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

Annotation and post-analyses of microarray data

Background

A joint SABRE and EADGENE working group was established to evaluate existing methods and software, and propose new methods to deal with the **annotation and post-analyses of microarray data**.

The area of biology addressed by this working group was gene expression studies in livestock looking at transcriptomic differences between treatments as well as genotypes and combinations of these.

Two years ago, we organised a workshop to discuss the best approaches to analyze two-colour DNA microarray data in our area of research and the outcomes of that workshop have been published in 4 open access publications (GSE 39, 2007). While there is currently a reasonable amount of consensus on the statistical analyses of a microarray experiment (i.e. getting a gene list), the subsequently analysis of the gene list is still an area of much confusion to many scientists.

Methods

During a three-day workshop in November 2008, we discussed five aspects of these so-called post analyses of microarray data: 1) **re-annotation** of the probe set on DNA microarrays, 2) **pathway analyses** to identify significantly affected biological processes from microarray results, 3) reverse engineering of regulatory **networks** from microarray results, 4) the integration of **gene expression studies with QTL detection studies** and 5) the prediction of **phenotypic** outcomes using gene expression results.

Prior to the workshop, we distributed two sets of data to the workshop participants.

The first set of gene expression data dealt with experimental challenge of chicken with two types of *Eimeria.* This experiment is described in some detail in one of the summary papers, while the actual data is available from ArrayExpress http://www.ebi.ac.uk/microarray-as/ae/ under accession number E-MEXP-1972.

The second experiment deals with the transcriptomic effects of adrenocorticotropic hormone (ACTH) treatment in two breeds of pigs. These gene expression results are available from Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo, GSE8377 – DH06 Adrenal ACTH Sus scrofa).

Results

The **EADGENE microarray annotations** are available at: www.eadgene.info > tools and resources.

The **15 papers** arising from this working group are published in **BMC Proceedings 2009, Vol. 3 (Suppl. 4)**

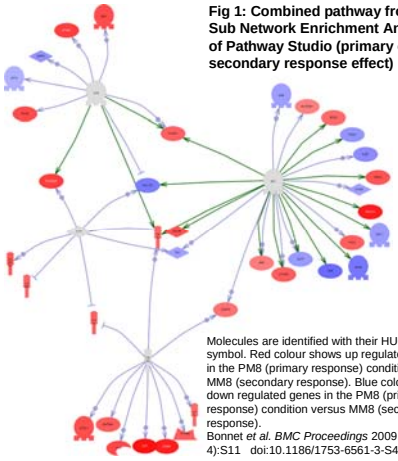


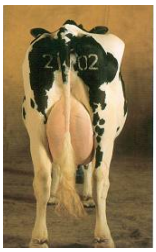
Fig 1: Combined pathway from the Sub Network Enrichment Analysis of Pathway Studio (primary or secondary response effect)

Molecules are identified with their HUGO symbol. Red colour shows up regulated genes in the PM8 (primary response) condition versus MM8 (secondary response). Blue colour shows down regulated genes in the PM8 (primary response) condition versus MM8 (secondary response).
Bonnet *et al. BMC Proceedings* 2009 **3** (Suppl 4):S11 doi:10.1186/1753-6561-3-S4-S11

Comparative transcriptomic studies of mastitis

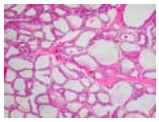
Background

The EADGENE **mastitis working group** has brought together the research activities and results from institutes located in seven European countries, working on mastitis in **cattle, goats and sheep**. Immune defence in the udder is organ specific, but is also modulated by the species of the host. The mastitis working group characterizes the genetic basis of host specific properties of udder defence, using both *in vivo* and *in vitro* techniques.

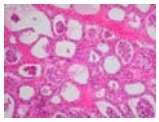


Healthy:
50 litres milk/day

Mastitis:
<30% of secretion (~no milk)



Healthy:
Mammary Epithelial Cells



Mastitis:
MEC, Granulocytes

Methods

Whole genome covering transcriptome profilings were conducted to identify key genes in udder and immune cells counteracting the attack of three pathogens: ***Escherichia coli, Staphylococcus aureus*** or ***Streptococcus uberis***. Experimental infections of cow udders, and also of immune cells from cows, were conducted with these three pathogens. The contribution of milk cells to counteract *S. aureus* infections of goat udders was also recorded. Selection lines with different genetic predisposition for suffering from mastitis were established in sheep and the selection line specific response of their milk cells towards an *S. aureus* challenge has been recorded.

The data are currently being exploited for gene mining in different ways. Biostatistical and bioinformatical tools are being used to identify host genes contributing to the pathogen-specific, and eventually udder or cell type specific immune response. The data have also been used for hypothesis driven analysis into organ specific immune mechanisms and their pathogen-specific modulation.

Table 1: Summary of the mastitis working group experiments

Infected for:		<0h	0h	1h	2h	3h	6h	8h	12h	24h	36h	48h	72h	Meta-analysis
Expt 1	a	*E. coli in vivo* in cattle					E,O		O	L,O				✓
	b	*S. aureus* 24h *in vivo* in cattle					O		O	E,O				✓
	c	*S. aureus* 72h *in vivo* in cattle							E,O				L,O	✓
Expt 2		*S. uberis in vivo* in cattle										L,O		✓
Expt 3		*S. aureus* in macrophages *in vitro* in cattle			E,O		L,O							✓
Expt 4		*S. aureus in vivo* in goats							E,O	O		L,O		✓
Expt 5		*S. aureus* in dendritic cells *in vitro* in sheep				E,O		I,O						✓
Expt 6		*E. coli in vitro* in primary Mammary Epithelial Cells (MEC) in cattle												
Expt 7		*S. aureus in vitro* in primary Mammary Epithelial Cells (MEC) in cattle												
Expt 8		*S. aureus* virulence factors *in vitro* in primary Mammary Epithelial Cells (MEC) in cattle												

E = Early time response (no signs of mastitis), L = Late time response (clear signs of mastitis), O = Used in "overall" meta-analysis, Green = data

Results: Meta-analysis

One way to integrate results obtained from different groups under different experimental conditions is to adopt a meta-analysis approach, with the goal of identifying commonalities between data sets which would not be evident by single analyses (e.g. genes commonly induced by a pathogical state). This is possible thanks to the higher statistical power and diminution of false positives when data sets are integrated by meta-analysis.

Data from the various mastitis experiments were collected, combined and re-analysed in the same way taking into account the contrasting experimental designs and sizes, host species and tissues, pathogens, arrays and time points (Table 1). *Pointillist* (Hwang et al. 2005 Proc. Natl. Acad. Sci. 102, 17302-7 and 17296-301) was chosen as meta-analysis tool in order to deal effectively with the high heterogeneity of the final dataset.

The 11 gene lists obtained (see below) were analyzed by IPA (Ingenuity Package Analysis) to infer which genes and gene pathways were modulated in different mastitis biological systems:

- Bovine specific response
- Goat specific response
- Sheep specific response
- General, overall response (bovine, goat and sheep)
- Early time response (bovine, goat and sheep)
- Late time response (bovine, goat and sheep)
- Early time specific (early but not late, bovine, goat and sheep)
- Late time specific (late but not early, bovine, goat and sheep)
- General, overall in vitro response (bovine macrophages and sheep dendritic cells (DC))
- Late time in vitro response (bovine macrophages and sheep DC)
- Early time in vitro response (bovine macrophages and sheep DC)

The first results indicate that meta-analysis was able to cope efficiently with data heterogeneity and has allowed us to identify important features, for example immune response was highlighted as the most emerging commonality between data sets, and genes with a role in signal transduction were highly represented.

Technology transfer poultry project

Background:

Technology transfer (TT) between science and industry is a particular priority for EADGENE. The TT poultry project is guided by a working group with equal numbers of representatives from science and industry. The project aims to: 1) investigate **epistasis effects on resistance to disease traits in chicken**; and 2) search for **SNP markers associated with resistance to Salmonella**.

Experiment 1

For the first part of the project, MarkerSet (Demeure and Lecerf, 2008), a tool created for SNP selection based on their informativity and location was used to select 1536 SNP covering the genome. 2380 animals with phenotypes for disease, composition or quality traits have been genotyped. The first QTL analyses are undergoing using the QTLMAP software which has been adapted to handle large number of markers. Different methods for epistasis analysis will be tested to estimate its impact on these traits. Results will then be transferred to a high quality, slow growing commercial line targeting the previously detected regions by genotyping 1000 animals (10 families) for 384 SNP.

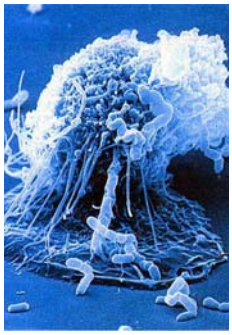
Experiment 2

For the second part of the project, 1536 SNP were chosen. 194 of them had been formerly identified within the three QTL regions (GGA1, GGA2, GGA5) known for their effect on resistance to Salmonella carrier-state. Indeed GGA1 has been confirmed in these commercial lines at the younger age, GGA2 has been confirmed in the F2 cross between the inbred N and 6 lines (Calenge et al., 2009) and it has been shown that GGA5 is close to Sal1. The other 1342 SNPs were chosen for their informativity giving a good coverage of the rest of the genome.

This SNP set will be used to genotype 650 animals measured for resistance to Salmonella carrier-state, either at a younger age (at 1 week of age, to mimic infection of broilers) or at the adult age (at the peak of lay when hens may lay contaminated eggs which are the main source of human toxi-infections). Considering both traits is of importance since they appear to negatively and moderately genetically correlated (Beaumont et al., 2009).

For further information please visit http://www.eadgene.info > Industry

Laying hen
Photo: Roslin Institute & R(D)SVS of University of Edinburgh



Salmonella typhimurium
Photo: Rocky Mountain Laboratories, NIAID, NIH

For more information about EADGENE's results, please visit www.eadgene.info > results

Co-ordinator: Marie-Hélène Pinard-van der Laan, UMR Génétique Animale et Biologie Intégrative - bât 211, Equipe Génétique, Immunité, Santé, INRA - Dépt. de Génétique Animale, F-78352 Jouy-en-Josas Cedex, France; **Email:** eadgene@dga.jouy.inra.fr; **Website:** www.eadgene.info. EADGENE is a Network of Excellence supported by funding under the 6th Research Framework Programme of the European Commission. EU Contract No. FOOD-CT-2004-506416. 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.