

**The chicken data set analysed
by DJF**

 Li Jiang
 Lene Nagstrup Conley
 Axel Skarman
 Bart Buitenhuis
 Henrik Hornshøj
 Jakob Hedegaard
 Peter Sørensen

EADGENE and SABRE Post-Analyses Workshop
 12-14th November 2008, Animal Sciences Group, Wageningen UR, Lelystad

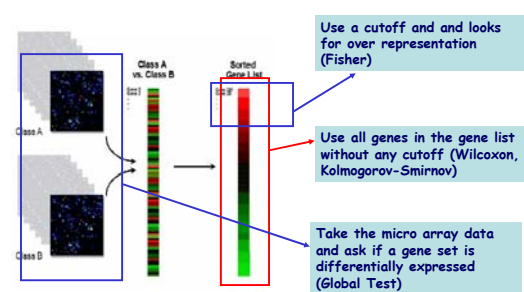
Gene set Analysis

- Focus on sets of genes instead of single genes

Methods

- Fisher
- Wilcoxon
- Kolmogorov-Smirnov
- topGO
 - Elimination
 - Weight
- Global Test

Use of data



Use a cutoff and look for overrepresentation (Fisher)

Use all genes in the gene list without any cutoff (Wilcoxon, Kolmogorov-Smirnov)

Take the micro array data and ask if a gene set is differentially expressed (Global Test)

Methods

- Fisher
- Wilcoxon
- Kolmogorov-Smirnov
- topGO
 - Elimination
 - Weight
- Global Test

Fisher

- Choose a cutoff that defines the differentially expressed genes.

	# genes above cutoff	# genes below cutoff	sum
# genes in G	10	40	50
# genes not in G	90	9900	9950
sum	100	9900	10000

- Test for overrepresentation of these genes in the gene set.

Methods

- Fisher
- Wilcoxon
- Kolmogorov-Smirnov
- topGO
 - Elimination
 - Weight
- Global Test

Wilcoxon

- Rank all genes without using any cutoff.
- Use the rank of the genes and compare the distributions of the ranks.

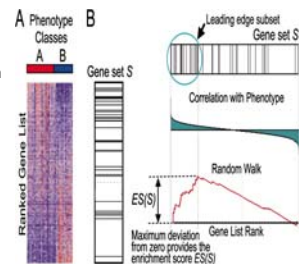
Gene	Score	G member
Gene 1	score 1	+
Gene 2	score 2	-
Gene 3	score 3	+
Gene 4	score 4	+
...		
Gene 100	score 100	-
Gene 101	score 101	+
...		
Gene 6000	score 6000	-

Methods

- Fisher
- Wilcoxon
- Kolmogorov-Smirnov
- topGO
 - Elimination
 - Weight
- Global Test

Kolmogorov-Smirnov

- Use all genes without any cutoff
- Use a running sum over the ranked gene list to compute an enrichment score.
- The running sum increase when the gene is part of the current geneset and decrease otherwise. The maximum deviation gives the enrichment score.
- Subramanian et al. (2005)



Methods

- Fisher
- Wilcoxon
- Kolmogorov-Smirnov
- topGO
 - Elimination
 - Weight
- Global Test

Top GO

- Use the GO graph structure in the analysis.
- Two methods, *Elim* and *Weight*, used.
- Alexa et al. (2006)



Elim

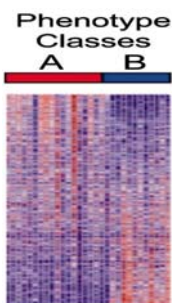
- A bottom-up strategy.
- Removes the genes in significant child nodes.
- Finds more specific nodes.



Methods

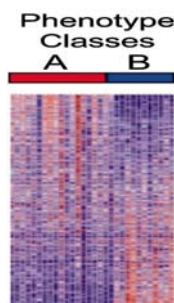
- Fisher
- Wilcoxon
- Kolmogorov-Smirnov
- topGO
 - Elimination
 - Weight
- Global Test

Global Test



- Use all expression data $X = (x_{ij})$
- Self contained and subject sampling
- Goeman et al. (2004)

Global Test

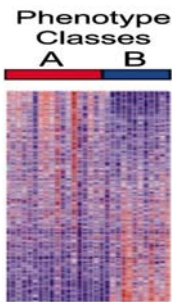


- The model

$$E(Y_i | \beta) = h^{-1} \left(\alpha + \sum_{j=1}^m x_{ij} \beta_j \right)$$
- Uses a the null hypothesis

$$h_0 : \beta_1 = \beta_2 = \dots = \beta_m = 0$$
- Not possible to test in a classical way

Global Test

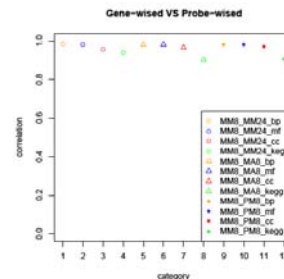


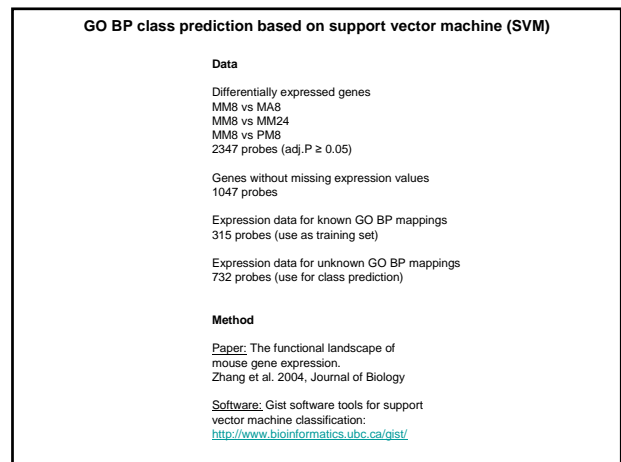
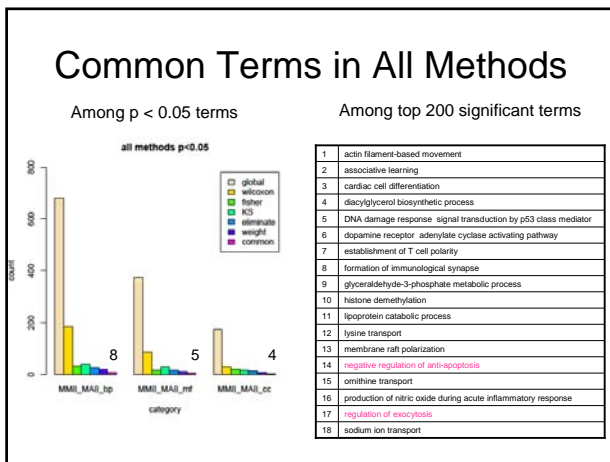
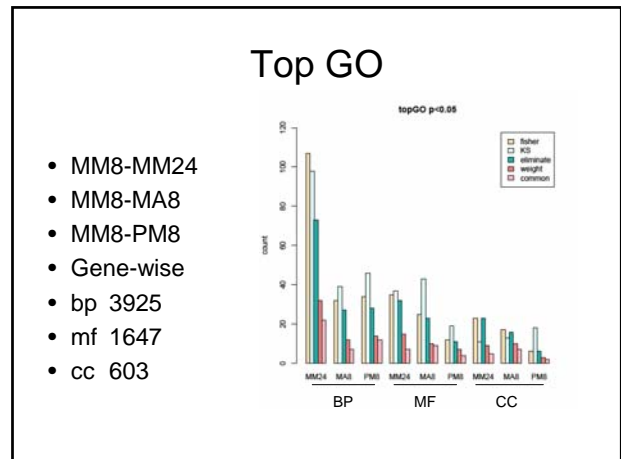
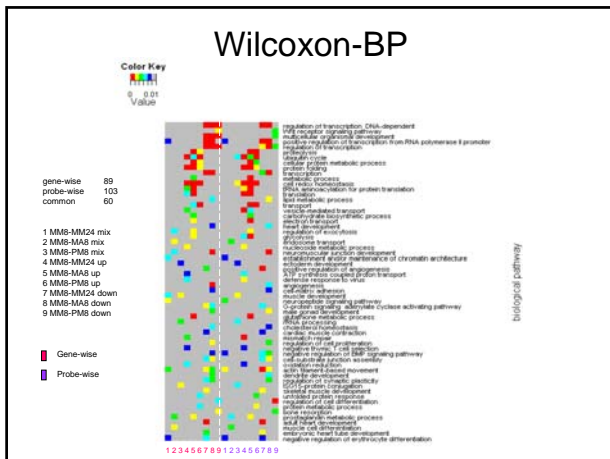
- If $\beta_1, \beta_2, \dots, \beta_m$ are assumed to come from the same distribution, with the same variance τ^2 and mean equal to zero, the null hypothesis becomes

$$h_0 : \tau^2 = 0$$

Gene-wise VS Probe-wise

- Wilcoxon test
- bp 2553
- mf 1436
- cc 481
- Maximum 0.98
- Minimum 0.90
- Mean 0.96

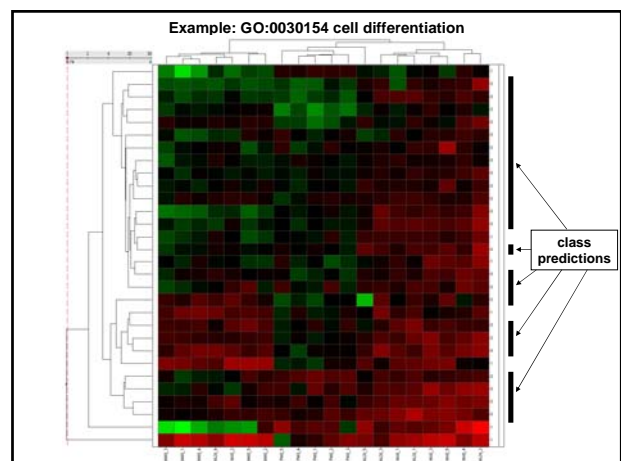


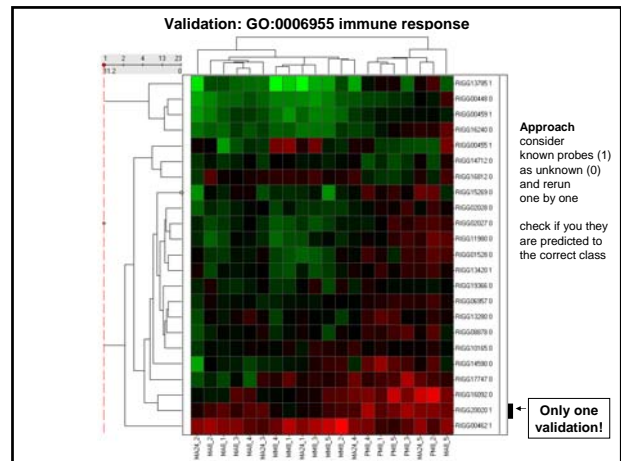
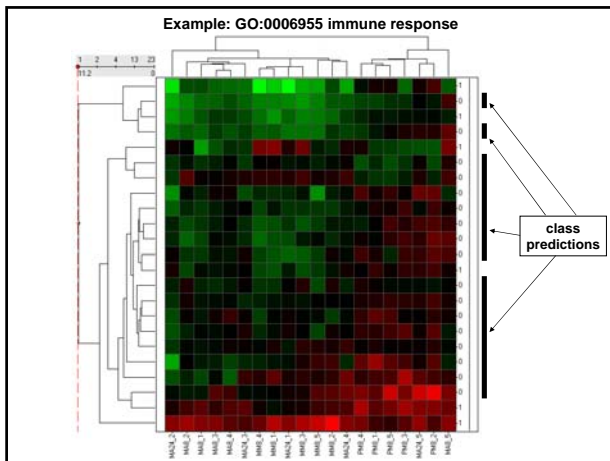


GO BP class prediction based on support vector machine (SVM)

Discriminant ≥ 1.0

GO.ID	definition	probe count known	probe count predicted
GO:0030154	cell differentiation	12	23
GO:0007155	cell adhesion	9	16
GO:0070910	cell cycle	8	16
GO:0006955	immune response	7	17
GO:0045892	negative regulation of transcription, DNA-dependent	5	17
GO:0006412	translation	11	16
GO:0006468	protein amino acid phosphorylation	13	16
GO:0002835	negative regulation of cell proliferation	8	14
GO:0016055	Wnt receptor signaling pathway	5	14
GO:0007165	signal transduction	22	12
GO:0007275	multicellular organismal development	17	12
GO:0002634	positive regulation of cell proliferation	8	12
GO:0007242	intracellular signaling cascade	12	11
GO:0007605	sensory perception of sound	5	11
GO:0045893	positive regulation of transcription, DNA-dependent	6	11
GO:0015031	protein transport	9	10
GO:0043066	negative regulation of apoptosis	6	10
GO:0001501	skeletal development	7	9
GO:0006350	transcription	24	8
GO:0007294	small GTPase mediated signal transduction	8	8
GO:0006152	metabolic process	9	8
GO:0016491	negative regulation of transcription	6	8
GO:0006457	protein folding	9	7
GO:0006886	intracellular protein transport	5	6
GO:0001525	angiogenesis	7	5
GO:0006118	electron transport	6	5
GO:0045449	regulation of transcription	12	5





AARHUS UNIVERSITY

Thank you for your attention!

EADGENE and SABRE Post-Analyses Workshop
12-14th November 2008, Animal Sciences Group, Wageningen UR, Lelystad

Det Jørgensen-Instituttet