



HAL
open science

The EADGENE and SABRE post-analyses workshop

Florence Jaffrezic, Jakob Hedegaard, Magali San Cristobal, Christophe C. Klopp, Dirk-Jan de Koning

► **To cite this version:**

Florence Jaffrezic, Jakob Hedegaard, Magali San Cristobal, Christophe C. Klopp, Dirk-Jan de Koning. The EADGENE and SABRE post-analyses workshop. BMC Proceedings, 2009, 3, online (suppl. 4), Non paginé. 10.1186/1753-6561-3-S4-I1 . hal-01193449

HAL Id: hal-01193449

<https://hal.science/hal-01193449>

Submitted on 30 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Introduction

The EADGENE and SABRE post-analyses workshop

Florence Jaffrezic¹, Jakob Hedegaard², Magali SanCristobal³,
Christophe Klopp⁴ and Dirk-Jan de Koning*⁵

Address: ¹INRA AgroParisTech, Animal Genetics and Integrative Biology, Populations Statistics Genomes, 78350 Jouy-en-Josas, France, ²Aarhus University, Faculty of Agricultural Sciences, Department of Genetics and Biotechnology, P.O. Box 50 DK-8830 Tjele, Denmark, ³INRA, UMR444 Laboratoire de Génétique Cellulaire, F-31326 Castanet-Tolosan, France, ⁴Sigenae UR875 Biométrie et Intelligence Artificielle, INRA, BP 52627, 31326 Castanet-Tolosan Cedex, France and ⁵Roslin Institute and R(D)SVS, University of Edinburgh, Roslin, EH25 9PS, UK

Email: Florence Jaffrezic - florence.jaffrezic@jouy.inra.fr; Jakob Hedegaard - Jakob.Hedegaard@agrsci.dk; Magali SanCristobal - magali.sancristobal@toulouse.inra.fr; Christophe Klopp - christophe.klopp@toulouse.inra.fr; Dirk-Jan de Koning* - DJ.deKoning@roslin.ed.ac.uk

* Corresponding author

from EADGENE and SABRE Post-analyses Workshop
Lelystad, The Netherlands. 12–14 November 2008

Published: 16 July 2009

BMC Proceedings 2009, 3(Suppl 4):11 doi:10.1186/1753-6561-3-S4-11

This article is available from: <http://www.biomedcentral.com/1753-6561/3/S4/11>

© 2009 Jaffrezic et al; licensee BioMed Central Ltd.

This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Background

Analysis of genome-wide gene expression using DNA microarrays has become pervasive in almost all areas of biology. The area of biology addressed by this workshop is gene expression studies in livestock looking at transcriptional differences between treatments as well as genotypes and combinations of these. Two years ago, we organized 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 [1-4]. 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.

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 deals with experimental challenge of chicken with two types of *Eimeria*. This experiment is described in some detail in one of the summary papers [5], 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 adrenocorticotrophic 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).

Observations

Re-annotation of microarray probe set

Up-to-date annotation and target specificity is essential for functional analysis of microarray data. Three annotation pipelines were used to re-annotate 791 selected probes from the chicken microarray [6-8] and subsequently compared [9]. The main difference between annotation pipelines came from differences between the thresholds that were applied in order to link a probe to a certain type of annotation. It was recommended to have flexible thresholds in order to evaluate the effect of strin-

gency and strike the right balance between reliability and coverage of the annotation.

The application of pathway analyses

Several conceptually different analytical approaches, using both commercial and public available software, were applied by the participating groups to interpret the affected probes from the chicken experiment [10-15]. A total of twelve pathway related software tools were tested on the chicken data. The main focus of the approaches was to utilise the relation between probes/genes and their gene ontology and pathways to interpret the affected probes/genes. The lack of a well annotated chicken genome did limit the possibilities to fully explore the tools. The main results from these analyses showed that the biological interpretation is highly dependent on the statistical method used but that some common biological conclusions could be reached [5].

Reverse engineering of regulatory networks

Graphical Gaussian models, as implemented in the R library GeneNet, were applied to 85 gene transcripts from the chicken experiment that were selected for their significance and lack of missing data. While a large number of significant relationships (edges) were found between these 85 genes, they could not be confirmed using pathway analyses because of limited annotation [16].

Integration of microarrays with QTL results

Using the pig experiment, three groups evaluated different ways to link the gene expression results to QTL results: 1) co-location between differentially expressed genes and QTL results from the same experiments [17,18], 2) co-location between differentially expressed genes and QTL from the public domain, and 3) overlap between genes and QTL regions at the Pathway level: genes and QTL may not co-locate but differentially expressed genes have enriched pathways with genes in the QTL region [19]. Because the pig has only a preliminary draft genome sequence, comparative mapping approaches were also used to compare QTL locations and differentially expressed genes. Because of very limited annotations, no meaningful pathway comparisons could be made.

Phenotypic prediction from microarray data

The pig data has two treatments and two genotypes. In order to predict these grouping using the microarray data the authors used a Random Forest approach and also compared the classical Partial Least Squares regression (PLS) with a novel approach called sparse PLS [20]. All methods performed well on this data set. The sparse PLS outperformed the PLS in terms of prediction performance and improved the interpretability of the results. Both approaches are well adapted to transcriptomic data where the number of features is much greater than the number

of individuals. Only a small number of genes (<20) was required to give perfect prediction of the four groups.

Take home message

The central theme of the meeting was the lack of annotation. This was not in terms of bioinformatics tools to link sequences between species but a clear lack of knowledge regarding gene function. This was not specific for livestock species and considerable efforts are required before pathway based approaches will really come to fruition. In this context, there is a clear benefit for methods that do not require any level of annotation such as reverse engineering of networks and phenotypic prediction from microarray data. One challenging opportunity is to catalogue this level of experimental annotation (e.g. 'up-regulated after infection with *Eimeria*') as an alternative means to derive functional links over time.

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

The authors gratefully acknowledge the local workshop organisers in Lelystad as well as crucial coordination by Caroline Channing. The authors acknowledge the EC-funded Integrated Project SABRE (EC contract number FOOD-CT-2006-01625) and the EC-funded Network of Excellence EADGENE (EC contract number FOOD-CT-2004-506416) for supporting the workshop and publication of this manuscript.

This article has been published as part of *BMC Proceedings* Volume 3 Supplement 4, 2009: EADGENE and SABRE Post-analyses Workshop. The full contents of the supplement are available online at <http://www.biomedcentral.com/1753-6561/3?issue=S4>.

References

1. de Koning DJ, Jaffrezic F, Lund MS, Watson M, Channing C, Hulsege I, Pool MH, Buitenhuis B, Hedegaard J, Hornshoj H, et al.: **The EADGENE microarray data analysis workshop (open access publication)**. *Genet Sel Evol* 2007, **39**:621-631.
2. Jaffrezic F, de Koning DJ, Boettcher PJ, Bonnet A, Buitenhuis B, Closset R, Dejean S, Delmas C, Detilleux JC, Dovc P, et al.: **Analysis of the real EADGENE data set: Comparison of methods and guidelines for data normalisation and selection of differentially expressed genes (Open Access publication)**. *Genet Sel Evol* 2007, **39**:633-650.
3. Sorensen P, Bonnet A, Buitenhuis B, Closset R, Déjean S, Delmas C, Duval M, Glass L, Hedegaard J, Hornshoj H, et al.: **Analysis of the real EADGENE data set: Multivariate approaches and post analysis (Open Access publication)**. *Genet Sel Evol* 2007, **39**:651-668.
4. Watson M, Alegre MP, Baron MD, Delmas C, Dovc P, Duval M, Foulley JL, Pavon JG, Hulsege I, Jaffrezic F, et al.: **Analysis of a simulated microarray dataset: Comparison of methods for data normalisation and detection of differential expression (Open Access publication)**. *Genet Sel Evol* 2007, **39**:669-683.
5. Hedegaard J, Arce C, Bicciato S, Bonnet A, Ramerez-Boo M, Buitenhuis AJ, Collado-Romero M, Conley LN, SanCristobal M, Ferrari F, et al.: **Methods for interpreting lists of affected genes obtained in a DNA microarray experiment**. *BMC Proceedings* 2009, **3(Suppl 4)**:S5.
6. Casel P, Moreews F, Lagarrigue S, Klopp C: **sigReannot: an oligo-set re-annotation pipeline based on similarities with the Ensembl transcripts and Unigene clusters**. *BMC Proceedings* 2009, **3(Suppl 4)**:S3.

7. Neerincx PBT, Rauwerda H, Nie H, Groenen MAM, Breit TM, Leunissen JAM: **OligoRAP – An Oligo Re-Annotation Pipeline to improve annotation and estimate target specificity.** *BMC Proceedings* 2009, **3(Suppl 4):S4.**
8. Prickett D, Watson M: **IMAD: Flexible annotation of microarray sequences.** *BMC Proceedings* 2009, **3(Suppl 4):S2.**
9. Neerincx PBT, Casel P, Prickett D, Nie H, Watson M, Leunissen JAM, Groenen MAM, Klopp C: **Comparison of three Microarray Probe Annotation Pipelines: Differences in Strategies and their Effect on Downstream Analysis.** *BMC Proceedings* 2009, **3(Suppl 4):S1.**
10. Bonnet A, Lagarrigue S, Liaubet L, Robert-Granié C, SanCristobal M, Tosser-Klopp G: **Pathway results from the chicken data set using GOTM, Pathway Studio and Ingenuity software.** *BMC Proceedings* 2009, **3(Suppl 4):S11.**
11. Jimenez-Marin A, Collado-Romero M, Ramerez-Boo M, Arce-Jimenez C, Garrido JJ: **Biological pathway analysis by ArrayUnlock and Ingenuity Pathway Analysis.** *BMC Proceedings* 2009, **3(Suppl 4):S6.**
12. Prickett D, Watson M: **Use of GenMAPP and MAPPFinder to analyse pathways involved in chickens infected with the protozoan parasite Eimeria.** *BMC Proceedings* 2009, **3(Suppl 4):S7.**
13. Nie H, Neerincx PBT, Poel JJ van der, Ferrari F, Biccato S, Leunissen JAM, Groenen MAM: **Microarray data mining using Bioconductor packages.** *BMC Proceedings* 2009, **3(Suppl 4):S9.**
14. Hulsege IB, Kommadath A, Smits MA: **Globaltest and GOEAST: Two different approaches for Gene Ontology analysis.** *BMC Proceedings* 2009, **3(Suppl 4):S10.**
15. Skarman A, Jiang L, Hornshoj H, Buitenhuis AJ, Hedegaard J, Conley LN, Sorensen P: **Gene set analysis methods applied to chicken microarray expression data.** *BMC Proceedings* 2009, **3(Suppl 4):S8.**
16. Jaffrezic F, Tosser-Klopp G: **Gene network reconstruction from microarray data.** *BMC Proceedings* 2009, **3(Suppl 4):S12.**
17. Désautès C, Bidanel JP, Milant D, Iannuccelli N, Amigues Y, Bourgeois F, Caritez JC, Renard C, Chevalet C, Mormède P: **Genetic linkage mapping of quantitative trait loci for behavioral and neuroendocrine stress response traits in pigs.** *J Anim Sci* 2002, **80:2276-2285.**
18. Hazard D, Liaubet L, Sancristobal M, Mormède P: **Gene array and real time PCR analysis of the adrenal sensitivity to adrenocorticotrophic hormone in pig.** *BMC Genomics* 2008, **9:101.**
19. Jouffe V, Rowe SJ, Liaubet L, Buitenhuis AJ, Hornshoj H, SanCristobal M, Mormède P, de Koning DJ: **Using microarrays to identify positional candidate genes for QTL: the case study of ACTH response in pigs.** *BMC Proceedings* 2009, **3(Suppl 4):S14.**
20. Robert-Granié C, Le Cao K-A, SanCristobal M: **Predicting qualitative phenotypes from microarray data – the Eadgene pig data set.** *BMC Proceedings* 2009, **3(Suppl 4):S13.**

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:

http://www.biomedcentral.com/info/publishing_adv.asp

