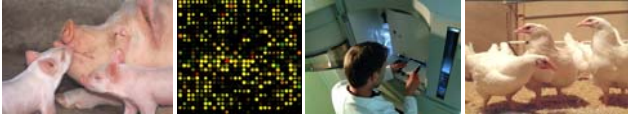




EADGENE and SABRE Post-Analyses Workshop

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



Array data mining using Bioconductor

Haisheng Nie et al



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.



Introduction: data

- Immune reactions after a homo- or heterologous challenge of broilers primed with *Eimeria maxima*
- Differentially expressed gene list:
 - ❖ Three differentially expressed gene lists
 - ❖ Cut-off: adjusted p values ≤ 0.05

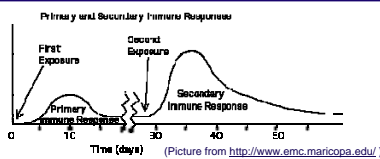
| Contrast | UP | DOWN | total DE |
|----------|-----|------|----------|
| MM8.PM8 | 923 | 803 | 1726 |
| MM8.MA8 | 23 | 58 | 81 |
| MM8.MM24 | 152 | 639 | 791 |



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



Introduction: three contrasts



- MM8.PM8: differences between primary and secondary response?
- MM8.MA8: will E.ac trigger a secondary response directly?
whether the secondary response to E.ac is the same to E.max?
- MM8.MM24: what changes during time for secondary response?



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



Introduction: questions need to be answered

- What is the biological information in those six gene lists (3 up and 3 down) by looking at:
 - i. Gene Ontology (GO)
 - ii. Biological pathways (KEGG)
 - iii. Others
- How can we extract those information?



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



Introduction: Bioconductor packages

There are several R-bioconductor packages

- i. GOstats (Tools for manipulating GO and microarrays)
- ii. topGO (Enrichment analysis for Gene Ontology)
- iii. sigPathways (pathway analysis)
- iv. other packages...



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



GOstats

➤ GOstats

A set of tools for interacting with GO and microarray data.

➤ Hypergeometric testing

- ❖ Are there any GO terms that have a larger than expected subset of our selected genes in their annotation list?
- ❖ The common test is for **over-representation**, but one can test for **under-representation**

➤ Array-based annotation R package

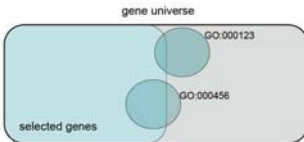


EADGENE and SABRE Post-Analyses Workshop
12-14th Nov 2008, Animal Sciences Group, Wageningen UR, Lelystad

GOstats

(example from Seth Falcon 2007)

Testing a GO Term



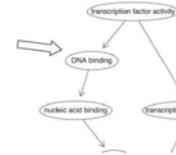
| | Selected (white) | Not (black) |
|-----------|------------------|-------------|
| In GO | n11 | n12 |
| Not in GO | n21 | n22 |

Gene Universe: for example, all the genes on the array

GOstats

(example from Seth Falcon 2007)

➤ The Hypergeometric test assumes independence of categories, but...



➤ We only want to call a GO term significant if there is evidence beyond that provided by its significant children

GOstats

(example from Seth Falcon 2007)

Conditional Hypergeometric testing

test a term conditional on its children, remove genes inherited from significant children before testing.

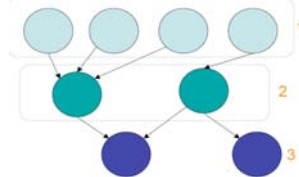


GOstats

(example from Seth Falcon 2007)

A Conditional Algorithm

1. Walk leaves of the GO DAG, compute Hypergeometric as usual.
2. When computing the next level, remove genes from significant children.



For further details:

Falcon and Gentleman, Bioinformatics, 2007

Alexa et al, Bioinformatics, 2007

AnnotationDbi

AnnotationDbi:

Create maps that allow access from R to underlying annotation databases.

- It creates standard array annotation packages used in other bioconductor packages, e.g. GOstats
- Using human orthologs (INRA re-annotation) to create a kind of "human-chip" annotation package for chicken oligos

Results1: MM8_PM8.UP up in MM8

| GOBPID | Pvalue | OddsRatio | ExpCount | Count | Size | ONLY immune-related significant terms are shown (p<0.05) |
|------------|---------|-----------|----------|-------|------|--|
| GO:0051251 | 0.00793 | 4.66511 | 1.28630 | 5 | 25 | positive regulation of lymphocyte activation |
| GO:0045577 | 0.01478 | 18.54321 | 0.20581 | 2 | 4 | regulation of B cell differentiation |
| GO:0042102 | 0.01635 | 6.96827 | 0.56597 | 3 | 11 | positive regulation of T cell proliferation |
| GO:0002694 | 0.01736 | 3.20089 | 2.10953 | 6 | 41 | regulation of leukocyte activation |
| GO:0017157 | 0.02380 | 12.36008 | 0.25726 | 2 | 5 | regulation of exocytosis |
| GO:0045058 | 0.02380 | 12.36008 | 0.25726 | 2 | 5 | T cell selection |
| GO:0050863 | 0.02533 | 3.32777 | 1.69792 | 5 | 33 | regulation of T cell activation |
| GO:0032946 | 0.02627 | 5.57276 | 0.66888 | 3 | 13 | positive regulation of mononuclear cell proliferation |
| GO:0045582 | 0.03449 | 9.26852 | 0.30871 | 2 | 6 | positive regulation of T cell differentiation |
| GO:0046651 | 0.03699 | 3.54274 | 1.28630 | 4 | 25 | lymphocyte proliferation |
| GO:0045072 | 0.04667 | 7.41358 | 0.36016 | 2 | 7 | regulation of Interferon-gamma biosynthetic process |



Results1: MM8_PM8.DOWN up in PM8

| GOBPID | Pvalue | OddsRatio | ExpCount | Count | Size | ONLY immune-related significant terms are shown (p<0.05) |
|------------|---------|-----------|----------|-------|------|--|
| GO:0050860 | 0.00197 | Inf | 0.08902 | 2 | 2 | negative regulation of T cell receptor signaling pathway |
| GO:0002429 | 0.01423 | 7.22222 | 0.53409 | 3 | 12 | immune response-activating cell surface receptor signaling pathway |
| GO:0002764 | 0.01790 | 6.49892 | 0.57860 | 3 | 13 | immune response-regulating signal transduction |
| GO:0050854 | 0.01805 | 14.40714 | 0.22254 | 2 | 5 | regulation of antigen receptor-mediated signaling pathway |
| GO:0051603 | 0.02544 | 2.07781 | 5.65246 | 11 | 127 | proteolysis involved in cellular protein catabolic process |
| GO:0048041 | 0.02629 | 10.80357 | 0.26705 | 2 | 6 | focal adhesion formation |
| GO:0002683 | 0.04337 | 4.32903 | 0.90114 | 3 | 18 | negative regulation of immune system process |
| GO:0006110 | 0.04627 | 7.20000 | 0.35606 | 2 | 8 | regulation of glycolysis |
| GO:0030048 | 0.04627 | 7.20000 | 0.35606 | 2 | 8 | actin filament-based movement |



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



Preliminary Conclusion 1: MM8_PM8

- Contrast 1
 - 1) PBS-E.max treatment mainly triggers primary immune response:
 - more general immune-related GO BP terms
 - 2) E.max-E.max treatment mainly triggers secondary immune response
 - more GO terms related to T & B cell regulation



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



Results2: MM8_MA8.UP up in MM8

| GOBPID | Pvalue | OddsRatio | ExpCount | Count | Size | All significant terms (p<0.05) |
|------------|---------|-----------|----------|-------|------|---|
| GO:0006996 | 0.00904 | 6.71921 | 0.90751 | 4 | 575 | organelle organization and biogenesis |
| GO:0051056 | 0.01519 | 12.82025 | 0.19413 | 2 | 123 | regulation of small GTPase mediated signal transduction |
| GO:0007265 | 0.01916 | 11.29380 | 0.21938 | 2 | 139 | Ras protein signal transduction |
| GO:0010605 | 0.03402 | 8.20722 | 0.29830 | 2 | 189 | negative regulation of macromolecule metabolic process |
| GO:0031324 | 0.04057 | 7.42718 | 0.32828 | 2 | 208 | negative regulation of cellular metabolic process |



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



Results2: MM8_MA8.DOWN up in MA8

| GOBPID | Pvalue | OddsRatio | ExpCount | Count | Size | All significant terms (p<0.05) |
|------------|---------|-----------|----------|-------|------|--|
| GO:0006817 | 0.00720 | 17.86060 | 0.12926 | 2 | 39 | phosphate transport |
| GO:0051239 | 0.00769 | 8.46038 | 0.41430 | 3 | 125 | regulation of multicellular organismal process |
| GO:0007626 | 0.02387 | 9.25723 | 0.24195 | 2 | 73 | locomotory behavior |
| GO:0006915 | 0.03526 | 3.66466 | 1.27604 | 4 | 385 | apoptosis |
| GO:0006820 | 0.04497 | 6.47629 | 0.34138 | 2 | 103 | anion transport |
| GO:0008219 | 0.04566 | 3.35379 | 1.38542 | 4 | 418 | cell death |



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



Preliminary Conclusion 2: MM8_MA8

➤ Contrast 2

E.max-E.max and E.max-E.ac treatments do not show big differences on triggering secondary immune responses.



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



Results3: MM8_MM24.UP up in MM8

| GOBPID | Pvalue | OddsRatio | ExpCount | Count | Size | ONLY immune-related significant terms are shown (p<0.05) |
|------------|---------|-----------|----------|-------|------|--|
| GO:0051092 | 0.00602 | 20.34740 | 0.11900 | 2 | 13 | positive regulation of NF-kappaB transcription factor activity |
| GO:0043388 | 0.01274 | 13.15336 | 0.17393 | 2 | 19 | positive regulation of DNA binding |
| GO:0051090 | 0.01998 | 10.15584 | 0.21970 | 2 | 24 | regulation of transcription factor activity |
| GO:0005996 | 0.03304 | 4.57297 | 0.70486 | 3 | 77 | monosaccharide metabolic process |
| GO:0031145 | 0.04253 | 6.55882 | 0.32955 | 2 | 36 | anaphase-promoting complex-dependent proteasomal ubiquitin-dependent protein catabolic process |
| GO:0045596 | 0.04470 | 6.37041 | 0.33870 | 2 | 37 | negative regulation of cell differentiation |
| GO:0051352 | 0.04470 | 6.37041 | 0.33870 | 2 | 37 | negative regulation of ligase activity |
| GO:0051436 | 0.04470 | 6.37041 | 0.33870 | 2 | 37 | negative regulation of ubiquitin-protein ligase activity during mitotic cell cycle |
| GO:0051098 | 0.04691 | 6.19246 | 0.34785 | 2 | 38 | regulation of binding |
| GO:0051437 | 0.04917 | 6.02413 | 0.35701 | 2 | 39 | positive regulation of ubiquitin-protein ligase activity during mitotic cell cycle |



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

Results: MM8_MM24.DOWN up in MM24

| GOBPID | Pvalue | OddsRatio | ExpCount | Count | Size | ONLY immune-related significant terms are shown (p<0.05) |
|------------|---------|-----------|----------|-------|------|--|
| GO:0030334 | 0.00480 | 4.25888 | 1.59722 | 6 | 46 | regulation of cell migration |
| GO:0006310 | 0.00657 | 3.95979 | 1.70139 | 6 | 49 | DNA recombination |
| GO:0006928 | 0.00914 | 2.25103 | 6.21528 | 13 | 179 | cell motility |
| GO:0050918 | 0.01120 | 18.69419 | 0.17361 | 2 | 5 | positive chemotaxis |
| GO:0046457 | 0.01642 | 14.01835 | 0.20833 | 2 | 6 | prostanoid biosynthetic process |
| GO:0006909 | 0.01655 | 6.49025 | 0.55556 | 3 | 16 | phagocytosis |
| GO:0007257 | 0.02246 | 11.21284 | 0.24306 | 2 | 7 | activation of JNK activity |
| GO:0017156 | 0.02927 | 9.34251 | 0.27778 | 2 | 8 | calcium ion-dependent exocytosis |
| GO:0008286 | 0.03044 | 4.95888 | 0.69444 | 3 | 20 | insulin receptor signaling pathway |
| GO:0032623 | 0.03678 | 8.00655 | 0.31250 | 2 | 9 | interleukin-2 production |
| GO:0043542 | 0.03678 | 8.00655 | 0.31250 | 2 | 9 | endothelial cell migration |
| GO:0050852 | 0.03678 | 8.00655 | 0.31250 | 2 | 9 | T cell receptor signaling pathway |
| GO:0006693 | 0.04495 | 7.00459 | 0.34722 | 2 | 10 | prostaglandin metabolic process |
| GO:0007155 | 0.04842 | 1.62351 | 11.00694 | 17 | 317 | cell adhesion |

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

Preliminary Conclusion 3: MM8_MM24

➤ Contrast 3

E.max-E.max treatment activates secondary immune response, and immune activities are increasing from 8h to 24h.



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

Comparison between two re-annotations

➤ Re-annotation sources:

- oligoRAP re-annotation (WUR)
- Sigenae re-annotation (INRA)
- IAH re-annotation (IAH)

➤ Gene Ontology:

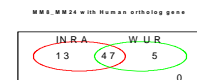
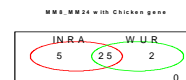
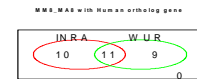
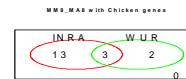
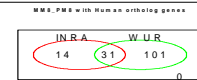
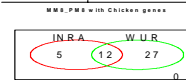
- using GO info directly linked to chicken genes
- linking chicken oligos to GO info via human orthologs genes



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

Comparison between two re-annotations

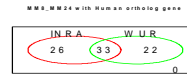
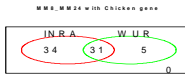
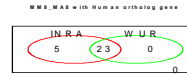
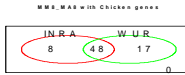
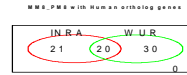
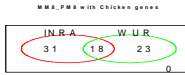
| Contrast | UP-regulation | INRA_Chick | INRA_Human | WUR_Chick | WUR_Human |
|----------|---------------|------------|------------|-----------|-----------|
| MM8_PM8 | MM8 | 17 | 45 | 39 | 132 |
| MM8_MAS | MM8 | 16 | 21 | 5 | 20 |
| MM8_MM24 | MM8 | 30 | 60 | 27 | 52 |



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

Comparison between two re-annotations

| Contrast | UP-regulation | INRA_Chick | INRA_Human | WUR_Chick | WUR_Human |
|----------|---------------|------------|------------|-----------|-----------|
| MM8_PM8 | PM8 | 49 | 41 | 41 | 50 |
| MM8_MAS | MAS | 56 | 28 | 65 | 23 |
| MM8_MM24 | MM24 | 65 | 59 | 36 | 55 |



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

Gene Universe: what to choose?



➤ In this case, the gene universe differs by using different re-annotation sources (INRA & WUR)

➤ The choice of gene universe has a big impact on Hypergeometric test results.

➤ A hybrid re-annotation pipeline by combining three different methods will help?



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



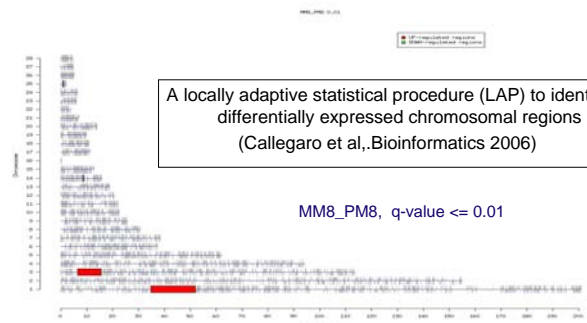
Overlapped GO terms: MM8_PM8.up

| GOBPID | Count.x | Size.x | Count.y | Size.y | ONLY immune-related |
|------------|---------|--------|---------|--------|---|
| | | | | | significant terms |
| GO:0002694 | 6 | 41 | 7 | 45 | regulation of leukocyte activation |
| GO:0032946 | 3 | 13 | 4 | 17 | positive regulation of mononuclear cell proliferation |
| GO:0042102 | 3 | 11 | 4 | 15 | positive regulation of T cell proliferation |
| GO:0045072 | 2 | 7 | 2 | 8 | regulation of interferon-gamma biosynthetic process |
| GO:0045577 | 2 | 4 | 2 | 3 | regulation of B cell differentiation |
| GO:0050867 | 5 | 26 | 6 | 31 | positive regulation of cell activation |
| GO:0051251 | 5 | 25 | 6 | 30 | positive regulation of lymphocyte activation |

Did we worry about slightly different
re-annotation methods too much ??



Extra: Differentially expressed chr locations



Which genes are there?

| UP-regulated regions | | | | |
|------------------------|-----|----------|----------|------------------|
| Cluster number | chr | start | end | genes in cluster |
| Cluster1: | 1 | 34739763 | 51993116 | 203 |
| Cluster1: | 3 | 6458897 | 15588694 | 84 |
| Cluster1: | 14 | 8506101 | 8626211 | 5 |
| DOWN-regulated regions | | | | |
| Cluster number | chr | start | end | genes in cluster |
| Cluster1: | 25 | 803105 | 904932 | 9 |



Other authors of this work

- Pathway analysis
Pieter Neerincx and Martien Groenen
(Wageningen University, Netherlands)
- Differentially expressed Chr regions (LAP)
Francesco Ferrari and Silvio Bicciato
(University of Padova, Italy)



Acknowledgment

- EADGENE, SABRE, ASG-Lelystad
- Caroline Channing and Sandrine Ayuso
- Annemarie Rebel and colleagues
- Jan van der Poel

