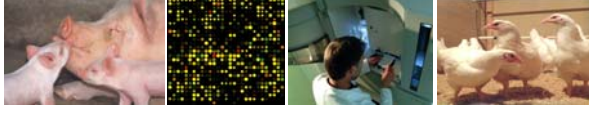


Annotation of a subset of the Arkgenomics Chicken microarray  
 Dennis Prickett and Michael Watson  
 Institute for Animal Health, UK

**EADGENE** **SABRE**  
 CUTTING EDGE GENOMICS FOR SUSTAINABLE ANIMAL BREEDING

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



Annotation of a subset of the Arkgenomics Chicken microarray  
 Dennis Prickett Michael Watson  
 Institute for Animal Health

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.

**EADGENE** **SABRE**  
 CUTTING EDGE GENOMICS FOR SUSTAINABLE ANIMAL BREEDING

**OUTLINE**

- Annotation methodology
- Results
  - ENSEMBL
  - Unigene
- Other annotation
  - Intelligent parsing of HSPs
  - Current GO gene association files from EBI

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

**EADGENE** **SABRE**  
 CUTTING EDGE GENOMICS FOR SUSTAINABLE ANIMAL BREEDING

**Methodology**

- Based on BLAST with parsing of HSPs
- Used ENSEMBL and UniGene databases
- Parameters used for the BLAST
  - $E \leq 1e-01$
  - 20% of query length at 80% identity

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

**EADGENE** **SABRE**  
 CUTTING EDGE GENOMICS FOR SUSTAINABLE ANIMAL BREEDING

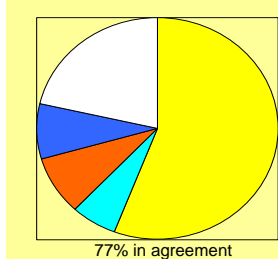
**IAH Microarray Database (IMAD)**

- MySQL database
- Perl object for accessing the database
- CGI script for web-queries
- Scripts for downloading and inserting databases
- Scripts for parsing BLAST output into database
- Scripts for extracting data in spreadsheet

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

**EADGENE** **SABRE**  
 CUTTING EDGE GENOMICS FOR SUSTAINABLE ANIMAL BREEDING

**Results - ENSEMBL comparisons**

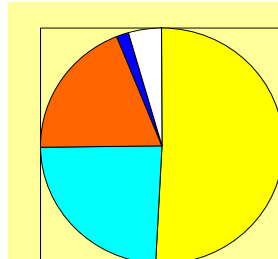


77% in agreement

- Genelds agree
- both hits-no agreement
- IMAD only
- Initial only
- neither with a hit

**EADGENE** **SABRE**  
 CUTTING EDGE GENOMICS FOR SUSTAINABLE ANIMAL BREEDING

**Results - Unigene comparisons**



55% in agreement

- Agreement
- both hits-no agreement
- IMAD hit only
- Initial only
- both no hits

**IMAD Multiple hits**

- 167 total - ENSEMBL
- 243 total - Unigene
- Possible source of some disagreements
- What to do with the data from multiple hits?
- Caveats for the end user?
- Madness?

**Resources to improve annotation**

- Intelligent parsing of HSPs
- Current Chicken GO annotation

**Intelligent Parsing of multiple HSPs**

- Checks HSPs on same strand as best HSP
- Checks that HSPs are not overlapping
- Parses HSPs to create percentage identity over the entire length of both query and hit

Multiple good HSPs, small e-value, but probably not the same gene

**Example**

- Pig oligo from Operon
  - 100% identity along entire probe with "gnl|UG|Ssc#S17500974"
  - However, not good microarray hit!

Perhaps not a good hit in terms of array hybridisation!

**Current Chicken GO associations**

- The current GO gene association files are available from the EBI at <http://www.geneontology.org/GO.current.annotations.shtml>
- From the chicken annotation community
- Updated every 3-4 weeks
- Manually transferred from humans
- Inferred by sequence or structural similarity
- Associations with ENSEMBL, UniProtKB, and RefSeq
- Annotations for 214 probes for this subset



**Chicken GOA**

Current Composition of Chicken GOA

The following table describes the latest version of GOA Chicken 38.0, released on October 17th, 2008 and assembled using the publicly released data available on the source databases on October 15th.

GO Annotation Source	Number of Annotations	Number of Binary Probes
Electronic GO annotation using UniProtKB (UniProtKB)	43214	14488
Electronic GO annotation using UniProtKB (UniProtKB)	7368	3025
Electronic GO annotation using UniProtKB (UniProtKB)	215	188
Electronic GO annotation using UniProtKB (UniProtKB)	132	108
Electronic GO annotation using UniProtKB (UniProtKB)	8745	1031
Total Electronic GO annotations	51735	15220
Manual GO annotation by UniProt	1568	174
Manual GO annotation by UniProt	279	102
Manual GO annotation by UniProt	17	7
Manual GO annotation by UniProt	17	4
Manual GO annotation by UniProt	84	18
Manual GO annotation by UniProt	83	18
Manual GO annotation by UniProt	26	25
Total Manual GO annotations	5917	182
Total Chicken GOA annotations	57652	15402
Orphan Protein References	107	
Total Protein References	1618	


Annotation of a subset of the Arkgenomics Chicken microarray  
Dennis Prickett and Michael Watson  
Institute for Animal Health, UK



CUTTING EDGE GENOMICS FOR SUSTAINABLE ANIMAL BREEDING

### How much annotation do we really need?

- Spreadsheet with 78 columns of information
- Do we really need annotation from human, rat, and mouse?
- What do we really need for the end user beyond database identifiers and gene descriptions?



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

