1-1-2004

A mixed-model approach for the analysis of cDNA microarray gene expression data from extreme-performing pigs after infection with Actinobacillus pleuropneumoniae

R J. Moser
CSIRO

A Reverter
CSIRO

Caroline A. Kerr
University of Wollongong, carkerr@uow.edu.au

K J. Beh
CSIRO

A mixed-model approach for the analysis of cDNA microarray gene expression data from extreme-performing pigs after infection with Actinobacillus pleuropneumoniae

Abstract

We proposed a novel statistical approach for the analysis of cDNA experiments based on mixed-model methodology combined with mixtures of distributions. Our objective was to detect genes that may be involved in conferring heritable differences in susceptibility to common infections in intensive pig production. We employed a microarray expression profiling strategy and a mixed-model approach to the analysis of the expression data. A cDNA microarray of pig with 6,420 probes from immune tissues and cells was used to compare gene expression in peripheral blood leukocytes of two pigs showing extreme performance in their response to infection with Actinobacillus pleuropneumoniae. Principal components analyses were used to identify the two most extreme-performing pigs after infection (i.e., pigs whose measured responses to infection fell at the extremes). Blood samples and expression profiles from 0 to 24 h after infection were compared using a bivariate, mixed-model approach, in which the effect gene × immunological status interaction was treated as a random effect. Bayesian model-based clustering via mixtures of normal distributions of the resulting BLUP of the random interaction was approached and resulted in a list of 307 differentially expressed genes, of which 179 were down-regulated in the susceptible pig. The majority of the differentially expressed genes were derived from a cDNA library of leukocytes of A. pleuropneumoniae-challenged pigs that were subtracted against leukocytes before the challenge. These results provide evidence that the proposed statistical approach was useful in enhancing the knowledge of the mechanisms involved in the genetics of the immune response.

Keywords

pigs, performing, extreme, data, expression, gene, microarray, after, cdna, pleuropneumoniae, analysis, approach, model, mixed, infection, actinobacillus

Disciplines

Medicine and Health Sciences

Publication Details

A mixed-model approach for the analysis of cDNA microarray gene expression data from extreme-performing pigs after infection with *Actinobacillus pleuropneumoniae*

R. J. Moser, A. Reverter, C. A. Kerr, K. J. Beh and S. A. Lehnert

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://www.journalofanimalscience.org/content/82/5/1261

American Society of Animal Science

www.asas.org

A mixed-model approach for the analysis of cDNA microarray gene expression data from extreme-performing pigs after infection with *Actinobacillus pleuropneumoniae*[1]

R. J. Moser, A. Reverter[2], C. A. Kerr, K. J. Beh, and S. A. Lehnert

CSIRO Livestock Industries, Queensland Bioscience Precinct, St. Lucia, Queensland 4067, Australia

ABSTRACT: We proposed a novel statistical approach for the analysis of cDNA experiments based on mixed-model methodology combined with mixtures of distributions. Our objective was to detect genes that may be involved in conferring heritable differences in susceptibility to common infections in intensive pig production. We employed a microarray expression profiling strategy and a mixed-model approach to the analysis of the expression data. A cDNA microarray of pig with 6,420 probes from immune tissues and cells was used to compare gene expression in peripheral blood leukocytes of two pigs showing extreme performance in their response to infection with *Actinobacillus pleuropneumoniae*. Principal components analyses were used to identify the two most extreme-performing pigs after infection (i.e., pigs whose measured responses to infection fell at the extremes). Blood samples and expression profiles from 0 to 24 h after infection were compared using a bivariate, mixed-model approach, in which the effect gene × immunological status interaction was treated as a random effect. Bayesian model-based clustering via mixtures of normal distributions of the resulting BLUP of the random interaction was approached and resulted in a list of 307 differentially expressed genes, of which 179 were down-regulated in the susceptible pig. The majority of the differentially expressed genes were derived from a cDNA library of leukocytes of *A. pleuropneumoniae*-challenged pigs that were subtracted against leukocytes before the challenge. These results provide evidence that the proposed statistical approach was useful in enhancing the knowledge of the mechanisms involved in the genetics of the immune response.

Key Words: Complementary DNA, Gene Expression, Microarray, Swine

Introduction

Gene expression microarray technology can expedite the identification of genes responsible for infectious disease pathogenesis in livestock as well as an elucidation of their mechanisms (Kato-Maeda et al., 2001). *Actinobacillus pleuropneumoniae*, the etiologic agent of contagious pleuropneumonia in pigs, has a great impact on the pig industry worldwide. In Australia, the chronic form of the disease has been estimated to cost $64 per sow and year (Blackall, 2001). These economic losses relate to the poor performance of infected pigs expressed as decreased weight gain and low feed efficiency or increased death rate (Taylor, 1999; Hoflack et al., 2001).

One challenge presented by the large data sets from microarray experiments is the development of accurate data processing and analysis techniques pointing towards the reliable detection of transcripts differentially expressed as a result of treatment effects. New statistical methods for the analysis of microarray data are being proposed in the literature at an incredible pace. A recent review of statistical tests for differentially expressed genes in microarray experiments can be found in Cui and Churchill (2003). The authors identify the mixed-model ANOVA method as a general and powerful approach for microarray experiments with multiple factors and several sources of variation.

The objective of this article is to present a mixed-model approach for the analysis of cDNA microarray gene expression data from pigs following bacterial challenge. The resulting BLUP for the gene expression × immunological status interaction were processed through model-based clustering via mixtures of normal distributions to identify differentially expressed genes.

The approach was tested on RNA from the peripheral blood leukocytes from two extreme-performing pigs in terms of susceptibility to pleuropneumonia at 0 and 24 h postchallenge with *A. pleuropneumoniae*.

Materials and Methods

Animals, Bacterial Challenge, Selection Criteria, and RNA Extraction

An animal ethics committee approved all animal procedures. Eighteen 28-wk-old male pigs (predominantly Large white × Landrace) free of *A. pleuropneumoniae* and weighing on average 33 ± 5 kg were housed in individual pens in two climate-controlled separate air spaces with two rooms per air space. These animals were part of a larger experiment conducted at the Livestock Industries division of the Commonwealth Scientific and Industrial Research Organization (CSIRO) of Australia. Briefly, pigs were fed a commercial diet in individual feed troughs containing 10 g lysine/kg and 13.5 MJ digestible energy/kg (DM basis) at 0900 each day with the residual food removed for the calculation of daily feed intake.

Pigs were acclimatized to these environmentally controlled rooms for 8 d before challenge. Then pigs were fitted with venous catheters 1 d before the challenge. At challenge, pigs were given 1×10^5 cfu of *A. pleuropneumoniae* (serotype 1, HS54) endotracheally. In addition to monitoring clinical signs like coughing, lethargy, anorexia and labored breathing, a blood sample (15 mL) was collected from each pig immediately before infection (T0) and 24 h after infection (T24) into tubes containing 150 μL of 15% EDTA. The blood was centrifuged ($453 \times g$ for 15 min at 4°C) and stored at –20°C following removal of 2 mL of plasma. After 8 d, all pigs were weighed, killed, and necropsied to determine the type and extent of lung pathology.

To identify the two most extreme-performing individuals among the 18 infected pigs, a combination of seven key performance measurements was used, including: 1) percentage of weight gain (**WTP**); 2) calcitonin-receptor expression level (**CTR**); 3) percentage of lung score damage (**LSP**); 4) total clinical score (**TCS**); 5) feed intake on d 2 (**FI2**); 6) feed intake on d 5 (**FI5**); and 7) feed intake on d 7 (**FI7**) after the challenge. The PRINCOMP procedure of SAS (Version 8.2; SAS Inst. Inc., Cary, NC) was used to perform principal components analysis of the seven performance measurements. The most extreme pigs were labeled as resistant (**RES**) and susceptible (**SUS**). Blood samples, taken at time 0 and at 24 h after infection, from these extreme-performing pigs, were used for subsequent microarray experimentation.

Leukocyte RNA was extracted from frozen blood samples and DNAse treated using the Qiagen RNeasy kit (Qiagen, Hilden, Germany), using the manufacturer's instructions. Muscle RNA was extracted from porcine skeletal muscle using TRIzol (Invitrogen, Carlsbad,

CA) and DNAse treated. Amplified RNA was prepared using Ambion's aRNA kit (Ambion, Austin, TX), following manufacturer's instructions.

Microarray Design

The main objective of the gene expression experiment was to develop a focused microarray platform to identify which genes were differentially expressed between the RES and the SUS animals over the 24-h trial period. The array contained 7,776 elements (subsequently referred to as simply *genes*) comprising 6,420 anonymous cDNA clones obtained from 10 subtracted libraries using the PCR-Select cDNA subtraction kit (Clontech, Palo Alto, CA), 68 cDNA clones of known function, and six scorecard controls (Amersham Bioscience, Piscataway, NJ) including elements from other print-facility users. The tissues that were explored for the generation of the subtracted libraries were from spleen, liver, lymph node, lymphocytes, leukocytes, and muscle. The prepared cDNA fragments were arrayed in single spots in duplicated grids at the Institute of Molecular Bioscience (University of Queensland, Brisbane, Australia) on glass slides (Eppendorf, Hamburg, Germany) using a Virtek Microarrayer with the Stealth 48 Pin head and Micro Spotting Pins (Telechem International Inc., Sunnyvale, CA) at a spacing of 220 μm. Each array contained 15,552 spots arranged in 48 blocks of 18 rows × 18 columns.

For the microarray experiments, a control reference design was developed. The availability of reference vs. treatment RNA as well as the cost of the arrays themselves was taken into consideration for the development of the design. In a reference design, an extraneous reference sample is (usually) used along with the RNA of interest. Every sample of interest is compared with this sample in a hybridization. The design is intuitive: every RNA of interest can be compared indirectly because each is compared directly to the reference. Kerr and Churchill (2001b) provide details of and alternatives to the reference design in microarray experiments.

Two control reference samples were chosen using total RNA or amplified RNA from either pure leukocytes (**Lk**) or a 1:2 mixture of muscle tissue and leukocytes (**MLk**). Reference RNA samples were labeled with fluorescent green dye (**G**), using direct incorporation of Cye3-dUTP during the Superscript II reverse-transcribed cDNA synthesis step (Invitrogen, Carlsbad, CA). For each combination of animal (RES and SUS) and time (T0 and T24) two aRNA samples (i.e., technical replicates) were directly labeled with fluorescent red dye (**R**) using direct incorporation of Cye5-dUTP. In total, 16 cDNA microarray slides were used in this study.

The GenePix 4000 optical scanner and the image analysis software GenePixPro 3.0, both from Axon Instruments Inc. (Union City, CA), were used to quantify the expression intensities. This software provides a distinct quality reading for bad-quality spots, which were

subsequently excluded from the analyses. Also, only genes that scored twice within each microarray slide were used in this study. Finally, the accuracy of the resulting R and G signals was assessed from the correlation between the mean and median signal intensities and estimated by dividing the smaller of the mean or median by the larger for both fluorescent channels in a spot (Tran et al., 2002). These editing criteria resulted in 414,936 intensity readings (half R and half G, and each with a foreground and a background component) across the 16 microarray slides and from 6,456 genes out of the original 7,776. For these readings, the R and G intensity levels were background-corrected by subtracting the background (R_b and G_b) from the foreground (R_f and G_f) intensities. Thus, for analyses, $R = R_f - R_b$ and $G = G_f - G_b$.

Measurement of Differential Gene Expression

A base-2 logarithmic ($\log_2$) transformation was applied to the background-corrected intensity levels R and G so that they approximated a normal distribution. In order to reduce the influence of within-slide spot variation without estimating it, the average of the two readings for each gene in each array slide was used as a measure of M_i as follows:

$$M_i = 0.5 \times \left(\log_2 \frac{R_{i1}}{G_{i1}} + \log_2 \frac{R_{i2}}{G_{i2}} \right) = 0.5 \qquad [1]$$

$$\times [\log_2(R_{i1}R_{i2}) - \log_2(G_{i1}G_{i2})]$$

where R_{i1} and G_{i1} are the red and green intensity readings, respectively, for gene i from spot 1, and similarly for R_{i2} and G_{i2} from spot 2.

Location normalization was applied to the ratio values in M_i from Eq. [1] by subtracting the within-slide median. The median was used because it is more robust against outliers than the mean. This within-slide location standardization assumes that most genes, at least a half, will not be differentially expressed. It further assumes that, within each microarray, changes are roughly symmetric around the median and across all intensities.

For each animal (RES and SUS) and reference design (Lk and MLk), a measure of possible differential gene expression from T0 to T24 was obtained by subtracting the median-corrected M ratio values at T0 from those at T24. Two variables, one for each animal type, were computed for each gene and sample combination as follows:

$$RES_{ijkm} = M_{ijk} - M_{ijm} \qquad [2]$$

$$SUS_{ijkm} = M_{ijk} - M_{ijm} \qquad [3]$$

In this notation, subscript i refers to gene ($i = 1$ to 6,456), j refers to control reference design ($j = 1$ and 2 for Lk and MLk, respectively), k refers to sample 1 or 2 evaluated at T24, and m refers to sample 1 or 2 evaluated at T0. A positive (or negative) value in RES_{ijkm} indicates an increase (or decrease) in the expression of gene i from T0 to T24 in the resistant animal. Similarly, a positive (negative) value in SUS_{ijkm} indicates an increase (decrease) in the expression of gene i from T0 to T24 in the susceptible animal.

Mixed-Model Equations and Estimation of (Co)Variance Components

The mixed-model for the bivariate analysis of observations in Eq. [2] and [3] can be written as

$$\begin{bmatrix} y_R \\ y_S \end{bmatrix} = \begin{bmatrix} X_R & 0 \\ 0 & X_S \end{bmatrix} \begin{bmatrix} \beta_R \\ \beta_S \end{bmatrix} + \begin{bmatrix} Z_R & 0 \\ 0 & Z_S \end{bmatrix} \begin{bmatrix} g_R \\ g_S \end{bmatrix} + \begin{bmatrix} e_R \\ e_S \end{bmatrix} \qquad [4]$$

where

y_R (y_S) = vector of observations in RES_{ijkm} (SUS_{ijkm})

X_R (X_S) = incidence matrix relating observations in y_R (y_S) with comparison group fixed effect levels in β_R (β_S)

Z_R (Z_S) = incidence matrix relating observations in y_R (y_S) with gene expression random effect levels in g_R (g_S)

g_R (g_S) = vector of random gene effects in the RES (SUS) individual

e_R (e_S) = vector of residual effects for the observations in RES_{ijkm} (SUS_{ijkm})

In total, there were 49,366 observations of which 44,136 corresponded to observations on gene expression change in both y_R and y_S, 2,596 corresponded to observations on gene expression change in only y_R, and 2,634 corresponded to observations on gene expression change in only y_S.

There were 136 levels for comparison group fixed effects, defined by those gene expressions corresponding to genes from the same plate library (with 17 levels), the same control reference design (with two levels) and the same sample contrast in the T0-with-T24 comparison (two technical replicates by two time points). The 17 levels of plate library originated from 10 subtracted libraries of anonymous cDNA clones, plus the set of 68 cDNA clones of known function, plus the six scorecard controls.

For the model in Eq. [4], the (co)variance matrix for the random effects was assumed to be as follows

$$\text{var} \begin{bmatrix} g_R \\ g_S \\ e_R \\ e_S \end{bmatrix} = \begin{bmatrix} \sigma_{g_R}^2 I_G & \sigma_{g_R g_S} I_G & 0 & 0 \\ \sigma_{g_R g_S} I_G & \sigma_{g_S}^2 I_G & 0 & 0 \\ 0 & 0 & \sigma_{e_R}^2 I_N & \sigma_{e_R e_S} I_N \\ 0 & 0 & \sigma_{e_R e_S} I_N & \sigma_{e_S}^2 I_N \end{bmatrix}$$

$$= \begin{bmatrix} G_{11} & G_{12} & 0 & 0 \\ G_{12} & G_{22} & 0 & 0 \\ 0 & 0 & E_{11} & E_{12} \\ 0 & 0 & E_{12} & E_{22} \end{bmatrix} = \begin{bmatrix} G & 0 \\ 0 & E \end{bmatrix} \quad [5]$$

where

$\sigma^2_{g_R}$ $(\sigma^2_{g_S})$ is the variance of gene expression effects in g_R (g_S)

$\sigma_{g_R g_S}$ is the gene expression covariance between g_R and g_S resulting from those genes being observed in both individuals

$\sigma^2_{e_R}$ $(\sigma^2_{e_S})$ is the variance of residual effects in e_R (e_S)

$\sigma_{e_R e_S}$ is the residual covariance between e_R and e_R resulting from spatial correlation due to proximity on plate

I_G and I_N indicate an identity matrix of dimension equal the number of gene expression levels (i.e., 6,456) and number of observations (i.e., 49,366).

Writing out the equations for each variable in the model separately, the mixed-model equations become

$$\begin{bmatrix} X_R^T E^{11} X_R & X_R^T E^{12} X_S & X_R^T E^{11} Z_R & X_R^T E^{12} Z_S \\ X_R^T E^{12} X_R & X_S^T E^{22} X_S & X_R^T E^{12} Z_R & X_S^T E^{22} X_S \\ Z_R^T E^{11} X_R & Z_R^T E^{12} X_S & Z_R^T E^{11} X_R + G^{11} & Z_R^T E^{11} X_S + G^{12} \\ Z_S^T E^{12} X_R & Z_S^T E^{22} X_S & Z_S^T E^{11} X_R + G^{12} & Z_S^T E^{22} X_S + G^{22} \end{bmatrix}$$

$$\begin{bmatrix} \beta_R \\ \beta_S \\ g_R \\ g_S \end{bmatrix} = \begin{bmatrix} X_R^T E^{11} y_R + X_R^T E^{12} y_S \\ X_S^T E^{12} y_R + X_S^T E^{22} y_S \\ Z_R^T E^{11} y_R + Z_R^T E^{12} y_S \\ Z_S^T E^{12} y_R + Z_S^T E^{22} y_S \end{bmatrix} \quad [6]$$

where G^{ij} (E^{ij}) are elements of G^{-1} (E^{-1}). The total number of equations represented in Eq. [6] is 13,184 coming from twice 136 comparison group fixed effects plus twice 6,456 gene random effects.

Estimates of (co)variance components and solutions to the mixed-model equations were obtained by Method $\Re$ (Reverter et al., 1994) using an acceleration method described by Druet et al. (2001). The multiplicative iterative algorithm to update (co)variances for traits i and j was

$$G^{(n+1)} = \begin{bmatrix} \hat{g}_{i_p}^T \hat{g}_i & \hat{g}_{j_p}^T \hat{g}_i \\ \hat{g}_{i_p}^T \hat{g}_j & \hat{g}_{j_p}^T \hat{g}_j \end{bmatrix} \begin{bmatrix} \hat{g}_{i_p}^T \hat{g}_{j_p} & \hat{g}_{j_p}^T \hat{g}_{i_p} \\ \hat{g}_{i_p}^T \hat{g}_{j_p} & \hat{g}_{j_p}^T \hat{g}_{j_p} \end{bmatrix}^{-1} G^{(n)} \quad [7]$$

where $\hat{g}_i$ and $\hat{g}_j$ represent vectors of estimated solutions from the complete data set, and $\hat{g}_{i_p}$ and $\hat{g}_{j_p}$ represent vectors of estimated solution from a 50% random partial data. At convergence, the G matrix is symmetric, and during iteration symmetry was forced by averaging the off-diagonal elements.

Data from each animal were analyzed 25 times with different 50% partial data subsets selected randomly to obtain BLUP in $\hat{g}_{i_p}$ and $\hat{g}_{j_p}$. The convergence criterion was r_i = 1.0 ± 0.0001, where r_i is the regression for variance component i. Sampling standard errors of the estimates were calculated from the SD of 25 estimates from different subsamples. The ABTK2.0, Animal Breeder's ToolKit, software (Golden et al., 1992) was used both for the estimates of variance components and for the solutions to the mixed-model equations.

Bayesian Model-Based Clustering

Solutions to the mixed-model equations are BLUP of gene effects $\hat{g}_R$ and $\hat{g}_S$ for RES and SUS, respectively. For each gene in i, the difference

$$d_i = \hat{g}_{R_i} - \hat{g}_{S_i} \quad [8]$$

provides a measure to be investigated in order to identify differentially expressed genes in RES relative to SUS. Large positive d_i values are likely to belong to genes whose expression is up-regulated in the RES animal and down-regulated in the SUS animal during the 24-h postinfection period. Similarly, large negative d_i values are likely to belong to down- (or up-)regulated genes in the RES (or SUS) animal in the same period.

Bayesian model-based clustering with a known number of components was approached to isolate differentially expressed genes with d_i values assumed to be independent observations from a mixture density with k (possible unknown but finite) components and with probability density function:

$$f(d; \Phi_k) = \sum_{j=1}^{k} \pi_j \phi(d; \mu_j, V_j) \quad [9]$$

where $\phi(d; \mu_j, V_j)$ denotes the normal density function with mean μ_j and covariance matrix V_j, and the mixing proportions π_j are constrained to be nonnegative and sum to unity. All unknown parameters are represented in Φ_k for a k-component (or k-cluster) mixture model. In the present study, mixture models with up to five components (or clusters) were contemplated.

Following Raftery (1996) the following prior densities were used:

$$V_j^{-1} \propto \Gamma \left(\frac{w_j}{2}, \frac{\lambda_j}{2} \right)$$

$$\mu_j \propto N \left(\xi_j, \frac{\nu_j}{\tau_j} \right) \quad [10]$$

$$\pi \propto \mathrm{Dir} \left(\alpha_1, ..., \alpha_k \right)$$

where $\Gamma(a, b)$ denotes a gamma density with mean a/b and variance a/b^2; $N(a, b)$ denotes a normal distribution with mean a and variance b; $\pi = (\pi_1, ..., \pi_k)$; and $\mathrm{Dir}(\alpha_1,..., \alpha_k)$ denotes the Dirichlet distribution with parameter α.

Following Richardson and Green (1997), prior hyperparameters were chosen to be data-dependent-constant so that the prior distribution was relatively flat over the range of values that could be expected. Finally, the Gibbs sampler proceeded by sampling successively from the following conditional distributions:

$$p(u_i = j|\pi,k,\theta,d) \propto \pi_j\, N(\mu_j, V_j)$$

$$p(\mu_j|\pi,k,u,\mu_{-j},V_j,d) \propto N \times$$

$$\left[(n_j V_j^{-1} + k)^{-1}\, (n_j V_j^{-1}\, \overline{d}_j + k\, \xi_j),\, (n_j V_j^{-1} + k)^{-1}\right]$$

$$p(V_j|\pi,k,u,\mu,d) \propto \chi^{-2}\left(2\alpha + n_j, \left[2\nu_j + \sum_{i:d_i=j}(d_i - \mu_j)^2\right]^{-1}\right)$$

$$p(\pi|k,u,\theta,d) \propto \mathrm{Dir}(t + n_1,...,t + n_k)$$

where the latent data $u = (u_1, ..., u_n)$ are indicative of the mixture component from which d_i were generated. The Gibbs sampler was run until a Markov chain of length 12,000 was generated, the first 2,000 (burn-in) were discarded, and averages from the remaining 10,000 samples were used to obtain point estimates for parameters in Eq. [9].

Up to five components (clusters) were explored for the model in Eq. [9]. Criteria for model selection include a combination of the likelihood evaluation (log$L$) as well as the Akaike information criterion (Akaike, 1969) and the Bayesian information criterion (Schwartz, 1978).

Once the mixture model of choice has been identified, the probability of each data point belonging to each cluster was given by the posterior probability in

$$\tau_{ij}^{(m)} = \frac{\pi_i^{(m)}\, \phi(d_i;\, \mu_j^{(m)},\, V_j^{(m)})}{f(d_i;\, \Phi^{(m)})} \qquad [11]$$

Further, a data point in d_i (and thus the ith gene) was classified to a given cluster if its posterior probability was the largest.

Software Availability

Software named BAYESMIX was developed using FORTRAN90 to perform Bayesian model-based clustering via mixtures of normal distributions and is available from the corresponding author on request and for non-commercial use only. Further details of the software are given in Reverter et al. (2003b).

Results

Table 1 provides summary statistics of key performance traits and principal components analysis results for the 18 infected pigs involved in this study. The first principal component explained 53% of the total variation and had a clear biological interpretation with positive weights for favorable traits (weight gain and feed intakes) and negative weights for unfavorable (i.e., clinical) traits. Therefore, this first principal component was used to rank individuals in terms of susceptibility to *Actinobacillus pleuropneumoniae*.

The ability of the first principal component to rank animals is further illustrated in Figure 1, in which the performance and clinical data for the 18 pigs are represented by plotting the WTP over LSP. The median of each measure was inserted to divide the graph into four phenotypic quadrants, namely (clockwise from top left): resistant, resilient, susceptible, and reactive. The two most extreme pigs from the resistant (RES pig with LSP = 0.8% and WTP = 34.7%) and the susceptible (SUS pig with LSP = 39.7% and WTP = 12.8%) quadrants were selected for the subsequent microarray experiment.

Table 2 presents details of the experimental design and summary statistics for the 16 microarray slides used in this study. After data edits, the total number of unique genes across the 16 slides was 6,456. Within slide, the number of genes with valid readings ranged from 5,540 to 6,426. Spot accuracy as measured by correlation between mean and median signal intensities was above 0.85 in 14 and 12 arrays for the red and green channels, respectively. Tran et al. (2002) suggest that a correlation of 0.85 or higher not only retains more data than other methods, but retained data are more accurate than traditional thresholds or common spot flagging algorithms.

Reference sample and plate of origin were the two major sources of systematic variation, although combined they accounted for only 11% and 12% of total variation in RES_{ijkm} and SUS_{ijkm}, respectively. Small percentages of total variation accounted for by design effects are not uncommon in microarray studies. For example, Kerr and Churchill (2001a) report a combined effect of array, dye, and array × dye interaction that accounted for 8.8% of total variation, and Kerr et al. (2002) report that the combined effect of spot, array, dye, and array × dye interaction accounted for 5.9% of total variation.

Table 3 presents the number of observations and least square means for each level of reference sample and plate of origin as obtained using the GLM procedures of SAS (SAS Inst. Inc., Cary, NC). The 46,732 observations in y_R averaged 0.129 and ranged from −4.033 to 5.017 with a SD of 0.843. The 46,770 observations in y_S averaged 0.028 and ranged from −4.791 to 7.633 with a SD of 1.008.

Table 4 presents Method $\mathfrak{R}$ estimates of (co)variance components. The number of iterations to achieve convergence averaged 31 and ranged from 28 to 45. In this situation, with few levels of fixed effects relative to the large amount of observations in each, and assuming an identity matrix to describe dispersion in random gene

Table 1. Summary statistics of key performance traits and principal components analysis (PCA) results for the 18 infected pigs involved in this study

Trait[a]	Mean	SD	Minimum	Maximum	PCA[b]
WTP, %	21.5	10.1	−2.5	37.5	0.37
CTR, units	137,139	116,162	0	420,099	−0.20
LSP, %	9.2	9.7	0.0	39.7	−0.36
TCS, units	40.2	7.7	22.0	51.0	−0.31
FI2, g	218	345	0	1,075	0.38
FI5, g	1,026	595	49	2,000	0.50
FI7, g	1,853	493	1,140	2,800	0.45

[a]WTP = percentage of weight gain; CTR = calcitonin-receptor expression level; LSP = percentage of lung score damage; TCS = total clinical score; FI2 = feed intake (as-fed basis) on d 2 after infection; FI5 = feed intake (as-fed basis) on d 5 after infection; FI7 = feed intake (as-fed basis) on d 7 after infection.

[b]Coefficients for the first principal component explaining 53% of the total variation and used to rank individuals in terms of susceptibility to *Actinobacillus pleuropneumoniae*.

expression effects, Method $\mathfrak{R}$ was found to be particularly stable (judged by sampling variation, Table 4) as compared with applications in genetic parameter estimation studies (see, for example, Cantet et al., 2000) and likely to be free of bias as judged by simulations studies of Druet et al. (2001) and Duangjinda et al. (2001).

Larger variation both at genetic and at residual levels was found for gene expression intensities from the sus-

ceptible animal as compared with the resistant animal. The amounts of total variation accounted for by g_R and g_S were 0.461 ± 0.013 and 0.581 ± 0.015, respectively. Correlation estimates between g_R and g_S, and between e_R and e_S were 0.027 ± 0.026 and 0.256 ± 0.009, respectively.

Mean, SD, minimum and maximum values for resulting BLUP differences in d_i from Eq. [8] were 0.000, 0.793, −5.214 and 3.378, respectively. Fitting of mix-
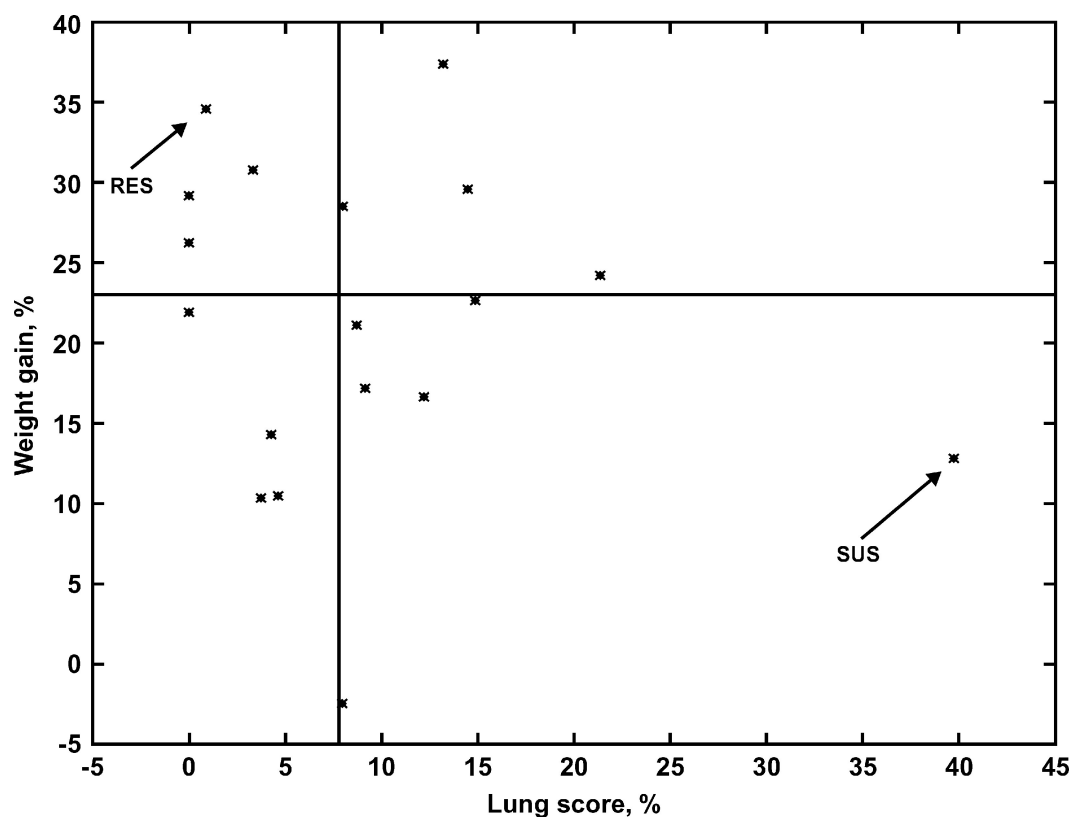
Figure 1. Scatter plot of the percentage of weight gain over the percentage of lung score damage for the 18 *Actinobacillus pleuropneumoniae*-infected pigs. Arrows indicate the coordinates for the two most extreme pigs, resistant (RES) and susceptible (SUS). Horizontal and vertical lines correspond to the median weight gain and lung score, respectively.

Table 2. Design details and summary statistics for the 16 microarrays used in this study

Array	Ref.	Animal	Time	Sample	Spots	Genes	Red	Green	Median
		Design details[a]			Summary statistics[b]		Correlation		
1	Lk	RES	T0	1	13,526	6,351	0.95	0.95	−0.09
2	Lk	RES	T0	2	13,340	6,261	0.95	0.95	−0.02
3	Lk	RES	T24	1	13,701	6,426	0.95	0.95	−0.13
4	Lk	RES	T24	2	12,936	6,078	0.92	0.92	−0.12
5	Lk	SUS	T0	1	13,556	6,363	0.95	0.96	−0.03
6	Lk	SUS	T0	2	13,367	6,274	0.92	0.92	0.14
7	Lk	SUS	T24	1	13,527	6,346	0.95	0.96	0.02
8	Lk	SUS	T24	2	12,006	5,664	0.90	0.89	−0.13
9	MLk	RES	T0	1	12,876	6,085	0.89	0.87	−0.21
10	MLk	RES	T0	2	11,662	5,540	0.89	0.80	−0.10
11	MLk	RES	T24	1	12,879	6,087	0.88	0.87	−0.24
12	MLk	RES	T24	2	12,610	5,908	0.80	0.69	−0.10
13	MLk	SUS	T0	1	12,880	6,073	0.91	0.89	−0.27
14	MLk	SUS	T0	2	12,288	5,740	0.90	0.84	0.29
15	MLk	SUS	T24	1	13,016	6,135	0.89	0.87	−0.05
16	MLk	SUS	T24	2	13,298	6,227	0.87	0.83	0.14

[a]Ref. (reference); Lk = leukocyte RNA; MLk = muscle and leukocyte RNA; Animal: RES = resistant pig, SUS = susceptible pig; time: T0 = 0 h, T24 = 24 h after disease challenge; sample = RNA replicates (Note: within animal × time cell, Samples 1 and 2 were cross-classified across reference).
[b]Spots = number of valid spots per array; genes = number of unique elements per array; correlation = correlation between the mean and median signal intensities by red and green channels; median = median of base-2 log ratio of red to green intensities.

tures of normal distributions to the quantities in d_i revealed the following three-component (or cluster) model as the one of choice:

$$f(y;\hat{\Phi}) = 0.02 \times N(-3.05, 0.84) + 0.95 \times N(0.00, 0.33) + 0.03 \times N(2.14, 0.20)$$

The majority of records fell into the middle cluster with zero mean and intermediate variance and corresponded to genes not differentially expressed between the RES and the SUS animals. The first cluster corresponded to genes whose expression from T0 to T24 was decreased in the RES and increased in the SUS. There

Table 3. Number of records and least squares means (LSM) for gene expression change from 0 to 24 h postchallenge for the effects of reference sample and plate of origin and for the resistant and susceptible *Actinobacillus pleuropneumoniae*–infected pigs

Effect	Level	No.	LSM	No.	LSM
		Resistant		Susceptible	
Reference[a]	Lk	24,558	0.112 ± 0.005	23,672	−0.003 ± 0.006
	MLk	22,174	0.145 ± 0.005	23,098	0.052 ± 0.006
Plate[b]	1	2,732	0.054 ± 0.015	2,660	0.007 ± 0.018
	2	2,576	0.190 ± 0.016	2,624	0.195 ± 0.018
	3	2,629	0.152 ± 0.015	2,650	0.327 ± 0.018
	4	2,861	−0.049 ± 0.015	2,913	0.754 ± 0.017
	5	2,738	0.015 ± 0.015	2,769	0.410 ± 0.018
	6	2,659	0.165 ± 0.015	2,552	−0.015 ± 0.019
	7	2,799	0.118 ± 0.015	2,841	0.292 ± 0.018
	8	2,875	0.124 ± 0.015	2,956	0.202 ± 0.017
	9	2,493	−0.073 ± 0.016	2,566	−0.029 ± 0.019
	10	2,577	−0.106 ± 0.016	2,631	−0.046 ± 0.018
	11	2,963	0.139 ± 0.015	2,911	−0.074 ± 0.017
	12	2,949	0.133 ± 0.015	2,835	−0.174 ± 0.018
	13	2,866	0.208 ± 0.015	2,707	−0.335 ± 0.018
	14	2,770	0.969 ± 0.015	2,855	0.235 ± 0.018
	15	2,828	0.562 ± 0.015	2,881	−0.083 ± 0.018
	16	3,052	−0.324 ± 0.014	3,041	−0.693 ± 0.017
	17	2,365	−0.095 ± 0.016	2,378	−0.550 ± 0.019

[a]Lk = leukocyte RNA; MLk = muscle and leukocyte RNA.
[b]The 17 levels of plate originated from 10 subtracted libraries of anonymous cDNA clones, plus a set of 68 cDNA clones of known function, plus six scorecard controls.

Moser et al.

Table 4. Method $\mathfrak{R}$ estimates of (co)variance components.

Component[a]	Mean	SD	Minimum	Maximum
$\sigma^2_{g_{\mathrm{R}}}$	0.250	0.011	0.228	0.273
$\sigma^2_{g_{\mathrm{S}}}$	0.473	0.023	0.436	0.535
$\sigma_{g_{\mathrm{R}}g_{\mathrm{S}}}$	0.009	0.009	−0.007	0.025
$\sigma^2_{e_{\mathrm{R}}}$	0.292	0.005	0.283	0.302
$\sigma^2_{e_{\mathrm{S}}}$	0.340	0.005	0.332	0.350
$\sigma_{e_{\mathrm{R}}e_{\mathrm{S}}}$	0.081	0.003	0.074	0.086

[a]$\sigma^2_{g_{\mathrm{R}}}$ ($\sigma^2_{g_{\mathrm{S}}}$) is the genetic variance of gene expression change from 0 to 24 h in the resistant (susceptible) animal; $\sigma_{g_{\mathrm{R}}g_{\mathrm{S}}}$ is the genetic covariance of gene expression change from 0 to 24 h between the resistant and the susceptible animal; $\sigma^2_{e_{\mathrm{R}}}$ ($\sigma^2_{e_{\mathrm{S}}}$) is the residual variance of gene expression change from 0 to 24 h in the resistant (susceptible) animal; and $\sigma_{e_{\mathrm{R}}e_{\mathrm{S}}}$ is the residual covariance of gene expression change from 0 to 24 h between the resistant and the susceptible animal.

were 128 genes classified into this first cluster. The majority of these genes originated from the lymph node vs. liver tissue subtractive library. These 128 genes presented a d_i value that was ≤−1.851 (or −2.33 SD). The last cluster corresponded to 179 genes whose expression from T0 to T24 was increased in the RES and decreased in the SUS. The majority of these candidates resulted from the RNA of leukocytes of *A. pleuropneumoniae*–challenged pigs, where the leukocytes 24 h postchallenge were subtracted against leukocytes before the challenge. The d_i observed for these 179 differentially expressed genes was ≥1.626 (or 2.05 SD). Figure 2 illustrates the scatter plot of $\hat{g}_S$ vs. $\hat{g}_R$ for the 6,456 genes and where the ones identified as differentially expressed are highlighted. Note the presence of 11 genes classified into Cluster 1 although their expression was up-regulated in both RES and SUS pigs. Similarly, there were eight genes classified into Cluster 3 even though their expression was down-regulated in both pigs RES and SUS.

Discussion

Current management of *A. pleuropneumoniae* infections includes vaccination and use of antimicrobial growth promotants. However, these measures only slightly confine the commercial consequences of the pathogen. Although the continuous use of antibiotics is expensive and cannot prevent the establishment of *A. pleuropneumoniae* in the lungs, it is also speculated to increase the emergence of antibiotic resistance. Vaccination has only a marginally protective effect because several serotypes are prevalent in various countries and even vaccinated pigs can remain carriers of the pathogen in a subclinical or chronic state (Wongnarkpet et al., 1999; Chiers et al., 2002). By learning more about the pig immune system and modulating its responses, antibiotics and chemicals currently used to control disease may be reduced or replaced, with the added benefit of improving health and increasing disease resistance.

In the present study, not all infected pigs showed the same degree of symptomatic response to the *A. pleurop-*

neumoniae challenge. This is indicative of what happens during an outbreak in a commercial piggery. Pigs displaying extreme phenotypes measured in terms of performance (either high or low performers) after being challenged with *A. pleuropneumoniae* are obvious targets for a global functional genomic approach to reveal genes responsible for the degree of susceptibility to infection. In keeping with studies reported elsewhere (Balaji et al., 2002), pigs used in this experiment displayed anorexia, along with other relevant clinical signs (Table 1), within a few hours after *A. pleuropneumoniae* challenge. There is also considerable evidence for enhanced production of mRNA for inflammatory mediators in lung tissue of pigs infected with *A. pleuropneumoniae* (Baarsch et al., 2000). Therefore, it was anticipated that measuring changes in gene expression levels in peripheral blood leukocytes at 24 h after the challenge might distinguish genes involved in the pathogenesis of *A. pleuropneumoniae* pneumonia.

There is currently no consensus about how to best eliminate sources of error in the intensity readings of microarray experiments. Two commonly discussed approaches are normalization adjustment of data before statistical analysis (Yang et al., 2002) and adjusting for sources of bias and confounding with a linear model (Kerr and Churchill, 2001a; Wolfinger et al., 2001). Fitting fixed effects aims at normalizing gene expression data by accounting for systematic sources of variation, such as array, dye, variety, and samples. In a complex experiment in which the RNA samples have a nontrivial experiment design structure, it may be desirable to include both fixed and random terms in the treatment × gene interaction component of the model. Simultaneously solving for fixed and random effects as in the mixed model of Eq. [4] adjusts for the possibility of different expression intensities of a given gene across levels of comparison group as well as variability both across and within genes. In this situation, the recently reported two-stage approach of Dobbin and Simon (2002) would be most inadequate as it would ignore fixed by random off-diagonal elements of the information matrix in the left-hand side of Eq. [6]. Mixed models
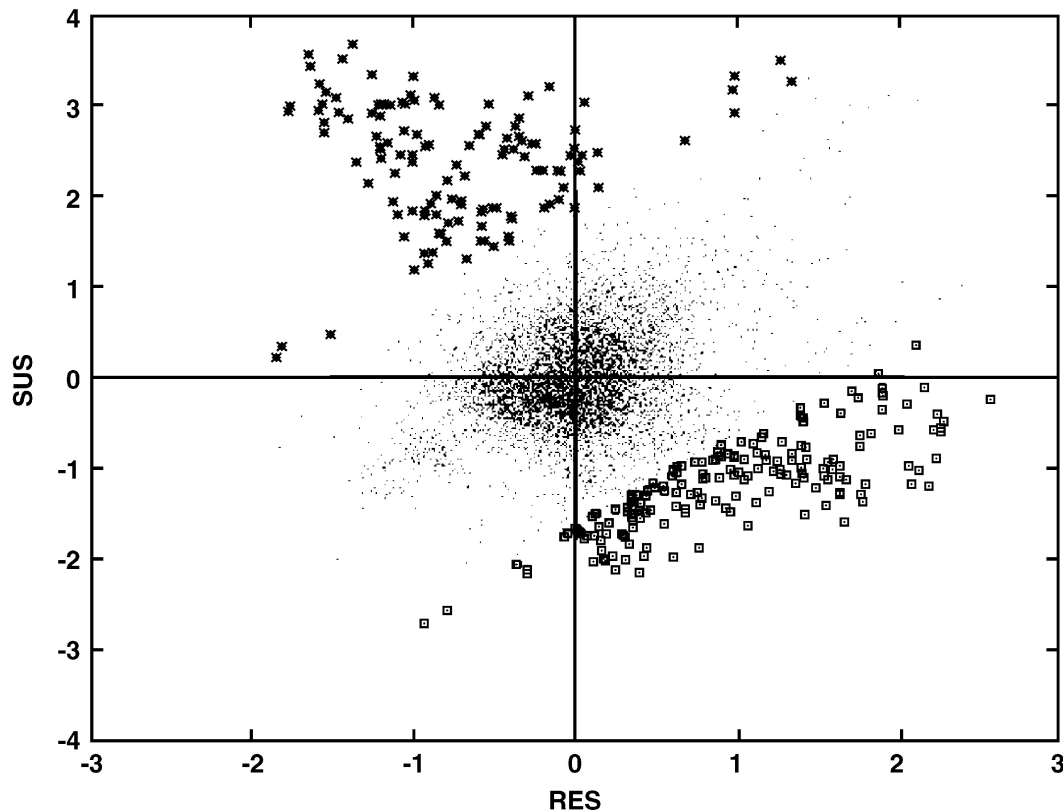
Figure 2. Scatter plot of BLUP of gene expression change from 0 to 24 h post-challenge for the resistant (RES) and susceptible (SUS) *Actinobacillus pleuropneumoniae*-infected pigs. Highlighted spots indicate differentially expressed genes.

can admit more general covariance structure and will provide shrinkage of estimated effects that can reduce bias. The optimality of mixed models to assess gene significance from cDNA microarray expression data was previously reported by Wolfinger et al. (2001).

The heterogeneity observed at the gene expression level between RES and SUS is not unexpected in microarray gene expression analysis given the wide variety of known gene functions. Gene expression heterogeneity has been documented in other studies, such as Golub et al. (1999) and Dudoit et al. (2002). From a biological viewpoint, the lack of correlation between g_R and g_S observed in this study is encouraging and indicates that the set of genes responsible for a given change in genetic expression (i.e., either up- or down-regulation) between T0 and T24 in the RES animal is not related to the set of genes responsible for the same type of change in genetic expression in the SUS animal and in the same time period. This indicates that distinct sets of genes may be induced by infection in the resistant and susceptible phenotypes.

Although observations on RES and SUS individuals were not compared directly in the same microarray, but through the same reference sample, an estimate of residual correlation was obtained and attributed to spatial correlation due to proximity on slide array. This positive residual correlation estimated at 0.256 ± 0.009

indicates that the nongenetic factors (i.e., array, block, plate of origin) responsible for an increase in gene expression in RES are also related to those responsible for an increase in SUS and vice versa.

Model-based clustering via mixture of distributions has been identified as a method of choice for identifying which genes have differential expression levels. In livestock species, this technique was recently used to identify differentially expressed genes in bovine muscle tissue in response to varying levels of energy and protein in the diet (Reverter et al., 2003a). Model-based clustering clearly defines cluster subpopulations with a certain distribution, and clustering results are stable. Several statistical methods can be applied to estimate the number of clusters (Yeung et al., 2001; Pan et al., 2002b). In addition, model-based clustering has been shown to provide an elegant framework to calculate the power of detecting a specified magnitude of change (Rekaya, 2002) as well as to estimate the number of replicates needed for precise inferences (Pan et al., 2002a). In all the above-mentioned studies, model-based clustering was applied directly to intensity ratios or to linear combinations of intensity ratios (in the form of a *t*-statistic). A novel approach was described by Allison et al. (2002), exploiting, under the null hypothesis, the uniform distribution of *P*-values on the interval [0, 1] regardless of the statistical test used. The authors then proposed

Table 5. Origin of 307 differentially expressed (DEXP) elements and proportion (%) compared with size of respective subtracted cDNA library of porcine descent

cDNA library	cDNA subtraction direction[a]	Library size	DEXP elements (profile)[b]		%
			Rup-Sdown	Rdown-Sup	
SPL	<u>Spleen</u> vs. liver tissue	576	6	—	1.04
LIVa	<u>Liver</u> vs. spleen tissue	480	2	—	0.4
LN	<u>Lymph node</u> vs. liver tissue	672	16	109	18.6
PLC-F[c]c	<u>Activated</u>[d] vs. non-activated lymphocytes	768	5	—	0.7
PLC-R	<u>Nonactivated</u> vs. activated lymphocytes	768	5	—	0.7
PLMU-F	<u>Activated lymphocytes</u> vs. muscle tissue	576	27	4	5.4
PLMU-R	<u>Muscle tissue</u> vs. activated lymphocytes	576	23	2	4.3
WBC-F	<u>Blood leukocytes after</u>[e] vs. blood leukocytes before challenge	714	91	13	14.6
WBC-R	<u>Blood leukocytes before</u> vs. blood leukocytes after challenge	714	4	—	0.6

[a]cDNA pool being selected for in the subtraction process is underlined.

[b]Rup-Sdown = up-regulated in the resistant and down-regulated in the susceptible animal; Rdown-Sup = down-regulated in the resistant and up-regulated in the susceptible animal.

[c]F = forward; R = reverse.

[d]Total RNA and subsequently mRNA was extracted from proliferated lymphocytes after 24-h activation with the lectin concanavalin A (conA).

[e]Total RNA (mRNA) was extracted from white blood cells (buffy coat) from 10 pigs before and 24 h after being challenged with *Actinobacillus pleuropneumoniae*.

the fitting of a mixture of beta distributions to these *P*-values to identify the gene comparisons of interest. To the best of our knowledge, the present study is the first report in which fitting a mixture of distributions is applied to the BLUP resulting from the fitting of mixed-model equations. An alternative and possibly preferable option could be to integrate the cluster analysis with the mixed-model analysis as recently described by Gianola et al. (2004) within the context of inferring susceptibility to mastitis in dairy cattle.

The statistical approach used in our study resulted in a list of differentially expressed genes that originated mostly from two distinct subtracted cDNA libraries (Table 5) adding biological meaning to the cluster results from the mixture analysis. For instance, 109 out of the 128 differentially expressed genes downregulated in RES and upregulated in SUS are part of a lymph node vs. liver cDNA library and 18.6% of clones from this library (with 672 clones) were identified as being differentially expressed. In contrast, the majority of differentially expressed genes (91 out of 179) that were upregulated in RES and downregulated in SUS were derived from an immune cell subtractive library in which pig blood leukocytes 24 h after challenge were subtracted against blood leukocytes before challenge (WBC-F in Table 5) and 14.6% of clones from the WBC-F library (with 714 clones) were identified as being differentially expressed. Additionally, another high proportion of differentially expressed genes showing this profile (i.e., upregulation in RES and downregulation in SUS) were identified as being part of immune vs. nonimmune tissue cDNA libraries generated by subtracting cDNA of activated lymphocytes with muscle tissue cDNA and vice versa (in Table 5, PLMU-F and PLMU-R).

Model-based clustering via mixture of distributions resulted in a list of putative differentially expressed

genes with a clear bias toward immune-relevant genes, as determined by the nature of the libraries they originated from. The resulting list of differentially expressed genes seems to reflect the contrasting effect of the disease challenge and the appearance of the extreme pig phenotypes in the trial. Furthermore, the size of this list (307 genes) could provide evidence of the polygenic nature of some immune responses and also the difficulty of selecting animals for immune response using genetic markers.

We are confident that the experimental approach and analytical methods presented here are likely to result in biologically relevant data. However, the fact that only the two most extreme-performing pigs (one RES and one SUS) were included in this analysis implies that caution must be taken when interpreting the results. Some of the differentially expressed genes identified in this work could be responses to things affecting the pig's life other than infection. Our intention is to direct further research to a new microarray experiment incorporating additional biological replicates and a superior design. This extended analysis will be followed by nucleotide sequencing of differentially expressed genes and the development of independent tests to confirm the expression patterns characterized by the microarray study.

Implications

The present study explores gene expression in peripheral blood leukocytes of two pigs showing extreme performance in their response to infection with *Actinobacillus pleuropneumoniae*. The statistical approach to identify differentially expressed genes is based on mixed-model methodology to analyze the gene expression levels combined with Bayesian model-based clustering of the solutions for the gene × immunological status inter-

action. The resulting list of differentially expressed elements is biased toward immune-relevant genes and seems to reflect the contrasting effect of the disease challenge and the appearance of the extreme pig phenotypes in the trial.

Literature Cited

Akaike, H. 1969. Fitting autoregressive models for prediction. Ann. Inst. Stat. Math. 21:243–247.

Allison, D. B., G. L. Gadbury, M. Heo, J. R. Fernández, C-K. Lee, T. A. Prolla, and R. Weindruch. 2002. A mixture model approach for the analysis of microarray gene expression data. Comp. Stat. Data Anal. 39:1–20.

Baarsch, M. J., D. L. Foss, and M. P. Murtaugh. 2000. Pathophysiologic correlates of acute porcine pleuropneumonia. Am. J. Vet. Res. 61:684–690.

Balaji, B., K. J. Wright, J. L. Turner, C. M. Hill, S. S. Dritz, B. Fenwick, J. A. Carroll, M. E. Zannelli, L. A. Beausang, and J. E. Minton. 2002. Circulating cortisol, tumor necrosis factor-α interleukin-1β, and interferon-γ in pigs infected with *Actinobacillus pleuropneumoniae*. J. Anim. Sci. 80:202–207.

Blackall, P. 2001. Pleuropneumonia in pigs. Agency for Food and Fibre Sciences. Department of Primary Industries, Queensland Government. Available: http://www.dpi.qld.gov.au/pigs/4463.html. Accessed Jan. 13, 2004.

Cantet, R. J. C., A. N. Birchmeier, M. G. Santos-Cristal, and V. S. de Avila. 2000. Comparison of restricted maximum likelihood and Method $\Re$ for estimating heritability and predicting breeding value under selection. J. Anim. Sci.78:2554–2560.

Chiers, K., E. Donné, I. Van Overbeke, R. Ducatelle, and F. Haesebrouck. 2002. *Actinobacillus pleuropneumoniae* infections in closed swine herds: Infection patterns and serological profiles. Vet. Microbiol. 85:343–352.

Cui, X., and G. A. Churchill. 2003. Statistical tests from differential expression in cDNA microarray experiments. Genome Biol. 4:210. Available: http://genomebiology.com/2003/4/4/210. Accessed Jan. 13, 2004.

Dobbin, K., and R. Simon. 2002. Comparison of microarray designs for class comparison and class discovery. Bioinformatics (Oxf.) 18:1438–1445.

Druet, T., I. Misztal, M. Duangjinda, A. Reverter, and N. Gengler. 2001. Estimation of genetic covariances with Method $\Re$. J. Anim. Sci. 79:605–615.

Duangjinda, M., I. Misztal, J. K. Bertrand, and S. Tsuruta. 2001. The empirical bias of estimates by restricted maximum likelihood, Bayesian method, and Method $\Re$ under selection for additive, maternal, and dominance models. J. Anim. Sci. 79:2991–2996.

Dudoit, S., Y. H. Yang, T. P. Speed, and M. J. Callow. 2002. Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments. Statistica Sinica 12:111–139.

Gianola, D., J. Ødegård, B. Heringstad, G. Klemetsdal, D. Sorensen, P. Madsen, J. Jensen, and J. Detilleux. 2004. Mixture model for inferring susceptibility to mastitis in dairy cattle: a procedure for likelihood-based inference. Genet. Sel. Evol. (Paris) 36:3–27.

Golden, B. L., W. M. Snelling, and C. H. Mallinckrodt. 1992. Animal breeder's tool kit user's guide and reference manual. Tech. Bull. No. LTB92-2, Colorado State Univ. Agric. Exp. Stn., Ft. Collins.

Golub, T. R., D. K. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J. P. Mesirov, H. Coller, M. L. Loh, J. R. Downing, M. A. Caligiuri, C. D. Bloomfield, and E. S. Lander. 1999. Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring. Science 286:531–537.

Hoflack, G., D. Maes, B. Mateusen, M. Verdonck, and A. De Kruif. 2001. Efficacy of Tilmicosin Phosphate (Pulmotil Premix) in feed for the treatment of a clinical outbreak of *Actinobacillus pleuro-*

pneumoniae infection in growing-finishing pigs. J. Vet. Med. B 48:655–664.

Kato-Maeda, M., Q. Gao, and P. M. Small. 2001. Microarray analysis of pathogens and their interaction with hosts. Cell. Microbiol. 3:713–719.

Kerr, M. K., C. Afshari, L. Bennett, P. Bushel, J. Martinez, N. Walker, and G. A. Churchill. 2002. Statistical analysis of gene expression microarray experiments with replication. Statistica Sinica 12:203–218.

Kerr, M. K., and G. A. Churchill. 2001a. Bootstrapping cluster analysis: Assessing the reliability of conclusions from microarray experiments. Proc. Natl. Acad. Sci. 98:8961–8965.

Kerr, M. K., and G. A. Churchill. 2001b. Experimental design for gene expression microarrays. Biostatistics 2:183–201.

Pan, W., J. Lin, and C. T. Le. 2002a. How many replicates of array are required to detect gene expression changes in microarray experiments? A mixture model approach. Genome Biol. 3:research0022.1–0022.10.

Pan, W., J. Lin, and C. T. Le. 2002b. Model-based cluster analysis of microarray gene-expression data. Genome Biol. 3:research0009.1–0009.8.

Raftery, A. E. 1996. Hypothesis testing and model selection. Pages 163–188 in Markov Chain Monte Carlo in Practice. 1st ed. W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, ed. Chapman & Hall, Suffolk, U.K.

Rekaya, R. 2002. Bayesian mixture models with unknown number of components: Application to power calculation is microarray experiments. Proc. 7th World Cong. Genet. Appl. Livest. Prod., Montpellier, France, CD-ROM Communication No. 16–12.

Reverter, A., K. A. Byrne, H. L. Bruce, Y. H. Wang, B. P. Dalrymple, and S. A. Lehnert. 2003a. A mixture model-based cluster analysis of cDNA microarray gene expression data on Brahman and Brahman composite steers fed high-, medium-, and low-quality diets. J. Anim. Sci. 81:1900–1910.

Reverter, A., K. A. Byrne, and B. P. Dalrymple. 2003b. BAYESMIX: A software program for Bayesian analysis of mixture models with an application to the analysis of microarray gene expression data. XV Proc. Assoc. Adv. Anim. Breed. Genet., Melbourne, VIC, Australia, 15:90–93.

Reverter, A., B. L. Golden, R. M. Bourdon, and J. S. Brinks. 1994. Method $\Re$ variance components procedure: Application on the simple breeding value model. J. Anim. Sci. 72:2247–2253.

Richardson, S., and P. J. Green. 1997. On Bayesian analysis of mixtures with an unknown number of components. J. Royal Stat. Soc. Series B 59:731–792.

Schwartz, G. 1978. Estimating the dimensions of a model. Ann. Stat. 6:461–464.

Taylor, R. 1999. *Actinobacillus pleuropneumoniae*. Pages 343–354 in Diseases of Swine. 8th ed. B. Straw, S. D'Allaire, W. Mengeling, and D. Taylor, ed. Iowa State University Press, Ames.

Tran, P. H., D. A. Peiffer, Y. Shin, L. M. Meek, J. P. Brody, and K. W. Y. Cho. 2002. Microarray optimisations: Increasing spot accuracy and automated identification of true microarray signals. Nucleic Acids Res. 30:e54.

Wolfinger, R. D., G. Gibson, E. D. Wolfinger, L. Bennett, H. Hamadeh, R. Bushel, C. Afshari, and R. S. Paules. 2001. Assessing gene significance from cDNA microarray expression data via mixed models. J. Comput. Bio. 8:625-637.

Wongnarkpet, S., D. U. Pfeiffer, R. S. Morris, and S. G. Fenwick. 1999. An on-farm study of the epidemiology of *Actinobacillus pleuropneumoniae* infection in pigs as part of a vaccine efficacy trial. Prev. Vet. Med. 39:1–11.

Yang, Y. H., S. Dudoit, P. Luu, D. M. Lin, V. Peng, J. Ngai, and T. P. Speed. 2002. Normalization for cDNA microarray data: A robust composite method addressing single and multiple slide systematic variation. Nucleic Acids Res. 30:e15.

Yeung, K. Y., C. Fraley, A. Murua, A. E. Raftery, and W. L. Ruzzo. 2001. Model-based clustering and data transformations for gene expression data. Bioinformatics 17:977–987.

References	This article cites 28 articles, 13 of which you can access for free at:
	http://www.journalofanimalscience.org/content/82/5/1261#BIBL
Citations	This article has been cited by 6 HighWire-hosted articles:
	http://www.journalofanimalscience.org/content/82/5/1261#otherarticles