





 European Animal Disease Genomics Network of Excellence
for Animal Health and Food Safety 

Genomics for Animal Health: Outlook for the Future
13- 14th October 2009, Muséum National d'Histoire Naturelle, Paris, France



Post-analyses of microarray data
Florence Jaffrézic
INRA, Jouy-en-Josas, France.

 This publication represents the views of the Authors, not the EC.
The EC is not liable for any use that may be made of the information. 


 



Introduction

Nov. 2006: Workshop organized by WP 1.4
Best practice for the analysis of two-colour microarray data.

- > Data quality control
- > Normalization
- > Detection of differentially expressed genes.

⇒ **What to do with these lists of genes ?**
⇒ Aim of the **post-analyses workshop** (Nov. 2008)

 Genomics for Animal Health: Outlook for the Future
13-14th October 2009, Muséum National d'Histoire Naturelle, Paris, France


 



Post-analyses workshop

Five aspects investigated:

- 1) **Re-annotation** of the probe set on DNA microarrays.
- 2) **Pathway** analyses to identify biological processes from microarray results.
- 3) **Reverse engineering** of gene **regulatory networks**.
- 4) Integration of gene expression and **QTL** studies.
- 5) **Prediction of phenotypic outcomes** using gene expression results.

⇒ **Papers in BMC Proceedings.**

 Genomics for Animal Health: Outlook for the Future
13-14th October 2009, Muséum National d'Histoire Naturelle, Paris, France


 



Pathway analysis

Data set distributed among the participants:
Experimental challenge of chicken with two types of *Eimeria*.
Available from ArrayExpress (E-MEXP-1972).

Same list of differentially expressed genes for all groups.

Twelve pathway related software tools (commercial or publicly available) were tested.


 Genomics for Animal Health: Outlook for the Future
13-14th October 2009, Muséum National d'Histoire Naturelle, Paris, France



 

Pathway analysis

Conclusions:

- > **Differences** in the specific GO-terms and pathways found by the different groups.
- > Some **common biological** conclusions could still be reached for each of the differentially expressed gene lists.

 Genomics for Animal Health: Outlook for the Future
13-14th October 2009, Muséum National d'Histoire Naturelle, Paris, France


 

Pathway analysis

Main challenge:
Lack of annotation of the microarray probes.

Only **half** of the probes could be mapped and contributed to the biological interpretation of the data.

⇒ **Inference of gene networks from microarray data.**

 Genomics for Animal Health: Outlook for the Future
13-14th October 2009, Muséum National d'Histoire Naturelle, Paris, France



Gene network reconstruction

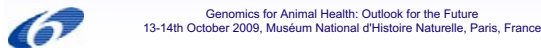
Two main approaches:

1) Bayesian networks.

- > Very computationally intensive.
- > No R package available.

2) Graphical Gaussian models.

- > Computationally efficient.
- > **R package GeneNet** (Schäfer and Strimmer, 2005).



Gene network reconstruction

Comparison study (Wehrli et al., 2006)

Both methods have **quite similar performances** for gene network reconstruction from microarray data.

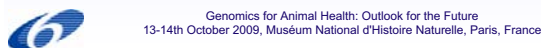
For simplicity of implementation
=> **Graphical Gaussian models.**



Graphical Gaussian models

Let **X** be the **observed data matrix** with N rows (samples) and G columns (genes).

X is supposed to be drawn from a **multivariate normal distribution** $N_G(\mu, \Sigma)$, with **mean vector** $\mu = (\mu_1, \dots, \mu_G)'$ and positive-definite **covariance matrix** $\Sigma = (\sigma_{ij})_{1 \leq i, j \leq G}$.



Graphical Gaussian models

Covariance parameters σ_{ij} can also be written as:

$$\sigma_{ij} = \rho_{ij} \sigma_i \sigma_j$$

where σ_i^2 and σ_j^2 are the **variance terms** for genes i and j, respectively and ρ_{ij} is the **Pearson correlation coefficient** between genes i and j.

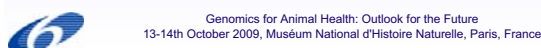


Graphical Gaussian models

Let **P** be the **Pearson correlation matrix** $P = (\rho_{ij})_{1 \leq i, j \leq G}$.

A high correlation coefficient ρ_{ij} **between two genes may indicate either:**

- 1) A **direct interaction** between genes i and j.
- 2) an **indirect interaction** between these two genes.
- 3) a **regulation** of the two genes by a common gene.



Graphical Gaussian models

For **network reconstruction** we are only interested in **direct interactions** represented by the **partial correlation matrix** $\Pi = (\pi_{ij})_{1 \leq i, j \leq G}$.

Coefficient π_{ij} **represents the correlation between two genes i and j conditioned on all the other genes.**





Graphical Gaussian models

It can be shown that the **partial correlation matrix Π** is related to the **inverse of the covariance matrix Σ** as follows:

$$\pi_{ij} = -\omega_{ij} / \sqrt{\omega_{ii}\omega_{jj}}$$

with $\Sigma^{-1}=(\omega_{ij})$, for $1 \leq i, j \leq G$.



Genomics for Animal Health: Outlook for the Future
13-14th October 2009, Muséum National d'Histoire Naturelle, Paris, France



Graphical Gaussian models

In microarray analyses, due to the problem of $N \ll G$ and small sample size N :

Shrinkage approach (Schäfer and Strimmer, 2005) to estimate the **empirical covariance matrix** and the **partial correlations**.



Genomics for Animal Health: Outlook for the Future
13-14th October 2009, Muséum National d'Histoire Naturelle, Paris, France



Graphical Gaussian models

Statistical tests to determine the **partial correlation coefficients significantly different from 0**, which correspond to **significant edges** of the graph.

Multiple test correction:

Calculation of an **edge-specific local FDR**, defined using the partial correlation coefficients.



Genomics for Animal Health: Outlook for the Future
13-14th October 2009, Muséum National d'Histoire Naturelle, Paris, France



Gene network reconstruction

Chicken infection data set.

Network reconstruction from the lists of **differentially expressed genes** between conditions **MM8** and **MM24**.

Condition MM: chickens were infected first with *E. maxima* and afterwards with the same parasite *E. maxima*.

Two time points sampled post infection: 8 and 24 hours.



Genomics for Animal Health: Outlook for the Future
13-14th October 2009, Muséum National d'Histoire Naturelle, Paris, France



Gene network reconstruction

Quite **small number of biological replicates** per condition (5 animals).

- ⇒ Network inference based only on a **few dozens of genes**.
- ⇒ 116 genes found DE between conditions MM8 and MM24 at a 1% Benjamini-Hochberg threshold.
- ⇒ No missing values allowed in *GeneNet* => **85 genes** considered.



Genomics for Animal Health: Outlook for the Future
13-14th October 2009, Muséum National d'Histoire Naturelle, Paris, France



Results

For condition MM8:

2356 significant edges among the 85 genes at a 20% local FDR threshold.

1964 significant edges at a 5% local FDR threshold.



Genomics for Animal Health: Outlook for the Future
13-14th October 2009, Muséum National d'Histoire Naturelle, Paris, France

Results

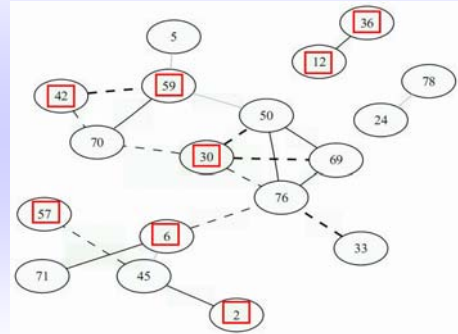
For condition MM24:

1760 significant edges among the 85 genes at a 20% local FDR threshold.

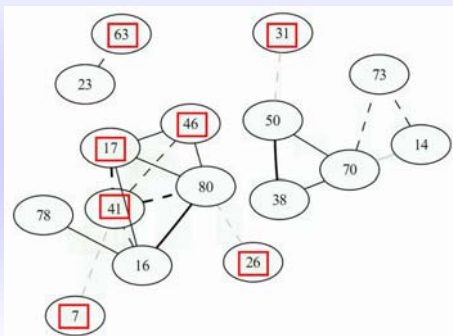
1156 significant edges at a 5% local FDR threshold.

=> Graph of the 20 most significant edges for both conditions.

Condition MM8



Condition MM24



Results

Very little overlapping between the two networks.
=> **Evolution** of the gene connections **over time**.

Lack of gene annotation.

For condition MM8: 8 annotated genes among 18.
For condition MM24: 7 annotated genes among 16.

Results

Links found here between the annotated genes were **not confirmed** with either **Ingenuity** or **Pathway Studio**.

=> **Further biological validation required** (using in vitro invalidation of genes, etc.).

Discussion

Advantages of network inference from microarray data.

- > Based on gene expression measurements.
- > **No need for annotations.**
- > Can allow to discover **new relationships** between genes not known in already existing data bases.

But need for **biological validation** of these new links.



Outlook for the future

Graphical Gaussian model based on **partial correlations**.
=> Model only **linear relationships**.

More complex model to account for **nonlinear relationships**
based on **entropy** (Hausser and Strimmer, 2009).



Genomics for Animal Health: Outlook for the Future
13-14th October 2009, Muséum National d'Histoire Naturelle, Paris, France



Outlook for the future

Here only two time points => static networks.

Several methods recently proposed for gene network reconstruction in **time-course studies** (VAR(1), dynamic bayesian networks, state-space models – R package EBDBN).



Genomics for Animal Health: Outlook for the Future
13-14th October 2009, Muséum National d'Histoire Naturelle, Paris, France



Outlook for the future

Microarray studies: Small number of biological replicates.

How to improve edge detection power and accuracy for gene network reconstruction ?

=> Combine both gene expression measurements and **prior biological knowledge**.

=> Combine **expression values from several microarray** studies in a meta-analysis.



Genomics for Animal Health: Outlook for the Future
13-14th October 2009, Muséum National d'Histoire Naturelle, Paris, France



Acknowledgements

Special thanks to DJ de Koning and Caroline Channing for all their work and efforts in organizing these two workshops.

Thank you also to the other coordinators:
Jakob Hedegaard (pathways), Christophe Klopp (re-annotation),
Magali SanCristobal (phenotypic prediction).

And to all the members of WP1.4 who actively participated.



Genomics for Animal Health: Outlook for the Future
13-14th October 2009, Muséum National d'Histoire Naturelle, Paris, France