

Overview of microarray workshop

Dirk-Jan de Koning



Tune, November 2006

EADGENE

Microarray analysis workshop

- Tune, 1-3 November 2006
- 32 participants, 13 groups, 10 countries
- Real and simulated data
- 21 presentations (12 real data)
- One very happy but confused chair



Tune, November 2006

EADGENE

Introduction

- **The Workshop Organisers (WP1.4)**
 - Dirk-Jan de Koning (RI)
 - Mogens Sandø Lund (DIAS)
 - Florence Jaffrézic (INRA)
 - Michael Watson (IAH)
 - Caroline Channing (RI)



Tune, November 2006

EADGENE

Real data

- 2 artificial infections (E. Coli and S. Aureus)
- 12 cows
 - 0, 6, 12, 24 hours p.i. E.Coli (n=4)
 - 0, 6, 12, 24 hours p.i. S. Aureus (n=4)
 - 0, 12, 72, 72 S. Aureus (n=4)
- 48 slides, reference design
- Bluefuse files + annotation



Tune, November 2006

EADGENE

Tools

- Commercial: Genespring, Spotfire
- Stats package: SAS (JUMP)
- Open Source: Bioconductor (Limma) in R
- Orange: open source
- GEPAS: web based software



Tune, November 2006

EADGENE

Many different analyses

- Within bacteria, between time points
- Between bacteria
- Across bacteria and time points fitting bacteria, time and bacteria x time interaction
- Too many to make a meaningful comparison



Tune, November 2006

EADGENE

Flavour of results

- **Most diversity in Quality Control**
 - Omit no spots
 - Omit many spots
 - Omit entire slides
 - Fit statistical weights to spots

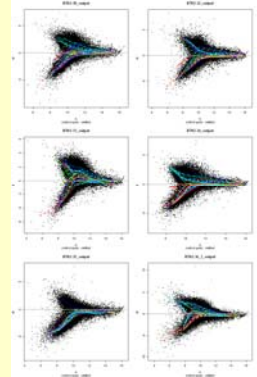


Tune, November 2006

EADGENE

Dye bias : raw data

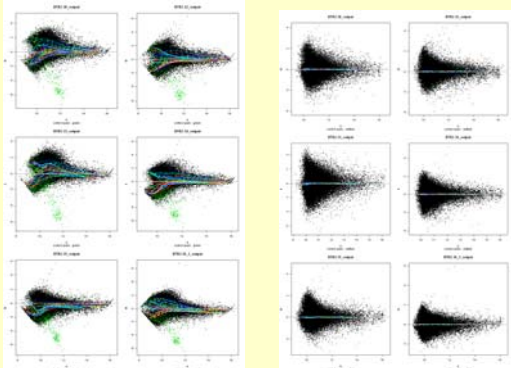
Dave
Waddington



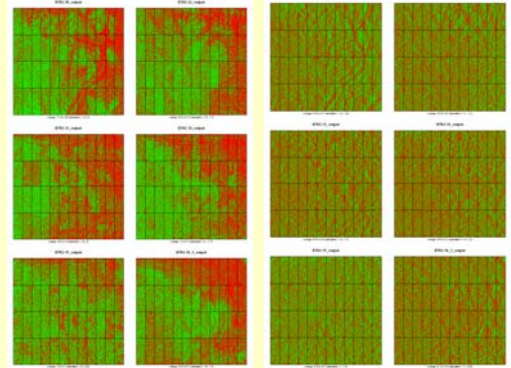
Tune, November 2006

EADGENE

Dye bias : raw + 2⁸ normed



Spatial : raw + 2⁸ normed



Some very advanced analyses

Gene profile clustering, selection of predictive genes using random forests and a stochastic algorithm, regulatory networks in a transcriptomic kinetics on bovine mastitis

K-A. Lê Cao, M. San Cristobal, G. Tosser-Klopp, C. Delmas, M. Duval, C. Robert-Granié

Kim-Anh Lê Cao

INRA and Univ. Paul Sabatier, Toulouse



Will be presented during Sabre-Eadgene days

Tune, November 2006

EADGENE

Co-expression, Mick Watson IAH

- Co-expression – **two or more genes that show similar expression patterns, signified by correlated expression profiles.**
- Differential co-expression – **two or more genes that are co-expressed in one set of experiments and not co-expressed in a second set of experiments**



Tune, November 2006

EADGENE

CoXpress: the method

- Cluster genes based on one set of experiments
- Cut tree to obtain groups of co-expressed genes
- Examine those groups in a second set of experiments
 - Are the genes in the first data set more highly correlated than could be expected by chance?
 - Method: if group has size n , resample the data set 1000 times, each time taking n random genes, calculate the correlation matrix, and summarise the pairwise correlations using the *t*-statistic
 - Compare the observed *t*-statistic with the distribution of "random" *t*-statistics
 - If correlation matrix is non-random in the *E.coli* data set, and random in the *S.aureus* data set, then the group is differentially co-expressed.



Tune, November 2006



Conclusions

- QC is still unresolved!
- Many different approaches
- Some findings very robust
- Overflow of results
 - Need extra hands
- Next workshop
 - More specific questions
 - Provide pre-analysed data for post-analyses



Tune, November 2006

