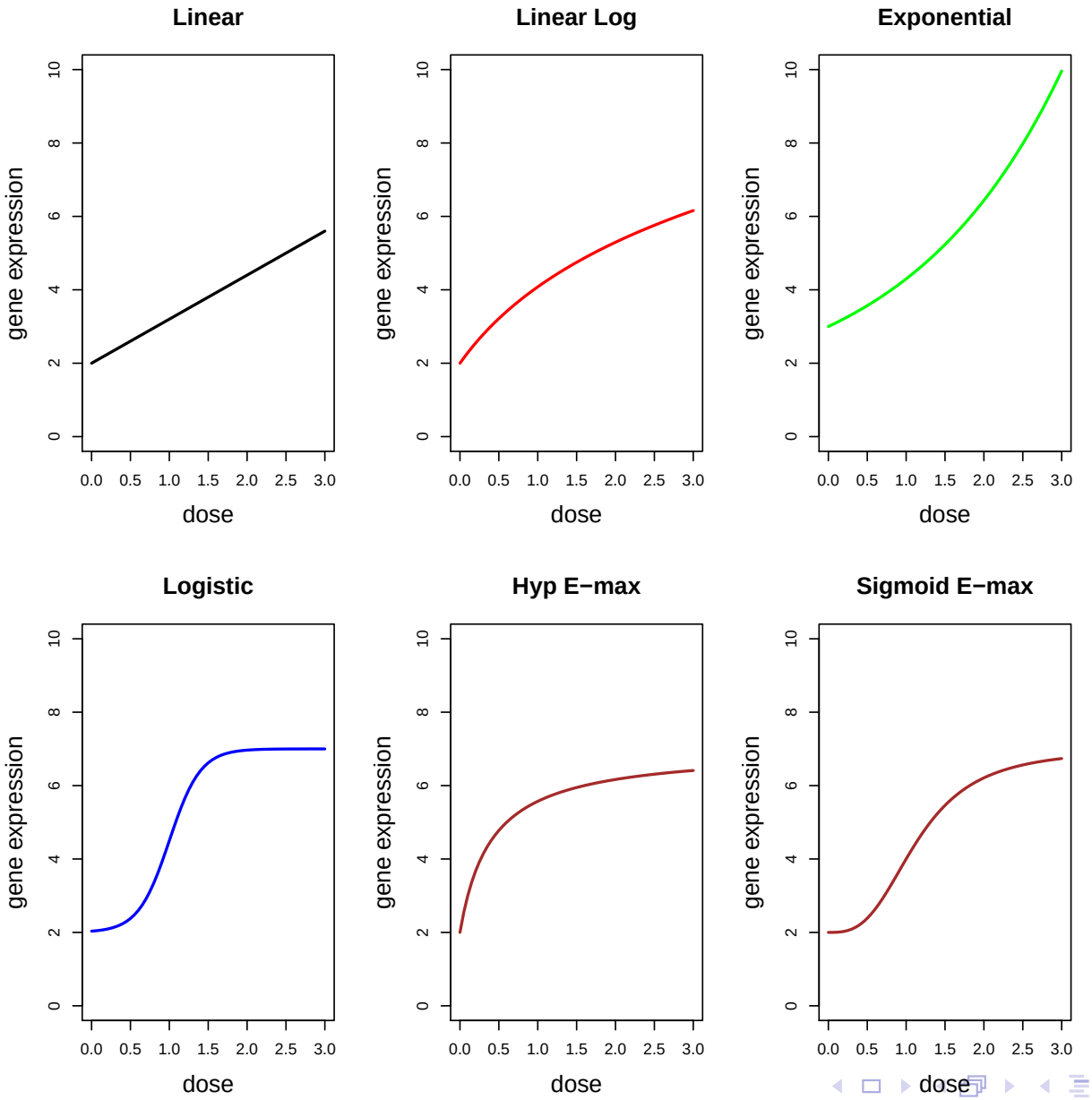
Dose-response Modeling: Pros and Cons

- Assume a functional relationship between the response and the dose according to a pre-specified parametric model.
- The dose is taken as a quantitative factor.
- Provides flexibility in investigating the effect of doses not used in the actual study.
- Its result validity depends on the correct choice of the DR model, which is a priori unknown.
- Multiple models describe the data equivalently, but the estimates target dose are different.

Step 3: Model Averaging

- Account for model uncertainty.
- All fits are taken into consideration.
- Combines results from different models.
- Poor fits receive small weights.

DR Model Candidates



Model Averaging

- Let θ be a quantity in which we are interested in and we can estimate θ from R models, the model averaged (MA) θ is defined as:

$$\hat{\theta}_{MA} = \sum_{r}^{R} \omega_r \times \hat{\theta}_r,$$

  where $\hat{\theta}_r$ is the estimate of θ from model r and ω_r the weights that sum to one assigned to model r.

- Given the fits of R models, we can estimate the MA dose-response curve as:

$$\hat{f}_{MA}(d) = \sum_{r}^{R} \omega_r \times \hat{f}(\theta, d)_r.$$

Model Averaging Weights

- Information Criterion

$$\omega_r = \frac{exp(-\Delta I_r/2)}{\sum exp(-\Delta I_r/2)}$$

- $\Delta I_r = I_r - I_{min}$, where I_{min} is the smallest Information Criterion value, $I_r = AIC, BIC$.
- Bootstrapping (Buckland et al. 1997).
- We implemented AIC

Model Averaged ED_{50}

- The model-averaged ED_{50} is defined as:

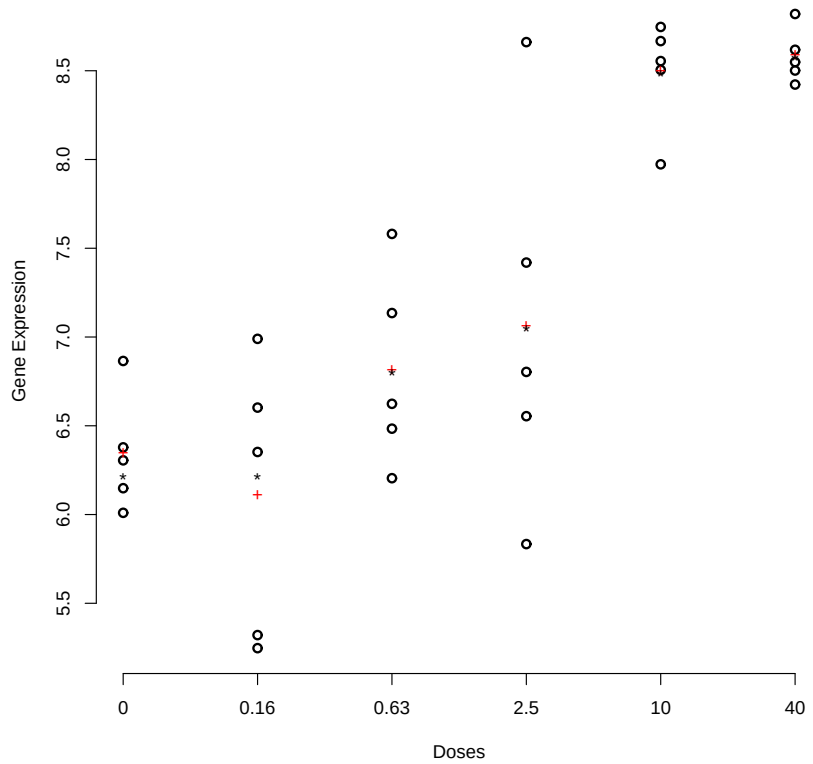
$$\widehat{\overline{ED}}_{50} = \sum_{r=1}^{R} \omega_r \widehat{ED}_{50,r}\,, \tag{1}$$

  where $\widehat{ED_{50,r}}$ is the estimate of ED_{50} of model r, and ω_r is the akaike's weight of model r.

- Since the distribution of the $\widehat{\overline{ED}}_{50}$ is unknown, the estimator for variance of $\widehat{\overline{ED}}_{50}$ is obtained using bootstrap method.

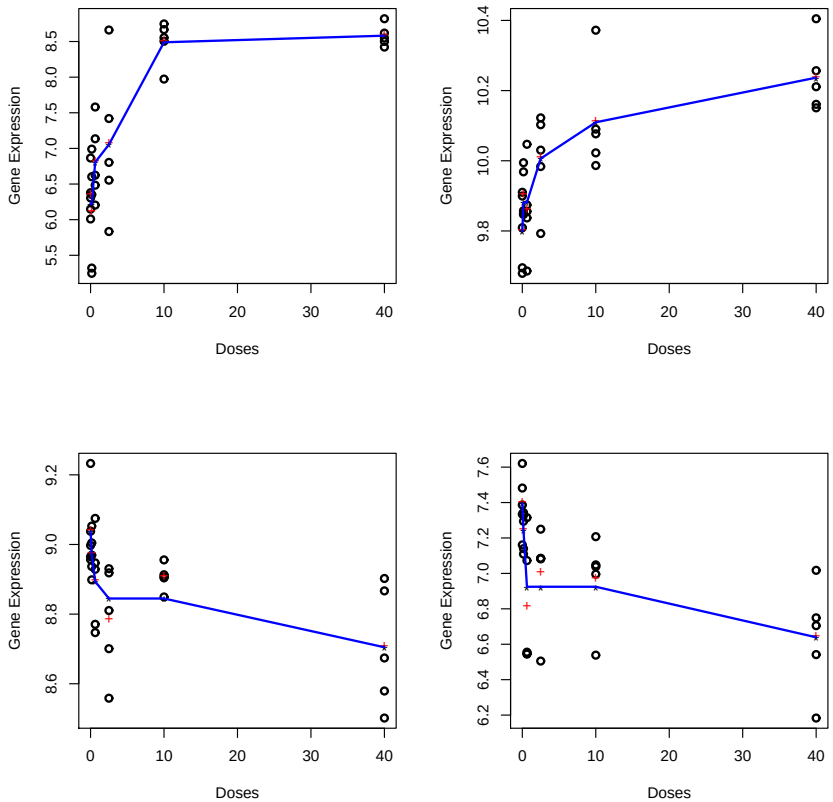
Antipsychotic Study

- Case study: a study focuses on antipsychotic compounds.
- 6 dose levels with 4-5 samples at each dose level.
- Each array consists of 11,565 genes.



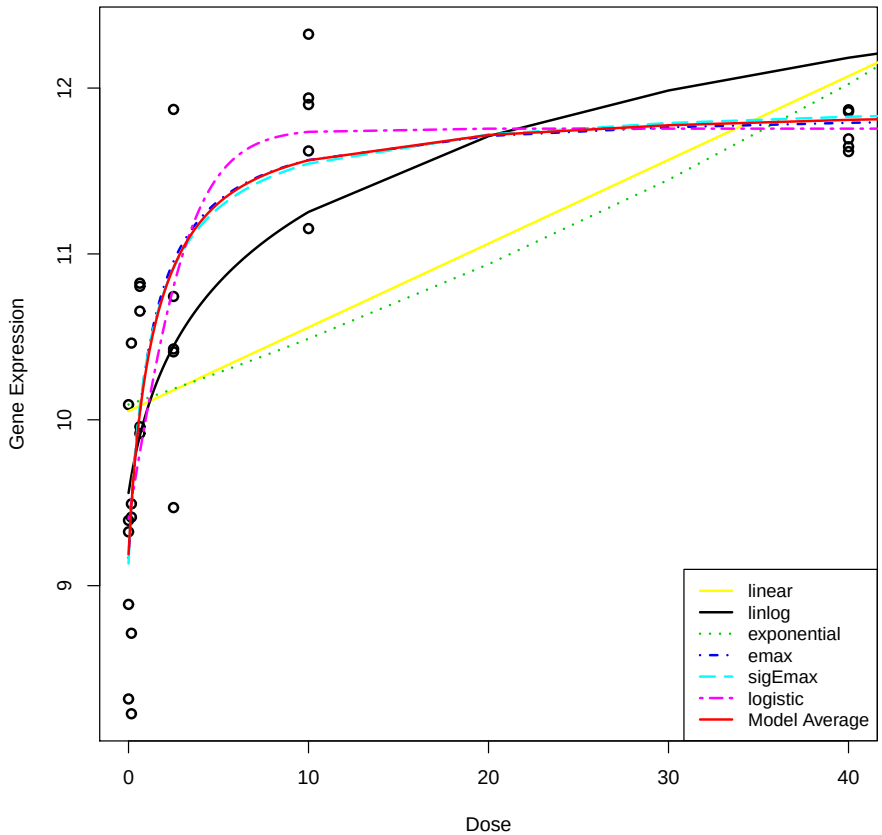
Results: Feature Selection

- 72 genes have a significant monotonic trend (FDR=0.05).
- Data and isotonic trend of four significant genes:



Results: Model-based

Data and fitted value for the one of the genes



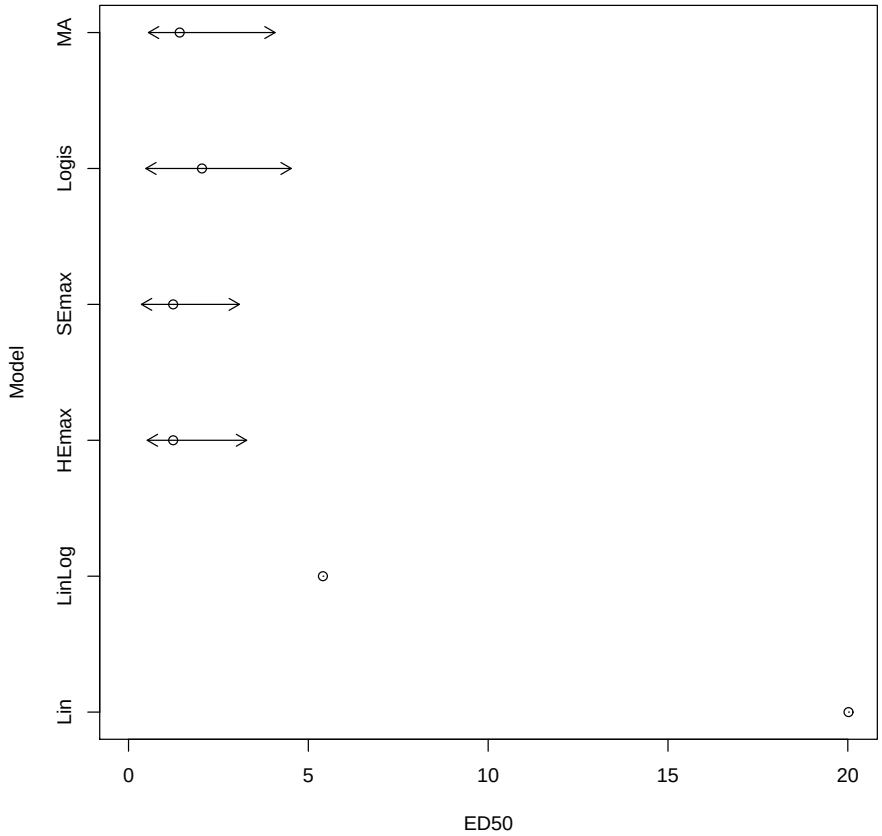
Results: Model Averaging

ED_{50}, AIC, and AIC weight for one of the genes

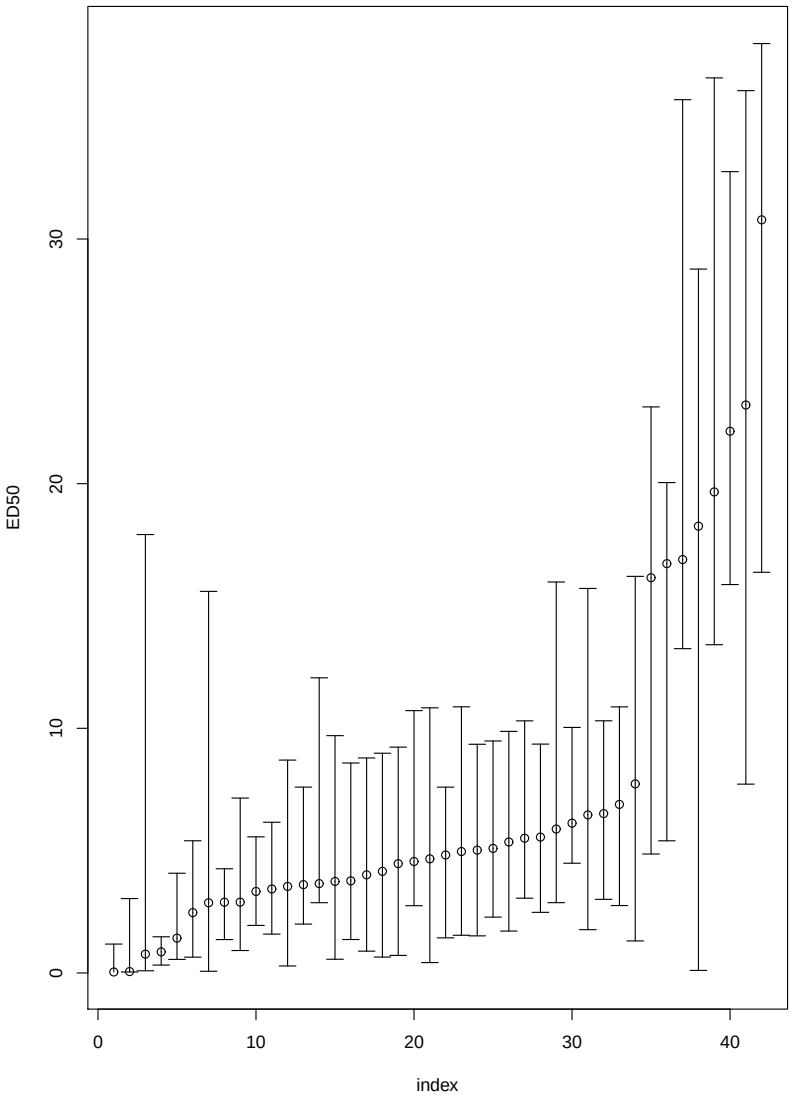
Model	ED_{50}	AIC	AIC weight
Linear	20.000	86.37	$<$0.0001
Linear log-dose	5.405	69.51	0.029
Exponential	22.502	89.78	$<$0.0001
4P Logistic	2.042	67.53	0.077
Hyperbolic E_{max}	1.241	63.30	0.640
Sigmoidal E_{max}	1.241	65.15	0.254
Model Average ED_{50}	1.423		

Results: Model Averaging

ED_{50} and its confidence interval for each model



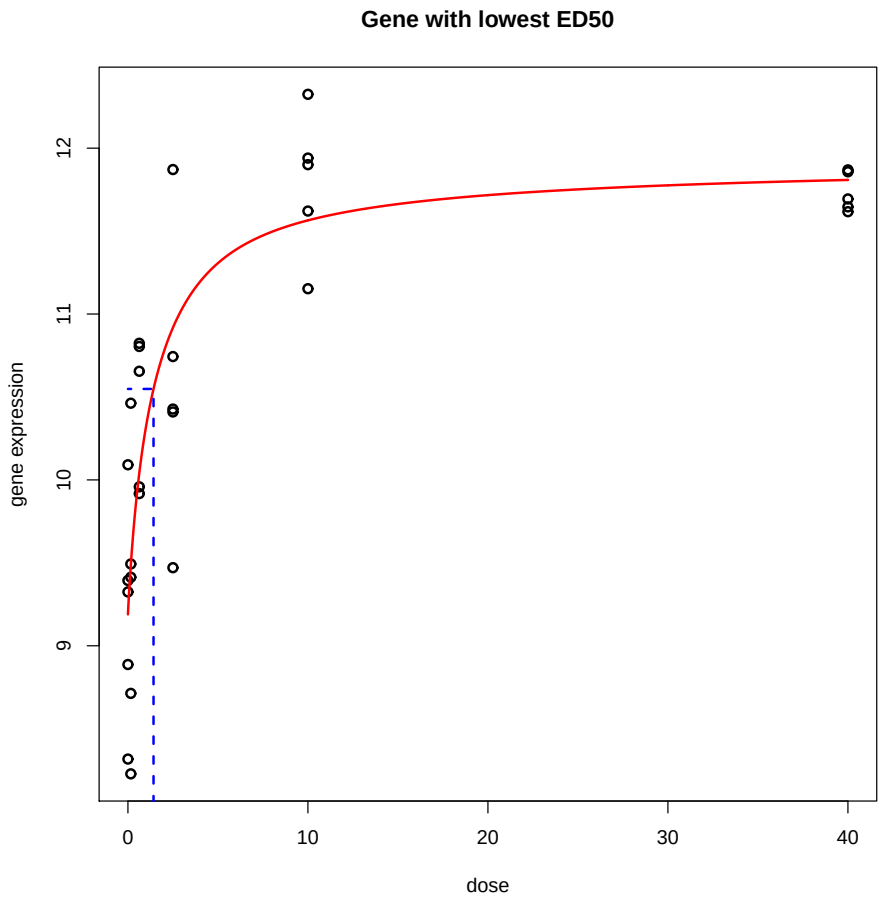
Results: Gene Ranking Based on MA ED_{50}

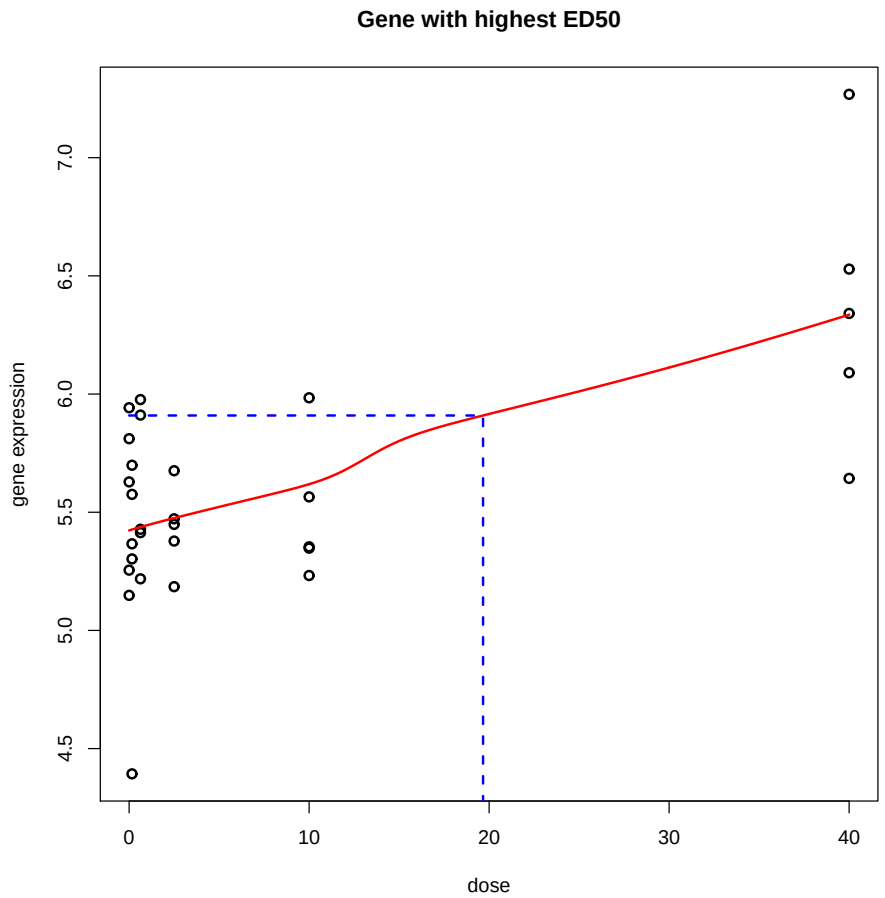


- Genes with a smaller ED_{50} react faster to the compound (genes need less dose to be expressed.).
- Genes with high ED_{50} are less interesting.

Results: Gene Ranking

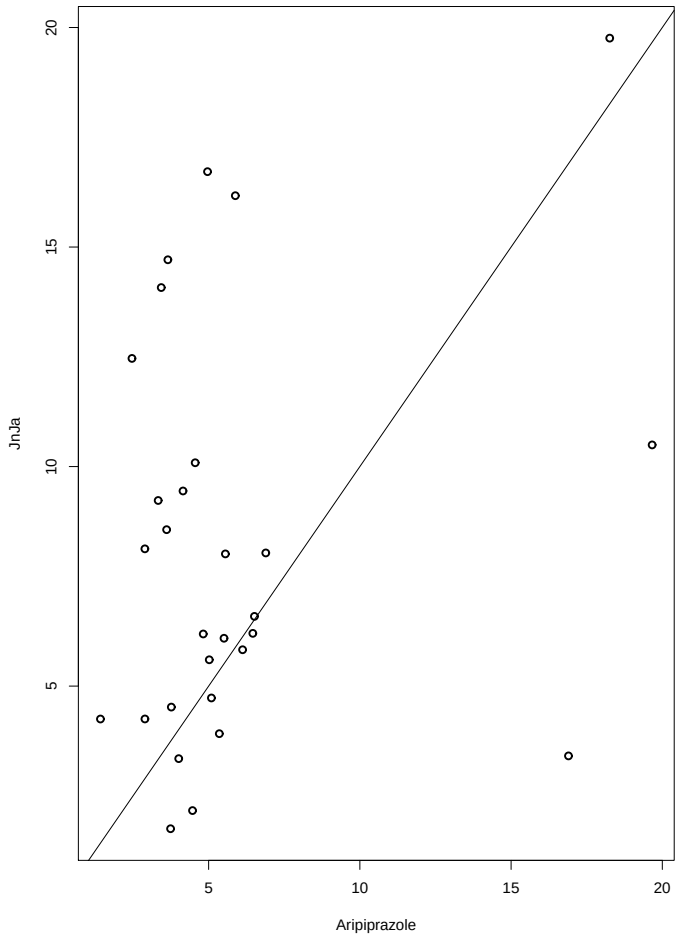
Genes profile with the smallest and highest ED_{50}.





Results: Comparison with Other Compounds

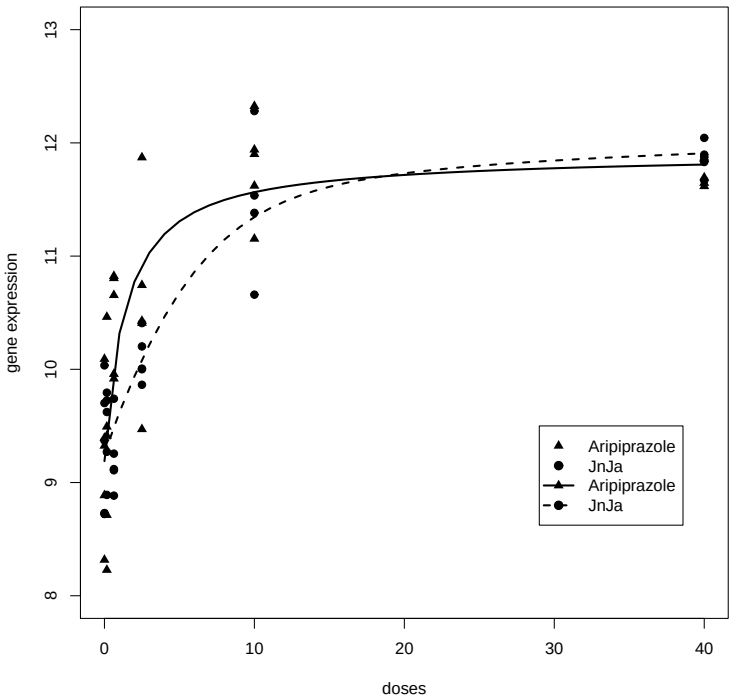
Plot of MA ED_{50} compound JnJa vs. Aripi



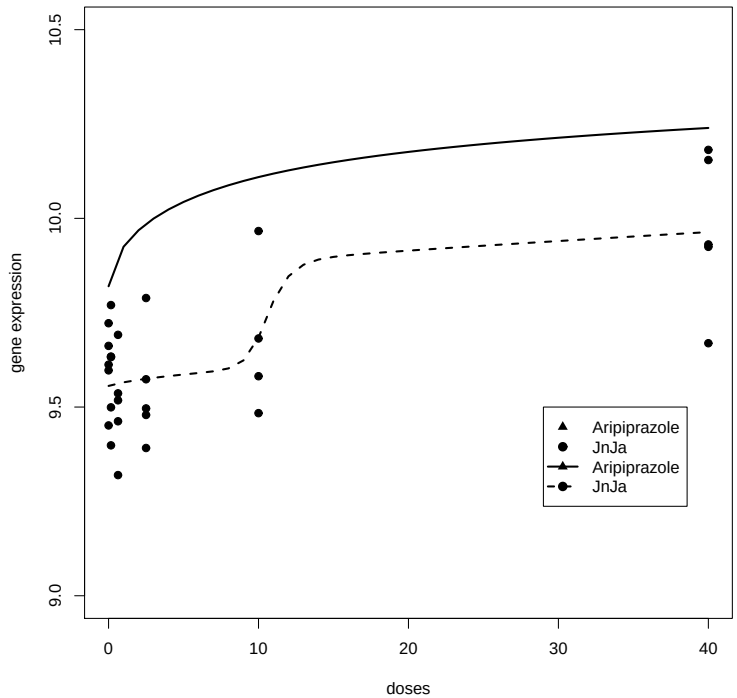
- Genes express differently over the two compounds.
- Most of the genes react slower to the compound JnJa than Aripiprazole.

Results: Comparison with other compounds

Aripiprazole: ED_{50}= 1.143
JnJa: ED_{50}= 4.25



Aripiprazole: ED_{50}=4.96
JnJa: ED_{50}= 16.72



Concluding Remarks

- In the DR modeling in microarray settings, fitting directly the proposed models to all genes (which can be tens thousands) can create problems, such as complexity and time consumption.
- There is no single model fits all genes.
- We propose a three steps approach:
  - Select the genes with a monotone trend using the E^2.
  - Fit the selected genes with the candidate models to get a target dose.
  - Average the target dose from the candidate models.
- These procedures combine the advantage of testing for monotone trend, model-based and model averaging.

Concluding Remarks

- The MA(ED_{50}) can be used to rank the genes and compare a specific gene over the tested compounds.
- Software:
  - `IsoGene` (CRAN) and `IsoGeneGUI` (bioconductor) R packages for testing for monotonic trend,
    `http://www.ibiostat.be/software/IsoGeneGUI/index.html`.
  - `DoseFinding` R package for non-linear DR modeling and model averaging.
- More details:
  Dan Lin, Ziv Shkedy, Daniel Yekutieli, Dhammika Amaratunga, and Luc Bijnens (Editors). (2011). *Modeling Dose-response Microarray Data in Early Drug Development Experiments Using* R. Springer.

Thank you for your attention....

" All things are poison and nothing is without poison; only the dose makes that a thing is no poison." (Paracelsus)