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

Animal Disease Genomics: Opportunities and Applications
10th - 11th June 2008, Edinburgh, UK

Meta-analysis of the EADGENE mastitis data

Sem Genini, Parco Tecnologico Padano, Italy

Starting points

- list of important livestock diseases
- people willing to collaborate
- availability of suitable data

SABRE provides additional opportunities …

- knowledge transfer agreement
- research project considering practical applications

Animal Disease Genomics: Opportunities and Applications
10th - 11th June 2008, Edinburgh, UK

Structure of the mastitis project

T
I
M
E

1. **Collection of data with contrasting**
   - **- designs and sizes**
   - **- host species (cattle, goat) and tissues**
   - **- pathogens (*E. coli*, *S. aureus*, *S. uberis*)**
   - **- arrays (Ark-Genomics, NBFGC)**
   - **- time points PI**
   - … **but analyzed in the same way!!!**
2. **Evaluation of data sets and methods**
   - **- tools: meta-analysis with pointillist**
3. **Inter- & intraspecies specific comparisons**
4. **Evaluation of results**

Animal Disease Genomics: Opportunities and Applications
10th - 11th June 2008, Edinburgh, UK

Common annotation between arrays

- **nomenclature**

ARK 17k cDNA array ≠ ARK 20k cDNA array

- **identity**

ARK cDNA arrays ≠ NBFGC bovine cDNA array

blast comparison between the sequences (threshold for similarity >100 bp without mismatches)

- total number of clones = 29,413
- clones in common = 5,772

Animal Disease Genomics: Opportunities and Applications
10th - 11th June 2008, Edinburgh, UK

Meta-analysis

- **meta-analysis: combination/integration of data across studies**
- **studies: address the same biological question**

Animal Disease Genomics: Opportunities and Applications
10th - 11th June 2008, Edinburgh, UK

Meta-analysis: benefits

- systems/ networks interactions
- combine information from different
  - sources
  - perspectives

↑ power to detect affected genes
↑ reliability of results
↓ number of false positives

Animal Disease Genomics: Opportunities and Applications
10th - 11th June 2008, Edinburgh, UK

Meta-analysis: problems

- Accuracy = f [technology]
- Bias = f [data set]
  - **abundance**
  - **scope**
  - **mode of analysis**

Pointillist

A data integration methodology for systems biology

Daehee Hwang[*], Alistair G. Rust[*], Stephen Ramsey[*], Jennifer J. Smith[*], Deena M. Leslie[*], Andrea D. Weston[*,†], Pedro de Atauri[*], John D. Aitchison[*], Leroy Hood[*,‡], Andrew F. Siegel[§], and Hamid Bolouri[*,‡]

A data integration methodology for systems biology: Experimental verification

Daehee Hwang[*,†], Jennifer J. Smith[*,†], Deena M. Leslie[*], Andrea D. Weston[*,‡], Alistair G. Rust[*], Stephen Ramsey[*], Pedro de Atauri[*], Andrew F. Siegel[§], Hamid Bolouri[*,¶], John D. Aitchison[*], and Leroy Hood[*,¶]

PNAS | **November 29, 2005** | vol. 102 | no. 48

Pointillist experimental verification

Galactose utilisation in yeast

- Integration of 18 different data types
  - transcription profiles from environment & genetic treatments
  - protein-protein interactions
  - protein-DNA interactions
  - domain-domain interactions
  - transcription-factor binding site predictions

Pointillist: evidence-weighted inferer

- compares multiple data sets to determine which elements are affected by a perturbation
- uses an algorithm to minimize the number of false positives and false negatives
- assigns weights (reliability parameter) to each data set, which represent the statistical power compared with the other data sets!

Pointillist: important steps

1. Forms C (affected, p<threshold) and Cc (non-affected) groups of elements
2. Weights data sets

Pointillist: important steps

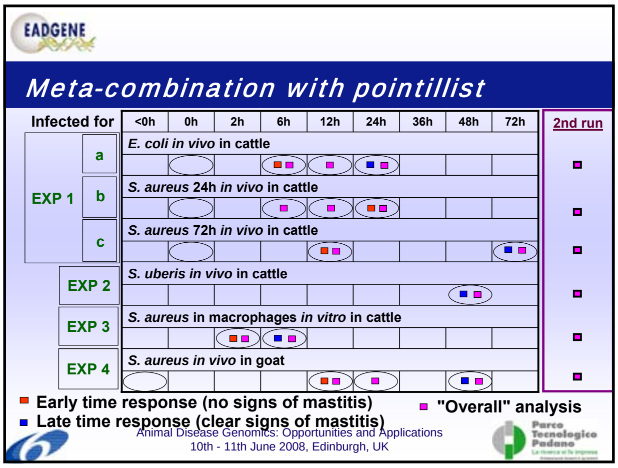
3) Transforms (3 options) the p-values using the weights (point 2)

- **Linear**: **p* = b + p x W**
- **Power**: **p* = p ^ W**
- **Uniform: p* = p ^ W, each experiment has equal weight**

p* = transformed p-value
p = untransformed p-value
b = bias term = (1-W)/ Σ W

4) Calculates combined p-values (overall significances)

POWER: Fisher's weighted F: $F_w = -2 \sum_{i=1}^{k} w_i \ln(P_i)$.

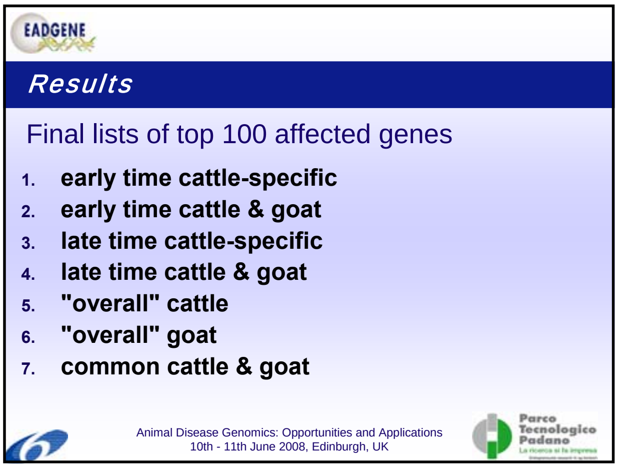


Meta-combination with pointillist

	Infected for	<0h	0h	2h	6h	12h	24h	36h	48h	72h	2nd run
EXP 1	a: *E. coli in vivo* in cattle		○		🟥 🟪	🟪	🟦 🟪				🟪
	b: *S. aureus* 24h *in vivo* in cattle		○		🟪	🟪	🟥 🟪				🟪
	c: *S. aureus* 72h *in vivo* in cattle		○			🟥 🟪				🟦 🟪	🟪
EXP 2	*S. uberis in vivo* in cattle								🟦 🟪		🟪
EXP 3	*S. aureus* in macrophages *in vitro* in cattle			🟥 🟪	🟦 🟪						🟪
EXP 4	*S. aureus in vivo* in goat	○				🟥 🟪	🟪		🟦 🟪		🟪

🟥 Early time response (no signs of mastitis)
🟦 Late time response (clear signs of mastitis)
🟪 "Overall" analysis

Animal Disease Genomics: Opportunities and Applications
10th - 11th June 2008, Edinburgh, UK



Results

Final lists of top 100 affected genes

1. **early time cattle-specific**
2. **early time cattle & goat**
3. **late time cattle-specific**
4. **late time cattle & goat**
5. **"overall" cattle**
6. **"overall" goat**
7. **common cattle & goat**

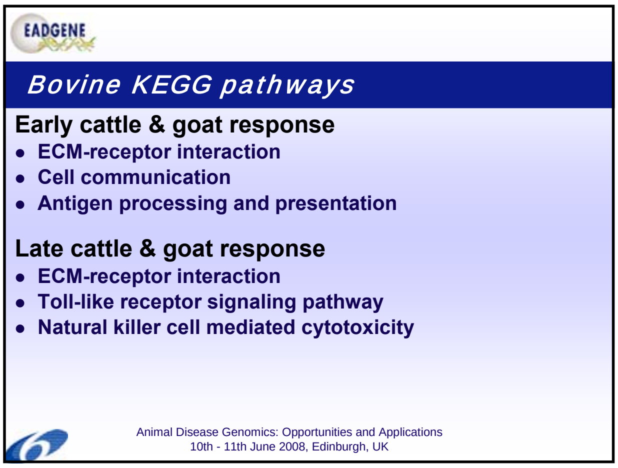
Animal Disease Genomics: Opportunities and Applications
10th - 11th June 2008, Edinburgh, UK

Bovine KEGG pathways

"Common" cattle & goat response

- **MAPK signaling pathway**
- **Jak-STAT signaling pathway**

Animal Disease Genomics: Opportunities and Applications
10th - 11th June 2008, Edinburgh, UK



Bovine KEGG pathways

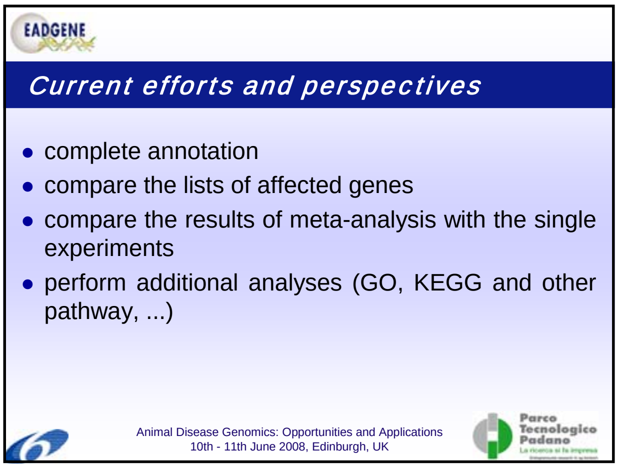
Early cattle & goat response

- **ECM-receptor interaction**
- **Cell communication**
- **Antigen processing and presentation**

Late cattle & goat response

- **ECM-receptor interaction**
- **Toll-like receptor signaling pathway**
- **Natural killer cell mediated cytotoxicity**

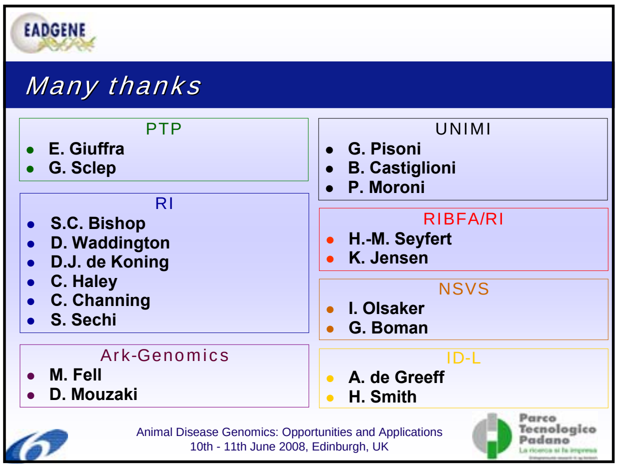
Animal Disease Genomics: Opportunities and Applications
10th - 11th June 2008, Edinburgh, UK



Current efforts and perspectives

- complete annotation
- compare the lists of affected genes
- compare the results of meta-analysis with the single experiments
- perform additional analyses (GO, KEGG and other pathway, ...)

Animal Disease Genomics: Opportunities and Applications
10th - 11th June 2008, Edinburgh, UK



Many thanks

PTP
- **E. Giuffra**
- **G. Sclep**

RI
- **S.C. Bishop**
- **D. Waddington**
- **D.J. de Koning**
- **C. Haley**
- **C. Channing**
- **S. Sechi**

Ark-Genomics
- **M. Fell**
- **D. Mouzaki**

UNIMI
- **G. Pisoni**
- **B. Castiglioni**
- **P. Moroni**

RIBFA/RI
- **H.-M. Seyfert**
- **K. Jensen**

NSVS
- **I. Olsaker**
- **G. Boman**

ID-L
- **A. de Greeff**
- **H. Smith**

Animal Disease Genomics: Opportunities and Applications
10th - 11th June 2008, Edinburgh, UK