Arch. Tierz., Dummerstorf **50** (2007) Special Issue, 27-31

Animal Breeding and Genetics Centre, Animal Sciences Group of Wageningen University and Research Centre, Lelystad, The Netherlands

MARINUS F.W. TE PAS, INA HULSEGGE and MARCO H. POOL

Pathway analysis: Combining microarray data and physiological data to study myogenesis

Abstract

Microarray experiments investigate the changes in the expression of the transcriptome of a tissue during biological processes such as development of the tissue. Analysis usually produces a list of up and down regulated genes. While this in itself may highlight important biological processes taking place much information about relations between the genes may remain hidden in the huge amount of data. Databases can be searched via the internet might contain physiological information that can be invaluable for the understanding of the microarray results. However, the databases contain more information on humans and model animals such as mice and rats than other species. Therefore, we developed a set of PERL scripts enabling the automated search of the database for pathways using the list of expressed genes present on the microarray. A previously reported microarray experiment investigating prenatal myogenesis in pigs was used to search the KEGG database. Pathways returned by the KEGG database indicated that the gene expression patterns in several pathways suggest a single regulatory mechanism. Furthermore, pathways may be active in a specific myogenesis process such as proliferation of myoblasts or differentiation indicated by up regulated expression of most genes in that pathway. The results also indicate that pathways act together forming networks of pathways. This may give insight in higher level regulatory mechanisms taking place in the cell. We conclude that combining microarray and physiological data such as biochemical pathways in databases accessible via the internet is an important tool for gaining biological knowledge from microarray experiments.

Key Words: Bioinformatics, microarrays, Pathway analysis, KEGG Database, Myogenesis

Zusammenfassung

Titel der Arbeit: **Biochemische Pfadanalyse: Kombination von microarray und physiologischen Daten zum Studium von Myogenese**

Microarry Experimente untersuchen Änderungen in der Expression der Transcription eines Gewebes während der biologischen Prozesse zum Beispiel bei der Entwicklung dieser. Normalerweise erzeugt die Analyse eine Liste von auf und ab regulierten Genen. Bei diesen relativ großen Datenmengen können viele Informationen wichtiger biologischer Prozesse der Beziehungen zwischen den Genen unerkannt bleiben. Die über das Internet zu findenden Datenbanken enthalten wichtige physiologische Informationen, die für das Verständnis der microcarry Ergebnisse von unschätzbarem Wert sein können. Diese Datenbanken enthalten jedoch mehr Informationen über Menschen und Modelltiere wie z.B. Mäuse und Ratten als über andere Spezies. Daher entwickelten wir eine Reihe von PERL-Indizes für die automatische Datenbanksuche für biochemische Pfade, welche die Liste expressionierter Gene für microarry Experimente nutzen lassen. In einem zuvor beschriebenen microcarry Experiment, in welchem die pränatale Myogese bei Schweinen untersucht wurde, konnten diese Indizes für die Suche der KEGG Datenbank genutzt werden. Die durch die biochemischen Pfade mittels der KEGG aufgezeigten Wege zeigten einfache Regulationsmechanismen für das Muster der Genexpression. Außerdem können in spezifischen Myogeneseprozessen, wie der Zellteilung von Myoblasten oder der Differenzierung biochemische Pfade aktiv sein, und diese Pfade die Regulation der Genexpression der meisten Gene anzeigen. Die Ergebnisse zeigen auch, dass die biochemischen Pfade in einem Netzwerk fungieren. Dies gibt Einblick in das in der Zelle wirkende hohe Niveau der Regulationsmechanismen. Es wird geschlussfolgert, dass die Kombination von microarry und physiologischen Daten in den über das Internet zugänglichen Datenbanken ein wichtiges Werkzeug für biologische Erkenntnisse von microarry Experimenten bilden.

Schlüsselwörter: Bioinformation, Microarry, biochemische Pfadanalyse, KEGG Datenbank, Myogenese

Introduction

Mammalian myogenesis or muscle fibre formation is an exclusive prenatal process under strict genetic regulation (REHFELDT et al., 2004; STICKLAND et al., 2004; TE PAS and SOUMILLION, 2001). Livestock meat producing animals such as pigs

are important model organisms because of a selection history of several decades for increased meat (muscle) growth (MERKS, 2000). Previously we reported on studies aiming to elucidate Transcriptome regulation during myogenesis in prenatal pigs using microarrays (TE PAS et al., 2005A, 2005B, 2006). We also reported on differential Transcriptome regulation in pig breeds differing in muscularity (CAGNAZZO et al., 2006) indicating the genetic bases underlying selection. Microarraying is a powerful technique to simultaneously analyse the changes in the mRNA expression of all genes in a cell type or tissue. Analysis of up or down regulation of groups of genes allowed to indicate processes taking place including not only proliferation and differentiation including the expression of muscle structural genes, but also energy metabolism.
Such microarray analyse however is limited to regulated individual genes and does not necessarily provide biological knowledge of processes taking place. Presently an overwhelming amount of information about physiological and biochemical processes in humans and model animals is available in databases on the internet. Similar physiological information from other species is more limited. This study aims to develop a methodology to combine the power of the microarray technique and the information available in the KEGG (Kyoto Encyclopaedia of Genes and Genomes) database containing information on a large number of physiological / biochemical pathways taking place in the cell (KANEHISA et al., 2006), called pathway analysis (PAPIN et al., 2004). The aim of this analysis is to gain more biological relevant information about the genetic regulation of myogenesis in pigs using physiological information from other species.

Materials and Methods

Development of PERL scripts to search the KEGG pathway database

The KEGG data base (http://www.genome.ad.jp/kegg/) contains general and species-specific information on biological pathways (KANEHISA et al., 2006). While searching the KEGG database for known pathways we found that genes were not found because they may be represented with several synonyms that were not all linked to the pathways in the KEGG data base. Therefore, we first linked the microarray data with a local MySQL installation of the Gene Ontology (GO) database (http://www.godatabase.org/cgi-bin/amigo/go.cgi) containing data of the monthly release of the GO database to collect all the synonyms of gene names (some of them obsolete) and added these to the file before searching the KEGG database. To automate the searching and retrieving of pathway data from the KEGG database a PERL script (http://www.perl.com/) was written using the KEGG API (KAWASHIMA et al., 2003). Direct links to each pathway for each gene were added to the file.

Microarray data

Microarray data have been previously reported (TE PAS et al., 2005A, 2005B, 2006).

Results and Discussion

Development of the pathway analysis tool

The amount of physiological information known and present in the databases varies between species. Physiological studies have been mostly conducted with human material and model (laboratory) species such as mice and rats, but less with livestock

animals. Comparative genomics suggests that data derived from these species can be extrapolated to other – related – species. Using such information the KEGG database contains both species-specific and general pathways called reference pathways. Especially for non-human and non-model-animal species it may be important that analyses can use this information. However, many software tools for pathways analysis such as Whole Pathway Scope (http://www.abcc.ncifcrf.gov/wps/wps_login.php?typ=download) and GOminer (http://discover.nci.nih.gov/gominer/) and others use species-specific ID's to search the pathway database resulting in a limited number of pathways retrieved for non-human and non-model animal-species including livestock. Therefore, we use gene names to search the KEGG and other pathway databases. Since many genes have multiple names often related to their history (e.g. the muscle regulatory gene MRF4 has been previously named myf-6 and herculin) and not all databases recognize all names, it is important to collect all known names of all genes. Therefore, we first searched the GO database and updated the gene list with all synonyms. The pathway database was searched with the updated gene list. If the gene was found in the database a link to all pathways where the gene was found was added to the file (Fig. 1).

It should be noted that physiological information for many genes is still lacking. Most genes with known pathway information were found in a single pathway. However, there are genes found in more than one pathway, and a limited number in over ten pathways. Presently it is unknown whether the lack of physiological information is causing this difference between genes, or that many genes are specific to a single pathway while others are more generally acting in multiple pathways. The latter genes may be also important to connect pathways into a network.

The pathway information was combined with the microarray information (Fig. 1). This study analyses a time series of microarrays expressing patterns of the genes related to different steps in the myogenesis process. The results showed that most genes within involved pathways had similar expression patterns during development suggesting a single regulatory mechanism. If we assume that up regulated expression means that the pathway is actively involved in regulation of the myogenesis process (e.g. proliferation of myoblasts or differentiation) we gain biological knowledge on how myogenesis is proceeding during the development of the embryo, and therefore about the mechanism of genetic regulation of myogenesis.

The information of pathways returned by the KEGG database search suggested that pathways connect into a network (Fig. 1). This higher level of analysis may provide insight into how biochemical reactions and physiological pathways interact inside and between the cells. Pathways are human made representations of biochemical processes, which may not fully be described by a single pathway. The overall is aim is enhancing physiological understanding of biology.

Myogenesis-related pathways – first results

The KEGG database search returned 88 pathways, 21 of which showed sufficient information (i.e. number of genes on the pathway with microarray information) for further analysis. The results indicated pathways specifically up regulated during proliferation of myoblasts. Others were specifically up regulated during differentiation. Primary and secondary waves of muscle fibre formation (REHFELDT et al., 2004; STICKLAND et al., 2004) differed in expression levels of pathways.

The results suggested two larger networks of pathways (Fig. 2). One network consisted of 14 pathways involved in the regulation of the expression of the actin cytoskeleton in the cells containing four independent links to processes initiating proliferation of cells, and one to processes inducing contraction of the cells. Another complex network of ten pathways may highlight the regulation of the complex pattern of glycolytic and oxidative energy metabolism observed during myogenesis (TE PAS et al., 2005A, 2005B, 2006) and the differences in energy metabolism expression between pig breeds during myogenesis (CAGNAZZO et al., 2006).

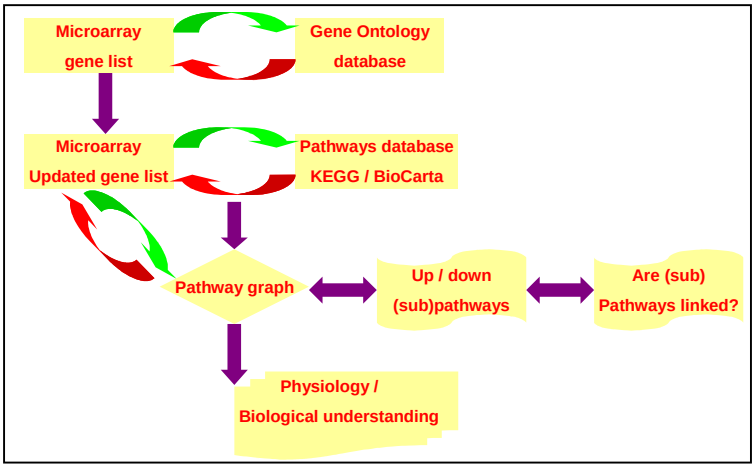

Fig. 1: Lay out of the pathway analysis method

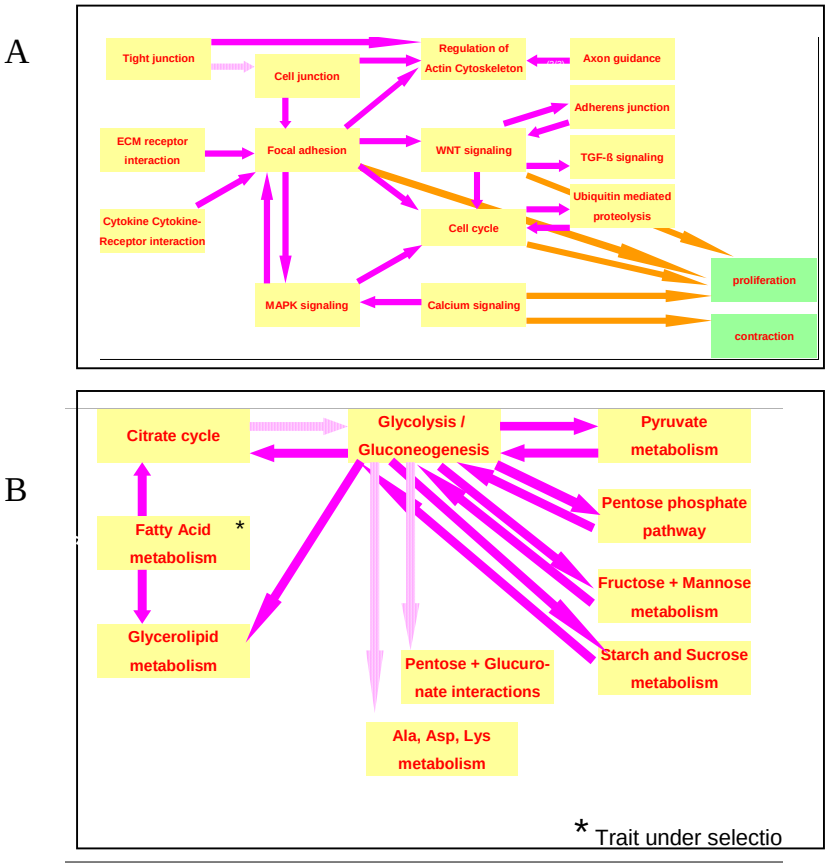

Fig. 2: Networks of pathways involved in Regulation of Actin Cytoskeleton (A) and Regulation of energy metabolism (B). Pathways are indicated in boxes. Arrows denote interactions between pathways in the network.

Conclusion: New knowledge derived from the microarrays using pathways analysis
The pathway analysis provided information about physiological processes taking place during myogenesis in pigs. Furthermore, higher level regulatory mechanisms of pathways acting together in networks are suggested that may provide insight in the network of biochemical reactions taking place in and between cells. Similar analyses in other experiments (e.g. *Salmonella enteritidis* infection in chicken) using the same tool confirm the usefulness of meta-analysis of large microarray datasets.

Acknowledgement

This research was supported by internal WUR funds (grant numbers 213.213.4000 and 4434.6057.00). Microarray data were from FP5 EU project PorDictor, contract number QLTR-1999-31363.

References

CAGNAZZO, M.; TE PAS, M.F.W.; PRIEM, J.; DE WIT, A.; POOL, M.H.; DAVOLI, R.; RUSSO, V.:
Comparison of prenatal muscle tissue expression profiles of two pig breeds differing in muscle characteristics. J. Anim. Sci. **84** (2006), 1-10

KANEHISA, M.; GOTO, S.; HATTORI, M.; AOKI-KINOSHITA, K.F., ITOH, M.; KAWASHIMA, S.; KATAYAMA, T ; ARAKI, M.; HIRAKAWA, M.:
From genomics to chemical genomics: new developments in KEGG. Nucl. Acids Res. **34** (2006), D354-D357

KAWASHIMA, S.; KATAYAMA, T.; SATO, Y.; KANEHISA, M.:
KEGG API: A web service using SOAP/WSDL to access the KEGG system. Genome informatics **14** (2003), 673-674

MERKS, J.W.M.:
One century of genetic changes in pigs and the future needs. In The Challenge of Genetic Change in Animal Production. Occasional Publication. Brit Soc Anim Sci, Edited by Hill WG, Bishop SC, McGuirk B, McKay JC, Simm G, Webb AJ. Edinburgh **27** (2000), 8-19

PAPIN, J.A.; STELLING, J.; PRICE, N.D.; KLAMT, S. ; SCHUSTER, S. ; PALSSON, B.O.:
Comparison of network-based pathway analysis methods. Trends Biotechnol **22** (2004), 400-405

REHFELDT, C. ; FIEDLER, I. ; STICKLAND, N.C. :
Number and size of muscle fibres in relation to meat production. In Muscle Development of Livestock Animals - Physiology, Genetics, and Meat Quality Edited by te Pas MFW, Everts ME, Haagsman HP. CABI publishers, Wallingford, Oxfordshire, UK, (2004), 1-38.

STICKLAND, N.C., BAYOL, S., ASHTON, C., REHFELDT, C. :
Manipulation of muscle fibre number during prenatal development. In Muscle Development of Livestock Animals - Physiology, Genetics, and Meat Quality Edited by te Pas MFW, Everts ME, Haagsman HP. CABI publishers, Wallingford, Oxfordshire, UK, (2004), 69-82.

TE PAS, M.F.W.; DE WIT, A.A.C.; PRIEM, J.; CAGNAZZO, M.; DAVOLI, R.; RUSSO, V.; POOL, M.H.:
Transcriptome expression profiles in prenatal pigs in relation to myogenesis. J. Muscle Res. Cell Motil. **26** (2005A), 157-165

TE PAS, M.F.W.; CAGNAZZO, M.; DE WIT, A.A.C.; PRIEM, J.; POOL, M.H.; DAVOLI, R.:
Muscle transcriptomes of Duroc and Pietrain pig breeds during prenatal formation of skeletal muscle tissue using microarray technology. Arch. Tierz., Dummerstorf **48** (2005B) Special Issue, 141-147

TE PAS, M.F.W.; POOL, M.H.; HULSEGGE, I.; JANSS, L.L.G.:
Analysis of the differential transcriptome expression profiles during prenatal muscle tissue development. Arch. Tierz., Dummerstorf **49** (2006) Special Issue, 110-115

TE PAS, M.F.W.; SOUMILLION, A.:
The use of physiologic and functional genomic information of the regulation of the determination of skeletal muscle mass in livestock breeding strategies to enhance meat production. Current Genomics **2** (2001), 285-304

Corresponding Author
MARINUS F.W. TE PAS, Wageningen University and Research Centre, Animal Sciences Group, Division of Animal Resource Development,
ID-Lelystad, PO Box 62, 8200 AN Lelystad, The Netherlands