

## Research Article

# A Pathway Analysis Tool for Analyzing Microarray Data of Species with Low Physiological Information

M. F. W. te Pas,<sup>1</sup> S. van Hemert,<sup>2</sup> B. Hulsege,<sup>1</sup> A. J. W. Hoekman,<sup>1</sup> M. H. Pool,<sup>1</sup>  
J. M. J. Rebel,<sup>2</sup> and M. A. Smits<sup>1</sup>

<sup>1</sup>Animal Breeding and Genetics Centre (ABGC), Animal Sciences Group, Wageningen University and Research Centre, P.O. Box 65, 8200 AB, Lelystad, The Netherlands

<sup>2</sup>Department of Infectious Biology, Animal Sciences Group, Wageningen University and Research Centre, P.O. Box 65, 8200 AB, Lelystad, The Netherlands

Correspondence should be addressed to M. F. W. te Pas, marinus.tepas@wur.nl

Received 18 February 2008; Revised 25 July 2008; Accepted 17 October 2008

Recommended by Pierre Rouzé

Pathway information provides insight into the biological processes underlying microarray data. Pathway information is widely available for humans and laboratory animals in databases through the internet, but less for other species, for example, livestock. Many software packages use species-specific gene IDs that cannot handle genomics data from other species. We developed a species-independent method to search pathways databases to analyse microarray data. Three PERL scripts were developed that use the names of the genes on the microarray. (1) Add synonyms of gene names by searching the Gene Ontology (GO) database. (2) Search the Kyoto Encyclopaedia of Genes and Genomes (KEGG) database for pathway information using this GO-enriched gene list. (3) Combine the pathway data with the microarray data and visualize the results using color codes indicating regulation. To demonstrate the power of the method, we used a previously reported chicken microarray experiment investigating line-specific reactions to Salmonella infection as an example.

Copyright © 2008 M. F. W. te Pas et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## 1. Introduction

Microarray technology can simultaneously measure the expression of large numbers of genes in a tissue and thereby identify the genes involved in a process. Typically, microarray experiments produce long lists of genes that are differentially expressed between two different situations. In order to better understand the biology behind these data, it is relevant to include the available biological information of the genes under study [1]. Many databases such as the KEGG contain information on biochemical pathways [2]. Combination of microarray data and pathway information may highlight the processes taking place in the cell and tissue and provide biological knowledge on the tissue- and process-specific functioning of the genome.

Pathway databases contain information mainly based on research performed with human and laboratory animal material. Most pathway information is displayed species-

specific. Livestock animals and animals with less information on genome sequence and/or physiology are less represented. Comparative genomics suggests that most of the genetics and physiology of the less well-represented species will be similar or comparable with the data of human and laboratory animal species stored in the database. However, many software tools to analyse microarray data use species-specific gene identification. This makes it difficult to use pathway information for other animal species. The development of software tools that allow the use of pathway information across other species is therefore necessary.

The present study aimed to develop software tools using species-independent gene IDs that streamline the process of searching for pathways information in online databases using lists of genes represented on microarrays followed by combining pathway information with microarray data. This enabled us to identify relevant pathways from the KEGG database (see [2] and <http://www.genome.ad.jp/kegg/>) for

livestock species. Part of the software has been tested and published before [3]. A new powerful module has been added since then enabling the direct quantitative visualization of microarray results on the pathway file obtained through the internet. To demonstrate the power of the method, we used a dataset of a previously reported chicken microarray experiment investigating line-specific host reaction to Salmonella infection. Combination of the microarray data with the pathway information highlighted line-specific biological processes underlining the added value of the developed method.

## 2. Materials and Methods

### 2.1. Database Searches

The KEGG database (see <http://www.genome.ad.jp/kegg/>) contains general information on biological pathways including gene names and information on species-specific pathways [2]. While searching the KEGG database with known pathways, we found that genes may be represented with several synonyms that were not all linked to the pathways in the KEGG database. Therefore, we first linked the microarray data with a local MySQL (see <http://www.mysql.com/>) installation of the Gene Ontology database (<http://www.godatabase.org/cgi-bin/amigo/go.cgi>) which contains the monthly release to collect all the common 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 [2], a PERL script was written using the KEGG API [4]. Direct links to each pathway for each gene were added to the file. A third PERL script quantitatively visualizes the microarray results in the obtained pathways. All database searches were performed with homemade PERL scripts (<http://www.perl.com/>). The software can be used for free at <http://www.ASGbioinformatics.wur.nl/>. Free registration is required to use the software.

### 2.2. The Example: Animals, Experiment, and Microarray Analysis

Two chicken lines differently selected for growth rate were used. The lines also differed for Salmonella host response. Five one-day-old chickens were orally inoculated with  $10^5$  CFU *S. enteritidis*, five animals served as controls. Twenty-four hours after the infection, the chickens were killed and parts of the jejunum were snapped frozen in liquid nitrogen and used for RNA isolation. For further details, see van Hemert et al. [5, 6].

RNA pools were hybridized on an Affymetrix chicken whole genome Genechip array. The annotation file of the microarrays was provided by the supplier. See van Hemert et al. [6] for further details on the microarrays used, the hybridizations, and the first analysis. For raw data see

NCBI (<http://www.ncbi.nlm.nih.gov/geo/>), accession number GSE3702.

## 3. Results

### 3.1. Flow Diagram of the Developed Tool

The pathway analysis tool is a four-step procedure; see Figure 1. The first three steps are automated with PERL scripts that are available as additional information (see Additional File 1 available online at doi:10.1155/2008/719468). The tool uses gene names to find pathway information in the database. Since genes may be known by different names, it is important to have all the synonymous names of all the genes in the gene name list of the microarray. Therefore, the first step (script 1) of the procedure is to collect all the synonymous names of all genes. The Gene Ontology (GO) database (<http://amigo.geneontology.org/cgi-bin/amigo/go.cgi>) is a web-based resource that contains all synonyms. The PERL-script tool searches a local download of the GO database using a txt file of a list of all gene names on the microarray and the results were added to the list. The second script uses this updated gene list to search the pathway database. There are several pathway databases accessible via the internet such as KEGG (<http://www.genome.ad.jp/kegg/>), BioCarta (<http://www.biocarta.com/>), Reactome (<http://www.reactome.org/>), and others. Presently, the PERL script to search the KEGG database is available as a proven tool and the BioCarta tool is under development. Searching the database returns for all recognized gene names, the names of the pathways in which the genes are found, and a link to the reference pathway. These data are added to the file. The reference pathways are developed by the KEGG database comparing the species-specific pathways. Thus, the reference pathways represent pathways that seem to be similar in different species. In our projects, we often use the reference pathways after checking for similarity with the human and mouse pathways. The third step (script) downloads the figure/diagram of the pathway and visualizes the microarray results in the pathways figure. The PERL script places a colored oval around the gene name in the figure. It is suggested to use green for upregulated genes, red for downregulated genes, and black for not regulated genes. The tool can visualize separately together the data from more than one microarray in the figure. A practical step by step guide to use the pathway analysis tool is given in Box 1.

The fourth step is not automated by a PERL script, and is probably not automatable in a research setting because it comprises the biological analysis of the results generating physiological or biological knowledge from the data. Part of this analysis is the generation of networks of pathways, which may be automated in the future. Networks of pathways are generated using two types of data: (1) KEGG pathway figures may indicate input from, or output to another pathway, or (2) genes may be found in more than one pathway suggesting direct links between pathways. The final outcome

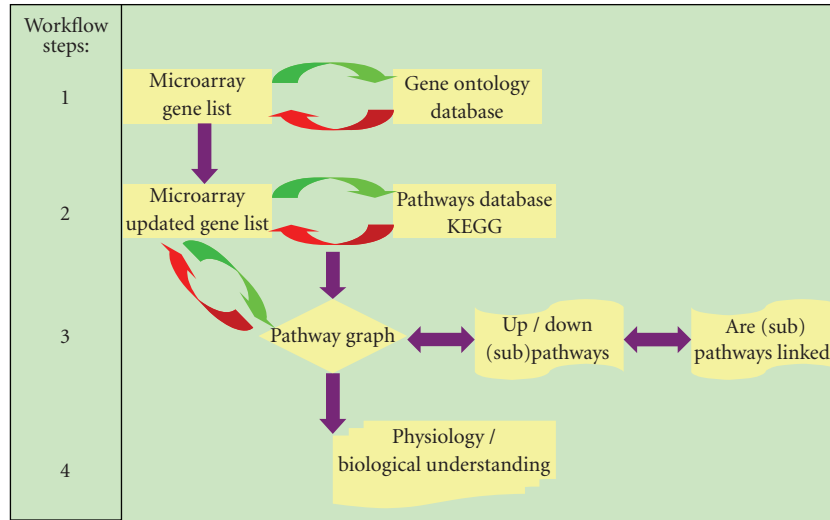


FIGURE 1: Flow diagram of the pathway analysis tool. The method comprises four steps, of which the first three steps are automated by PERL scripts. Step 1. Use the gene names list of the microarray annotation to search the Gene Ontology database for synonyms of the gene names and add to the list. Step 2. Use the updated gene name list to search the pathway database for pathway information of the genes and add pathway names and web links to the file. Step 3. Visualize the microarray results on the pathways. Step 4. Construct networks of pathways and generate biological knowledge from the results.

of the method is that biological knowledge is generated from microarray data.

### 3.2. An Example: Chicken Line-Specific Reaction to Salmonella Infection

To show the power of the developed tools, we used a dataset of a previously reported experiment. The study focussed on the early host gene expression response to Salmonella infection in the intestine of newly hatched chicken of two chicken lines.

*Step 1.* The Affymetrix GeneChip chicken genome array contained 14 343 entries (38,449 Gallus gallus probe sets, with 11 pairs of 25 mers). However, these entries contained known genes, EST sequences, and chromosomal locations. Only known genes with names can be used to search the GO database for synonyms and to search the KEGG database for pathways. In this example, a total of 4666 gene names were found in the GO database and updated.

*Step 2.* The KEGG database search retrieved 178 pathways. From the known genes 3520 gene—pathway combinations were found in the KEGG database. Of these, 1203 gene—pathway combinations showed up or down regulation of the gene expression in at least one of the four experiments. The number of genes per pathway with microarray information varied from 1 (14 pathways) to 109 (one pathway) (Figure 2).

*Step 3.* Visualization of microarray data on the pathways revealed that the pathways could be categorized (see Table 1). *Category A:* 57 pathways with relevant microarray information were observed, of which 22 have suggested linkages with one or more other KEGG pathways. KEGG pathways may

have connections with other pathways through input/output of biochemical products or via protein sharing forming a biochemical network of pathways. See Additional Files 2 and 3 for information on each individual gene within each individual pathway. Additional File 3 shows the visualization of microarray data, which is the output of the newly added module as compared to the previous publication [3]. *Category B:* fourteen pathways with relevant microarray information but none of the genes showed differential expression. This indicates that these pathways are active but not involved in the regulation of the traits under investigation. *Category C–F:* For several reasons pathways cannot be analysed further, or can be analysed only partly: (i) Especially for the pathways represented by a few genes (e.g., less than five) on the microarray the information content is low and the participation of that pathway in the processes studied was considered highly uncertain (Category E). However, it was observed that for several pathways limited microarray information was present. Such limited information may be found localized on a single biochemical path. We named this a subpathway, and analyzed it further; (ii) Other pathways returned by the KEGG database search were clearly false positive hits, for example, a photosynthetic pathway was returned by the KEGG database search for one gene (Category D).

*Step 4* (constructing networks of pathways). Networks of pathways can be generated using two types of data. (i) KEGG links between pathways indicate where the information in one pathway ends and continues in the next pathway. (ii) Genes may be active in more than one pathway (see Figure 3). Most genes were found in a single pathway. The KEGG database search returned more than one pathway for 698 genes, ranging from two pathways up to a maximum of

**Practical step-by-step guide to use the Pathways analysis tool**

- (1) Log in to the website (www.asgbioinformatics.wur.nl). If you did not previously please register and you will receive a free log in name and password. This is only for registration that will ensure that you will receive an answer to questions asked to our help desk.
- (2) Prepare a flat text file (.txt) containing two columns with (1) a gene ID and (2) the names of the genes on the microarray (usually derived from the annotation database). Each row should contain a SINGLE name. Rules for presenting gene names are in the help file on the website: select microarray analysis/pathway/GO gene name synonyms and press “[here](#)” is an example with 6 keywords (gene names).
- (3) Upload the file to the website: and press the search button. This may take a while – for large files even days. The results field can be bookmarked to your favorites so that you can return and collect the results later.
- (4) The results file is in .txt format and consists of four columns: your original columns, a copy of your gene name column that has been used to search the Gene Ontology database, and the results. The results column may contain a single row if the gene is not found, or if the gene is known under a single name. However, it may contain multiple rows if multiple synonyms for the gene have been found.
- (5) Prepare a new .txt file with the fourth column (the synonyms) of the file in step 4.
- (6) Upload the file to the website: select microarray analysis/pathway/Pathway\_Kegg and press the button “search pathway”. This again may take a while (even longer than step 3)—and the same procedure can be followed.
- (7) The results file is a .txt file with two columns added to the input file: one column containing the name of a pathway that has been found in the KEGG database, and one column that contains a link to the pathway and highlights the position of the gene in the pathway. Please note that each gene can have either zero or up to many pathways. Please note that if a gene returns more than 100 reactions the run is aborted and the message displayed is “NUMBER OF HITS FOUND > 100 !!!!” in the column for pathway names.
- (8) Prepare a file containing the columns of the file in 7 (see above) and add the results of the microarray to the genes. The results of the microarrays can be displayed as the M-values with or without P-values to indicate significance of the results. The results of more than one microarray may be added to the file in separate columns. If a high number of microarray results should be added (e.g. 100) please contact the help desk and they will increase the number of columns available.
- (9) Upload the file to the website: select microarray analysis/pathway/Visualization Kegg\_microarray data and press the button “continue”. Fill in the form to indicate the colours and criteria for up- or down regulation, or for non-regulated genes. Please note that this may take some time—follow the same procedure as above.
- (10) The output is a figure file for each pathway with all genes on the microarray encircled in the required colour visualizing the microarray results.

Box 1

35. Altogether networks of pathways are constructed for “mechanisms of cytoskeletal changes” (ten KEGG pathways), “apoptosis mechanism” (five KEGG pathways) and “regulation of energy metabolism” (six KEGG pathways), and several pathways that could be grouped without direct network associations (see te Pas et al.s’ Additional File 4 for details).

### 3.3. Biological Knowledge Derived from the Pathways Analysis

For each of the networks of pathways and some other pathways, the chicken lines differed in reaction to Salmonella infection. The data of all pathways of the networks are summarized in the table in Additional File 5. Detailed information is given in the file following the table. In summary, chicken lines A and B differ in their Salmonella

susceptibility phenotype. The faster growing line A shows more severe illness as measured with growth and colony-forming units in the liver to Salmonella infection than the slower growing line B. The results of the pathway analysis of the microarray data provide insight into differential line-specific biological processes that may explain the difference in host response to Salmonella. Three major networks of pathways that differ between the lines are discussed in more detail below.

#### *(i) Mechanisms of Cytoskeletal Changes*

Pathways analysis indicated that the chicken lines differ in their expression of the genes in the network regulating the cytoskeleton, at least in the intestinal tissues used for microarray analysis. The genes in the pathways in this network show higher expression levels in line B compared to line A. Following Salmonella infection, line A increases

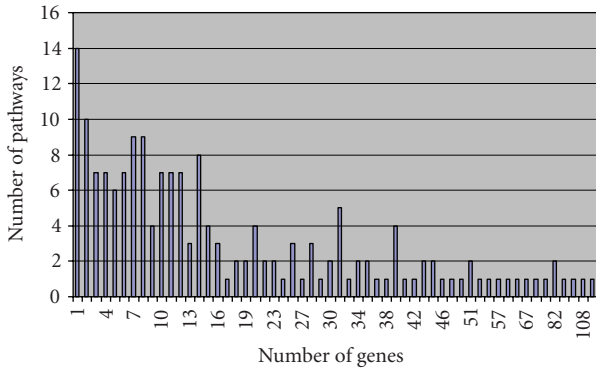


FIGURE 2: The number of genes on the microarray found per pathway.

TABLE 1: Pathways returned by searching the KEGG database. The list of gene names of the microarray was first updated for all synonyms using the Gene Ontology database before searching the KEGG database.

Pathways category	Number of pathways
A: total number of pathways returned by KEGG search	178
B: positive associated pathways (with links to other pathways)	57 (22)
C: no regulation*	14
D: false positive**	30
E: (too) limited information to conclude***	35
F: uncertain****	42

\*None of the genes in the pathways present on the microarray were regulated. These pathways were considered not related to the chicken line differences for Salmonella infection response.

\*\*Plant or prokaryote-specific pathways—recognized due to gene name similarity. Almost always these pathways were recognized by a very limited number of genes on the microarray.

\*\*\*The genes with microarray data on the pathway are too limited in number or their locations on the pathway are too scattered to conclude on the relevance of the pathway for the studied traits.

\*\*\*\*The available microarray data of the genes on the pathway show either limited or confusing up- or downregulatory patterns.

its expression to the same (basal) level of line B, while the expression level in line B was unchanged. Thus, selection for a production trait has influenced the reaction mechanism of the animals to respond to Salmonella infection. This result may be related to the reaction time of the lines to Salmonella infection: line B reacts faster than line A. Apparently, the different selection background of the chicken lines which created a difference in growth rate between the chicken lines is accompanied with a difference in basal expression levels of the cytoskeletal network pathways. Part of these results confirmed the results of previous reports [5, 6] or what is known about the involvement of the cytoskeleton in cellular uptake of Salmonella [7].

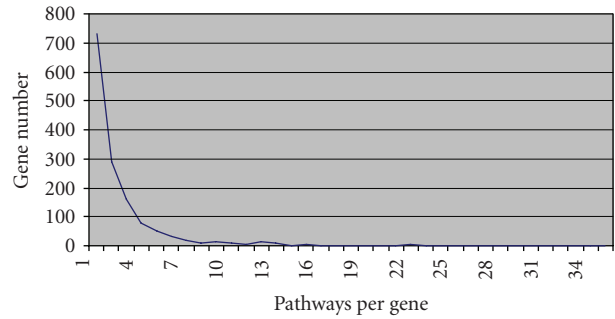


FIGURE 3: The number of pathways per gene returned by the KEGG database search.

(ii) Apoptosis Mechanism

Pathway analysis suggests that the intestinal tissue of line A reacts to a Salmonella infection with an apoptotic mechanism, while line B resumes growing (i.e., proliferation and differentiation mechanisms).

(iii) Regulation of Energy Metabolism

The apoptosis versus growth mechanism may be supported by the observed higher expression of genes related to energy metabolism in line B making energy available for the mechanisms of cell proliferation and differentiation. However, it should be noted that the expression levels of the genes in the energy metabolism pathways were reduced in both lines in response to Salmonella infection, but the effect was more severe in line A than in line B.

4. Discussion

High throughput postgenomics methods generate large-scale dataset. The conversion of these data into knowledge requires approaches that allow integration of the results into models describing how the cell or its components, or tissues, organs, and so forth, work. This approach is called systems biology [8], which requires computerized methods to build integrated models on all levels. A good example of this is the Physiome Project [9, 10]. Placing transcriptomics and proteomics results in the physiological knowledge is a first step on this road. Different approaches consist to continue in this direction. Based on the functional genomics, data genetic networks can be identified [1, 11, 12]. Models for gene regulation and regulation of networks of genes can be determined [13]. Such approaches are valuable for understanding the large datasets, but these results might or might not be related to the physiology of the cell. Other approaches more directly use the wealth of validated physiological knowledge available through the internet [14]. However, such data is often presented in a species-specific way for good reasons. To enable us to use the physiological information for less well-studied species, software tools need to be developed. The pathway analysis tool described here generates physiological data from microarray results of

species less well represented in the physiological pathway databases. Previously, a manuscript using only the first two modules of the pathway analysis software tool has been published [3]. In the latter manuscript, we used these modules to analyze microarray data on the prenatal development of muscle tissue in pigs. The data consisted of a time series. Results were not recorded as up or downregulation, but as differential expression in time. The latter paper is an example of a specific case of using only part of the pathway analysis tool made possible by keeping the software as free—nonintegrated—software components. In contrast, the present description of the software and the example given show how quantitative results of microarray experiments can be integrated directly into pathway structures and visualized. This enables to draw conclusions about up or downregulated pathways. Furthermore, it suggests how information flows may go from one pathway to another and thus how pathways can be integrated into networks.

#### 4.1. The Pathway Analysis Tool: Merging Microarray Data with Biological Database Information

Pathway analysis is a tool to produce biological meaningful knowledge from the huge amount of data resulting from microarray experiments. Biochemical pathways such as those stored in the KEGG database describe physiological processes. However, one should keep in mind that the description of the biological process may be species-specific. Furthermore, the gene list of the microarray may be incomplete for a pathway, due to inadequate annotation of genes. These considerations may hamper the analysis of the microarray results for the pathways. The physiology of a process may also differ between lines due to selection background. Therefore, both general (called reference) pathways and species-specific pathways can be searched. Chicken-specific pathways are often not available. Therefore, we used pathway data from other species and always compared these with the reference pathways, which were used for further analysis.

A set of PERL scripts written to extract data from databases via the internet was developed. The need for biological interpretation of microarray data has been recognized by many research groups, and consequently the presented pathway analysis tool is not the only tool that can be applied. However, many software packages work well with human or model (laboratory) animal species, but less well with organisms lacking (much) physiologic knowledge. Nevertheless, applying the principle of comparative genomics could make the knowledge of the model species available for other species if the software tool guaranties the use of species-independent gene identifiers. For example, one could try the use of Entrez IDs and HomoloGene. On the other hand, this software tool uses the identifier recognized by each investigator: the gene name or abbreviation. We feel this gives an advantage to this software. Differently from other similar software that can be found on the internet (Bioconductor (<http://www.bioconductor.org/>), Whole Pathway Scope ([<http://discover.nci.nih.gov/gominer/>\), GenMapp \(<http://www.genmapp.org/>\)\) we use the names and synonyms of the genes rather than the species-specific gene-IDs. Therefore, this tool allows analysing microarray data from species with relative low physiological and/or sequencing information in the database. As a consequence, it will remain important to screen manually for obvious false positive pathways. In the example, a chlorophyll metabolism pathway was discarded as false positive result \(1 gene in the pathway, data not shown\).](http://www.abcc.ncifcrf.gov/wps/wps_login</a></p>
</div>
<div data-bbox=)

From all entries listed on the microarray and used for searching the KEGG pathway database, approximately 25 percent were found on one or more pathways. The reasons for not finding pathway information for an entry may be diverse, but often related to the annotation of the microarray: (1) gene entries indicated by a locus (LOC) number were rarely found on a pathway; (2) hypothetical protein entries were also rarely found on a pathway; (3) entries indicated by a chromosomal position were never found on a pathway; and (4) in the example given in Section 3.2 with details in the additional file, the annotation of chicken genes is often poor. These results strongly support the importance of the continuously ongoing efforts to improve the annotation of microarrays. We have used the annotation provided by the supplier, which was similar to the annotation used in the initial analysis [5]. This implies that the improvement of the current analysis is solely due to the use of the new developed pathway analysis tool. Of course, updating the annotation to a more up-to-date level could have increased the number of pathways found, and thus improved the analysis even further.

Apart from these reasons, there were also known genes without known pathway information in the KEGG database. However, the method used in this study returned substantially more results than the software where species-specific gene IDs have to be used (data not shown) because many pathways in the KEGG and other databases do not have a chicken-specific pathway included although it is clear that many metabolism-specific pathways do exist in the chicken. However, one should keep in mind that species-specific differences in pathways do exist. We keep this risk to a minimum using the reference pathways of the KEGG database, which are created combining the species-specific pathways of as many as possible species.

Approximately one third of the pathways found by searching the KEGG database in our example proved to be regulated in the experiment. The majority of the pathways were excluded for several reasons as outlined in Table 1. Apart from false positive pathways that may be found due to genes that may have similar gene name, synonyms about one third of the pathways were supposed to be irrelevant for the regulation of the traits under investigation because of the absence of differential expression in the experiments or too limited information content in the microarray dataset.

In the example given, the pathways analysis resulted in not less than 57 pathways which were found to be differentially regulated between the two chicken lines with or without Salmonella infection. Furthermore, from the data, four networks were designed that describe the reaction of the chicken to Salmonella at a higher level. The results

confirm and extend the previously reported results and extend the knowledge to a higher physiological level—networks of regulated pathways in intestine tissue. This example shows that the developed tool is able to increase biological insight in processes studied with microarrays, especially for species with either little genomic information or with little physiologic information available.

## Acknowledgments

This research was supported by internal WUR funds (Grants no. 213.213.4000 and no. 4434.6057.00). The research was also supported by a grant from the EADGENE FP6 EU Network of Excellence.

## References

- [1] J. A. Papin, J. Stelling, N. D. Price, S. Klamt, S. Schuster, and B. O. Palsson, "Comparison of network-based pathway analysis methods," *Trends in Biotechnology*, vol. 22, no. 8, pp. 400–405, 2004.
- [2] M. Kanehisa, S. Goto, M. Hattori, et al., "From genomics to chemical genomics: new developments in KEGG," *Nucleic Acids Research*, vol. 34, database issue, pp. D354–D357, 2006.
- [3] M. F. W. te Pas, I. Hulsege, A. Coster, M. H. Pool, H. H. Heuven, and L. L. G. Janss, "Biochemical pathways analysis of microarray results: regulation of myogenesis in pigs," *BMC Developmental Biology*, vol. 7, article 66, pp. 1–15, 2007.
- [4] S. Kawashima, T. Katayama, Y. Sato, and M. Kanehisa, "KEGG API: a web service using SOAP/WSDL to access the KEGG system," *Genome Informatics*, vol. 14, pp. 673–674, 2003.
- [5] S. van Hemert, A. J. W. Hoekman, M. A. Smits, and J. M. J. Rebel, "Gene expression responses to a *Salmonella* infection in the chicken intestine differ between lines," *Veterinary Immunology and Immunopathology*, vol. 114, no. 3–4, pp. 247–258, 2006.
- [6] S. van Hemert, A. J. W. Hoekman, M. A. Smits, and J. M. J. Rebel, "Early host gene expression responses to a *Salmonella* infection in the intestine of chickens with different genetic background examined with cDNA and oligonucleotide microarrays," *Comparative Biochemistry and Physiology Part D*, vol. 1, no. 3, pp. 292–299, 2006.
- [7] D. G. Guiney and M. Lesnick, "Targeting of the actin cytoskeleton during infection by *Salmonella* strains," *Clinical Immunology*, vol. 114, no. 3, pp. 248–255, 2005.
- [8] T. Ideker, T. Galitski, and L. Hood, "A new approach to decoding life: systems biology," *Annual Review of Genomics and Human Genetics*, vol. 2, pp. 343–372, 2001.
- [9] P. Hunter, P. Robbins, and D. Noble, "The IUPS human Physiome project," *Pflügers Archiv European Journal of Physiology*, vol. 445, no. 1, pp. 1–9, 2002.
- [10] P. J. Hunter and T. K. Borg, "Integration from proteins to organs: the Physiome project," *Nature Reviews Molecular Cell Biology*, vol. 4, no. 3, pp. 237–243, 2003.
- [11] M. Xiong, J. Li, and X. Fang, "Identification of genetic networks," *Genetics*, vol. 166, no. 2, pp. 1037–1052, 2004.
- [12] X. J. Zhou, M.-C. J. Kao, H. Huang, et al., "Functional annotation and network reconstruction through cross-platform integration of microarray data," *Nature Biotechnology*, vol. 23, no. 2, pp. 238–243, 2005.
- [13] N. J. Guido, X. Wang, D. Adalsteinsson, et al., "A bottom-up approach to gene regulation," *Nature*, vol. 439, no. 7078, pp. 856–860, 2006.
- [14] H. Pang, A. Lin, M. Holford, et al., "Pathway analysis using random forests classification and regression," *Bioinformatics*, vol. 22, no. 16, pp. 2028–2036, 2006.