APPLICATION OF MIXED MODELS TO THE EXPRESSION MICROARRAY DATA FOR THE IDENTIFICATION OF GENES RESPONSIBLE FOR THE INTRAMUSCULAR FAT CONTENT IN PIGS

Joanna Szyda[1], Ina Widerøe[2], Eli Grindflek[2], Przemysław Biecek[3], Jakob Hedegaard[4], Christian Bendixen[4] and Sigbjørn Lien[2]

1. Institute of Animal Genetics, Wroclaw University of Life Sciences, Wroclaw, Poland
2. Norwegian University of Life Sciences and Centre for Integrative Genetics, Aas, Norway
3. Institute of Mathematics, Polish Academy of Sciences, Wrocław, Poland
4. Department of Genetics and Biotechnology, University of Aarhus, Tjele, Denmark

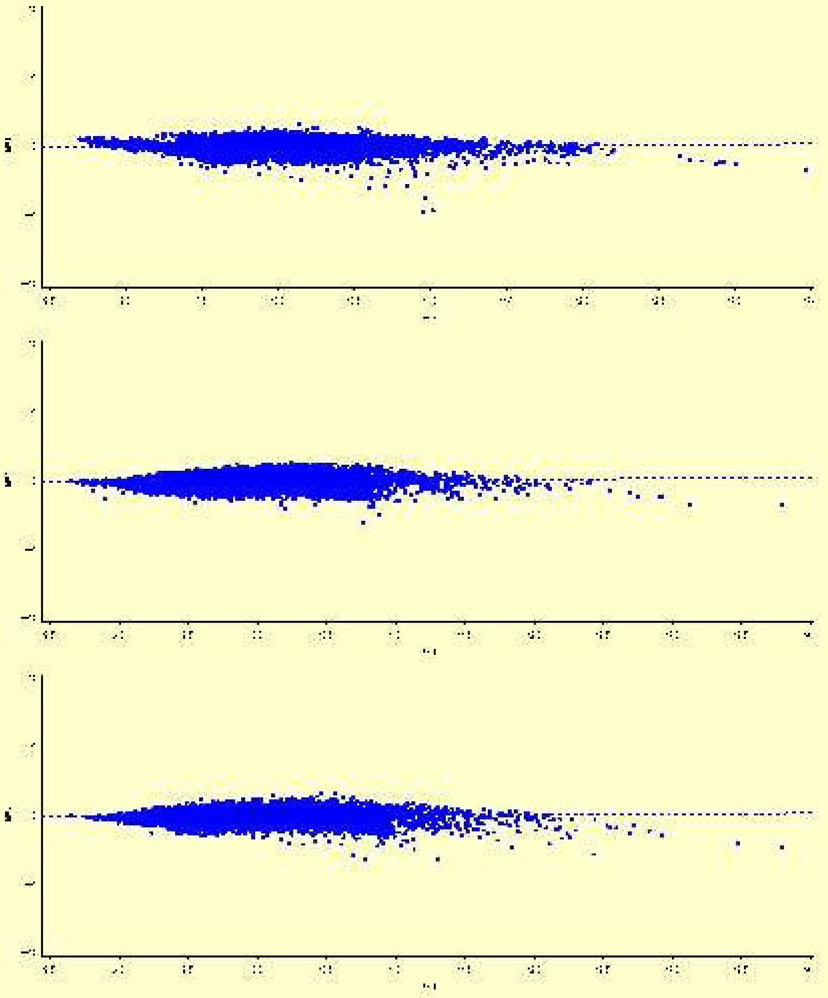
1. **Data**
2. **Normalisation**
3. **Modelling**
4. **Results**
5. **Conclusions**

EXPERIMENTAL DESIGN

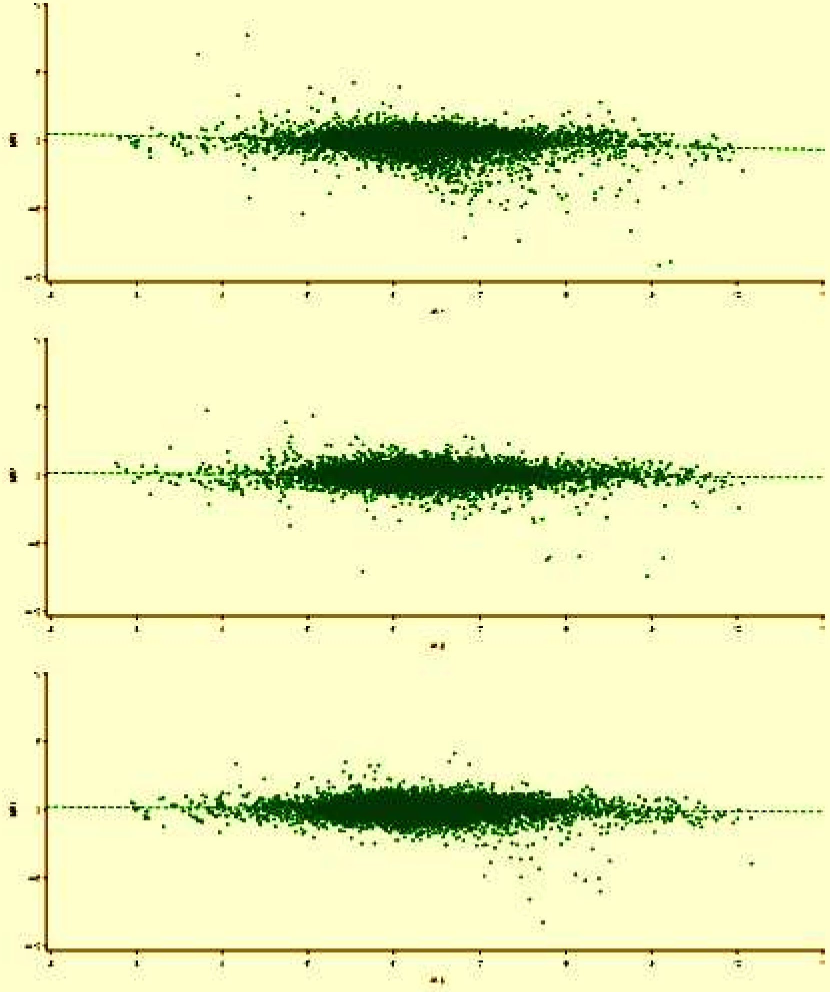
ARRAY	FAMILY	FAMILY
1	A	D
2	E	C
3	E	D
4	A	B
5	C	C
6	D	E
7	C	B
8	D	A
9	E	A
10	E	D
11	D	B
12	D	D
13	D	B
14	C	B

NORMALISATION (MA plot)

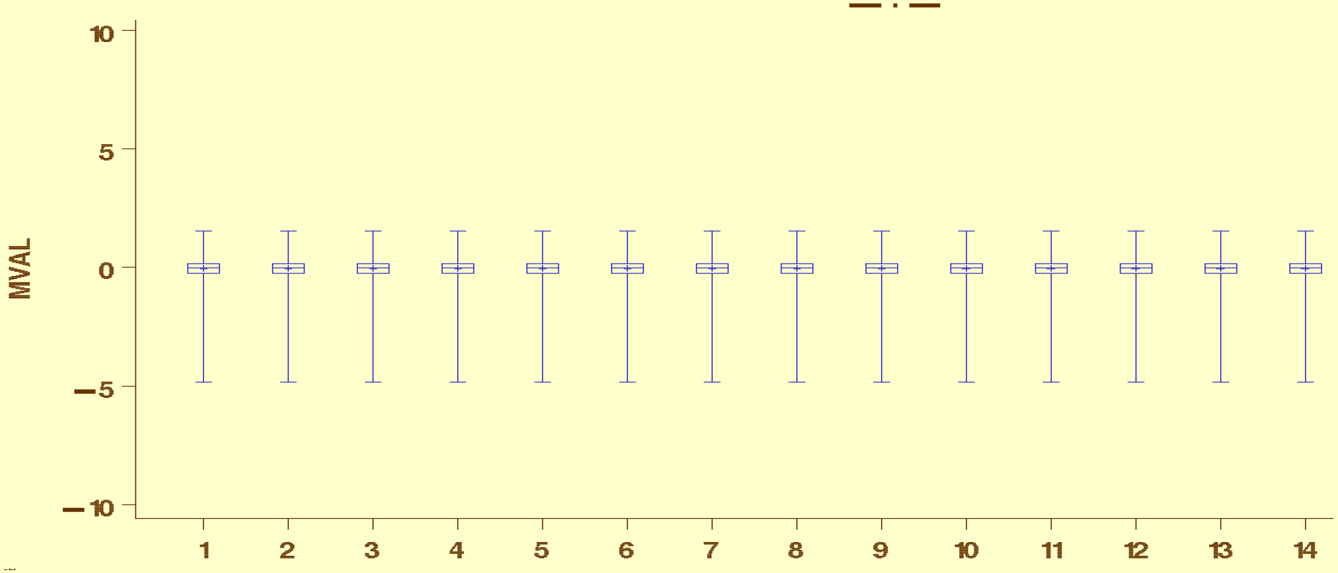
LIMMA



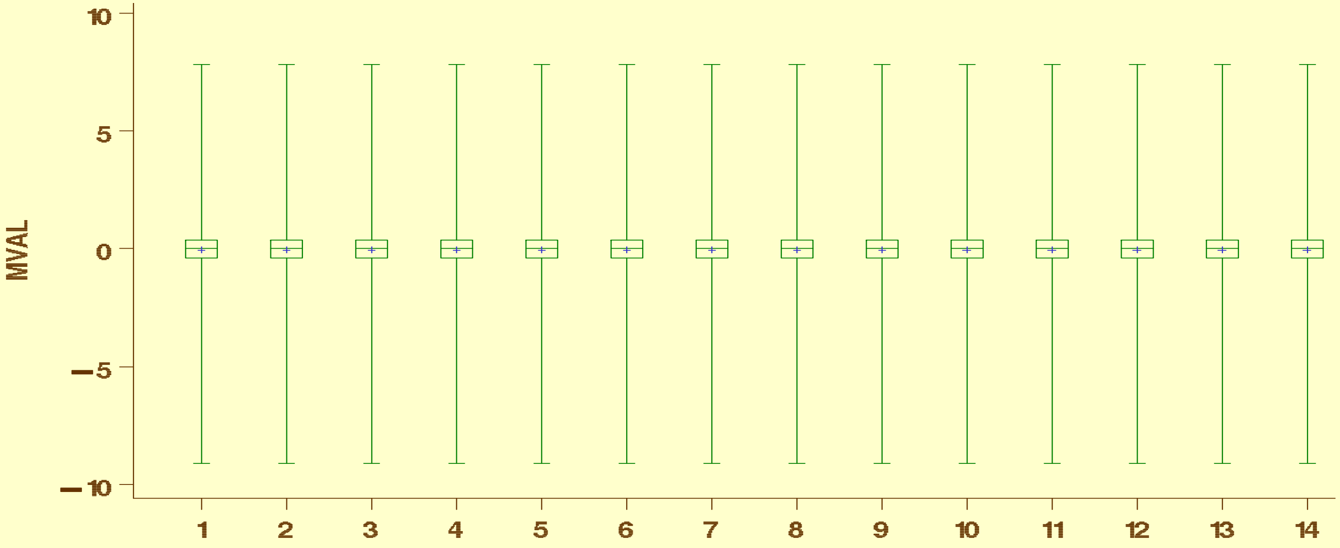
MARRAY



NORMALISATION (boxplot)



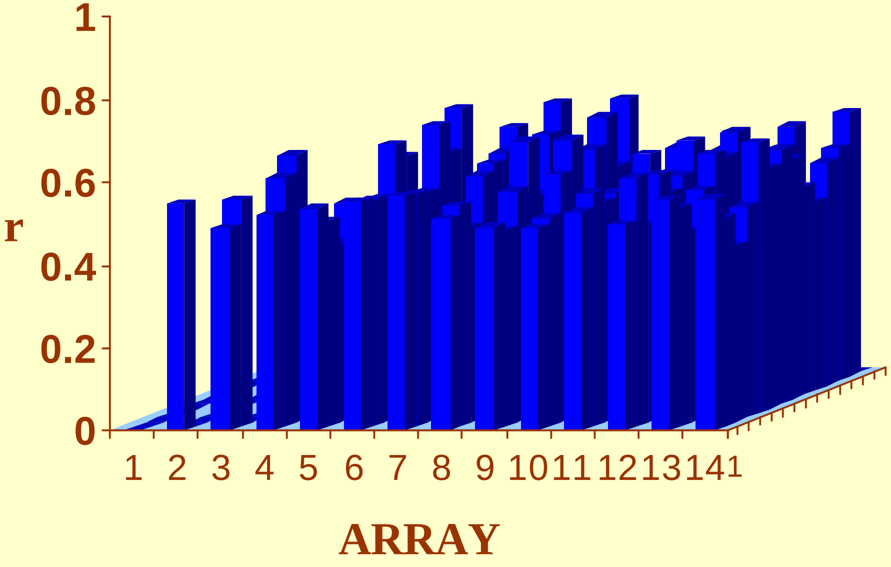
LIMMA



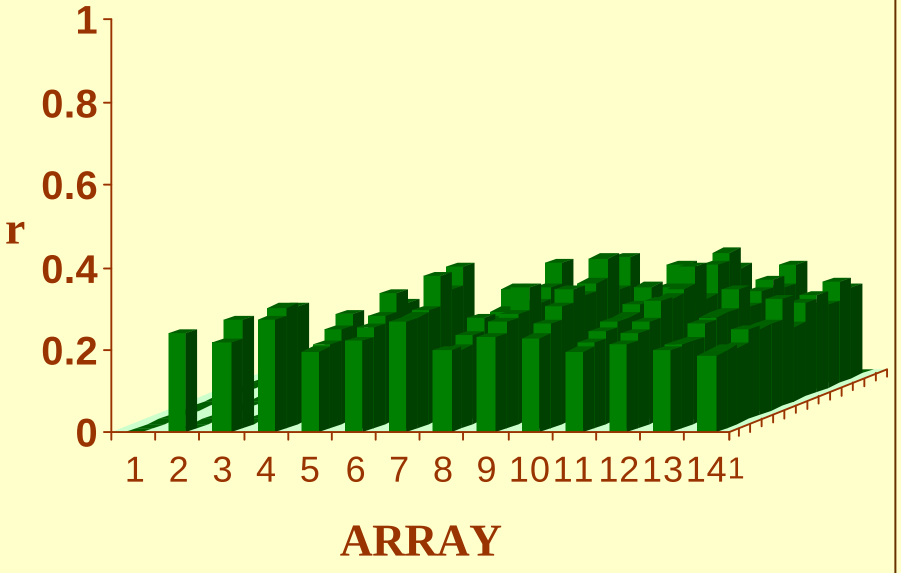
MARRAY

NORMALISATION (correlations between M values)

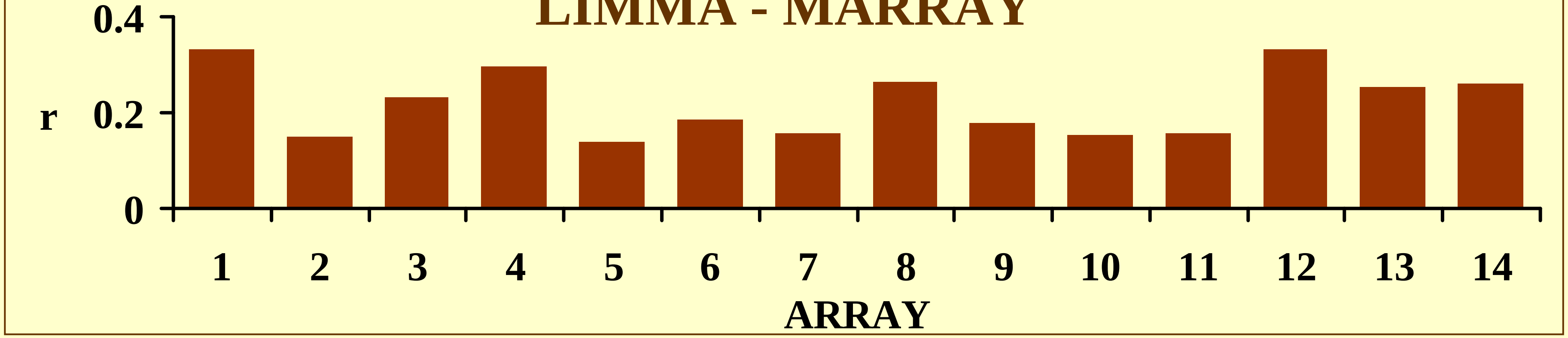
LIMMA



MARRAY



LIMMA - MARRAY



ANALYSIS (models)

GENE ONTOLOGY

Homo sapiens dbase 19.01.08; 912 terms

GENE

11 732 genes

univariate

$$y = \mu + s + go + e$$

$$y = \mu + s + gene + e$$

$$gene \sim N\left(0, \mathbf{G}\sigma^2_{gene}\right)$$

multivariate

$$y = \mu + s + go + ar + e$$

$$ar \sim N\left(0, \mathbf{A}\sigma^2_{array}\right)$$

$$y = \mu + s + gene + ar + e$$

$$gene \sim N\left(0, \mathbf{G}\sigma^2_{gene}\right)$$

$$ar \sim N\left(0, \mathbf{A}\sigma^2_{array}\right)$$

y – M value

s – spot (2)

go – gene ontology term (912)

gene – gene ID (11 732)

ar – array (14)

ANALYSIS (covariance matrices)

$$ar \sim N\left(0, \mathbf{A}\sigma^2_{array}\right)$$

A 14×14

- **based on similarity in experimental design (families A-E)**

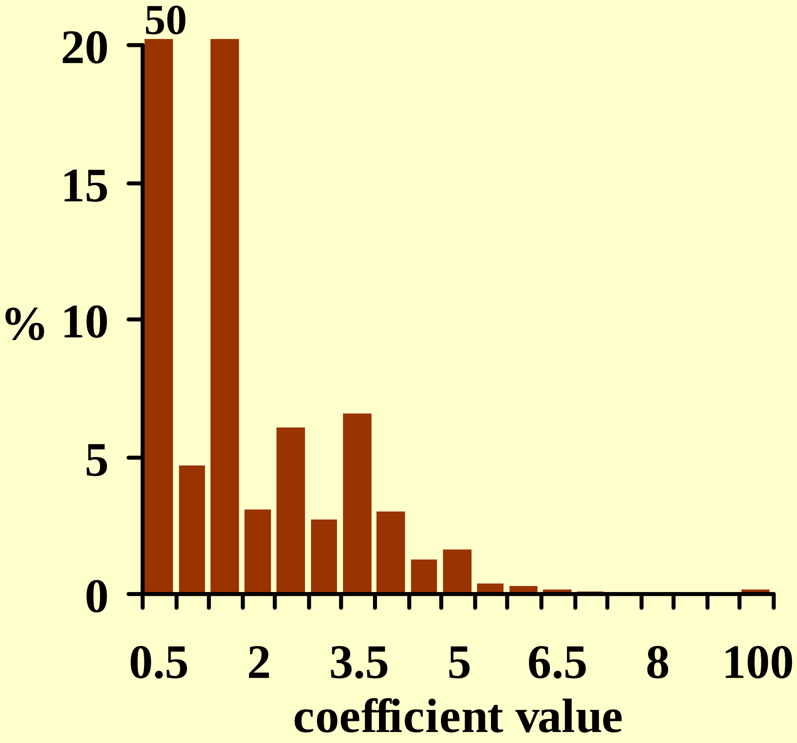
$$gene \sim N\left(0, \mathbf{G}\sigma^2_{gene}\right)$$

G 11732×11732

- **based on maximum pairwise semantic similarity between GO term sets**

Schlicker et al. 2006

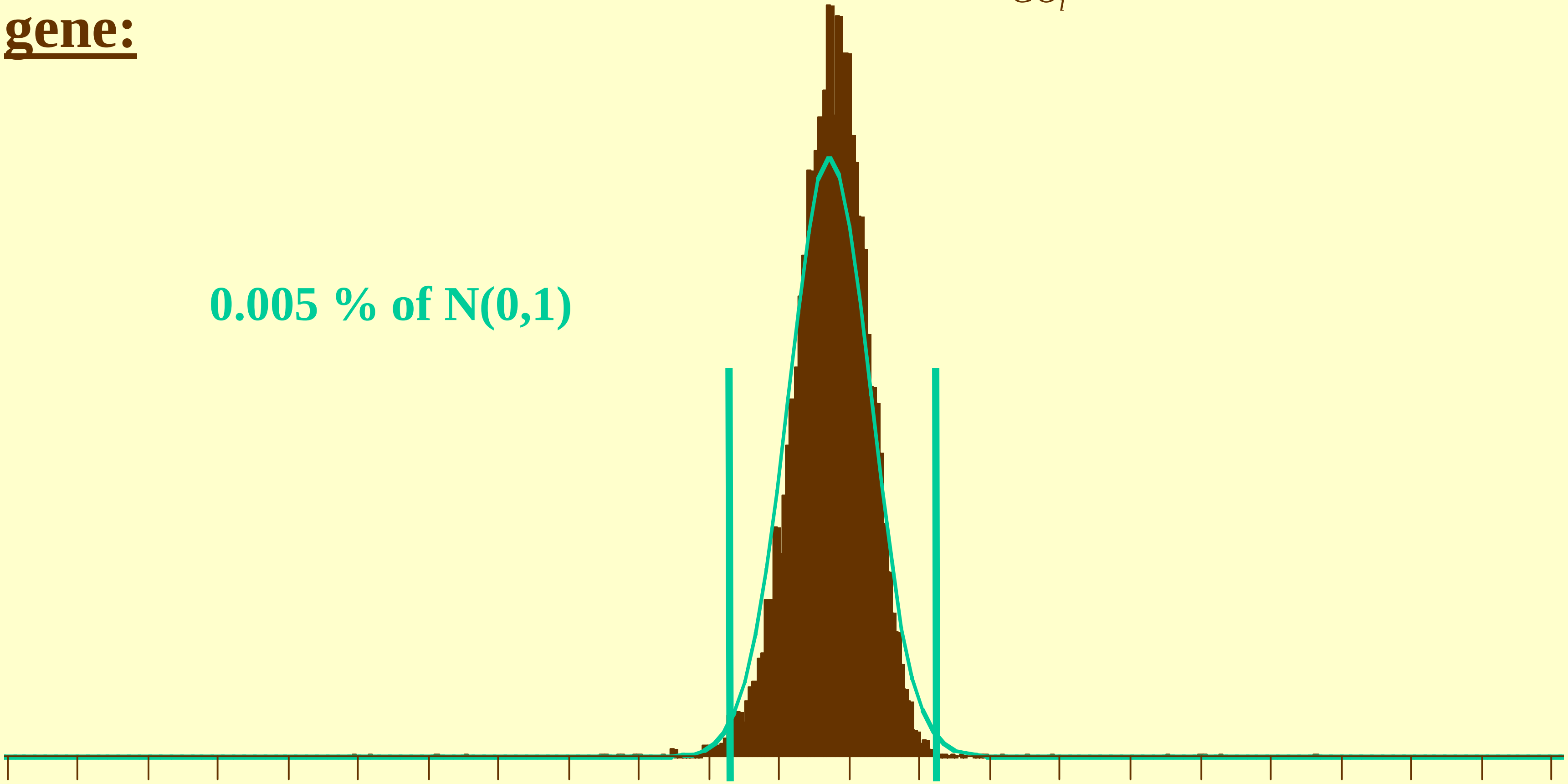
- **identity matrix**



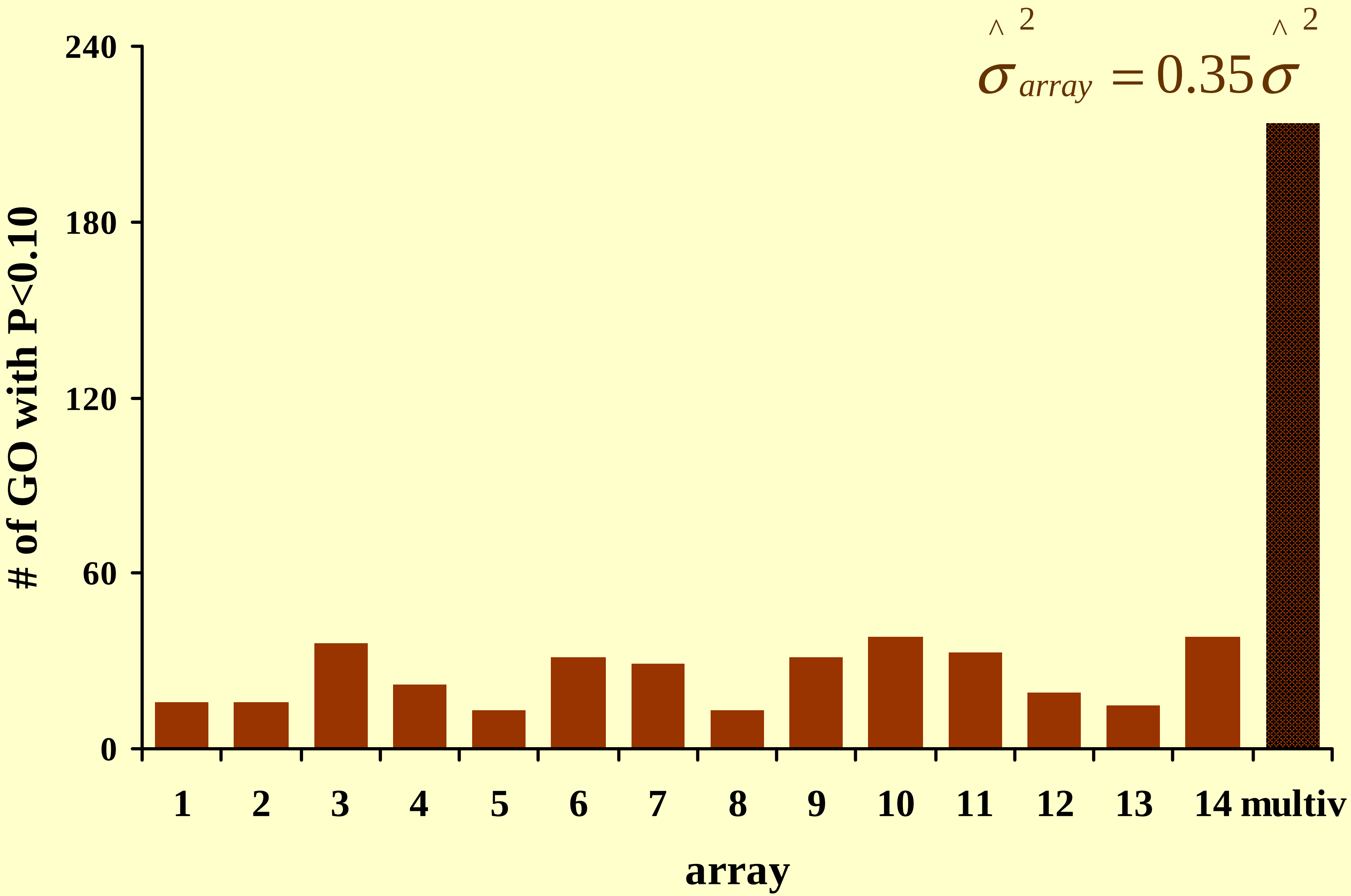
ANALYSIS (effect selection)

gene ontology: $H_0:\ GO_i = 0 \qquad W = \dfrac{\hat{GO}_i^{\,2}}{\hat{\sigma}^2_{GO_i}} \sim \chi^2_1 \qquad P^* = P\,N_{GO}$

gene:

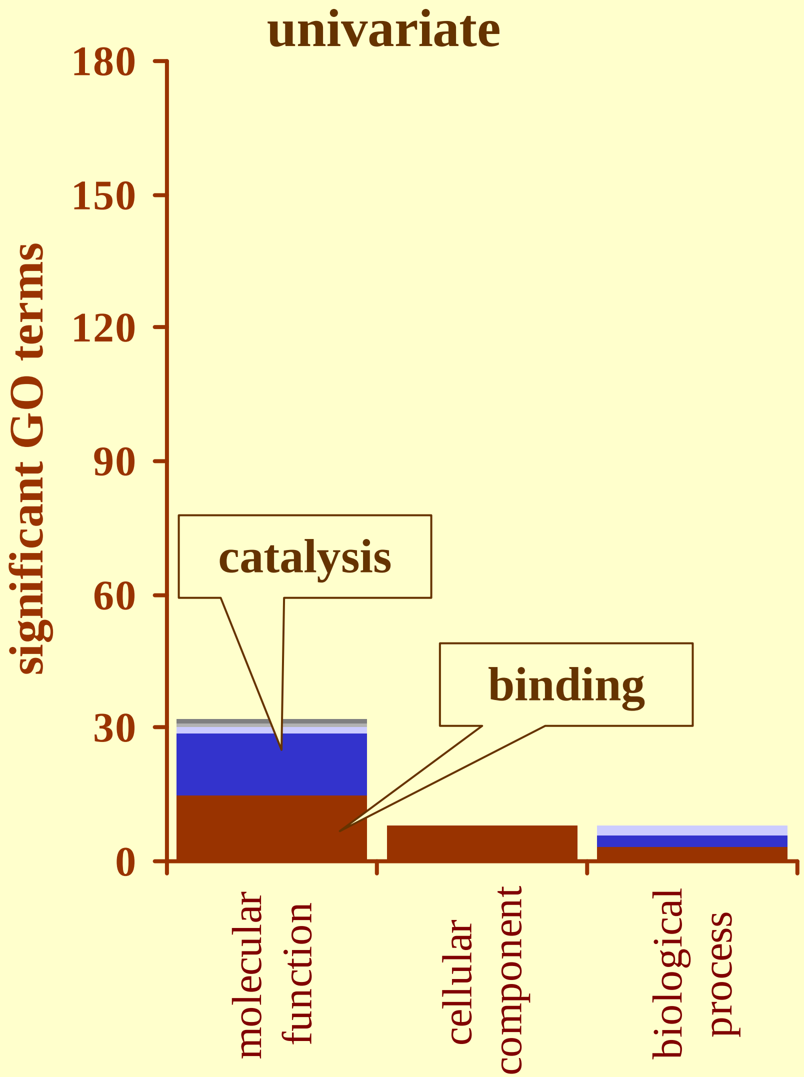

0.005 % of N(0,1)

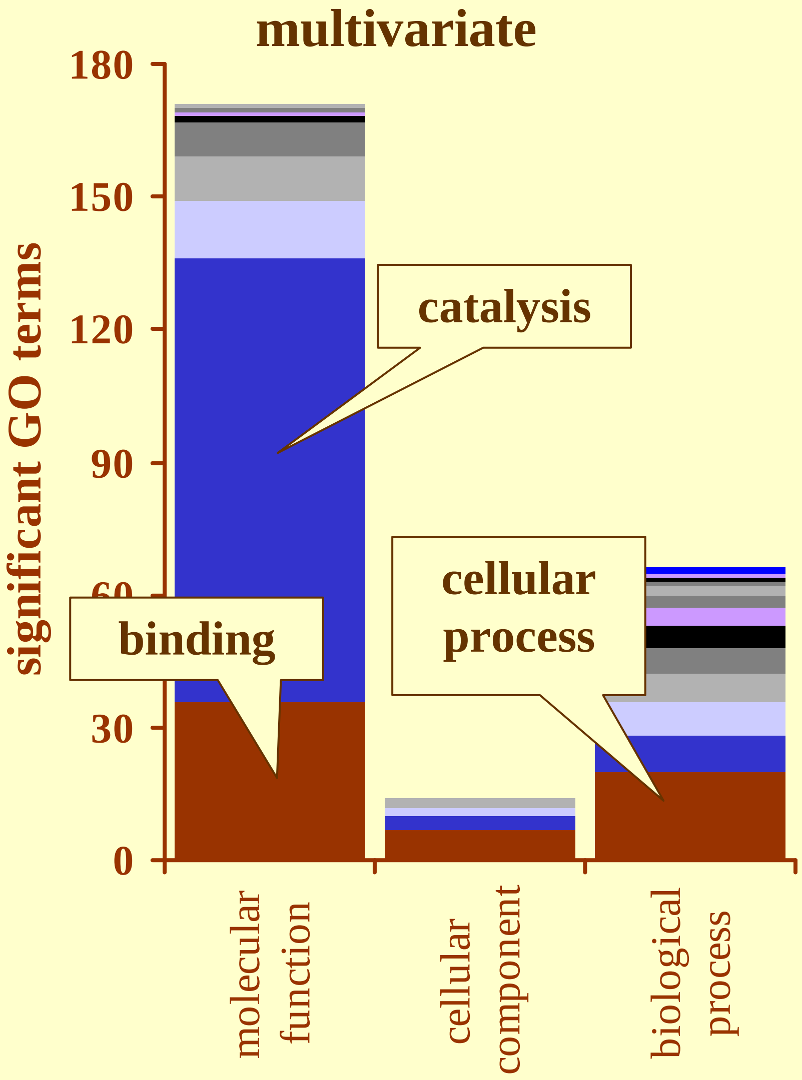
RESULTS (gene ontology)



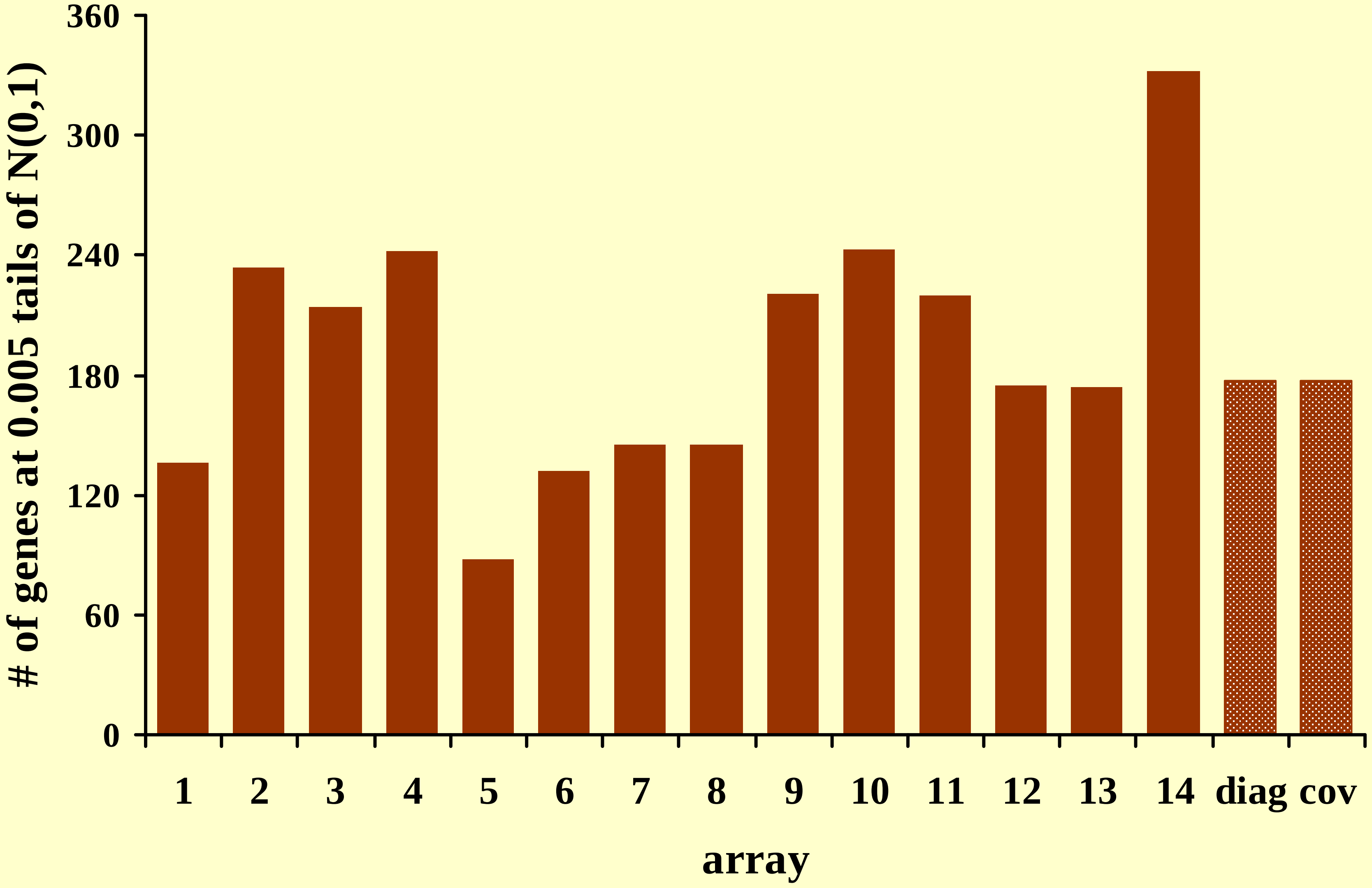
$$\hat{\sigma}^2_{array} = 0.35\,\hat{\sigma}^2$$

RESULTS (gene ontology)

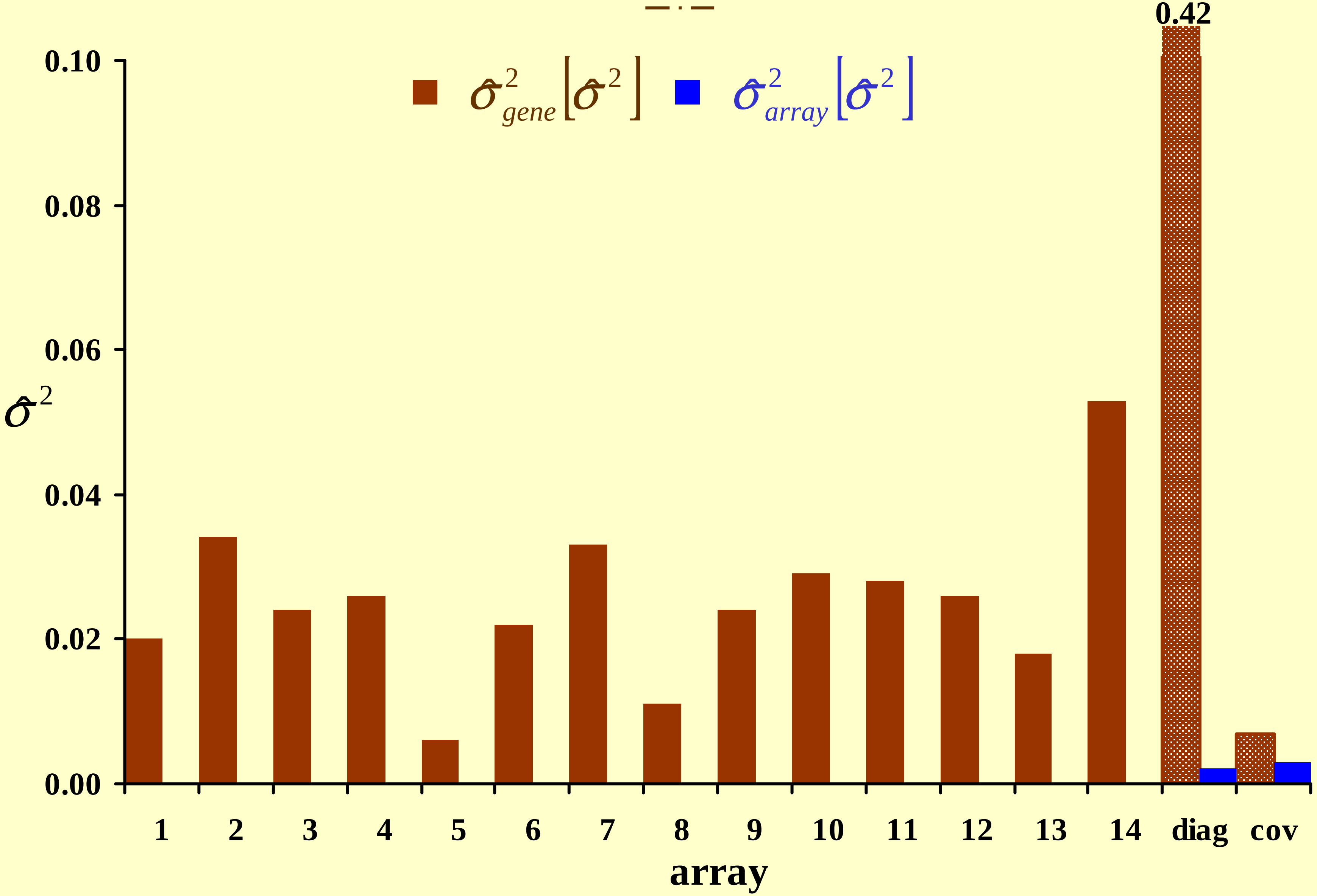




RESULTS (genes)



RESULTS (gene)



CONCLUSIONS

1. Large difference between normalisations
2. Low correlations between arrays
3. Good correspondence in effect estimation
    - GO terms significant on: 11+1, 10+1, 6+1 models
    - Genes significant on: <u>2×(14+1)</u>, 3×(10+1) models

Thank you !