
A random coefficients model for regional co-expression associated with DNA copy number aberrations

Wessel van Wieringen
wvanwie@few.vu.nl

Dept. of Epidemiology and Biostatistics, VUmc
& Dept. of Mathematics, VU University
Amsterdam, The Netherlands

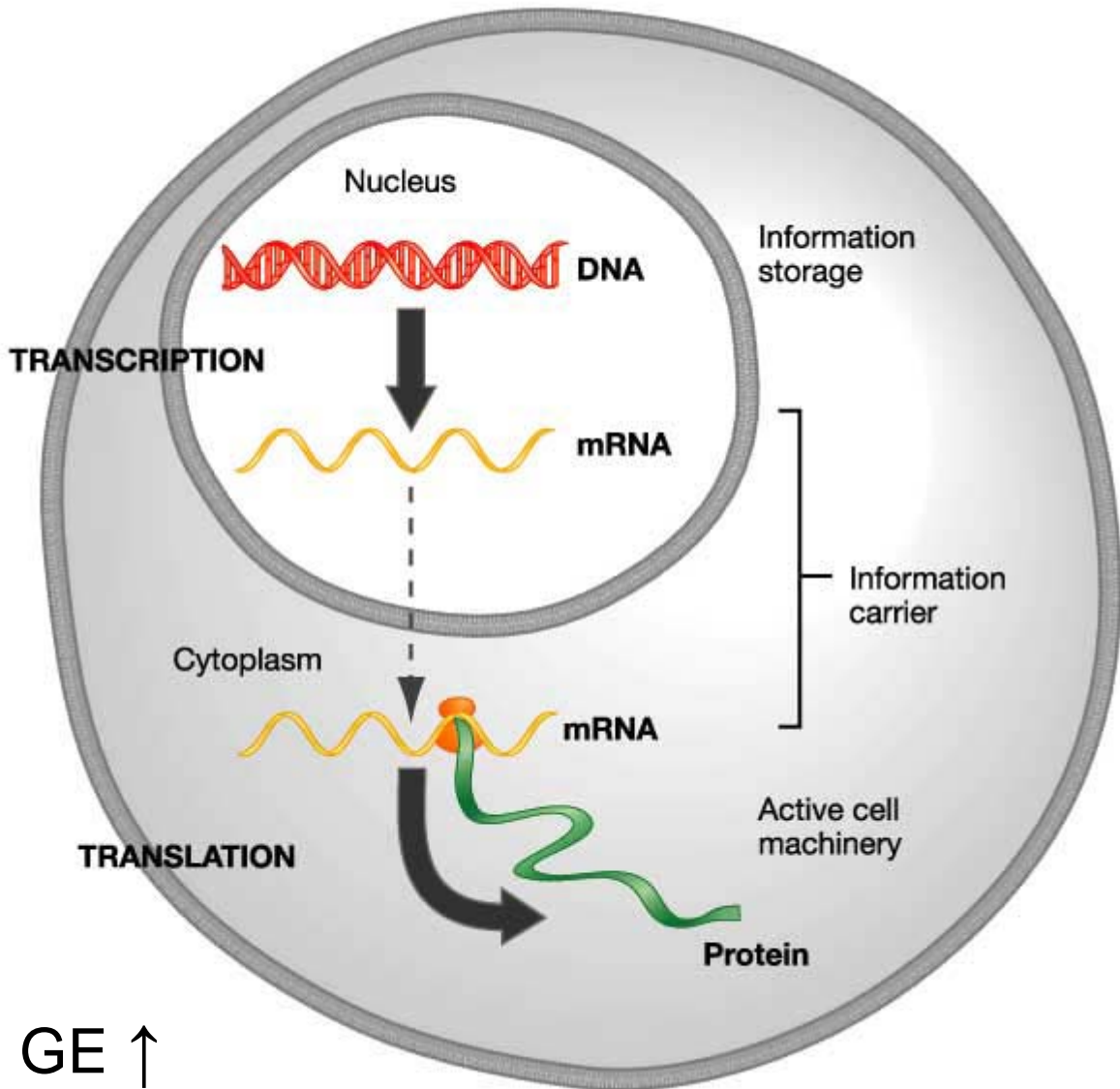
Integrative analysis

aCGH measures
DNA copy number
changes

Expression arrays
measure mRNA
levels

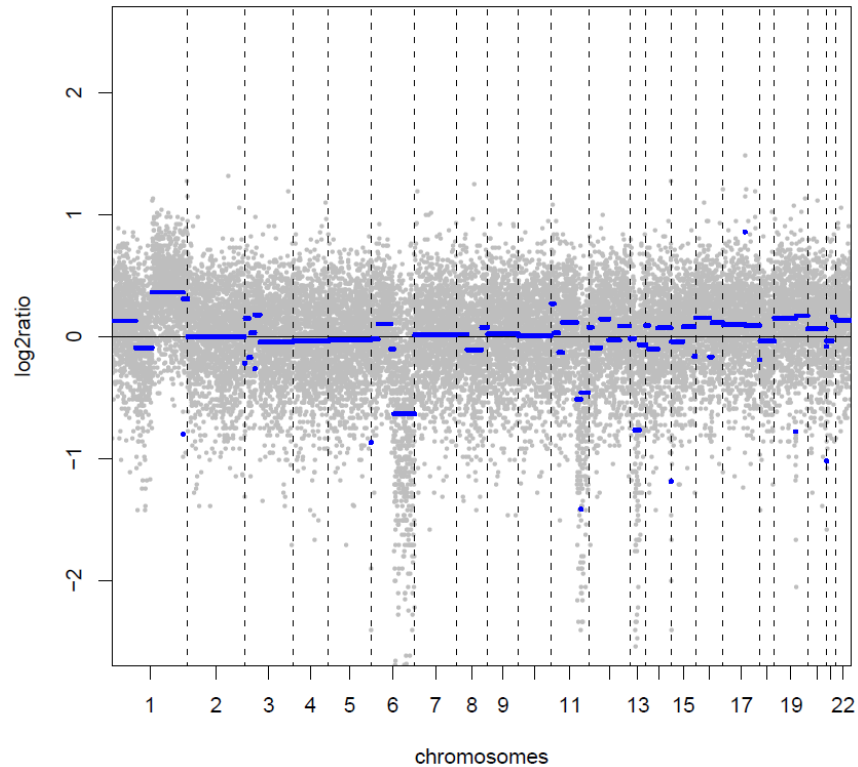
Principle

CN ↓ GE ↓ or CN ↑ GE ↑

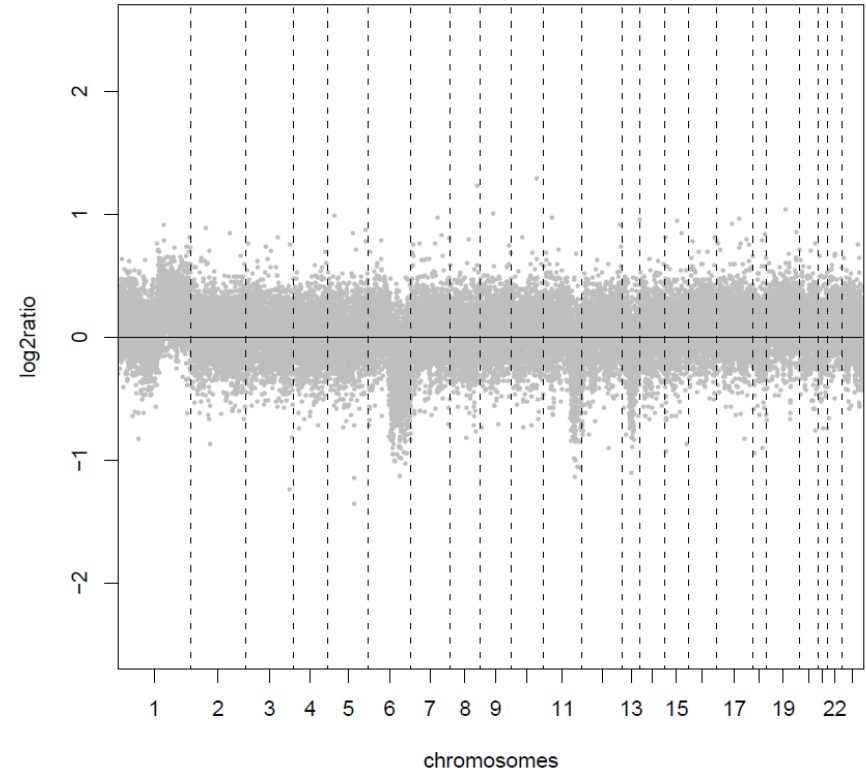


Integrative analysis

Copy number (CN)



Gene expression (GE)

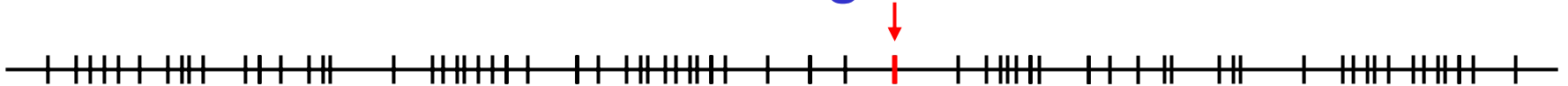


Genomic order = spatial structure

Units of integration

DNA

gene



DNA

region



- set of contiguous genes with the same CN signature
- determined by data of multiple samples

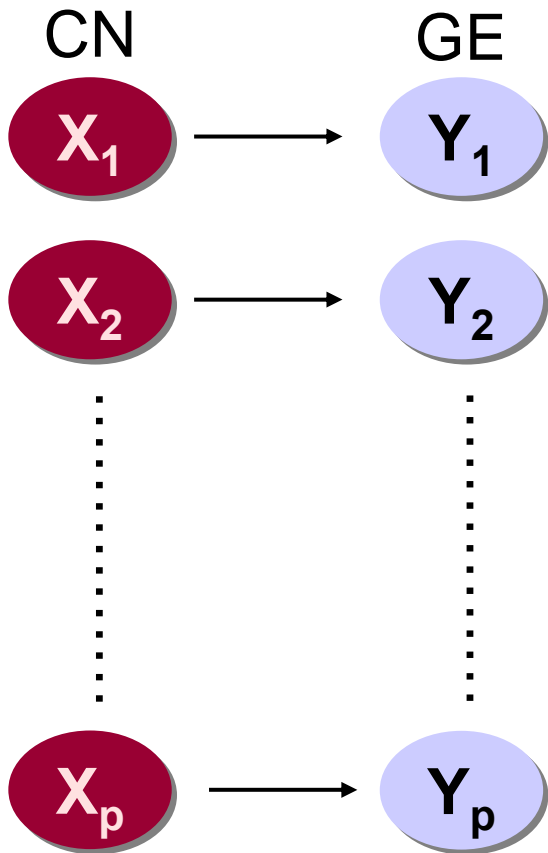
DNA

pathway

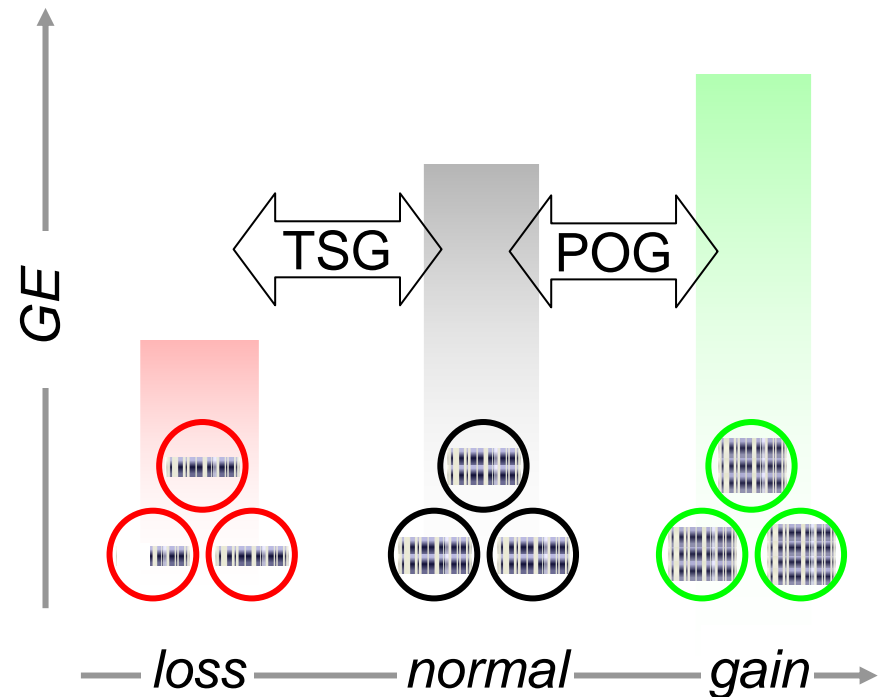


- set of genes from all over the genome
- determined by literature and presence

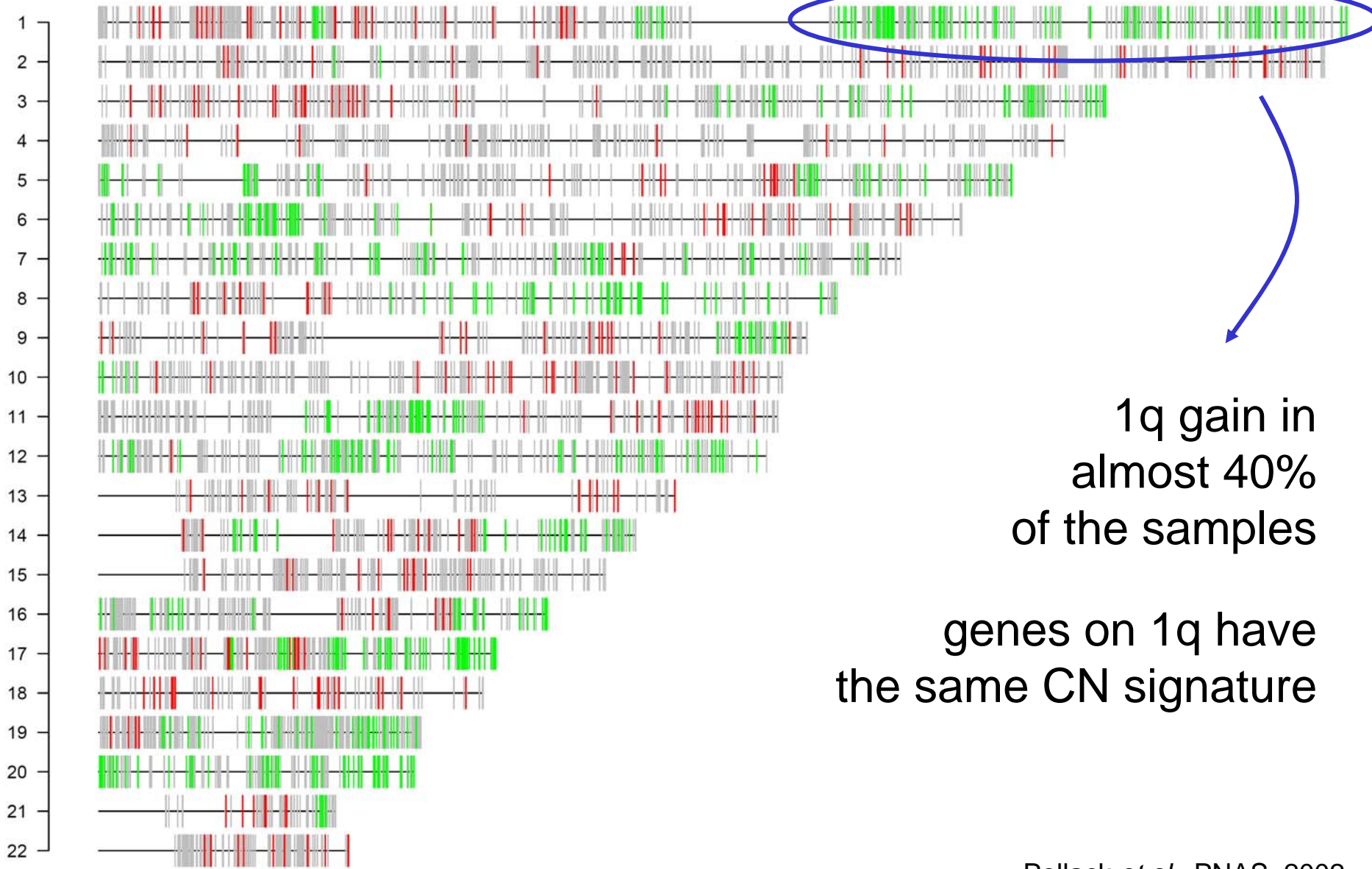
Univariate analysis



Tumorsuppressor (TSG)
and *onco genes (POG)*

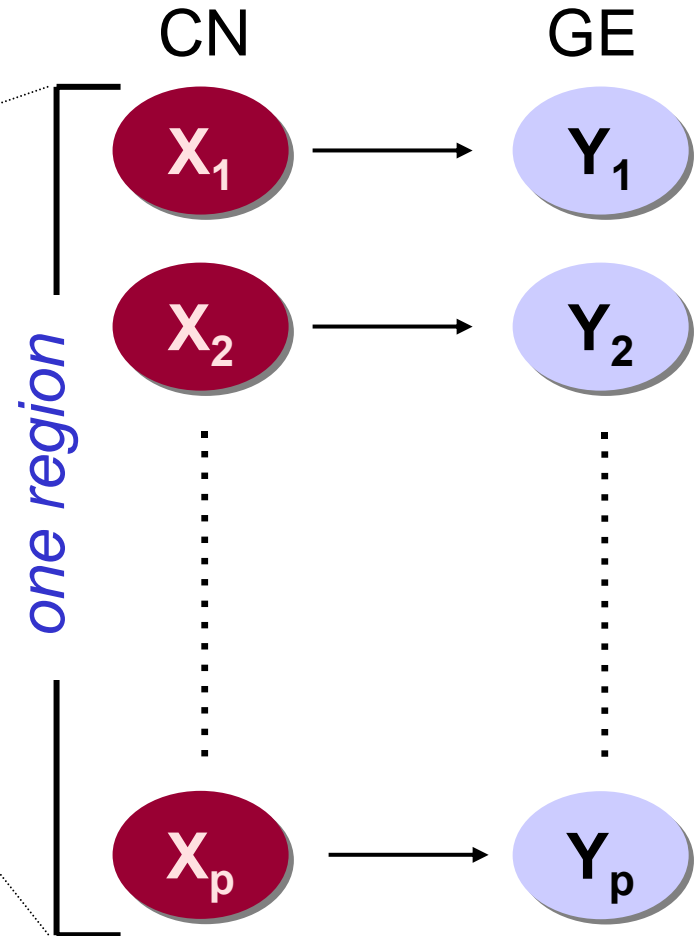
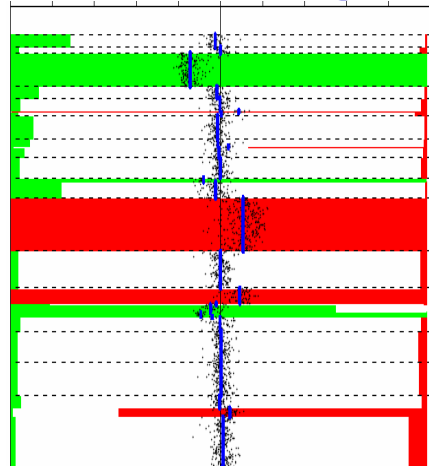


Univariate analysis



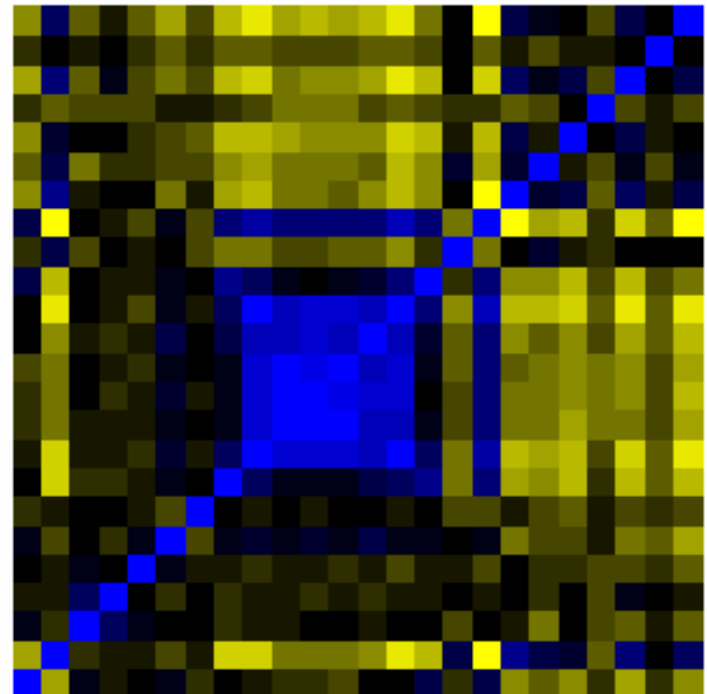
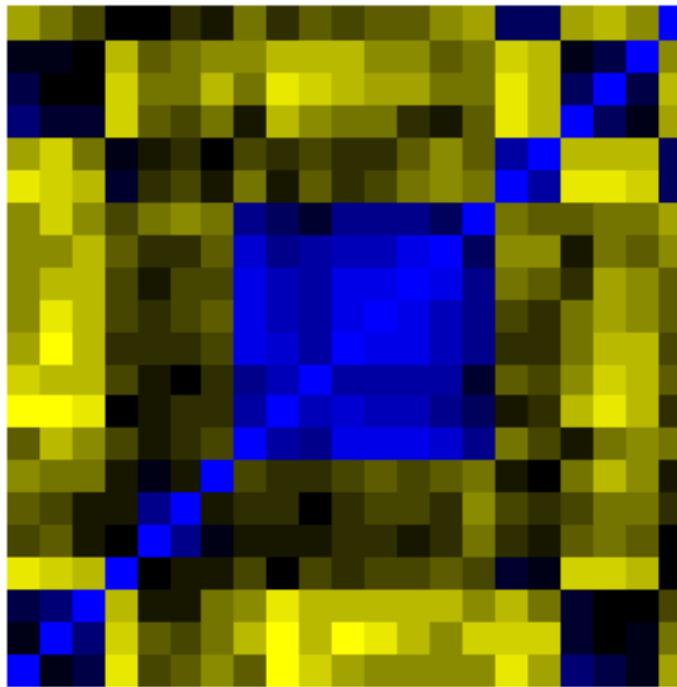
Univariate analysis

Look at a *region*



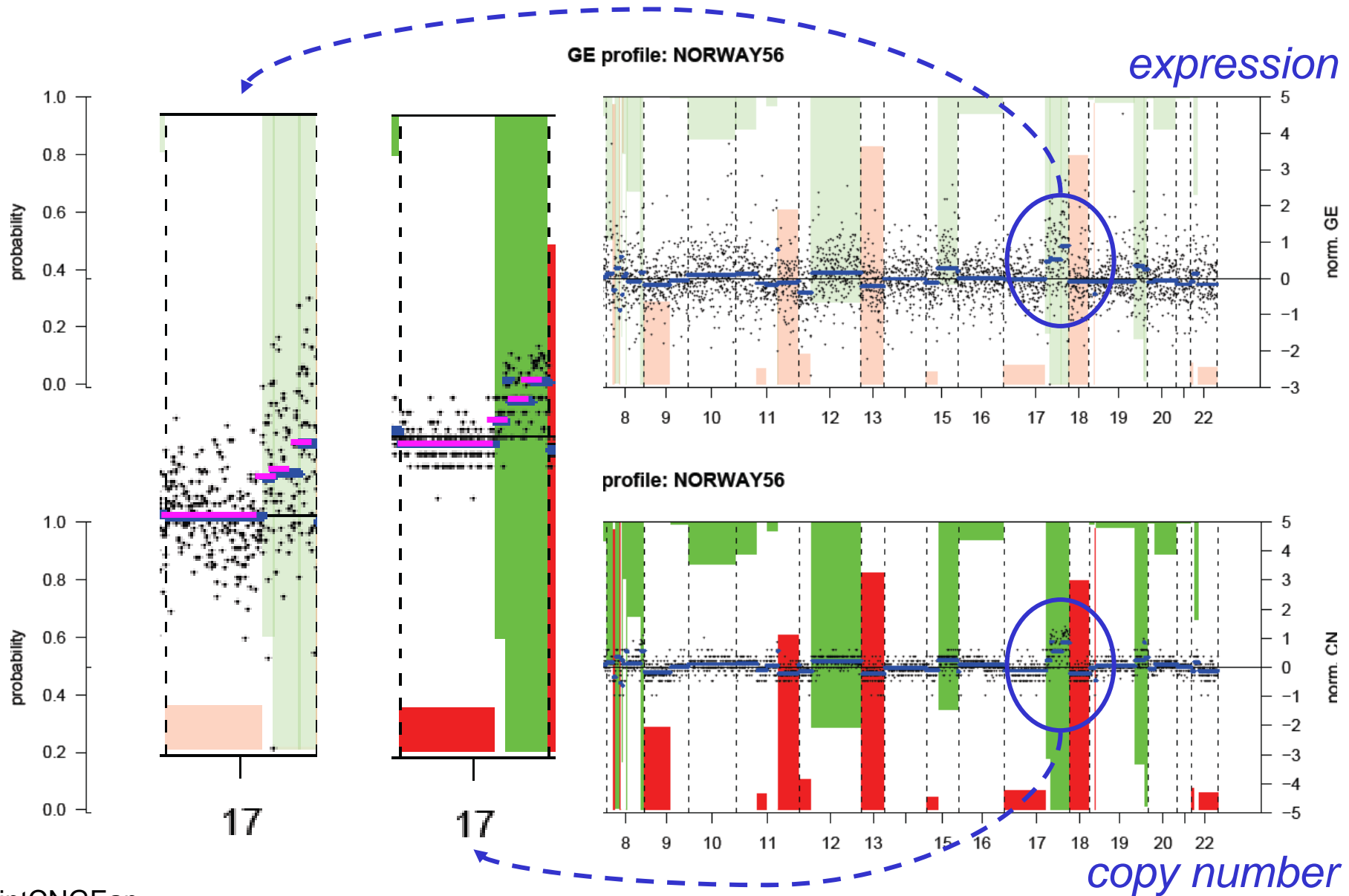
Shrinkage: borrow info within a region to increase test power.

Regional co-expression



(Caron *et al.*, Science, 2001; Zhou *et al.*, Cancer Res., 2003; Furge *et al.*, Cancer Res., 2004; Reyat *et al.*, Cancer Res., 2005; Wang *et al.*, GC&C, 2006; many others)

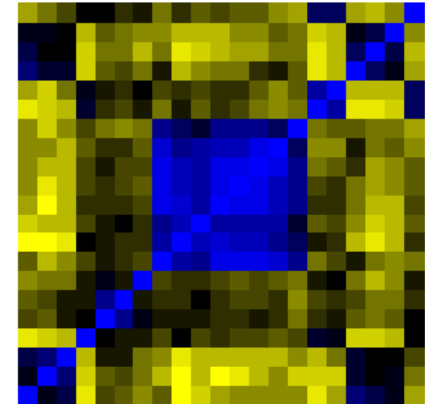
Regional co-expression



Regional co-expression

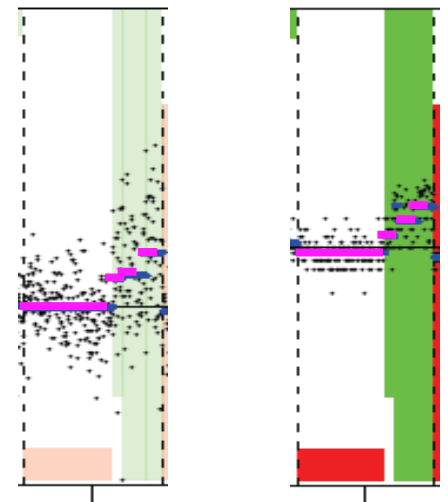
Regional co-expression due to:

- Copy number changes
- Common transcription factor
- Methylation
- ...



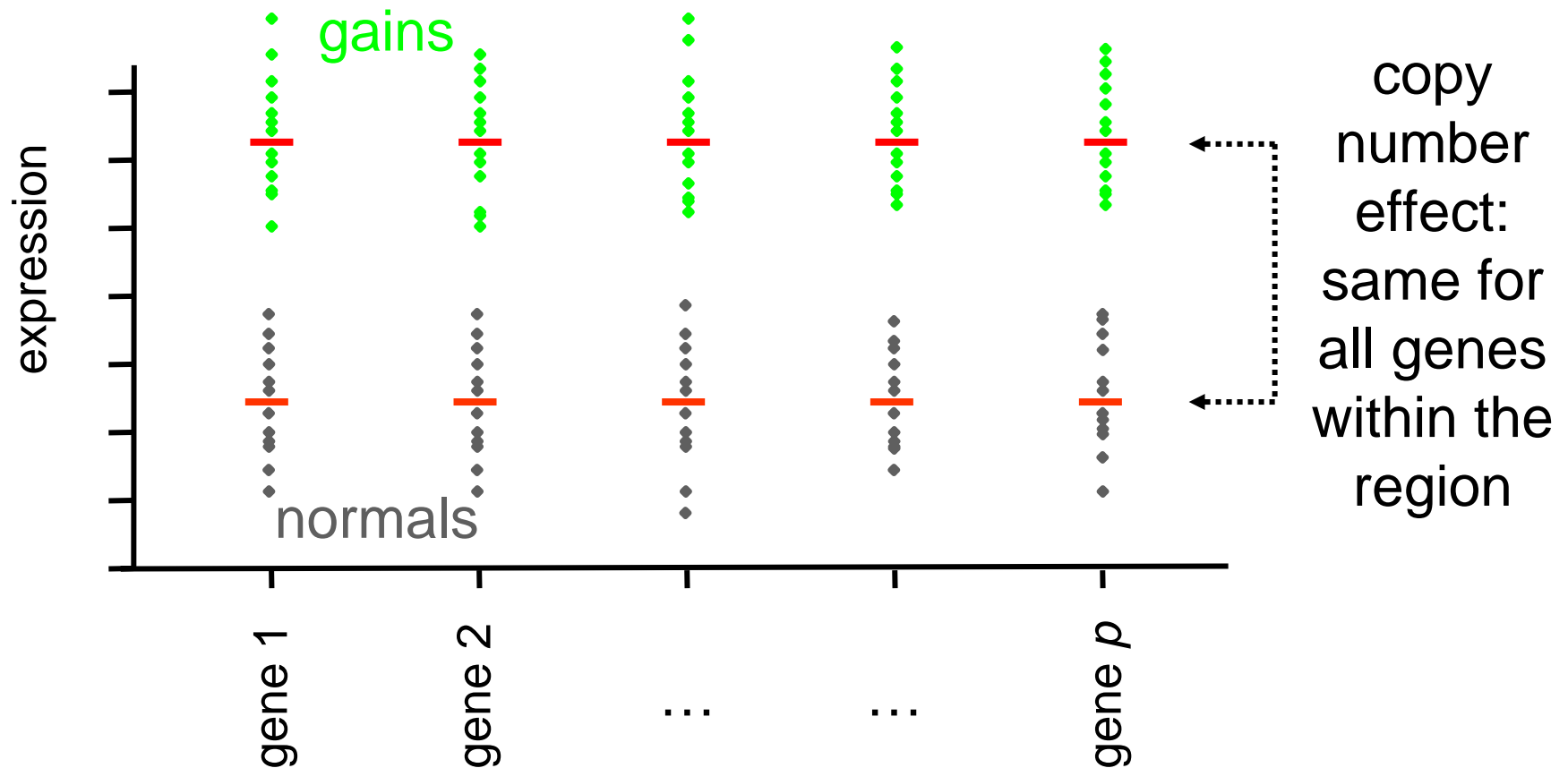
Goal

Test whether regional co-expression may be attributed to gene dosage.



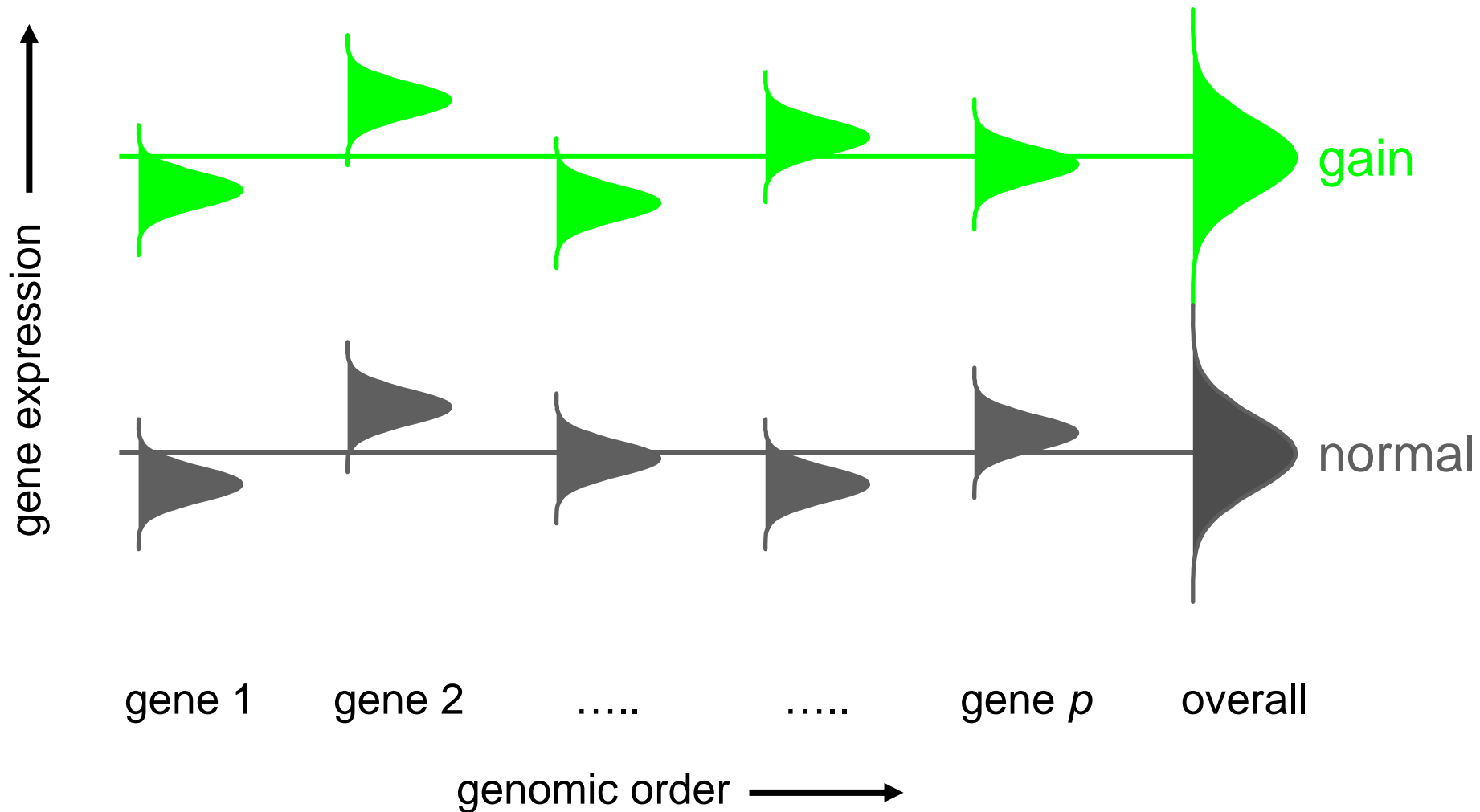
Model: graphically

Same CN signature, same effect on GE.



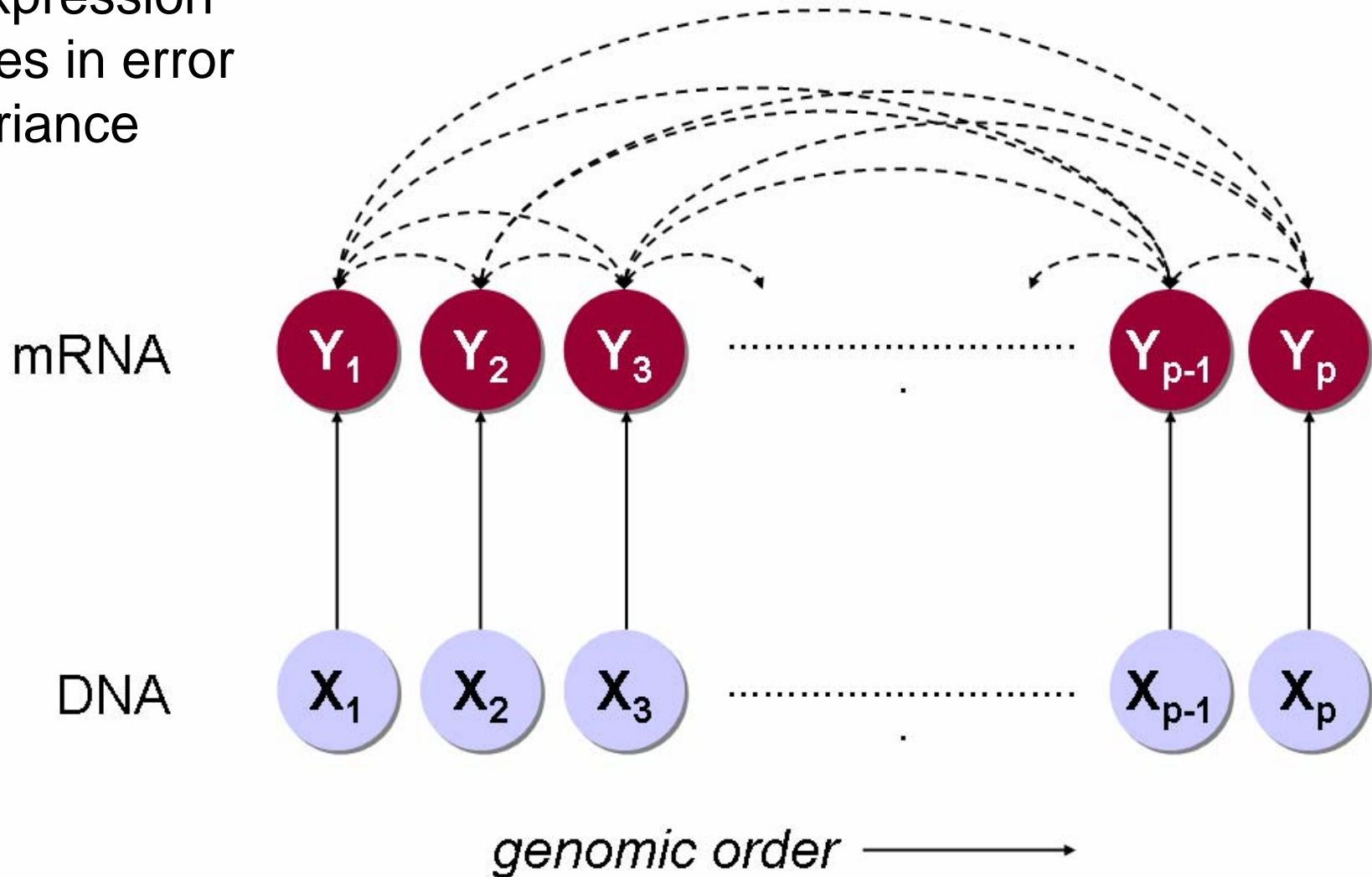
Model: graphically

Effect may be heterogeneous over genes.



Model: graphically

Accommodate other
co-expression
causes in error
covariance



Model

For single gene, assume:

$$Y_{ij} | \mathbf{X}_i = \mathbf{X}_i \boldsymbol{\beta}_j + \varepsilon_{ij} \quad \text{with } \varepsilon_{ij} \sim N(0, \sigma_j^2)$$

Prior knowledge requires:

$$\mathbf{R} \boldsymbol{\beta}_j \geq \mathbf{0}$$

In matrix notation, for all genes within a region:

$$\mathbf{Y}_1 | \mathbf{X} = \mathbf{X} \boldsymbol{\beta}_1 + \boldsymbol{\varepsilon}_1 \quad \text{with } \boldsymbol{\varepsilon}_1 \sim N(0, \sigma_1^2 \mathbf{I}_n)$$

$$\mathbf{Y}_2 | \mathbf{X} = \mathbf{X} \boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}_2 \quad \text{with } \boldsymbol{\varepsilon}_2 \sim N(0, \sigma_2^2 \mathbf{I}_n)$$

...

$$\mathbf{Y}_p | \mathbf{X} = \mathbf{X} \boldsymbol{\beta}_p + \boldsymbol{\varepsilon}_p \quad \text{with } \boldsymbol{\varepsilon}_p \sim N(0, \sigma_p^2 \mathbf{I}_n)$$

Model

As a SURE model:

$$\underbrace{\begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \vdots \\ \mathbf{Y}_p \end{pmatrix}}_{\overset{\circ}{\mathbf{Y}}} \Big| \mathbf{X} = \underbrace{\begin{pmatrix} \mathbf{X} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{X} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{X} \end{pmatrix}}_{\overset{*}{\mathbf{X}}} \underbrace{\begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix}}_{\overset{\circ}{\beta}} + \underbrace{\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_p \end{pmatrix}}_{\overset{\circ}{\varepsilon}}$$

with $\mathbf{R} \beta_j \geq \mathbf{0}$ in place for all genes, and

$$\text{Var}(\overset{\circ}{\varepsilon}) = \Sigma \otimes \mathbf{I}_n$$

with $\Sigma = \Upsilon \Lambda \Upsilon$, $\text{diag}(\Upsilon) = (\sigma_1, \dots, \sigma_p)$, and

$$\Lambda = \rho \mathbf{1}_{p \times p} + (1 - \rho) \mathbf{I}_{p \times p}$$

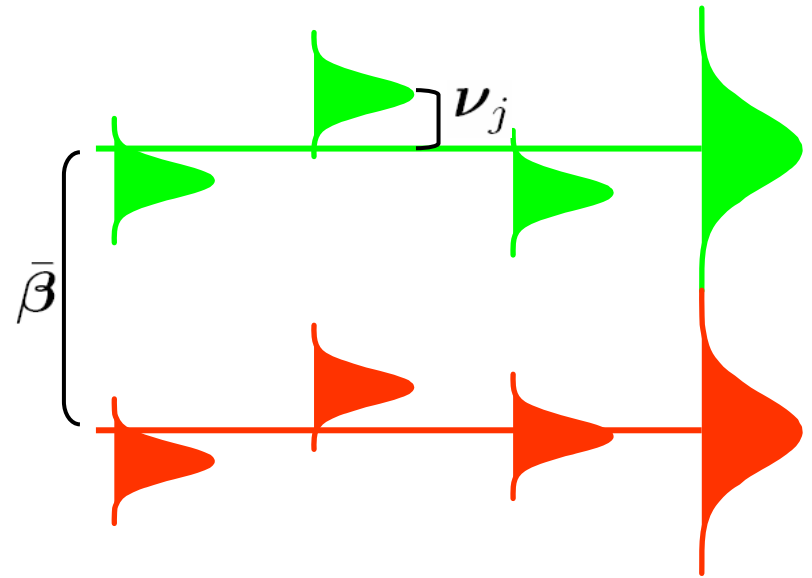
Model

Within SURE model treat coefficients as random:

$$\beta_j = \bar{\beta} + \nu_j$$

with:

- $\mathbf{R} \bar{\beta} \geq 0$
- $\nu_j \sim N(\mathbf{0}, \Theta)$
- $\text{diag}(\Theta) = (\tau_1^2, \dots, \tau_m^2)$
- $\text{Cov}(\nu_{jk}, \epsilon_{ij}) = 0$



Random coefficients also in GlobalTest.

Model

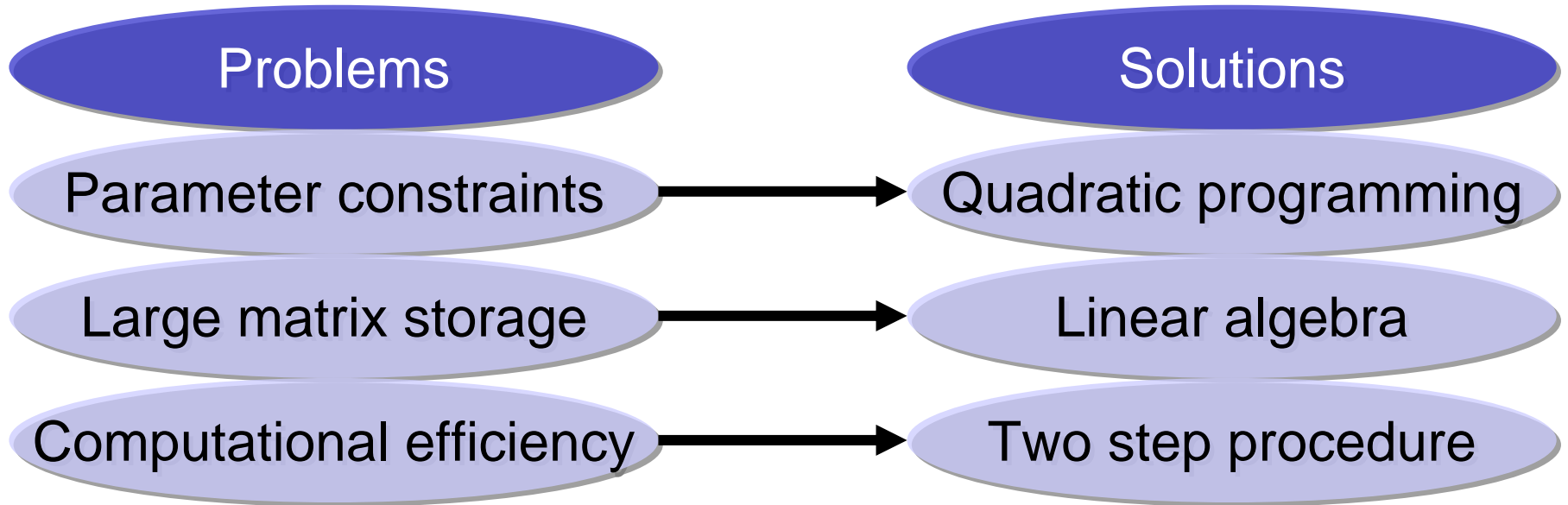
A random coefficients model:

$$\underbrace{\begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \vdots \\ \mathbf{Y}_p \end{pmatrix}}_{\overset{\circ}{\mathbf{Y}}} \Big| \mathbf{X} = \underbrace{\begin{pmatrix} \mathbf{X} \\ \mathbf{X} \\ \vdots \\ \mathbf{X} \end{pmatrix}}_{\overset{\circ}{\mathbf{X}}} \bar{\boldsymbol{\beta}} + \underbrace{\begin{pmatrix} \mathbf{X} & 0 & \dots & 0 \\ 0 & \mathbf{X} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{X} \end{pmatrix}}_{\overset{*}{\mathbf{X}}} \underbrace{\begin{pmatrix} \boldsymbol{\nu}_1 \\ \boldsymbol{\nu}_2 \\ \vdots \\ \boldsymbol{\nu}_p \end{pmatrix}}_{\overset{\circ}{\boldsymbol{\nu}}} + \underbrace{\begin{pmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \vdots \\ \boldsymbol{\varepsilon}_p \end{pmatrix}}_{\overset{\circ}{\boldsymbol{\varepsilon}}}$$

Its error component $\overset{\circ}{\boldsymbol{\delta}} = \overset{*}{\mathbf{X}} \overset{\circ}{\boldsymbol{\nu}} + \overset{\circ}{\boldsymbol{\varepsilon}}$ has variance:

$$\begin{aligned} \boldsymbol{\Omega} &= \text{Var}(\overset{\circ}{\boldsymbol{\delta}} \mid \overset{*}{\mathbf{X}}) = \boldsymbol{\Sigma} \otimes \mathbf{I}_n + \overset{*}{\mathbf{X}} (\mathbf{I}_p \otimes \boldsymbol{\Theta}) \overset{*}{\mathbf{X}}^T \\ &= \boldsymbol{\Sigma} \otimes \mathbf{I}_n + \mathbf{I}_p \otimes \mathbf{X} \boldsymbol{\Theta} \mathbf{X}^T = \boldsymbol{\Sigma} \oplus \mathbf{X} \boldsymbol{\Theta} \mathbf{X}^T \end{aligned}$$

Estimation



- Extra's:
- Robustification
 - Variance shrinkage

Testing

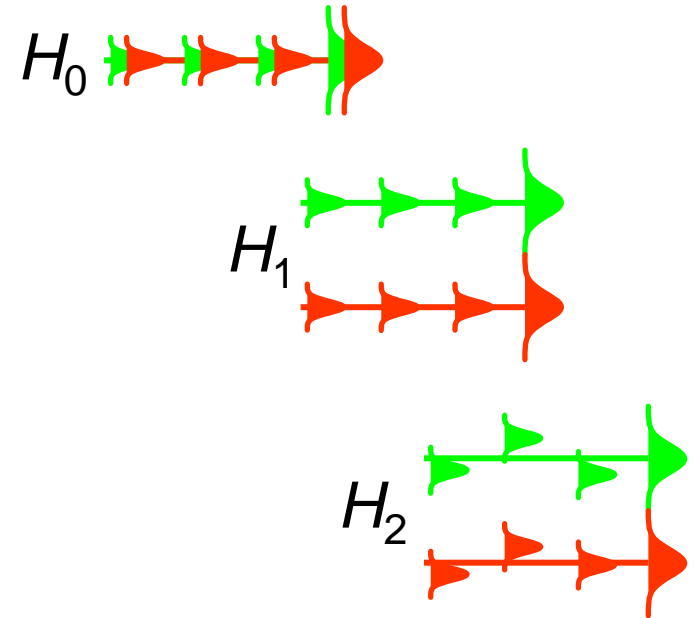
Three hypotheses of interest:

$$H_0 : \mathbf{R} \bar{\boldsymbol{\beta}} = \mathbf{0}, \tau^2 = 0$$

$$H_1 : \mathbf{R} \bar{\boldsymbol{\beta}} \geq \mathbf{0}, \tau^2 = 0$$

$$H_2 : \mathbf{R} \bar{\boldsymbol{\beta}} \geq \mathbf{0}, \tau^2 \geq 0$$

At least one inequality strict in H_1 and H_2 .



Test statistic : likelihood ratio statistic

p -value calculation : re-sampling

Breast cancer data

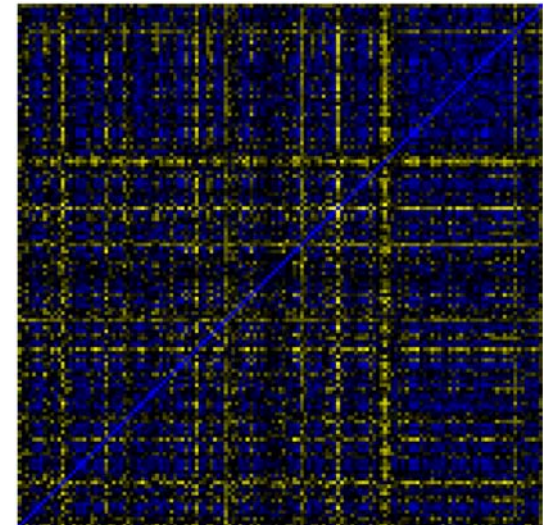
GENES, CHROMOSOMES & CANCER 45:1033–1040 (2006)

Distinct Patterns of DNA Copy Number Alteration Are Associated with Different Clinicopathological Features and Gene-Expression Subtypes of Breast Cancer

Anna Bergamaschi,^{1,2} Young H. Kim,² Pei Wang,^{3†} Therese Sørlie,¹ Tina Hernandez-Boussard,⁴ Per E. Lonning,⁵ Robert Tibshirani,^{3,6} Anne-Lise Børresen-Dale,^{1,7} and Jonathan R. Pollack^{2*}

- 85 breast cancers:
 - copy number data
 - gene expression data
- region on 16p
- 145 features
- gained in 31 samples

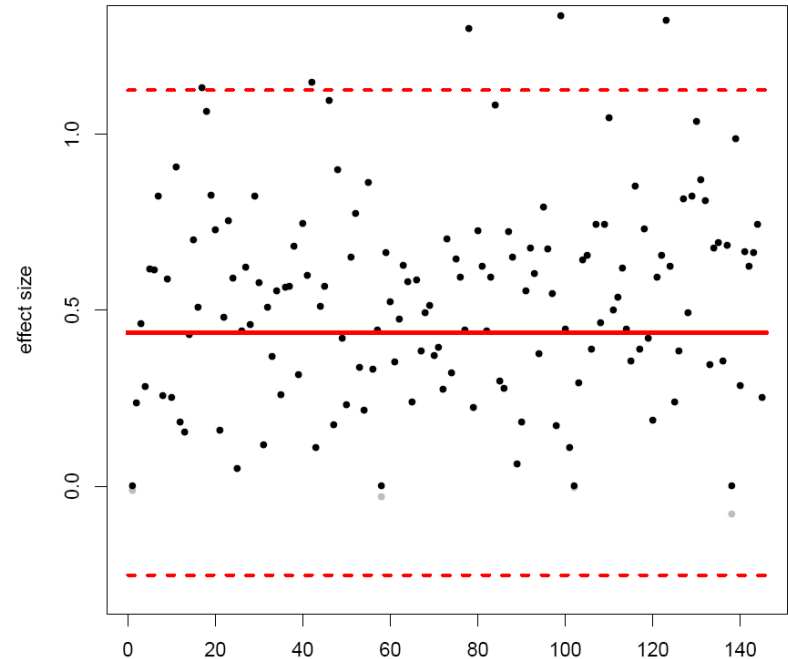
region on 16p



genes pre-dominantly positively correlated

Breast cancer data

- sign. shared CN effect
- sign. heterogeneity
- residual co-expression



Est.	$\hat{\beta}$	$\hat{\tau}^2$	$\hat{\rho}$	$\hat{\kappa}$	p -value H_0 vs H_1	p -value H_1 vs H_2	p -value H_0 vs H_2
normal	0.436	0.119	0.140	0.000	< 0.001	< 0.001	< 0.001
robust	0.374	0.150	0.155	0.000	< 0.001	< 0.001	< 0.001
normal	0.437	0.118	0.140	0.039	< 0.001	< 0.001	< 0.001
robust	0.376	0.149	0.155	0.056	< 0.001	< 0.001	< 0.001

Thanks to...

Hans Berkhof
Aad van der Vaart

Mark van de Wiel
Bauke Ylstra

