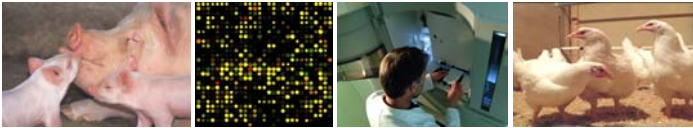




EADGENE and SABRE Post-Analyses Workshop

12-14th November 2008, Animal Sciences Group, Wageningen UR, Lelystad



Annotation Session Introduction



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.



The aims of the workshop

- Presenting the different annotation strategies and pipelines used by the teams collaborating in EADGENE.
- Comparing the pipelines on one or two data sets in order to present the impact on the annotation files provided to the biologists.
- Discussing the possible new needs, options and evolutions of these tools with other/new users.



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

2



Some general ideas about probe sets

- Microarrays use different types of probes
 - cDNA
 - Oligo-nucleotides
 - ...
- The rules to select probes are often :
 - the chosen probes should represent the transcriptome as broadly as possible or at least the area we want to study,
 - the chosen probes should be specific to a ...,
 - the chosen probes should match the technical specifications of the technology



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

3



Ideal conditions to design an oligo-set

- Complete genome sequence, with highly specific sequences for the transcripts.
- Complete transcript annotation.
- Other type of RNA annotations (like ncRNA).



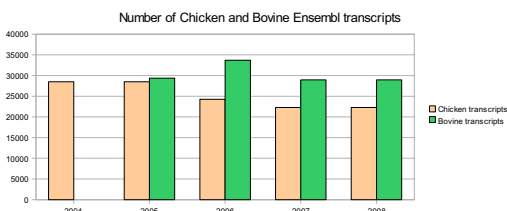
EADGENE and SABRE Post-Analyses Workshop
 12-14th Nov 2008, Animal Sciences Group, Wageningen UR, Lelystad

4



Why should we re-annotated the oligo-sets?

- Genomic information evolves (new genome assembly versions, new gene build versions,...)
- Annotation information changes (GO, orthologous genes, pathways, ...)

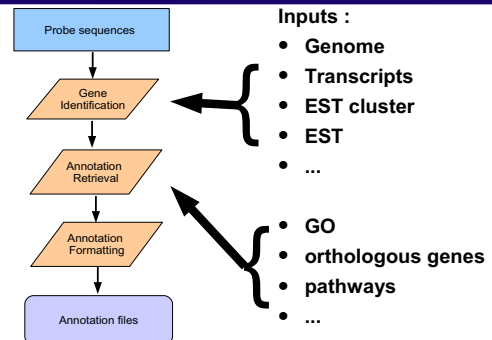


EADGENE and SABRE Post-Analyses Workshop
 12-14th Nov 2008, Animal Sciences Group, Wageningen UR, Lelystad

5



Re-annotation pipelines



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

6

How do we link an oligo to a gene

- Sequence alignment
- Kane specificity criteria :
 - >75% similarity over 50 bp for 50 mers
 - No stretch of more than 15 contiguous bp

#552-#557 Nucleic Acids Research, 2000, Vol. 28, No. 22 © 2000 Oxford University Press
Assessment of the sensitivity and specificity of oligonucleotide (50mer) microarrays
 Michael D. Kane, Timothy A. Jatkoe, Craig R. Stumpf, Jia Lu¹, Jeffrey D. Thomas and Steven J. Madore^{*}
 Department of Molecular Biology and Genomics and ¹Department of Infectious Diseases, Pfizer Global Research and Development, Ann Arbor, MI 48105, USA
 Received July 5, 2000; Revised and Accepted September 21, 2000

What is a probe set for us?

cDNA Oligo-nucleotide

Where do we get the information from?

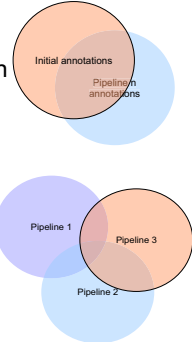


What are the biologists looking for?

- Information :**
- Chromosomal position (eQTL)
 - Orthologous genes (access to other software)
 - Gene Ontology
 - KEGG pathways
 - ...
- Formats :**
- MS-Excel
 - Tab delimited
 - ...

Annotation results comparison

The pipelines are build using different strategies, data sources, technical architectures and therefore the results can be different.



The aim is to find and try to explain the differences and select the best possible method for each case.

This is done on a limited dataset.

The comparison data-sets

• **Chicken oligos :**

Species	Origin	Nb oligos	Design year	Number of slides
Chicken	ARK-Genomics	20,460	2005	RI 445 / INRA 200
Cattle	BOMC / ARK Genomics	24,000	2003	RI 85
Pig	DIAS	25 210	2003	0

- Subset of the oligo-set : 791
- Up and down regulated from the "MM8_MM24.txt" file

• **Pig cDNA :**

- cDNA present in the "Supplemental_data_Table1.xls" file,
- Subset of 241 probes



Chicken oligo probe-set

The chicken oligos were designed against a mixed panel of

- ESTs from Genbank/EMBL,
- Ensembl release 30 genes and transcripts,
- UMIST chicken ChEST cDNAs,
- miRBase RNAs
- and contributed sequence.



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

13



Pig cDNA probe-set

- Multi tissue cDNA library
- 24 449 high-quality sequences
- Positioned in 7 749 different contigs
- 9 216 selected clones (partly from singlets)

BMC Genomics



Research article

A pig multi-tissue normalised cDNA library: large-scale sequencing, cluster analysis and 9K micro-array resource generation
Agnes Bonnet¹, Eddie Iannuccelli², Karine Hugot^{3,4}, Francis Benne¹, Maria F Bonaldo⁵, Marcelo B Soares⁵, François Hatey¹ and Gwenola Tosser-Klopp*¹

Open Access



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

14



The presentations

- Pieter Neerinx (Wageningen University)
[OligoRAP: A dual dual dual strategy for optimal oligo annotation](#)
- Dennis Prickett (IAH)
[Annotation results](#)
- Pierrot Casel (INRA)
[Annotation results](#)



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

15